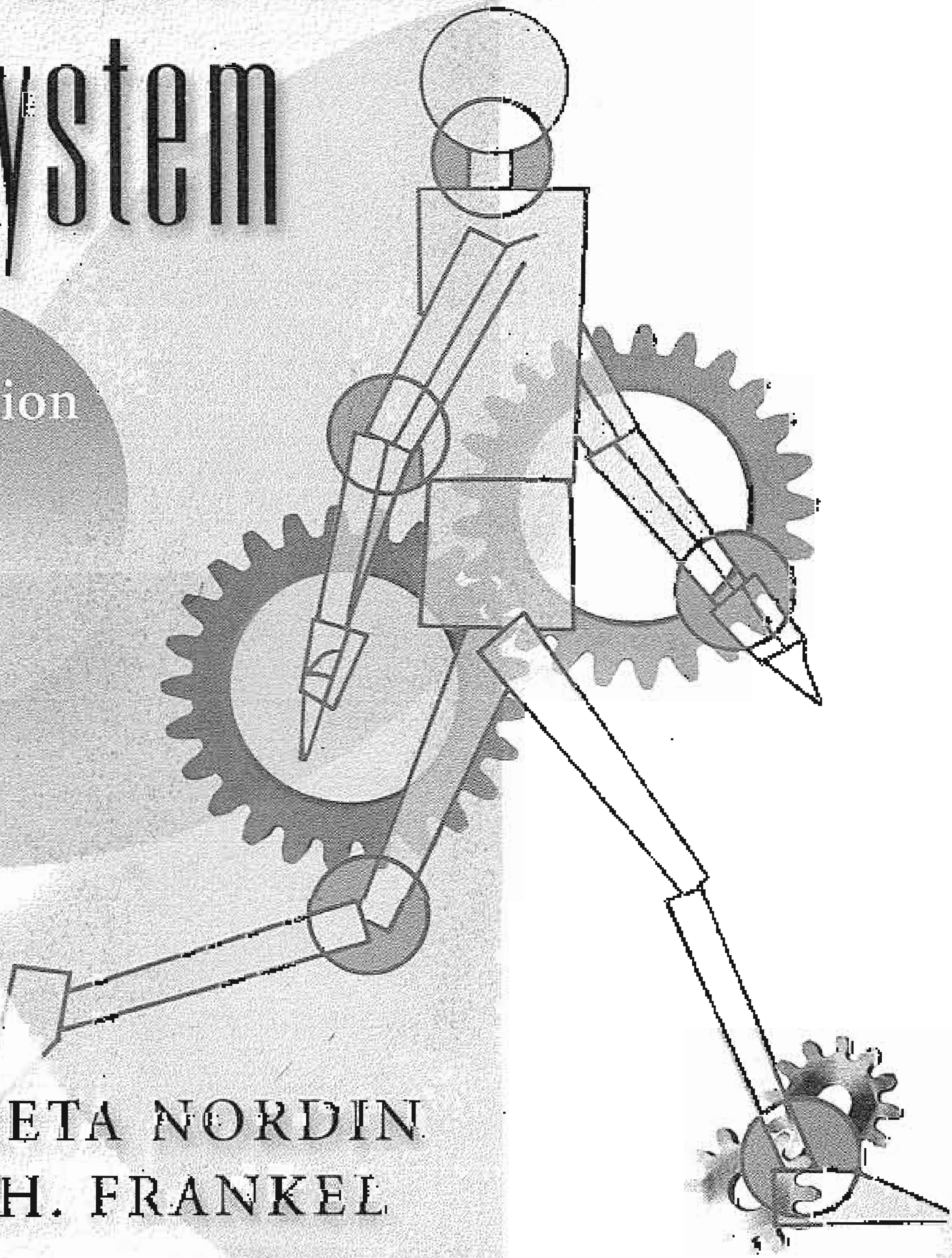
Basic Biomechanics of the Musculoskeletal System

third edition



MARGARETA NORDIN
VICTOR H. FRANKEL

EDITION

BASIC BIOMECHANICS *of the* MUSCULOSKELETAL SYSTEM

Margareta Nordin, P.T., Dr. Sci.
Director, Occupational and Industrial Orthopaedic Center (OIOC)
Hospital for Joint Diseases Orthopaedic Institute
Mt. Sinai NYU Health
Program of Ergonomics and Biomechanics, New York University
Research Professor
Department of Orthopaedics and Environmental Health Science
School of Medicine, New York University
New York, New York

Victor H. Frankel, M.D., Ph.D., KNO
President Emeritus
Hospital for Joint Diseases Orthopaedic Institute
Professor of Orthopaedic Surgery
New York University School of Medicine
New York, New York

Dawn Leger, Ph.D., Developmental Editor
Kajsa Forssen, Illustrator
Angela Lis, P.T., M.A., Editorial Assistant

LIPPINCOTT WILLIAMS & WILKINS
A Wolters Kluwer Company
Philadelphia • Baltimore • New York • London
Buenos Aires • Hong Kong • Sydney • Tokyo

Foreword

Mechanics and biology have always fascinated humankind. The importance of understanding the biomechanics of the musculoskeletal system cannot be underestimated. Much attention has been paid in recent years to genetic and biomolecular research, but the study of the mechanics of structure and of the whole body system is still of immense importance. Musculoskeletal ailments are among the most prevalent disorders in the world and will continue to grow as the population ages.

Since the days when I first studied biomechanics in Sweden with Carl Hirsch, through my years as an orthopaedic surgeon, teacher, and researcher, I have always emphasized combining basic and applied research with clinical experience. This text represents my full effort to integrate biomechanical knowledge into clinical training for patient care. It is not a simple task, but by relating the basic concepts of biomechanics to everyday life, rehabilitation, orthopaedics, traumatology, and patient care are greatly enhanced. Biomechanics is a multidisciplinary specialty, and so we have made a special effort to invite contributors from many disciplines so that individuals from different fields may feel comfortable reading this book.

Together with an invaluable team, Margareta Nordin and I have produced this third edition of *Basic Biomechanics of the Musculoskeletal System*. The new edition is sharpened and improved thanks to the input from the students and residents in orthopaedics that during the past 10 years have used the text. This book is written for students and with a major input from students and will hopefully be used to educate students and residents for many years to come. Although the basic information contained in the book remains largely unchanged, a considerable amount of extra information has been provided throughout. We have also made a special point to document with the key references any significant changes in the field of biomechanics and rehabilitation.

It has always been my interest to bridge the gap between engineering knowledge and clinical care and practice. This book is written primarily for clinicians such as orthopaedists, physical and occupational therapists, clinical ergonomists, chiropractors, and other health professionals who are acquiring a working knowledge of biomechanical principles for use in the evaluation and treatment of musculoskeletal dysfunction. We only hope that if you find this book interesting, you will seek more in-depth study in the field of biomechanics. Enjoy it, discuss it, and become a better clinician and/or researcher.

We are extremely proud that *Basic Biomechanics of the Musculoskeletal System* has been designated "A Classic" by the publishers, Lippincott Williams & Wilkins. We thank the readers, students, professors, and all who acquire the text and use it.

Victor H. Frankel, M.D., Ph.D., KNO

Preface

Biomechanics uses physics and engineering concepts to describe the motion undergone by the various body segments and forces acting on these body parts during normal activities. The interrelationship of force and motion is important and must be understood if rational treatment programs are to be applied to musculoskeletal disorders. Deleterious effects may be produced if the forces acting on the areas with disorders rise to high levels during exercise or other activities of daily living.

The purpose of this text is to acquaint the readers with the force-motion relationship within the musculoskeletal system and the various techniques used to understand these relationships. The third edition of *Basic Biomechanics of the Musculoskeletal System* is intended for use as a textbook either in conjunction with an introductory biomechanics course or for independent study. The third edition has been changed in many ways, but it is still a book that is designed for use by students who are interested in and want to learn about biomechanics. It is primarily written for students who do not have an engineering background but who want to understand the most basic concepts in biomechanics and physics and how these apply to the human body.

Input from students has greatly improved this third edition. We have used the book for 10 years in the Program of Ergonomics and Biomechanics at New York University, and it is the students and residents who have suggested the changes and who have continuously shown an interest in developing and improving this book. This edition has been further strengthened by the contribution of the students over the past year. We formed focus groups to understand better what the students wanted and applied their suggestions whenever possible. We retained the selected examples to illustrate the concepts needed for basic knowledge of the musculoskeletal biomechanics; we also have kept the important engineering concepts throughout the volume. We have added four chapters on applied biomechanics topics. Patient case studies and calculation boxes have been added to each chapter. We incorporated flowcharts throughout the book as teaching tools.

The text will serve as guide to a deeper understanding of musculoskeletal biomechanics gained through further reading and independent research. The information presented should also guide the reader in assessing the literature on biomechanics. We have attempted to provide therapeutic examples, but it was not our purpose to cover this area; instead, we have described the underlying basis for rational therapeutic or exercise programs.

An introductory chapter describes the importance of the study of biomechanics, and an appendix on the international system of measurements serves as an introduction to the physical measurements used throughout the book. The reader needs no more than basic knowledge of mathematics to fully comprehend the material in the book, but it is important to review the appendix on the SI System and its application to biomechanics.

The body of the third edition is then divided into three sections. The first section is the Biomechanics of Tissues and Structures of the Musculoskeletal System and covers the basic biomechanics of bone, ligaments, cartilage, tendons, muscles, and nerves. The second section covers the Biomechanics of Joints, including every joint system in the human body. Chapters range from the foot and ankle through the cervical spine, and cover every joint in between. The third section covers some topics in Applied Biomechanics, including chapters on fracture fixation; arthroplasty; sitting, standing and lying; and gait. These are basic chapters that serve to introduce topics in applied biomechanics; they are not in-depth explanations of the subject.

Finally, we hope that the revision and expansion of this third edition of *Basic Biomechanics of the Musculoskeletal System* will bring about an increased awareness of the importance of biomechanics. It has never been our intention to completely cover the subject, but instead provide a basic introduction to the field that will lead to further study of this important topic.

Margareta Nordin and Victor H. Frankel

Acknowledgments

This book was made possible through the outstanding contributions of many individuals. The chapter authors' knowledge and understanding of the basic concepts of biomechanics and their wealth of experience have brought both breadth and depth to this work. Over the past 10 years, questions raised by students and residents have made this book a better teaching tool. The Third Edition could not have been done without the students who have shared their comments and really scrutinized the Second Edition. There are too many names to list here, but we thank each student who asked a question or made a suggestion during the course of his or her studies. Special thanks to the students who participated in several focus groups, whose input was invaluable in finalizing the contents and design of the text.

We are honored and grateful for the contributions of everyone who has worked to prepare this new edition. We can honestly say that this third edition is written for the student and by students and residents who leave the classroom with the knowledge to enhance our life and existence.

A book of this size with its large number of figures, legends, and references cannot be produced without a strong editorial team. As project editor, Dawn Leger's continuous effort and perseverance and thoughtfulness shines through the entire book. She has contributed not just to the editing but also to logistics, and as a stylist, as an innovator, and a friend. Our editorial assistant, Angela Lis, is a physical therapist and recent recipient of the MA degree in Ergonomics and Biomechanics from NYU. As a recent graduate, Angela was also a recent user of the book, and she devoted several months to help finalize this edition. She created the flowcharts and scrutinized all the figures, patient cases, and calculation boxes. Angela took this book to her heart, and we are all the better for her passion and attention to detail.

The illustrator, Kajsa Forssen, has now worked on all three editions of this text. Her never-failing grasp of biomechanical illustrations, her simplicity and exactness of figures, is always appreciated. In drawing all the figures and graphs, she considers how they would translate into a slide or into a computer-generated presentation. Kajsa Forssen is one of the top illustrators that we have ever worked with, and she has been an important member of the publication team.

This book was also made when publication companies merged and merged again, and in the end we are deeply grateful to Ulita Lushnycky, who has with her team at Lippincott Williams & Wilkins been responsible for the production. She has worked with tremendous energy and positive thinking, put the book together in record speed, and we forward our sincerest gratitude to her. We are also thankful for a development grant provided by Lippincott Williams & Wilkins to finance this effort.

Our colleagues at the Occupational and Industrial Orthopaedic Center and the Department of Orthopaedics of the Hospital for Joint Diseases Orthopaedic Institute functioned as critical reviewers and contributors to the chapters. Special thanks is extended to David Goldsheyder for assistance in reviewing the biomechanical calculation boxes, to Marco Campello as a contributor and reviewer, and to Shira Schecter Weiner for contributing to the spine chapter. Much thanks to Dr. Mark Pitman for supplying vital x-rays for the new edition. We are particularly grateful to Dr. Markus Pietrek for contributing with the latest on intra-abdominal pressure, to Dr. Ali Sheikhzadeh for reviewing chapters and contributing new references, to Dr. Tobias Lorenz for his work on the first section, and to all other staff at the Occupational and Industrial Orthopaedic Center who have been managing the center while we are absorbed with the book.

We are most grateful to Drs. Bejjani, Lindh, Pitman, Peterson, and Stuchin for their contributions to the second edition which served as a framework for the updated third edition.

The third edition of *Basic Biomechanics of the Musculoskeletal System* was supported throughout its production by the Research and Development Foundation of the Hospital for Joint Diseases Orthopaedic Institute and the hospital administration, to whom we forward our sincere gratitude.

To all who helped, we say again, thank you and TACK SA MYCKET.

Margareta Nordin and Victor H. Frankel

Contributors

Gunnar B. J. Andersson, M.D., Ph.D.
Professor and Chairman
Department of Orthopaedic Surgery
Rush-Presbyterian-St. Luke's Medical Center
Chicago, IL

Thomas P. Andriacchi, Ph.D.
Biomechanical Engineering Division
Stanford University
Stanford, CA

Sherry I. Backus, M.D., P.T.
Senior Research Physical Therapist and Research Associate
Motion Analysis Laboratory
Hospital for Special Surgery
New York, NY

Ann E. Barr, Ph.D., P.T.
Assistant Professor
Physical Therapy Department
College of Allied Health Professionals
Temple University
Philadelphia, PA

Fadi Joseph Bejjani, M.D., Ph.D.
Director of Occupational Musculoskeletal Diseases
Department
University Rehabilitation Association
Newark, NJ

Maureen Gallagher Birdzell, Ph.D.
Department of Orthopaedic Surgery
Hospital for Joint Diseases/Mt. Sinai NYU Health
New York, NY

Marco Campello, P.T., M.A.
Associate Clinical Director
Occupational and Industrial Orthopaedic Center
Hospital for Joint Diseases/Mt. Sinai NYU Health
New York, NY

Dennis R. Carter, Ph.D.
Professor
Biomechanical Engineering Program
Stanford University
Stanford, CA

Craig J. Della Valle, M.D.
NYU-HJD Department of Orthopaedic Surgery
Hospital for Joint Diseases
School of Medicine
New York University
New York, NY

Victor H. Frankel, M.D., Ph.D., KNO
President Emeritus
Hospital for Joint Diseases Orthopaedic Institute
Professor of Orthopaedic Surgery
New York University School of Medicine
New York, NY

Ross Todd Hockenbury, M.D.
River City Orthopaedic Surgeons
Louisville, KY

Clark T. Hung, Ph.D.
Assistant Professor
Department of Mechanical Engineering and Center for Biomedical Engineering
Columbia University
New York, NY

Debra E. Hurwitz, Ph.D.
Assistant Professor
Department of Orthopaedics
Rush-Presbyterian-St. Luke's Medical Center
Chicago, IL

Laith M. Jazrawi, M.D.
NYU-HJD Department of Orthopaedic Surgery
Hospital for Joint Diseases
School of Medicine
New York University
New York, NY

Frederick J. Kummer, Ph.D.
Associate Director, Musculoskeletal Research Center
Hospital for Joint Diseases/Mt. Sinai NYU Health
Research Professor, NYU-HJD Department of Orthopaedic Surgery
School of Medicine
New York University
New York, NY

Dawn Leger, Ph.D.
Adjunct Assistant Professor
NYU-HJD Department of Orthopaedics
School of Medicine
New York University
New York, NY

Jane Bear-Lehman, Ph.D., OTR, FAOTA
Assistant Professor of Clinical Occupational Therapy
Department of Occupational Therapy
Columbia University College of Physicians and Surgeons
New York, NY

Margareta Lindh, M.D., Ph.D.
Associate Professor
Department of Physical Medicine and Rehabilitation
Sahlgren Hospital
Gothenburg University
Gothenburg, Sweden

Angela Lis, M.A., P.T.
Research Physical Therapist
Occupational and Industrial Orthopaedic Center
Hospital for Joint Diseases/Mt. Sinai NYU Health
New York, NY
Associate Professor
Physical Therapy Program
Corporación Universitaria Iberoamericana
Bogotá, COLOMBIA

Tobias Lorenz, M.D.
Fellow
Occupational and Industrial Orthopaedic Center
Hospital for Joint Diseases/Mt. Sinai NYU Health
New York, NY

Goran Lundborg, M.D.
Professor
Department of Hand Surgery
Lunds University
Malmo Allmanna Sjukhus
Malmo, Sweden

Ronald Moskovich, M.D.
Associate Chief
Spine Surgery
NYU-HJD Department of Orthopaedic Surgery
Hospital for Joint Diseases
School of Medicine
New York University
New York, NY

Van C. Mow, Ph.D.
Director
Orthopaedic Research Laboratory
Department of Orthopaedic Surgery
Columbia University
New York, NY

Robert R. Myers, Ph.D.
Associate Professor
Department of Anesthesiology
University of California San Diego
La Jolla, CA

Margareta Nordin, P.T., Dr. Sci.
Director, Occupational and Industrial Orthopaedic Center (OIOC)
Hospital for Joint Diseases Orthopaedic Institute
Mt. Sinai NYU Health
Program of Ergonomics and Biomechanics
New York University
Research Professor
Department of Orthopaedics and Environmental Health Science
School of Medicine, New York University
New York, NY

Kjell Olmarker, M.D., Ph.D.
Associate Professor
Department of Orthopaedics
Sahlgren Hospital
Gothenburg University
Gothenburg, Sweden

Nihat Özkaya (deceased)
Associate Professor
Occupational and Industrial Orthopaedic Center
Hospital for Joint Diseases
Research Associate Professor
Department of Environmental Medicine
New York University
New York, NY

Lars Peterson, M.D., Ph.D.
Gruvgat 6
Vastra Frolunda
Sweden

Mark I. Pitman, M.D.
Clinical Associate Professor
NYU-HJD Department of Orthopaedic Surgery
School of Medicine
New York University
New York, NY

Andrew S. Rokito, M.D.
Associate Chief, Sports Medicine Service
Assistant Professor
NYU-HJD Department of Orthopaedic Surgery
School of Medicine
New York University
New York, NY

Bjorn Rydevik, M.D., Ph.D.
Professor and Chairman
Department of Orthopaedics
Sahlgren Hospital
Gothenburg University
Gothenburg, Sweden

G. James Sammarco, M.D.
Program Director
Fellowship in Adult Reconstructive Surgery
Foot and Ankle Orthopaedic Surgery Program
The Center for Orthopaedic Care, Inc.
Volunteer Professor of Orthopaedic Surgery
Department of Orthopaedics
University of Cincinnati Medical Center
Cincinnati, OH

Chris J. Snijders, Ph.D.
Professor
Biomedical Physics and Technology
Faculty of Medicine
Erasmus University
Rotterdam, The Netherlands

Steven Stuchin, M.D.
Director Clinical Orthopaedic Services
Director Arthritis Service
Associate Professor
NYU-HJD Department of Orthopaedics
School of Medicine
New York University
New York, NY

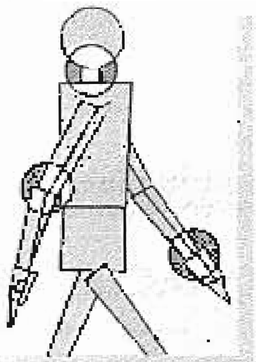
Shira Schecter Weiner, M.A., P.T.
Research Physical Therapist
Occupational and Industrial Orthopaedic Center
Hospital for Joint Diseases/Mt. Sinai NYU Health
New York, NY

Joseph D. Zuckerman, M.D.
Professor and Chairman
NYU-HJD Department of Orthopaedic Surgery
Hospital for Joint Diseases
School of Medicine
New York University
New York, NY

Contents

THIRD EDITION

BASIC BIOMECHANICS *of the* MUSCULOSKELETAL SYSTEM



CHAPTER 1

Introduction to Biomechanics: Basic Terminology and Concepts

Nihat Özkaya, Dawn Leger

Introduction

Biomechanics is considered a branch of bioengineering and biomedical engineering. Bioengineering is an interdisciplinary field in which the principles and methods from engineering, basic sciences, and technology are applied to design, test, and manufacture equipment for use in medicine and to understand, define, and solve problems in physiology and biology. Bioengineering is one of several specialty areas that come under the general field of biomedical engineering.

Biomechanics considers the applications of classical mechanics to the analysis of biological and physiological systems. Different aspects of biomechanics utilize different parts of applied mechanics. For example, the principles of statics have been applied to analyze the magnitude and nature of forces involved in various joints and muscles of the musculoskeletal system. The principles of dynamics have been utilized for motion description, gait analysis, and segmental motion analysis and have many applications in sports mechanics. The mechanics of solids provides the necessary tools for developing the field constitutive equations for biological systems that are used to evaluate their functional behavior under different load conditions. The principles of fluid mechanics have been used to investigate blood flow in the circulatory system, air flow in the lung, and joint lubrication.

Research in biomechanics is aimed at improving our knowledge of a very complex structure—the human body. Research activities in biomechanics can be divided into three areas: experimental studies, model analyses, and applied research. Experimental studies in biomechanics are done to determine the mechanical properties of biological materials, including the bone, cartilage, muscle, tendon, ligament, skin, and blood as a whole or as parts constituting them. Theoretical studies involving mathematical model analyses have also been an important component of research in biomechanics. In general, a model that is based on experimental findings can be used to predict the effect of environmental and operational factors without resorting to laboratory experiments.

Applied research in biomechanics is the application of scientific knowledge to benefit human beings. We know that musculoskeletal injury and illness is one of the primary occupational hazards in industrialized countries. By learning how the musculoskeletal system adjusts to common work conditions and by developing guidelines to assure that manual work conforms more closely to the physical limitations of the human body and to natural body movements, these injuries may be combatted.

Basic Concepts

Biomechanics of the musculoskeletal system requires a good understanding of basic mechanics. The basic terminology and concepts from mechanics and physics are utilized to describe internal forces of the human body. The objective of studying these forces is to understand the loading condition of soft tissues and their mechanical responses. The purpose of this section is to review the basic concepts of applied mechanics that are used in biomechanical literature and throughout this book.

SCALARS, VECTORS, AND TENSORS

Most of the concepts in mechanics are either scalar or vector. A scalar quantity has a magnitude only. Concepts such as mass, energy, power, mechanical work, and temperature are scalar quantities. For example, it is sufficient to say that an object has 80 kilograms (kg) of mass. A vector quantity, conversely, has both a magnitude and a direction associated with it. Force, moment, velocity, and acceleration are examples of vector quantities. To describe a force fully, one must state how much force is applied and in which direction it is applied. The magnitude of a vector is also a scalar quantity. The magnitude of any quantity (scalar or vector) is always a positive number corresponding to the numerical measure of that quantity.

Graphically, a vector is represented by an arrow. The orientation of the arrow indicates the line of action and the arrowhead denotes the direction and sense of the vector. If more than one vector must be shown in a single drawing, the length of each arrow must be proportional to the magnitude of the vector it represents. Both scalars and vectors are special forms of a more general category of all quantities in mechanics called tensors. Scalars are also known as "zero-order tensors," whereas vectors are "first-order tensors." Concepts such as stress and strain, conversely, are "second-order tensors."

FORCE VECTOR

Force can be defined as mechanical disturbance or load. When an object is pushed or pulled, a force is applied on it. A force is also applied when a ball is thrown or kicked. A force acting on an object may deform the object, change its state of motion, or both. Forces may be classified in various ways according to their effects on the objects to which they are applied or according to their orientation as compared with one another. For example, a force may be internal or external, normal (perpendicular) or tangential; tensile, compressive, or shear; gravitational (weight); or frictional. Any two or more forces acting on a single body may be coplanar (acting on a two-dimensional plane surface); collinear (have a common line of action), concurrent (lines of action intersecting at a single point), or parallel. Note that weight is a special form of force. The weight of an object on Earth is the gravitational force exerted by Earth on the mass of that object. The magnitude of the weight of an object on Earth is equal to the mass of the object times the magnitude of the gravitational acceleration, which is approximately 9.8 meters per second squared (m/s^2). For example, a 10-kg object weighs approximately 98 newtons (N) on Earth. The direction of weight is always vertically downward.

TORQUE AND MOMENT VECTORS

The effect of a force on the object it is applied upon depends on how the force is applied and how the object is supported. For example, when pulled, an open door will swing about the edge along which it is hinged to the wall. What causes the door to swing is the torque generated by the applied force about an axis that passes through the hinges of the door. If one stands on the free-end of a diving board, the board will bend. What bends the board is the moment of the body weight about the fixed end of the board. In general, torque is associated with the rotational and twisting action of applied forces, while moment is related to the bending action. However, the mathematical definition of moment and torque is the same.

Torque and moment are vector quantities. The magnitude of the torque or moment of a force about a point is equal to the magnitude of the force times the length of the shortest distance between the point and the line of action of the force, which is known as the lever or moment arm. Con-

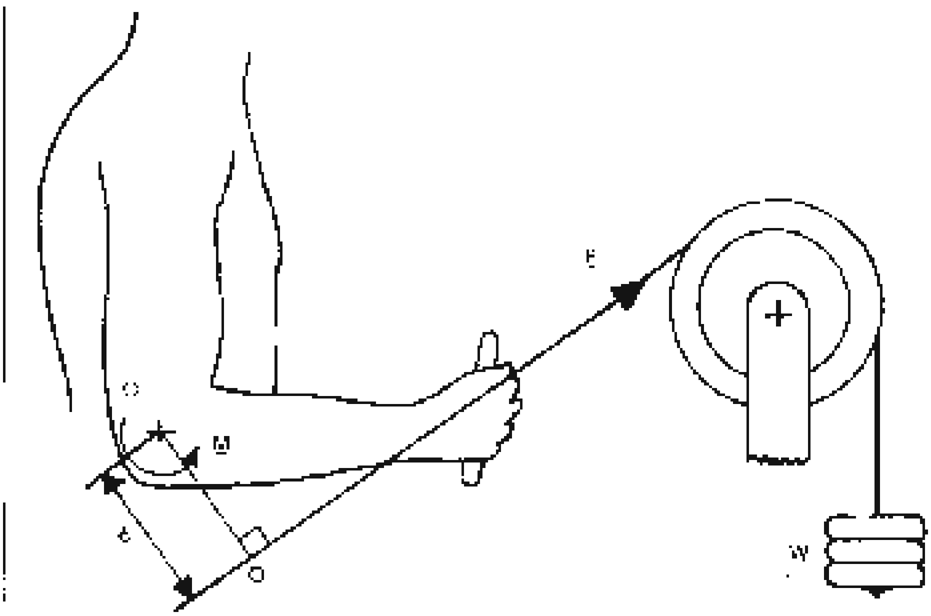


FIG. 1-1

Definition of torque. *Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom, Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.*

sider a person on an exercise apparatus who is holding a handle that is attached to a cable (Fig. 1-1). The cable is wrapped around a pulley and attached to a weight pan. The weight in the weight pan stretches the cable such that the magnitude F of the tensile force in the cable is equal to the weight of the weight pan. This force is transmitted to the person's hand through the handle. At this instant, if the cable attached to the handle makes an angle θ with the horizontal, then the force $\underline{F}$ exerted by the cable on the person's hand also makes an angle θ with the horizontal. Let O be a point on the axis of rotation of the elbow joint. To determine the magnitude of the moment due to force $\underline{F}$ about O, extend the line of action of force $\underline{F}$ and draw a line from O that cuts the line of action of $\underline{F}$ at right angles. If the point of intersection of the two lines is Q, then the distance d between O and Q is the lever arm, and the magnitude of the moment $\underline{M}$ of force $\underline{F}$ about the elbow joint is $M = dF$. The direction of the moment vector is perpendicular to the plane defined by the line of action of $\underline{F}$ and line OQ, or for this two-dimensional case, it is counterclockwise.

NEWTON'S LAWS

Relatively few basic laws govern the relationship between applied forces and corresponding motions. Among these, the laws of mechanics introduced by Sir Isaac Newton (1642–1727) are the most important. Newton's first law states that an object at rest will remain at rest or an object in motion will move in a straight line with constant velocity if the net force acting on the object is zero. Newton's second law states that an object with a nonzero net force acting on it will accelerate in the direction of the net force and that the magnitude of the acceleration will be proportional to the magnitude of the net force. Newton's second law can be formulated as $\underline{F} = m\,\underline{a}$. Here, $\underline{F}$ is the applied force, m is the mass of the object, and $\underline{a}$ is the linear (translational) acceleration of the object on which the force is applied. If more than one force is acting on the object, then $\underline{F}$ represents the net or the resultant force (the vector sum of all forces). Another way of stating Newton's second law of motion is $\underline{M} = I\,\underline{\alpha}$, where $\underline{M}$ is the net or resultant moment of all forces acting on the object, I is the mass moment of inertia of the object, and $\underline{\alpha}$ is the angular (rotational) acceleration of the object. The mass m and mass moment of inertia I in these equations of motion are measures of resistance to changes in motion. The larger the inertia of an object, the more difficult it is to set in motion or to stop if it is already in motion.

Newton's third law states that to every action there is a reaction and that the forces of action and reaction between interacting objects are equal in magnitude, opposite in direction, and have the same line of action. This law has important applications in constructing free-body diagrams.

FREE-BODY DIAGRAMS

Free-body diagrams are constructed to help identify the forces and moments acting on individual parts of a system and to ensure the correct use of the equations of mechanics to analyze the system. For this purpose, the parts constituting a system are isolated from their surroundings and the effects of surroundings are replaced by proper forces and moments.

The human musculoskeletal system consists of many parts that are connected to one another through a complex tendon, ligament, muscle, and joint structure. In some analyses, the objective may be to investigate the forces involved at and around various joints of the human body for different postural and load conditions. Such analyses can be carried out by separating the body into two parts at the joint of interest and drawing the free-body diagram of one of the parts. For example, consider the arm illustrated in Figure 1-2. Assume that the forces involved at the elbow joint are to be analyzed. As illustrated in Figure 1-2, the entire body is separated into two at the elbow joint and the free-body diagram of the forearm is drawn (Fig. 1-2B). Here

$\underline{F}$ is the force applied to the hand by the handle of the cable attached to the weight in the weight pan,

$\underline{W}$ is the total weight of the lower arm acting at the center of gravity of the lower arm,

$\underline{F}_{M1}$ is the force exerted by the biceps on the radius,

$\underline{F}_{M3}$ is the force exerted by the brachioradialis muscles on the radius,

$\underline{F}_{M2}$ is the force exerted by the brachialis muscles on the ulna, and

$\underline{F}_J$ is the resultant reaction force at the humeroulnar and humeroradial joints of the elbow. Note that the muscle and joint reaction forces represent the mechanical effects of the upper arm on the lower arm. Also note that as illustrated in Figure 1-2A (which is not a complete free-body diagram), equal magnitude but opposite muscle and joint reaction forces act on the upper arm as well.

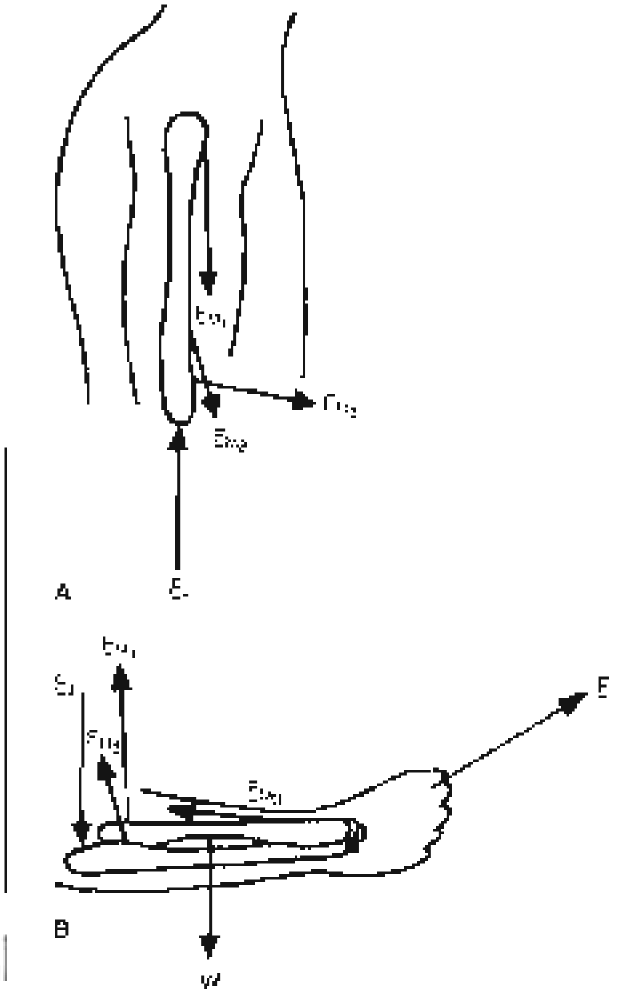


FIG. 1-2

Forces involved at and around the elbow joint and the free body diagram of the lower arm. *Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom.* Environmental and Occupational Medicine *(3rd ed., pp. 1437–1454). New York: Lippincott-Raven.*

CONDITIONS FOR EQUILIBRIUM

Statics is an area within applied mechanics that is concerned with the analysis of forces on rigid bodies in equilibrium. A rigid body is one that is assumed to undergo no deformations. In reality, every object or material may undergo deformation to an extent when acted on by forces. In some cases, the amount of deformation may be so small that it may not affect the desired analysis and the object is assumed to be rigid. In mechanics, the term equilibrium implies that the body of concern is either at rest or moving with constant velocity. For a body to be in a state of equilibrium, it has to be both in translational and rotational equilibrium. A body is in translational equilibrium if the net force (vector sum of all forces) acting on it is zero. If the net force is zero, then the linear acceleration (time rate of change of linear velocity) of the body is zero, or the linear velocity of the body is either constant or zero. A body is in rotational equilibrium if the net moment (vector sum of the moments of all forces) acting on it is zero. If the net moment is zero, then the angular acceleration (time rate of change of angular velocity) of the body is zero, or the angular velocity of the body is either constant or zero. Therefore, for a body in a state of equilibrium, the equations of motion (Newton's second law) take the following special forms:

$$\Sigma \underline{F} = 0 \text{ and } \Sigma \underline{M} = 0$$

It is important to remember that force and moment are vector quantities. For example, with respect to a rectangular (Cartesian) coordinate system, force and moment vectors may have components in the x, y, and z directions. Therefore, if the net force acting on an object is zero, then the sum of forces acting in each direction must be equal to zero ($\Sigma F_x = 0$, $\Sigma F_y = 0$, $\Sigma F_z = 0$). Similarly, if the net moment on an object is zero, then the sum of moments in each direction must also be equal to zero ($\Sigma M_x = 0$, $\Sigma M_y = 0$, $\Sigma M_z = 0$). Therefore, for three-dimension force systems there are six conditions of equilibrium. For two-dimensional force systems in the xy-plane, only three of these conditions ($\Sigma F_x = 0$, $\Sigma F_y = 0$, and $\Sigma M_z = 0$) need to be checked.

STATICS

The principles of statics (equations of equilibrium) can be applied to investigate the muscle and joint forces involved at and around the joints for various postural positions of the human body and its segments. The immediate purpose of static analysis is to provide answers to questions such as: What tension must the neck extensor muscles exert on the head to support the head in a specified position? When a person bends, what would be the force exerted by the erector spinae on the fifth lumbar vertebra? How does the compression at the elbow, knee, and ankle joints vary with externally applied forces and with different segmental arrangements? How does the force on the femoral head vary with loads carried in the hand? What are the forces involved in various muscle groups and joints during different exercise conditions?

In general, the unknowns in static problems involving the musculoskeletal system are the magnitudes of joint reaction forces and muscle tensions. The mechanical analysis of a skeletal joint requires that we know the vector characteristics of tensions in the muscles, the proper locations of muscle attachments, the weights of body segments, and the locations of the centers of gravity of the body segments. Mechanical models are obviously simple representations of complex systems. Many models are limited by the assumptions that must be made to reduce the system under consideration to a statically determinate one. Any model can be improved by considering the contributions of other muscles, but that will increase the number of unknowns and make the model a statically indeterminate one. To analyze the improved model, the researcher would need additional information related to the muscle forces. This information can be gathered through electromyography measurements of muscle signals or by applying certain optimization techniques. A similar analysis can be made to investigate forces involved at and around other major joints of the musculoskeletal system.

MODES OF DEFORMATION

When acted on by externally applied forces, objects may translate in the direction of the net force and rotate in the direction of the net torque acting on them. If an object is subjected to externally applied forces but is in static equilibrium, then it is most likely that there is some local shape change within the object. Local shape change under the effect of applied forces is known as deformation. The extent of deformation an object may undergo depends on many factors, including the material properties, size, and shape of the object; environmental factors such as heat and humidity; and the magnitude, direction, and duration of applied forces.

One way of distinguishing forces is by observing their tendency to deform the object they are applied upon. For example, the object is said to be in tension if the body tends to elongate and in compression if it tends to shrink in the direction of the applied forces. Shear loading differs from tension and compression in that it is caused by forces acting in directions tangent to the area resisting the forces causing shear, whereas both tension and compression are caused by collinear forces applied perpendicular to the areas on which they act. It is common to call tensile and compressive forces normal or axial forces; shearing forces are tangential forces. Objects also deform when they are subjected to forces that cause bending and torsion, which are related to the moment and torque actions of applied forces.

A material may respond differently to different loading configurations. For a given material, there may be different physical properties that must be considered while analyzing the response of that material to tensile loading as compared with compressive or shear loading. The mechanical properties of materials are established through stress analysis by subjecting them to various experiments such as uniaxial tension and compression, torsion, and bending tests.

NORMAL AND SHEAR STRESSES

Consider the whole bone in Figure 1-3A that is subjected to a pair of tensile forces of magnitude F. The bone is in static equilibrium. To analyze the forces induced within the bone, the method of sections can be applied by hypothetically cutting the bone into two pieces through a plane perpendicular to the long axis of the bone. Because the bone as a whole is in equilibrium, the two pieces must individually be in equilibrium as well. This requires that at the cut section of each piece there is an internal force that is equal in magnitude but opposite in direction to the externally applied force (Fig. 1-3B). The internal force is distributed over the entire cross-sectional area of the cut section, and F represents the resultant of the distributed force (Fig. 1-3C). The intensity of this distributed force (force per unit area) is known as stress. For the case shown in Figure 1-3, because the force resultant at the cut section is perpendicular to the plane of the cut, the corresponding stress is called a normal or axial stress. It is customary to use the symbol σ (sigma) to refer to normal stresses. Assuming that the intensity of the distributed force at the cut section is uniform over the cross-sectional area A of the bone, then $\sigma = F/A$. Normal stresses that are caused by forces that tend to stretch (elongate) materials are more specifically known as tensile stresses; those that tend to shrink them are known as compressive stresses. According to the Standard International (SI) unit system (see Appendix), stresses are measured in newton per square meter (N/m^2), which is also known as pascal (Pa).

There is another form of stress, shear stress, which is a measure of the intensity of internal forces acting tangent (parallel) to a plane of cut. For

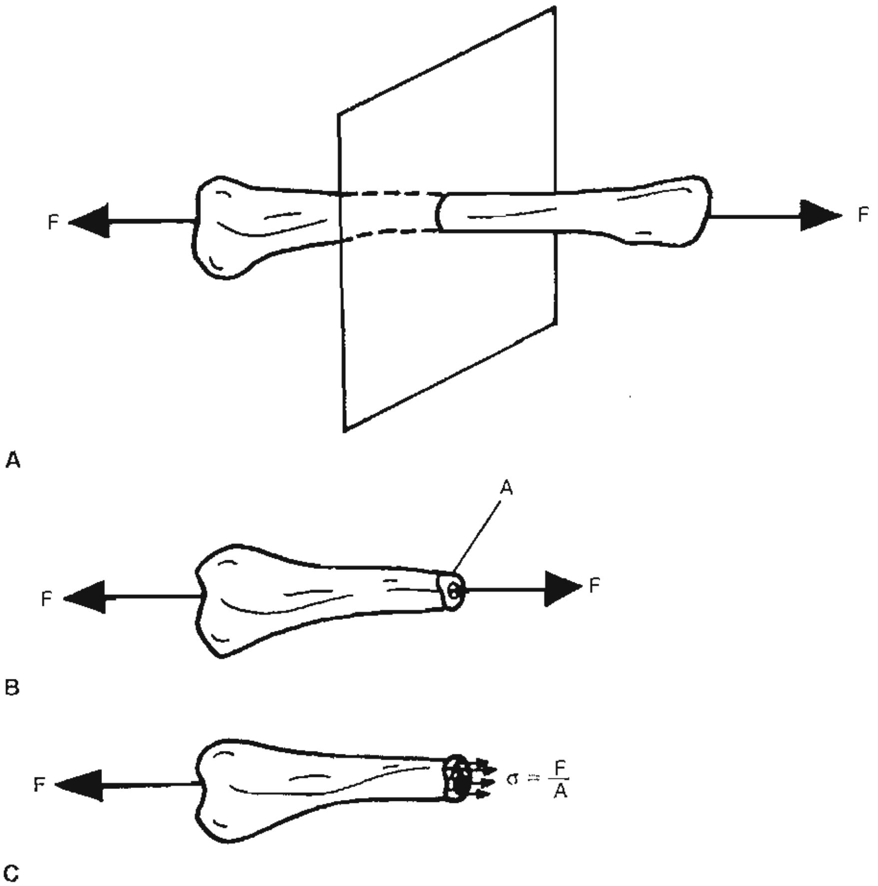


FIG. 1-3

Definition of normal stress. *Reprinted with permission from Özkaya, N. (1998). Biomechanics.* In W.N. Rom, Environmental and Occupational Medicine *(3rd ed., pp. 1437–1454). New York: Lippincott-Raven.*

example, consider the whole bone in Figure 1-4*A*. The bone is subject to a number of parallel forces that act in planes perpendicular to the long axis of the bone. Assume that the bone is cut into two parts through a plane perpendicular to the long axis of the bone (Fig. 1-4*B*). If the bone as a whole is in equilibrium, its individual parts must be in equilibrium as well. This requires that there must be an internal force at the cut section that acts in a direction tangent to the cut surface. If the magnitudes of the external forces are known, then the magnitude F of the internal force can be calculated by considering the translational and rotational equilibrium of one of the parts constituting the bone. The intensity of the internal force tangent to the cut section is known as the shear stress. It is customary to use the symbol τ (tau) to refer to shear stresses (Fig. 1-4*C*). Assuming that the intensity of the force tangent to the cut section is uniform over the cross-sectional area A of the bone, then τ = F/A.

NORMAL AND SHEAR STRAINS

Strain is a measure of the degree of deformation. As in the case of stress, two types of strains can be distinguished. A normal strain is defined as the ratio of the change (increase or decrease) in length to the original (undeformed) length, and is commonly denoted with the symbol ϵ (epsilon). Consider the whole bone in Figure 1-5. The total length of the bone is l. If the bone is subjected to a pair of tensile forces, the length of the bone may increase to l′ or by an amount $\Delta\lambda = l' - l$. The normal strain is the ratio of the amount of elongation to the original

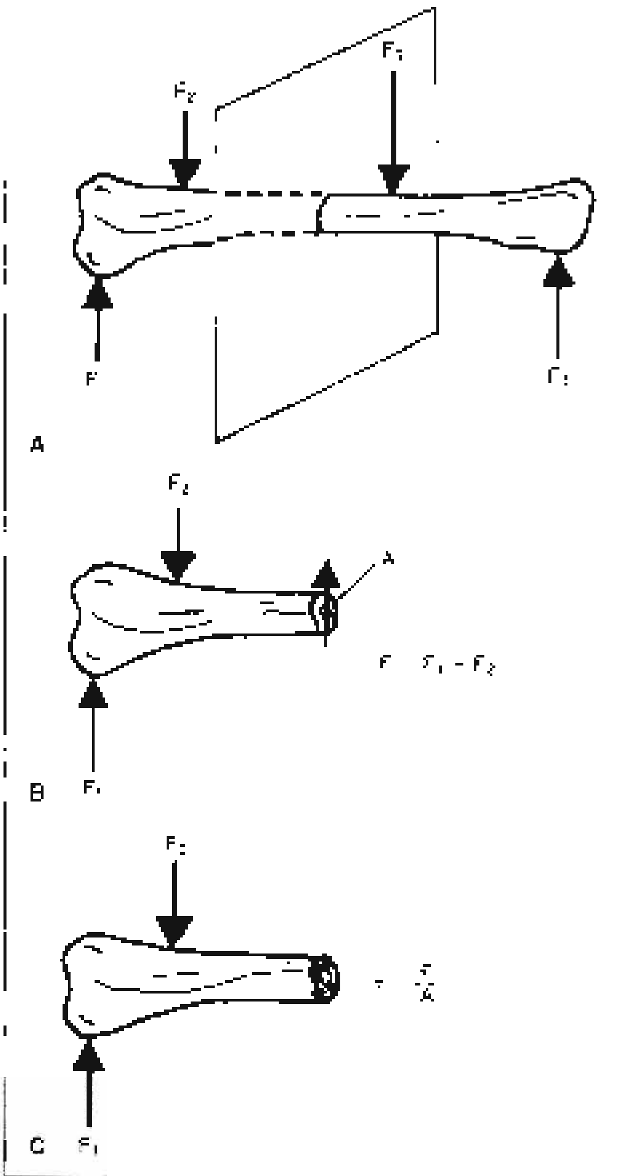


FIG. 1-4

Definition of shear stress. Adapted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom, Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.

length, or $\varepsilon = \Delta l / l$. If the length of the bone increases in the direction in which the strain is calculated, then the strain is tensile and positive. If the length of the bone decreases in the direction in which the strain is calculated, then the strain is compressive and negative.

Shear strains are related to distortions caused by shear stresses and are commonly denoted with the symbol γ (gamma). Consider the rectangle (ABCD) shown in Figure 1-6 that is acted on by a pair of tangential forces that deform the rectangle into a parallelogram (AB'C'D). If the relative horizontal displacement of the top and the bottom of the rectangle is d and the height of the rectangle is h, then the average shear strain is the ratio of d and h, which is equal to the tangent of angle γ. The angle γ is usually very small. For small angles, the tangent of the angle is approximately equal to the angle itself measured in radians. Therefore, the average shear strain is $\gamma = d/h$.

Strains are calculated by dividing two quantities measured in units of length. For most applications, the deformations and consequently the strains involved may be very small (e.g., 0.001). Strains can also be given in percentages (e.g., 0.1%).

STRESS-STRAIN DIAGRAMS

Different materials may demonstrate different stress-strain relationships. Consider the stress-strain diagram shown in Figure 1-7. There are six distinct points on the curve, which are labeled as O, P, E, Y, U, and R. Point O is the origin of the stress-strain diagram, which corresponds to the initial (no load, no deformation) state. Point P represents the proportionality limit. Between O and P, stress and strain are linearly proportional and the stress-strain diagram is a straight line. Point E represents the elastic limit. Point Y is the yield point, and the stress σ_y corresponding to the yield point is called the yield strength of the material. At this stress level, considerable elongation (yielding) can occur without a corresponding increase of load. U is the highest stress point on the stress-strain diagram. The stress σ_u is the ultimate strength of the material. The last point on the stress-strain diagram is R, which represents the rupture or failure point. The stress at which the failure occurs is called the rupture strength of the material. For some materials, it may not be easy to distinguish the elastic limit and the yield point. The yield strength of such materials is determined by the offset method, which is applied by drawing a line parallel to the linear section of the stress-strain diagram that passes through a strain level of approximately 0.2%. The intersection of this line with the stress-strain curve is taken to be the yield point, and the stress corresponding to this point is called the apparent yield strength of the material.

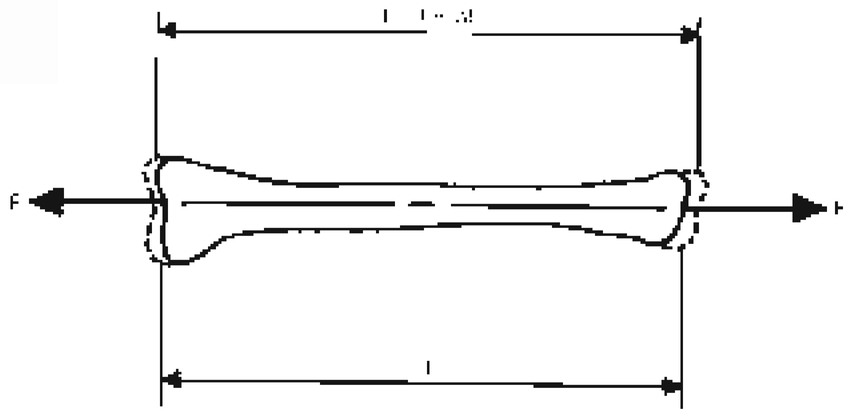


FIG. 1-5

Definition of normal strain. *Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom, Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.*

Note that a given material may behave differently under different load and environmental conditions. If the curve shown in Figure 1-7 represents the stress-strain relationship for a material under tensile loading, there may be a similar but different curve representing the stress-strain relationship for the same material under compressive or shear loading. Also, temperature is known to alter the relationship between stress and strain. For some materials, the stress-strain relationship may also depend on the rate at which the load is applied on the material.

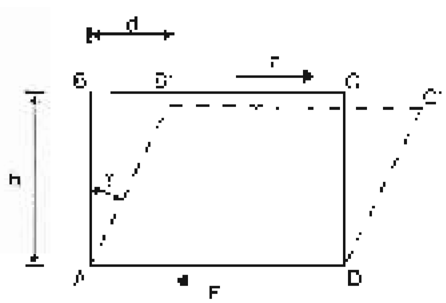


FIG. 1-6

Definition of shear strain. *Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom, Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.*

ELASTIC AND PLASTIC DEFORMATIONS

Elasticity is defined as the ability of a material to resume its original (stress-free) size and shape on removal of applied loads. In other words, if a load

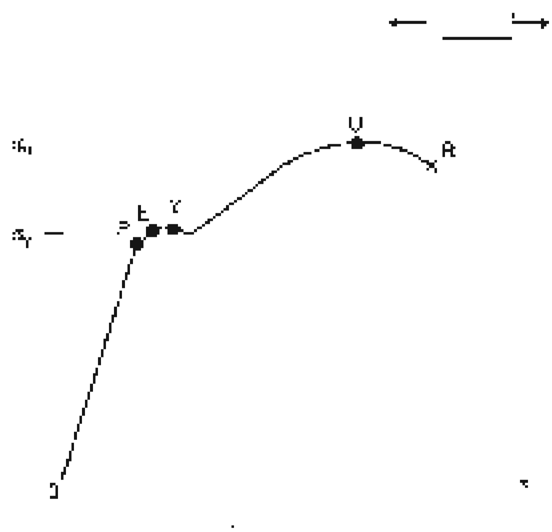


FIG. 1-7

Stress-strain diagrams. *Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom, Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.*

is applied on a material such that the stress generated in the material is equal to or less than the elastic limit, the deformations that took place in the material will be completely recovered once the applied loads are removed. An elastic material whose stress-strain diagram is a straight line is called a linearly elastic material. For such a material, the stress is linearly proportional to strain. The slope of the stress-strain diagram in the elastic region is called the elastic or Young's modulus of the material, which is commonly denoted by E. Therefore, the relationship between stress and strain for linearly elastic materials is $\sigma = E\varepsilon$. This equation that relates normal stress and strain is called a material function. For a given material, different material functions may exist for different modes of deformation. For example, some materials may exhibit linearly elastic behavior under shear loading. For such materials, the shear stress τ is linearly proportional to the shear strain γ, and the constant of proportionality is called the shear modulus, or the modulus of rigidity. If G represents the modulus of rigidity, then $\tau = G\gamma$. Combinations of all possible material functions for a given material form the constitutive equations for that material.

Plasticity implies permanent deformations. Materials may undergo plastic deformations following elastic deformations when they are loaded beyond their elastic limits. Consider the stress-strain diagram of a material under tensile loading (Fig. 1-7). Assume that the stresses in the specimen are brought to a level greater than the yield strength of the material. On removal of the applied load, the material will recover the elastic deformation that had taken place by following an unloading path parallel to the initial linearly elastic region. The point where this path cuts the strain axis is called the plastic strain, which signifies the extent of permanent (unrecoverable) shape change that has taken place in the material.

Viscoelasticity is the characteristic of a material that has both fluid and solid properties. Most materials are classified as either fluid or solid. A solid material will deform to a certain extent when an external force is applied. A continuously applied force on a fluid body will cause a continuous deformation (also known as flow). Viscosity is a fluid property that is a quantitative measure of resistance to flow. Viscoelasticity is an example of how areas in applied mechanics can overlap, because it utilizes the principles of both fluid and solid mechanics.

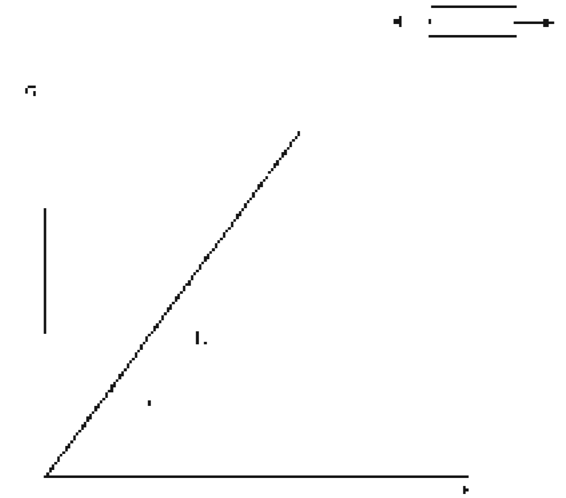


FIG. 1-8

Linearly elastic material behavior. *Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom, Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.*

VISCOELASTICITY

When they are subjected to relatively low stress levels, many materials such as metals exhibit elastic material behavior. They undergo plastic deformations at high stress levels. Elastic materials deform instantaneously when they are subjected to externally applied loads and resume their original shapes almost instantly when the applied loads are removed. For an elastic material, stress is a function of strain only, and the stress-strain relationship is unique (Fig. 1-8). Elastic materials do not exhibit time-dependent behavior. A different group of materials, such as polymer plastics, metals at high temperatures, and almost all biological materials, exhibits gradual deformation and recovery when subjected to loading and unloading. Such materials are called viscoelastic. The response of viscoelastic materials is dependent on how quickly the load is applied or removed. The extent of deformation that viscoelastic materials undergo is dependent on the rate at which the deformation-causing loads are applied. The stress-strain relationship for a viscoelastic material is not unique but is a function of time or the rate at which the stresses and strains are developed in the material (Fig. 1-9). The word "viscoelastic" is made of two words. Viscosity is a fluid property and

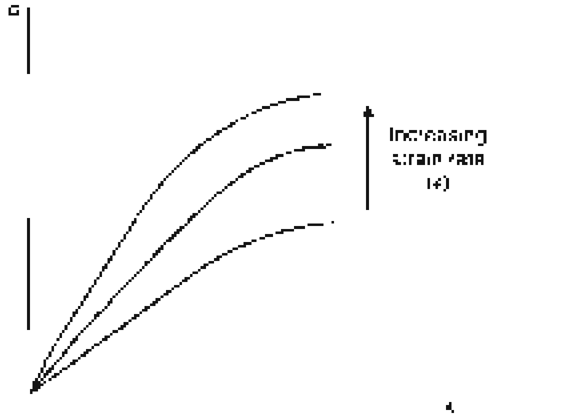


FIG. 1-9

Strain rate-dependent viscoelastic material behavior. Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom, Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.

is a measure of resistance to flow. Elasticity is a solid material property. Therefore, viscoelastic materials possess both fluid- and solid-like properties.

For an elastic material, the energy supplied to deform the material (strain energy) is stored in the material as potential energy. This energy is available to return the material to its original (unstressed) size and shape once the applied load is removed. The loading and unloading paths for an elastic material coincide, indicating no loss of energy. Most elastic materials exhibit plastic behavior at high stress levels. For elasto-plastic materials, some of the strain energy is dissipated as heat during plastic deformations. For viscoelastic materials, some of the strain energy is stored in the material as potential energy and some of it is dissipated as heat regardless of whether the stress levels are small or large. Because viscoelastic materials exhibit time-dependent material behavior, the differences between elastic and viscoelastic material responses are most evident under time-dependent loading conditions.

Several experimental techniques have been designed to analyze the time-dependent aspects of material behavior. As illustrated in Figure 1-10A, a creep and recovery test is conducted by applying a load on the material, maintaining the load at a constant level for a while, suddenly removing the load, and observing the material response. Under a creep and recovery test, an elastic material will respond with an instantaneous strain that would remain at a constant level until the load is removed (Fig. 1-10B). At the instant when the load is removed, the deformation will instantly and completely recover. To the same constant loading condition, a viscoelastic material will respond with a strain increasing and decreasing gradually. If the material is a viscoelastic solid, the recovery will eventually be complete (Fig. 1-10C). If the material is a viscoelastic fluid, complete recovery will never be achieved and there will be a residue of deformation left in the material (Fig. 1-10D). As illustrated in Figure 1-11A, a stress-relaxation experiment is conducted

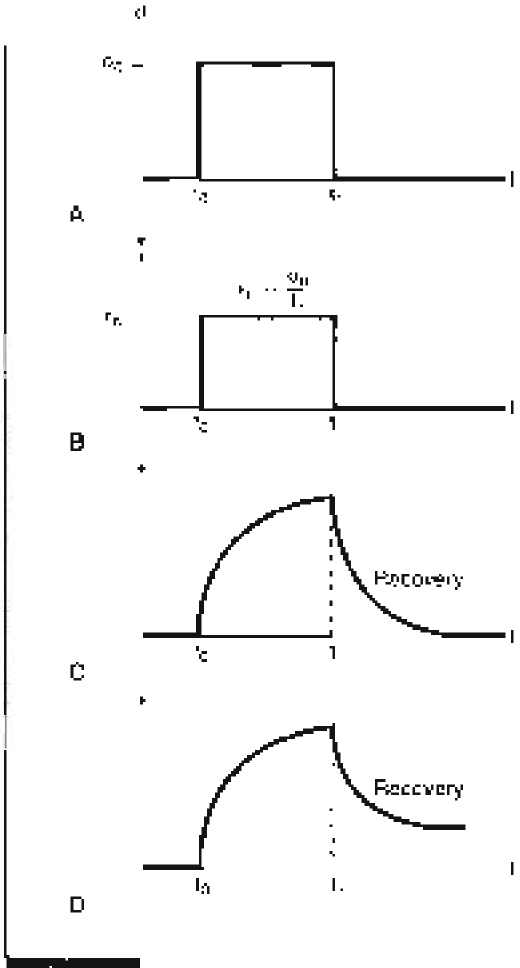


FIG. 1-10

Creep and recovery test. Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom, Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.

by straining the material to a level and maintaining the constant strain while observing the stress response of the material. Under a stress-relaxation test, an elastic material will respond with a stress developed instantly and maintained at a constant level (Fig. 1-11B). That is, an elastic material will not exhibit a stress-relaxation behavior. A viscoelastic material, conversely, will respond with an initial high stress level that will decrease over time. If the material is a viscoelastic solid, the stress level will never reduce to zero (Fig. 1-11C). As illustrated in Figure 1-11D, the stress will eventually reduce to zero for a viscoelastic fluid.

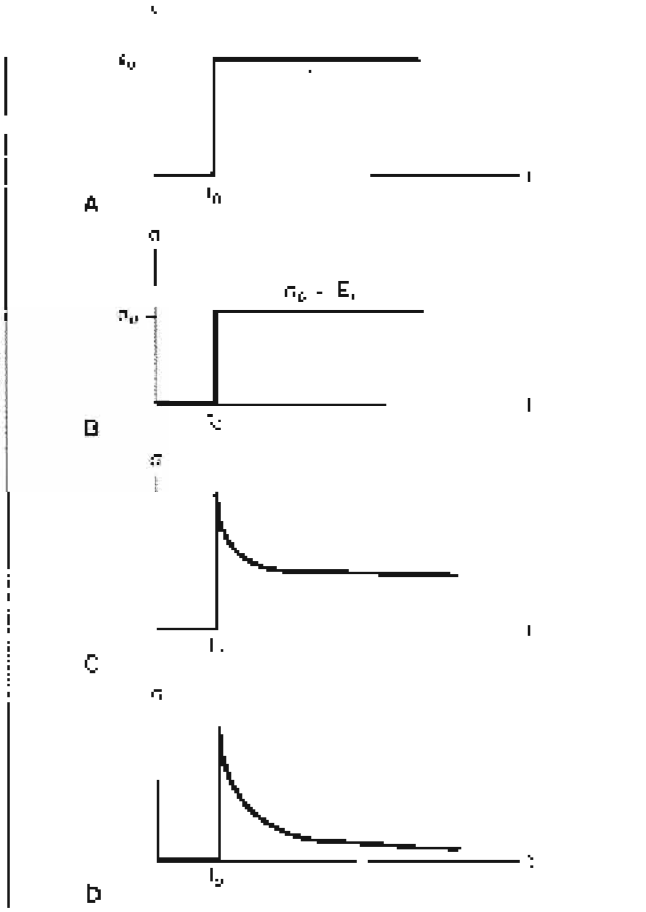


FIG. 1-11

Stress-relaxation experiment. Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom. Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.

MATERIAL PROPERTIES BASED ON STRESS-STRAIN DIAGRAMS

The stress-strain diagrams of two or more materials can be compared to determine which material is relatively stiffer, harder, tougher, more ductile, or more brittle. For example, the slope of the stress-strain diagram in the elastic region represents the elastic modulus that is a measure of the relative stiffness of materials. The higher the elastic modulus, the stiffer the material and the higher its resistance to deformation. A ductile material is one that exhibits a large plastic deformation prior to failure. A brittle material, such as glass, shows a sudden failure (rupture) without undergoing a considerable plastic deformation. Toughness is a measure of the capacity of a material to sustain permanent deformation. The toughness of a material is measured by considering the total area under its stress-strain diagram. The larger this area, the tougher the material. The ability of a material to store or absorb energy without permanent deformation is called the resilience of the material. The resilience of a material is measured by its modulus of resilience, which is equal to the area under the stress-strain curve in the elastic region.

Although they are not directly related to the stress-strain diagrams, other important concepts are used to describe material properties. For example, a material is called homogeneous if its properties do not vary from location to location within the material. A material is called isotropic if its properties are independent of direction. A material is called incompressible if it has a constant density.

PRINCIPAL STRESSES

There are infinitely many possibilities of constructing elements around a given point within a structure. Among these possibilities, there may be one element for which the normal stresses are maximum and minimum. These maximum and minimum normal stresses are called the principal stresses, and the planes whose normals are in the directions of the maximum and minimum stresses are called the principal planes. On a principal plane, the normal stress is either maximum or minimum, and the shear stress is zero. It is known that fracture or material failure occurs along the planes of maximum stresses, and structures must be designed by taking into consideration the maximum stresses involved. Failure by yielding (excessive deformation) may occur whenever the largest principal stress is equal to the yield strength of the material or failure by rupture may occur whenever the largest principal stress is equal to the ultimate strength of the material. For a given structure and loading condition, the principal stresses may be within the limits of operational safety. However, the structure must also be checked for critical shearing stress, called the maximum shear stress. The maximum shear stress occurs on a material element for which the normal stresses are equal.

FATIGUE AND ENDURANCE

Principal and maximum shear stresses are useful in predicting the response of materials to static loading configurations. Loads that may not cause the failure of a structure in a single application may cause fracture when applied repeatedly. Failure may occur after a few or many cycles of loading and unloading, depending on factors such as the amplitude of the applied load, mechanical properties of the material, size of the structure, and operational conditions. Fracture resulting from repeated loading is called fatigue.

Several experimental techniques have been developed to understand the fatigue behavior of materials. Consider the bar shown in Figure 1-12A. Assume that the bar is made of a material whose ultimate strength is σ_u. This bar is first stressed to a mean stress level σ_m and then subjected to a stress fluctuating over time, sometimes tensile and other times compressive (Fig. 1-12B). The amplitude σ_a of the stress is such that the bar is subjected to a maximum tensile stress less than the ultimate strength of the material. This reversible and periodic stress is applied until the bar fractures and the number of cycles N to fracture is recorded. This experiment is repeated on specimens having the same material properties by applying stresses of varying amplitude. A typical result of a fatigue test is plotted in Figure 1-12C on a diagram showing stress amplitude versus number of cycles to failure. For a given N, the corresponding stress value is called the fatigue strength of the material at that number of cycles. For a given stress level, N represents the fatigue life of the material. For some materials the stress amplitude versus number of cycles curve levels off. The stress σ_e at which the fatigue curve levels off is called the endurance limit of the material. Below the endurance limit, the material has a high probability of not failing in fatigue, regardless of how many cycles of stress are imposed on the material.

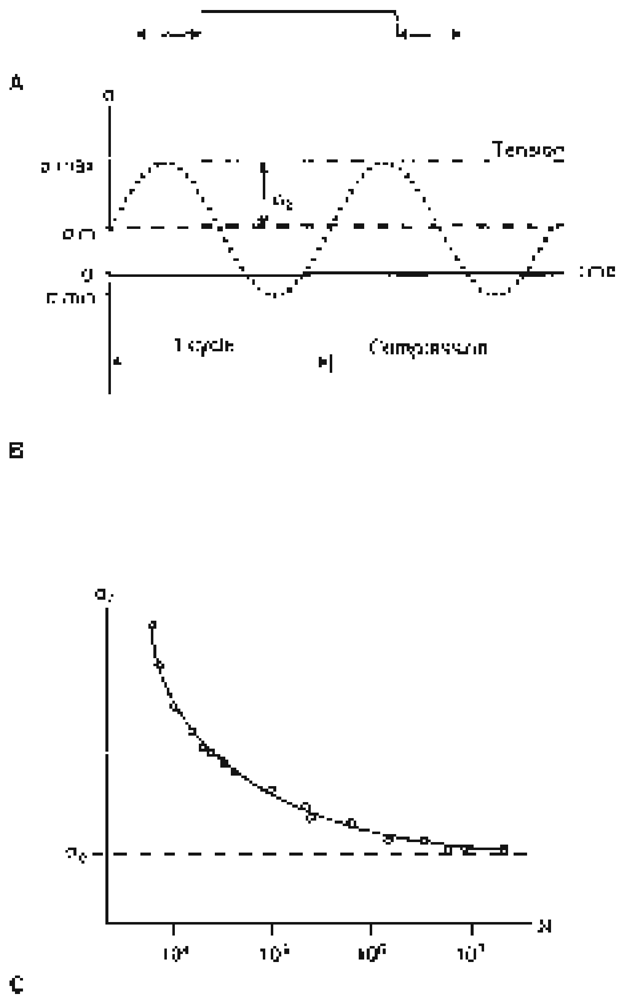


FIG. 1-12

Fatigue and endurance. Reprinted with permission from Özkaya, N. (1998). Biomechanics. In W.N. Rom, *Environmental and Occupational Medicine* (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.

The fatigue behavior of a material depends on several factors. The higher the temperature in which the material is used, the lower the fatigue strength. The fatigue behavior is sensitive to surface imperfections and the presence of discontinuities within the material that can cause stress concentrations. The fatigue failure starts with the creation of a small crack on the surface of the material, which can propagate under the effect of repeated loads, resulting in the rupture of the material.

Orthopaedic devices undergo repeated loading and unloading as a result of the activities of the patients and the actions of their muscles. Over a period of years, a weight-bearing prosthetic device or a fixation device can be subjected to a considerable number of cycles of stress reversals as a result of normal daily activity. This cyclic loading and unloading can cause fatigue failure of the device.

Basic Biomechanics of the Musculoskeletal System

Understanding even a simple task executed by the musculoskeletal system requires a broad, in-depth knowledge of various fields that may include motor control, neurophysiology, physiology, physics, and biomechanics. For example, based on the purpose and intention of a task and the sensory information gathered from the physical environment and orientation of the body and joints, the central nervous system plans a strategy for a task execution. According to the strategy adopted, muscles will be recruited to provide the forces and moments required for the movement and balance of the system. Consequently, the internal forces will be changed and soft tissues will experience different load conditions.

The purpose of this book is to present a well-balanced synthesis of information gathered from various disciplines, providing a basic understanding of biomechanics of the musculoskeletal system. The material presented here is organized to cover three areas of musculoskeletal biomechanics.

PART I: BIOMECHANICS OF TISSUES AND STRUCTURES

The material presented throughout this textbook provides an introduction to basic biomechanics of the musculoskeletal system. Part I includes chapters on the biomechanics of bone, articular cartilage, tendons and ligaments, peripheral nerves, and skeletal muscle. These are augmented with case studies to illustrate the important concepts for understanding the biomechanics of biological tissues.

PART II: BIOMECHANICS OF JOINTS

Part II of this textbook covers the major joints of the human body, from the spine to the ankle. Each chapter contains information about the structure and functioning of the joint, along with case studies illustrating the clinical diagnosis and management of joint injury and illness. The chapters are written by clinicians to provide an introductory level of knowledge about each joint system.

PART III: APPLIED BIOMECHANICS

A new section in the third edition of this book introduces important issues in applied biomechanics. These include the biomechanics of fracture fixation, arthroplasty, sitting, standing, and lying, and gait. It is important for the beginning student to understand the application of biomechanical principles in different clinical areas.

Summary

1 Biomechanics is a young and dynamic field of study based on the recognition that conventional engineering theories and methods can be useful for understanding and solving problems in physiology and medicine. Biomechanics considers the applications of classical mechanics to biological problems. The field of biomechanics flourishes from the cooperation among life scientists, physicians, engineers, and basic scientists. Such cooperation requires a certain amount of common vocabulary. An engineer must learn some anatomy and physiology, and medical personnel need to understand some basic concepts of physics and mathematics.

2 The information presented throughout this textbook is drawn from a large scholarship. The authors aim to introduce some of the basic concepts of biomechanics related to biological tissues and joints. The book does not intend to provide a comprehensive review of the literature, and readers are encouraged to consult the list of suggested reading below to supplement their knowledge. Some basic textbooks are listed here, and students should consult peer-reviewed journals for in-depth presentations of the latest research in specialty areas.

SUGGESTED READING

Black, J. (1988). Orthopaedic Biomaterials in Research and Practice. New York: Churchill Livingstone.

Bronzino, J.D. (Ed.) (1995). The Biomedical Engineering Handbook. Boca Raton, FL: CRC Press.

Burstein, A.H., & Wright, T.M. (1994). Fundamentals of Orthopaedic Biomechanics. Baltimore: Williams & Wilkins.

Chaffin, D.B., & Andersson, G.B.J. (1991). Occupational Biomechanics (2nd ed.). New York: John Wiley & Sons.

Fung, Y.C. (1981). Biomechanics: Mechanical Properties of Living Tissues. New York: Springer-Verlag.

Fung, Y.C. (1990). Biomechanics: Motion, Flow, Stress, and Growth. New York: Springer-Verlag.

Hay, J.G., & Reid, J.G. (1988). Anatomy, Mechanics and Human Motion (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kelly, D.L. (1971). Kinesiology: Fundamentals of Motion Description. Englewood Cliffs, NJ: Prentice-Hall.

Mow, V.C., & Hayes, W.C. (1997). Basic Orthopaedic Biomechanics (2nd ed.). New York: Raven Press.

Mow, V.C., Ratcliffe, A., & Woo, S.L-Y. (Eds.) (1990). Biomechanics of Diarthrodial Joints. New York: Springer-Verlag.

Nahum, A.M., & Melvin, J. (Eds.) (1985). The Biomechanics of Trauma. Norwalk, CT: Appleton-Century-Crofts.

Nordin, M., Andersson, G.B.J., & Pope, M.H. (Eds.) (1997). Musculoskeletal Disorders in the Workplace. Philadelphia: Mosby-Year Book.

Nordin, M., & Frankel, V.H. (Eds.) (1989). Basic Biomechanics of the Musculoskeletal System (2nd ed.). Philadelphia: Lea & Febiger.

Özkaya, N. (1998). Biomechanics. In W.N. Rom, Environmental and Occupational Medicine (3rd ed., pp. 1437–1454). New York: Lippincott-Raven.

Özkaya, N., & Nordin, M. (1999). Fundamentals of Biomechanics: Equilibrium, Motion, and Deformation (2nd ed.). New York: Springer-Verlag.

Schmid-Schönbein, G.W., Woo, S.L-Y., & Zweifach, B.W. (Eds.) (1985). Frontiers in Biomechanics. New York: Springer-Verlag.

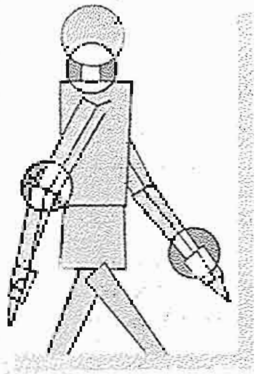
Skalak, R., & Chien, S. (Eds.) (1987). Handbook of Bioengineering. New York: McGraw-Hill.

Thompson, C.W. (1989). Manual of Structural Kinesiology (11th ed.). St. Louis, MO: Times Mirror/Mosby.

Williams, M., & Lissner, H.R. (1992). Biomechanics of Human Motion (3rd ed.). Philadelphia: Saunders.

Winter, D.A. (1990). Biomechanics and Motor Control of Human Behavior (2nd ed.). New York: John Wiley & Sons.

Winters, J.M., & Woo, S.L-Y. (Eds.) (1990). Multiple Muscle Systems. New York: Springer-Verlag.



APPENDIX 1

The System International d'Unites (SI)

Dennis R. Carter

The SI Metric System

The System International d'Unites (SI), the metric system, has evolved into the most exacting system of measures devised. In this section, the SI units of measurement used in the science of mechanics are described. SI units used in electrical and light sciences have been omitted for the sake of simplicity.

BASE UNITS

The SI units can be considered in three groups: 1. the base units, 2. the supplementary units, and 3. the derived units (Fig. App-1). The base units are a small group of standard measurements that have been arbitrarily defined. The base unit for length is the meter (m), and the base unit for mass is the kilogram (kg). The base units for time and temperature are the second (s) and the kelvin (K), respectively. Definitions of the base units have become increasingly sophisticated in response to the expanding needs and capabilities of the scientific community (Table App-1). For example, the meter is now defined in terms of the wavelength of radiation emitted from the krypton 86 atom.

SUPPLEMENTARY UNITS

The radian (rad) is a supplementary unit to measure plane angles. This unit, like the base units, is arbitrarily defined (Table App-1). Although the radian is the SI unit for plane angle, the unit of the degree has been retained for general use because it is firmly established and is widely used around the world. A degree is equivalent to $\pi/180$ rad.

DERIVED UNITS

Most units of the SI system are derived units, meaning that they are established from the base units in accordance with fundamental physical principles. Some of these units are expressed in terms of the base units from which they are derived. Examples are area, speed, and acceleration, which are expressed in the SI units of square meters (m^2), meters per second (m/s), and meters per second squared (m/s^2), respectively.

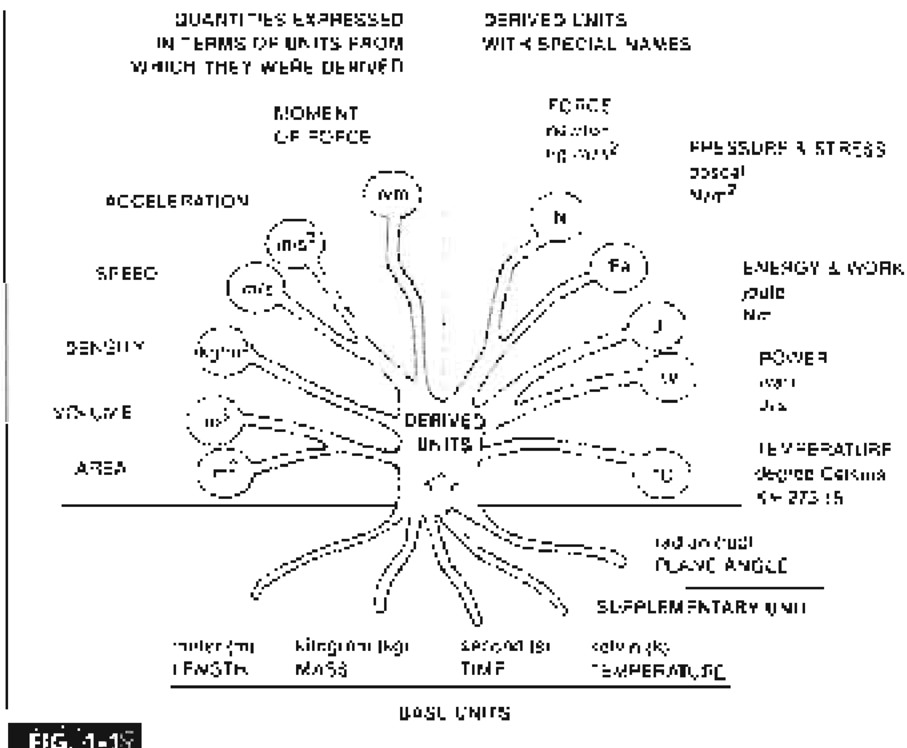


FIG. 1-17

The International System of Units.

Specially Named Units

Other derived units are similarly established from the base units but have been given special names (Fig App-1 and Table App-1). These units are defined through the use of fundamental equations of physical laws in conjunction with the arbitrarily defined SI base units. For example, Newton's second law of motion states that when a body that is free to move is subjected to a force, it will experience an acceleration proportional to that force and inversely proportional to its own mass. Mathematically, this principle can be expressed as:

force = mass × acceleration

The SI unit of force, the newton (N), is therefore defined in terms of the base SI units as:

$$1\ N = 1\ kg \times 1\ m/s^2$$

The SI unit of pressure and stress is the pascal (Pa). Pressure is defined in hydrostatics as the force divided by the area of force application. Mathematically, this can be expressed as:

pressure = force/area

The SI unit of pressure, the pascal, is therefore defined in terms of the base SI units as:

$$1\ Pa = 1\ N / 1\ m^2$$

Although the SI base unit of temperature is the kelvin, the derived unit of degree Celsius (°C or c) is much more commonly used. The degree Celsius is equivalent to the kelvin in magnitude, but the absolute value of the Celsius scale differs from that of the Kelvin scale such that °C = K − 273.15.

When the SI system is used in a wide variety of measurements, the quantities expressed in terms of the base, supplemental, or derived units may be either very large or very small. For example, the area on the head of a pin is an extremely small number when expressed in terms of square meters. Conversely, the weight of a whale is an extremely large number when expressed in terms of newtons. To accommodate the convenient representation of small or large quantities, a system of prefixes has been incorporated into the SI system (Table App-2). Each prefix has a fixed meaning and can be used with all SI units. When used with the name of the unit, the prefix indicates that the quantity described is being expressed in some multiple of

TABLE 1

Definitions of SI Units

Base SI Units	
meter (m)	The meter is the length equal to 1,650,763.73 wavelengths in vacuum of the radiation corresponding to the transition between the levels $2p_{10}$ and $5d_5$ of the krypton-86 atom
kilogram (kg)	The kilogram is the unit of mass and is equal to the mass of the international prototype of the kilogram
second (s)	The second is the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom
kelvin (K)	The kelvin, a unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water
Supplementary SI Unit	
radian (rad)	The radian is the plane angle between two radii of a circle that subtend on the circumference of an arc equal in length to the radius
Derived SI Units With Special Names	
newton (N)	The newton is that force which, when applied to a mass of 1 kilogram, gives it an acceleration of 1 meter per second squared. $1\ N = 1\ kg\ m/s^2$
pascal (Pa)	The pascal is the pressure produced by a force of 1 newton applied, with uniform distribution, over an area of 1 square meter. $1\ Pa = 1\ N/m^2$
joule (J)	The joule is the work done when the point of application of a force of 1 newton is displaced through a distance of 1 meter in the direction of the force. $1\ J = 1\ Nm$
watt (W)	The watt is the power that in 1 second gives rise to the energy of 1 joule. $1\ W = 1\ J/s$
degree Celsius (°C)	The degree Celsius is a unit of thermodynamic temperature and is equivalent to K − 273.15

TABLE 2

SI Multiplication Factors and Prefixes

Multiplication Factor	SI Prefix	SI Symbol
1 000 000 000 = 10^9	giga	G
1 000 000 = 10^6	mega	M
1 000 = 10^3	kilo	k
100 = 10^2	hecto	h
10 = 10	deca	da
.1 = 10^{-1}	deci	d
.01 = 10^{-2}	centi	c
.001 = 10^{-3}	milli	m
.000 001 = 10^{-6}	mikro	μ
.000 000 001 = 10^{-9}	nano	n
.000 000 000 001 = 10^{-12}	pico	p

Reprinted with permission from Ozkaya, N., & Nordin, M. (1999). Fundamentals of Biomechanics: Equilibrium, Motion, and Deformation (2nd ed.). New York: Springer-Verlag, p. 10.

ten times the unit used. For example, the millimeter (mm) is used to represent one thousandth (10^{-3}) of a meter and a gigapascal (Gpa) is used to denote one billion (10^9) pascals.

Standard Units Named for Scientists

One of the more interesting aspects of the SI system is its use of the names of famous scientists as standard units. In each case, the unit was named after a scientist in recognition of his contribution to the field in which that unit plays a major role. Table App-3 lists a number of SI units and the scientist for which each was named.

For example, the unit of force, the newton, was named in honor of the English scientist Sir Isaac Newton (1624–1727). He was educated at Trinity College at Cambridge and later returned to Trinity College as a professor of mathematics. Early in his career Newton made fundamental contributions to mathematics that formed the basis of differential and integral calculus. His other major discoveries were in the fields of optics, astronomy, gravitation, and mechanics. His work in gravitation was purportedly spurred by being hit on the head by an apple falling from a tree. It is perhaps poetic justice that the SI unit of one newton is approximately equivalent to the weight of a medium-sized apple. Newton was knighted in 1705 by Queen Mary for his monumental contributions to science.

TABLE 3

SI Units Named After Scientists

Symbol	Unit	Quantity	Scientist	Country of Birth	Dates
A	ampere	electric current	Ampère, André-Marie	France	1775–1836
C	coulomb	electric charge	Coulomb, Charles Augustin de	France	1736–1806
°C	degree celsius	temperature	Celsius, Anders	Sweden	1701–1744
F	farad	electric capacity	Faraday, Michael	England	1791–1867
H	henry	inductive resistance	Henry, Joseph	United States	1797–1867
Hz	hertz	frequency	Hertz, Heinrich Rudolph	Germany	1857–1894
J	joule	energy	Joule, James Prescott	England	1818–1889
K	kelvin	temperature	Thomson, William (Lord Kelvin)	England	1824–1907
N	newton	force	Newton, Sir Isaac	England	1642–1727
Ω	ohm	electric resistance	Ohm, Georg Simon	Germany	1787–1854
Pa	pascal	pressure/stress	Pascal, Blaise	France	1623–1662
S	siemens	electric conductance	Siemens, Karl Wilhelm (Sir William)	Germany (England)	1823–1883
T	tesla	magnetic flux density	Tesla, Nikola	Croatia (US)	1856–1943
V	volt	electrical potential	Volta, Count Alessandro	Italy	1745–1827
W	watt	power	Watt, James	Scotland	1736–1819
Wb	weber	magnetic flux	Weber, Wilhelm Eduard	Germany	1804–1891

BOX APP-1 Conversion of Units

Length

1 centimeter (cm) = 0.01 meter (m)

1 inch (in) = 0.0254 m

1 foot (ft) = 0.3048 m

1 yard (yd) = 0.9144 m

1 mile = 1609 m

1 angstrom (Å) = 10^{-10} m

Time

1 minute (min) = 60 seconds (s)

1 hour (h) = 3600 s

1 day (d) = 86400 s

Mass

1 pound mass (lbm) = 0.4536 kilogram (kg)

1 slug = 14.59 kg

Force

1 kilogram force (kgf) = 9.807 newton (N)

1 pound force (lbf) = 4.448 N

1 dyne (dyn) = 10^{-5} N

Pressure and Stress

1 kg/m·s² = 1 N/m² = 1 Pascal (Pa)

1 lb/in² (psi) = 6895 Pa

1 lbf/ft² (psf) = 92956 Pa

1 dyn/cm² = 0.1 Pa

Moment (Torque)

1 dyn-cm = 10^{-7} N-m

1 lbf-ft = 1.356 N-m

Work and Energy

1 kg·m²/s² = 1 N-m = 1 Joule (J)

1 dyne-cm = 1 erg = 10^{-7} J

1 lbf-ft = 1.356 J

Power

1 kg·m²/s³ = 1 J/s = 1 Watt (W)

1 horsepower (hp) = 550 lbf-ft/s = 746 W

Plane Angle

1 degree (°) = π/180 radian (rad)

1 revolution (rev) = 360°

1 rev = 2π rad = 6.283 rad

Temperature

°C = °K − 273.2

°C = 5 (°F − 32)/9

Reprinted with permission from Özkaya, N., & Nordin, M. (1999). Fundamentals of Biomechanics: Equilibrium, Motion, and Deformation (2nd ed.). New York: Springer-Verlag, p. 11.

The unit of pressure and stress, the pascal, was named after the French physicist, mathematician, and philosopher Blaise Pascal (1623–1662). Pascal conducted important investigations on the characteristics of vacuums and barometers and also invented a machine that would make mathematical calculations. His work in the area of hydrostatics helped lay the foundation for the later development of these scientific fields. In addition to his scientific pursuits, Pascal was passionately interested in religion and philosophy and thus wrote extensively on a wide range of subjects.

The base unit of temperature, the kelvin, was named in honor of Lord William Thomson Kelvin (1824–1907). Named William Thomson, he was educated at the University of Glasgow and at Cambridge University. Early in his career, Thomson investigated the thermal properties of steam at a scientific laboratory in Paris. At the age of 22, he returned to Glasgow to accept the chair of Natural Philosophy. His meeting with James Joule in 1847 stimulated interesting discussions on the nature of heat, which eventually led to the establishment of Thomson's absolute scale of temperature, the Kelvin scale. In recognition of Thomson's contributions to the field of thermodynamics, King Edward VII conferred on him the title of Lord Kelvin.

The commonly used unit of temperature, the degree Celsius, was named after the Swedish astronomer and inventor Anders Celsius (1701–1744).

Celsius was appointed professor of astronomy at the University of Uppsala at the age of 29 and remained at the university until his death 14 years later. In 1742, he described the centigrade thermometer in a paper prepared for the Swedish Academy of Sciences. The name of the centigrade temperature scale was officially changed to Celsius in 1948.

Converting to SI From Other Units of Measurement

Box App-1 contains the formulae for the conversion of measurements expressed in English and non-SI metric units into SI units. One fundamental source of confusion in converting from one system to another is that two basic types of measurement systems exist. In the "physical" system (such as SI), the units of length, time, and mass are arbitrarily defined, and other units (including force) are derived from these base units. In "technical" or "gravitational" systems (such as the English system), the units of length, time, and *force* are arbitrarily defined, and other units (including mass) are derived from these base units. Because the units of force in gravitational systems are in fact the *weights* of standard masses, conversion to SI is dependent on the acceleration of mass due to the Earth's gravity. By international agreement, the acceleration due to gravity is 9.806650 m/s². This value has been used in establishing some of the conversion factors in Box App-1.

REFERENCES

Lerner, J.L. (1977). *SI Metric Handbook*. New York: Charles Scribner's Sons.

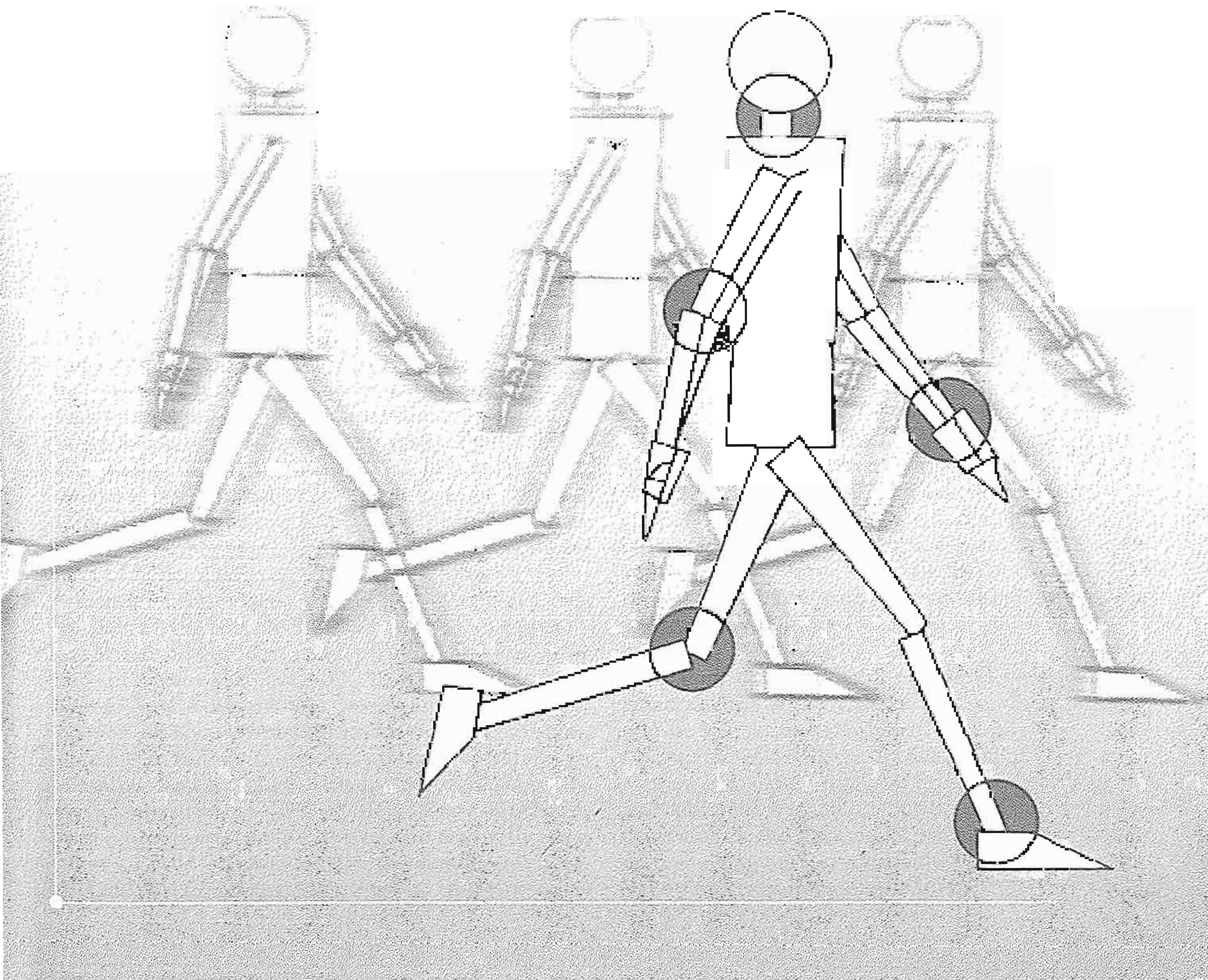
Özkaya, N., & Nordin, M. (1999). *Fundamentals of Biomechanics: Equilibrium, Motion, and Deformation* (2nd ed.). New York: Springer-Verlag.

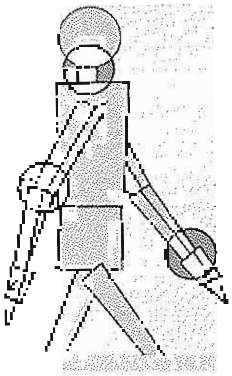
Pennycuick, C.J. (1974). *Handy Matrices of Unit Conversion Factors for Biology and Mechanics*. New York: John Wiley & Sons.

World Health Organization. (1977). *The SI for the Health Professions*. Geneva: WHO.

I

Biomechanics of Tissue and Structures of the Musculoskeletal System





CHAPTER 2

Biomechanics of Bone

Victor H. Frankel, Margareta Nordin

Introduction

The purpose of the skeletal system is to protect internal organs, provide rigid kinematic links and muscle attachment sites, and facilitate muscle action and body movement. Bone has unique structural and mechanical properties that allow it to carry out these roles. Bone is among the body's hardest structures; only dentin and enamel in the teeth are harder. It is one of the most dynamic and metabolically active tissues in the body and remains active throughout life. A highly vascular tissue, it has an excellent capacity for self-repair and can alter its properties and configuration in response to changes in mechanical demand. For example, changes in bone density are commonly observed after periods of disuse and of greatly increased use; changes in bone shape are noted during fracture healing and after certain operations. Thus, bone adapts to the mechanical demands placed on it.

This chapter describes the composition and structure of bone tissue, the mechanical properties of bone, and the behavior of bone under different loading conditions. Various factors that affect the mechanical behavior of bone in vitro and in vivo also are discussed.

Bone Composition and Structure

Bone tissue is a specialized connective tissue whose solid composition suits it for its supportive and protective roles. Like other connective tissues, it consists of cells and an organic extracellular matrix of fibers and ground substance produced by the cells. The distinguishing feature of bone is its high content of inorganic materials, in the form of mineral salts, that combine intimately with the organic matrix (Buckwalter et al., 1995). The inorganic component of bone makes the tissue hard and rigid, while the organic component gives bone its flexibility and resilience. The composition of bone differs depending on site, animal age, dietary history, and the presence of disease (Kaplan et al., 1994).

In normal human bone, the mineral or inorganic portion of bone consists primarily of calcium and phosphate, mainly in the form of small crystals resembling synthetic hydroxyapatite crystals with the composition $Ca_{10}(PO_4)_6(OH)_2$. These minerals, which account for 60 to 70% of its dry weight, give bone its solid consistency. Water accounts for 5 to 8% and the organic matrix makes up the remainder of the tissue. Bone serves as a reservoir for essential minerals in the body, particularly calcium.

Bone mineral is embedded in variously oriented fibers of the protein collagen, the fibrous portion of the extracellular matrix—the inorganic matrix. Collagen fibers (type I) are tough and pliable, yet they resist stretching and have little extensibility. Collagen composes approximately 90% of the extracellular matrix and accounts for approximately 25 to 30% of the dry weight of bone. A universal building block of the body, collagen also is the chief fibrous component of other skeletal structures. (A detailed description of the microstructure and mechanical behavior of collagen is provided in Chapters 3 and 4.)

The gelatinous ground substance surrounding the mineralized collagen fibers consists mainly of protein polysaccharides, or glycosaminoglycans (GAGs), primarily in the form of complex macromolecules called proteoglycans (PGs). The GAGs serve as a cementing substance between layers of mineralized collagen fibers. These GAGs, along with various noncollagenous glycoproteins, constitute approximately 5% of the extracellular matrix. (The structure of PGs, which are vital components of articular cartilage, is described in detail in Chapter 3.)

Water is fairly abundant in live bone, accounting for up to 25% of its total weight. Approximately 85% of the water is found in the organic matrix, around the collagen fibers and ground substance, and in the hydration shells surrounding the bone crystals. The other 15% is located in the canals and cavities that house bone cells and carry nutrients to the bone tissue.

At the microscopic level, the fundamental structural unit of bone is the osteon, or haversian system (Fig. 2-1). At the center of each osteon is a small channel, called a haversian canal, that contains blood vessels and nerve fibers. The osteon itself consists of a concentric series of layers (lamellae) of mineralized matrix surrounding the central canal, a configuration similar to growth rings in a tree trunk.

Along the boundaries of each layer, or lamella, are small cavities known as lacunae, each containing one bone cell, or osteocyte (Fig. 2-1C). Numerous small channels, called canaliculi, radiate from each lacuna, connecting the lacunae of adjacent lamellae and ultimately reaching the haversian canal. Cell processes extend from the osteocytes into the canaliculi, allowing nutrients from the blood vessels in the haversian canal to reach the osteocytes.

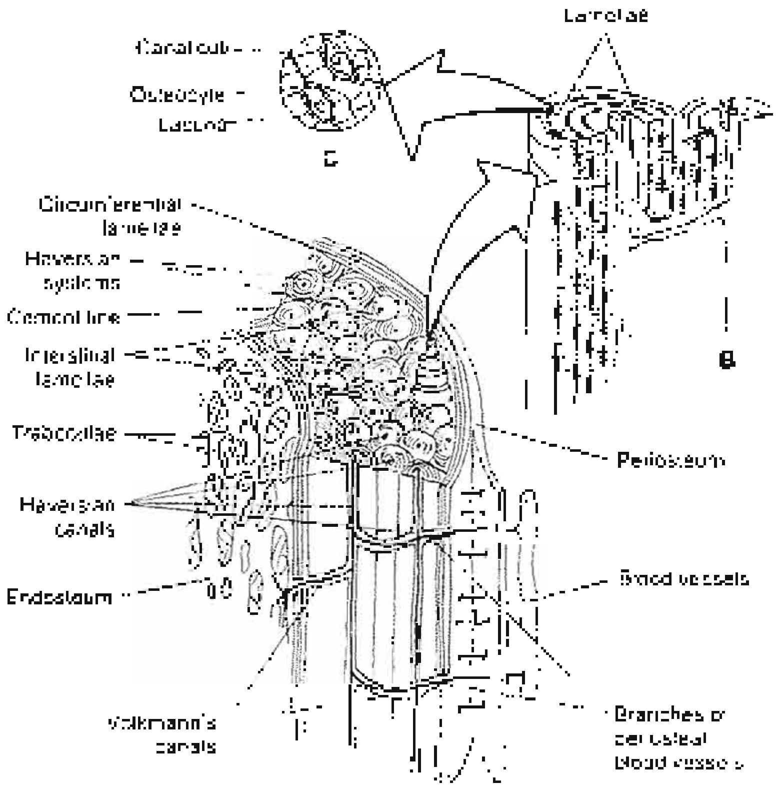


FIG. 2-1

A. The fine structure of bone is illustrated schematically in a section of the shaft of a long bone depicted without inner marrow. The osteons, or haversian systems, are apparent as the structural units of bone. The haversian canals are in the center of the osteons, which form the main branches of the circulatory network in bone. Each osteon is bounded by a cement line. One osteon is shown extending from the bone (20×). *Adapted from Bassett, C.A.L. (1965). Electrical effects in bone. Sci Am, 213,18.* **B.** Each osteon consists of lamellae, concentric rings composed of a mineral matrix surrounding the haversian canal. *Adapted from Tortora G.J., & Anagnostakos, N.P. (1984). Principles of Anatomy and Physiology (4th ed.). New York: Harper & Row.* **C.** Along the boundaries of the lamellae are small cavities known as lacunae, each of which contains a single bone cell, or osteocyte. Radiating from the lacunae are tiny canals, or canaliculi, into which the cytoplasmic processes of the osteocytes extend. *Adapted from Tortora G.J., & Anagnostakos, N.P. (1984). Principles of Anatomy and Physiology (4th ed.). New York: Harper & Row.*

At the periphery of each osteon is a cement line, a narrow area of cement-like ground substance composed primarily of GAGs. The canaliculi of the osteon do not pass this cement line. Like the canaliculi, the collagen fibers in the bone matrix interconnect from one lamella to another within an osteon but do not cross the cement line. This intertwining of collagen fibers within the osteon undoubtedly increases the bone's resistance to mechanical stress and probably explains why the cement line is the weakest portion of the bone's microstructure.

A typical osteon is approximately 200 micrometers (μ) in diameter. Hence, every point in the osteon is no more than 100 μm from the centrally located blood supply. In the long bones, the osteons usually run longitudinally, but they branch frequently and anastomose extensively with each other.

Interstitial lamellae span the regions between complete osteons (Fig. 2-1A). They are continuous with the osteons and consist of the same material in a different geometric configuration. As in the os-

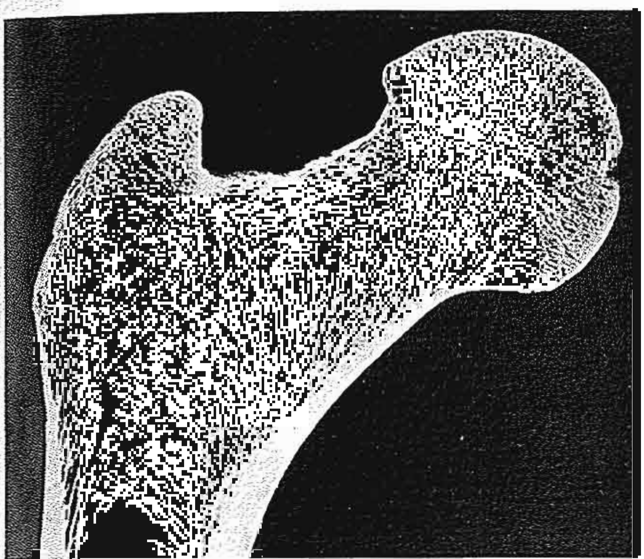


FIG. 2-2

Frontal longitudinal section through the head, neck, greater trochanter, and proximal shaft of an adult femur. Cancellous bone, with its trabeculae oriented in a lattice, lies within the shell of cortical bone. *Reprinted with permission from Gray, H. (1985). Anatomy of the Human Body (13th American ed.). Philadelphia: Lea & Febiger.*

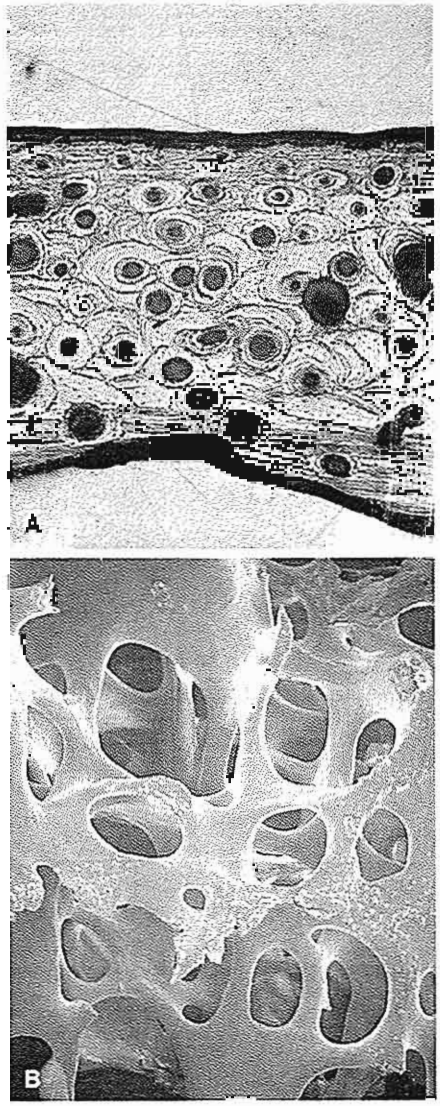


FIG. 2-3

A, Reflected-light photomicrograph of cortical bone from a human tibia (40×). B, Scanning electron photomicrograph of cancellous bone from a human tibia (30×). *Reprinted with permission from Carter, D.R., & Hayes, W.C. (1977). Compact bone fatigue damage: A microscopic examination. Clin Orthop, 127, 265.*

teons, no point in the interstitial lamellae is farther than 100 µm from its blood supply. The interfaces between these lamellae contain an array of lacunae in which osteocytes lie and from which canaliculi extend.

At the macroscopic level, all bones are composed of two types of osseous tissue: cortical, or compact, bone and cancellous, or trabecular, bone (Fig. 2-2). Cortical bone forms the outer shell, or cortex, of the bone and has a dense structure similar to that of ivory. Cancellous bone within this shell is composed of thin plates, or trabeculae, in a loose mesh structure; the interstices between the trabeculae are filled with red marrow (Fig. 2-3). Cancellous bone tissue is arranged in concentric lacunae-containing lamellae but does not contain haversian canals. The osteocytes receive nutrients through canaliculi from blood vessels passing through the red marrow. Cortical bone always surrounds cancellous bone, but the relative quantity of each type varies among bones and within individual bones according to functional requirements.

On a microscopic level, bone consists of woven and lamellar bone (Fig. 2-4). Woven bone is considered immature bone. This type of bone is found in the embryo, in the newborn, in the fracture callus, and in the metaphyseal region of growing bone as well as in tumors, osteogenesis imperfecta, and pagetic bone. Lamellar bone begins to form 1 month after birth and ue pols replaces woven bone. Lamellar bone is therefore a more mature bone.

All bones are surrounded by a dense fibrous membrane called the periosteum (Fig. 2-1B). Its outer layer is permeated by blood vessels (Fig. 2-5) and nerve fibers that pass into the cortex via Volkmann's canals, connecting with the haversian canals and extending to the cancellous bone. An inner, osteogenic layer contains bone cells responsible for generating new bone during growth and repair (osteoblasts). The periosteum covers the entire bone except for the joint surfaces, which are covered with articular cartilage. In the long bones, a thinner membrane, the endosteum, lines the central (medullary) cavity, which is filled with yellow fatty marrow. The endosteum contains osteoblasts and also giant multinucleated bone cells called osteoclasts, both of which play important roles in the remodeling and resorption of bone.

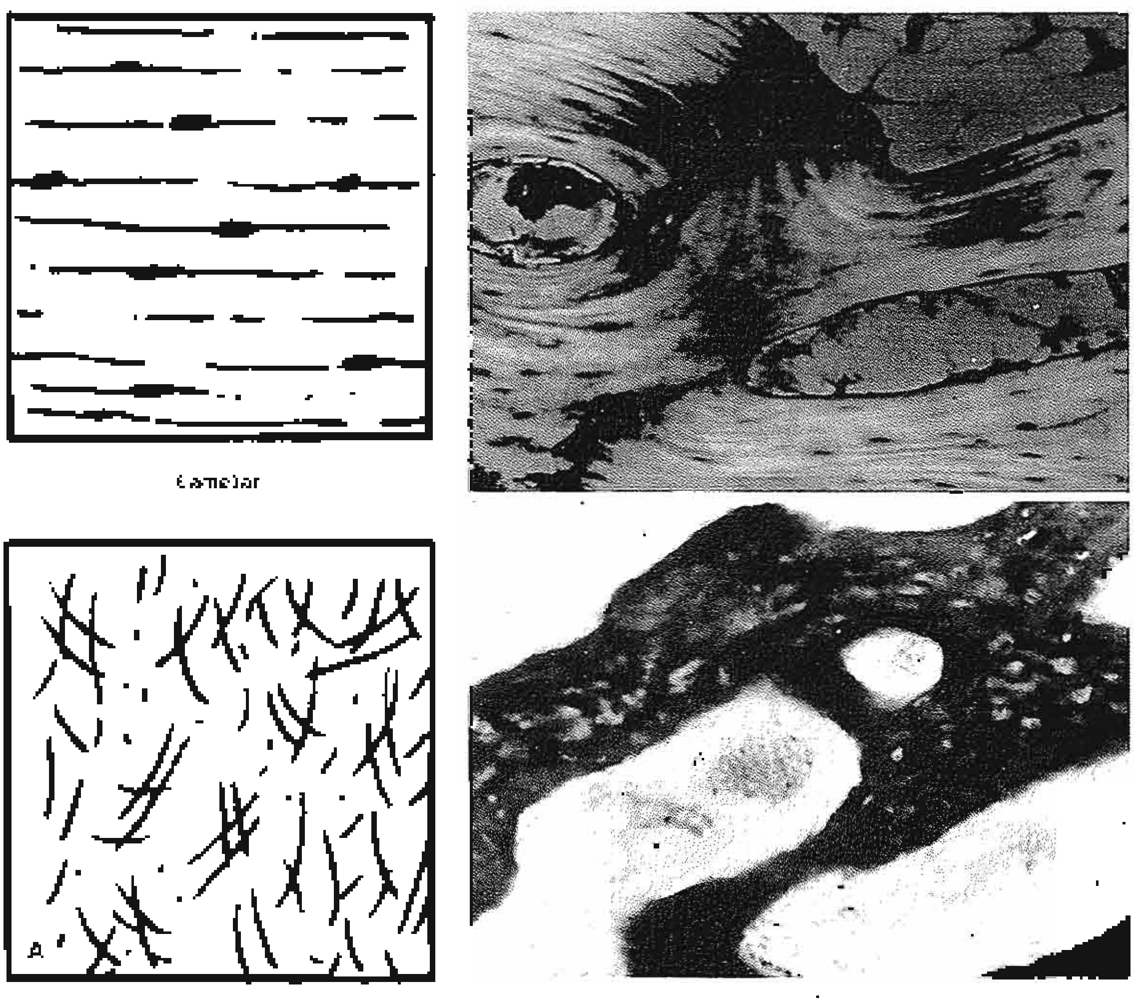


FIG. 2-4

Schematic drawing and photomicrographs of lamellar and woven bone. *Adapted from Kaplan, F.S., Hayes, W.C., Keaveny, T.M., et al. (1994). Form and function of bone. In S.R. Simon (Ed.). Orthopaedic Basic Science (pp. 129, 130). Rosemont, IL: AAOS.*

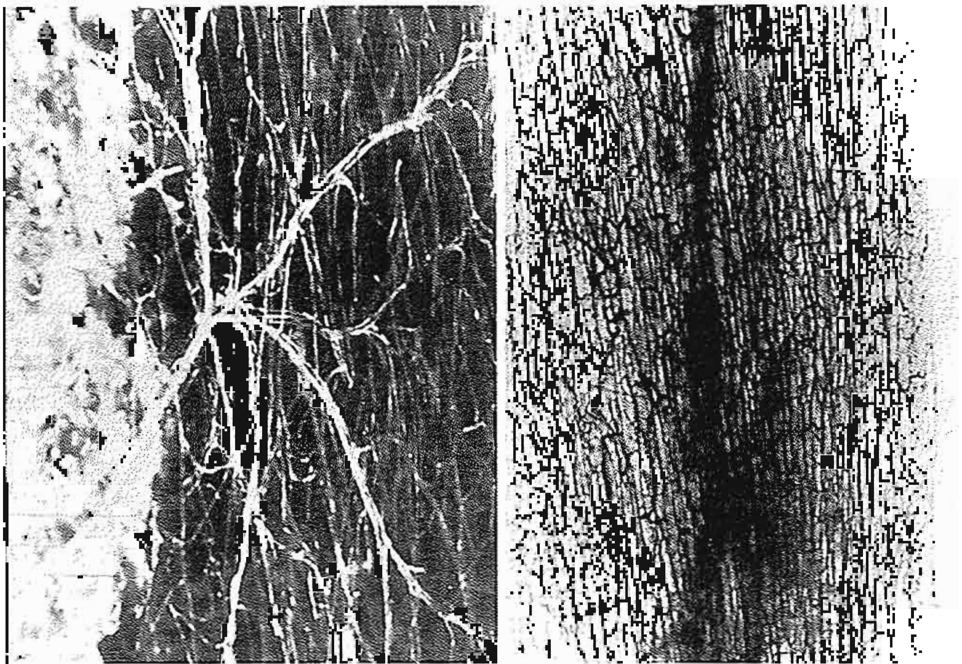


FIG. 2-5

Photomicrograph showing the vasculature of cortical bone. *Adapted from Kaplan, F.S., Hayes, W.C., Keaveny, T.M., et al. (1994). Form and function of bone. In S.R. Simon (Ed.). Orthopaedic Basic Science (p. [illegible]). Rosemont, IL: AAOS.*

Biomechanical Properties of Bone

Biomechanically, bone tissue may be regarded as a two-phase (biphasic) composite material, with the mineral as one phase and the collagen and ground substance as the other. In such materials (a nonbiological example is fiberglass) in which a strong, brittle material is embedded in a weaker, more flexible one, the combined substances are stronger for their weight than is either substance alone (Bassett, 1965).

Functionally, the most important mechanical properties of bone are its strength and stiffness. These and other characteristics can best be understood for bone, or any other structure, by examining its behavior under loading, that is, under the influence of externally applied forces. Loading causes a deformation, or a change in the dimensions, of the structure. When a load in a known direction is imposed on a structure, the deformation of that structure can be measured and plotted on a load-deformation curve. Much information about the strength, stiffness, and other mechanical properties of the structure can be gained by examining this curve.

A hypothetical load-deformation curve for a somewhat pliable fibrous structure, such as a long bone, is shown in Figure 2-6. The initial (straight line) portion of the curve, the elastic region, reveals the elasticity of the structure, that is, its capacity for returning to its original shape after the load is removed. As the load is applied, deformation occurs but is not permanent; the structure recovers its original shape when unloaded. As loading continues, the outermost fibers of the structure begin to yield at some point. This yield point signals the elastic limit of the structure. As the load exceeds this limit, the structure exhibits plastic behavior, reflected in the second (curved) portion of the curve, the plastic region. The structure will no longer return to its original dimensions when the load has been released; some residual deformation will be permanent. If loading is progressively increased, the structure will fail at some point (bone will fracture). This point is indicated by the ultimate failure point on the curve.

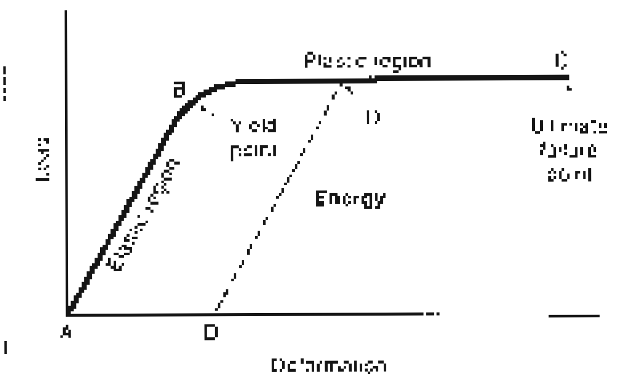


FIG. 2-6

Load-deformation curve for a structure composed of a somewhat pliable material. If a load is applied within the elastic range of the structure (A to B on the curve) and is then released, no permanent deformation occurs. If loading is continued past the yield point (B) and into the structure's plastic range (B to C on the curve) and the load is then released, permanent deformation results. The amount of permanent deformation that occurs if the structure is loaded to point D in the plastic region and then unloaded is represented by the distance between A and D'. If loading continues within the plastic range, an ultimate failure point (C) is reached.

Three parameters for determining the strength of a structure are reflected on the load-deformation curve: 1, the load that the structure can sustain before failing; 2, the deformation that it can sustain before failing; and 3, the energy that it can store before failing. The strength in terms of load and deformation, or ultimate strength, is indicated on the curve by the ultimate failure point. The strength in terms of energy storage is indicated by the size of the area under the entire curve. The larger the area, the greater the energy that builds up in the structure as the load is applied. The stiffness of the structure is indicated by the slope of the curve in the elastic region. The steeper the slope, the stiffer the material.

The load-deformation curve is useful for determining the mechanical properties of whole structures such as a whole bone, an entire ligament or tendon, or a metal implant. This knowledge is helpful in the study of fracture behavior and repair, the response of a structure to physical stress, or the effect of various treatment programs. However, characterizing a bone or other structure in terms of the material of which it is composed, independent of its geometry, requires standardization of the testing conditions and the size and shape of the test specimens. Such standardized testing is useful for comparing the mechanical properties of two or more materials, such as the relative strength of bone and tendon tissue or the relative stiffness of various materials used in prosthetic implants. More precise units of measurement can be used when standardized samples are tested—that is, the load per unit of area of the sample (stress) and the amount of deformation in terms of the percentage of change in the sample's dimensions (strain). The curve generated is a stress-strain curve.

Stress is the load, or force, per unit area that develops on a plane surface within a structure in response to externally applied loads. The three units most commonly used for measuring stress in standardized samples of bone are newtons per centimeter squared (N/cm^2); newtons per meter squared, or pascals (N/m^2, Pa); and meganewtons per meter squared, or megapascals (MN/m^2, MPa).

Strain is the deformation (change in dimension) that develops within a structure in response to externally applied loads. The two basic types of strain are linear strain, which causes a change in the length of the specimen, and shear strain, which causes a change in the angular relationships within the structure. Linear strain is measured as the amount of linear deformation (lengthening or shortening) of the sample divided by the sample's original length. It is a nondimensional parameter expressed as a percentage (e.g., centimeter per centimeter). Shear strain is measured as the amount of angular change (γ) in a right angle lying in the plane of interest in the sample. It is expressed in radians (one radian equals approximately 57.3°) (International Society of Biomechanics, 1987).

Stress and strain values can be obtained for bone by placing a standardized specimen of bone tissue in a testing jig and loading it to failure (Fig. 2-7). These values can then be plotted on a stress-strain curve (Fig. 2-8). The regions of this curve are similar to those of the load-deformation curve. Loads in the elastic region do not cause permanent deformation, but once the yield point is exceeded, some deformation is permanent. The strength of the material in terms of energy storage is represented by the area under the entire curve. The stiffness is represented by the slope of the curve in the elastic region. A value for stiffness is obtained by dividing the stress at a point in the elastic (straight line) portion of the curve by the strain at that point. This value is called the modulus of elasticity (Young's modulus). Young's modulus (E) is derived from the relationship between stress (σ) and strain (ϵ):

$$E = \sigma/\epsilon$$

The elasticity of a material or the Young's modulus E is equal to the slope of the stress (σ) and strain (ϵ) diagram in the elastic linear region. E represents the stiffness of the material, such that the higher the elastic modulus or Young's modulus, the stiffer the material (Özkaya & Nordin, 1999).

Mechanical properties differ in the two bone types. Cortical bone is stiffer than cancellous bone, withstanding greater stress but less strain before failure. Cancellous bone in vitro may sustain up to 50% of strains before yielding, while cortical bone yields and fractures when the strain exceeds 1.5 to 2.0%. Because of its porous structure, cancellous bone has a large capacity for energy storage (Keaveny & Hayes, 1993). The physical difference between the two bone tissues is quantified in terms of the apparent density of bone, which is defined as the mass of bone tissue present in a unit of bone volume (gram per cubic centimeter [g/cc]). Figure 2-9 depicts typical stress-strain qualities of cortical and trabecular bone with different bone densities tested under similar conditions. In general, it is not enough to describe bone strength with a single number. A better way is to examine the stress-strain curve for the bone tissue under the circumstances tested.

To better understand the relationship of bone to other materials, schematic stress-strain curves for bone, metal, and glass illustrate the differences in mechanical behavior among these materials (Fig. 2-10). The variations in stiffness are reflected in the different slopes of the curves in the elastic region. Metal has the steepest slope and is thus the stiffest material.

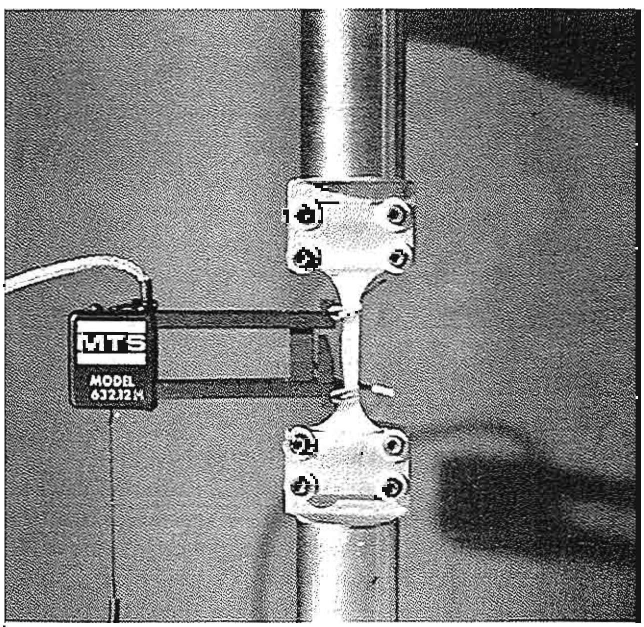


FIG. 2-7

Standardized bone specimen in a testing machine. The strain in the segment of bone between the two gauge arms is measured with a strain gauge. The stress is calculated from the total load measured. *Courtesy of Dennis R. Carter, PhD.*

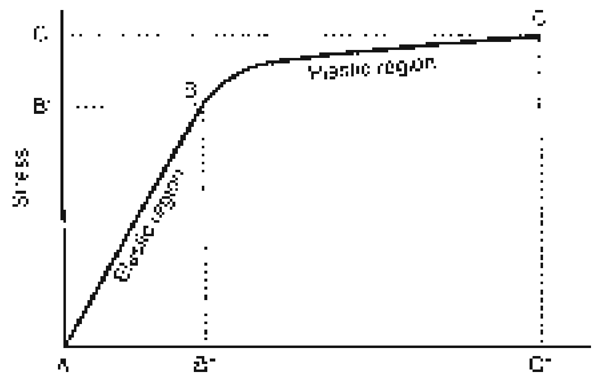


FIG. 2-8

Stress-strain curve for a cortical bone sample tested in tension (pulled). Yield point (B): point past which some permanent deformation of the bone sample occurred. Yield stress (B'): load per unit area sustained by the bone sample before plastic deformation took place. Yield strain (B"): amount of deformation withstood by the sample before plastic deformation occurred. The strain at any point in the elastic region of the curve is proportional to the stress at that point. Ultimate failure point (C): the point past which failure of the sample occurred. Ultimate stress (C'): load per unit area sustained by the sample before failure. Ultimate strain (C"): amount of deformation sustained by the sample before failure.

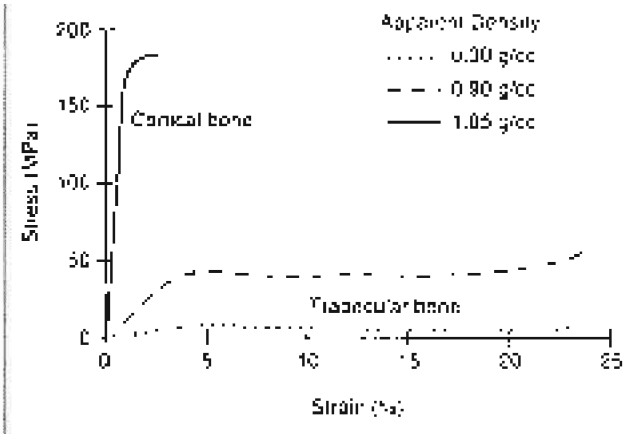


FIG. 2-9

Example of stress-strain curves of cortical and trabecular bone with different apparent densities. Testing was performed in compression. The figure depicts the difference in mechanical behavior for the two bone structures. *Reprinted with permission from Keaveny, T.M., & Hayes, W.C. (1993). Mechanical properties of cortical and trabecular bone. Bone, 7, 285–344.*

The elastic portion of the curve for glass and metal is a straight line, indicating linearly elastic behavior; virtually no yielding takes place before the yield point is reached. By comparison, precise testing of cortical bone has shown that the elastic portion of the curve is not straight but instead slightly curved, indicating that bone is not linearly elastic in its behavior but yields somewhat during loading in the elastic region (Bonefield & Li, 1967). Table 2-1 depicts the mechanical properties of selected biomaterials for comparison. Materials are classified as brittle or ductile depending on the extent of deformation before failure. Glass is a typical brittle material, and soft metal is a typical ductile material. The difference in the amount of deformation is reflected in the fracture surfaces of the two materials (Fig. 2-11). When pieced together after fracture, the ductile material will not conform to its original shape whereas the brittle material will. Bone exhibits more brittle or more ductile behavior depending on its age (younger bone being more ductile) and the rate at which it is loaded (bone being more brittle at higher loading speeds).

After the yield point is reached, glass deforms very little before failing, as indicated by the absence of a plastic region on the stress-strain curve. By contrast, metal exhibits extensive deformation before failing, as indicated by a long plastic region on the curve. Bone also deforms before failing but to a much lesser extent than metal. The difference in the plastic behavior of metal and bone is the result of differences in micromechanical events at yield. Yielding in metal (tested in tension, or pulled) is caused by plastic flow and the formation of plastic slip lines; slip lines are formed when the molecules of the lattice structure of metal dislocate. Yielding in bone (tested in tension) is caused

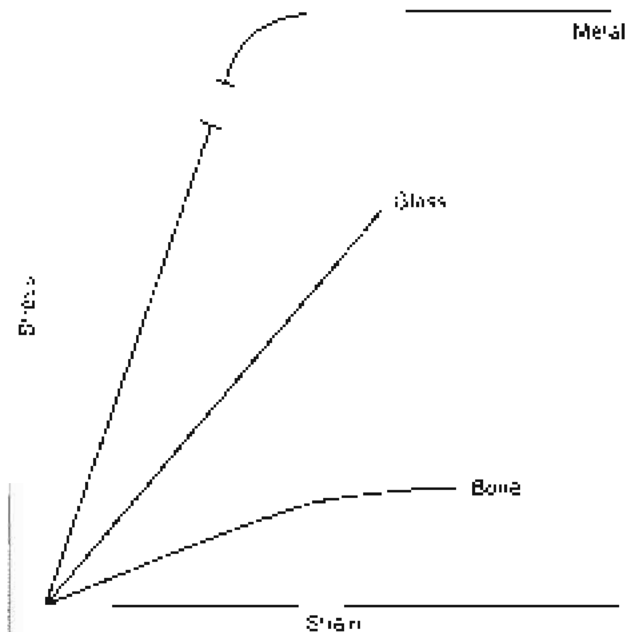


FIG. 2-10

Schematic stress-strain curves for three materials. Metal has the steepest slope in the elastic region and is thus the stiffest material. The elastic portion of the curve for metal is a straight line, indicating linearly elastic behavior. The fact that metal has a long plastic region indicates that this typical ductile material deforms extensively before failure. Glass, a brittle material, exhibits linearly elastic behavior but fails abruptly with little deformation, as indicated by the lack of a plastic region on the stress-strain curve. Bone possesses both ductile and brittle qualities demonstrated by a slight curve in the elastic region, which indicates some yielding during loading within this region.

TABLE 2-1

Mechanical Properties of Selected Biomaterials

	Ultimate Strength (MPa)	Modulus (GPa)	Elongation (%)
Metals			
Co-Cr Alloy			
Cast	600	220	8
Forged	950	220	15
Stainless steel	850	210	10
Titanium	900	110	15
Polymers			
Bone cement	20	2.0	2–4
Ceramic			
Alumina	300	350	<2
Biological			
Cortical bone	100–150	10–15	1–3
Trabecular bone	8–50		2–4
Tendon, ligament	20–35	2.0–4.0	10–25

Adapted from [illegible] 1989 [illegible] Biomaterials [illegible] Spector [illegible] [illegible] New York: McGraw-Hill.

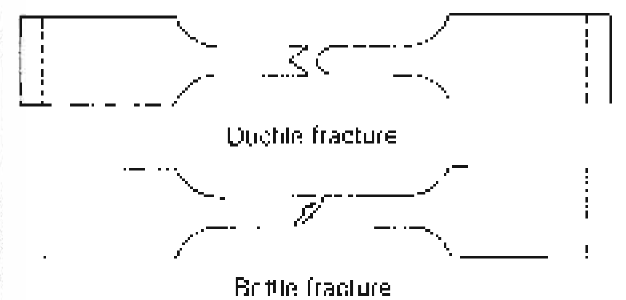


FIG. 2-11

Fracture surfaces of samples of a ductile and a brittle material. The broken lines on the ductile material indicate the original length of the sample, before it deformed. The brittle material deformed very little before fracture.

by debonding of the osteons at the cement lines and microfracture (Fig. 2-12), while yielding in bone as a result of compression is indicated by cracking of the osteons (Fig. 2-13).

Because the structure of bone is dissimilar in the transverse and longitudinal directions, it exhibits different mechanical properties when loaded along different axes, a characteristic known as anisotropy.

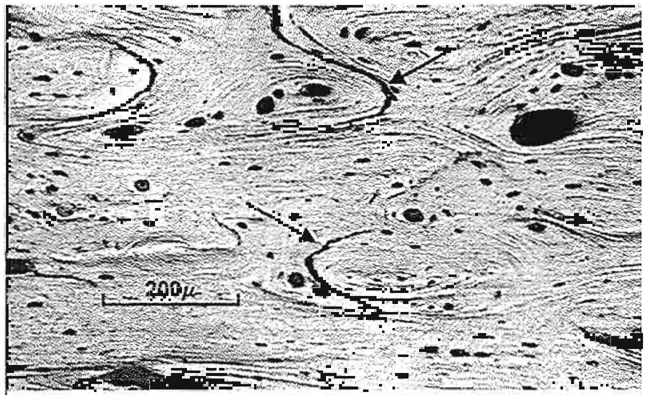


FIG. 2-12

Reflected light photomicrograph of a human cortical bone specimen tested in tension (30×). Arrows indicate debonding at the cement lines and pulling out of the osteons. *Courtesy of Dennis R. Carter, Ph.D.*

FIG. 2-13

Scanning electron photomicrograph of a human cortical bone specimen tested in compression (30×). Arrows indicate oblique cracking of the osteons. *Courtesy of Dennis R. Carter, Ph.D.*

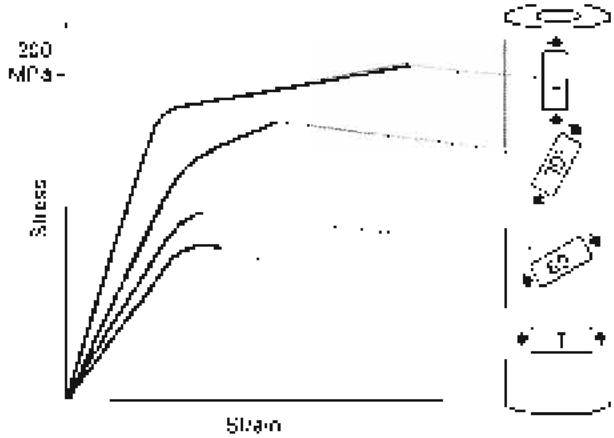


FIG. 2-14

Anisotropic behavior of cortical bone specimens from a human femoral shaft tested in tension (pulled) in four directions: longitudinal (L), tilted 30° with respect to the neutral axis of the bone, tilted 60°, and transverse (T). Data from Frankel, V.H., & Burstein, A.H. (1970). Orthopaedic Biomechanics. Philadelphia: Lea & Febiger.

Figure 2-14 shows the variations in strength and stiffness for cortical bone samples from a human femoral shaft, tested in tension in four directions (Frankel & Burstein, 1970; Carter, 1978). The values for both parameters are highest for the samples loaded in the longitudinal direction. Figures 2-9 and 2-15 show trabecular bone strength and stiffness tested in two directions: compression and tension. Trabecular or cancellous bone is approximately 25% as dense, 5 to 10% as stiff, and five times as ductile as cortical bone.

Although the relationship between loading patterns and the mechanical properties of bone throughout the skeleton is extremely complex, it generally can be said that bone strength and stiffness are greatest in the direction in which daily loads are most commonly imposed.

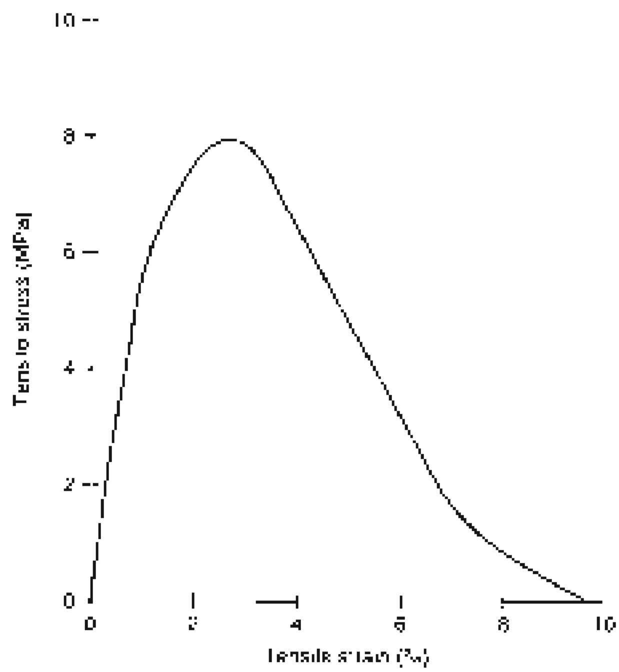


FIG. 2-15

Example of tensile stress-strain behavior of trabecular bone tested in the longitudinal axial direction of the bone. *Adapted from Gibson, L.J., & Ashby, M.F. (1988). Cellular Solids: Structure and Properties. New York: Pergamon Press.*

Biomechanical Behavior of Bone

The mechanical behavior of bone—its behavior under the influence of forces and moments—is affected by its mechanical properties, its geometric characteristics, the loading mode applied, direction of loading, rate of loading, and frequency of loading.

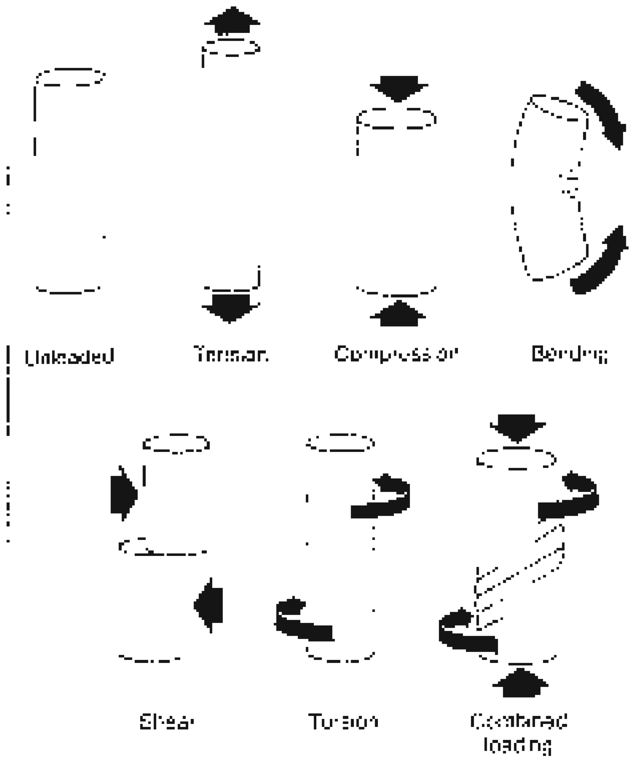


FIG. 2-16

Schematic representation of various loading modes.

BONE BEHAVIOR UNDER VARIOUS LOADING MODES

Forces and moments can be applied to a structure in various directions, producing tension, compression, bending, shear, torsion, and combined loading (Fig. 2-16). Bone in vivo is subjected to all of these loading modes. The following descriptions of these modes apply to structures in equilibrium (at rest or moving at a constant speed); loading produces an internal, deforming effect on the structure.

Tension

During tensile loading, equal and opposite loads are applied outward from the surface of the structure, and tensile stress and strain result inside the structure. Tensile stress can be thought of as many small forces directed away from the surface of the structure. Maximal tensile stress occurs on a plane perpendicular to the applied load (Fig. 2-17). Under tensile loading, the structure lengthens and narrows.

Clinically, fractures produced by tensile loading are usually seen in bones with a large proportion of cancellous bone. Examples are fractures of the base of the fifth metatarsal adjacent to the attachment of the peroneus brevis tendon and fractures of the calcaneus adjacent to the attachment of the Achilles tendon. Figure 2-18 shows a tensile fracture through the calcaneus; intense contraction of the triceps surae muscle produces abnormally high tensile loads on the bone.

Compression

During compressive loading, equal and opposite loads are applied toward the surface of the structure and compressive stress and strain result inside the structure. Compressive stress can be thought of as many small forces directed into the surface of the structure. Maximal compressive stress occurs on a plane perpendicular to the applied load (Fig. 2-19). Under compressive loading, the structure shortens and widens.

Clinically, compression fractures are commonly found in the vertebrae, which are subjected to high compressive loads. These fractures are most often seen in the elderly with osteoporotic bone tissue. Figure 2-20 shows the shortening and widening

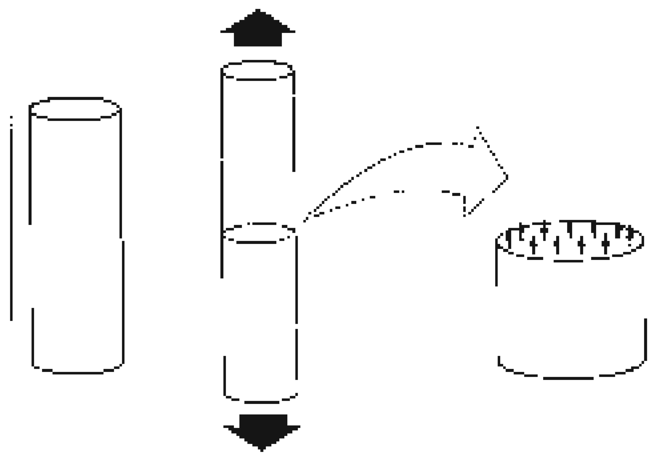


FIG. 2-17

Tensile loading.

FIG. 2-18

Tensile fracture through the calcaneus produced by strong contraction of the triceps surae muscle during a tennis match. Courtesy of Robert A. Winquist, MD.

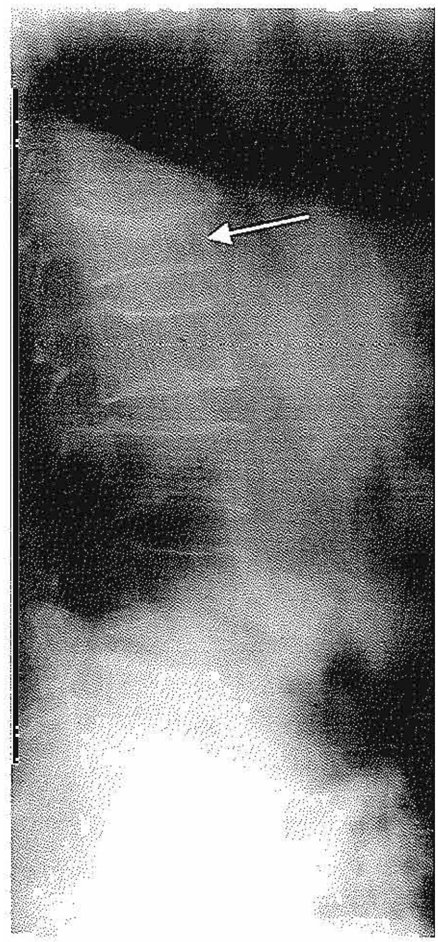


FIG. 2-20

Compression fracture of a human first lumbar vertebra. The vertebra has shortened and widened.

that takes place in a human vertebra subjected to a high compressive load. In a joint, compressive loading to failure can be produced by abnormally strong contraction of the surrounding muscles. An example of this effect is presented in Figure 2-21; bilateral subcapital fractures of the femoral neck were sustained by a patient undergoing electroconvulsive therapy: strong contractions of the muscles around the hip joint compressed the femoral head against the acetabulum.

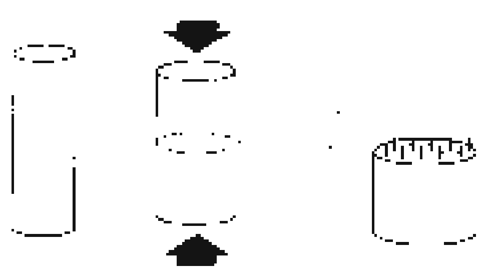


FIG. 2-19

Compressive loading.

Shear

During shear loading, a load is applied parallel to the surface of the structure, and shear stress and strain result inside the structure. Shear stress can be thought of as many small forces acting on the surface of the structure on a plane parallel to the applied load (Fig. 2-22). A structure subjected to a shear load deforms internally in an angular manner; right angles on a plane surface within the structure

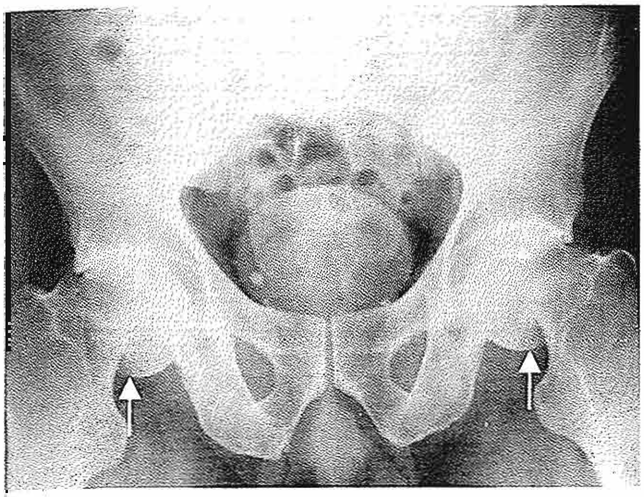


FIG. 2-21

Bilateral subcapital compression fractures of the femoral neck in a patient who underwent electroconvulsive therapy.

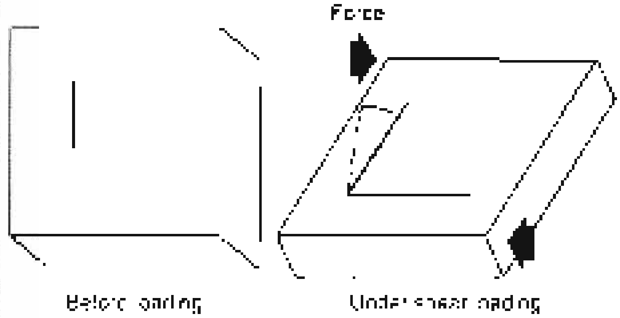


FIG. 2-23

When a structure is loaded in shear, lines originally at right angles on a plane surface within the structure change their orientation, and the angle becomes obtuse or acute. This angular deformation indicates shear strain.

become obtuse or acute (Fig. 2-23). Whenever a structure is subjected to tensile or compressive loading, shear stress is produced. Figure 2-24 illustrates angular deformation in structures subjected to these loading modes. Clinically, shear fractures are most often seen in cancellous bone.

Human adult cortical bone exhibits different values for ultimate stress under compressive, tensile, and shear loading. Cortical bone can withstand greater stress in compression (approximately 190 Mpa) than in tension (approximately 130 Mpa) and greater stress in tension than in shear (70 Mpa). The elasticity (Young's modulus) is approximately 17 GPa in longitudinal or axial loading and approximately 11 GPa in transverse loading. Human trabecular bone values for testing in compression are approximately 50 Mpa and are reduced to approxi-

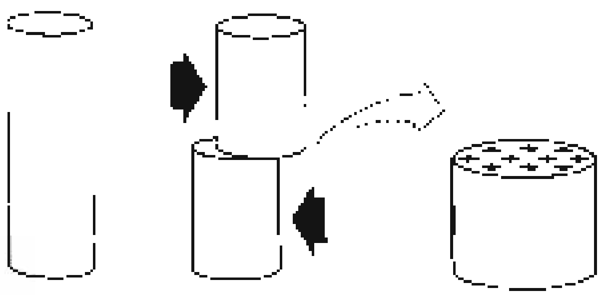


FIG. 2-22

Shear loading

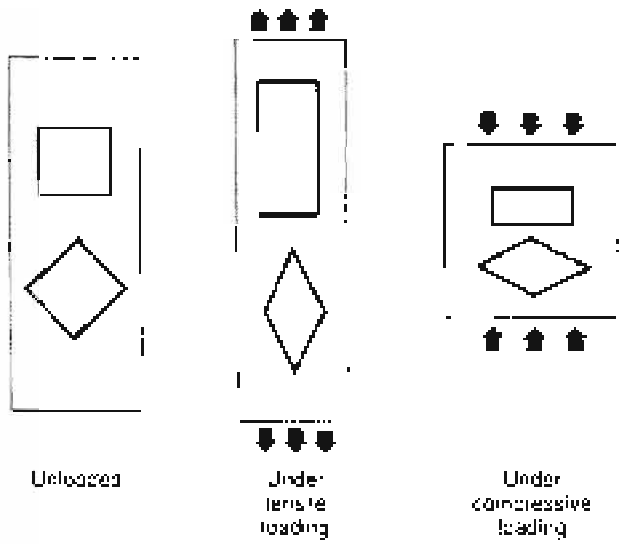


FIG. 2-24

The presence of shear strain in a structure loaded in tension and in compression is indicated by angular deformation.

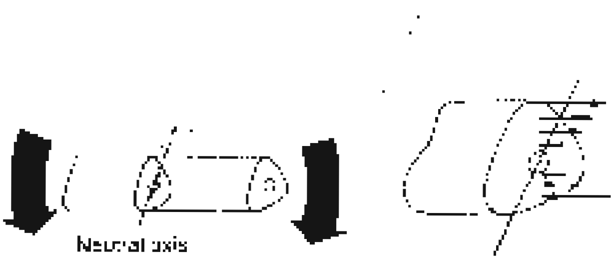


FIG. 2-25

Cross-section of a bone subjected to bending, showing distribution of stresses around the neutral axis. Tensile stresses act on the superior side, and compressive stresses act on the inferior side. The stresses are highest at the periphery of the bone and lowest near the neutral axis. The tensile and compressive stresses are unequal because the bone is asymmetrical.

mately 8 Mpa if loaded in tension. The modulus of elasticity is low (0.0–0.4 GPa) and dependent on the apparent density of the trabecular bone and direction of loading. The clinical biomechanical consequence is that the direction of compression failure results in general in a stable fracture, while a fracture initiated by tension or shear may have catastrophic consequences.

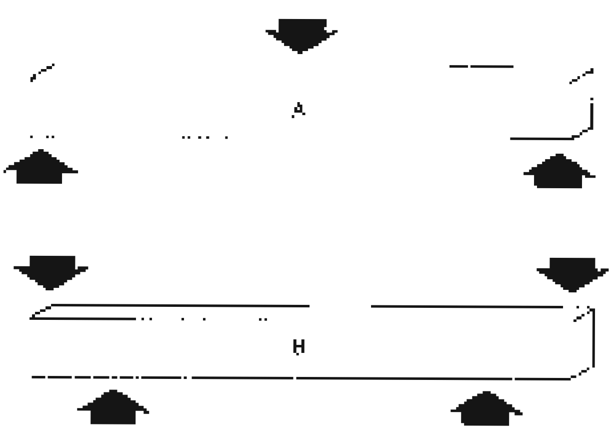


FIG. 2-26

Two types of bending. A, Three-point bending. B, Four-point bending.

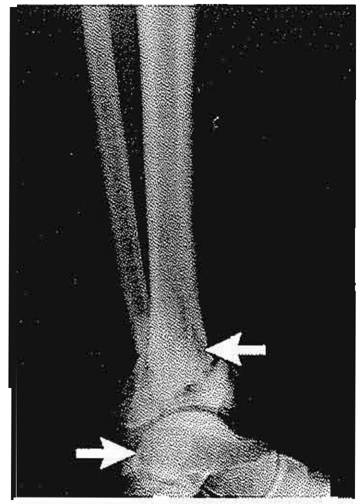


FIG. 2-27

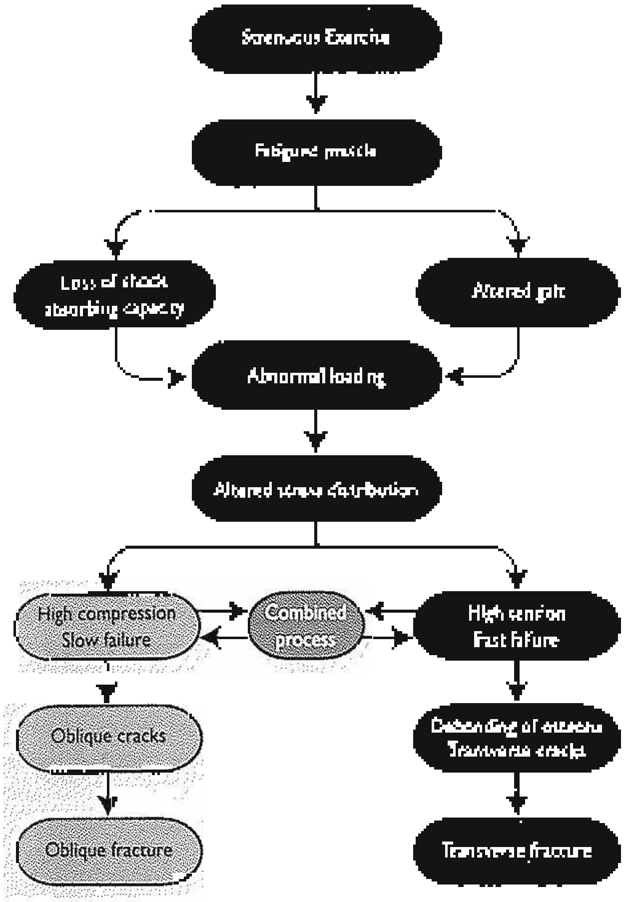
Lateral roentgenogram of a "boot top" fracture produced by three-point bending. Courtesy of Robert A. Winquist, MD.

Bending

In bending, loads are applied to a structure in a manner that causes it to bend about an axis. When a bone is loaded in bending, it is subjected to a combination of tension and compression. Tensile stresses and strains act on one side of the neutral axis, and compressive stresses and strains act on the other side (Fig. 2-25); there are no stresses and strains along the neutral axis. The magnitude of the stresses is proportional to their distance from the neutral axis of the bone. The farther the stresses are from the neutral axis, the higher their magnitude. Because a bone structure is asymmetrical, the stresses may not be equally distributed.

Bending may be produced by three forces (three-point bending) or four forces (four-point bending) (Fig. 2-26). Fractures produced by both types of bending are commonly observed clinically, particularly in the long bones.

Three-point bending takes place when three forces acting on a structure produce two equal moments, each being the product of one of the two peripheral forces and its perpendicular distance from the axis of rotation (the point at which the middle force is applied) (Fig. 2-26A). If loading continues to the yield point, the structure, if homogeneous, symmetrical, and with no structural or tissue de-



FLOW CHART 2-1

ever, will break at the point of application of the middle force.

A typical three-point bending fracture is the "boot top" fracture sustained by skiers. In the "boot top" fracture shown in Figure 2-27, one bending moment acted on the proximal tibia as the skier fell forward over the top of the ski boot. An equal moment, produced by the fixed foot and ski, acted on the distal tibia. As the proximal tibia was bent forward, tensile stresses and strains acted on the posterior side of the bone and compressive stresses and strains acted on the anterior side. The tibia and fibula fractured at the top of the boot. Because adult bone is weaker in tension than in compression, failure begins on the side subjected to tension. Because immature bone is more ductile, it may fail first in compression, and a buckle fracture may result on the compressive side (Flow Chart 2-1).

Four-point bending takes place when two force couples acting on a structure produce two equal moments. A force couple is formed when two parallel forces of equal magnitude but opposite direction are applied to a structure (Fig. 2-28A). Because the magnitude of the bending moment is the same throughout the area between the two force couples, the structure breaks at its weakest point. An example of a four-point bending fracture is shown in Figure 2-28B. A stiff knee joint was manipulated incorrectly during rehabilitation of a patient with a postsurgical infected femoral fracture. During the

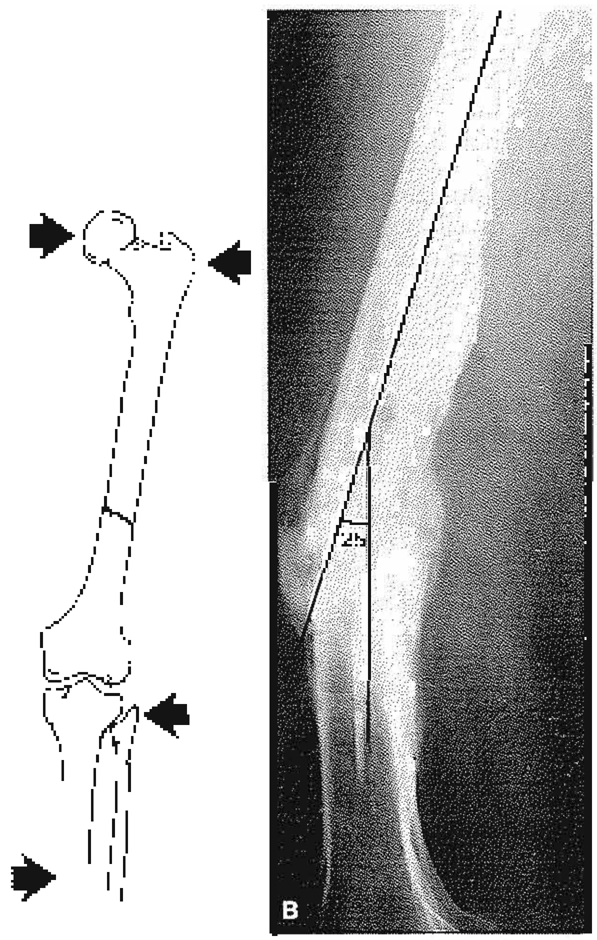


FIG. 2-28

A, During manipulation of a stiff knee joint during fracture rehabilitation, four-point bending caused the femur to refracture at its weakest point, the original fracture site. B, Lateral radiograph of the fractured femur. Courtesy of [illegible]

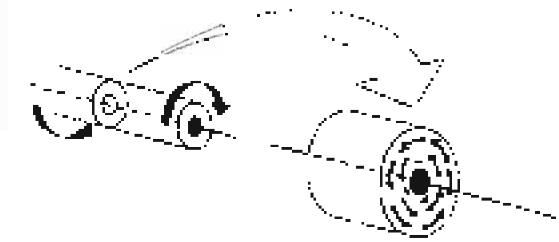


FIG. 2-29

Cross-section of a cylinder loaded in torsion, showing the distribution of shear stresses around the neutral axis. The magnitude of the stresses is highest at the periphery of the cylinder and lowest near the neutral axis.

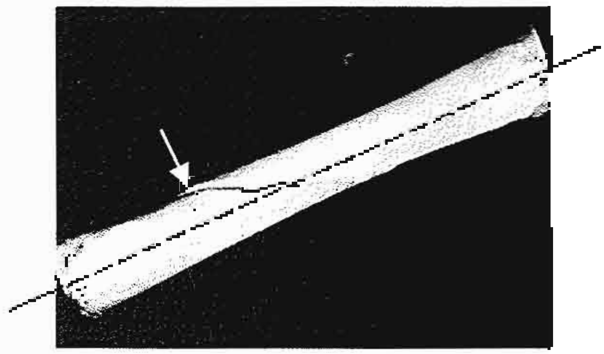


FIG. 2-31

Experimentally produced torsional fracture of a canine femur. The short crack (arrow) parallel to the neutral axis represents shear failure; the fracture line at a 30° angle to the neutral axis represents the plane of maximal tensile stress.

manipulation, the posterior knee joint capsule and tibia formed one force couple and the femoral head and hip joint capsule formed the other. As a bending moment was applied to the femur, the bone failed at its weakest point, the original fracture site.

Torsion

In torsion, a load is applied to a structure in a manner that causes it to twist about an axis, and a torque (or moment) is produced within the structure. When a structure is loaded in torsion, shear stresses are distributed over the entire structure. As in bending, the magnitude of these stresses is proportional to their distance from the neutral axis (Fig. 2-29). The farther the stresses are from the neutral axis, the higher their magnitude.

Under torsional loading, maximal shear stresses act on planes parallel and perpendicular to the neutral axis of the structure. In addition, maximal tensile and compressive stresses act on a plane diagonal to the neutral axis of the structure. Figure 2-30 illustrates these planes in a small segment of bone loaded in torsion.

The fracture pattern for bone loaded in torsion suggests that the bone fails first in shear, with the formation of an initial crack parallel to the neutral axis of the bone. A second crack usually forms along the plane of maximal tensile stress. Such a pattern can be seen in the experimentally produced torsional fracture of a canine femur shown in Figure 2-31.

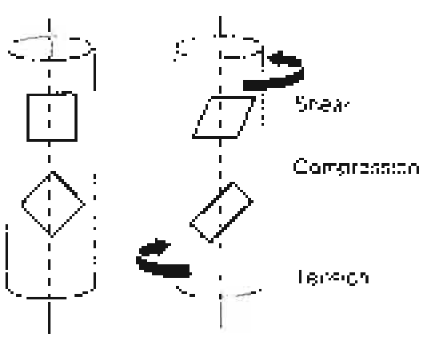


FIG. 2-30

Schematic representation of a small segment of bone loaded in torsion. Maximal shear stresses act on planes parallel and perpendicular to the neutral axis. Maximal tensile and compressive stresses act on planes diagonal to this axis.

Combined Loading

Although each loading mode has been considered separately, living bone is seldom loaded in one mode only. Loading of bone in vivo is complex for two principal reasons: bones are constantly subjected to multiple indeterminate loads and their geometric structure is irregular. In vivo measurement of the strains on the anteromedial surface of a human adult tibia during walking and jogging demonstrates the complexity of the loading patterns during these common physiological activities (Lanyon et al., 1975). Stress values calculated from these strain measurements by Carter (1978) showed that during normal walking, the stresses were compressive during heel strike, tensile during the stance phase, and again compressive during push-off (Fig. 2-32A). Values for shear stress were relatively high in the later portion of the gait cycle, denoting significant torsional loading. This torsional loading was associated with external rotation of the tibia during stance and push-off.

During jogging, the stress pattern was quite different (Fig. 2-32B). The compressive stress predominating at toe strike was followed by high tensile stress during push-off. The shear stress was low throughout the stride, denoting minimal torsional loading produced by slight external and internal rotation of the tibia in an alternating pattern. The increase in speed from slow walking to jogging increased both the stress and the strain on the tibia (Lanyon et al., 1975). This increase in strain with greater speed was confirmed in studies of locomotion in sheep, which demonstrated a fivefold increase in strain values from slow walking to fast running (Lanyon & Bourn, 1979).

INFLUENCE OF MUSCLE ACTIVITY ON STRESS DISTRIBUTION IN BONE

When bone is loaded in vivo, the contraction of the muscles attached to the bone alters the stress distribution in the bone. This muscle contraction decreases or eliminates tensile stress on the bone by producing compressive stress that neutralizes it either partially or totally.

The effect of muscle contraction can be illustrated in a tibia subjected to three-point bending. Figure 2-33A represents the leg of a skier who is falling forward, subjecting the tibia to a bending

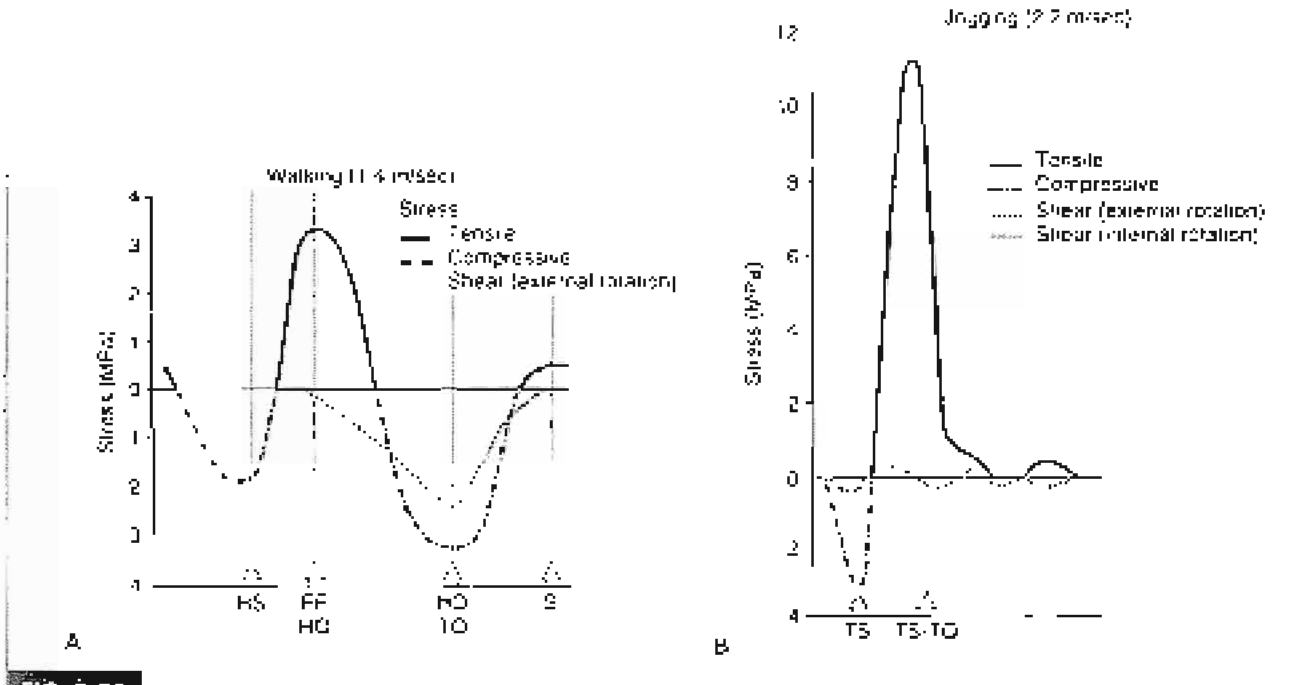


FIG. 2-32

A. Calculated stresses on the anterolateral cortex of a human tibia during walking. HS, heel strike; FF, foot flat; HO, heel-off; TO, toe off; S, swing. *Calculated from Lanyon, L.E., Hampson, W.G.J., Goodship, A.E., et al. (1975). Bone deformation recorded in vivo from strain gauges attached to the human tibial shaft. Acta Orthop Scand, 46, 256. Figure courtesy of Dennis R. Carter, Ph.D.*

B. Calculated stresses on the anterolateral cortex of a human tibia during jogging. TS, toe strike; TO, toe off. *Calculated from Lanyon, L.E., Hampson, W.G.J., Goodship, A.E., et al. (1975). Bone deformation recorded in vivo from strain gauges attached to the human tibial shaft. Acta Orthop Scand, 46, 256. Figure courtesy of Dennis R. Carter, Ph.D.*

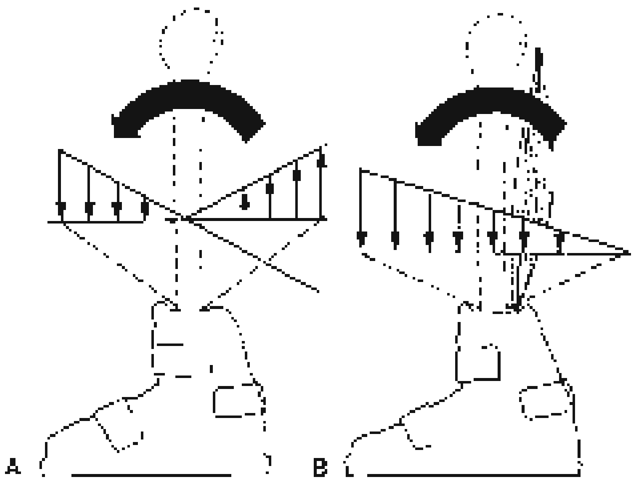


FIG. 2-33

A, Distribution of compressive and tensile stresses in a tibia subjected to three-point bending. B, Contraction of the triceps surae muscle produces high compressive stress on the posterior aspect, neutralizing the high tensile stress.

moment. High tensile stress is produced on the posterior aspect of the tibia, and high compressive stress acts on the anterior aspect. Contraction of the triceps surae muscle produces great compressive stress on the posterior aspect (Fig. 2-33B), neutralizing the great tensile stress and thereby protecting the tibia from failure in tension. This muscle contraction may result in higher compressive stress on the anterior surface of the tibia and thus protect the bone from failure. Adult bone can usually withstand this stress, but immature bone, which is weaker, may fail in compression.

*Muscle contraction produces a similar effect in the hip joint (Fig. 2-34). During locomotion, bending moments are applied to the femoral neck and tensile stress is produced on the superior cortex. Contraction of the gluteus medius muscle produces compressive stress that neutralizes this tensile stress, with the net result that neither compressive nor tensile stress acts on the superior cortex. Thus, the muscle contraction allows the femoral neck to sustain higher loads than would otherwise be possible.

STRAIN RATE DEPENDENCY IN BONE

Because bone is a viscoelastic material, its biomechanical behavior varies with the rate at which the bone is loaded (i.e., the rate at which the load is applied and removed). Bone is stiffer and sustains a higher load to failure when loads are applied at higher rates. Bone also stores more energy before failure at higher loading rates, provided that these rates are within the physiological range.

The in vivo daily strain can vary considerably. The calculated strain rate for slow walking is 0.001 per second, while slow running displays a strain rate of 0.03 per second.

In general, when activities become more strenuous, the strain rate increases (Keaveny & Hayes, 1993). Figure 2-35 shows cortical bone behavior in tensile testing at different physiological strain rates. As can be seen from the figure, the same change in strain rate produces a larger change in ultimate stress (strength) than in elasticity (Young's modulus). The data indicates that the bone is approximately 30% stronger for brisk walking than for slow

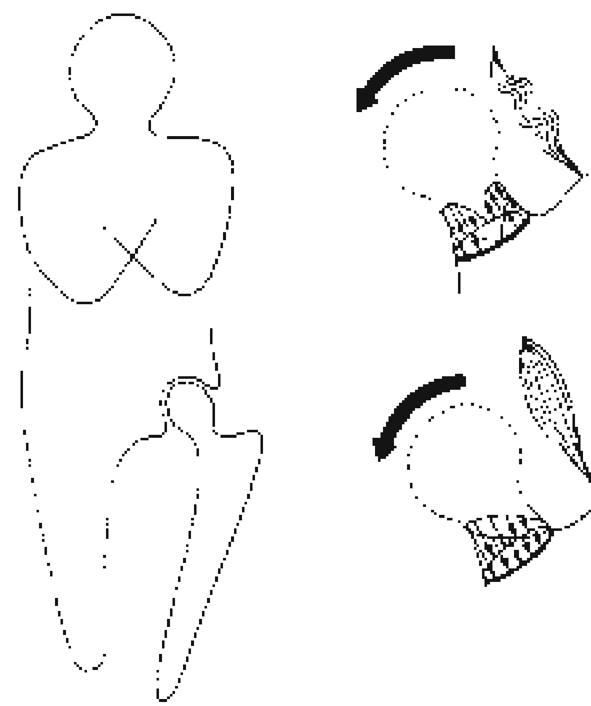


FIG. 2-34

Stress distribution in a femoral neck subjected to bending. When the gluteus medius muscle is relaxed (top), tensile stress acts on the superior cortex and compressive stress acts on the inferior cortex. Contraction of this muscle (bottom) neutralizes the tensile stress.

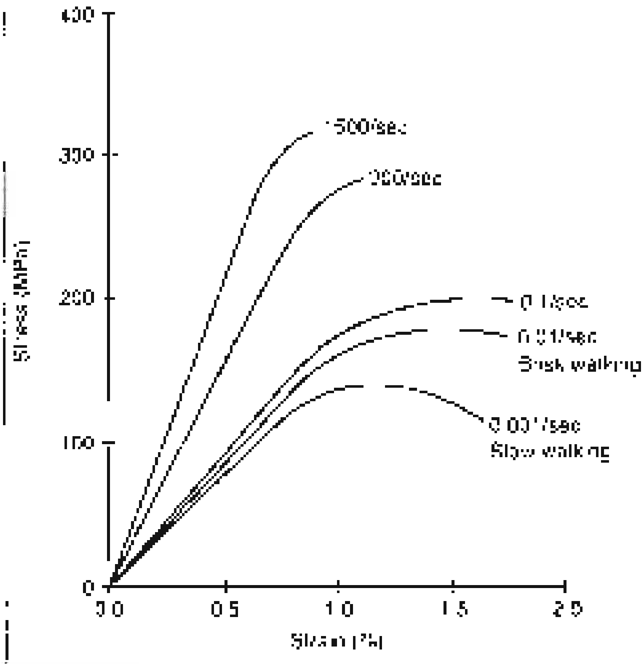


FIG. 2-35

Rate dependency of cortical bone is demonstrated at five strain rates. Both stiffness (modulus) and strength increase considerably at increased strain rates. *Adapted from McElhaney, J.H. (1966). Dynamic response of bone and muscle tissue. J Appl Physiol, 21, 1231–1236.*

walking. At very high strain rates (>1 per second) representing impact trauma, the bone becomes more brittle. In a full range of experimental testing for ultimate tensile strength and elasticity of cortical bone, the strength increases by a factor of three and the modulus by a factor of two (Keaveny & Hayes, 1993).

The loading rate is clinically significant because it influences both the fracture pattern and the amount of soft tissue damage at fracture. When a bone fractures, the stored energy is released. At a low loading rate, the energy can dissipate through the formation of a single crack; the bone and soft tissues remain relatively intact, with little or no displacement of the bone fragments. At a high loading rate, however, the greater energy stored cannot dissipate rapidly enough through a single crack, and comminution of bone and extensive soft tissue damage result. Figure 2-36 shows a human tibia tested in vitro in torsion at a high loading rate; numerous bone fragments were produced, and displacement of the fragments was pronounced.

Clinically, bone fractures fall into three general categories based on the amount of energy released at fracture: low-energy, high-energy, and very high-energy. A low-energy fracture is exemplified by the simple torsional ski fracture; a high-energy fracture is often sustained during automobile accidents; and a very high-energy fracture is produced by very high-muzzle velocity gunshot.

FATIGUE OF BONE UNDER REPETITIVE LOADING

Bone fractures can be produced by a single load that exceeds the ultimate strength of the bone or by repeated applications of a load of lower magnitude. A fracture caused by a repeated load application is called a fatigue fracture and is typically produced either by few repetitions of a high load or by many repetitions of a relatively normal load (Case Study 2-1).

The interplay of load and repetition for any material can be plotted on a fatigue curve (Fig. 2-37). For some materials (some metals, for example), the fatigue curve is asymptotic, indicating that if the load is kept below a certain level, theoretically the material will remain intact no matter how many repetitions. For bone tested in vitro, the curve is not asymptotic. When bone is subjected to repetitive low loads, it may sustain microfractures. Testing of bone in vitro also reveals that bone fatigues rapidly when the load or deformation approaches its yield strength; that is, the number of repetitions needed to produce a fracture diminishes rapidly.

In repetitive loading of living bone, the fatigue process is affected not only by the amount of load and the number of repetitions but also by the number of applications of the load within a given time (frequency of loading). Because living bone is self-repairing, a fatigue fracture results only when the remodeling process is outpaced by the fatigue process—that is, when loading is so frequent that it precludes the remodeling necessary to prevent failure.

Fatigue fractures are usually sustained during continuous strenuous physical activity, which causes the muscles to become fatigued and reduces their ability to contract. As a result, they are less able to store energy and thus to neutralize the stresses imposed on the bone. The resulting alteration of the stress distribution in the bone causes

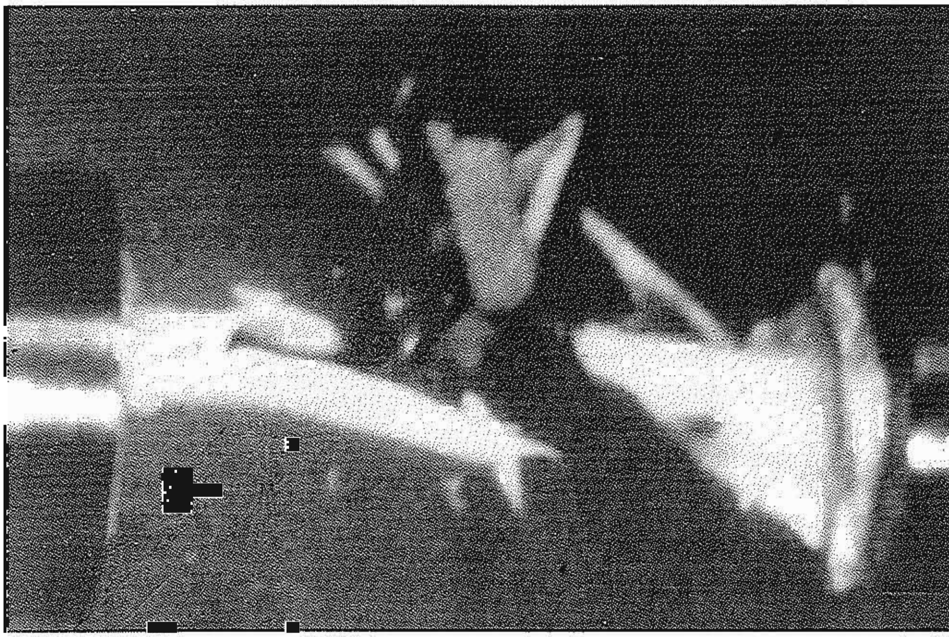


FIG. 2-36

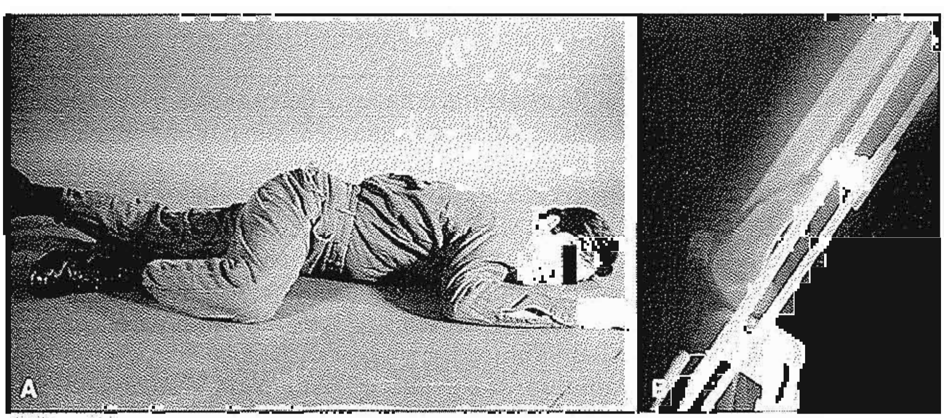
Human tibia experimentally tested to failure in torsion at a high loading rate. Displacement of the numerous fragments was pronounced.

CASE STUDY 2-1

Bone Overloading

A 23-year-old military recruit was exposed to an intensive heavy physical training program that included repetitive continuous crawling in an awkward position for several weeks (Case Study Fig. 2-1-1A). The repeated application of loads (high repetitions) and the number of applications of a load during a short period of time (high frequency of loading) surpassed the time for the bone remodeling process to prevent failure. [illegible] fatigue occurred as a result of the abnormal [illegible] pattern and the extreme [illegible]; it affected the [illegible] [illegible] redistribution of the stress increased, leading to abnormal loading and altered stress distribution (Case Study Fig. 2-1-1B).

After 4 weeks of strenuous physical activity, the damage occurred. A stress fatigue at the femoral shaft lead to an oblique fracture.



Case Study Figure 2-1-1A. Abnormal loads at the femoral shaft occurred

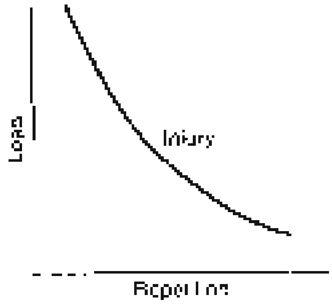


FIG. 2-37

The interplay of load and repetition is represented on a fatigue curve.

abnormally high loads to be imposed, and a fatigue damage accumulation occurs that may lead to a fracture. Bone may fail on the tensile side, on the compressive side, or on both sides. Failure on the tensile side results in a transverse crack, and the bone proceeds rapidly to complete fracture. Fatigue fractures on the compressive side appear to be pro-

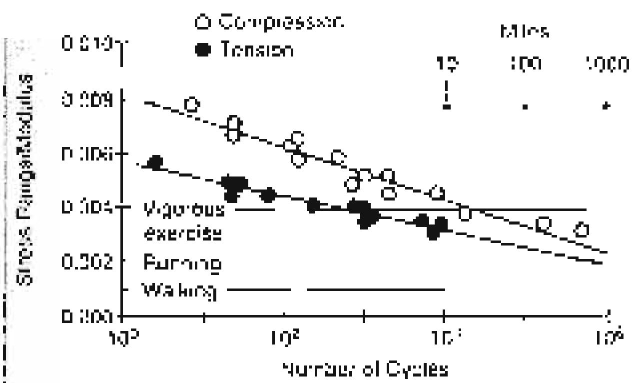


FIG. 2-38

Fatigue testing showing the number of cycles (x-axis) and strain range (y-axis) expressed as stress range/modulus in human cortical bone specimens loaded in tension and compression. Typical strain ranges are shown for walking, running, and vigorous exercises. Note that resistance to fatigue fracture is greater in compressive loading. Ten miles represent approximately 5,000 cycles, corresponding to the number of steps running during that distance. *Adapted from Carter, D.R., Caler, W.E., Spengler, D.M., Frankel, V.H. (1981). Fatigue behavior of adult cortical bone: the influence of mean strain and strain range. Acta Orthop Scand, 52, 481–490.*

duced more slowly; the remodeling is less easily outpaced by the fatigue process and the bone may not proceed to complete fracture.

This theory of muscle fatigue as a cause of fatigue fractures in the lower extremities is outlined in the schema in Flowchart 2-1 on p. 41.

Figure 2-38 shows typical strain ranges for human femoral cortical bone during different activities and distances. Resistance to fatigue behavior is greater in compression than in tension (Keaveny & Hayes, 1993). On average, approximately 5,000 cycles of experimental loading correspond to the number of steps in 10 miles of running. One million cycles corresponds to approximately 1,000 miles. A total distance of less than 1,000 miles could cause a fracture of the cortical bone tissue. This is consistent with stress fractures reported among military recruits undergoing strenuous training of up to 1,000 miles of running over a short period of time (6 weeks). Fractures of individual trabeculae in cancellous bone have been observed in postmortem human specimens and may be caused by fatigue accumulation. Common sites are the lumbar vertebrae, the femoral head, and the proximal tibia. It has been suggested that these fractures may play a role in bone remodeling as well as in age-related fractures, collapse of subchondral bone, degenerative joint diseases, and other bone disorders.

INFLUENCE OF BONE GEOMETRY ON BIOMECHANICAL BEHAVIOR

The geometry of a bone greatly influences its mechanical behavior. In tension and compression, the load to failure and the stiffness are proportional to the cross-sectional area of the bone. The larger the area, the stronger and stiffer the bone. In bending, both the cross-sectional area and the distribution of bone tissue around a neutral axis affect the bone's mechanical behavior. The quantity that takes into account these two factors in bending is called the area moment of inertia. A larger moment of inertia results in a stronger and stiffer bone. Figure 2-39 shows the influence of the area moment of inertia on the load to failure and the stiffness of three rectangular structures that have the same area but different shapes. In bending, beam III is the stiffest of the three and can withstand the highest load because the greatest amount of material is distributed at a distance from the neutral axis. For rectangular cross-sections, the formula for the area moment of

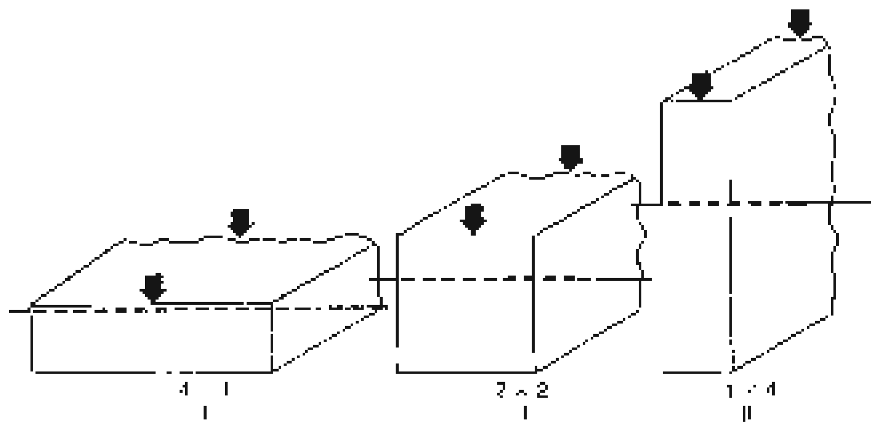


FIG. 2-39

Three beams of equal area but different shapes subjected to bending. The area moment of inertia for beam I is 4/12; for beam II, 16/12; and for beam III, 64/12. *Adapted from Frankel, V.H., & Burstein, A.H. (1970). Orthopaedic Biomechanics. Philadelphia: Lea & Febiger.*

inertia is the width (B) multiplied by the cube of the height (H^3) divided by 12:

$$\frac{B \cdot H^3}{12}$$

Because of its large area moment of inertia, beam III can withstand four times more load in bending than can beam I.

A third factor, the length of the bone, influences the strength and stiffness in bending. The longer the bone, the greater the magnitude of the bending moment caused by the application of a force. In a rectangular structure, the magnitude of the stresses produced at the point of application of the bending moment is proportional to the length of the structure. Figure 2-40 depicts the forces acting on two beams with the same width and height but different lengths: beam B is twice as long as beam A. The bending moment for the longer beam is twice that for the shorter beam; consequently, the stress magnitude throughout the beam is twice as high. Because of their length, the long bones of the skeleton are subjected to high bending moments and, therefore, to high tensile and compressive stresses. Their tubular shape gives them the ability to resist bending moments in all directions. These bones have a large area moment of inertia because much of the bone tissue is distributed at a distance from the neutral axis.

The factors that affect bone strength and stiffness in torsion are the same that operate in bending: the cross-sectional area and the distribution of bone tissue around a neutral axis. The quantity that takes into account these two factors in torsional loading is the polar moment of inertia. The larger the polar moment of inertia, the stronger and stiffer the bone.

Figure 2-41 shows distal and proximal cross-sections of a tibia subjected to torsional loading. Although the proximal section has a slightly smaller bony area than does the distal section, it has a much higher polar moment of inertia because much of the bone tissue is distributed at a distance from the neutral axis. The distal section, while it has a larger bony area, is subjected to much higher shear stress because much of the bone tissue is distributed close to the neutral axis. The magnitude of the shear stress in the distal section is approximately double that in the proximal section. Clinically, torsional fractures of the tibia commonly occur distally.

When bone begins to heal after fracture, blood vessels and connective tissue from the periosteum migrate into the region of the fracture, forming a cuff of dense fibrous tissue, or callus (woven bone), around the fracture site, stabilizing that area (Fig. 2-42A). The callus significantly increases the area and polar moments of inertia, thereby increasing the strength and stiffness of the bone in bending and torsion during the healing period. As the frac-

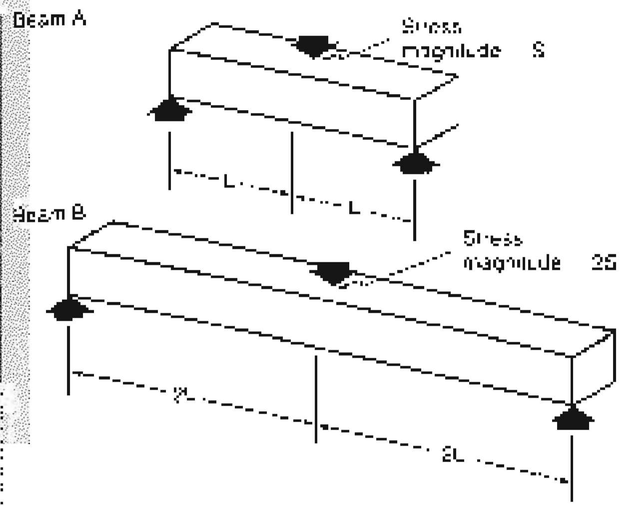


FIG. 2-40

Beam B is twice as long as beam A and sustains twice the bending moment. Hence, the stress magnitude throughout beam B is twice as high. *Adapted from Frankel, V.H., & Burstein, A.H. (1970). Orthopaedic Biomechanics. Philadelphia: Lea & Febiger.*

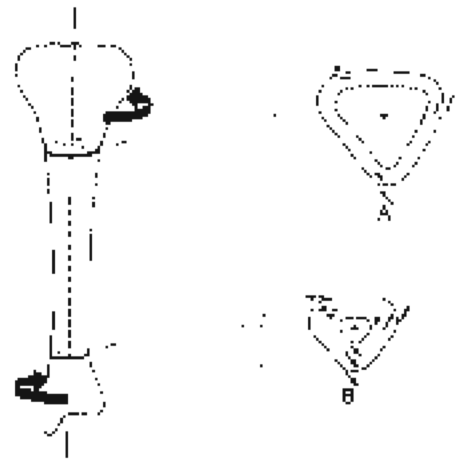


FIG. 2-41

Distribution of shear stress in two cross sections of a tibia subjected to torsional loading. The proximal section (A) has a higher moment of inertia than does the distal section (B) because more bony material is distributed away from the neutral axis. *Adapted from Frankel, V.H., & Burstein, A.H. (1970). Orthopaedic Biomechanics. Philadelphia: Lea & Febiger.*

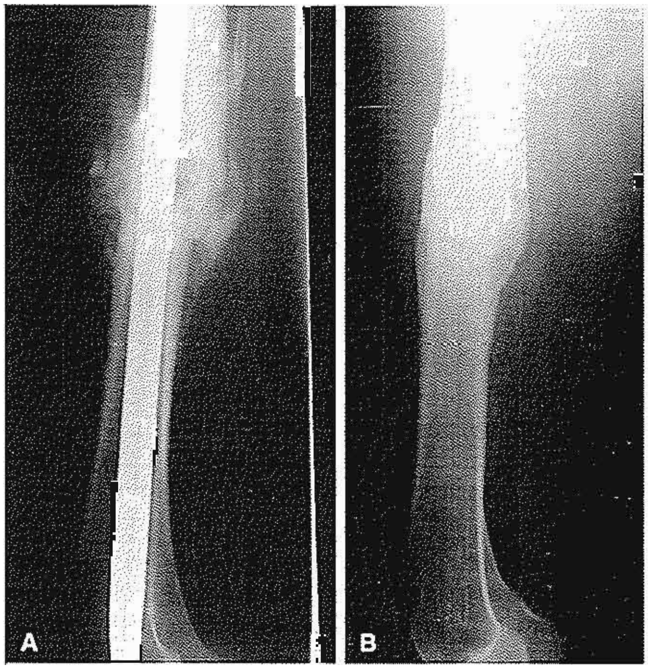


FIG. 2-42

A. Early callus formation in a femoral fracture fixed with an intramedullary nail. **B.** Nine months after injury, the fracture has healed and most of the callus cuff has been resorbed. *Courtesy of Robert A. Winquist, M.D.*

ture heals and the bone gradually regains its normal strength, the callus cuff is progressively resorbed and the bone returns to as near its normal size and shape as possible (Fig. 2-42B).

Certain surgical procedures produce defects that greatly weaken the bone, particularly in torsion. These defects fall into two categories: those whose length is less than the diameter of the bone (stress raisers) and those whose length exceeds the bone diameter (open section defects).

A stress raiser is produced surgically when a small piece of bone is removed or a screw is inserted. Bone strength is reduced because the stresses imposed during loading are prevented from being distributed evenly throughout the bone and instead become concentrated around the defect. This defect is analogous to a rock in a stream, which diverts the water, producing high water turbulence around it. The weakening effect of a stress raiser is particularly marked under torsional loading; the total decrease in bone strength in this loading mode can reach 60%.

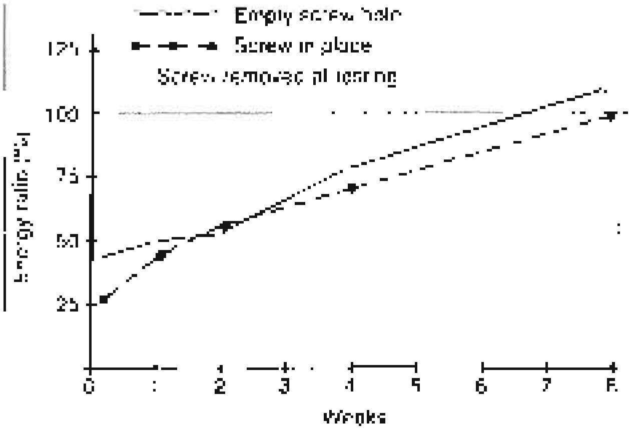


FIG. 2-43

Effect of screws and of empty screw holes on the energy storage capacity of rabbit femora. The energy storage for experimental animals is expressed as a percentage of the total energy storage capacity for control animals. When screws were removed immediately before testing, the energy storage capacity decreased by 50%. *Adapted from Burstein, A.H., et al. (1972). Bone strength: The effect of screw holes. J Bone Joint Surg, 54A, 1143.*

Burstein and associates (1972) showed the effect of stress raisers produced by screws and by empty screw holes on the energy storage capacity of rabbit bones tested in torsion at a high loading rate. The immediate effect of drilling a hole and inserting a screw in a rabbit femur was a 74% decrease in energy storage capacity. After 8 weeks, the stress raiser effect produced by the screws and by the holes without screws had disappeared completely because the bone had remodeled. Bone had been laid down around the screws to stabilize them, and the empty screw holes had been filled in with bone. In femurs from which the screws had been removed immediately before testing, however, the energy storage capacity of the bone decreased by 50%, mainly because the bone tissue around the screw sustained microdamage during screw removal (Fig. 2-43).

An open section defect is a discontinuity in the bone caused by the surgical removal of a piece of bone longer than the bone's diameter (e.g., by the cutting of a slot during a bone biopsy). Because the outer surface of the bone's cross-section is no longer continuous, its ability to resist loads is altered, particularly in torsion.

In a normal bone subjected to torsion, the shear stress is distributed throughout the bone and acts to resist the torque. This stress pattern is illustrated in the cross-section of a long bone shown in Figure 2-44A. (A cross-section with a continuous outer surface is called a closed section.) In a bone with an

FIG. 2-44

Stress pattern in an open and closed section under torsional loading. **A.** In the closed section, all the shear stress resists the applied torque. **B.** In the open section, only the shear stress at the periphery of the bone resists the applied torque.

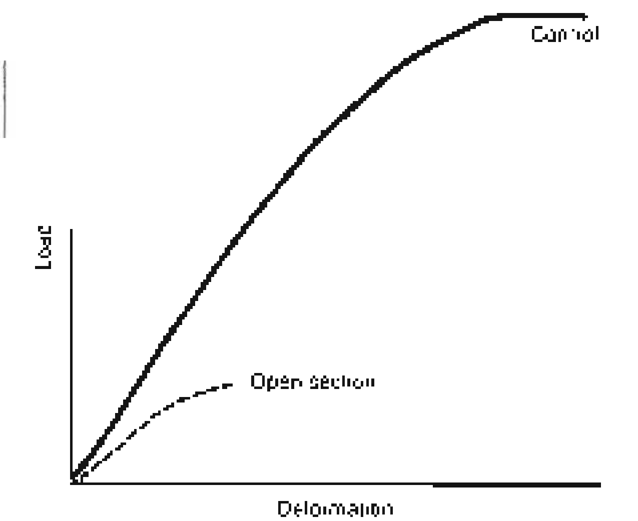


FIG. 2-45

Load-deformation curves for human adult tibiae tested in vitro under torsional loading. The control curve represents a tibia with no defect; the open section curve represents a tibia with an open section defect. *Adapted from Frankel, V.H., & Burstein, A.H. (1970). Orthopaedic Biomechanics. Philadelphia: Lea & Febiger.*

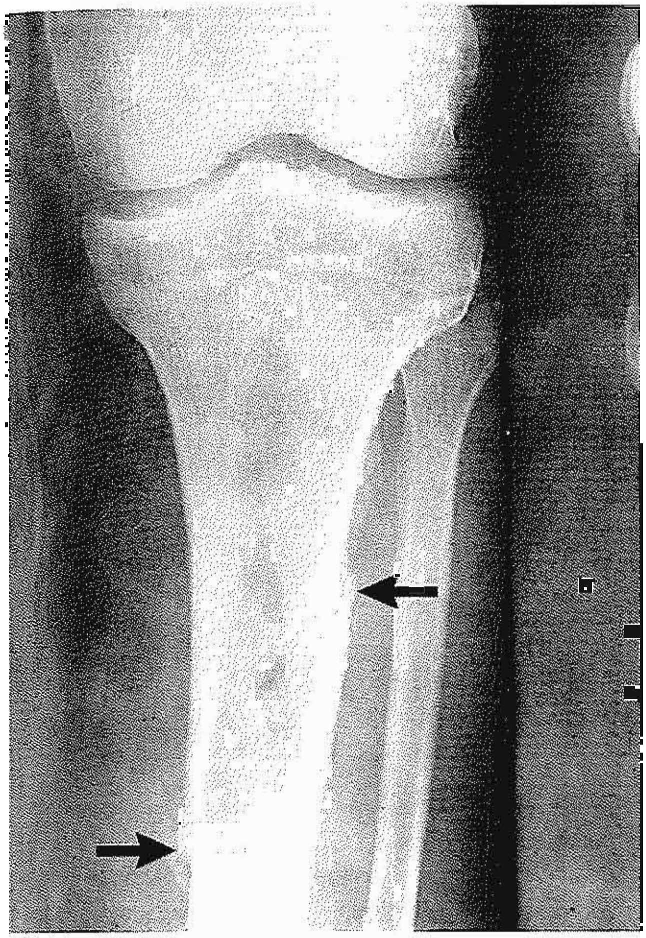


FIG. 2-46

A patient sustained a tibial fracture through a surgically produced open section defect when she tripped a few weeks after the biopsy.

open section defect, only the shear stress at the periphery of the bone resists the applied torque. As the shear stress encounters the discontinuity, it is forced to change direction (Fig. 2-44B). Throughout the interior of the bone, the stress runs parallel to the applied torque, and the amount of bone tissue resisting the load is greatly decreased.

In torsion tests in vitro of human adult tibiae, an open section defect reduced the load to failure and energy storage to failure by as much as 90%. The deformation to failure was diminished by approximately 70% (Frankel & Burstein, 1970) (Fig. 2-45).

Clinically, the surgical removal of a piece of bone can greatly weaken the bone, particularly in torsion. Figure 2-46 is a radiograph of a tibia from which a graft was removed for use in an arthrodesis of the hip. A few weeks after operation, the patient tripped while twisting, and the bone fractured through the defect.

Bone Remodeling

Bone has the ability to remodel, by altering its size, shape, and structure, to meet the mechanical demands placed on it (Buckwalter et al., 1995). This phenomenon, in which bone gains or loses cancellous and/or cortical bone in response to the level of stress sustained, is summarized as Wolff's law, which states that the remodeling of bone is influenced and modulated by mechanical stresses (Wolff, 1892).

Load on the skeleton can be accomplished by either muscle activity or gravity. A positive correlation exists between bone mass and body weight. A greater body weight has been associated with a larger bone mass (Exner et al., 1979). Conversely, a prolonged condition of weightlessness, such as that experienced during space travel, has been found to result in decreased bone mass in weight-bearing bones. Astronauts experience a fast loss of calcium and consequent bone loss (Rambaut & Johnston, 1979; Whedon, 1984). These changes are not completely reversible.

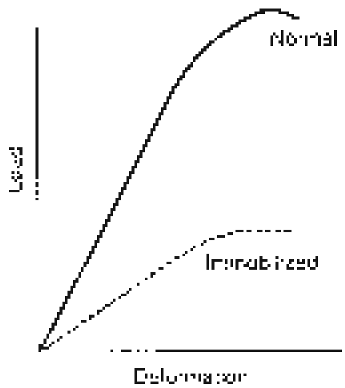


FIG. 2-47

Load-deformation curves for vertebral segments L5 to L7 from normal and immobilized Rhesus monkeys. Note the extensive loss of strength and stiffness in the immobilized specimens. (Adapted from Kazarian, L.L., & Von Gierke, H.E. (1969). Bone loss as a result of immobilization and chelation. Preliminary results in Macaca mulatta. Clin Orthop, 65, 67.)

CASE STUDY 2-2

Bone Remodeling

A 30-year-old man underwent surgical removal of a metal plate after stabilization of a displaced ulnar fracture. Figure 2-48 shows anteroposterior (A) and lateral (B) roentgenograms of the ulna after plate removal.

The implant is used to stabilize the fracture for initial healing. However, in situations such as this one, the plate removal decreased the amount of mechanical stresses necessary for bone remodeling. This is of concern when the plate carries most or all of the mechanical load and remains after fracture healing. Thus, according to Wolff's law, it will promote localized osseous resorption as a result of decreased mechanical stress and stimulus of the bone under the plate, resulting in a decrease in strength and stiffness of the bone.

Disuse or inactivity has deleterious effects on the skeleton. Bed rest produces a bone mass decrease of approximately 1% per week (Jenkins & Cochran, 1969; Krolner & Toft, 1983). In partial or total immobilization, bone is not subjected to the usual mechanical stresses, which leads to resorption of the periosteal and subperiosteal bone and a decrease in the mechanical properties of bone (i.e., strength and stiffness). This decrease in bone strength and stiffness was shown by Kazarian and Von Gierke (1969), who immobilized Rhesus monkeys in full-body casts for 60 days. Subsequent compressive testing in vitro of the vertebrae from the immobilized monkeys and from controls showed up to a threefold decrease in load to failure and energy storage capacity in the vertebrae that had been immobilized; stiffness was also significantly decreased (Fig. 2-47).

An implant that remains firmly attached to a bone after a fracture has healed may also diminish the strength and stiffness of the bone. In the case of a plate fixed to the bone with screws, the plate and the bone share the load in proportions deter-

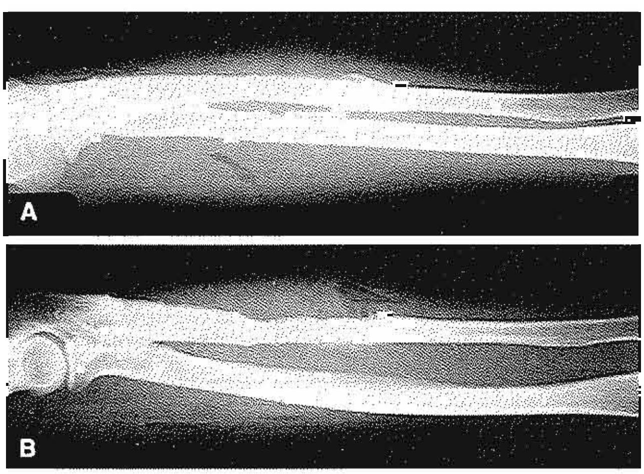


FIG. 2-48

Anteroposterior (A) and lateral (B) roentgenograms of an ulna after plate removal show a decreased bone diameter caused by resorption of the bone under the plate. Cancellization of the cortex and the presence of screw holes also weaken the bone. Courtesy of Marc Martens, M.D.

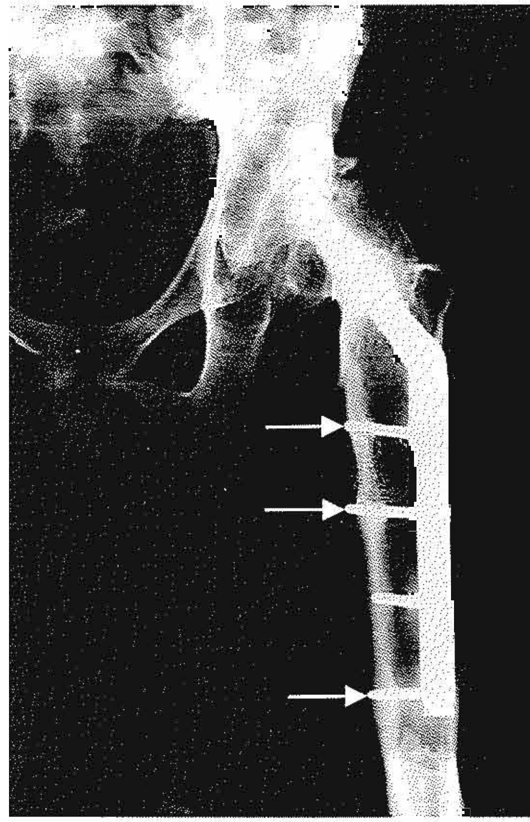


FIG. 2-49

Roentgenogram of a fractured femoral neck to which a nail plate was applied. Loads are transmitted from the plate to the bone via the screws. Bone has been laid down around the screws to bear these loads.

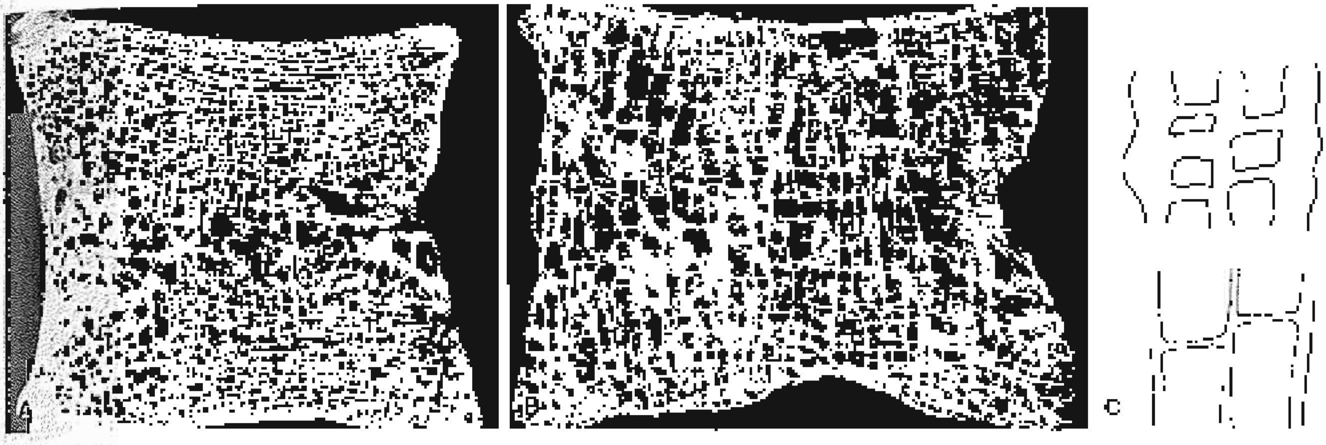


FIG. 2-50

Vertebral cross-sections from autopsy specimens of young (A) and old (B) bone show a marked reduction in cancellous bone in the latter. *Reprinted with permission from Nordin, B.E.C. (1973). Metabolic Bone and Stone Disease. Edinburgh: Churchill Livingstone.* C. Bone reduction with aging is schematically depicted. As normal bone (top) is subjected to absorption (shaded area) during the aging process, the longitudinal trabeculae become thinner and some transverse trabeculae disappear (bottom). *Adapted from Siffert, R.S., & Levy, R.N. (1981). Trabecular patterns and the internal architecture of bone. Mt Sinai J Med, 48, 221.*

mined by the geometry and material properties of each structure (Case Study 2-2). A large plate, carrying high loads, unloads the bone to a great extent; the bone then atrophies in response to this diminished load. (The bone may hypertrophy at the bone-screw interface in an attempt to reduce the micromotion of the screws.)

Bone resorption under a plate is illustrated in Figure 2-48. A compression plate made of a material approximately 10 times stiffer than the bone was applied to a fractured ulna and remained after the fracture had healed. The bone under the plate carried a lower load than normal; it was partially resorbed, and the diameter of the diaphysis became markedly smaller. A reduction in the size of the bone diameter greatly decreases bone strength, particularly in bending and torsion, as it reduces the area and polar moments of inertia. A 30% decrease in bone diameter may reduce the strength in torsion by 60%. Changes in bone size and shape illustrated in Figure 2-48 suggest that rigid plates should be removed shortly after a fracture has healed and before the bone has markedly diminished in size. Such a decrease in bone size is usually accompanied by secondary osteoporosis, which further weakens the bone (Slätis et al., 1980).

An implant may cause bone hypertrophy at its attachment sites. An example of bone hypertrophy around screws is illustrated in Figure 2-49. A nail plate was applied to a femoral neck fracture and the bone hypertrophied around the screws in response to the increased load at these sites. Hypertrophy may also result if bone is repeatedly subjected to high mechanical stresses within the normal physiological range. Hypertrophy of normal adult bone in response to strenuous exercise has been observed (Dalén & Olsson, 1974; Huddleston et al., 1980; Jones et al., 1977), as has an increase in bone density (Nilsson & Westlin, 1971).

Degenerative Changes in Bone Associated With Aging

A progressive loss of bone density has been observed as part of the normal aging process. The longitudinal trabeculae become thinner, and some of the transverse trabeculae are resorbed (Siffert & Levy, 1981) (Fig. 2-50). The result is a marked reduction in the amount of cancellous bone and a thinning of cortical bone. The relationship between bone mass, age, and gender is shown in Figure 2-51. The decrease in bone tissue and the slight decrease in the size of the bone reduce bone strength and stiffness.

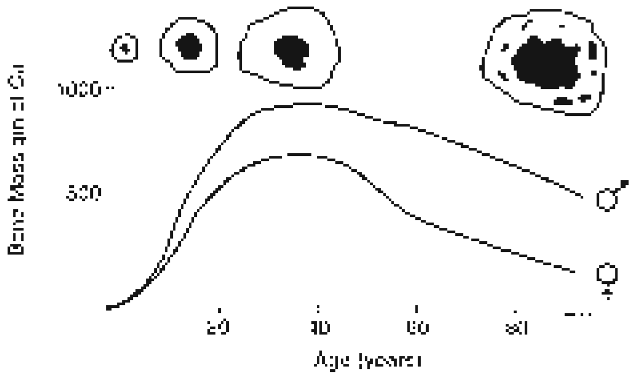


FIG. 2-51

Graph showing the relationship between bone mass, age, and gender. On the top of the figure, a cross-section of the diaphysis of the femur and the bone mass configuration is shown. *Reprinted with permission from Kaplan, F.S., Hayes, W.C., Keaveny, T.M., et al. (1994). Form and function of bone. In S.R. Simon (Ed.). Orthopaedic Basic Science (pp. 127–185). Rosemont, IL: AAOS.*

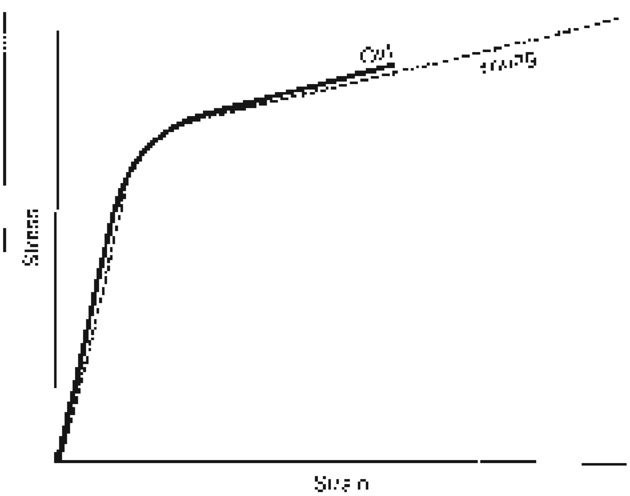


FIG. 2-52

Stress-strain curves for samples of adult young and old human tibiae tested in tension. Note that the bone strength is comparable but that the old bone is more brittle and has lost its ability to deform. *Adapted from Burstein, A.H., Reilly, D.T., & Martens, M. (1976). Aging of bone tissue: Mechanical properties. J Bone Joint Surg, 58A, 82.*

Stress-strain curves for specimens from human adult tibiae of two widely differing ages tested in tension are shown in Figure 2-52. The ultimate stress was approximately the same for the young and the old bone. The old bone specimen could withstand only half the strain that the young bone could, indicating greater brittleness and a reduction in energy storage capacity. The reduction in bone density, strength, and stiffness results in increased bone fragility. Age-related bone loss depends on a number of factors, including gender, age, post-menopause, endocrine abnormality, inactivity, disuse, and calcium deficiency. Over several decades, the skeletal mass may be reduced to 50% of original trabecular and 25% of cortical mass. In the fourth decade, women lose approximately 1.5 to 2% a year while men lose only approximately half that rate (0.5 to 0.75%) yearly. Regular physical activity and exercise (Zetterberg et al., 1990), calcium, and possibly estrogen intake may decrease the rate of bone mineral loss during aging.

Summary

1 Bone is a complex two-phase composite material. One phase is composed of inorganic mineral salts and the other is an organic matrix of collagen and ground substance. The inorganic component makes bone hard and rigid, whereas the organic component gives bone its flexibility and resilience.

2 Microscopically, the fundamental structural unit of bone is the osteon, or haversian system, composed of concentric layers of a mineralized matrix surrounding a central canal containing blood vessels and nerve fibers.

3 Macroscopically, the skeleton is composed of cortical and cancellous (trabecular) bone. Cortical bone has high density while trabecular bone varies in density over a wide range.

4 Bone is an anisotropic material, exhibiting different mechanical properties when loaded in different directions. Mature bone is strongest and stiffest in compression.

5 Bone is subjected to complex loading patterns during common physiological activities such as walking and jogging. Most bone fractures are produced by a combination of several loading modes.

6 Muscle contraction affects stress patterns in bone by producing compressive stress that partially

or totally neutralizes the tensile stress acting on the bone.

7 Bone is stiffer, sustains higher loads before failing, and stores more energy when loaded at higher physiological strain rates.

8 Living bone fatigues when the frequency of loading precludes the remodeling necessary to prevent failure.

9 The mechanical behavior of a bone is influenced by its geometry (length, cross-sectional area, and distribution of bone tissue around the neutral axis).

10 Bone remodels in response to the mechanical demands placed on it; it is laid down where needed and resorbed where not needed.

11 With aging comes a marked reduction in the amount of cancellous bone and a decrease in the thickness of cortical bone. These changes diminish bone strength and stiffness.

REFERENCES

Bassett, C.A.L. (1965). Electrical effects in bone. *Sci Am, 213*, 18.

Bonfield, W., & Li, C.H. (1967). Anisotropy of nonelastic flow in bone. *J Appl Physics, 38*, 2450.

Buckwalter, J.A., Glimcher, M.J., Cooper, R.R., et al. (1995). Bone biology. Part I: Structure, blood supply, cells, matrix and mineralization. Part II: Formation, form, remodeling and regulation of cell function. (Instructional Course Lecture). *J Bone Joint Surg, 77A*, 1256–1289.

Burstein, A.H., Reilly, D.T., & Martens, M. (1976). Aging of bone tissue: Mechanical properties. *J Bone Joint Surg, 58A*, 82.

Burstein, A.H., et al. (1972). Bone strength: The effect of screw holes. *J Bone Joint Surg, 54A*, 1143.

Carter, D.R. (1978). Anisotropic analysis of strain rosette information from cortical bone. *J Biomech, 11*, 199–202.

Carter, D.R., & Hayes, W.C. (1977). Compact bone fatigue damage: A microscopic examination. *Clin Orthop, 127*, 265.

Dalen, N., & Olsson, K.E. (1974). Bone mineral content and physical activity. *Acta Orthop Scand, 45*, 170.

Exner, G.U., et al. (1979). Bone densitometry using computed tomography. Part I: Selective determination of trabecular bone density and other bone mineral parameters. Normal values in children and adults. *Br J Radiol, 52*, 14.

Frankel, V.H., & Burstein, A.H. (1970). *Orthopaedic Biomechanics*. Philadelphia: Lea & Febiger.

Huddleston, A.L., Rockwell, D., Kulund, D.N., et al. (1980). Bone mass in lifetime tennis athletes. *JAMA, 244*, 1107.

International Society of Biomechanics (1987). *Quantities and Units of Measurements in Biomechanics* (unpublished).

Jenkins, D.P., & Cochran, T.H. (1969). Osteoporosis: The dramatic effect of disuse of an extremity. *Clin Orthop, 64*, 128.

Jones, H.H., Priest, J.D., Hayes, W.C., et al. (1977). Humeral hypertrophy in response to exercise. *J Bone Joint Surg, 59A*, 204.

Kaplan, F.S., Hayes, W.C., Keaveny, T.M., et al. (1994). Form and function of bone. In S.R. Simon (Ed.). *Orthopaedic Basic Science* (pp.127–184). Rosemont, IL: AAOS.

Kazarian, L.L., & Von Gierke, H.E. (1969). Bone loss as a result of immobilization and chelation: Preliminary results in *Macaca mulatta*. *Clin Orthop, 65*, 67.

Keaveny, T.M., & Hayes, W.C. (1993). Mechanical properties of cortical and trabecular bone. *Bone, 7*, 285–344.

Krølner, B., & Toft, B. (1983). Vertebral bone loss: An unheeded side effect of therapeutic bed rest. *Clin Sci, 64*, 537–540.

Kummer, J.K. (1999). Implant biomaterials. In J.M. Spivak, P.E. DiCesare, D.S. Feldman, K.J. Koval, A.S. Rokito, & J.D. Zuckerman (Eds.). *Orthopaedics: A Study Guide* (pp. 45–48). New York: McGraw-Hill.

Lanyon, L.E., & Bourn, S. (1979). The influence of mechanical function on the development and remodeling of the tibia: An experimental study in sheep. *J Bone Joint Surg, 61A*, 263.

Lanyon, L.E., Hampson, W.G.J., Goodship, A.E., et al. (1975). Bone deformation recorded in vivo from strain gauges attached to the human tibial shaft. *Acta Orthop Scand, 46*, 256.

Nilsson, B.E., & Westlin, N.E. (1971). Bone density in athletes. *Clin Orthop, 77*, 179.

Ozkaya, N., & Nordin, M. (1999). *Fundamentals of Biomechanics: Equilibrium, Motion, and Deformation* (2nd ed.). New York: Springer-Verlag.

Rambaut, P.C., & Johnston, R.S. (1979). Prolonged weightlessness and calcium loss in man. *Acta Astronautica, 6*, 1113.

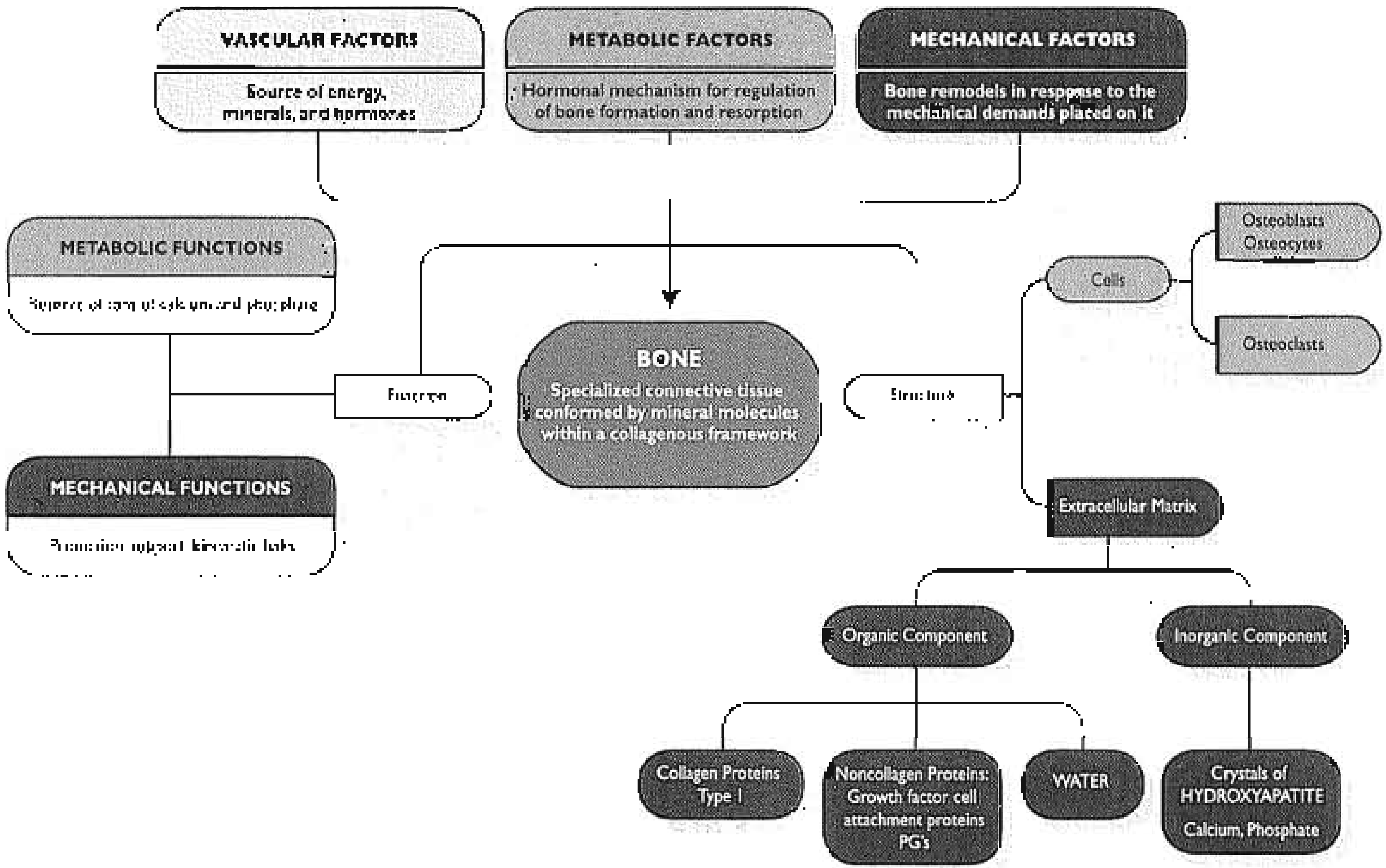
Siffert, R.S., & Levy, R.N. (1981). Trabecular patterns and the internal architecture of bone. *Mt Sinai J Med, 48*, 221.

Slatis, P., Paavolainen, P., Karaharju, E., et al. (1980). Structural and biomechanical changes in bone after rigid plate fixation. *Can J Surg, 23*, 247.

Whedon, G.D. (1984). Disuse osteoporosis: Physiological aspects. *Calcif Tissue Int, 36*, 146–150.

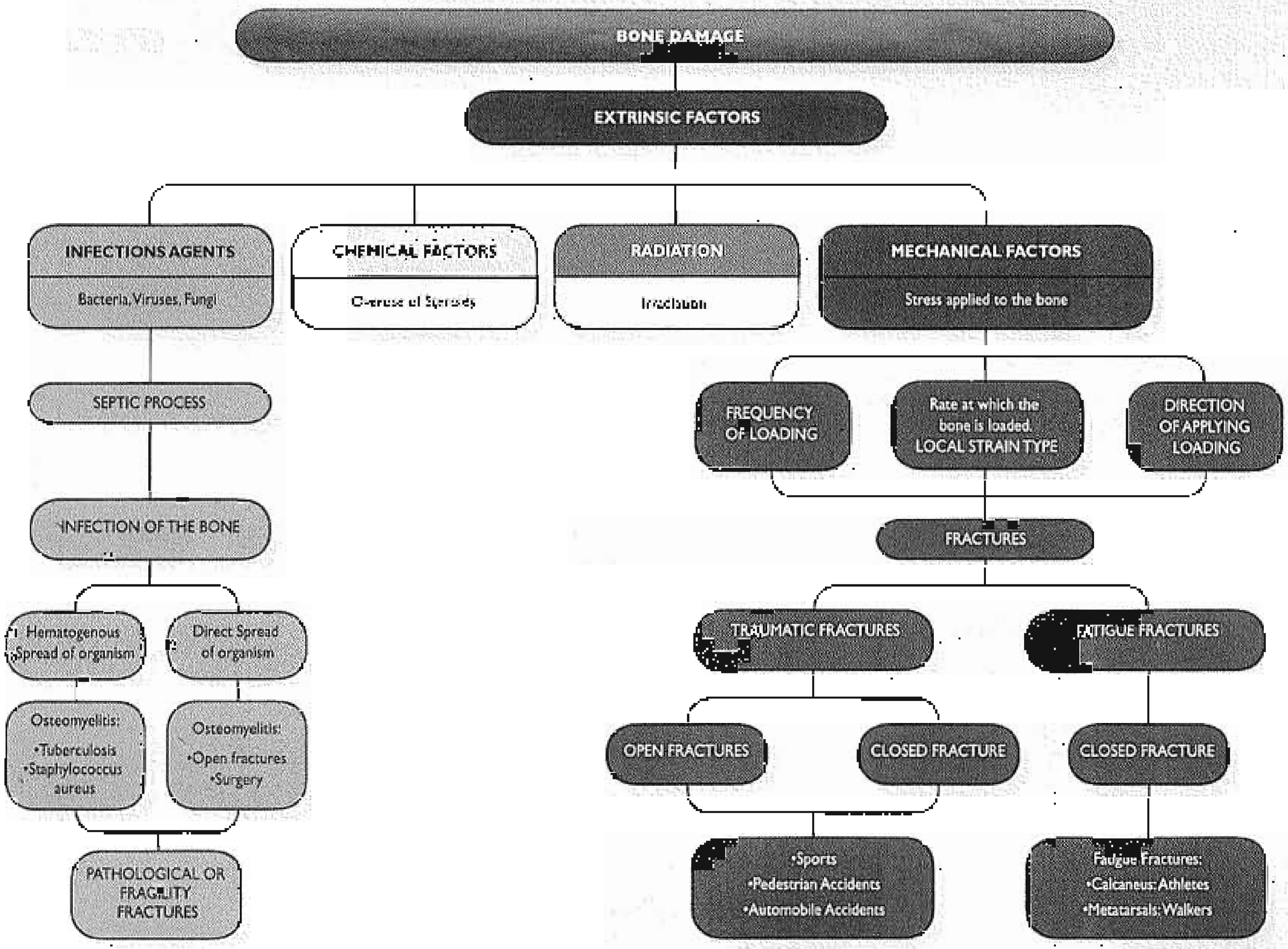
Wolff, J. (1892). *Das Gesetz der Transformation der Knochen*. Berlin: Hirschwald.

Zetterberg, C., Nordin, M., Skovron, M.L., et al. (1990). Skeletal effects of physical activity. *Geriatrika, 7*(14), 17–24.



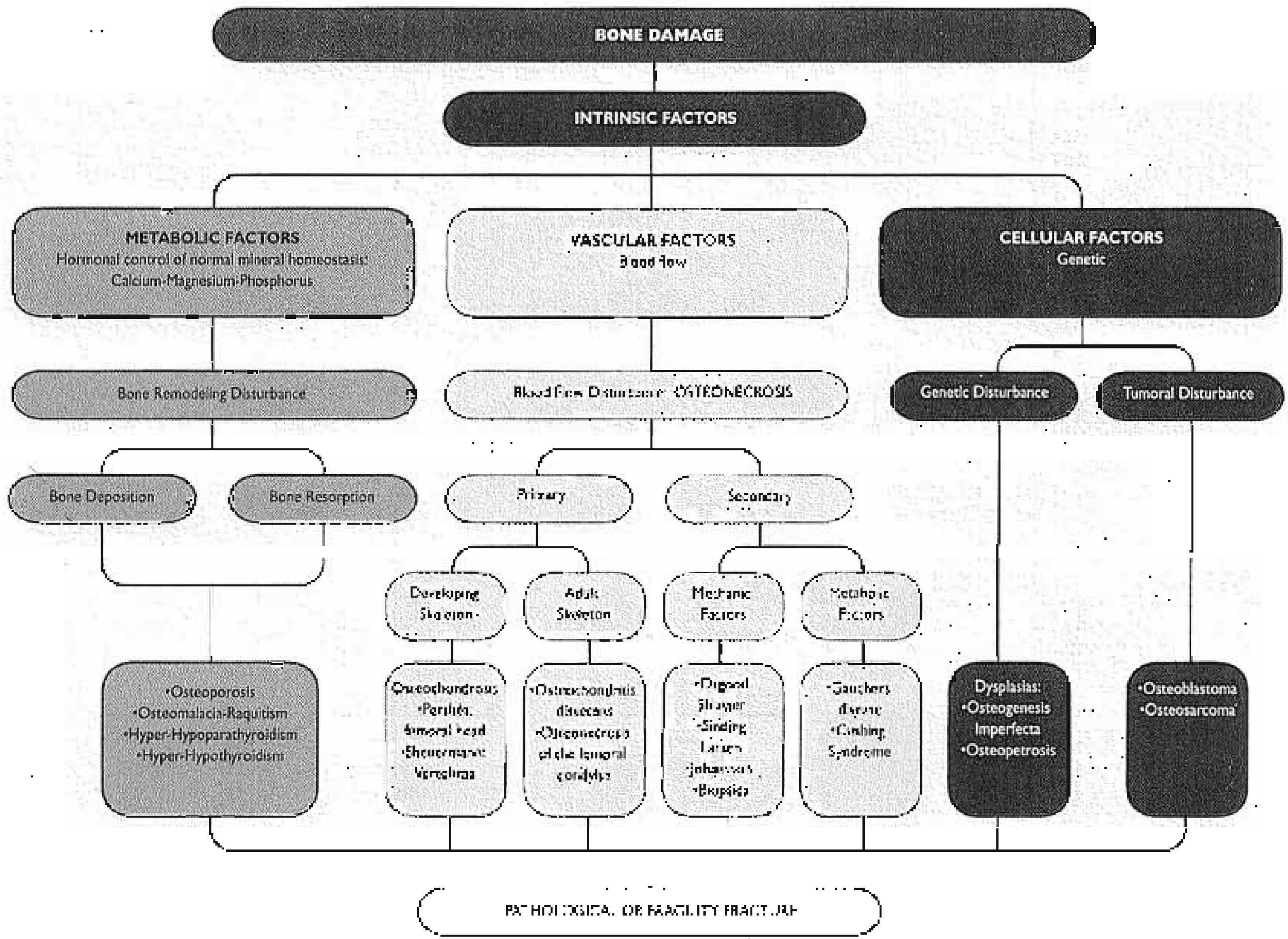
FLOW CHART 2-2 Bone composition, structure, and functions.* (PG's, proteoglycans)

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.



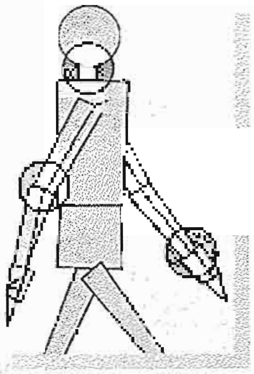
FLOW CHART 2-3 Extrinsic factors associated with bone damage. Clinical examples.*

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.



FLOW CHART 3-4 Intrinsic factors associated with bone damage. Clinical examples.*

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.



CHAPTER 3

Biomechanics of Articular Cartilage

Van C. Mow, Clark T. Hung

Introduction

Three types of joints exist in the human body: fibrous, cartilaginous, and synovial. Only one of these, the synovial, or diarthrodial, joint, allows a large degree of motion. In young normal joints, the articulating bone ends of diarthrodial joints are covered by a thin (1–6 mm), dense, translucent, white connective tissue called hyaline articular cartilage (Box 3-1). Articular cartilage is a highly specialized tissue precisely suited for withstanding the highly loaded joint environment without failure during an average individual's lifetime. Physiologically, however, it is virtually an isolated tissue, devoid of blood vessels, lymphatic channels, and neurological innervation. Furthermore, its cellular density is less than that of any other tissue (Stockwell, 1979).

In diarthrodial joints, articular cartilage has two primary functions: (1) to distribute joint loads over a wide area, thus decreasing the stresses sustained by the contacting joint surfaces (Ateshian et al., 1995; Helminen et al., 1987) and (2) to allow relative movement of the opposing joint surfaces with minimal friction and wear (Mow & Ateshian, 1997). In this chapter, we will describe how the biomechanical properties of articular cartilage, as determined by its composition and structure, allow for the optimal performance of these functions.

BOX 3-1 Hyaline Articular Cartilage

A notable exception to the definition of hyaline articular cartilage is the temporomandibular joint, a synovial joint in which fibrocartilage is found covering the bone ends. Fibrocartilage and a third type of cartilage, elastic cartilage, are closely related to hyaline cartilage embryologically and histologically but are vastly different in mechanical and biochemical properties. Fibrocartilage represents a transitional cartilage found at the margins of some joint cavities, in the joint capsules, and at the insertions of ligaments and tendons into bone.

Fibrocartilage also forms the menisci interposed between the articular cartilage of some joints and composes the outer covering of the intervertebral discs, the anulus fibrosus. Elastic cartilage is found in the external ear, in the cartilage of the eustachian tube, in the epiglottis, and in certain parts of the larynx.

Composition and Structure of Articular Cartilage

Chondrocytes, the sparsely distributed cells in articular cartilage, account for less than 10% of the tissue's volume (Stockwell, 1979). Schematically, the zonal arrangement of chondrocytes is shown in Figure 3-1. Despite their sparse distribution, chondrocytes manufacture, secrete, organize, and maintain the organic component of the extracellular matrix (ECM) (Fosang & Hardingham, 1996; Muir, 1983). The organic matrix is composed of a dense network of fine collagen fibrils (mostly type II collagen, with minor amounts of types V, VI, IX, and XI) that are enmeshed in a concentrated solution of proteoglycans (PGs) (Bateman et al., 1996; Eyre, 1980; Muir, 1983). In normal articular cartilage, the collagen content ranges from 15 to 22% by wet weight and the PG content from 4 to 7% by wet weight; the remaining 60 to 85% is water, inorganic salts, and small amounts of other matrix proteins, glycoproteins, and lipids (Mow & Ratcliffe, 1997). Collagen fibrils and PGs, each being capable of forming structural networks of significant strength (Broom & Silyn-Roberts, 1990; Kempson et al., 1976; Schmidt et al., 1990; Zhu et al., 1991, 1993), are the structural components supporting the internal mechanical stresses that result from loads being applied to the articular cartilage. Moreover, these structural components, together with water, determine the biomechanical behavior of this tissue (Ateshian et al., 1997; Maroudas, 1979; Mow et al., 1980, 1984; Mow & Ateshian, 1997).

COLLAGEN

Collagen is the most abundant protein in the body (Bateman et al., 1996; Eyre, 1980). In articular cartilage, collagen has a high level of structural organization that provides a fibrous ultrastructure (Clark, 1985; Clarke, 1971; Mow & Ratcliffe, 1997). The basic biological unit of collagen is tropocollagen, a structure composed of three procollagen polypeptide chains (alpha chains) coiled into left-handed helices (Fig. 3-2A) that are further coiled about each other into a right-handed triple helix (Fig. 3-2B). These rod-like tropocollagen molecules, 1.4 nanometers (nm) in diameter and 300 nm long (Fig. 3-2, *C & D*), polymerize into larger collagen fibrils (Bateman et al., 1996; Eyre, 1980). In articular cartilage, these fibrils have an average diameter of 25 to 40 nm (Fig. 3-2E; Box 3-2); however, this is highly variable.

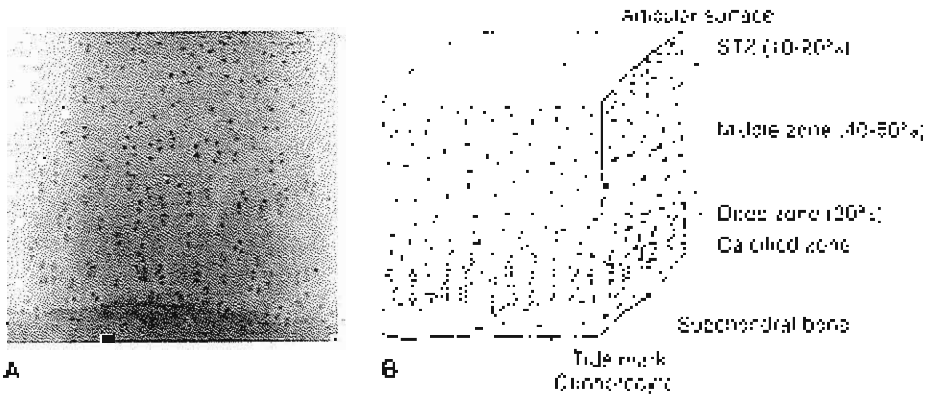


FIG. 3-1

Photomicrograph (A) and schematic representation (B) of the chondrocyte arrangement throughout the depth of noncalcified articular cartilage. In the superficial tangential zone chondrocytes are oblong with their long axes aligned parallel to the articular surface. In the middle zone, the chondrocytes are "round" and randomly distributed. Chondrocytes in the deep zone are arranged in a columnar fashion oriented perpendicular to the tidemark, the demarcation between the calcified and noncalcified tissue.

Scanning electron microscope studies, for instance, have described fibers with diameters ranging up to 200 nm (Clarke, 1971). Covalent cross-links form between these tropocollagen molecules, adding to the fibrils high tensile strength (Bateman et al., 1996).

The collagen in articular cartilage is inhomogeneously distributed, giving the tissue a layered character (Lane & Weiss, 1975; Mow & Ratcliffe, 1997). Numerous investigations using light, transmission electron, and scanning electron microscopy have identified three separate structural zones. For example, Mow et al. (1974) proposed a zonal arrangement for the collagen network shown schematically in Figure 3-3A. In the superficial tangential zone, which represents 10 to 20% of the total thickness, are sheets of fine, densely packed fibers randomly woven in planes parallel to the articular surface (Clarke, 1971; Redler & Zimny, 1970; Weiss et al., 1968). In the middle zone (40 to 60% of the total thickness), there are greater distances between the randomly oriented and homogeneously dispersed fibers. Below this, in the deep zone (approximately 30% of the total thickness), the fibers come together, forming larger, radially oriented fiber bundles (Redler et al., 1975). These bundles then cross the tidemark, the interface between articular cartilage and the calcified cartilage beneath it, to enter the calcified cartilage, thus forming an interlocking "root" system anchoring the cartilage to the underlying bone (Bullough & Jagannath, 1983; Redler et al., 1975). This anisotropic fiber orientation is mirrored by the inhomogeneous zonal variations in the collagen content, which is highest at the surface and then remains relatively constant throughout the deeper zones (Lipshitz et al., 1975). This compositional layering appears to provide an important biomechanical function by distributing the stress more uniformly across the loaded regions of the joint tissue (Setton et al., 1995).

Cartilage is composed primarily of type II collagen. In addition, an array of different collagen (types V, VI, IX, XI) can be found in quantitatively minor amounts within articular cartilage. Type II collagen is present primarily in articular cartilage, the nasal septum, and sternal cartilage as well as in

BOX 3-2 Differences in Collagen Types

Differences in the types of collagen and their ratios in various body tissues are due to specific molecular species, or types of collagen. The collagen type in hyaline cartilage, type II collagen, is finer than type I collagen found in bone, tendons, and skin. Type II collagen forms smaller fibrils than that of type I, permitting maximum dispersion of collagen throughout the cartilage tissue.

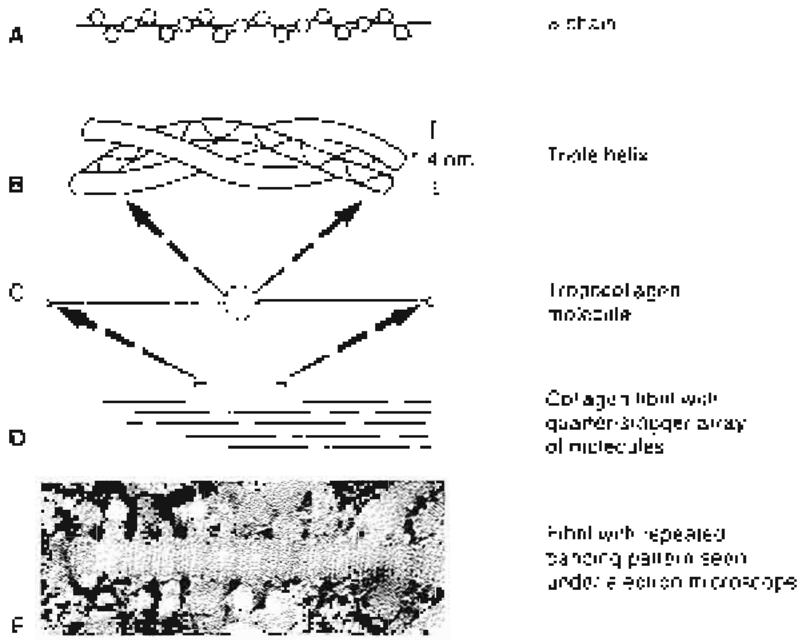


FIG. 3-2

Molecular features of collagen structure from the alpha chain (α) to the fibril. The flexible amino acid sequence in the alpha chain (A) allows these chains to wind tightly into a right-handed triple helix configuration (B), thus forming the tropocollagen molecule (C). This tight triple helical arrangement of the chains contributes to the high tensile strength of the collagen fibril. The parallel alignment of the individual tropocollagen molecules, in which each molecule overlaps the other by about one quarter of its length (D), results in a repeating banded pattern of the collagen fibril seen by electron microscopy (×20,000) (E). Reprinted with permission from Donohue, J.M., Buss, D., Oegema, T.R., et al. (1983). The effects of indirect blunt trauma on adult canine articular cartilage. J Bone Joint Surg, 65A, 948.

the inner regions of the intervertebral disc and meniscus. For reference, type I is the most abundant collagen in the human body and can be found in bone and soft tissues such as intervertebral discs (mainly in the annulus fibrosis), skin, meniscus, tendons, and ligaments. The most important mechanical properties of collagen fibers are their tensile stiffness and their strength (Fig. 3-4A). Although a single collagen fibril has not been tested in tension, the tensile strength of collagen can be inferred from tests on structures with high collagen content. Tendons, for example, are about 80% collagen (dry weight) and have a tensile stiffness of 10^3 MPa and a tensile strength of 50 MPa (Akizuki et al., 1986; Kempson, 1976, 1979; Woo et al., 1987, 1997). Steel, by comparison, has a tensile stiffness of approximately 220×10^3 MPa. Although strong in tension, collagen fibrils offer little resistance to compression because their large slenderness ratio, the ratio of length to thickness, makes it easy for them to buckle under compressive loads (Fig. 3-4B).

Like bone, articular cartilage is anisotropic; its material properties differ with the direction of loading (Akizuki et al., 1986; Kempson, 1979; Mow & Ratcliffe, 1997; Roth & Mow, 1980; Woo et al., 1987). It is thought that this anisotropy is related to the varying collagen fiber arrangements within the planes parallel to the articular surface. It is also thought, however, that variations in collagen fiber cross-link density, as well as variations in collagen-PG interactions, also contribute to articular cartilage tensile anisotropy. In tension, this anisotropy is usually described with respect to the direction of the articular surface split lines. These split lines are elongated fissures produced by piercing the articular surface with a small, round awl (Fig. 3-5;

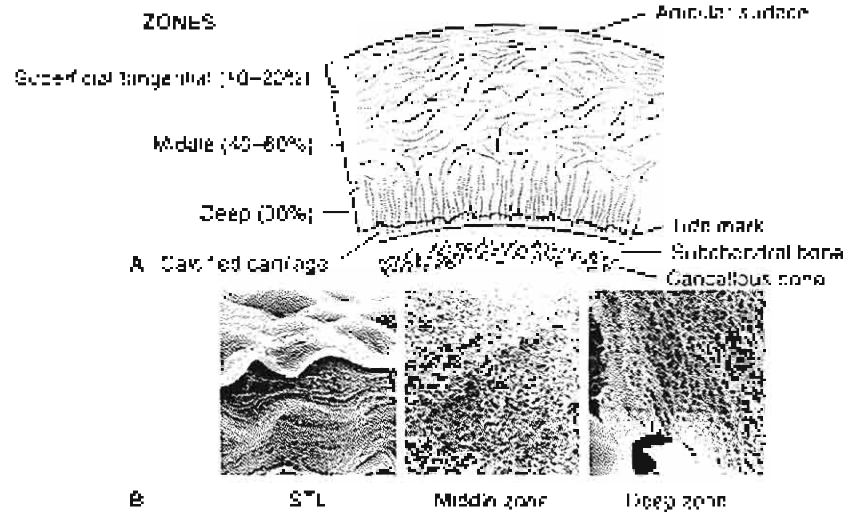


FIG. 3-3

A, Schematic representation. (*Reprinted with permission from Mow VC, et al. (1974). Some surface characteristics of articular cartilage. A scanning electron microscopy study and a theoretical model for the dynamic interaction of synovial fluid and articular cartilage. J Biomech, 7, 449.*) B, Photomicrographs (× 3000) (provided through the courtesy of Dr. T. Takei, Nagano, Japan) of the ultrastructural arrangement of the collagen network throughout the depth of articular cartilage. In the superficial tangential zone (STZ), collagen fibrils are tightly woven into sheets arranged parallel to the articular surface. In the middle zone, randomly arrayed fibrils are less densely packed to accommodate the high concentration of proteoglycans and water. The collagen fibrils of the deep zone form larger radially oriented fiber bundles that cross the tidemark, enter the calcified zone, and anchor the tissue to the underlying bone. Note the correspondence between this collagen fiber architecture and the spatial arrangement of the chondrocytes shown in Figure 3-1. In the above photomicrographs (B), the STZ is shown under compressive loading while the middle and deep zones are *unloaded*.

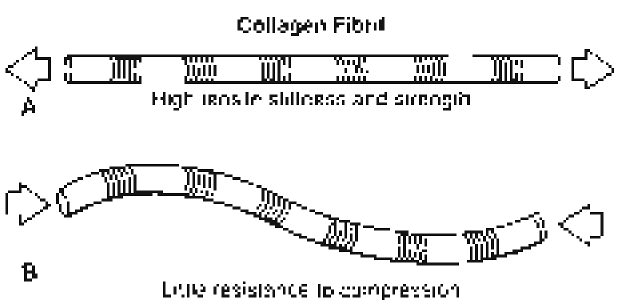


FIG. 3-4

Illustration of the mechanical properties of collagen fibrils: (A) stiff and strong in tension, but (B) weak and buckling easily with compression. Adapted from *Myers, E.R., Lai, W.M., & Mow, V.C. (1984). A continuum theory and an experiment for the ion-induced swelling behavior cartilage. J Biomech Eng, 106(2), 151–158.*

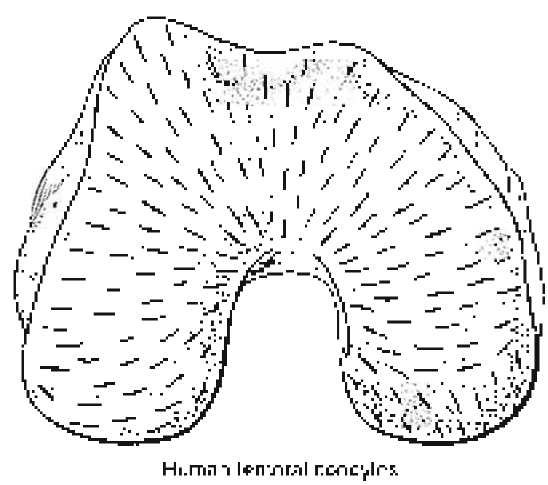


FIG. 3-5

Diagrammatic representation of a split line pattern on the surface of human femoral condyles. (*Reprinted with permission from Hultkrantz, W. (1898). Ueber die Spaltrichtungen der Gelenkknorpel. Verhandlungen der Anatomischen Gesellschaft, 12, 248.*)

Hultkrantz, 1898). The origin of the pattern is related to the directional variation of the tensile stiffness and strength characteristics of articular cartilage described above. To date, however, the exact reasons as to why articular cartilage exhibits such pronounced anisotropies in tension is not known, nor is the functional significance of this tensile anisotropy.

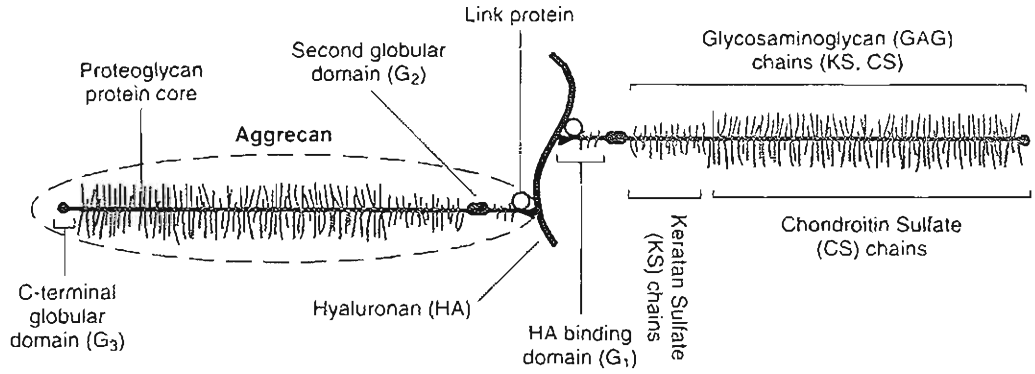
PROTEOGLYCAN

Many types of PGs are found in cartilage. Fundamentally, it is a large protein-polysaccharide molecule composed of a protein core to which one or more glycosaminoglycans (GAGs) are attached (Fosang & Hardingham, 1996; Muir, 1983; Ratcliffe & Mow, 1996). Even the smallest of these molecules, biglycan and decorin, are quite large (approximately 1×10^4 mw), but they comprise less than 10% of all PGs present in the tissue. Aggrecans are much larger ($1–4 \times 10^6$ mw), and they have the remarkable capability to attach to a hyaluronan molecule (HA; 5×10^5 mw) via a specific HA-binding region (HABR). This binding is stabilized by a link protein (LP) ($40–48 \times 10^3$ mw). Stabilization is critical to the function of normal cartilage; without it, the components of the PG molecule would rapidly escape from the tissue (Hardingham & Muir, 1974; Hascall, 1977; Muir, 1983).

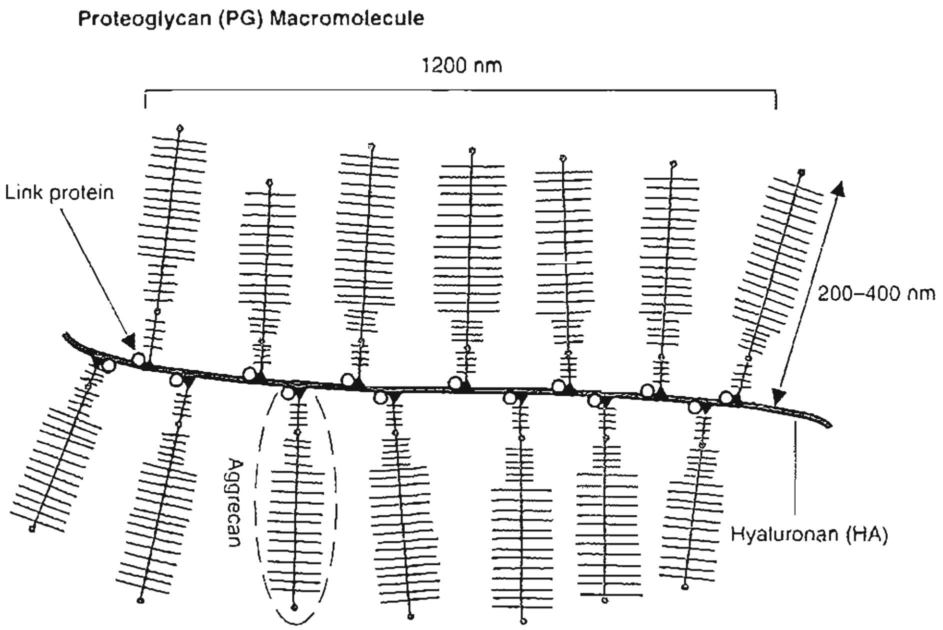
Two types of GAGs comprise aggrecan: chondroitin sulfate (CS) and keratan sulfate (KS). Each CS chain contains 25 to 30 disaccharide units, while the shorter KS chain contains 13 disaccharide units (Muir, 1983). Aggrecans (previously referred to as subunits in the American literature or as monomers in the UK and European literature) consist of an approximately 200-nanometer-long protein core to which approximately 150 GAG chains, and both O-linked and N-linked oligosaccharides, are covalently attached (Fosang & Hardingham, 1996; Muir, 1983). Furthermore, the distribution of GAGs along the protein core is heterogeneous; there is a region rich in KS and O-linked oligosaccharides and a region rich in CS (Fig. 3-6A). Figure 3-6A depicts the famous "bottle-brush" model for an aggrecan (Muir, 1983). Also shown in Figure 3-6A is the heterogeneity of the protein core that contains three globular regions: G_1, the HABR located at the N-terminus that contains a small amount of KS (Poole, 1986) and a few N-linked oligosaccharides, G_2, located between the HABR- and the KS-rich region (Hardingham et al., 1987), and G_3, the core protein C-terminus. A 1:1 stoichiometry exists between the LP and the G_1 binding region in cartilage. More recently, the other two globular regions have been extensively studied (Fosang & Hardingham, 1996), but their functional significance has not yet been elucidated. Figure 3-6B is the accepted molecular conformation of a PG aggregate; Rosenberg et al. (1975) were the first to obtain an electron micrograph of this molecule (Fig. 3-6C).

In native cartilage, most aggrecans are associated with HA to form the large PG aggregates (Fig. 3-6C). These aggregates may have up to several hundred aggrecans noncovalently attached to a central HA core via their HABR, and each site is stabilized by an LP. The filamentous HA core molecule is a nonsulfated disaccharide chain that may be as long as 4 μm in length. PG biochemists have dubbed the HA an "honorary" PG, as it is so intimately involved in the structure of the PG aggregate in articular cartilage. The stability afforded by the PG aggregates has a major functional significance. It is accepted now that PG aggregation promotes immobilization of the PGs within the fine collagen meshwork, adding structural stability and rigidity to the ECM (Mow et al., 1989b; Muir, 1983; Ratcliffe et al., 1986). Furthermore, two additional forms of dermatan sulfate PG have been identified in the ECM of articular cartilage (Rosenberg et al., 1985). In tendon, dermatan sulfate PGs have been shown to bind noncovalently to the surfaces of collagen fibrils (Scott & Orford, 1981); however, the role of dermatan sulfate in articular cartilage is unknown, biologically and functionally.

Although aggrecans generally have the basic structure as described above, they are not structurally identical (Fosang & Hardingham, 1996). Aggrecans vary in length, molecular weight, and composition in a variety of ways; in other words, they are polydisperse. Studies have demonstrated two distinct populations of aggrecans (Buckwalter et al., 1985; Heinegard et al., 1985). The first population is present throughout life and is rich in CS; the second contains PGs rich in KS and is present only in adult cartilage. As articular cartilage matures, other age-related changes in PG composition and structure occur. With cartilage maturation, the water content (Armstrong & Mow, 1982; Bollet & Nance, 1965; Linn & Sokoloff, 1965; Maroudas, 1979; Venn, 1978) and the carbohydrate/protein ratio progressively decrease (Garg & Swann, 1981; Roughley & White, 1980). This decrease is mirrored by a decrease in the CS content. Conversely, KS, which is present only in small amounts at birth, increases throughout development and aging. Thus, the



A



B

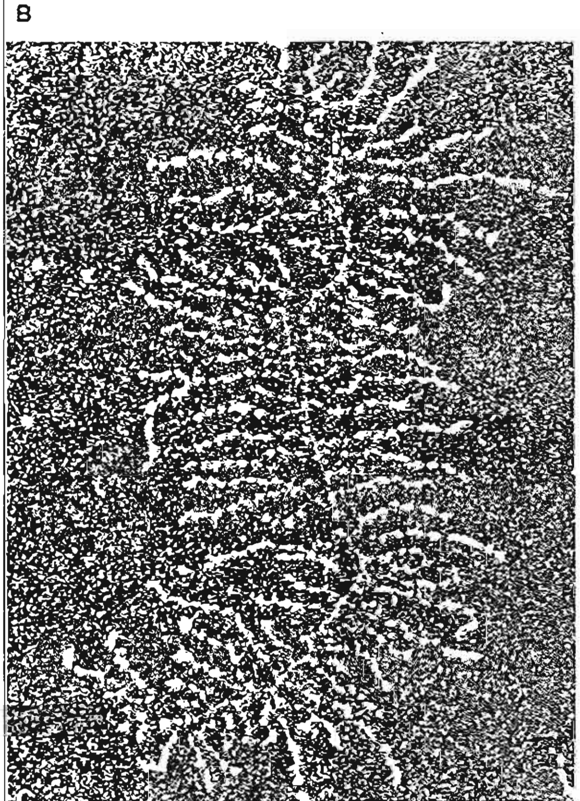


FIG. 3-6

A, Schematic depiction of aggrecan, which is composed of keratan sulfate and chondroitin sulfate chains bound covalently to a protein core molecule. The proteoglycan protein core has three globular regions as well as keratan sulfate-rich and chondroitin sulfate-rich regions. B, Schematic representation of a proteoglycan macromolecule. In the matrix, aggrecan noncovalently binds to HA to form a macromolecule with a molecular weight of approximately 200 × 106. Link protein stabilizes this interaction between the binding region of the aggrecan and the HA core molecule. C, Dark field electron micrograph of a proteoglycan aggregate from bovine humeral articular cartilage (×120,000). *Horizontal line* at lower right represents 0.5μm. *Reprinted with permission from Rosenberg, L., Hellmann, W., & Kleinschmidt, A.K. (1975). Electron microscopic studies of proteoglycan aggregates from bovine articular cartilage.* J Biol Chem, 250, *1877.*

CS/KS ratio, which is approximately 10:1 at birth, is only approximately 2:1 in adult cartilage (Roughley & White, 1980; Sweet et al., 1979; Thonar et al., 1986). Furthermore, sulfation of the CS molecules, which can occur at either the 6 or the 4 position, also undergoes age-related changes. In utero, chondroitin-6-sulfate and chondroitin-4-sulfate are present in equal molar amounts; however, by maturity, the chondroitin-6-sulfate:chondroitin-4-sulfate ratio has increased to approximately 25:1 (Roughley et al., 1981). Other studies have also documented an age-related decrease in the hydrodynamic size of the aggrecan. Many of these early changes seen in articular cartilage may reflect cartilage maturation, possibly as a result of increased functional demand with increased weight-bearing. However, the functional significance of these changes, as well as those occurring later in life, is as yet undetermined.

WATER

Water, the most abundant component of articular cartilage, is most concentrated near the articular surface (~80%) and decreases in a near-linear fashion with increasing depth to a concentration of approximately 65% in the deep zone (Lipshitz et al., 1976; Maroudas, 1979). This fluid contains many free mobile cations (e.g., Na^+, K^+, and Ca^{2+}) that greatly influence the mechanical and physicochemical behaviors of cartilage (Gu et al., 1998; Lai et al., 1991; Linn & Sokoloff, 1965; Maroudas, 1979). The fluid component of articular cartilage is also essential to the health of this avascular tissue because it permits gas, nutrient, and waste product movement back and forth between chondrocytes and the surrounding nutrient-rich synovial fluid (Bollet & Nance, 1965; Linn & Sokoloff, 1965; Mankin & Thrasher, 1975; Maroudas, 1975, 1979).

A small percentage of the water in cartilage resides intracellularly, and approximately 30% is strongly associated with the collagen fibrils (Maroudas et al., 1991; Torzilli et al., 1982). The interaction between collagen, PG, and water, via Donnan osmotic pressure, is believed to have an important function in regulating the structural organization of the ECM and its swelling properties (Donnan, 1924; Maroudas, 1968, 1975). Most of the water thus occupies the interfibrillar space of the ECM and is free to move when a load or pressure gradient or other electrochemical motive forces are applied to the tissue (Gu et al., 1998; Maroudas, 1979). When loaded by a compressive force, approximately 70% of the water may be moved. This interstitial fluid movement is important in controlling cartilage mechanical behavior and joint lubrication (Ateshian et al., 1997, 1998; Hlavacek, 1995; Hou et al., 1992; Mow et al., 1980; Mow & Ateshian, 1997).

STRUCTURAL AND PHYSICAL INTERACTION AMONG CARTILAGE COMPONENTS

The chemical structure and physical interactions of the PG aggregates influence the properties of the ECM (Ratcliffe & Mow, 1996). The closely spaced (5–15 angstroms) sulfate and carboxyl charge groups on the CS and KS chains dissociate in solution at physiological pH (Fig. 3-7), leaving a high concentration of fixed negative charges that create strong intramolecular and intermolecular charge-charge repulsive forces; the colligative sum of these forces (when the tissue is immersed in a physiological saline solution) is equivalent to the Donnan osmotic pressure (Buschmann & Grodzinsky, 1995; Donnan, 1924; Gu et al., 1998; Lai et al., 1991). Structurally, these charge-charge repulsive forces tend to extend and stiffen the PG macromolecules into the interfibrillar space formed by the surrounding collagen network. To appreciate the magnitude of this force, according to Stephen Hawkings (1988), this electrical repulsion is one million, million, million, million, million, million, million times (42 zeros) greater than gravitational forces.

In nature, a charged body cannot persist long without discharging or attracting counter-ions to maintain electroneutrality. Thus, the charged sulfate and carboxyl groups fixed along the PGs in articular cartilage must attract various counter-ions and co-ions (mainly Na^+, Ca^{2+}, and Cl^-) into the tissue to maintain electroneutrality. The total concentration of these counter-ions and co-ions is given by the well-known Donnan equilibrium ion distribution law (Donnan, 1924). Inside the tissue, the mobile counter-ions and co-ions form a cloud surrounding the fixed sulfate and carboxyl charges, thus shielding these charges from each other. This charge shielding acts to diminish the very large electrical repulsive forces that otherwise would exist. The net result is a swelling pressure given by the Donnan osmotic pressure law (Buschmann & Grodzinsky, 1995; Donnan, 1924; Gu et al., 1998; Lai et al., 1991; Schubert & Hamerman, 1968). The Donnan osmotic pressure theory has been extensively used to calculate the

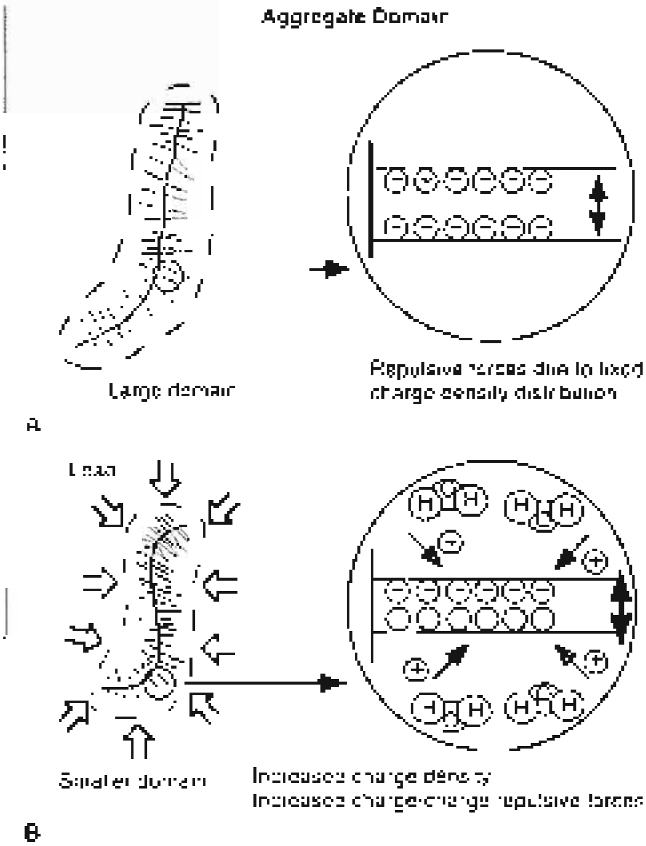


FIG. 3-7

A, Schematic representation of a proteoglycan aggregate solution domain (*left*) and the repelling forces associated with the fixed negative charge groups on the GAGs of aggrecan (*right*). These repulsive forces cause the aggregate to assume a stiffly extended conformation, occupying a large solution domain. B, Applied compressive stress decreases the aggregate solution domain (*left*), which in turn increases the charge density and thus the intermolecular charge repulsive forces (*right*).

swelling pressures of articular cartilage and the intervertebral disc (Maroudas, 1979; Urban & McMullin, 1985). By Starling's law, this swelling pressure is, in turn, resisted and balanced by tension developed in the collagen network, confining the PGs to only 20% of their free solution domain (Maroudas, 1976; Mow & Ratcliffe, 1997; Setton et al., 1995). Consequently, this swelling pressure subjects the collagen network to a "pre-stress" of significant magnitude even in the absence of external loads (Setton et al., 1995, 1998).

Cartilage PGs are inhomogeneously distributed throughout the matrix, with their concentration generally being highest in the middle zone and lowest in the superficial and deep zones (Lipshitz et al., 1976; Maroudas, 1968, 1979; Venn, 1978). The biomechanical consequence of this inhomogeneous swelling behavior of cartilage (caused by the varying PG content throughout the depth of the tissue) has recently been quantitatively assessed (Setton et al., 1998). Also, results from recent finite element calculations based on models incorporating an inhomogeneous PG distribution show that it has a profound effect on the interstitial counter-ion distribution throughout the depth of the tissue (Sun et al., 1998).

When a compressive stress is applied to the cartilage surface, there is an instantaneous deformation caused primarily by a change in the PG molecular domain, Figure 3-7B. This external stress causes the internal pressure in the matrix to exceed the swelling pressure and thus liquid will begin to flow out of the tissue. As the fluid flows out, the PG concentration increases, which in turn increases the Donnan osmotic swelling pressure or the charge-charge repulsive force and bulk compressive stress until they are in equilibrium with the external stress. In this manner, the physicochemical properties of the PG gel trapped within the collagen network enable it to resist compression. This mechanism complements the role played by collagen that, as previously described, is strong in tension but weak in compression. The ability of PGs to resist compression thus arises from two sources: (1) the Donnan osmotic swelling pressure associated with the tightly packed fixed anionic groups on the GAGS and (2) the bulk compressive stiffness of the collagen-PG solid matrix. Experimentally, the Donnan osmotic pressure ranges from 0.05 to 0.35 MPa (Maroudas, 1979), while the elastic modulus of the collagen-PG solid matrix ranges from 0.5 to 1.5 MPa (Armstrong & Mow, 1982; Athanasiou et al., 1991; Mow & Ratcliffe, 1997).

It is now apparent that collagen and PGs also interact and that these interactions are of great functional importance. A small portion of the PGs have been shown to be closely associated with collagen and may serve as a bonding agent between the collagen fibrils, spanning distances that are too great for collagenic crosslinks to develop (Bateman et al., 1996; Mow & Ratcliffe, 1997; Muir, 1983).

PGs are also thought to play an important role in maintaining the ordered structure and mechanical properties of the collagen fibrils (Muir, 1983; Scott & Orford, 1981). Recent investigations show that in concentrated solutions, PGs interact with each other to form networks of significant strength (Mow et al., 1989b; Zhu et al., 1991, 1996). Moreover, the density and strength of the interaction sites forming the network were shown to depend on the presence of LP between aggrecans and aggregates as well as collagen. Evidence suggests that there are fewer aggregates, and more biglycans and decorins than aggrecans, in the superficial zone of articular cartilage. Thus, there must be a difference in the interaction between these PGs and the collagen fibrils from the superficial zone than from those of the deeper zones (Poole et al., 1986). Indeed, the interaction between PG and collagen not only plays a direct role in the organization of the ECM but also contributes directly to the mechanical properties of the tissue (Kempson et al., 1976; Schmidt et al., 1990; Zhu et al., 1993).

The specific characteristics of the physical, chemical, and mechanical interactions between collagen and PG have not yet been fully determined. Nevertheless, as discussed above, we know that these structural macromolecules interact to form a porous-permeable fiber-reinforced composite matrix possessing all the essential mechanical characteristics of a solid that is swollen with water and ions and that is able to resist the high stresses and strains of joint articulation (Andriacchi et al., 1997; Hodge et al., 1986; Mow & Ateshian, 1997; Paul, 1976). It has been demonstrated that these collagen-PG interactions involve an aggrecan, an HA filament, type II collagen, other minor collagen types, an unknown bonding agent, and possibly smaller cartilage components such as collagen type IX, recently identified glycoproteins, and/or polymeric HA (Poole et al., 1986). A schematic diagram depicting the structural arrangement within a small volume of articular cartilage is shown in Figure 3-8.

When articular cartilage is subjected to external loads, the collagen-PG solid matrix and interstitial fluid function together in a unique way to protect against high levels of stress and strain developing in the ECM. Furthermore, changes to the biochemical composition and structural organization of the ECM, such as during osteoarthritis (OA), are paralleled by changes to the biomechanical properties of cartilage. In the following section, the behavior of articular cartilage under loading and the mechanisms of cartilage fluid flow will be discussed in detail.

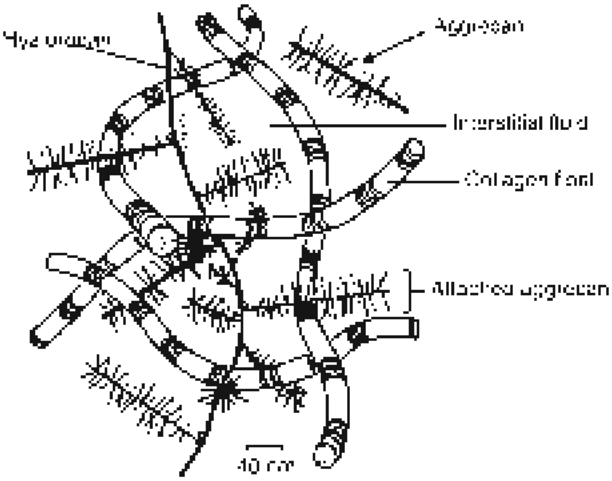


FIG. 3-8

Schematic representation of the molecular organization of cartilage. The structural components of cartilage, collagen, and proteoglycans, interact to form a porous composite fiber-reinforced organic solid matrix that is swollen with water. Aggrecans bind covalently to HA to form large proteoglycan macromolecules.

Biomechanical Behavior of Articular Cartilage

The biomechanical behavior of articular cartilage can best be understood when the tissue is viewed as a multiphasic medium. In the present context, articular cartilage will be treated as biphasic material consisting of two intrinsically incompressible, immiscible, and distinct phases (Bachrach et al., 1998; Mow et al., 1980): an interstitial fluid phase and a porous-permeable solid phase (i.e., the ECM). For explicit analysis of the contribution of the PG charges and ions, one would have to consider three distinct phases: a fluid phase, an ion phase, and a charged solid phase (Gu et al., 1998; Lai et al., 1991). For understanding how the water contributes to its mechanical properties. In the present context, articular cartilage may be considered as a fluid-filled porous-permeable (uncharged) biphasic medium, with each constituent playing a role in the functional behavior of cartilage.

During joint articulation, forces at the joint surface may vary from almost zero to more than ten times body weight (Andriacchi et al., 1997; Paul, 1976). The contact areas also vary in a complex manner, and typically they are only of the order of several square centimeters (Ahmed & Burke, 1983; Ateshian et al., 1994). It is estimated that the peak contact stress may reach 20 MPa in the hip while rising from a chair and 10 MPa during stair climbing (Hodge et al., 1986; Newberry et al., 1997). Thus, articular cartilage under physiological loading conditions is a highly stressed material. To understand how this tissue responds under these high physiological loading conditions, its intrinsic mechanical properties in compression, tension, and shear must be determined. From these properties, one can understand the load-carrying mechanisms within the ECM. Accordingly, the following subsections will characterize the tissue behavior under these loading modalities.

NATURE OF ARTICULAR CARTILAGE VISCOELASTICITY

If a material is subjected to the action of a constant (time-independent) load or a constant deformation and its response varies with time, then the mechanical behavior of the material is said to be viscoelastic. In general, the response of such a material can be theoretically modeled as a combination of the response of a viscous fluid (dashpot) and an elastic solid (spring), hence viscoelastic.

The two fundamental responses of a viscoelastic material are creep and stress relaxation. Creep occurs when a viscoelastic solid is subjected to the action of a constant load. Typically, a viscoelastic solid responds with a rapid initial deformation followed by a slow (time-dependent), progressively increasing deformation known as creep until an equilibrium state is reached. Stress relaxation occurs when a viscoelastic solid is subjected to the action of a constant deformation. Typically, a viscoelastic solid responds with a rapid, high initial stress followed by a slow (time-dependent), progressively decreasing stress required to maintain the deformation; this phenomenon is known as stress relaxation.

Creep and stress relaxation phenomena may be caused by different mechanisms. For single-phase solid polymeric materials, these phenomena are the result of internal friction caused by the motion of the long polymeric chains sliding over each other within the stressed material (Fung, 1981). The viscoelastic behavior of tendons and ligaments is primarily caused by this mechanism (Woo et al., 1987, 1997). For bone, the long-term viscoelastic behavior is thought to be caused by a relative slip of lamellae within the osteons along with the flow of the interstitial fluid (Lakes & Saha, 1979). For articular cartilage, the compressive viscoelastic behavior is primarily caused by the flow of the interstitial fluid and the frictional drag associated with this flow (Ateshian et al., 1997; Mow et al., 1980, 1984). In shear, as in single-phase viscoelastic polymers, it is primarily caused by the motion of long polymer chains such as collagen and PGs (Zhu et al., 1993, 1996). The component of articular cartilage viscoelasticity caused by interstitial fluid flow is known as the biphasic viscoelastic behavior (Mow et al., 1980), and the component of viscoelasticity caused by macromolecular motion is known as the flow-independent (Hayes & Bodine, 1978) or the intrinsic viscoelastic behavior of the collagen-PG solid matrix.

Although the deformational behavior has been described in terms of a linear elastic solid (Hirsch, 1944) or viscoelastic solid (Hayes & Mockros, 1971), these models fail to recognize the role of water in the viscoelastic behavior and the significant contribution that fluid pressurization plays in joint load support and cartilage lubrication (Ateshian et al., 1998; Elmore et al., 1963; Mow & Ratcliffe, 1997; Sokoloff, 1966). Recently, experimental measurements have determined that interstitial fluid pressurization supports more than 90% of the applied load to the cartilage surface (Soltz & Ateshian, 1998) immediately following loading. This effect can persist for more than 1,000 seconds and thus shields the ECM and chondrocytes from the crushing deformations of the high stresses (20 MPa) resulting from joint loading.

CONFINED COMPRESSION EXPLANT LOADING CONFIGURATION

The loading of cartilage in vivo is extremely complex. To achieve a better understanding of the deformational behavior of the tissue under load, an explant loading configuration known as confined compression (Mow et al., 1980) has been adopted by researchers. In this configuration, a cylindrical cartilage specimen is fitted snugly into a cylindrical, smooth-walled (ideally frictionless) confining ring that prohibits motion and fluid loss in the radial direction. Under an axial loading condition via a rigid porous-permeable loading platen (Fig. 3-9A), fluid will flow from the tissue into the porous-permeable platen, and, as this occurs, the cartilage sample will compress in creep. At any time the amount of compression equals the volume of fluid loss because both the water and the ECM are each intrinsically incompressible (Bachrach et al., 1998). The advantage of the confined compression test is that it creates a uniaxial, one-dimensional flow and deformational field within the tissue, which does not depend on tissue anisotropy or properties in the radial direction. This greatly simplifies the mathematics needed to solve the problem.

It should be emphasized that the stress, strain, pressure, fluid, and ion flow fields generated within the tissue during loading can only be calculated; however, these calculations are of idealized models and testing conditions. There are many confounding factors, such as the time-dependent nature and magnitude of loading and alterations in the natural state of prestress (acting within the tissue), that arise from disruption of the collagen network during specimen harvesting. Despite limitations in determining the natural physiological states of stress and strain within the tissue in vivo, a number of researchers have made gains toward an understanding of potential mechanosignal transduction mechanisms in cartilage through the use of explant loading studies (Bachrach et al., 1995; Buschmann et al., 1992; Kim et al., 1994; Valhmu et al., 1998) based on the biphasic constitutive law for soft hydrated tissues (Mow et al., 1980).

BIPHASIC CREEP RESPONSE OF ARTICULAR CARTILAGE IN COMPRESSION

The biphasic creep response of articular cartilage in a one-dimensional confined compression experiment is depicted in Figure 3-9. In this case, a constant compressive stress (σ_0) is applied to the tissue at time t_0 (point A in Fig. 3-9B) and the tissue is allowed to creep to its final equilibrium strain (ϵ_∞). For articular cartilage, as illustrated in the top diagrams, creep is caused by the exudation of the interstitial fluid. Exudation is most rapid initially, as evidenced by the early rapid rate of increased deformation, and it diminishes gradually until flow cessation occurs. During creep, the load applied at the surface is balanced by the compressive stress developed within the collagen-PG solid matrix and the frictional drag generated by the flow of the interstitial fluid during exudation. Creep ceases when the compressive stress developed within the solid matrix is sufficient to balance the applied stress alone; at this point no fluid flows and the equilibrium strain ϵ_∞ is reached.

Typically, for relatively thick human and bovine articular cartilages, 2 to 4 mm, it takes 4 to 16 hours to reach creep equilibrium. For rabbit cartilage, which is generally less than 1.0 mm thick, it takes approximately 1 hour to reach creep equilibrium. Theoretically, it can be shown that the time it takes to reach creep equilibrium varies inversely with the square of the thickness of the tissue (Mow et al., 1980). Under relatively high loading conditions, >1.0 MPa, 50% of the total fluid content may be squeezed from the tissue (Edwards, 1967). Furthermore, in vitro studies demonstrate that if the tissue is immersed in physiological saline, this exuded fluid is fully recoverable when the load is removed (Elmore et al., 1963; Sokoloff, 1963).

Because the rate of creep is governed by the rate of fluid exudation, it can be used to determine the permeability coefficient of the tissue (Mow et al., 1980, 1989a). This is known as the indirect measurement for tissue permeability (k). Average values of normal human, bovine, and canine patellar groove articular cartilage permeability k obtained in this manner are 2.17×10^{-15} m^4/N·s, 1.42×10^{-15} m^4/N·s, and 0.9342×10^{-15} m^4/N·s, respectively (Athanasiou et al., 1991). At equilibrium, no fluid flow occurs and thus the equilibrium deformation can be used to measure the intrinsic compressive modulus (H_A) of the collagen-PG solid matrix (Armstrong & Mow, 1982; Mow et al., 1980). Average values of normal human, bovine, and canine patellar groove articular cartilage compressive modulus H_A are 0.53, 0.47, and 0.55 megapascal (MPa; note 1.0 MPa = 145 lb/in²), respectively. Because these coefficients are a measure of the intrinsic material properties of the solid matrix, it is therefore meaningful to determine how they vary with matrix composition. It was determined that k varies directly, while H_A varies inversely with water content and varies directly with PG content (Mow & Ratcliffe, 1997).

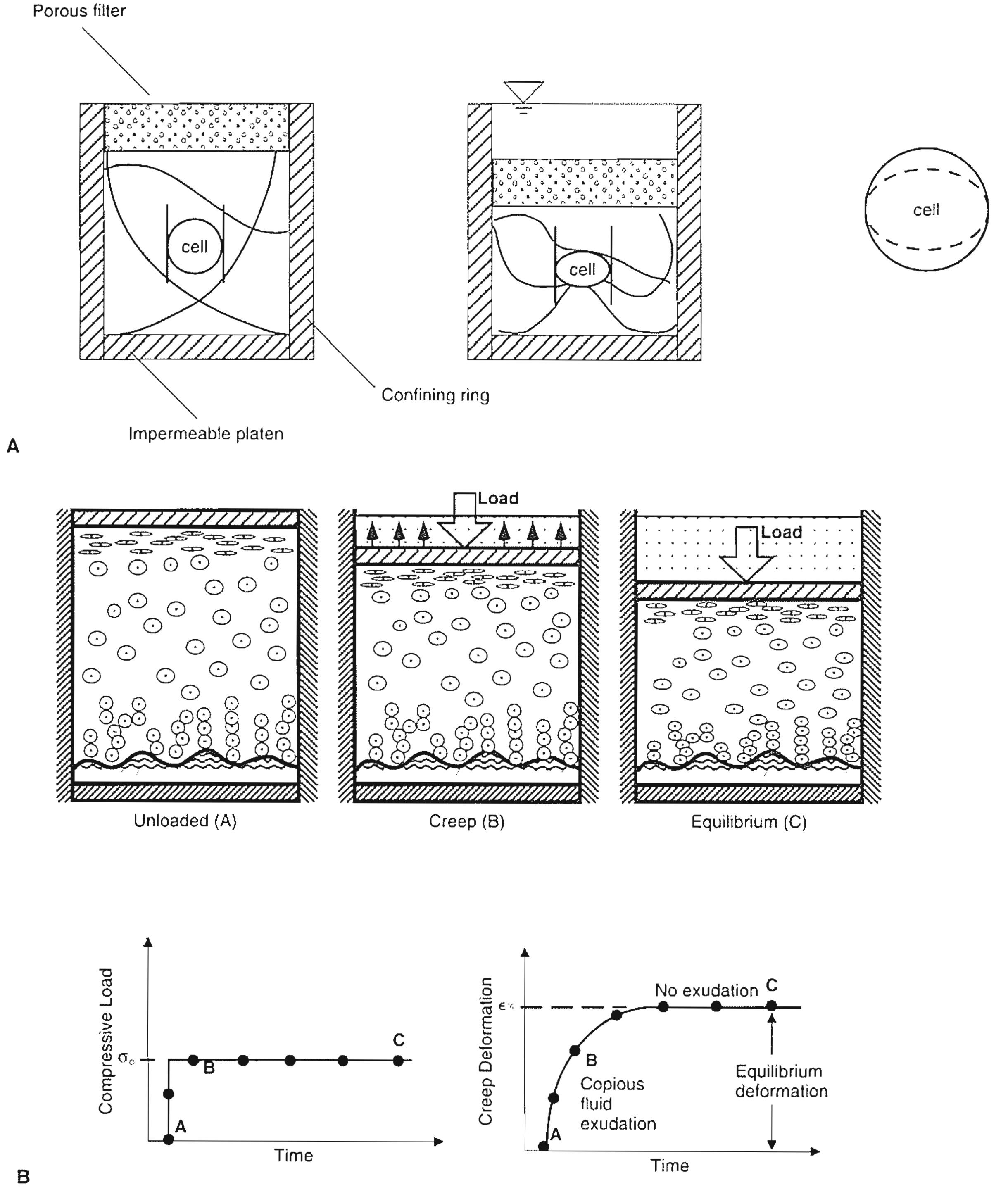


FIG. 3-9

A, A schematic of the confined compression loading configuration. A cylindrical tissue specimen is positioned tightly into an impermeable confining ring that does not permit deformation (or fluid flow) in the radial direction. Under loading, fluid exudation occurs through the porous platen in the vertical direction. **B,** A constant stress σ_o applied to a sample of articular cartilage (*bottom left*) and creep response of the sample under the constant applied stress (*bottom right*). The drawings of a block of tissue above the curves illustrate that creep is accompanied by copious exudation of fluid from the sample and that the rate of exudation decreases over time from points A to B to C. At equilibrium (ϵ_∞), fluid flow ceases and the load is borne entirely by the solid matrix (point C). *Adapted from Mow, V.C., Kuei, S.C., Lai, W.M., et al. (1980). Biphasic creep and stress relaxation of articular cartilage in compression: Theory and experiments.* J Biomech Eng, 102, 73–84.

BIPHASIC STRESS-RELAXATION RESPONSE OF ARTICULAR CARTILAGE IN COMPRESSION

The biphasic viscoelastic stress-relaxation response of articular cartilage in a 1D compression experiment is depicted in Figure 3-10. In this case, a constant compression rate (line t_0-A-B of lower left figure) is applied to the tissue until u_0 is reached; beyond point B, the deformation u_0 is maintained. For articular cartilage, the typical stress response caused by this imposed deformation is shown in the lower right figure (Holmes et al., 1985; Mow et al., 1984). During the compression phase, the stress rises continuously until σ_0 is reached, corresponding to u_0, while during the stress-relaxation phase, the stress continuously decays along the curve B-C-D-E until the equilibrium stress (σ_∞) is reached.

The mechanisms responsible for the stress rise and stress relaxation are depicted in the lower portion of Figure 3-10. As illustrated in the top diagrams, the stress rise in the compression phase is associated with fluid exudation, while stress relaxation is associated with fluid redistribution within the porous solid matrix. During the compressive phase

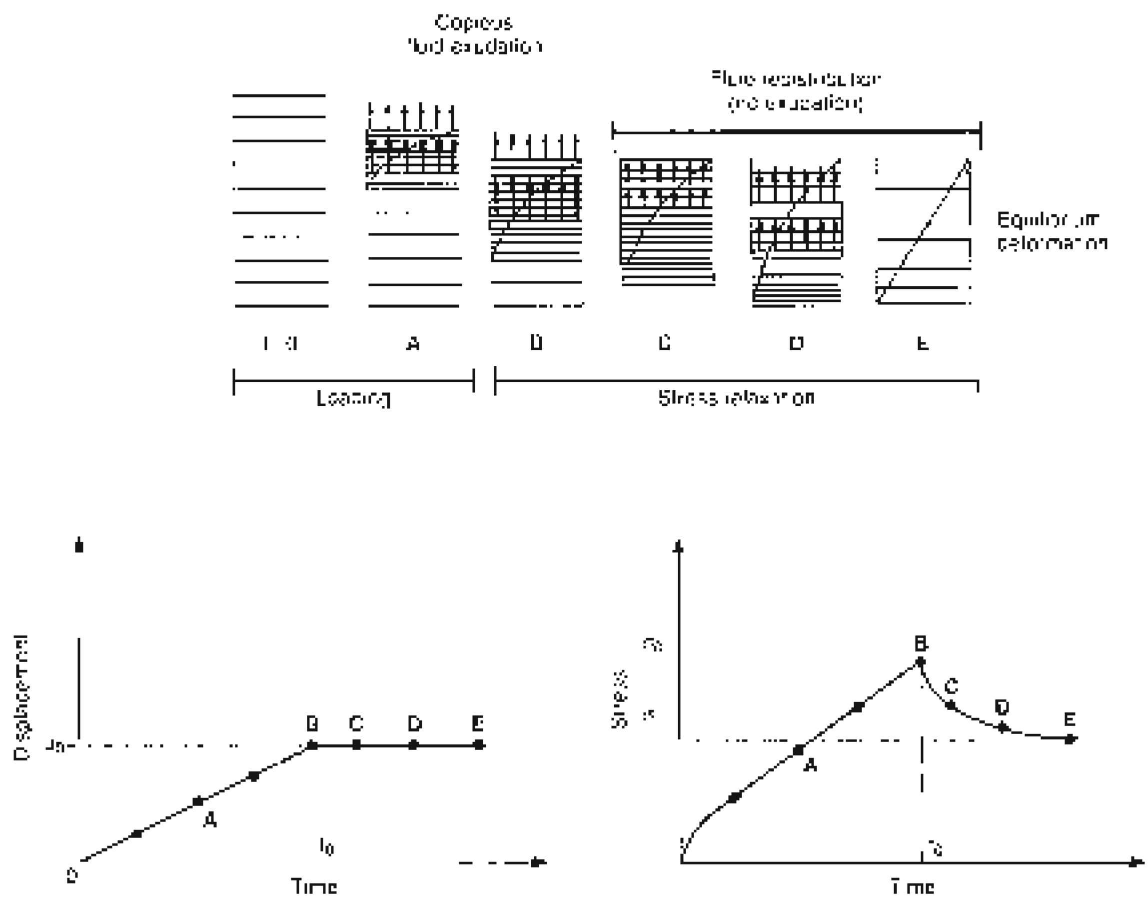


FIG. 3-10

Controlled ramp displacement curve imposed on a cartilage specimen commencing at t_0 (bottom left) and the stress-response curve of the cartilage in this uniaxial confined compression experiment (bottom right). The sample is compressed to point B and maintained over time (points B to E). The history of the stress and response shows a characteristic stress that rises during the compressive phase (points t_0 to B) and then decreases during the relaxation phase (points B to E) until an equilibrium is reached (point E). Above these two curves, schematics illustrate interstitial fluid flow (represented by arrows) and solid matrix deformation during this compressive process. Fluid exudation gives rise to the peak stress (point B), and fluid redistribution gives rise to the stress relaxation phenomena.

the high stress is generated by forced exudation of the interstitial fluid and the compaction of the solid matrix near the surface. Stress relaxation is in turn caused by the relief or rebound of the high compaction region near the surface of the solid matrix. This stress-relaxation process will cease when the compressive stress developed within the solid matrix reaches the stress generated by the intrinsic compressive modulus of the solid matrix corresponding to u_0 (Holmes et al., 1985; Mow et al., 1980, 1984). Analysis of this stress-relaxation process leads to the conclusion that under physiological loading conditions, excessive stress levels are difficult to maintain because stress relaxation will rapidly attenuate the stress developed within the tissue; this must necessarily lead to the rapid spreading of the contact area in the joint during articulation (Ateshian et al., 1995, 1998; Mow & Ateshian, 1997).

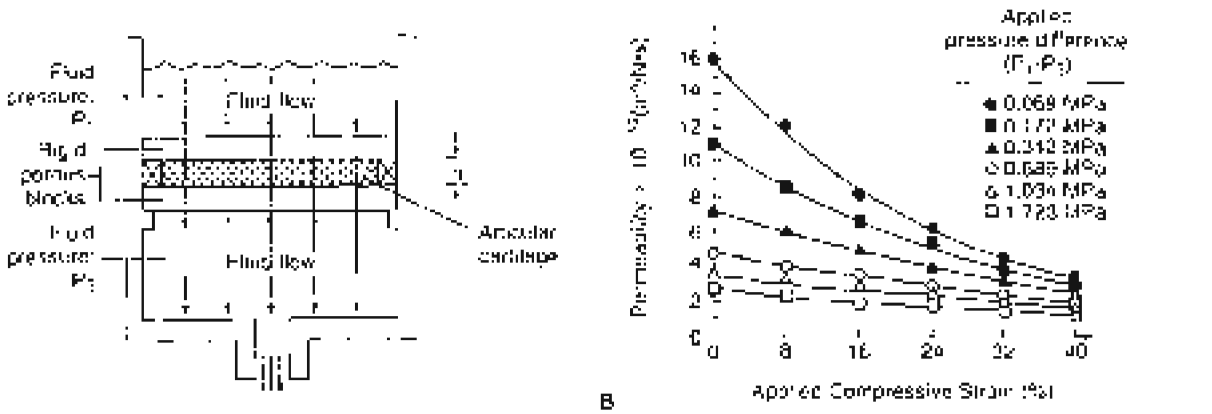
Recently, much focus has been on the inhomogeneity of HA with cartilage depth (Schinagl et al., 1996, 1997). Based on this data, from an analysis of the stress-relaxation experiment it was found that an inhomogeneous tissue would relax at a faster rate than would the uniform tissue (Wang & Mow, 1998). Moreover, the stress, strain, pressure, and fluid flow fields within the tissue were significantly altered as well. Thus it seems that the variations in biochemical and structural composition in the layers of cartilage provide another challenge to understanding the environment of chondrocytes in situ.

PERMEABILITY OF ARTICULAR CARTILAGE

Fluid-filled porous materials may or may not be permeable. The ratio of fluid volume (V^f) to the total volume (V^T) of the porous material is known as the porosity ($\beta = V^f/V^T$); thus, porosity is a geometric concept. Articular cartilage is therefore a material of high porosity (approximately 80%). If the pores are interconnected, the porous material is permeable. Permeability is a measure of the ease with which fluid can flow through a porous material, and it is inversely proportional to the frictional drag exerted by the fluid flowing through the porous-permeable material. Thus, permeability is a physical concept; it is a measure of the resistive force that is required to cause the fluid to flow at a given speed through the porous-permeable material. This frictional resistive force is generated by the interaction of the interstitial fluid and the pore walls of the porous permeable material. The permeability coefficient k is related to the frictional drag coefficient K by the relationship $k = 1/K$ (Lai & Mow, 1980). Articular cartilage has a very low permeability and thus high frictional resistive forces are generated when fluid is caused to flow through the porous solid matrix.

In the previous sections on cartilage viscoelasticity we discussed the process of fluid flow through articular cartilage induced by solid matrix compression and how this process influences the viscoelastic behavior of the tissue. This process also provides an indirect method to determine the permeability of the tissue. In this section, we discuss the experimental method used to directly measure the permeability coefficient. Such an experiment is depicted in Figure 3-11A. Here, a specimen of the tissue is held fixed in a chamber subjected to the action of a pressure gradient; the imposed upstream pressure P_1 is greater than the downstream pressure P_2. The thickness of the specimen is denoted by h and the cross-sectional area of permeation is defined by A. Darcy's law, used to determine the permeability k from this simple experimental setup, yields $k = Qh/A(P_1 - P_2)$, where Q is the volumetric discharge per unit time through the specimen whose area of permeation is A (Mow & Ratcliffe, 1997). Using low pressures, approximately 0.1 MPa, this method was first used to determine the permeability of articular cartilage (Edwards, 1967; Maroudas, 1975). The value of k obtained in this manner ranged from $1.1 \times 10^{-15}\ m^4/N{\cdot}s$ to $7.6 \times 10^{-15}\ m^4/N{\cdot}s$. In addition, using a uniform straight tube model, the average "pore diameter" has been estimated at 6 nm (Maroudas, 1979). Thus, the "pores" within articular cartilage are of molecular size.

The permeability of articular cartilage under compressive strain and at high physiological pressures (3 MPa) was first obtained by Mansour and Mow (1976) and later analyzed by Lai and Mow (1980). The high pressure and compressive strain conditions examined in these studies more closely resemble those conditions found in diarthrodial joint loading. In these experiments, k was measured as a function of two variables: the pressure gradient across the specimen and the axial compressive strain applied to the sample. The results from these experiments are shown in Figure 3-11B. Permeability decreased exponentially as a function of both increasing compressive strain and increasing applied fluid pressure. It was later shown, however, that the dependence of k on the applied fluid pressure derives from compaction of the solid matrix that, in turn, results from the frictional drag caused by the permeating fluid (Lai & Mow, 1980). From the point of view of pore



A B

FIG. 3-11

A. Experimental configuration used in measuring the permeability of articular cartilage, involving the application of a pressure gradient (P_1-P_2)/h across a sample of the tissue (h: tissue thickness). Because the fluid pressure (P_1) above the sample is greater than that beneath it (P_2), fluid will flow through the tissue. The permeability coefficient k in this experiment is given by the expression Qh/A(P_1-P_2), where Q is the volumetric discharge per unit time and A is the area of permeation. *Adapted from Torzilli, P.A., Mow, V.C. (1976). On the fundamental fluid transport mechanisms through normal and pathologic cartilage during function. I. The formulation. J Biomech, 9(8), 541–552.* **B.** Experimental curves for articular cartilage permeability show its strong dependence on compressive strain and applied pressure. Measurements were taken at applied pressure differential (P_1-P_2) and applied strains. The permeability decreased in an exponential manner as a function of both increasing applied compressive strain and increasing applied pressure. *Adapted from Lai, W.M., & Mow, V.C. (1980). Drag-induced compression of articular cartilage during a permeation experiment. J Biorheology, 17, 111.*

structure, compaction of the solid matrix decreases the porosity and hence the average "pore diameter" within the solid matrix; thus, solid matrix compaction increases frictional resistance (Mow et al., 1984).

The nonlinear permeability of articular cartilage demonstrated in Figure 3-11B suggests that the tissue has a mechanical feedback system that may serve important purposes under physiological conditions. When subjected to high loads through the mechanism of increased frictional drag against interstitial fluid flow, the tissue will appear stiffer and it will be more difficult to cause fluid exudation. Recent analyses of articular cartilage compressive stress relaxation behavior have validated this concept and its importance in the capacity of the interstitial fluid to support load (Ateshian et al., 1998; Soltz & Ateshian, 1998). Moreover, this mechanism also is important in joint lubrication.

BEHAVIOR OF ARTICULAR CARTILAGE UNDER UNIAXIAL TENSION

The mechanical behavior of articular cartilage in tension is highly complex. In tension, the tissue is strongly anisotropic (being stiffer and stronger for specimens harvested in the direction parallel to the split line pattern than those harvested perpendicular to the split line pattern) and strongly inhomogeneous (for mature animals, being stiffer and stronger for specimens harvested from the superficial regions than those harvested deeper in the tissue) (Kempson, 1979; Roth & Mow, 1980). Interestingly, articular cartilage from immature bovine knee joints does not exhibit these layered inhomogeneous variations; however, the superficial zones of both mature and immature bovine cartilage appear to have the same tensile stiffness (Roth & Mow, 1980). These anisotropic and inhomogeneous characteristics in mature joints are believed to be caused by the varying collagen and PG structural organization of the joint surface and the layering structural arrangements found within the tissue. Thus, the collagen-rich superficial zone appears to provide the joint cartilage with a tough wear-resistant protective skin (Setton et al., 1993) (Fig. 3-3A).

Articular cartilage also exhibits viscoelastic behavior in tension (Woo et al., 1987). This viscoelastic behavior is attributable to both the internal friction associated with polymeric motion and the flow of the interstitial fluid. To examine the intrinsic mechanical response of the collagen-PG solid matrix in tension, it is necessary to negate the biphasic fluid flow effects. To do this, one must perform slow, low strain-rate experiments (Akizuki et al., 1986; Roth & Mow, 1980; Woo et al., 1987) or perform an incremental strain experiment in which stress relaxation is allowed to progress toward equilibration at each increment of strain (Akizuki et al., 1986). Typically, in a low strain-rate (or near-equilibrium tensile) experiment, a displacement rate of 0.5 cm/minute is used and the specimens usually are pulled to failure. Unfortunately, using these procedures to negate the effect of interstitial fluid flow also negates the manifestation of the intrinsic viscoelastic behavior of the solid matrix. Thus, only the equilibrium intrinsic mechanical properties of the solid matrix may be determined from these tensile tests. The intrinsic viscoelastic properties of the solid matrix must be determined from a pure shear study.

The "equilibrium" stress-strain curve for a specimen of articular cartilage tested under a constant low strain-rate condition is shown in Figure 3-12. Like other fibrous biological tissues (tendons and ligaments), articular cartilage tends to stiffen with increasing strain when the strain becomes large. Thus, over the entire range of strain (up to 60%) in tension, articular cartilage cannot be described by a single Young's modulus. Rather, a tangent modulus, defined by the tangent to the stress-strain curve, must be used to describe the tensile stiffness of the tissue. This fundamental result has given rise to the wide range of Young's modulus, 3 to 100 MPa, reported for articular cartilage in tension (Akizuki et al., 1986; Kempson, 1979; Roth & Mow, 1980; Woo et al., 1987). At physiological strain levels, however, less than 15% (Armstrong et al., 1979) of the linear Young's modulus of articular cartilage ranges between 5 and 10 MPa (Akizuki et al., 1986).

Morphologically, the cause for the shape of the tensile stress-strain curve for large strains is depicted in the diagrams on the right of Figure 3-12. The initial toe region is caused by collagen fiber pull-out and realignment during the initial portion of the tensile experiment, and the final linear region is caused by the stretching of the straightened-aligned collagen fibers. Failure occurs when all the collagen fibers contained within the specimen are ruptured. Figure 3-13A depicts an unstretched articular cartilage specimen, while Figure 3-13B depicts a stretched specimen. Figure 3-14, A & B shows scanning electron micrographs of cartilage blocks under 0 and 30% stretch (right) and the corresponding histograms of collagen fiber orientation

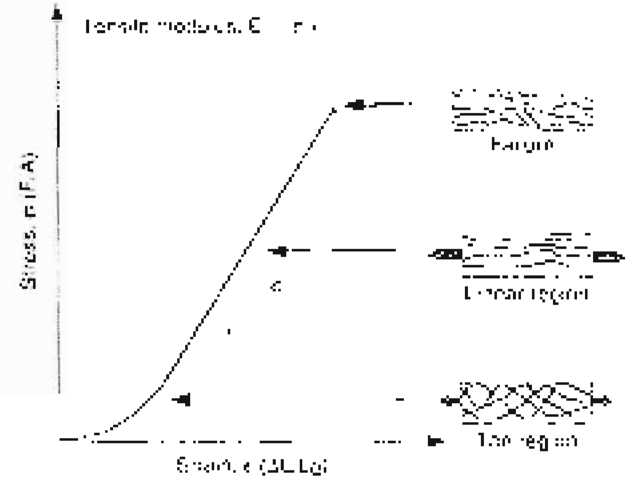


FIG. 3-12

Typical tensile stress-strain curve for articular cartilage. The drawings on the right of the curve show the configuration of the collagen fibrils at various stages of loading. In the toe region, collagen fibril pull-out occurs as the fibrils align themselves in the direction of the tensile load. In the linear region, the aligned collagen fibers are stretched until failure occurs.

determined from the scanning electron micrograph pictures (left). Clearly, it can be seen that the collagen network within cartilage responds to tensile stress and strain (Wada & Akizuki, 1987).

If the molecular structure of collagen, the organization of the collagen fibers within the collagenous network, or the collagen fiber cross-linking is altered (such as that occurring in mild fibrillation or OA), the tensile properties of the network will change. Schmidt et al. (1990) have shown a definitive relationship between collagen hydroxypyridinium cross-linking and tensile stiffness and strength of normal bovine cartilage. Akizuki et al. (1986) showed that progressive degradation of human knee joint cartilage, from mild fibrillation to OA, yields a progressive deterioration of the intrinsic tensile properties of the collagen-PG solid matrix. Similar results have been observed recently in animal models of OA (Guilak et al., 1994; Setton et al., 1994). Together, these observations support the belief that disruption of the collagen network is a key factor in the initial events leading to the development of OA. Also, loosening of the collagen network is generally believed to be responsible for the increased swelling, hence water content, of osteoarthritic cartilage (Maroudas & Venn, 1977; Maroudas, 1979). We have already discussed how increased water content leads

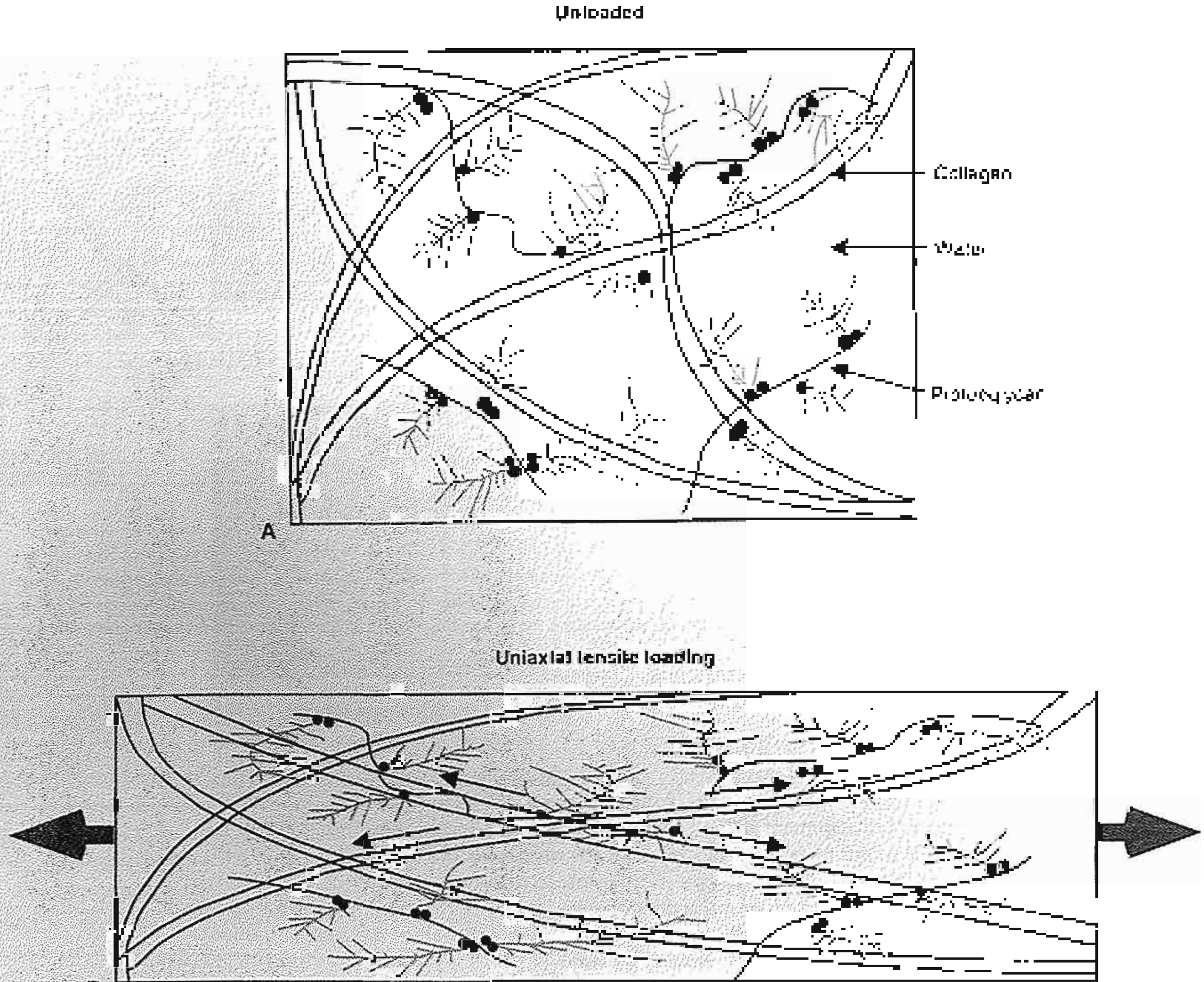


FIG. 3-13

Schematic depiction of the main components of articular cartilage when the tissue is unloaded (A) and when a tensile load is applied (B). Loading would result in an alignment of collagen fibrils along the axis of tension. *Adapted from Myers, E.R., Lai, W.M., & Mow, V.C. (1984). A continuum theory and an experiment for the ion-induced swelling behavior cartilage. J Biomech Eng, 106(2), 151–158.*

to decreased compressive stiffness and increased permeability of articular cartilage.

BEHAVIOR OF ARTICULAR CARTILAGE IN PURE SHEAR

In tension and compression, only the equilibrium intrinsic properties of the collagen-PG solid matrix can be determined. This is because a volumetric change always occurs within a material when it is subjected to uniaxial tension or compression. This volumetric change causes interstitial fluid flow and induces biphasic viscoelastic effects within the tissue. If, however, articular cartilage is tested in pure shear under infinitesimal strain conditions, no pressure gradients or volumetric changes will be produced within the material; hence, no interstitial fluid flow will occur (Hayes & Bodine, 1978; Zhu et al., 1993) (Fig. 3-15). Thus, a steady dynamic pure shear experiment can be used to assess the

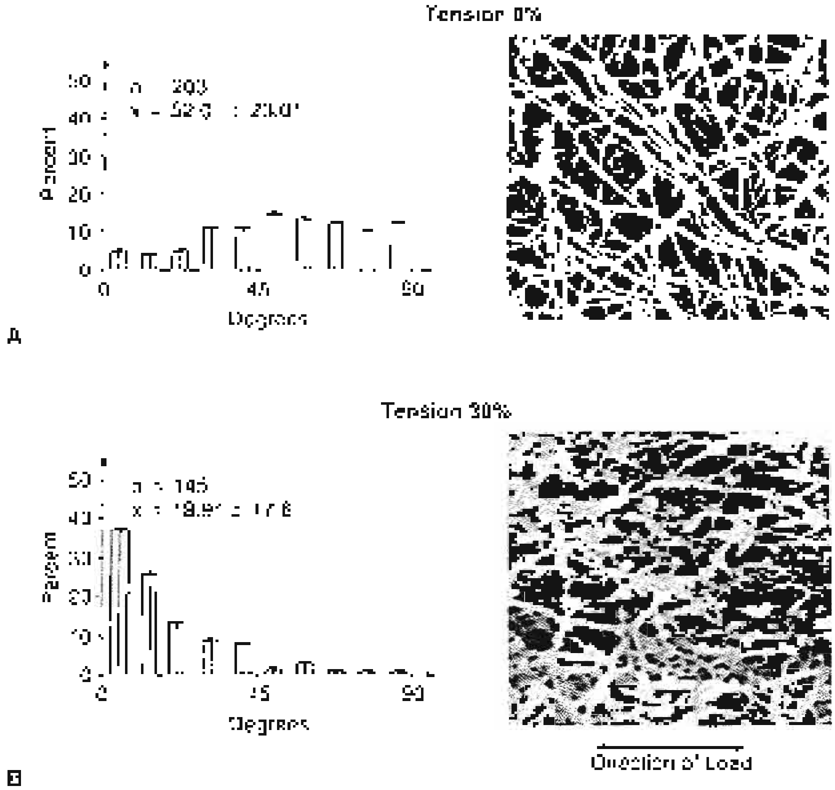


FIG. 3-14

Collagen fibril alignment is clearly demonstrated by the scanning electron micrographs (× 10,000) (*right*) of cartilage blocks under 0% stretch (A) and 30% stretch (B). The histograms (*left*), calculated from the micrographs, represent the percent of collagen fibers oriented in the direction of the applied tension. At 0% stretch the fibers have a random orientation; however, at 30% they are aligned in the direction of the applied tension. *Reprinted with permission from Wada, T., & Akizuki, S. (1987). An ultrastructural study of solid matrix of articular cartilage under uniaxial tensile stress. J Jpn Orthop Assn, 61.*

intrinsic viscoelastic properties of the collagen-PG solid matrix.

In a steady dynamic shear experiment, the viscoelastic properties of the collagen-PG solid matrix are determined by subjecting a thin circular wafer of tissue to a steady sinusoidal torsional shear, shown in Figure 3-16. In an experiment of this type, the tissue specimen is held by a precise amount of compression between two rough porous platens. The lower platen is attached to a sensitive torque transducer and the upper platen is attached to a precision mechanical spectrometer with a servo-controlled dc motor. A sinusoidal excitation signal may be provided by the motor in a frequency of excitation range of 0.01 to 20 hertz (Hz). For shear strain magnitudes ranging from 0.2 to 2.0%, the viscoelastic properties are equivalently defined by the elastic storage modulus G′, the viscous loss modulus G″ of the collagen-PG solid matrix may be determined as a function of frequency (Fung, 1981; Zhu et al., 1993).

Sometimes it is more convenient to determine the magnitude of the dynamic shear modulus $|G^*|$ given by

$$|G^*|^2 = (G')^2 + (G'')^2$$

and the phase shift angle given by

$$\delta = \tan^{-1}(G''/G')$$

The magnitude of the dynamic shear modulus is a measure of the total resistance offered by the viscoelastic material. The value of δ, the angle between

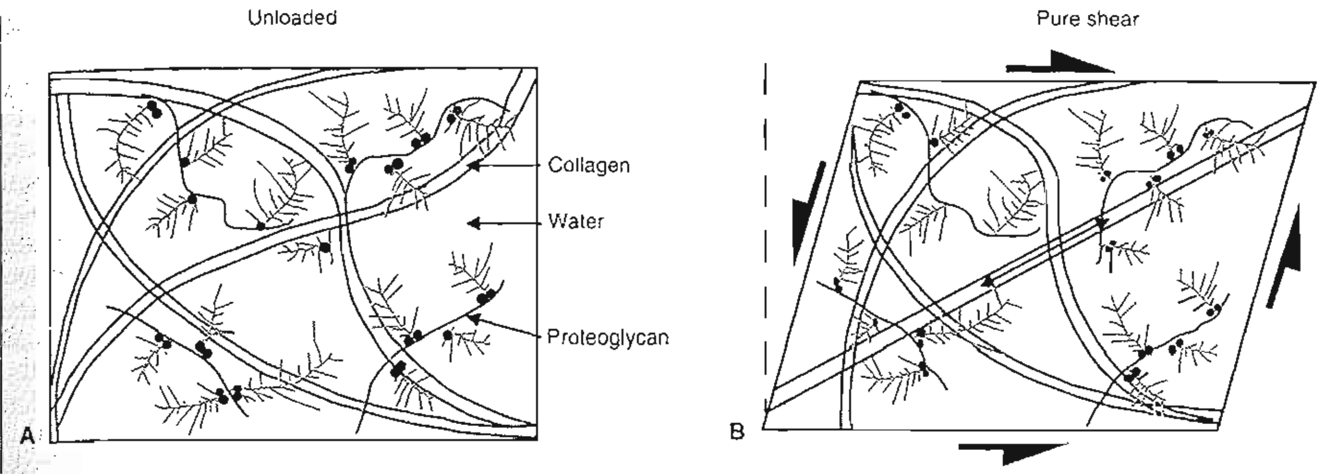


FIG. 3-15

Schematic depiction of unloaded cartilage (A), and cartilage subjected to pure shear (B). When cartilage is tested in pure shear under infinitesimal strain conditions, no volumetric changes or pressure gradients are produced; hence, no interstitial fluid flow occurs. This figure also demonstrates the functional role of collagen fibrils in resisting shear deformation.

the steady applied sinusoidal strain and the steady sinusoidal torque response, is a measure of the total frictional energy dissipation within the material. For a pure elastic material with no internal frictional dissipation, the phase shift angle δ is zero; for a pure viscous fluid, the phase shift angle δ is 90°.

The magnitude of the dynamic shear modulus for normal bovine articular cartilage has been measured to range from 1 to 3 MPa, while the phase shift angle has been measured to range from 9 to 20° (Hayes & Bodine, 1978; Zhu et al., 1993). The intrinsic transient shear stress-relaxation behavior of the collagen-PG solid matrix along with the steady dynamic shear properties also has been measured (Zhu et al., 1986). With both the steady dynamic and the transient results, the latter investigators showed that the quasilinear viscoelasticity theory proposed by Fung (1981) for biological materials provides an accurate description of the flow-independent viscoelastic behavior of the collagen-PG solid matrix. Figure 3-17 depicts a comparison of the theoretical prediction of the transient stress-relaxation phenomenon in shear with the results from Fung's 1981 quasilinear viscoelasticity theory.

From these shear studies, it is possible to obtain some insight as to how the collagen-PG solid matrix functions. First, we note that measurements of PG solutions at concentrations similar to those found in articular cartilage in situ yield a magnitude of shear modulus to be of the order of 10 Pa and phase shift angle ranging up to 70° (Mow et al., 1989b; Zhu et al., 1991, 1996). Therefore, it appears that the magnitude of the shear modulus of concentrated PG so-

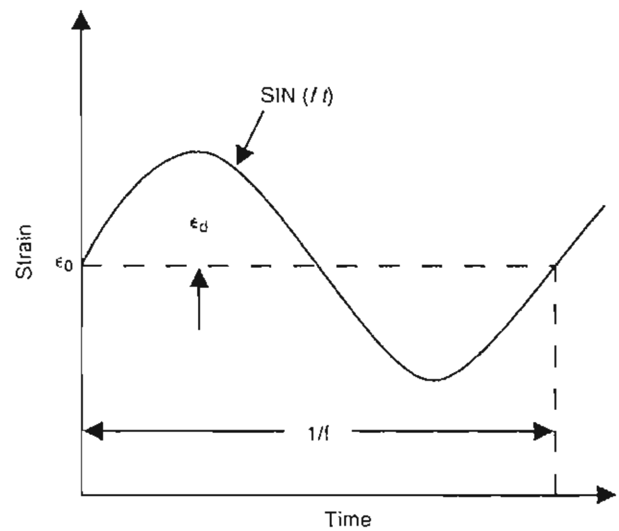


FIG. 3-16

Steady sinusoidal torsional shear imposed on a specimen in pure shear. The fluctuating strain in the form of a sine wave with a strain amplitude ϵ_a and frequency f.

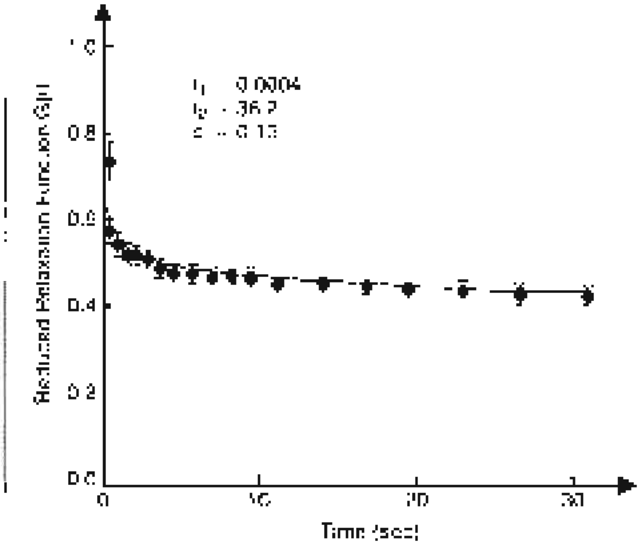


FIG. 3-17

Typical stress-relaxation curve after a step change in shear strain, expressed in terms of the mean of ten cycles of stress relaxation normalized by the initial stress. The solid line represents the theoretical prediction of the quasilinear viscoelasticity theory. *Adapted from Zhu, W.B., Lai, W.M., & Mow, V.C. (1986). Intrinsic quasi-linear viscoelastic behavior of the extracellular matrix of cartilage. Trans Orthop Res Soc, 11, 407.*

function is one hundred thousand times less and the phase angle is six to seven times greater than that of articular cartilage solid matrix. This suggests that PGs do not function in situ to provide shear stiffness for articular cartilage. The shear stiffness of articular cartilage must therefore derive from its collagen content, or from the collagen-PG interaction (Mow & Ratcliffe, 1997). From this interpretation, an increase in collagen, which is a much more elastic element than PG and the predominant load-carrying element of the tissue in shear, would decrease the frictional dissipation and hence the observed phase angle.

SWELLING BEHAVIOR OF ARTICULAR CARTILAGE

The Donnan osmotic swelling pressure, associated with the densely packed fixed anionic groups (SO_3^- and COO^-) on the GAG chains as well as the bulk compressive stiffness of the PG aggregates entangled in the collagen network, permits the PG gel in the collagen network to resist compression (Donnan, 1924; Maroudas, 1979; Mow & Ratcliffe, 1997). To account for such Fixed Charge Density (FCD) effects in cartilage, a triphasic mechano-electrochemical mixture theory was developed that models cartilage as a mixture of three miscible phases: a charged solid phase representing the collagen-PG network, a fluid phase representing the interstitial water, and an ion phase comprising the monovalent cation Na^+ and anion Cl^- as well as other multivalent species such as Ca^{2+} (Gu et al., 1998; Lai et al., 1991). In this theory, the total stress is given by the sum of two terms: $\sigma^{total} = \sigma^s - pI$, where σ^s and p are the solid matrix stress and interstitial fluid pressure, respectively. A significant portion of p is given by the Donnan osmotic pressure, π (see discussion below). Derived from all of the fundamental laws of mechanics and thermodynamics rather than through the ad hoc combination of existing specialized theories (e.g., Frank & Grodzinsky, 1987a,b), this triphasic theory provides a set of thermodynamically permissible constitutive laws to describe the time-dependent physicochemical, mechanical, and electrical properties of charged-hydrated soft tissues. Moreover, the triphasic mixture theory has been shown to be entirely consistent with the specialized classical osmotic pressure theory for charged polymeric solutions, phenomenological transport theories, and the biphasic theory (Donnan, 1924; Katchalsky & Curran, 1975; Mow et al., 1980; Onsager, 1931), all of which have been frequently used to study specific facets of articular cartilage.

The triphasic theory has been used successfully to describe many of the mechano-electrochemical behaviors of articular cartilage. These include the prediction of free swelling under chemical load, nonlinear dependence of hydraulic permeability with FCD, nonlinear dependence of streaming potentials with FCD, curling of cartilage layers, prestress, osmotic and negative osmotic flows, swelling and electrical responses of cells to osmotic shock loading, and the influence of inhomogeneous fixed charge density (Gu et al., 1993, 1997, 1998; Lai et al., 1991; Mow et al., 1998; Setton et al., 1998; Sun et al., 1998). Providing more versatility, the triphasic theory has been generalized to include multielectrolytes in the tissue (Gu et al., 1998).

From analyses using the triphasic theory, it becomes clear that the swelling behavior of the tissue can be responsible for a significant fraction of the compressive load-bearing capacity of articular cartilage at equilibrium (Mow & Ratcliffe, 1997). For example, the triphasic theory predicts for confined-compression at equilibrium that the total stress (σ^{t}) acting on the cartilage specimen is the sum of the stress in the solid matrix (σ^{s}) and the Donnan osmotic pressure ($\sigma^{t} = \sigma^{s} - \pi$). The Donnan osmotic pressure is the swelling pressure caused by the ions in association with the FCD and represents the physicochemical motive force for cartilage swelling (Fig. 3-18). From the classical theory for osmotic pressure, the Donnan osmotic pressure caused by the excess of ion particles inside the tissue is given by:

$$\pi = RT[\phi(2c + c^{F}) - 2\phi^{*}c^{*}] + P_{\infty}$$

where c is the interstitial ion concentration, c^{*} is the external ion concentration, c^{F} is the FCD, R is the universal gas constant, T is the absolute temperature, ϕ and ϕ^{*} are osmotic coefficients, and P_{∞} is the osmotic pressure caused by the concentration of PG particles in the tissue, usually assumed to be negligible (Lai et al., 1991). For a lightly loaded tissue, the swelling pressure may contribute significantly to the load support. But for highly loaded tissues, such as those found under physiological conditions and certainly for dynamically loaded tissues, the interstitial fluid pressurization (σ^{f}) would dominate; the contribution of this swelling pressure to load support would be less than 5% (Soltz & Ateshian, 1998).

As with the biphasic theory, the triphasic mechano-electrochemical theory can be used to elucidate potential mechano-signal transduction mechanisms in cartilage. For example, because of their potential effects on chondrocyte function, it is important to describe and predict electrokinetic phenomena such as streaming potentials and streaming currents (Gu et al., 1993, 1998; Katchalsky & Curran, 1975; Kim et al., 1994) that arise from ion movement caused by the convection of interstitial fluid flow past the FCD of the solid matrix. As a second example, the pressure produced in the interstitial fluid by polyethylene glycol-induced osmotic loading of cartilage explants (Schneiderman et al., 1986) was recently shown to be theoretically nonequivalent to the pressure produced in any other commonly used mechanically loaded explant experiment or by hydrostatic loading (Lai et al., 1998). In light of this finding, earlier interpretations of biological data from studies making such an assumption of equivalence should be revisited.

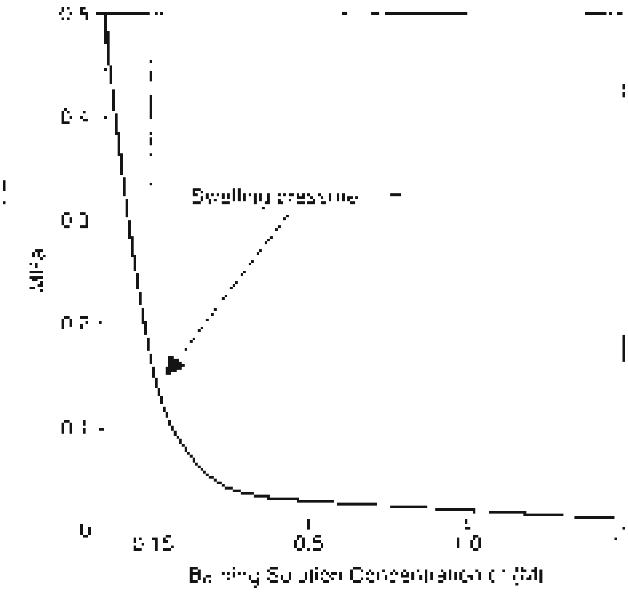


FIG. 3-18

Swelling pressure of articular cartilage versus bathing solution concentration (c*). At equilibrium, the interstitial fluid pressure is equal to the swelling pressure, which is defined by the tissue Donnan osmotic pressure (π).

Lubrication of Articular Cartilage

As already discussed, synovial joints are subjected to an enormous range of loading conditions, and under normal circumstances the cartilage surface sustains little wear. The minimal wear of normal cartilage associated with such varied loads indicates that sophisticated lubrication processes are at work within the joint and within and on the surface of the tissue. These processes have been attributed to a lubricating fluid film forming between the articular cartilage surface and to an adsorbed boundary lubricant on the surface during motion and loading. The variety of joint demands also suggests that a number of mechanisms are responsible for diarthrodial joint lubrication. To understand diarthrodial joint lubrication, one should use basic engineering lubrication concepts.

From an engineering perspective, there are two fundamental types of lubrication. One is boundary lubrication, which involves a single monolayer of lubricant molecules adsorbed on each bearing surface.

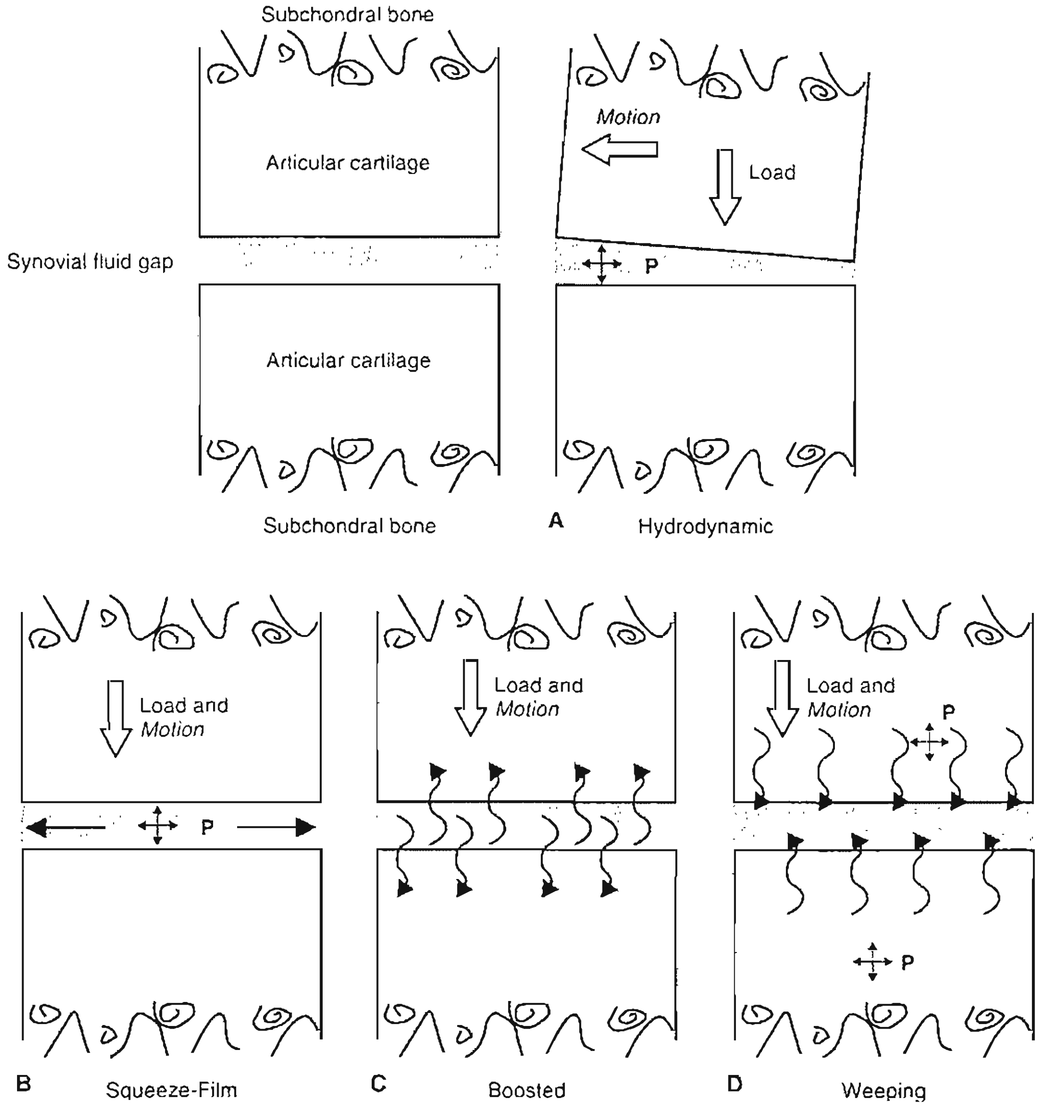


FIG. 3-19

A, In hydrodynamic lubrication, viscous fluid is dragged into a convergent channel, causing a pressure field to be generated in the lubricant. Fluid viscosity, gap geometry, and relative sliding speed determine the load-bearing capacity. B, As the bearing surfaces are squeezed together, the viscous fluid is forced from the gap into the transverse direction. This squeeze action generates a hydrodynamic pressure in the fluid for load support. The load-bearing capacity depends on the size of the surfaces, velocity of approach, and fluid viscosity. C, The direction of fluid flow under squeeze-film lubrication in the boosted mode for joint lubrication. D, Depicts the Weeping lubrication hypothesis for the uniform exudation of interstitial fluid from the cartilage. The driving mechanism is a self-pressurization of the interstitial fluid when the tissue is compressed.

The other is fluid-film lubrication, in which a thin fluid-film provides greater surface-to-surface separation (Bowden & Tabor, 1967). Both lubrication types appear to occur in articular cartilage under varying circumstances. Intact synovial joints have an extremely low coefficient of friction, approximately 0.02 (Dowson, 1966/1967; Linn, 1968; McCutchen, 1962; Mow & Ateshian, 1997). Boundary-lubricated surfaces typically have a coefficient of friction one or two orders of magnitude higher than surfaces lubricated by a fluid-film, suggesting that synovial joints are lubricated, at least in part, by the fluid-film mechanism. It is quite possible that synovial joints use the mechanism that will most effectively provide lubrication at a given loading condition. Unresolved, though, is the manner by which synovial joints generate the fluid lubricant film.

FLUID-FILM LUBRICATION

Fluid-film lubrication utilizes a thin film of lubricant that causes a bearing surface separation. The load on the bearing is then supported by the pressure that is developed in this fluid-film. The fluid-film thickness associated with engineering bearings is usually less than 20 μm. Fluid-film lubrication requires a minimum fluid-film thickness (as predicted by a specific lubrication theory) to exceed three times the combined statistical surface roughness of cartilage (e.g., 4 to 25 μm; Clarke, 1971; Walker et al., 1970). If fluid-film lubrication is unachievable because of heavy and prolonged loading, incongruent gap geometry, slow reciprocating-grinding motion, or low synovial fluid viscosity, boundary lubrication must exist (Mow & Ateshian, 1997).

The two classical modes of fluid-film lubrication defined in engineering are hydrodynamic and squeeze-film lubrication (Fig. 3-19, *A & B*). These modes apply to rigid bearings composed of relatively undeformable material such as stainless steel. Hydrodynamic lubrication occurs when nonparallel rigid bearing surfaces lubricated by a fluid-film move tangentially with respect to each other (i.e., slide on each other), forming a converging wedge of fluid. A lifting pressure is generated in this wedge by the fluid viscosity as the bearing motion drags the fluid into the gap between the surfaces, as shown in Figure 3-19A. In contrast, squeeze-film lubrication occurs when the bearing surfaces move perpendicularly toward each other. A pressure is generated in the fluid-film as a result of the viscous resistance of the fluid that acts to impede its escape from the gap (Fig. 3-19B). The squeeze-film mechanism is sufficient to carry high loads for short durations. Eventually, however, the fluid-film becomes so thin that contact between the asperities (peaks) on the two bearing surfaces occurs.

Calculations of the relative thickness of the fluid-film layer and the surface roughness are valuable in establishing when hydrodynamic lubrication may exist. In hydrodynamic and squeeze-film lubrication, the thickness and extent of the fluid-film, as well as its load-carrying capacity, are characteristics independent of the (rigid) bearing surface material properties. These lubrication characteristics are instead determined by the lubricant's properties, such as its rheological properties, viscosity and elasticity, the film geometry, the shape of the gap between the two bearing surfaces, and the speed of the relative surface motion.

Cartilage is unlike any man-made material with respect to its near frictionless properties. Classical theories developed to explain lubrication of rigid and impermeable bearings (e.g., steel) cannot fully explain the mechanisms responsible for lubrication of the natural diarthrodial joint. A variation of the hydrodynamic and squeeze-film modes of fluid-film lubrication, for example, occurs when the bearing material is not rigid but instead relatively soft, such as with the articular cartilage covering the joint surface. This type of lubrication, termed elastohydrodynamic, operates when the relatively soft bearing surfaces undergo either a sliding (hydrodynamic) or squeeze-film action and the pressure generated in the fluid-film substantially deforms the surfaces (Fig. 3-19, *A & B*). These deformations tend to increase the surface area and congruency, thus beneficially altering film geometry. By increasing the bearing contact area, the lubricant is less able to escape from between the bearing surfaces, a longer-lasting lubricant film is generated, and the stress of articulation is lower and more sustainable. Elastohydrodynamic lubrication enables bearings to greatly increase their load-carrying capacity (Dowson, 1966/1967, 1990).

Note that several studies have shown that hyaluronidase treatment of synovial fluid, which decreases its viscosity (to that of saline) by causing depolymerization of HA, has little effect on lubrication (Linn, 1968; Linn & Radin, 1968). Because fluid-film lubrication is highly dependent on lubricant viscosity, these results strongly suggest that an alternative mode of lubrication is the primary mechanism responsible for the low frictional coefficient of joints.

BOUNDARY LUBRICATION

During diarthrodial joint function, relative motion of the articulating surfaces occurs. In boundary lubrication, the surfaces are protected by an adsorbed layer of boundary lubricant, which prevents direct, surface-to-surface contact and eliminates most of the surface wear. Boundary lubrication is essentially independent of the physical properties of either the lubricant (e.g., its viscosity) or the bearing material (e.g., its stiffness), instead depending almost entirely on the chemical properties of the lubricant (Dowson, 1966/67). In synovial joints, a specific glycoprotein, "lubricin," appears to be the synovial fluid constituent responsible for boundary lubrication (Swann et al., 1979, 1985). Lubricin (25 × 10⁴ mw) is adsorbed as a macromolecular monolayer to each articulating surface (Fig. 3-20). These two layers, ranging in combined thickness from 1 to 100 nm, are able to carry loads and appear to be effective in reducing friction (Swann et al., 1979). More recently, Hills (1989) suggested that the boundary lubricant found in synovial fluid was more likely to be a phospholipid named dipalmitoyl phosphatidylcholine. Although experiments demonstrate that a boundary lubricant can account for a reduction of the friction coefficient by a factor of threefold to sixfold (Swann et al., 1985; Williams et al., 1993), this reduction is quite modest compared with the much greater range (e.g., up to 60-fold) reported earlier (McCutchen, 1962). Even so, these results do suggest that boundary lubrication exists as a complementary mode of lubrication.

MIXED LUBRICATION

There are two joint lubrication scenarios that can be considered a combination of fluid-film and boundary lubrication or simply mixed lubrication (Dowson, 1966). The first case refers to the temporal coexistence of fluid-film and boundary lubrication at spatially distinct locations, whereas the second case, termed "boosted lubrication," is characterized by a shift of fluid film to boundary lubrication with time over the same location (Walker et al., 1970).

The articular cartilage surface, like all surfaces, is not perfectly smooth; asperities project out from the surface (Clarke, 1971; Gardner & McGillivray, 1971; Redler & Zimny, 1970) (Figs. 3-18 and 3-21). In synovial joints, situations may occur in which the fluid-film thickness is of the same order as the mean articular surface asperity (Walker et al., 1970). During such instances, boundary lubrication between the asperities may come into play. If this occurs, a mixed mode of lubrication is operating, with the joint surface load sustained by both the fluid-film pressure in areas of noncontact and by the boundary lubricant lubricin in the areas of asperity contact (shown in Figure 3-22). In this mode of mixed lubrication, it is probable that most of the friction (which is still extremely low) is generated in the

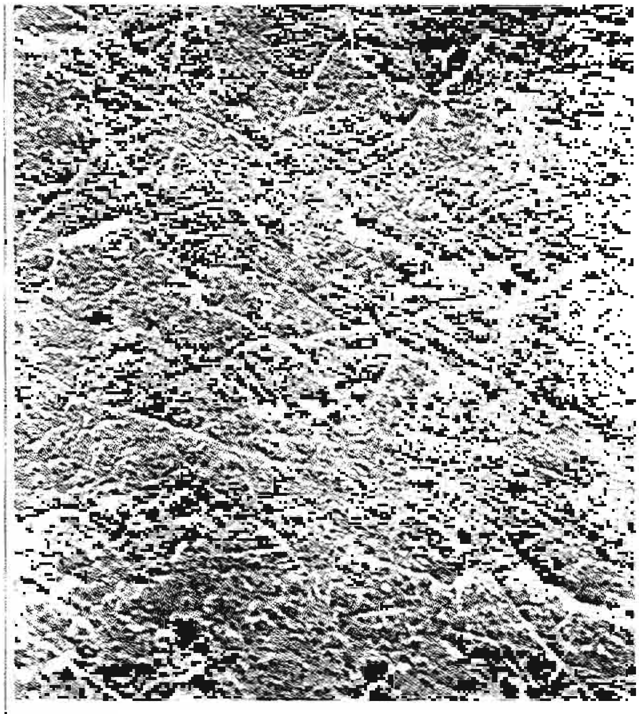


FIG. 3-21

Scanning electron micrograph of the surface of human articular cartilage from a normal young adult showing the typical irregularities characteristic of this tissue (×3,000). Adapted from Armstrong, C.G., & Mow, V.C. (1980). Friction, lubrication and wear of synovial joints. In R. Owen, J. Goodfellow, and P. Bullough (Eds.), Scientific Foundations of Orthopaedics and Traumatology (pp. 223–232). London: William Heinemann.

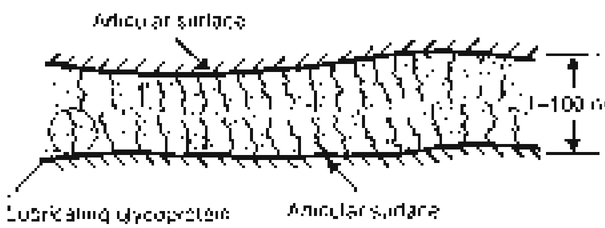


FIG. 3-20

Boundary lubrication of articular cartilage. The load is carried by a monolayer of the lubricating glycoprotein (LGP), which is adsorbed onto the articular surfaces. The monolayer effectively serves to reduce friction and helps to prevent cartilaginous wear. Adapted from Armstrong, C.G., & Mow, V.C. (1980). Friction, lubrication and wear of synovial joints. In R. Owen, J. Goodfellow, and P. Bullough (Eds.), Scientific Foundations of Orthopaedics and Traumatology (pp. 223–232). London: William Heinemann.

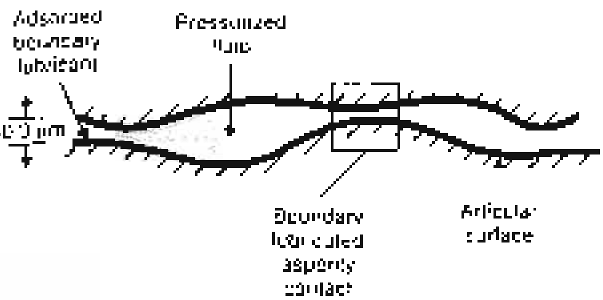


FIG. 3-22

Schematic depiction of mixed lubrication operating in articular cartilage. Boundary lubrication occurs when the thickness of the fluid film is on the same order as the roughness of the bearing surfaces. Fluid film lubrication takes place in areas with more widely separated surfaces. *Adapted from Armstrong, C.G., & Mow, V.C. (1980). Friction, lubrication and wear of synovial joints. In R. Owen, J. Goodfellow, and P. Bullough (Eds.). Scientific Foundations of Orthopaedics and Traumatology (pp. 223–232). London: William Heinemann.*

boundary lubricated areas while most of the load is carried by the fluid film (Dowson, 1966/1967, 1990).

The second mode of mixed lubrication (boosted lubrication) proposed by Walker et al. (1968, 1970) and Maroudas (1966/1967) is based on the movement of fluid from the gap between the approaching articular surfaces into the articular cartilage (Fig. 3-19C). Specifically, in boosted lubrication, articular surfaces are believed to be protected during joint loading by the ultrafiltration of the synovial fluid through the collagen-PG matrix. This ultrafiltration permits the solvent component of the synovial fluid (water and small electrolytes) to pass into the articular cartilage during squeeze-film action, yielding a concentrated gel of HA protein complex that coats and lubricates the bearing surfaces (Lai & Mow, 1978). According to this theory, it becomes progressively more difficult, as the two articular surfaces approach each other, for the HA macromolecules in the synovial fluid to escape from the gap between the bearing surfaces because they are physically too large (0.22–0.65 μm), as shown in Figure 3-23. The water and small solute molecules can still escape into the articular cartilage through the cartilage surface and/or laterally into the joint space at the periphery of the joint. Theoretical results by Hou et al. (1992) predict that fluid entry into the cartilage-bearing surface is possible, leading them to suggest that boosted lubrication may occur. The role of this HA gel in joint lubrication remains unclear, however, particularly in view of the findings by Linn (1968), which demonstrated that purified HA acts as a poor lubricant.

To summarize, in any bearing, the effective mode of lubrication depends on the applied loads and on the relative velocity (speed and direction of motion) of the bearing surfaces. Adsorption of the synovial fluid glycoprotein, lubricin, to articular surfaces seems to be most important under severe loading conditions, that is, contact surfaces with high loads, low relative speeds, and long duration. Under these conditions, as the surfaces are pressed together, the boundary lubricant monolayers interact to prevent direct contact between the articular surfaces. Conversely, fluid-film lubrication operates under less severe conditions, when loads are low and/or oscillate in magnitude and when the contacting surfaces are moving at high relative speeds. In light of the varied demands on diarthrodial joints during normal function, it is unlikely that only a single mode of lubrication exists. As yet, it is impossible to state definitely under which conditions a particular lubrication mechanism may operate. Nevertheless, using the human hip as an example, some general statements are possible.

1. Elastohydrodynamic fluid-films of both the sliding (hydrodynamic) and the squeeze type probably play an important role in lubricating the joint. During the swing phase of walking, when loads on the joint are minimal, a substantial layer of synovial fluid-film is probably maintained. After the first peak force, at heel strike, a supply of fluid lubricant is generated by articular cartilage. However, this fluid-film thickness will begin to decrease under the high load of stance phase; as a result squeeze-film action occurs. The second peak force during the walking cycle, just before the toe leaves the ground, occurs when the joint is swinging in the opposite direction. Thus, it is possible that a fresh supply of fluid film could be generated at toe-off, thereby providing the lubricant during the next swing phase.
2. With high loads and low speeds of relative motion, such as during standing, the fluid film will decrease in thickness as the fluid is squeezed out from between the surfaces (fluid film). Under these conditions, the fluid exuded from the compressed articular cartilage could become the main contributor to the lubricating film.
3. Under extreme loading conditions, such as during an extended period of standing follow

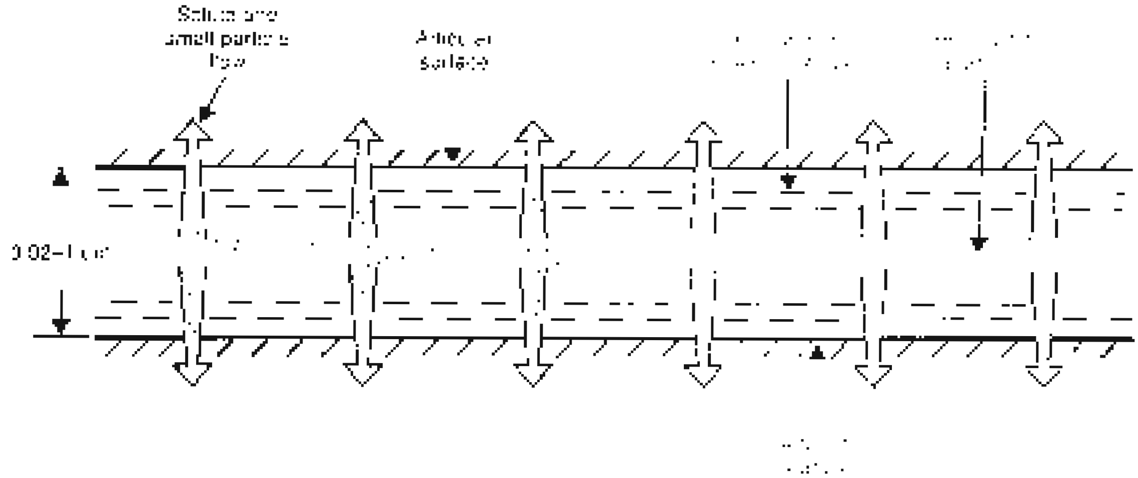


FIG. 3-23. Ultrafiltration of the synovial fluid into a highly viscous gel. As the articular surfaces come together, the small solute molecules escape into the articular cartilage and into the lateral joint space, leaving the large HA macromolecules that, be [illegible]

ing impact. The fluid-film may be eliminated, allowing surface-to-surface contact. The surfaces, however, will probably still be protected either by a thin layer of ultrafiltrated synovial fluid gel (boosted lubrication) or by the adsorbed lubricin monolayer (boundary lubrication).

ROLE OF INTERSTITIAL FLUID PRESSURIZATION IN JOINT LUBRICATION

During joint articulation, loads transmitted across a joint may be supported by the opposing joint surfaces via solid-to-solid contact, through a fluid-film layer, or by a mixture of both. Although fluid-film lubrication is achievable, its contribution to joint lubrication is transient, a consequence of the rapid dissipation of the fluid-film thickness by joint loads. With this caveat, Ateshian (1997), adopting the theoretical framework of the biphasic theory (Mow et al., 1980), proposed a mathematical formulation of a boundary friction model of articular cartilage to describe the underlying mechanism behind fluid-solid joint lubrication, in particular, the time-dependence of the friction coefficient of cartilage reported during creep and stress-relaxation experiments (Malcolm, 1976; McCutchen, 1962).

[illegible] solid and fluid phases of a biphasic material (Mow et al., 1980), Ateshian (1997) derived an expression for the effective friction coefficient [illegible] as pressure in the fluid. [illegible] applied loading [illegible] cartilage [illegible] and shown in Figures 3-24 and 3-25.

To validate this model, Ateshian and [illegible] developed a novel loading [illegible] superimposed [illegible] loads [illegible] cartilage [illegible] confined compression [illegible] (Fig. 3-24) (Ateshian et al., 1998). [illegible] cylindrical biphasic cartilage plug was compressed [illegible] radial [illegible] plug [illegible] surface of the plug opposite the platen was pressed against a [illegible] surface of the [illegible] compression. In this

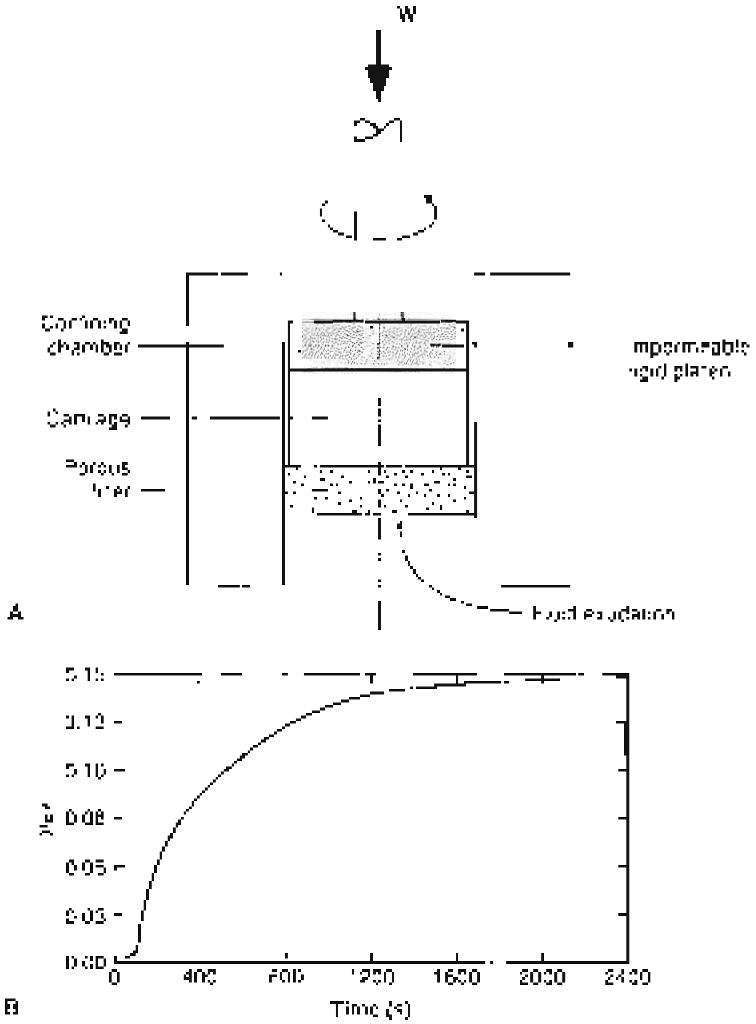


FIG. 3-24.

Experimental configuration superimposing a frictional torque with creep loading of an articular cartilage explant in confined compression (Ateshian et al., 1998). **A.** Note that fluid exudation occurs on the opposite face of the tissue exposed to the frictional load, indicating that the frictional properties of cartilage are not dependent on the weeping of interstitial fluid to the lubricating boundary. **B.** Note that effective friction coefficient (μ_{eff}) varies with increasing proportion of load on the solid matrix, as can be seen from the theoretical curve for μ_{eff} as a function of time during the experiment. *(Adapted from Mow, V.C., & Ateshian, G.A. [1997]. Lubrication and wear of diarthrodial joints. In V.C. Mow & W.C. Hayes [Eds.], Basic Biomechanics [2nd ed., pp. 275–315]. Philadelphia: Lippincott-Raven.)*

manner, a frictional torque was developed in the tissue. Because the application of a torque load that yields pure shear, under infinitesimal deformations, induces no volume change in the tissue or associated fluid exudation, the load generated by the frictional torque is independent of the biphasic creep behavior of the tissue. Theoretical predictions, which closely match experimental results, show that during initial loading, when interstitial pressurization is high, the friction coefficient can be very low (Fig. 3-24B). As creep equilibrium is reached and the load is transferred to the solid matrix, the friction coefficient becomes high (e.g., 0.15). The time constant for this transient response is in excellent agreement with observed experimental results (Malcolm, 1976; McCutchen, 1962). Another important result of this work is that fluid pressurization can function in joint lubrication without concomitant fluid exudation to the lubricating boundary as is proposed for weeping lubrication (McCutchen, 1962) (Fig. 3-19D). Equally significant, this lubrication theory is capable of explaining the observed decrease of the effective friction coefficient with increasing rolling and sliding joint velocities and with increasing joint load (Linn, 1968).

Recently, the interstitial fluid pressurization within cartilage during uniaxial creep and stress relaxation experiments was successfully measured (Soltz & Ateshian, 1998). As predicted by the biphasic theory, they found that interstitial fluid pressurization supported more than 90% of the load for several hundred seconds following loading in confined compression (Ateshian & Wang, 1995). The close agreement of their measurements with biphasic theoretical predictions represents a major advancement in the understanding of diarthrodial joint lubrication and provides compelling evidence for the role of interstitial fluid pressurization as a fundamental mechanism underlying the load bearing capacity of cartilage. It is emphasized that while the collagen-PG matrix is subjected to hydrostatic pressure in the surrounding interstitial fluid, it does not expose the solid matrix (nor encased chondrocytes) to deformation, presumably causing no mechanical damage.

Wear of Articular Cartilage

Wear is the unwanted removal of material from solid surfaces by mechanical action. There are two components of wear: interfacial wear resulting from the interaction of bearing surfaces and fatigue wear resulting from bearing deformation under load.

Interfacial wear occurs when bearing surfaces come into direct contact with no lubricant film (boundary or fluid) separating them. This type of wear can take place in either of two ways: adhesion or abrasion. Adhesive wear arises when, as the bearings come into contact, surface fragments adhere to each other and are torn off from the surface during sliding. Abrasive wear, conversely, occurs when a soft material is scraped by a harder one; the harder material can be either an opposing bearing or loose particles between the bearings. The low rates of interfacial wear observed in articular cartilage tested in vitro (Lipshitz & Glimcher, 1979) suggest that direct surface-to-surface contact between the two cartilage surfaces rarely occurs. Abrasive wear in these experiments, however, was not ruled out. The multiple modes of effective lubrication make interfacial wear unlikely in a normal healthy joint. Nevertheless, adhesive and abrasive wear may take place in an impaired or degenerated synovial joint. Once there is ultrastructural damage and/or decreases in mass, cartilage becomes softer and more permeable (Setton et al., 1993; Armstrong & Mow, 1982; Setton et al., 1994); thus fluid from the lubricant film separating the bearing surfaces may leak away more easily through the cartilage surface. This loss of lubricating fluid from between the surfaces increases the probability of direct contact of the asperities and exacerbates the abrasion process.

Fatigue wear of bearing surfaces results not from surface-to-surface contact, but from the accumulation of microscopic damage within the bearing material under repetitive stressing. Bearing surface failure may occur with the repeated application of high loads over a relatively short period or with the repetition of low loads over an extended period, even though the magnitude of those loads may be much lower than the material's ultimate strength. This fatigue wear resulting from cyclically repeated deformation of the bearing materials can take place even in well-lubricated bearings.

In synovial joints, the cyclical variation in total joint load during most physiological activities causes repetitive articular cartilage stressing (deformation). In addition, during rotation and sliding, a specific region of the articular surface "moves in and out of" the loaded contact area, repetitively stressing that articular region. Loads imposed on articular cartilage are supported by the collagen-PG matrix and by the resistance generated by fluid movement through the matrix. Thus, repetitive joint movement and loading will cause repetitive stressing of the solid matrix and repeated exudation and imbibition of the tissue's interstitial fluid (Mow & Ateshian, 1997). These processes give rise to two possible mechanisms by which fatigue damage may accumulate in articular cartilage: disruption of the collagen-PG solid matrix and PG "washout."

First, repetitive collagen-PG matrix stressing could disrupt the collagen fibers, the PG macromolecules, and/or the interface between these two components. A popular hypothesis is that cartilage fatigue is the result of tensile failure of the collagen fiber network (Freeman, 1975). Also, as discussed above; pronounced changes in the articular cartilage PG population have been observed with age and disease (Buckwalter et al., 1985; Muir, 1983; Roughley et al., 1980; Sweet et al., 1979). These PG changes could be considered as part of the accumulated tissue damage. These molecular structural changes would result in lower PG-PG interaction sites and thus lower network strength (Mow et al., 1989b; Zhu et al., 1991, 1996). Second, repetitive and massive exudation and imbibition of the interstitial fluid may cause the degraded PGs to "wash out" from the ECM, with a resultant decrease in stiffness and increase in permeability of the tissue that in turn defeats the stress-shielding mechanism of interstitial fluid load support and establishes a vicious cycle of cartilage degeneration.

A third mechanism of damage and resultant articular wear is associated with synovial joint impact loading—that is, the rapid application of a high load. With normal physiological loading, articular cartilage undergoes surface compaction during the compression with the lubricating fluid being exuded through this compacted region, as shown in Figure 3-10. As described above, however, fluid redistribution within the articular cartilage occurs over time, which relieves the stress in this compacted region. This process of stress relaxation takes place quickly; the stress may decrease by 63% within 2 to 5 seconds (Ateshian et al., 1998; Mow et al., 1980). If, however, loads are supplied so quickly that there is insufficient time for internal fluid redistribution to relieve the compacted region, the high stresses produced in the collagen-PG matrix may induce damage (Newberry et al., 1997; Thompson et al., 1991). This phenomenon could well explain why Radin and Paul (1971) found dramatic articular cartilage damage with repeated impact loads.

These mechanisms of wear and damage may be the cause of the commonly observed large range of structural defects observed in articular cartilage (Bullough & Goodfellow, 1968; Meachim & Fergie, 1975) (Fig. 3-25, A-C). One such defect is the splitting of the cartilage surface. Vertical sections of cartilage exhibiting these lesions, known as fibrillation, show that they eventually extend through the full depth of the articular cartilage. In other specimens, the cartilage layer appears to be eroded rather than split. This erosion is known as smooth-surfaced destructive thinning.

Considering the variety of defects noted in articular cartilage, it is unlikely that a single wear mechanism is responsible for all of them. At any given site, the stress history may be such that fatigue is the initiating failure mechanism. At another, the lubrication conditions may be so unfavorable that interfacial wear dominates the progression of cartilage failure. As yet, there is little experimental information on the type of defect produced by any given wear mechanism.

Once the collagen-PG matrix of cartilage is disrupted, damage resulting from any of the three wear mechanisms mentioned becomes possible: (1) further disruption of the collagen-PG matrix as a result of repetitive matrix stressing, (2) an increased "washing out" of the PGs as a result of violent fluid movement and thus impairment of articular cartilage's interstitial fluid load support capacity; and (3) gross alteration of the normal load carriage mechanism in cartilage, thus increasing frictional shear loading on the articular surface.

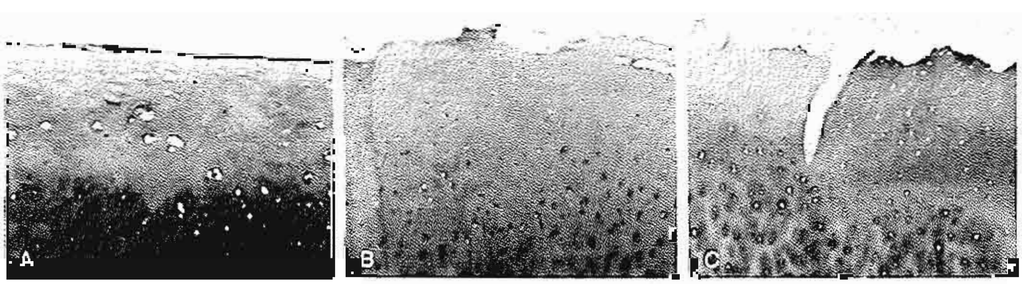


FIG. 3-25

Photomicrographs of vertical sections through the surface of articular cartilage showing a normal intact surface (A), an eroded articular surface (B), and a vertical split or fibrillation of the articular surface that will eventually extend through the full depth of the cartilage (C). *Photomicrographs provided through the courtesy of Dr. S. Akizuki, Nagano, Japan.*

All these processes will accelerate the rate of interfacial and fatigue wear of the already disrupted cartilage microstructure.

Hypotheses on the Biomechanics of Cartilage Degeneration

ROLE OF BIOMECHANICAL FACTORS

Articular cartilage has only a limited capacity for repair and regeneration, and if subjected to an abnormal range of stresses can quickly undergo total failure (Fig. 3-26). It has been hypothesized that failure progression relates to the following: (1) the magnitude of the imposed stresses; (2) the total number of sustained stress peaks; (3) the changes in the intrinsic molecular and microscopic structure of the collagen-PG matrix; and (4) the changes in the intrinsic mechanical property of the tissue. The most important failure-initiating factor appears to be the "loosening" of the collagen network that allows abnormal PG expansion and thus tissue swelling (Maroudas, 1976; McDevitt & Muir, 1976). Associated with this change is a decrease in cartilage stiffness and an increase in cartilage permeability (Altman et al., 1984; Armstrong & Mow, 1982; Guilak et al., 1994; Setton et al., 1994), both of which alter cartilage function in a diarthrodial joint during joint motion, as shown in Figure 3-27 (Mow & Ateshian, 1997).

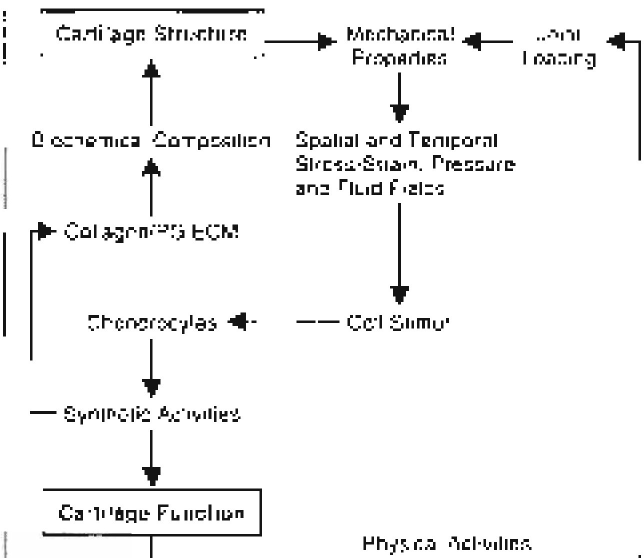


FIG. 3-26

Flow diagram of the events mediating the structure and function of articular cartilage. Physical activities result in joint loads that are transmitted to the chondrocyte via the extracellular matrix (ECM). The chondrocyte varies its cellular activities in response to the mechano-electrochemical stimuli generated by loading of its environment. The etiology of osteoarthritis is unclear but may be traced to intrinsic changes to the chondrocyte or to an altered ECM (e.g., resulting from injury or gradual wear) that leads to abnormal chondrocyte stimuli and cell activities.

The magnitude of the stress sustained by the articular cartilage is determined by both the total load on the joint and how that load is distributed over the articular surface contact area (Ahmed & Burke, 1983; Armstrong et al., 1979; Paul, 1976). Any intense stress concentration in the contact area will play a primary role in tissue degeneration. A large number of well-known conditions cause excessive stress concentrations in articular cartilage and result in cartilage failure. Most of these stress concentrations are caused by joint surface incongruity, resulting in an abnormally small contact area. Examples of conditions causing such joint incongruities include OA subsequent to congenital acetabular dysplasia, a slipped capital femoral epiphysis, and intra-articular fractures. Two further examples are knee joint meniscectomy, which eliminates the load-distributing function of the meniscus (Mow et al., 1992), and ligament rupture, which allows excessive movement and the generation of abnormal mechanical stresses in the affected joint (Altman et al., 1984; Guilak et al., 1994; McDevitt & Muir, 1976; Setton et al., 1994). In all the above cases, abnormal joint articulation increases the stress acting on the joint surface, which appears to predispose the cartilage to failure.

Macroscopically, stress localization and concentration at the joint surfaces have a further effect. High contact pressures between the articular surfaces decrease the probability of fluid-film lubrication (Mow & Ateshian, 1997). Subsequent actual surface-to-surface contact of asperities will cause microscopic stress concentrations that are responsible for further tissue damage (Ateshian et al., 1995, 1998; Ateshian & Wang, 1995) (Case Study 3-1).

The high incidence of specific joint degeneration in individuals with certain occupations, such as football players' knees and ballet dancers' ankles, can be explained by the increase in high and abnormal load frequencies and magnitude sustained by the joints of these individuals. It has been suggested

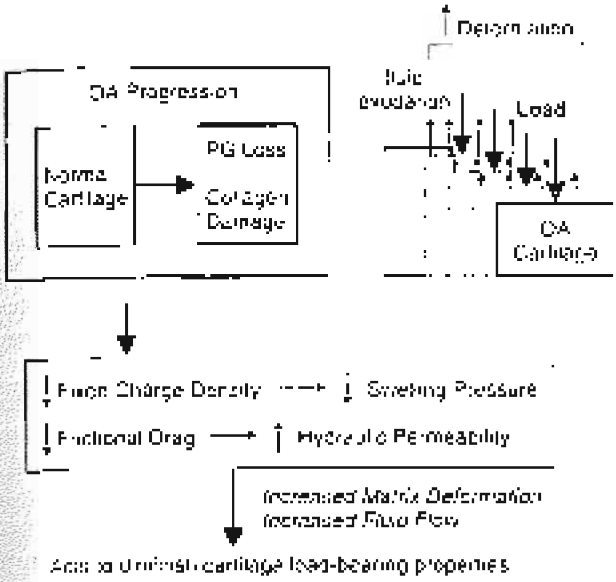


FIG. 3-27. A figure illustrating how osteoarthritic changes to the collagen-PG network can compromise the ability of articular cartilage to maintain interstitial fluid pressurization, which underlies the tissue's load-bearing and joint lubrication capacity. Loss of PG and damage to the collagen fibers result in an increased hydraulic permeability (decreased resistance to fluid flow) and supra-normal loads and strains on the solid matrix (and chondrocytes).

that, in some cases, OA may be caused by deficiencies in the mechanisms that act to minimize peak forces on the joints. Examples of these mechanisms include the active processes of joint flexion and muscle lengthening and the passive absorption of shocks by the subchondral bone (Radin, 1976) and meniscus (Mow et al., 1992).

Degenerative changes to the structure and composition of articular cartilage could lead to abnormal tissue swelling and functionally inferior biomechanical properties. In this weakened state, the cartilage ultrastructure will then be gradually destroyed by stresses of normal joint articulation (Fig. 3-27). OA may also arise secondarily from insult to the intrinsic molecular and microscopic structure of the collagen-PG matrix. Many conditions may promote such a breakdown in matrix integrity; these include degeneration associated with rheumatoid arthritis, joint space hemorrhage associated with hemophilia, various collagen metabolism disorders, and tissue degradation by proteolytic enzymes. The presence of soluble mediators such as cytokines (e.g., interleukin-1) (Ratcliffe et al., 1986) and growth factors (e.g., transforming growth factor beta 1) also appear to play an important role in OA. Another contributing factor to the etiology of OA may be age-related changes in the chondrocyte (Case Study 3-2).

IMPLICATIONS ON CHONDROCYTE FUNCTION

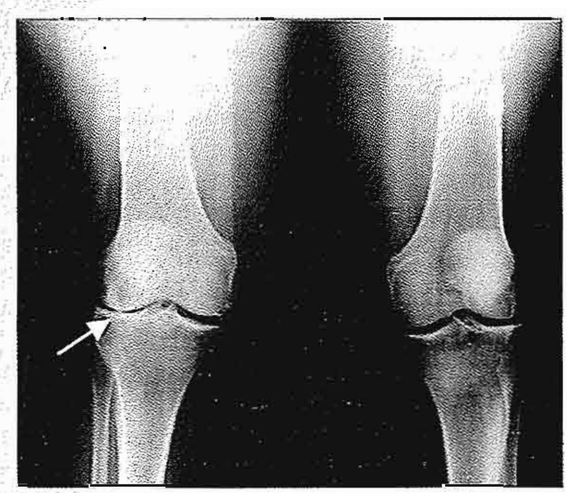
The ECM modulates the transmission of joint loads to the chondrocyte, acting as a transducer that converts mechanical loading to a plethora of environmental cues that mediate chondrocyte function. In healthy articular cartilage, loads from normal joint function motion result in the generation of

CASE STUDY 3-1

Knee Meniscectomy

Forty-year-old man who had a meniscectomy 10 years ago in his right knee. Currently he is suffering pain associated with movement, swelling, and limitations of knee motion (Fig. 3-1-1).

The history of knee meniscectomy not only implies an alteration in joint surface congruence but also the elimination of the load-distribution function of the meniscus. The effect is an abnormal joint, characterized by an increase in the stress acting on the joint surface that results in cartilage failure. Most of these stress concentrations are caused by joint surface incongruity resulting in an abnormally small contact area. This small contact area will suffer high contact pressure, decreasing the probability of fluid-film lubrication, and thus the actual surface-to-surface contact will cause microscopic stress concentrations that lead to damage.



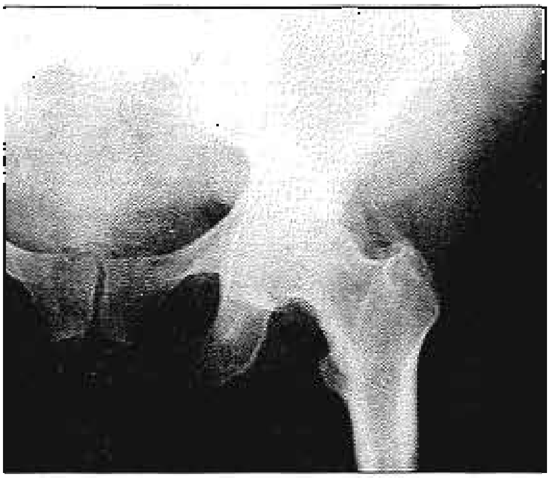
Case Study Figure 3-1-1

CASE STUDY 3-2

Osteoarthritis

Seventy-year-old woman, overweight, with OA of the right hip joint with associated symptoms of pain, limitation of motion, joint deformity, and abnormal gait (Fig. 3-2-1).

OA is characterized by erosive cartilage lesions, cartilage loss and destruction, subchondral bone sclerosis and cysts, and large osteophyte formation at the margins of the joint (Mow & Ratcliffe, 1997). In this case, roentgenography of the right hip of the patient shows a decrease in the interarticular space and changes in bone surfaces as sclerotic and osteophyte formations. The most severe alterations are found at the point of maximum pressure against the opposing cartilage surface, in this case at the superior aspect of the femoral head.



Case Study Figure 3-2-1

mechano-electrochemical stimuli (e.g., hydrostatic pressure, stress and strain fields, streaming potentials) that promote normal cartilage maintenance (by the chondrocytes) and normal tissue function (Fig. 3-26). However, when the integrity of the collagen-PG network (the transducer) of articular cartilage is compromised, such as from trauma or disease, normal joint articulation leads to abnormal mechano-electrochemical stimuli with ensuing abnormal ECM remodeling by the chondrocytes and debilitated tissue function.

In the absence of joint loading, the normal environment of the chondrocyte is characterized by the prestress established by the balance between tension in the collagen fibers and the Donnan osmotic pressure. During joint loading, by virtue of the tissue's low permeability, the normal environment of the chondrocyte is dominated by hydrostatic pressure of the interstitial fluid. Various speculations regarding interstitial fluid flow exist as well, implicated in enhancing nutrient diffusion. Interstitial fluid flow (i.e., of cationic counter ions) gives rise to cellular stimuli of an electrical nature, namely streaming potentials and currents (Frank & Grodzinsky, 1987; Gu et al., 1993, 1998). In addition, interstitial fluid flow through the small pores associated with the solid matrix (~50 nm) of normal cartilage, which offer considerable resistance to fluid flow (Maroudas, 1979; McCutchen, 1962; Mow et al., 1984), can give rise to a mechanical phenomenon termed fluid-induced matrix compaction (Lai & Mow, 1980). The frictional interaction between interstitial fluid and solid arises as a result of drag resistance to forced flow through the porous permeable cartilage matrix and a viscous shear stress exerted by the interstitial fluid. Given the normal flow rates of the interstitial fluid mentioned earlier and the low permeability of the cartilage matrix, chondrocyte perception of this frictional interaction force is likely to be dominated by the drag resistance of flow through the matrix rather than its direct viscous shear stress on the cell. This frictional drag force can produce solid matrix deformation on the order of 15 to 30%.

From the discussion above, chondrocyte deformation and function are considered to be governed by three coupled loading mechanisms: direct ECM deformation, fluid flow-induced compaction, and fluid pressurization. In OA, the increased tissue permeability diminishes cartilage's normal fluid pressure load support mechanism. Thus, there is a shift of load support onto the solid matrix, causing supraphysiological stresses and strains to be imposed on the chondrocytes (Fig. 3-27). These abnormally high stress and strain levels, and other mechano-electrochemical changes that are manifested with OA, can trigger an imbalance of chondrocyte anabolic and catabolic activities, further contributing to a vicious cycle of progressive cartilage degeneration. Indeed, changes to the biochemical composition and structure of cartilage can have a profound impact on tissue and chondrocyte function. With multidisciplinary collaborations and an appropriate theoretical framework, such as the biphasic theory, insights into the factors that govern chondrocyte function, cartilage structure and function, and the etiology of OA can be obtained.

Summary

1 The function of articular cartilage in diarthrodial joints is to increase the area of load distribution (thereby reducing the stress) and provide a smooth, wear-resistant bearing surface.

2 Biomechanically, articular cartilage should be viewed as a multiphasic material. In terms of a biphasic material, articular cartilage is composed of a porous-permeable collagen-PG solid matrix (approximately 25% by wet weight) filled by the freely movable interstitial fluid (approximately 75% by wet weight). In addition to solid and fluid, there exists an additional ion phase when considering articular cartilage as a triphasic medium. The ion phase is necessary to describe the swelling and other electromechanical behaviors of the tissue.

3 Important biomechanical properties of articular cartilage are the intrinsic material properties of the solid matrix and the frictional resistance to the flow of interstitial fluid through the porous-permeable solid matrix (a parameter inversely proportional to the tissue permeability). Together these parameters define the level of interstitial fluid pressurization, a major determinant of the load-bearing and lubrication capacity of the tissue, which can be generated in cartilage.

4 Damage to articular cartilage, from whatever cause, can disrupt the normal interstitial fluid load-bearing capacity of the tissue and thus the normal lubrication process operating within the joint. Therefore, lubrication insufficiency may be a primary factor in the etiology of OA.

5 When describing articular cartilage in the context of a rigorous theoretical framework such as the biphasic, triphasic, or multiphasic theories, it is possible to accurately predict the biomechanical behavior of articular cartilage under loading and to elucidate the underlying mechanisms that govern its load bearing and lubrication function. Furthermore, insights into the temporal and spatial nature of the physical stimuli that may affect chondrocyte function in situ can be gained.

ACKNOWLEDGMENTS

This work was sponsored by the National Institutes of Health grants AR41913 and AR42850.

REFERENCES

Ahmed, A.M., & Burke, D.L. (1983). In-vitro measurement of static pressure distribution in synovial joints—Part I: Tibial surface of the knee. *J Biomech Eng, 105*, 216.

Akizuki, S., Mow, V.C., Müller, F., et al. (1986). Tensile properties of knee joint cartilage: I. Influence of ionic conditions, weight bearing, and fibrillation on the tensile modulus. *J Orthop Res, 4*, 379.

Altman, R.D., Tenenbaum, J., Latta, L., et al. (1984). Biomechanical and biochemical properties of dog cartilage in experimentally induced osteoarthritis. *Ann Rheum Dis, 43*, 83.

Andriacchi, T.P., Natarajan, R.N., & Hurwitz, D.E. (1997). Musculoskeletal dynamics, locomotion, and clinical applications. In V.C. Mow & W.C. Hayes (Eds.), *Basic Orthopaedic Biomechanics* (2nd ed.) (pp. 37–68). Philadelphia: Lippincott-Raven.

Armstrong, C.G., Bahrani, A.S., & Gardner, D.L. (1979). In vitro measurement of articular cartilage deformations in the intact human hip joint under load. *J Bone Joint Surg, 61A*, 744.

Armstrong, C.G., & Mow, V.C. (1980). Friction, lubrication and wear of synovial joints. In Owen, R., Goodfellow, J., & Bullough, P. (Eds.), *Scientific Foundations of Orthopaedics and Traumatology* (pp. 223–232). London: William Heinemann.

Armstrong, C.G., & Mow, V.C. (1982). Variations in the intrinsic mechanical properties of human articular cartilage with age, degeneration, and water content. *J Bone Joint Surg, 64A*, 88.

Ateshian, G.A. (1997). Theoretical formulation for boundary friction in articular cartilage. *J Biomech Eng, 119*, 81.

Ateshian, G.A., Kwak, S.D., Soslowsky, L.J., et al. (1994). A new stereophotogrammetry method for determining in situ contact areas in diarthrodial joints: A comparison study. *J Biomechanics, 27*, 111.

Ateshian, G.A., Lai, W.M., Zhu, W.B., et al. (1994). An asymptotic solution for the contact of two biphasic cartilage layers. *J Biomechanics, 27*, 1347.

Ateshian, G.A., & Wang, H. (1995). A theoretical solution for the frictionless rolling contact of cylindrical biphasic articular cartilage layers. *J Biomechanics, 28*, 1341.

Ateshian, G.A., Wang, H., & Lai, W.M. (1998). The role of interstitial fluid pressurization and surface porosities on the boundary friction of articular cartilage. *J Tribology, 120*, 241.

Ateshian, G.A., Warden, W.H., Kim, J.J., et al. (1997). Finite deformation biphasic material properties of bovine articular cartilage from confined compression experiments. *J Biomechanics, 30*, 1157.

Athanasiou, K.A., Rosenwasser, M.P., Buckwalter, J.A., et al. (1991). Interspecies comparison of in situ mechanical properties of distal femoral cartilage. *J Orthop Res, 9*, 330.

Bachrach, N.M., Valhmu, W.B., Stazzone, E.J., et al. (1995). Changes in proteoglycan synthesis rates of chondrocytes in articular cartilage are associated with the time-dependent changes in the mechanical environment. *J Biomechanics, 28*, 1561.

Bachrach, N.M., Mow, V.C., & Guilak, F. (1998). Incompressibility of the solid matrix of articular cartilage under high hydrostatic pressure. *J Biomechanics, 31*, 445.

Bateman, J.F., Lamande, S.R., & Ramshaw, J.A.M. (1996). Collagen superfamily. In W.D. Comper (Ed.), *Extracellular Matrix* (Vol. 2, pp. 22–67). Amsterdam: Harwood Academic Pubs.

Bollet, A.J., & Nance, J.L. (1966). Biochemical findings in normal and osteoarthritic articular cartilage. II: Chondroitin sulfate concentration and chain length, water and ash content. *J Clin Invest, 45*, 1170.

Bowden, F.P., & Tabor, D. (1967). *Friction and Lubrication.* London: Methuen Pubs.

Broom, N.D., & Silyn-Roberts, H. (1990). Collagen-collagen versus collagen-proteoglycan interactions in the determination of cartilage strength. *Arthritis Rheum, 33*, 1512.

Buckwalter, J.A., Kuettner, K.E., & Thonar, E.J.-M.A. (1985). Age-related changes in articular cartilage proteoglycans: Electron microscopic studies. *J Orthop Res, 3*, 251.

Bullough, P.G., & Goodfellow, J. (1968). The significance of the fine structures of articular cartilage. *J Bone Joint Surg, 50B*, 852.

Bullough, P.G., & Jagannath, A. (1983). The morphology of the calcification front in articular cartilage. *J Bone Joint Surg, 65B*, 72.

Buschmann, M.D., Gluzband, Y.A., Grodzinsky, A.J., et al. (1992). Chondrocytes in agarose culture synthesize a mechanically functional extracellular matrix. *J Orthop Res, 10*, 745.

Buschmann, M.D., & Grodzinsky, A.J. (1995). A molecular model of proteoglycan-associated electrostatic forces in cartilage mechanics. *J Biomech Eng, 117*, 179.

Clark, J.M. (1985). The organization of collagen in cryofractured rabbit articular cartilage: A scanning electron microscopic study. *J Orthop Res, 3*, 17.

Clarke, I.C. (1971). Articular cartilage: A review and scanning electron microscope study—1. The interterritorial fibrillar architecture. *J Bone Joint Surg, 53B*, 732.

Donohue, J.M., Buss, D., Oegema, T.R., et al. (1983). The effects of indirect blunt trauma on adult canine articular cartilage. *J Bone Joint Surg, 65A*, 948.

Donnan, F.G. (1924). The theory of membrane equilibria. *Chemical Review, 1*, 73.

Dowson, D. (1966/1967). Modes of lubrication in human joints. *Proc Inst Mech Eng, 181J*, 45.

Dowson, D. (1990). Bio-tribology of natural and replacement joints. In V.C. Mow, A. Ratcliffe, & S.L.Y. Woo (Eds.), *Biomechanics of Diarthrodial Joints* (pp. 305–345). New York: Springer-Verlag.

Edwards, J. (1967). Physical characteristics of articular cartilage. *Proc Inst Mech Eng, 181J*, 16.

Elmore, S.M., Sokoloff, L., Norris, G., et al. (1963). Nature of "imperfect" elasticity of articular cartilage. *J Applied Physiol, 18*, 393.

Eyre, D.R. (1980). Collagen: Molecular diversity in the body's protein scaffold. *Science, 207*, 1315.

Fosang, A.J., & Hardingham, T.E. (1996). Matrix proteoglycans. In W.D. Comper (Ed.), *Extracellular Matrix* (Vol. 2, pp. 200–229). Amsterdam: Harwood Academic Pubs.

Frank, E.H., & Grodzinsky, A.J. (1987a). Cartilage electromechanics—I. Electrokinetic transduction and effects of pH and ionic strength. *J Biomechanics, 20*, 615.

Frank, E.H., & Grodzinsky, A.J. (1987b). Cartilage electromechanics—II. A continuum model of cartilage electrokinetics and correlation with experiments. *J Biomechanics, 20*, 629.

Freeman, M.A.R. (1975). The fatigue of cartilage in the pathogenesis of osteoarthrosis. *Acta Orthop Scand, 46*, 323.

Fung, Y.C. (1981). Quasi-linear viscoelasticity of soft tissues. In *Biomechanics: Mechanical Properties of Living Tissues* (p. 226). New York: Springer-Verlag.

Gardner, S.L., & McGillivray, D.C. (1971). Living articular cartilage is not smooth. The structure of mammalian and avian joint surfaces demonstrated in vivo by immersion incident light microscopy. *Ann Rheum Dis, 30*, 3.

Garg, H.G., & Swann, D.A. (1981). Age-related changes in the chemical composition of bovine articular cartilage. *Biochem J, 193*, 459.

Gu, W.Y., Lai, W.M., & Mow, V.C. (1993). Transport of fluid and ions through a porous-permeable charged-hydrated tissue, and streaming potential data on normal bovine articular cartilage. *J Biomechanics, 26*, 709.

Gu, W.Y., Lai, W.M., & Mow, V.C. (1997). A triphasic analysis of negative osmotic flows through charged hydrated soft tissues. *J Biomechanics, 30*, 71.

Gu, W.Y., Lai, W.M., & Mow, V.C. (1998). A mixture theory for charged-hydrated soft tissues containing multi-electrolytes: Passive transport and swelling behaviors. *J Biomech Eng, 120*, 169.

Guilak, F., Ratcliffe, A., Lane, N., et al. (1994). Mechanical and biochemical changes in the superficial zone of articular cartilage in a canine model of osteoarthritis. *J Orthop Res, 12*, 474.

Hardingham, T.E., & Muir, H. (1974). Hyaluronic acid in cartilage and proteoglycan aggregation. *Biochem J, 139*, 565.

Hardingham, T.E., Beardmore-Gray, M., & Dunham, D.G. (1987). Protein domain structure of the aggregating proteoglycan from cartilage. *Trans Orthop Res Soc, 12*, 61.

Hascall, V.C. (1977). Interactions of cartilage proteoglycans with hyaluronic acid. *J Supramol Structure, 7*, 101.

Hayes, W.C., & Bodine, A.J. (1978). Flow-independent viscoelastic properties of articular cartilage matrix. *J Biomechanics, 11*, 407.

Hayes, W.C., & Mockros, L.F. (1971). Viscoelastic properties of human articular cartilage. *J Appl Physiol, 31*, 562.

Hawkins, S.W. (1988). *A Brief History of Time: From the Big Bang to Black Holes.* New York: Bantam Books.

Heinegård, D., Wieslander, J., Sheehan, J., et al. (1985). Separation and characterization of two populations of aggregating proteoglycans from cartilage. *Biochem J, 225*, 95.

Helminen, H.J., Kiviranta, I., Tammi, M., et al. (Eds.) (1987). *Joint Loading: Biology and Health of Articular Structures.* Bristol, UK: Wright & Sons, Pubs.

Hills, B.A. (1989). Oligolamellar lubrication of joints by surface active phospholipid. *J Rheumatol, 16*, 82.

Hirsch, C. (1944). The pathogenesis of chondromalacia of the patella. *Acta Chir Scand, 83*(Suppl), 1.

Hlaváček, M. (1995). The role of synovial fluid filtration by cartilage in lubrication of synovial joints: IV. Squeeze-film lubrication for axial symmetry under high loading conditions. *J Biomechanics, 28*, 1199.

Hodge, W.A., Fijan, R.S., Carlson, K., et al. (1986). Contact pressures in the human hip joint measured in vivo. *Proc Natl Acad Sci USA, 83*, 2879.

Holmes, M.H., Lai, W.M., & Mow, V.C. (1985). Singular perturbation analysis of the nonlinear, flow-dependent, compressive stress-relaxation behavior of articular cartilage. *J Biomech Eng, 107*, 206.

Hou, J.S., Mow, V.C., Lai, W.M., et al. (1992). An analysis of the squeeze-film lubrication mechanism for articular cartilage. *J Biomechanics, 25*, 247.

Hultkrantz, W. (1898). Ueber die Spaltrichtungen der Gelenkknorpel. *Verhandlungen der Anatomischen Gesellschaft, 12*, 248.

Katchalsky, A., & Curran, P.F. (1975). *Nonequilibrium Thermodynamics in Biophysics* (4th ed.). Cambridge: Harvard University Press.

Kempson, G.E., Tuke, M.A., Dingle, J.T., et al. (1976). The effects of proteolytic enzymes on the mechanical properties of adult human articular cartilage. *Biochim Biophys Acta*, 428, 741.

Kempson, G.E. (1979). Mechanical properties of articular cartilage. In M.A.R. Freeman (Ed.), *Adult Articular Cartilage* (2nd ed., pp. 333–414). Tunbridge Wells, U.K.: Pitman Medical.

Kim, Y.J., Sah, R.L., Grodzinsky, A.J., et al. (1994). Mechanical regulation of cartilage biosynthetic behavior: Physical stimuli. *Arch Biochem Biophys*, 311, 1.

Lai, W.M., Gu, W.Y., & Mow, V.C. (1998). On the conditional equivalence of chemical loading and mechanical loading on articular cartilage. *J Biomechanics*, 31(12), 1181–1185.

Lai, W.M., & Mow, V.C. (1978). Ultrafiltration of synovial fluid by cartilage. *J Eng Mech Div ASCE*, 104, 79.

Lai, W.M., & Mow, V.C. (1980). Drag-induced compression of articular cartilage during a permeation experiment. *J Biorheology*, 17, 111.

Lai, W.M., Hou, J.S., & Mow, V.C. (1991). A triphasic theory for the swelling and deformational behaviors of articular cartilage. *J Biomech Eng*, 113, 245.

Lakes, R., & Saha, S. (1979). Cement line motion in bone. *Science*, 204, 501.

Lane, J.M., & Weiss, C. (1975). Review of articular cartilage collagen research. *Arthritis Rheum*, 18, 553.

Linn, F.C. (1968). Lubrication of animal joints: I. The mechanism. *J Biomechanics*, 1, 193.

Linn, F.C., & Radin, E.L. (1968). Lubrication of animal joints: III. The effect of certain chemical alterations of the cartilage and lubricant. *Arthritis Rheum*, 11, 674.

Linn, F.C., & Sokoloff, L. (1965). Movement and composition of interstitial fluid of cartilage. *Arthritis Rheum*, 8, 481.

Lipshitz, H., Etheredge, R., & Glimcher, M.J. (1975). In vitro wear of articular cartilage: I. Hydroxyproline, hexosamine, and amino acid composition of bovine articular cartilage as a function of depth from the surface; hydroxyproline content of the lubricant and the wear debris as a measure of wear. *J Bone Joint Surg*, 57A, 527.

Lipshitz, H., Etheredge, R., & Glimcher, M.J. (1976). Changes in the hexosamine content and swelling ratio of articular cartilage as functions of depth from the surface. *J Bone Joint Surg*, 58A, 1149.

Lipshitz, H., & Glimcher, M.J. (1979). In vitro studies of the wear of articular cartilage. *Wear*, 52, 297.

Malcolm, L.L. (1976). *An experimental investigation of the frictional and deformational responses of articular cartilage interfaces to static and dynamic loading*. Doctoral thesis, University of California, San Diego.

Mankin, H.J., & Thrasher, A.Z. (1975). Water content and binding in normal and osteoarthritic human cartilage. *J Bone Joint Surg*, 57A, 76.

Mansour, J.M., & Mow, V.C. (1976). The permeability of articular cartilage under compressive strain and at high pressures. *J Bone Joint Surg*, 58A, 509.

Maroudas, A. (1966–1967). Hyaluronic acid films. *Proc Inst Mech Eng, London*, 181J, 122.

Maroudas, A. (1968). Physicochemical properties of cartilage in light of ion-exchange theory. *Biophys J*, 8, 575.

Maroudas, A. (1975). Biophysical chemistry of cartilaginous tissues with special reference to solute and fluid transport. *Biorheology*, 12, 233.

Maroudas, A. (1976). Balance between swelling pressure and collagen tension in normal and degenerate cartilage. *Nature*, 260, 808.

Maroudas, A. (1979). Physicochemical properties of articular cartilage. In M.A.R. Freeman (Ed.), *Adult Articular Cartilage* (2nd ed., pp. 215–290). Tunbridge Wells, England: Pitman Medical.

Maroudas, A., Wachtel, E., Grushko, G., et al. (1991). The effect of osmotic and mechanical pressures on water partitioning in articular cartilage. *Biochim Biophys Acta*, 1073, 285.

McCutchen, C.W. (1962). The frictional properties of animal joints. *Wear*, 5, 1.

McDevitt, C.A., & Muir, H. (1976). Biochemical changes in the cartilage of the knee in experimental and natural osteoarthritis in the dog. *J Bone Joint Surg*, 58B, 94.

Meachim, G., & Fergie, I.A. (1975). Morphological patterns of articular cartilage fibrillation. *J Pathol*, 115, 231.

Mow, V.C., Arnoczky, S.P., & Jackson, D.W. (1992). *Knee Meniscus: Basic and Clinical Foundations*. New York: Raven Press.

Mow, V.C., & Ateshian, G.A. (1997). Lubrication and wear of diarthrodial joints. In V.C. Mow & W.C. Hayes (Eds.), *Basic Orthopaedic Biomechanics* (2nd ed., pp. 275–315). Philadelphia: Lippincott-Raven.

Mow, V.C., Ateshian, G.A., Lai, W.M., et al. (1998). Effects of fixed charges on the stress-relaxation behavior of hydrated soft tissues in a confined compression problem. *Int J Solids & Structures*, 35, 4945–4962.

Mow, V.C., Gibbs, M.C., Lai, W.M., et al. (1989a). Biphasic indentation of articular cartilage: Part II. A numerical algorithm and an experimental study. *J Biomechanics*, 22, 853.

Mow, V.C., Holmes, M.H., & Lai, W.M. (1984). Fluid transport and mechanical properties of articular cartilage: A review. *J Biomechanics*, 17, 377.

Mow, V.C., Kuei, S.C., Lai, W.M., et al. (1980). Biphasic creep and stress relaxation of articular cartilage in compression: Theory and experiments. *J Biomech Eng*, 102, 73.

Mow, V.C., Lai, W.M., & Redler, I. (1974). Some surface characteristics of articular cartilages. A scanning electron microscopy study and a theoretical model for the dynamic interaction of synovial fluid and articular cartilage. *J Biomechanics*, 7, 449.

Mow, V.C., & Ratcliffe, A. (1997). Structure and function of articular cartilage and meniscus. In V.C. Mow & W.C. Hayes (Eds.), *Basic Orthopaedic Biomechanics* (2nd ed., pp. 113–177). Philadelphia: Lippincott-Raven.

Mow, V.C., Zhu, W.B., Lai, W.M., et al. (1989b). The influence of link protein stabilization on the viscoelastic properties of proteoglycan aggregates. *Biochim Biophys Acta*, 992, 201.

Muir, H. (1983). Proteoglycans as organizers of the extracellular matrix. *Biochem Soc Trans*, 11, 613.

Myers, E.R., Lai, W.M., & Mow, V.C. (1984). A continuum theory and an experiment for the ion-induced swelling behavior of cartilage. *J Biomech Eng*, 106(2), 151–158.

Newberry, W.N., Zukosky, D.K., & Haut, R.C. (1997). Subfracture insult to a knee joint causes alterations in the bone and in the functional stiffness of overlying cartilage. *J Orthop Res*, 15, 450.

Onsager, L. (1931). Reciprocal relations in irreversible processes. I. *Phys Rev, 37*, 405.

Paul, J.P. (1976). Force actions transmitted by joints in the human body. *Proc R Soc Lond, 192B*, 163.

Poole, A.R. (1986). Proteoglycans in health and disease: Structure and function. *Biochem J, 236*, 1.

Radin, E.L., and Paul, I.L. (1971). Response of joints to impact loading. I. In vitro wear. *Arthritis Rheum, 14*, 356.

Radin, E.L. (1976). Aetiology of osteoarthrosis. *Clin Rheum Dis, 2*, 509.

Ratcliffe, A., & Mow, V.C. (1996). Articular cartilage. In W.D. Comper (Ed.), *Extracellular Matrix* (Vol. 1, pp. 234–302). Amsterdam: Harwood Academic Publs.

Ratcliffe, A., Tyler, J., & Hardingham, T.E. (1986). Articular cartilage cultured with interleukin 1: Increased release of link protein, hyaluronate-binding region and other proteoglycan fragments. *Biochem J, 238*, 571.

Redler, I., & Zimny, M.L. (1970). Scanning electron microscopy of normal and abnormal articular cartilage and synovium. *J Bone Joint Surg, 52A*, 1395.

Redler, I., Zimny, M.L., Mansell, J., et al. (1975). The ultrastructure and biomechanical significance of the tidemark of articular cartilage. *Clin Orthop Rel Res, 112*, 357.

Rosenberg, L., Choi, H.U., Tang, L-H., et al. (1985). Isolation of dermatan sulfate proteoglycans from mature bovine articular cartilage. *J Biol Chem, 260*, 6304.

Rosenberg, L., Hellmann, W., & Kleinschmidt, A.K. (1975). Electron microscopic studies of proteoglycan aggregates from bovine articular cartilage. *J Biol Chem, 250*, 1877.

Roth, V., & Mow, V.C. (1980). The intrinsic tensile behavior of the matrix of bovine articular cartilage and its variation with age. *J Bone Joint Surg, 62A*, 1102.

Roughley, P.J., & White, R.J. (1980). Age-related changes in the structure of the proteoglycan subunits from human articular cartilage. *J Biol Chem, 255*, 217.

Roughley, P.J., White, R.J., & Santer, V. (1981). Comparison of proteoglycans extracted from high and low weight-bearing human articular cartilage, with particular reference to sialic acid content. *J Biol Chem, 256*, 12699.

Schinagl, R.M., Gurskis, D., Chen, A.C., et al. (1997). Depth-dependent confined compression modulus of full-thickness bovine articular cartilage. *J Orthop Res, 15*, 499.

Schinagl, R.M., Ting, M.K., Price, J.H., et al. (1996). Video microscopy to quantitate the inhomogeneous equilibrium strain within articular cartilage during confined compression. *Ann Biomed Eng, 24*, 500.

Schmidt, M.B., Mow, V.C., Chun, L.E., et al. (1990). Effects of proteoglycan extraction on the tensile behavior of articular cartilage. *J Orthop Res, 8*, 353.

Schneiderman, R., Keret, D., & Maroudas, A. (1986). Effects of mechanical and osmotic pressure on the rate of glycosaminoglycan synthesis in adult femoral head cartilage: An in vitro study. *J Orthop Res, 4*, 393.

Schubert, M., & Hamerman, D. (1968). *A Primer on Connective Tissue Biochemistry*. Philadelphia: Lea & Febiger.

Scott, J.E., & Orford, C.R. (1981). Dermatan sulphate-rich proteoglycan associates with rat tail-tendon collagen at the d band in the gap region. *Biochem J, 197*, 213.

Setton, L.A., Gu, W.Y., Lai, W.M., et al. (1995). Predictions of the swelling-induced pre-stress in articular cartilage. In A.P.S. Selvadurai (Ed.), *Mechanics of Poroelastic Media* (pp. 299–322). Kluwer Academic Publs: Dordrecht, the Netherlands.

Setton, L.A., Mow, V.C., Müller, F.J., et al. (1994). Mechanical properties of canine articular cartilage are significantly altered following transection of the anterior cruciate ligament. *J Orthop Res, 12*, 451.

Setton, L.A., Tohyama, H., & Mow, V.C. (1998). Swelling and curling behavior of articular cartilage. *J Biomech Eng, 120*, 355.

Setton, L.A., Zhu, W.B., & Mow, V.C. (1993). The biphasic poroviscoelastic behavior of articular cartilage in compression: Role of the surface zone. *J Biomech, 26*, 581.

Sokoloff, L. (1963). Elasticity of articular cartilage: Effect of ions and viscous solutions. *Science, 141*, 1055.

Soltz, M.A., & Ateshian, G.A. (1998). Experimental verification and theoretical prediction of cartilage interstitial fluid pressurization at an impermeable contact interface in confined compression. *J Biomech, 31*(10), 927–934.

Stockwell, R.A. (1979). *Biology of Cartilage Cells*. Cambridge: Cambridge University Press.

Sun, D.N., Gu, W.Y., Guo, X.E., et al. (1999). The influence of inhomogeneous fixed charge density of cartilage mechanoelectrochemical behaviors. *Trans Orthop Res Soc, 24*, 484.

Swann, D.A., Radin, E.L., & Hendren, R.B. (1979). The lubrication of articular cartilage by synovial fluid glycoproteins. *Arthritis Rheum, 22*, 665.

Swann, D.A., Silver, F.H., Slayter, H.S., et al. (1985). The molecular structure and lubricating activity of lubricin isolated from bovine and human synovial fluids. *Biochem J, 225*, 195.

Sweet, M.B.E., Thonar, E.J-M.A., & Marsh, J. (1979). Age-related changes in proteoglycan structure. *Arch Biochem Biophys, 198*, 439–448.

Thompson, R.C., Oegema, T.R., Lewis, J.L., et al. (1991). Osteochondral changes after acute transarticular load. An animal model. *J Bone Joint Surg, 73A*, 990.

Thonar, E.J-M.A., Bjornsson, S., & Kuettner, K.E. (1986). Age-related changes in cartilage proteoglycans. In K. Kuettner, R.S. Schleyerbach, & V.C. Hascall (Eds.), *Articular Cartilage Biochemistry* (pp. 273–287). New York: Raven Press.

Torzilli, P.A., & Mow, V.C. (1976). On the fundamental fluid transport mechanisms through normal and pathological cartilage during function. I. The formulation. *J Biomech, 9*(8), 541–552.

Torzilli, P.A., Rose, D.E., & Dethmers, D.A. (1982). Equilibrium water partition in articular cartilage. *Biorheology, 19*, 519.

Urban, J.P.G., & McMullin, J.F. (1985). Swelling pressure of the intervertebral disc: Influence of collagen and proteoglycan content. *Biorheology, 22*, 145.

Valhmu, W.B., Stazzone, E.J., Bachrach, N.M., et al. (1998). Load-controlled compression of articular cartilage induces a transient stimulation of aggrecan gene expression. *Arch Biochem Biophys, 353*, 29.

Venn, M.F. (1978). Variation of chemical composition with age in human femoral head cartilage. *Ann Rheum Dis, 37*, 168.

Wada, T., & Akizuki, S. (1987). An ultrastructural study of solid matrix in articular cartilage under uniaxial tensile stress. *J Jpn Orthop Assoc, 61*.

Walker, P.S., Dowson, D., Longfield, M.D., et al. (1968). "Boosted lubrication" in synovial joints by fluid entrapment and enrichment. *Ann Rheum Dis, 27*, 512.

Walker, P.S., Unsworth, A., Dowson, D., et al. (1970). Mode of aggregation of hyaluronic acid protein complex on the surface of articular cartilage. *Ann Rheum Dis, 29*, 591.

Wang, C.B., & Mow, V.C. (1998). Inhomogeneity of aggregate modulus affects cartilage compressive stress-relaxation behavior. *Trans Orthop Res Soc, 23*(1), 481.

Weiss, C., Rosenberg, L., & Helfet, A.J. (1968). An ultrastructural study of normal young adult human articular cartilage. *J Bone Joint Surg, 50A*, 663.

Williams, P.F., Powell, G.L., & Laberge, M. (1993). Sliding friction analysis of phosphatidylcholine as a boundary lubricant for articular cartilage. *Proc Inst Mech Eng, 207*, 59.

Woo, S.L-Y., Mow, V.C., & Lai, W.M. (1987). Biomechanical properties of articular cartilage. In *Handbook of Bioengineering* (pp. 4.1–4.44). New York: McGraw-Hill.

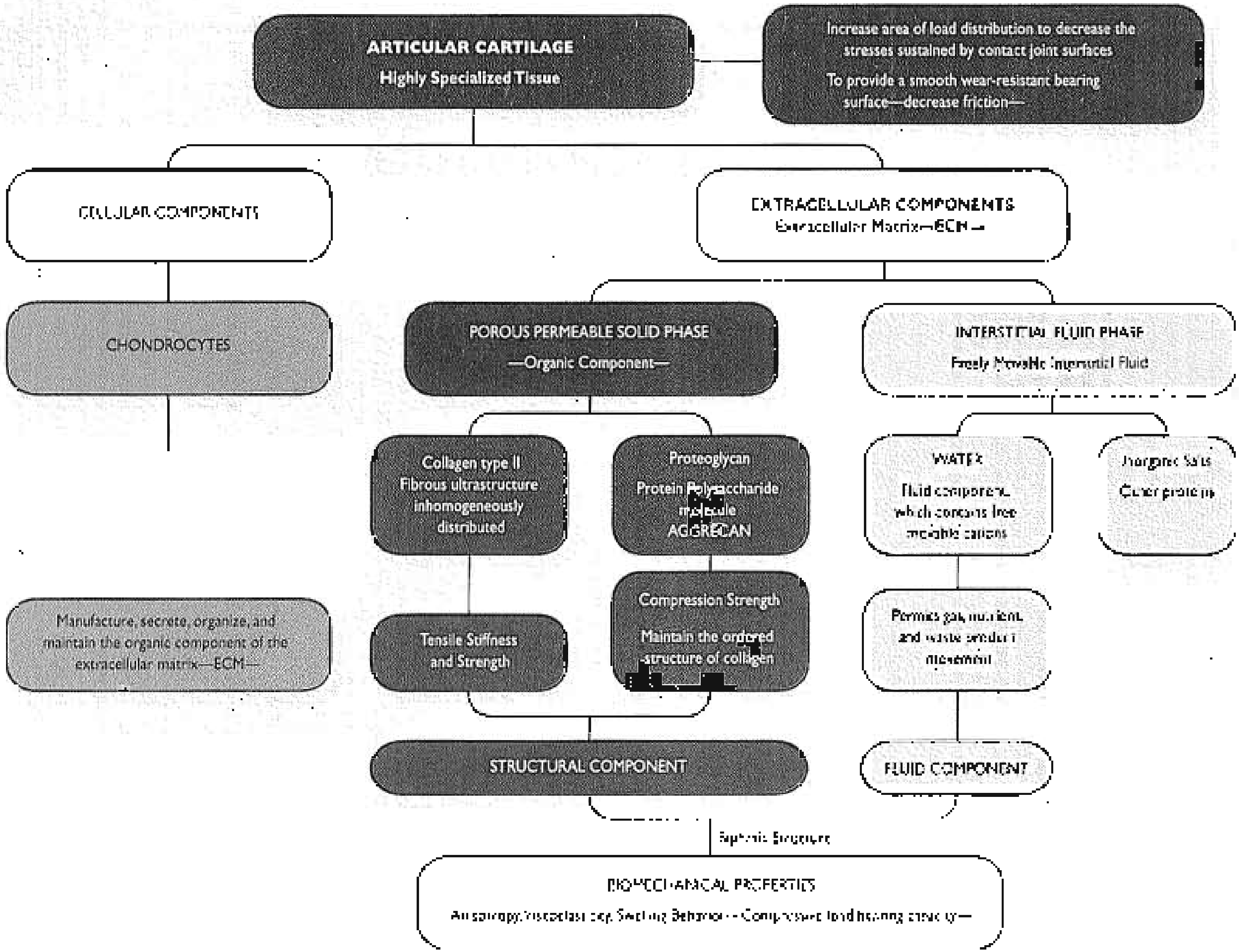
Woo, S.L-Y., Livesay, G.A., Runco, T.J., et al. (1997). Structure and function of tendons and ligaments. In V.C. Mow & W.C. Hayes (Eds.), *Basic Orthopaedic Biomechanics* (2nd ed., pp. 209–251). Philadelphia: Lippincott-Raven.

Zhu, W.B., Iatridis, J.C., Hlibczuk, V., et al. (1996). Determination of collagen-proteoglycan interactions in vitro. *J Biomechanics, 29*, 773.

Zhu, W.B., Lai, W.M., & Mow, V.C. (1986). Intrinsic quasi-linear viscoelastic behavior of the extracellular matrix of cartilage. *Trans Orthop Res Soc, 11*, 407.

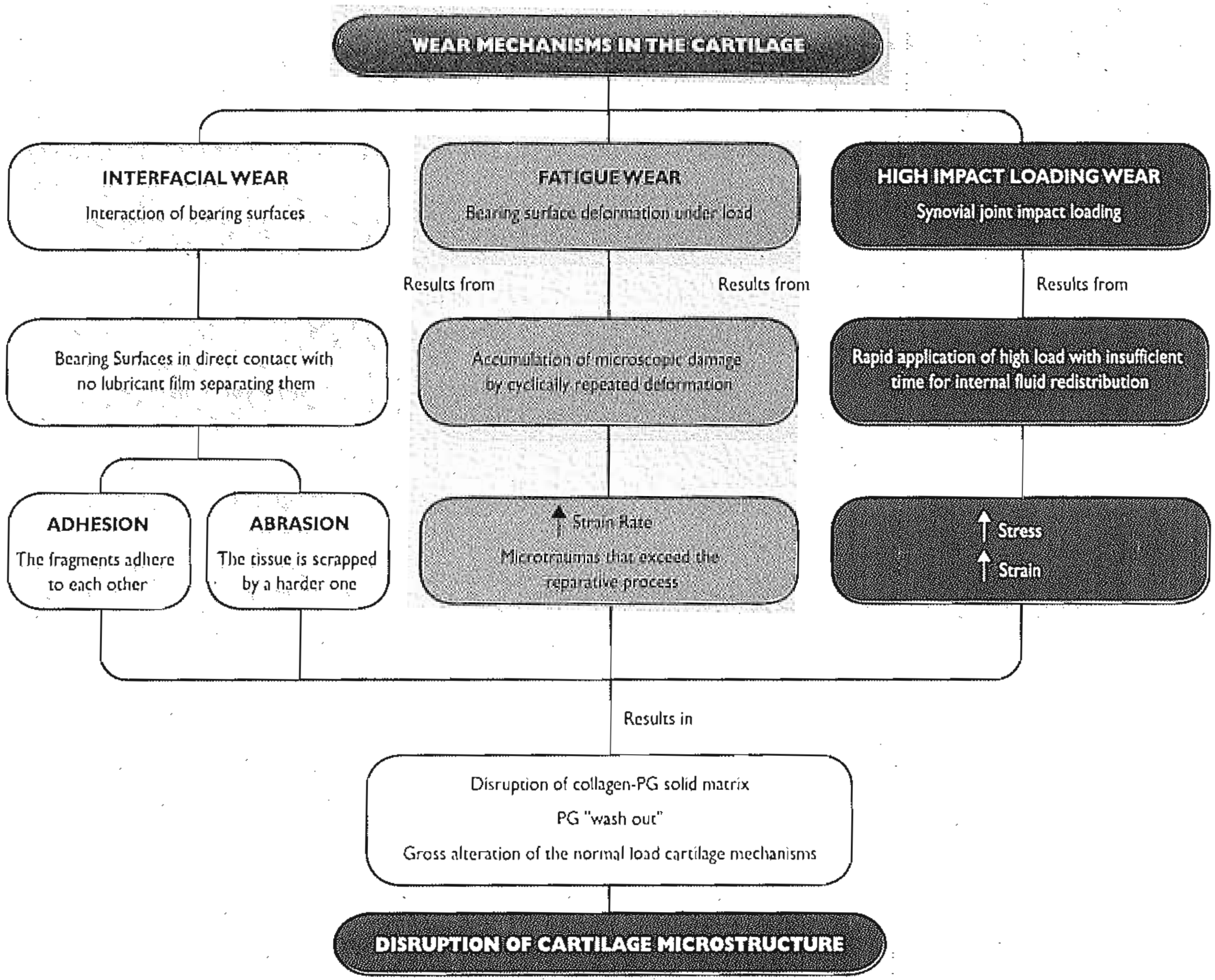
Zhu, W.B., Lai, W.M., & Mow, V.C. (1991). The density and strength of proteoglycan-proteoglycan interaction sites in concentrated solutions. *J Biomechanics, 24*, 1007.

Zhu, W.B., Mow, V.C., Koob, T.J., et al. (1993). Viscoelastic shear properties of articular cartilage and the effects of glycosidase treatments. *J Orthop Res, 11*, 771.



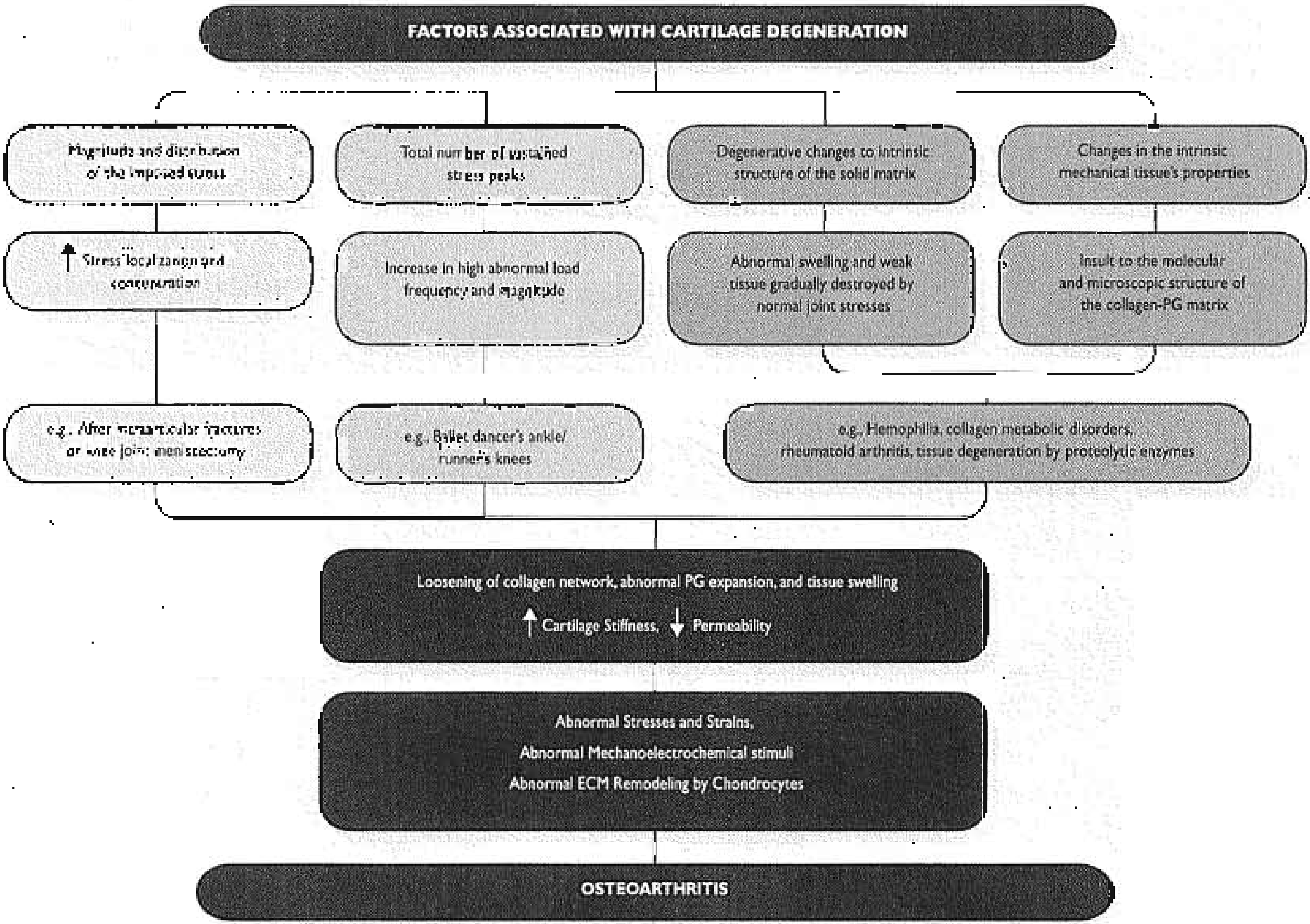
FLOW CHART 3-1 Articular cartilage structure and biomechanical properties.*

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.



FLOW CHART 3-2 Articular cartilage wear mechanisms.*

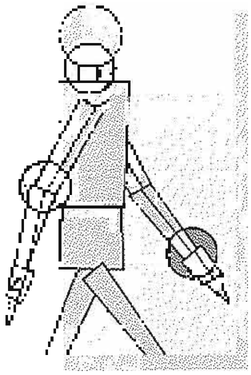
*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.



FLOW CHART 3-3 Factors associated with cartilage degeneration.* PG, proteoglycan; ECM, extracellular matrix.

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.

CHAPTER



Biomechanics of Tendons and Ligaments

Margareta Nordin, Tobias Lorenz, Marco Campello

Introduction

The three principal structures that closely surround, connect, and stabilize the joints of the skeletal system are tendons, ligaments, and joint capsules. Although these structures are passive (i.e., they do not actively produce motion as do the muscles), each plays an essential role in joint motion.

The role of the ligaments and joint capsules, which connect bone with bone, is to augment the mechanical stability of the joints, to guide joint motion, and to prevent excessive motion. Ligaments and joint capsules act as static restraints. The function of the tendons is to attach muscle to bone and to transmit tensile loads from muscle to bone, thereby producing joint motion, or to maintain the body posture. The tendons and the muscles form the muscle-tendon unit, which acts as a dynamic restraint. The tendon also enables the muscle belly to be at an optimal distance from the joint on which it acts without requiring an extended length of muscle between origin and insertion.

Tendon and ligament injuries and derangements are common. Proper management of these disorders requires an understanding of the mechanical properties and function of tendons and ligaments and their capacity for self-repair. This chapter discusses the following:

1. Composition and structure of tendons and ligaments
2. Biomechanical properties and behavior of normal tendon and ligament tissue
3. Biomechanical properties and behavior of injured tendon and ligament tissue

Several factors that affect the biomechanical function of tendons and ligaments are aging, pregnancy, mobilization and immobilization, diabetes, nonsteroidal anti-inflammatory drug (NSAID) use, and the effects of hemodialysis. Biomechanical considerations regarding grafts are also given.

Composition and Structure of Tendons and Ligaments

Tendons and ligaments are dense connective tissues known as parallel-fibered collagenous tissues. These sparsely vascularized tissues are composed largely of collagen, a fibrous protein constituting approximately one third of the total protein in the body (White, Handler, & Smith, 1964). Collagen constitutes a large portion of the organic matrix of bone and cartilage and has a unique mechanical supportive function in other connective tissues such as blood vessels, heart, ureters, kidneys, skin, and liver. The great mechanical stability of collagen gives the tendons and ligaments their characteristic strength and flexibility.

Like other connective tissues, tendons and ligaments consist of relatively few cells (fibroblasts) and an abundant extracellular matrix. In general, the cellular material occupies approximately 20% of the total tissue volume, while the extracellular matrix accounts for the remaining 80%. Approximately 70% of the matrix consists of water, and approximately 30% is solids. These solids are collagen, ground substance, and a small amount of elastin. The collagen content is generally over 75% and is somewhat greater in tendons than in ligaments (Kasser, 1996). In extremity tendons, the solid material may consist almost entirely of collagen (up to 99% of the dry weight) (Table 4-1).

The structure and chemical composition of ligaments and tendons are identical in humans and in many animal species such as rats, rabbits, dogs, and monkeys. Hence, extrapolations regarding these structures in humans can be made from the results of studies on these animal species.

COLLAGEN

The collagen molecule is synthesized by the fibroblast within the cell as a larger precursor (procollagen), which is then secreted and cleaved extracellularly to become collagen (Fitton-Jackson, 1965) (Fig. 4-1). Tendons and ligaments, like bone, are composed of the most common collagen molecule.

TABLE 4-1

Structural Composition of Tendons and Ligaments

Component	Ligament	Tendon
Cellular Material	20%	20%
Fibroblast		
Extracellular Matrix	80%	80%
Water	60–80%	60–80%
Solids	20–40%	20–40%
Collagen	70–80%	Slightly higher
Type I	90%	95–99%
Type 3	10%	1–5%
Ground substance	20–30%	Slightly lesser

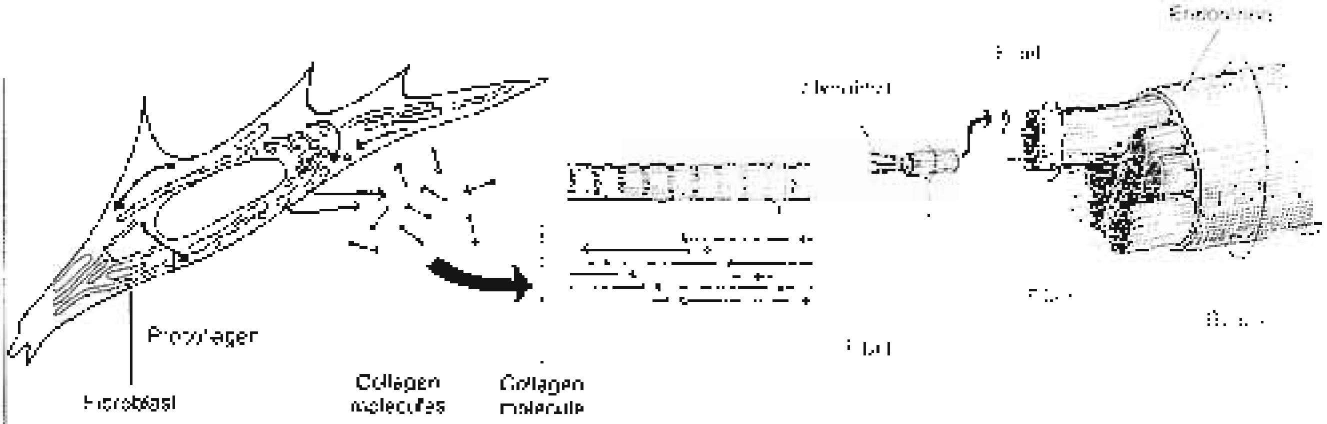


FIG. 4-1

Schematic representation of collagen fibrils, fibers, and bundles in tendons and collagenous ligaments (not drawn to scale). Collagen molecules, triple helices of coiled polypeptide chains, are synthesized and secreted by the fibroblasts. These molecules (depicted with "heads" and "tails" to represent positive and negative polar charges) aggregate in the extracellular matrix in a parallel arrangement to form microfibrils and then fibrils. The staggered array of the molecules, in which each overlaps the other, gives a banded appearance to the collagen fibrils under the electron microscope. The fibrils aggregate further into fibers, which come together into densely packed bundles.

type I collagen. This molecule consists of three polypeptide chains (α chains), each coiled in a left-handed helix with approximately 100 amino acids, which give it a total molecular weight of approximately 340,000 daltons (Rich & Crick, 1961) (Fig. 4-2). Two of the peptide chains (called α-1 chains) are identical and one differs slightly (the α-2 chain). The three α chains are combined in a right-handed triple helix, which gives the collagen molecule a rod-like shape. The length of the molecule is approximately 280 nanometers (nm), and its diameter is approximately 1.5 nm.

Almost two thirds of the collagen molecule consists of three amino acids: glycine (33%), proline (15%), and hydroxyproline (15%) (Ramachandran, 1963). Every third amino acid in each α chain is glycine, and this repetitive sequence is essential for the proper formation of the triple helix. The small size of this amino acid allows the tight helical packing of the collagen molecule. Moreover, glycine enhances the stability of the molecule by forming hydrogen bonds among the three chains of the superhelix. Hydroxyproline and proline form hydrogen bonds, or hydrogen-bonded water bridges, within each chain. The intra- and interchain bonding, or cross-linking, between specific groups on the chains is essential to the stability of the molecule.

Cross-links are also formed between collagen molecules and are essential to aggregation at the fibril level. It is the cross-linked character of the collagen fibrils that gives strength to the tissues they compose and allows these tissues to function under mechanical stress. Within the fibrils, the molecules are apparently cross-linked by "head-to-tail" interactions (Fig. 4-1), but interfibrillar cross-linking of a more complex nature also may occur.

In newly formed collagen, the cross-links are relatively few and reducible; the collagen is soluble in neutral salt solutions and in acid solutions, and the cross-links are, for the most part, denatured by heat. As collagen ages, the total number of reducible cross-links decreases to a minimum as a large number of stable, nonreducible cross-links are formed. Mature collagen is not soluble in neutral salt solutions or in acid solutions, and it survives a higher denaturation temperature. (For a review of cross-linkage in collagen, see Viidik, Danielsen & Oxlund, 1982.)

A fibril is formed by the aggregation of several collagen molecules in a quarter-stagger structure. This structure, in which each molecule overlaps the other, is responsible for the repeating bands observed on the fibrils under the electron microscope (Fig. 4-2; see also Fig. 3-3). The quarter-stagger structure of collagen relates to the organization of collagen molecules into a stable, low-energetic biological unit based on a regular association of adjacent molecules' basic and acidic amino acids. By arranging adjacent collagen molecules in a quarter-stagger re-

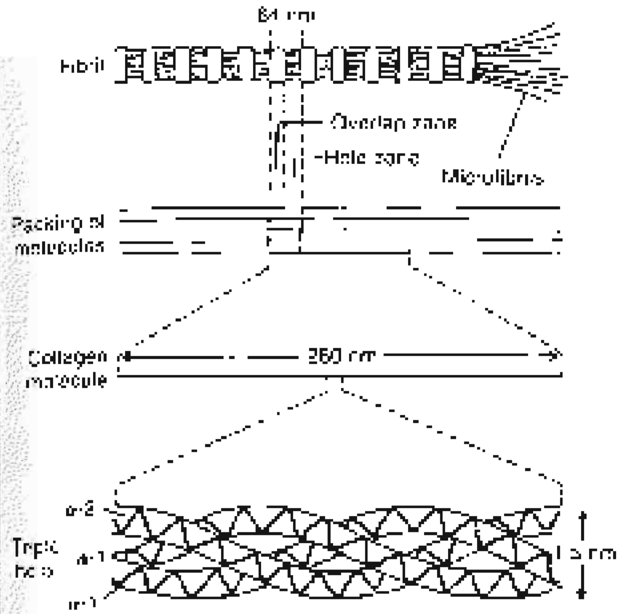


FIG. 4-2

Schematic drawing of collagen microstructure. The collagen molecule consists of three alpha chains in a triple helix (bottom). Several collagen molecules are aggregated into a staggered parallel array. This staggering, which creates hole zones and overlap zones, causes the cross-striation (banding pattern) visible in the collagen fibril under the electron microscope. *Adapted from Prockop, D.J., & Guzman, N.A. (1977). Collagen diseases and the biosynthesis of collagen. Hosp Pract, Dec, 61–68.*

positively charged amino acids are aligned. This stable structure will require a great amount of energy and force to separate its molecules, thus contributing to the strength of the structure. In this way organized collagen molecules (five) form units of microfibrils, subfibrils, and fibrils (Fig. 4-3) (Simon, 1994). The fibrils aggregate further to form collagen fibers, which are visible under the light microscope. These fibers, which range from 1 to 20 µm in diameter, do not branch and may be many centimeters long. They reflect a 64 nm periodicity of the fibrils and have a characteristic undulated form. The fibers aggregate further into bundles. Fibroblasts are aligned in rows between these bundles and are elongated along an axis in the direction of ligament or tendon function (Fig. 4-4).

The arrangement of the collagen fibers differs somewhat in the tendons and ligaments and is suited to the function of each structure. The fibers composing the tendons have an orderly, parallel arrangement, which equips the tendons to handle the high unidirectional (uniaxial) tensile loads to which they are subjected during activity (Fig. 4-4A). The ligaments generally sustain tensile loads in one predominant direction but may also bear smaller tensile loads in other directions; their fibers may not be completely parallel but are closely interlaced with one another (Fig. 4-4B). The specific orientation of the fiber bundles varies to some extent among the ligaments and is dependent on the function of the ligament (Amiel et al., 1984).

The metabolic turnover of collagen may be studied by tritium labeling of hydroxyproline or glycine and by autoradiographic methods. Studies in animals have shown that the half-life of collagen in mature animals is very long; the same collagen molecules may exist throughout the animal's adult life. However, in young animals and in physically altered (e.g., injured or immobilized) tissue, the turnover is accelerated. Rabbit studies have shown metabolic activity to be somewhat greater in ligaments than in tendons, probably because of different stress patterns (Amiel et al., 1984).

ELASTIN

The mechanical properties of tendons and ligaments are dependent not only on the architecture and properties of the collagen fibers but also on the proportion of elastin that these structures contain. The protein elastin is scarcely present in tendons and extremity ligaments, but in elastic ligaments such as the ligamentum flavum, the proportion of elastic fibers is substantial. Nachemson and Evans (1968) found a 2 to 1 ratio of elastic to collagen fibers in the ligamenta flava. These ligaments, which connect the laminae of adjacent vertebrae, appear to have a specialized role, which is to protect the spinal nerve roots from mechanical impingement, to pre-stress (preload) the motion segment (the functional unit of the spine), and to provide some intrinsic stability to the spine.

GROUND SUBSTANCE

The ground substance in ligaments and tendons consists of proteoglycans (PGs) (up to approximately 20% of the solids) along with structural glycoproteins, plasma proteins, and a variety of small molecules. The PG units, macromolecules composed of various sulfated polysaccharide chains (glycosaminoglycans) bonded to a core protein, bind to a long hyaluronic acid (HA) chain to form an extremely high molecular

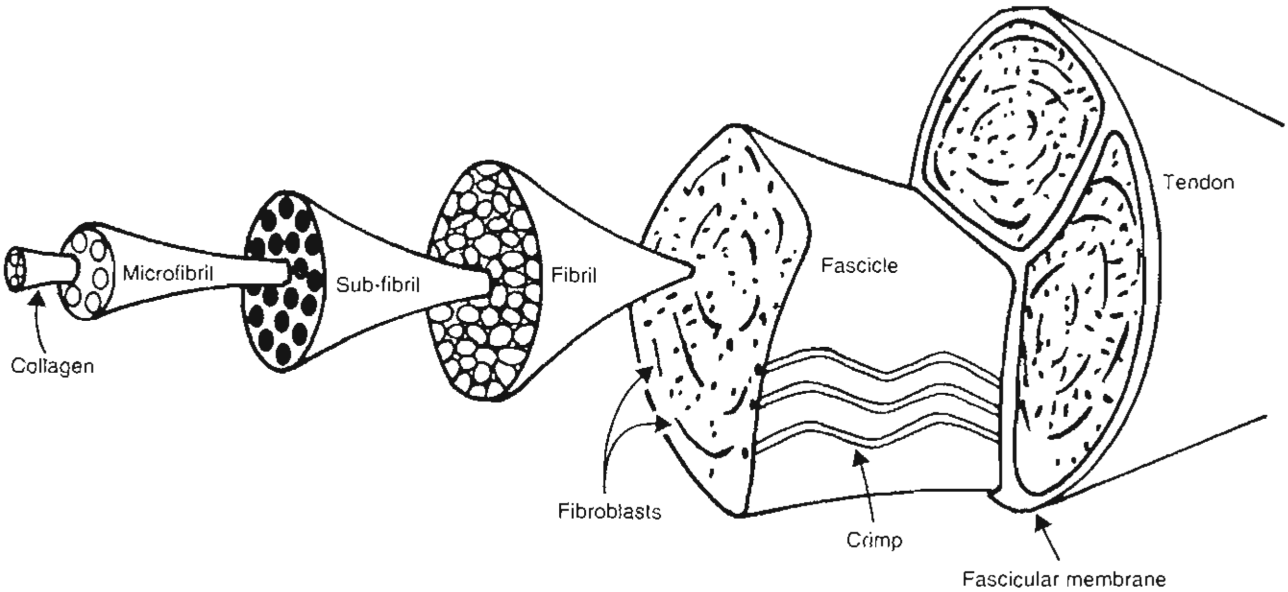


FIG. 4-3

Schematic representation of the microarchitecture of a tendon.

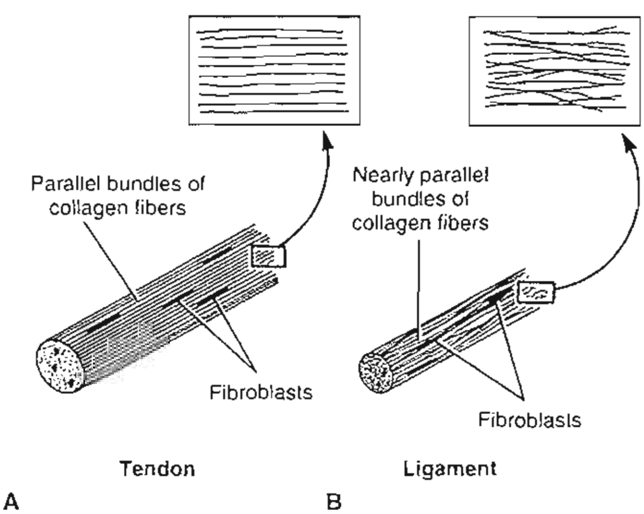


FIG. 4-4

Schematic diagram of the structural orientation of the fibers of tendon (A) and ligament (B); *insets* show longitudinal sections. In both structures the fibroblasts are elongated along an axis in the direction of function. *Adapted from Snell, R.S. (1984).* Clinical and Functional Histology for Medical Students. *Boston: Little, Brown.*

weight PG aggregate like that found in the ground substance of articular cartilage (see Fig. 3-6).

The PG aggregates bind most of the extracellular water of the ligament and tendon, making the matrix a highly structured gel-like material rather than an amorphous solution. Furthermore, by acting as a cement-like substance between the collagen microfibrils, they may help stabilize the collagenous skeleton of tendons and ligaments and contribute to the overall strength of these composite structures. Only a small number of these molecules exist in tendons, however, and their importance for its biomechanical properties has been questioned.

VASCULARIZATION

Tendons and ligaments have a limited vascularization, which affects directly their healing process and metabolic activity. Tendons receive their blood supply directly from vessels in the perimysium, the periosteal insertion, and the surrounding tissue via vessels in the paratenon or mesotenon. Tendons surrounded by paratenon have been referred to as vascular tendons, and those surrounded by a tendon sheath as avascular tendons. In tendons surrounded by a paratenon, vessels enter from many points on the periphery and anastomose with a longitudinal system of capillaries (Fig. 4-5).

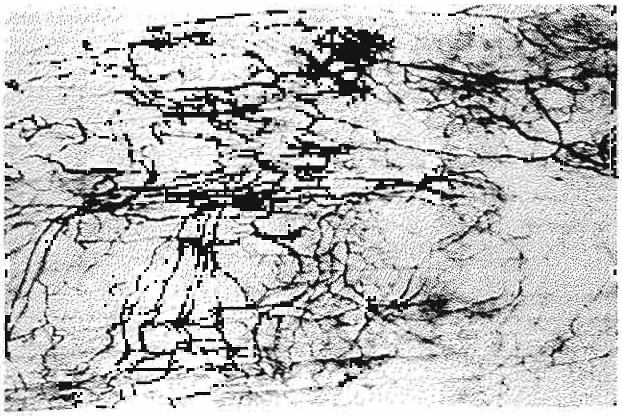


FIG. 4-5

India ink-injected (Spalteholz technique) into the calcaneal tendon of a rabbit, illustrating the vasculature of a paratenon-covered tendon. Vessels enter from many points on the periphery and anastomose with a longitudinal system of capillaries. *Reprinted with permission from Woo, S.L-Y., An, K.-N., Arnoczky, D.V.M., et al. (1994). Anatomy, biology, and biomechanics of the tendon, ligament, and meniscus. In S.R. Simon (Ed.): Orthopaedic Basic Science (p. 53). Rosemont, IL: AAOS.*

The vascular pattern for tendons surrounded by a tendon sheath is different. Here the mesotenons are reduced to vincula (Fig. 4-6). This avascular region led a variety of researchers to propose a dual pathway for tendon nutrition: a vascular pathway and, for the avascular regions, a synovial (diffusion) pathway. The concept of diffusional nutrition is of primary clinical significance in that it implies that tendon healing and repair can occur in the absence of adhesions (i.e., a blood supply). Conversely, ligaments in comparison with surrounding tissue appear to be hypovascular. However, histological studies reveal that throughout the ligament substance there is a uniform microvascularity, which originates from the insertion sites of the ligament. Despite the small size and limited blood flow of this vascular system, it is of primary importance in the maintenance of the ligament. Specifically, by providing nutrition for the cellular population, this vascular system maintains the continued process of matrix synthesis and repair. In its absence, damage from normal activities accumulates (fatigue) and the ligament is at risk for rupture (Woo et al., 1994).

Ligaments and tendons have been shown in both human and animal studies to have a variety of specialized nerve endings and mechanoreceptors. They play an important role in proprioception and nociception, which are directly related to the functionality of joints.

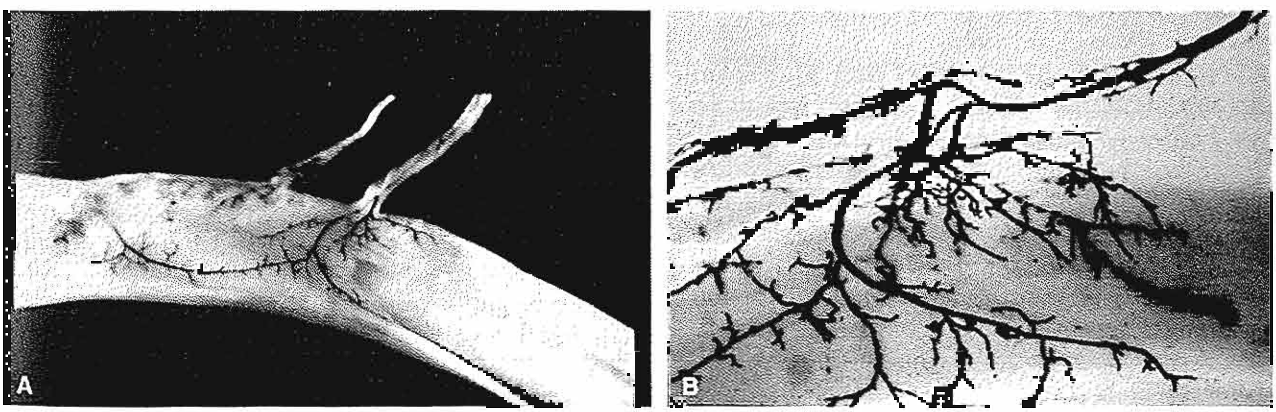


FIG. 4-6

A, India ink-injected specimen illustrating the vascular supply of the flexor digitorum profundus in a human through the vinculum longus. B, Close-up specimen (Spalteholz technique) showing the extent of the blood supply from the vinculum longus. The vessels in the vinculum divide into the dorsal, proximal, and distal branches, giving off vascular loops into the tendon substance. *Reprinted with permission from Woo, S.L-Y., An, K-N., Arnoczky, D.V.M., et al. (1994). Anatomy, biology and biomechanics of the tendon, ligament, and meniscus. In S.R. Simon (Ed.): Orthopaedic Basic Science (p. 52). Rosemont, IL: AAOS.*

OUTER STRUCTURE AND INSERTION INTO BONE

Certain similarities are found in the outer structure of tendons and ligaments, but there are also important differences related to function. Both tendons and ligaments are surrounded by a loose areolar connective tissue. In ligaments, this tissue has no specific name, but in tendons it is referred to as the paratenon. More structured than the connective tissue surrounding the ligaments, the paratenon forms a sheath that protects the tendon and enhances gliding. In some tendons, such as the flexor tendons of the digits, the sheath runs the length of the tendons, and in others the sheath is found only at the point where the tendon bends in concert with a joint.

In locations where the tendons are subjected to particularly high friction forces (e.g., in the palm, in the digits, and at the level of the wrist joint), a parietal synovial layer is found just beneath the paratenon; this synovium-like membrane, called the epitenon, surrounds several fiber bundles. The synovial fluid produced by the synovial cells of the epitenon facilitates gliding of the tendon. In locations where tendons are subjected to lower friction forces, they are surrounded by the paratenon only.

Each fiber bundle is bound together by the endotenon (Fig. 4-1), which continues at the musculotendinous junction into the perimysium. At the tendo-osseous junction, the collagen fibers of the endotenon continue into the bone as Sharpey's perforating fibers and become continuous with the periosteum (Woo et al., 1988).

The structure of the insertions into bone is similar in ligaments and tendons and consists of four zones; Figure 4-7 illustrates these zones in a tendon. At the end of the tendon (zone 1), the collagen fibers intermesh with fibrocartilage (zone 2). This fibrocartilage gradually becomes mineralized fibrocartilage (zone 3) and then merges into cortical bone (zone 4). The change from more tendinous to more bony material produces a gradual alteration in the mechanical properties of the tissue (i.e., increased stiffness), which results in a decreased stress concentration effect at the insertion of the tendon into the stiffer bone (Cooper & Misol, 1970).

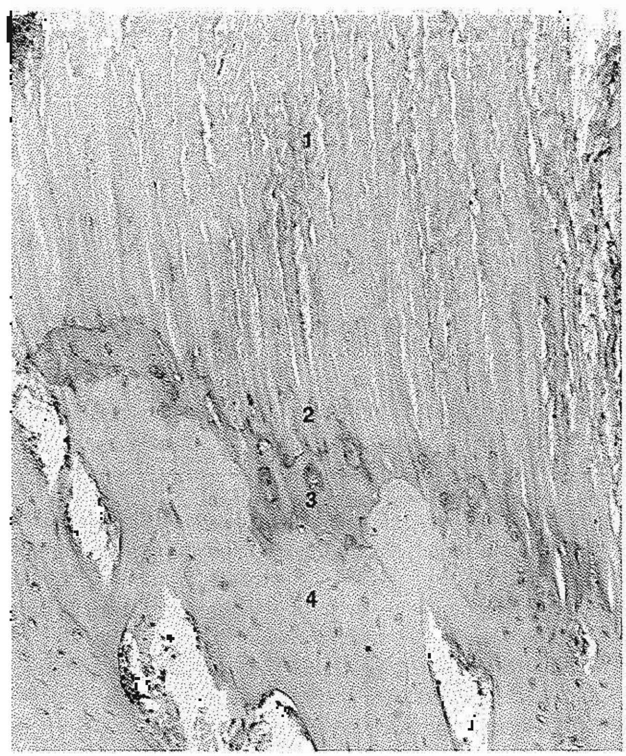


FIG. 4-7

Electron micrograph of a patellar tendon insertion from a dog, showing four zones (×25,000). Zone 1, parallel collagen fibers; zone 2, unmineralized fibrocartilage; zone 3, mineralized fibrocartilage; zone 4, cortical bone. The ligament-bone junction (not pictured) has a similar appearance. *Reprinted with permission from Cooper, R.R., & Misol, S. (1970). Tendon and ligament insertion: A light and electron microscopic study. J Bone Joint Surg, 52A, 1.*

Mechanical Behavior of Tendons and Ligaments

Tendons and ligaments are viscoelastic structures with unique mechanical properties. Tendons are strong enough to sustain the high tensile forces that result from muscle contraction during joint motion yet are sufficiently flexible to angulate around bone surfaces and to deflect beneath retinacula to change the final direction of muscle pull. The ligaments are pliant and flexible, allowing natural movements of the bones to which they attach, but are strong and inextensible so as to offer suitable resistance to applied forces.

Analysis of the mechanical behavior of tendons and ligaments provides important information for the understanding of injury mechanisms. Both structures sustain chiefly tensile loads during normal and excessive loading. When loading leads to injury, the degree of damage is affected by the rate of impact as well as the amount of load.

BIOMECHANICAL PROPERTIES

One means of analyzing the biomechanical properties of tendons and ligaments is to subject specimens to tensile deformation using a constant rate of elongation. The tissue is elongated until it ruptures, and the resulting force, or load (P), is plotted. The resulting load-elongation curve has several regions that characterize the behavior of the tissue (Fig. 4-8).

The first region of the load-elongation curve is called the "toe" region. The elongation reflected in this region is believed to be the result of a change in the wavy pattern of the relaxed collagen fibers. In this region, the tissue stretches easily, without much force, and the collagen fibers become straight and lose their wavy appearance as the loading progresses (Hirsch, 1974; Woo et al., 1994) (Fig. 4-9, A & B). Some data suggest, however, that this elongation may be caused mainly by interfibrillar sliding and shear of the interfibrillar gel (ground substance) (for review see Viidik, Danielson, & Oxlund, 1982).

As loading continues, the stiffness of the tissue increases and progressively greater force is required to produce equivalent amounts of elongation. The elongation is often expressed as strain (ε), which is the deformation of the tissue calculated as a percentage of the original length of the specimen. If

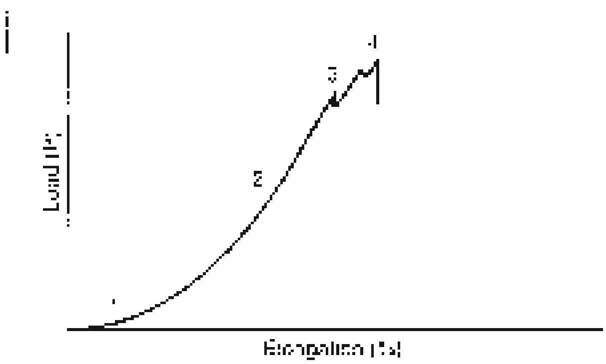


FIG. 4-8

Load-elongation curve for rabbit tendon tested to failure in tension. The numbers indicate the four characteristic regions of the curve. (1) Primary or "toe" region, in which the tissue elongated with a small increase in load as the wavy collagen fibers straightened out. (2) Secondary or "linear" region, in which the fibers straightened out and the stiffness of the specimen increased rapidly. Deformation of the tissue began and had a more or less linear relationship with load. (3) End of secondary region. The load value at this point is designated as P_{lin}. Progressive failure of the collagen fibers took place after P_{lin} was reached, and small force reductions (dips) occurred in the curve. (4) Maximum load (P_{max}) reflecting the ultimate tensile strength of the tissue. Complete failure occurred rapidly, and the specimen lost its ability to support loads. *Adapted from Carlstedt, C.A. (1987). Mechanical and chemical factors in tendon healing. Effects of indomethacin and surgery in the rabbit. Acta Orthop Scand Suppl, 224.*

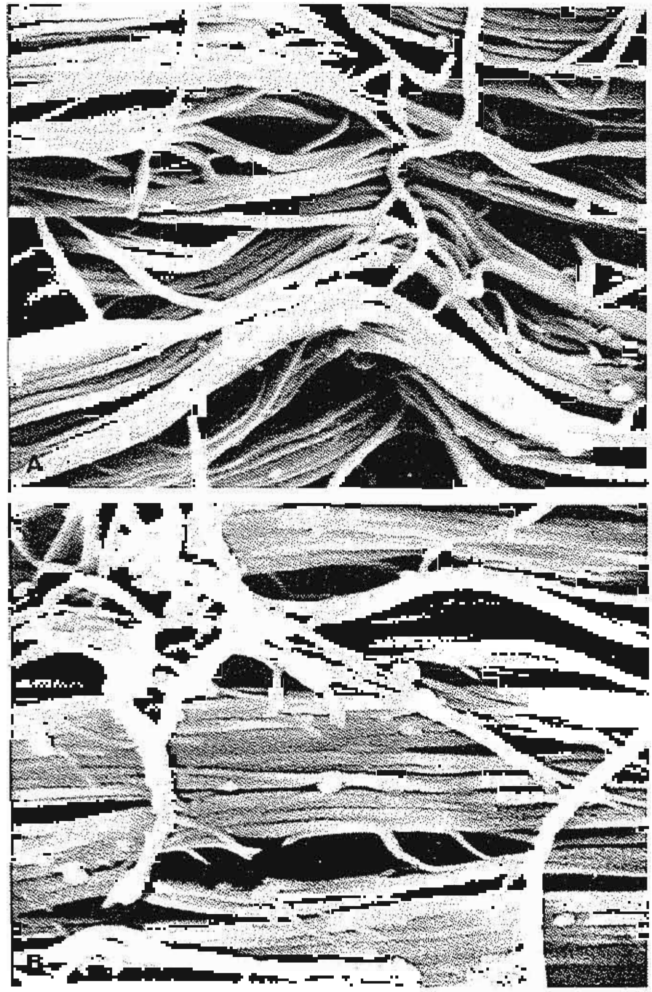


FIG. 4-9

Scanning electron micrographs of unloaded (relaxed) and loaded collagen fibers of human knee ligaments (×10,000). A. The unloaded collagen fibers have a wavy configuration. B. The collagen fibers have straightened out under load. *Reprinted with permission from Kennedy, J.C., Hawkins, R.J., Willis, R.B., et al. (1976). Tension studies of human knee ligaments: Yield point, ultimate failure, and disruption of the cruciate and tibial collateral ligaments. J Bone Joint Surg, 58A, 350.*

strains are increased (strain values of between 1.5 and 4% [Viidik, 1973]), a linear region will follow the toe region. This sudden increase in slope represents the second region in the diagram and corresponds to the response of the tissue to further elongation (Diamant et al., 1972).

Following the linear region, at large strains the stress-strain curve can end abruptly or curve downward as a result of irreversible changes (failure) (Woo et al., 1994). Where the curve levels off toward the strain axis, the load value is designated as P_{lin}. The point at which this value is reached is the yield point for the tissue. The energy uptake to P_{lin} is represented by the area under the curve up to the end of the linear region.

When the linear region is surpassed, major failure of fiber bundles occurs in an unpredictable manner. With the attainment of maximum load that reflects the ultimate tensile strength of the specimen, complete failure occurs rapidly, and the load-supporting ability of the tendon or ligament is substantially reduced.

The modulus of elasticity for tendons and ligaments has been determined in several investigations (Fung, 1967, 1972; Viidik, 1968). This parameter is based on a linear relationship between load and deformation (elongation), or stress and strain; that is, the stress (force per unit area) is proportional to the strain:

$$E = \sigma/\varepsilon$$

where E = modulus of elasticity

$$\sigma = \text{stress}$$
$$\varepsilon = \text{strain}$$

In the toe portion of the load-elongation curve (or stress-strain curve), the modulus of elasticity is not constant but increases gradually. The modulus stabilizes in the fairly linear secondary region of the curve.

The load-elongation curve depicted in Figure 4-8 generally applies to tendons and extremity ligaments. The curve for the ligamentum flavum, with its high proportion of elastic fibers, is entirely different (Fig. 4-10). In tensile testing of a human ligamentum flavum, elongation of the specimen reached 50% before the stiffness increased appreciably. Beyond this point, the stiffness increased greatly with additional loading and the ligament failed abruptly (reached P_{max}), with little further deformation (Nachemson & Evans, 1968).

The proportion of elastic proteins in ligaments and capsules is extremely important for the small elastic deformation that they endure under tensile strain and the storage and loss of energy. During loading and unloading of a ligament between two limits of elongation, the elastic fibers allow the material to return to its original shape and size after being deformed. Meanwhile, part of the energy spent is stored; what is left will represent the energy loss during the cycle and is called hysteresis. The area enclosed by the loop represents the energy loss (Fig. 4-11).

PHYSIOLOGICAL LOADING OF TENDONS AND LIGAMENTS

The ultimate tensile strength (P_{max}) of ligaments and tendons is of limited interest from a functional standpoint because under normal physiological conditions in vivo these structures are subjected to a stress magnitude that is only approximately one third of this value. The upper limit for physiological strain in tendons and ligaments (when running and jumping, for example) is from 2 to 5% (Fung, 1981).

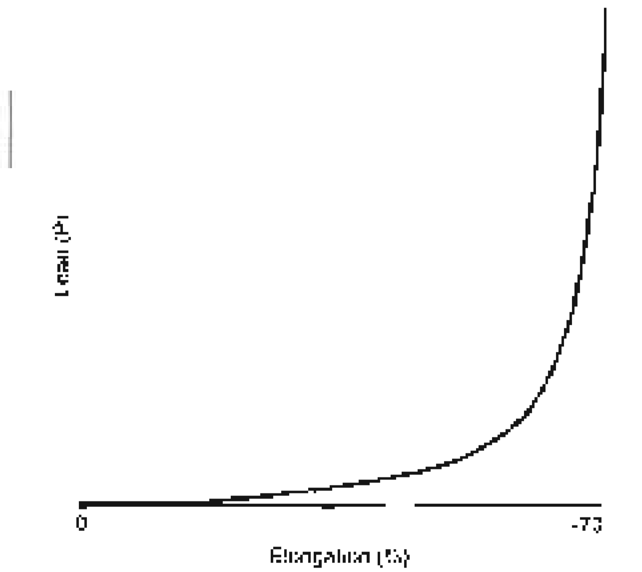


FIG. 4-10

Load-elongation curve for a human ligamentum flavum (60 to 70% elastic fibers) tested in tension to failure. At 70% elongation the ligament exhibited a great increase in stiffness with additional loading and failed abruptly without further deformation. *Adapted from Nachemson, A.L., & Evans, J.H. (1968). Some mechanical properties of the third human lumbar interlaminar ligament (ligamentum flavum). J Biomech, 1, 211–220.*

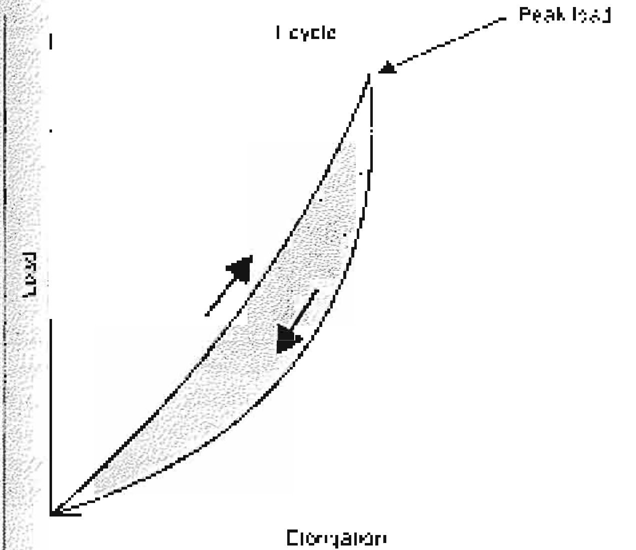


FIG. 4-11

Typical loading (top) and unloading (bottom) curves from tensile testing of knee ligaments. The two nonlinear curves form a hysteresis loop. The area between the curves, called the area of hysteresis, represents the energy losses within the tissue.

Few studies of loading of tendons or ligaments in vivo have been performed. Kear and Smith (1975), using the strain gauge method, measured the maximal strain in the lateral digital extensor tendons of sheep. The strain reached 2.6% while the sheep were trotting rapidly and decreased when the trotting speed decreased. This maximal strain occurred for only 0.1 second during each stride. The maximal load imposed on the entire tendon was approximately 45 newtons (N). These results suggest that during normal activity a tendon in vivo is subjected to less than one-fourth of its ultimate stress.

VISCOELASTIC BEHAVIOR (RATE DEPENDENCY) IN TENDONS AND LIGAMENTS

Ligaments and tendons exhibit viscoelastic, or rate-dependent (time-dependent), behavior under loading; their mechanical properties change with different rates of loading. When ligament and tendon specimens are subjected to increased strain rates (loading rates), the linear portion of the stress-strain curve becomes steeper, indicating greater stiffness of the tissue at higher strain rates. With higher strain rates, ligaments and tendons in isolation store more energy, require more force to rupture, and undergo greater elongation (Kennedy, Hawkins, Willis, & Danylchuk, 1976).

During cyclic testing of ligaments and tendons, where loads are applied and released at specific intervals, the stress-strain curve is displaced to the right along the deformation (strain) axis with each loading cycle, revealing the presence of a nonelastic (plastic) component. The amount of permanent (nonrecoverable) deformation is progressively greater with every loading cycle. As cyclic loading progresses, the specimen also shows an increase in elastic stiffness as a result of plastic deformation (molecular displacement). Microfailure can occur within the physiological range if frequent loading is imposed on an already damaged structure where the stiffness has decreased.

Two standard tests that reveal the viscoelasticity of ligaments and tendons are the stress-relaxation test and the creep test (Fig. 4-12). During a stress-relaxation test, loading is halted safely below the linear region of the stress-strain curve and the strain is kept constant over an extended period. The stress decreases rapidly at first and then more slowly. When the stress-relaxation test is repeated cyclically, the decrease in stress gradually becomes less pronounced.

During a creep test, loading is halted safely below the linear region of the stress-strain curve and the stress is kept constant over an extended period. The strain increases relatively quickly at first and then more and more slowly. When this test is performed cyclically, the increase in strain gradually becomes less pronounced.

The clinical application of a constant low load to the soft tissues over a prolonged period, which takes advantage of the creep response, is a useful treatment for several types of deformities. One example is the manipulation of a child's clubfoot by subjecting it to constant loads by means of a plaster cast. Another example is the treatment of idiopathic scoliosis with a brace, whereby constant loads are applied to the spinal area to elongate the soft tissues surrounding the abnormally curved spine.

More complex viscoelastic behavior is observed in the entire bone-ligament-bone complex. Anterior cruciate ligaments (ACLs) in knee specimens taken from 30 primates were tested in tension to failure at a slow and a fast loading rate (Noyes et al., 1974). At the slow loading rate (60 seconds), much slower than that of an injury mechanism in vivo, the bony insertion of the ligament was the weakest component of the bone-ligament-bone complex, and a tibial spine

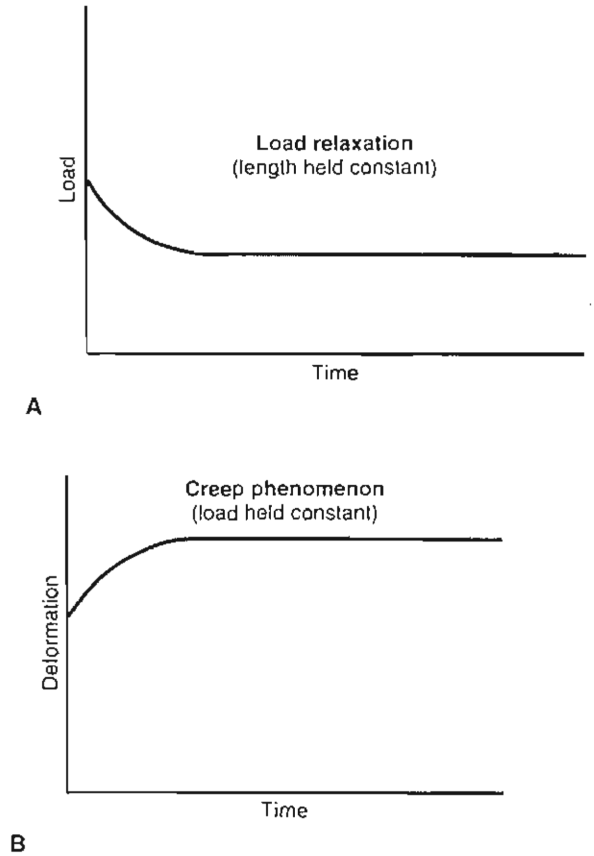


FIG. 4-12

The viscoelasticity (rate dependency, or time dependency) of ligaments and tendons can be demonstrated by two standard tests: the load-relaxation test and the creep test. **A,** Load relaxation is demonstrated when the loading of a specimen is halted safely below the linear region of the load-deformation curve and the specimen is maintained at a constant length over an extended period (i.e., the amount of elongation is constant). The load decreases rapidly at first (i.e., during the first 6 to 8 hours of loading) and then gradually more slowly, but the phenomenon may continue at a low rate for months. **B,** The creep response takes place when loading of a specimen is halted safely below the linear region of the load-deformation curve and the amount of load remains constant over an extended period. The deformation increases relatively quickly at first (within the first 6 to 8 hours of loading) but then progressively more slowly, continuing at a low rate for months.

avulsion was produced. At the fast loading rate (0.6 seconds), which simulated an injury mechanism in vivo, the ligament was the weakest component in two thirds of the specimens tested. At the slower rate, the load to failure decreased by 20%, and 30% less energy was stored to failure, but the stiffness of the bone-ligament-bone complex was nearly the same. These results suggest that as the loading rate is increased, bone shows a greater increase in strength than does ligament.

Ligament Failure and Tendon Injury Mechanisms

Injury mechanisms are similar for ligaments and tendons, therefore the following description of ligament injury and failure is generally applicable to tendons. When a ligament in vivo is subjected to loading that exceeds the physiological range, microfailure takes place even before the yield point (P_{lin}) is reached. When P_{lin} is exceeded, the ligament begins to undergo gross failure and simultaneously the joint begins to displace abnormally. This displacement can also result in damage to the sur-

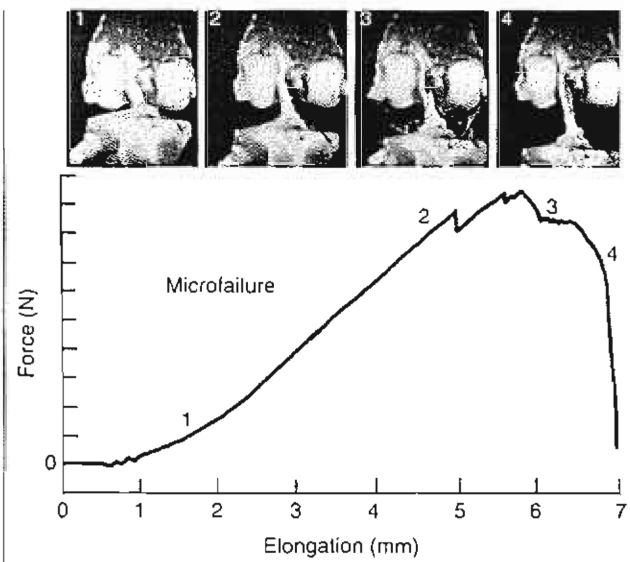


FIG. 4-13

Progressive failure of the anterior cruciate ligament from a cadaver knee tested in tension to failure at a physiological strain rate. The joint was displaced 7 mm before the ligament failed completely. The force-elongation curve generated during this experiment is correlated with various degrees of joint displacement recorded photographically; photos correspond to similarly numbered points on the curve. *Reprinted with permission from Noyes, F.R., and Grood, E.S. (1976). The strength of the anterior cruciate ligament in humans and Rhesus monkeys. Age-related and species-related changes.* J Bone Joint Surg. 58A, *1074–1082.*

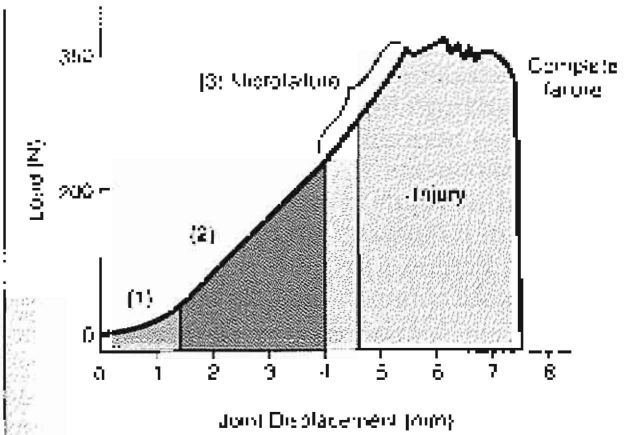


FIG. 4-14

The curve produced during tensile testing of a human anterior cruciate ligament in vitro (20) (see Fig. 4-13) has been converted to a load-displacement curve and divided into three regions correlating with clinical findings: (1) the load imposed on the anterior cruciate ligament during the anterior drawer test, (2) that placed on the ligament during physiological activity, and (3) that imposed on the ligament from partial injury to complete rupture. It should be noted that the divisions shown here represent a generalization. Microfailure is shown to begin toward the end of the physiological loading region, but it may take place well before this point in any given ligament.

rounding structures, such as the joint capsule, the adjacent ligaments, and the blood vessels that supply these structures.

Noyes (1977) demonstrated the progressive failure of the ACL and displacement of the tibiofemoral joint by applying a clinical test, the anterior drawer test, to a cadaver knee up to the point of ACL failure (Fig. 4-13). At maximum load, the joint had displaced several millimeters. The ligament was still in continuity even though it had undergone extensive macro- and microfailure and extensive elongation. In Figure 4-13, the force-elongation curve generated during the experiment, indicating where microfailure of the ligament began, is compared with various stages of joint displacement recorded photographically.

Correlation of the results of this test in vitro with clinical findings sheds light on the microevents that take place in the ACL during normal daily activity and during injuries of various degrees of severity. In Figure 4-14, the curve for the experimental study on cadaver knees that was presented in Figure 4-13 has been converted into a load-displacement curve and divided into three regions, corresponding respectively to (1) the load placed on the ACL during tests of knee joint stability performed clinically, (2) the load placed on this ligament during physiological activity, and (3) that imposed on the ligament during injury from the beginning of microfailure to complete rupture. Microfailure begins even before the physiological loading range is exceeded and can occur throughout the physiological range in any given ligament. In fact, under experimental testing, the ultimate tensile loads—or the load at failure for human ACL—was between 340 and 390 N (Case Study 4-1).

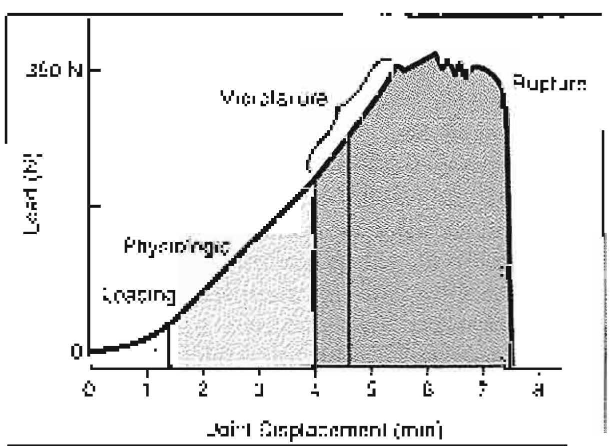
CASE STUDY 4-1

ACL Failure: Failure of the ACL Associated With High Strain and Stress

A 25-year-old male occasional soccer player injured his ACL as a result of an abnormal torque in rotation of the knee. The player locked his foot on the ground and pivoted on his lower limb to produce a high rotational torque on the knee, which increased tensile loads on the ACL.

The first region of the load-displacement curve shows a normal physiological loading response. In the microfailure region, the increase in strain deformation leads to high internal stress and finally a complete rupture. Experimental testing in vitro in human ACL yielded a point of failure between 340 and 390 N.

The knee with the ACL injury will increase intra-articular joint motion, producing abnormally high stresses on other joint structures such as cartilage, which can lead to osteoarthritis. A deficiency in joint stability that results from ACL impairment will increase the likelihood of experiencing the "giving way" sensation or functional instability, thus affecting activities of daily living such as gait, jogging, and squatting (Case Study Fig. 4-1-1).



Case Study Figure 4-1-1.

Ligament injuries are categorized clinically in three ways according to degree of severity. Injuries in the first category produce negligible clinical symptoms. Some pain is felt, but no joint instability can be detected clinically, even though microfailure of the collagen fibers may have occurred.

Injuries in the second category produce severe pain, and some joint instability can be detected clinically. Progressive failure of the collagen fibers has taken place, resulting in partial ligament rupture. The strength and stiffness of the ligament may have decreased by 50% or more, mainly because the amount of undamaged tissue has been reduced. The joint instability produced by the partial rupture of a ligament is often masked by muscle activity, and thus the clinical test for joint stability is usually performed with the patient under anesthesia.

Injuries in the third category produce severe pain during the course of trauma with less pain after injury. Clinically, the joint is found to be completely unstable. Most collagen fibers have ruptured, but a few may still be intact, giving the ligament the appearance of continuity even though it is unable to support any loads.

Loading of a joint that is unstable as a result of ligament or joint capsule rupture produces abnormally high stresses on the articular cartilage. This abnormal loading of the articular cartilage in the knee has been correlated with early osteoarthritis in humans and in animals.

Although injury mechanisms are generally comparable in ligaments and tendons, two additional factors become important in tendons because of their attachment to muscles: the amount of force produced by contraction of the muscle to which the tendon is attached and the cross-sectional area of the tendon in relation to that of its muscle. A tendon is subjected to increasing stress as its muscle contracts (see Fig. 6-10). When the muscle is maximally contracted, the tensile stress on the tendon reaches high levels. This stress can be increased further if rapid eccentric contraction of the muscle takes place (for example, rapid dorsiflexion of the ankle, which does not allow for reflex relaxation of the gastrocnemius and soleus muscles, increases the tension on the Achilles tendon). The load imposed on the tendon under these circumstances may exceed the yield point, causing Achilles tendon rupture (Case Study 4-2).

The strength of a muscle depends on its physiological cross-sectional area. The larger the cross-sectional area of the muscle, the higher the magnitude of the force produced by the contraction and thus the greater the tensile loads transmitted through the tendon. Similarly, the larger the cross-sectional area of the tendon, the greater the loads it can bear. Although the maximal stress to failure for a muscle has been difficult to compute accurately, such measurements have shown that the tensile strength of a healthy tendon may be more than twice that of its muscle (Elliott, 1967). This finding is supported clinically by the fact that muscle ruptures are more common than are ruptures through a tendon.

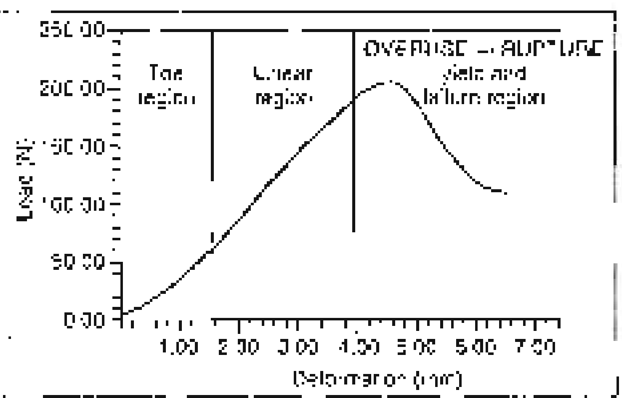
Large muscles usually have tendons with large cross-sectional areas. Examples are the quadriceps

CASE STUDY 4-2

Tendon Injuries: Achilles Tendon Injuries in Runners, Which Result From a High Strain Rate

A middle-aged male marathoner engaged in a strenuous running activity experienced pain and a popping sensation in his posterior calf. An overuse injury is diagnosed.

The first region of the load-deformation curve shows a normal physiological toe-loading response. In the linear region, high load is producing a higher deformation within the tendon structure. When the Achilles tendon is subjected to higher strain rates during frequent loading cycles and insufficient time is allowed for the healing process, the result is an overuse injury. Histological studies of these injuries reveal a pathological pattern described as "angiofibrotic hyperplasia," which suggests a degenerative process. This failure in tendon remodeling frequently occurs before the abrupt rupture of the tendon. Relative avascularity, inflammatory disease, and other local factors also contribute to midsubstance ruptures (Case Study 4-2-1).



Case Study Figure 4-2-1.

muscle with its patellar tendon and the triceps surae muscle with its Achilles tendon. Some small muscles have tendons with large cross-sectional areas, such as the plantaris, which is a tiny muscle with a large tendon.

Factors That Affect the Biomechanical Properties of Tendons and Ligaments

Numerous factors affect the biomechanical properties of tendons and ligaments. The most common are aging, pregnancy, mobilization and immobilization, diabetes, steroids, NSAID use, and hemodialysis. The biomechanical properties of grafts are also discussed because reconstruction, particularly of the anterior and posterior knee ligaments, is common.

MATURATION AND AGING

The physical properties of collagen and of the tissues it composes are closely associated with the number and quality of the cross-links within and between the collagen molecules. During maturation (up to 20 years of age), the number and quality of cross-links increases, resulting in increased tensile strength of the tendon and ligament (Viidik, Danielsen, & Oxlund, 1982). An increase in collagen fibril diameter is also observed (Parry et al., 1978) with high variability in size (range 20–180 nm) (Strocchi et al., 1996) noted in the young (<20 years). The diameter in adults (20–60 years) and in the elderly (>60 years) decreases remarkably (120 and 110 nm) but with a more even distribution. Strocchi et al. (1996) investigated age-related changes in human ACL collagen fibrils and reports an increase of fibril concentration from 68 fibrils/mm² in the young to 140 fibrils/mm² in the elderly. However, Amiel (1991) reports that the water content and the collagen concentration decreases significantly in the medial collateral ligament of 2-, 12-, and 39-month-old rabbits.

After maturation, as aging progresses, collagen reaches a plateau with respect to its mechanical properties, after which the tensile strength and stiffness of the tissue begin to decrease. This may be the result of an increase in small collagen fibrils. Conversely, when the ACL of younger donors (average age 30 years) was compared with the ACL of older donors (average age 64.7 years), the material properties (strain, elastic modules, and maximum stress) did not differ significantly (Kasperczyk et al., 1991). This may be due to the fact that only the ACL was taken from donors in whom no vascular or cardiopulmonary disease and no osteoarthritis of the knee was found on autopsy.

PREGNANCY AND THE POSTPARTUM PERIOD

A common clinical observation is the increased laxity of the tendons and ligaments in the pubic area during later stages of pregnancy and the postpartum period. This observation has been confirmed in animal studies. Rundgren (1974) found that the tensile strength of the tendons and the pubic symphysis in rats decreased at the end of pregnancy and during the postpartum period. Stiffness of these structures decreased in the early postpartum period but was later restored.

MOBILIZATION AND IMMOBILIZATION

Living tissues are dynamic and change their mechanical properties in response to stress, which leads to functional adaptation and optimal operation of the tissue.

Like bone, ligament and tendon appear to remodel in response to the mechanical demands placed on them; they become stronger and stiffer when subjected to increased stress and weaker and less stiff when the stress is reduced (Noyes et al., 1977a).

Physical training has been found to increase the tensile strength of tendons and of the ligament-bone interface (Woo et al., 1981). Tipton and coworkers (1970) compared the strength and stiffness of medial collateral ligaments from dogs that were exercised strenuously for 6 weeks with the values for ligaments from a control group of animals. The ligaments of the exercised dogs were stronger and stiffer than those of the control dogs, and the collagen fiber bundles had larger diameters.

Immobilization has been found to decrease the tensile strength of ligaments (Newton et al., 1995; Walsh et al., 1993). Noyes (1977a) demonstrated a reduction in the mechanical properties of the bone-ligament-bone complex in knees of primates immobilized in body casts for 8 weeks. When tested in tension to failure, the ACLs from these animals showed a 39% decrease in maximum load to failure and a 32% decrease in energy stored to failure compared with ligaments from a control group of animals (Fig. 4-13A). The immobilized ligaments also displayed more elongation and were significantly less stiff than the control specimens (Fig. 4-15B).

Amiel and coworkers (1982) showed a similar decrease in the strength and stiffness of lateral collateral ligaments in rabbits immobilized for 9 weeks. As the cross-sectional area of the specimens did not change significantly, the degeneration of mechanical properties was attributed to changes in the ligament substance itself. The tissue metabolism was noted to increase, leading to proportionally more immature collagen with a decrease in the amount and quality of the cross-links between collagen molecules. Newton et al. (1995) reported that the cross-sectional area of ligaments in immobilized rabbit knees was 74% of the control value.

In Noyes' (1977a) experiment, assessment of the effects of a reconditioning program initiated directly after the 8-week immobilization period demonstrated that considerable time was needed for the immobilized ligaments to regain their former strength and stiffness. After 5 months, the reconditioned ligaments still showed considerably less stiffness and 20% less strength than did ligaments from control animals. At 12 months, the reconditioned ligaments had strength and stiffness values comparable to those of control group ligaments (Fig. 4-15A). Woo et al. (1987) found that the stress-strain characteristics after remobilization return to normal, but that the energy-absorbing capabilities of the bone-ligament complex improved but did not return to normal.

DIABETES MELLITUS

The term diabetes refers to disorders characterized by excessive urine excretion. Diabetes mellitus is a metabolic disorder in which the ability to oxidize carbohydrates is more or less completely lost. This is usually caused by pancreas insufficiency and a disturbance of the normal insulin mechanism, resulting in hyperglycemia, glycosuria, and polyuria. Diabetes mellitus is known to cause musculoskeletal disorders. Diabetics compared with nondiabetics show higher rates of tendon contracture (29 vs. 9%), tenosynovitis (59 vs. 7%), joint stiffness (40 vs. 9%), and capsulitis (16 vs. 1%). Diabetes also causes osteoporosis (Carvallo et al., 1991; Lancaster et al., 1994).

Duquette (1996) examined the effects of diabetes on the properties of the collateral knee ligament in rats. The tissue elastic properties did not differ between the diabetic and the control group. The vis-

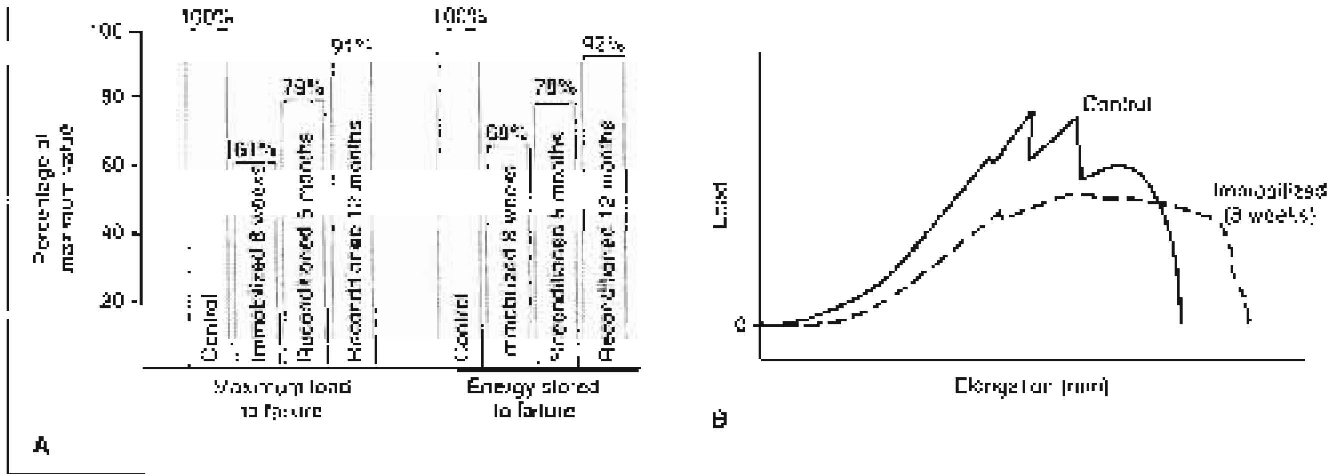


FIG. 4-15

A. Maximal load to failure and energy stored to failure for primate anterior cruciate ligaments tested in tension to failure. Values are shown as a percentage of control values for three groups of experimental animals: (1) those immobilized in body casts for 8 weeks, (2) those immobilized for 8 weeks and given a reconditioning program for 5 months, and (3) those immobilized for 8 weeks and given a reconditioning program for 12 months. B. Compared with controls, ligaments immobilized for 8 weeks were significantly less stiff (as indicated by the slope of the curve) and underwent greater elongation. *Adapted from Noyes, F.R. (1977a). Functional properties of knee ligaments and alterations induced by immobilization. Clin Orthop, 123, 210–242.*

cross component of the tissue response, however, was increased in the hyperglycemic group. Insulin therapy seems to lessen such alterations. Lancaster et al. (1994) examined the changes in the mechanical properties of the patellar tendon in diabetic dogs. The results showed the stiffness of the canine patellar tendon-tibia complex in a physiological range of loading was 13% greater than in the control group. There was no difference in the strength of the tendon between the groups, but the mode of failure was different. In the control group, failure was caused by substance and avulsion failure, whereas failure of the tendon in the diabetic group was caused by tensile fractures of the patella (Lancaster et al., 1994).

STEROIDS

Corticosteroids, when applied immediately after injury, may cause significant impairment of the biomechanical and histological properties in ligaments. Corticosteroids also are known to inhibit collagen synthesis in vitro (Walsh et al., 1995). Wiggins et al. (1994) described these results in rabbits and implied that an acute injured ligament treated with corticosteroid injections may not withstand the mechanical loads of an early vigorous rehabilitation. Noyes et al. (1977b) reported decreased ligament stiffness, failure load, and energy absorption in monkey ligaments after injection of long-acting corticosteroids. These findings were time- and dosage-dependent. After application of a dosage that was approximately 10 times an equivalent human dose, only minimal changes were found after 6 weeks, but after 15 weeks the maximum failure load (20%), energy absorption prior to failure (11%), and linear stiffness (10%) decreased significantly. After application of a dosage equivalent to the human dose, the maximum failure load (9%) and the energy absorption (8%) decreased significantly.

Campbell et al. (1996), however, showed that a single injection of long-acting corticosteroid does not cause histological differences in rats with acute injured ligaments as compared with rats with acute ligament injury and no injection of corticosteroids. Mechanical testing showed no significant difference in ultimate load or ultimate stress in the two groups. Oxlund et al. (1980) reports that local injections of corticosteroids every 3 days for 24 days increase the tensile strength and maximum load stiffness of muscle tendons but decrease the strength of the bone attachments of ligaments.

Laboratory investigations established the presence of estrogen receptors in human ACLs. Liu et al. (1997) reports that physiological levels of estrogen reduce the collagen production by 40% and at pharmacological levels of estrogen, collagen production is decreased by more than 50%. Estrogen fluctuations may alter ligament metabolism and may change the composition of ligament, rendering it more susceptible to injury.

NONSTEROIDAL ANTI-INFLAMMATORY DRUGS

NSAIDs (which include aspirin, acetaminophen, and indomethacin) are frequently used in the treatment of various painful conditions of the musculoskeletal system. NSAIDs are also widely used in the treatment of soft tissue injuries such as inflammatory disorders and partial ruptures of tendons and ligaments. Vogel (1977) found that treatment with indomethacin resulted in increased tensile strength in rat tail tendons. An increase in the proportion of insoluble collagen and in the total collagen content also was observed. Ohkawa (1982) found increased tensile strength in the periodontium of rats after indomethacin treatment. Carlstedt and associates (1986a, 1986b) found that indomethacin treatment increased the tensile strength in developing and healing plantaris longus tendons in the rabbit and noted that the mechanism for this increase was probably an increased cross-linkage of collagen molecules. These animal studies suggest that short-term administration of NSAIDs would not be deleterious for tendon healing but instead would increase the rate of biomechanical restoration of the tissue.

HEMODIALYSIS

Tendinous failure resulting from chronic renal failure does occur, with tendon rupture reaching 36% among individuals receiving hemodialysis. Hyperlaxity of tendons and ligaments was found in 74%, patellar tendon elongation in 49%, and articular hypermobility in 51% of individuals receiving long-term hemodialysis (Rillo et al., 1991). Dialysis-related amyloidosis may cause the deposition of amyloid in the synovium of tendons. The major constituent of the amyloid fibrils is the beta 2-microglobulin (Morita et al., 1995; Honda et al., 1990; Bardin et al., 1985).

GRAFTS

Reconstruction of torn ligaments, especially of the anterior and posterior cruciate ligament, is now a frequent procedure. The need for reconstruction is related to age, activity level, and associated injuries.

Grafts derived from different individuals of the same species are called allografts; grafts derived from the same individual are called autografts. Allograft tissue preservation is done through freeze-drying and low-dose irradiation to reduce rates of rejection and infection and to limit effects on the structural properties. Bone-patellar, tendon-bone, and Achilles tendon are usually used as allograft tissue, whereas the central tissue of the patellar tendon is commonly used as autograft tissue.

Shino et al. (1995) used fresh-frozen allogenic Achilles, tibialis anterior or posterior, and peroneus longus or brevis tendons for ACL reconstruction in humans. Specimens were procured during second-look arthroscopy. Several years after reconstruction, the allografts had collagen fibril profiles that did not resemble normal tendon grafts or normal ACL.

Sasaki et al. (1992) used patellar tendons that had been autografted to reconstruct torn ACLs. Follow-up biopsies were performed 6, 12, and 24 months after surgery. During this time, the autograft underwent considerable changes, and after 24 months the autograft had the appearance of normal ligament tissue. Sasaki suggested that the patellar tendon autograft is a valid functional ACL substitution for patients who desire to perform normal mechanical activity.

Corsetti et al. (1996) reports that replacement tissue undergoes extensive biological remodeling and incorporation. However, even a fully incorporated graft will never duplicate the native ACL but works instead as a check reign that increases the knee function. Tohyama et al. (1996) stated that the graft elongation at the time of implantation influences the long-term outcome of ACL reconstructions, at least in the canine model. They compared those cases where the graft elongation behavior was within the 95% confidence limit of normal ACL (group 1) with those cases where the graft elongation behavior was more than the 95% confidence limit of the normal ACL (group 2). Group 2 had significantly less inner stiffness of the graft than did group 1. Group 2 showed a significantly increased anteroposterior laxity, but there was no difference in ultimate failure load and absorbed energy.

Summary

1 Tendons and extremity ligaments are composed largely of collagen, whose mechanical stability gives these structures their characteristic strength and flexibility. The ligamentum flavum of the spine have a substantial proportion of elastin, which lends these structures their great elasticity.

2 The arrangement of the collagen fibers is nearly parallel in tendons, equipping them to withstand high unidirectional loads. The less parallel arrangement of the collagen fibers in ligaments allows these structures to sustain predominant tensile stresses in one direction and smaller stresses in other directions.

3 At the insertion of ligament and tendon into stiffer bone, the gradual change from a more fibrous to a more bony material results in a decreased stress concentration effect.

4 Tendons and ligaments undergo deformation before failure. When the ultimate tensile strength of these structures is surpassed, complete failure occurs rapidly, and their load-bearing ability is substantially decreased.

5 Studies suggest that during normal activity, a tendon in vivo is subjected to less than one fourth of its ultimate stress.

6 Injury mechanisms in a tendon are influenced by the amount of force produced by the contraction of the muscle to which the tendon is attached and the cross-sectional area of the tendon in relation to that of its muscle.

7 The biomechanical behavior of ligaments and tendons is viscoelastic, or rate dependent, so that these structures display an increase in strength and stiffness with an increased loading rate.

8 An additional effect of rate dependency is the slow deformation, or creep, that occurs when tendons and ligaments are subjected to a constant low load over an extended period; stress relaxation takes place when these structures sustain a constant elongation over time.

9 Aging results in a decline in the mechanical properties of tendons and ligaments, that is, their strength, stiffness, and ability to withstand deformation.

10 Pregnancy, immobilization, diabetes, steroids, NSAID use, and hemodialysis affect the biomechanical properties of ligaments and tendons.

11 Allografts and autografts are useful in ligament reconstruction but material properties do not return completely to normal levels.

12 Ligaments and tendons remodel in response to the mechanical demands placed on them.

REFERENCES

Amiel, D., Kuiper, S.D., Wallace, C.D., Harwood, F., VandeBerg, J.S. (1991). Age-related properties of medial collateral ligament and anterior cruciate ligament: A morphologic and collagen maturation study in the rabbit. *J Gerontol, 46*(4), B159–B165.

Amiel, D., Frank, C., Harwood, F., et al. (1984). Tendons and ligaments: A morphological and biochemical comparison. *J Orthop Res, 1*, 257.

Amiel, D., Woo, S.L-Y., Harwood, F.L., et al. (1982). The effect of immobilization on collagen turnover in connective tissue: A biochemical-biomechanical correlation. *Acta Orthop Scand, 53*, 325.

Bardin, T., Kuntz, D., Zingraff, J., Voisin, M.C., Zelmar, A. (1985). Synovial amyloidosis in patients undergoing long-term hemodialysis. *Arthritis Rheum, 28*(9), 1052–1058.

Campbell, R.B., Wiggins, M.E., Cannistra, L.M., Fadale, P.D., Akelman, E. (1996). Influence of steroid injection on ligament healing in the rat. *Clin Orthop, 332*, 242–253.

Carlstedt, C.A. (1987). Mechanical and chemical factors in tendon healing: Effects of indomethacin and surgery in the rabbit. *Acta Orthop Scand Suppl*, 224.

Carlstedt, C.A., Madsén, K., & Wredmark, T. (1986a). The influence of indomethacin on collagen synthesis during tendon healing in the rabbit. *Prostaglandins, 32*, 353.

Carlstedt, C.A., Madsén, K., & Wredmark, T. (1986b). The influence of indomethacin on tendon healing: A biomechanical and biochemical study. *Arch Orthop Trauma Surg, 105*, 332.

Cavallo, A., Ordóñez, M.A., Gracia, F., Lucas, J.C., Ballesteros, E., Valencia, J., Moreno, M. (1991). Adult, non-insulin-dependent diabetes. Limitation of articular mobility and soft tissue involvement. *Rev Med Chil, 119*(9), 1016–1021.

Cooper, R.R., & Misol, S. (1970). Tendon and ligament insertion: A light and electron microscopic study. *J Bone Joint Surg, 52A*, 1.

Corsetti, J.R., & Jackson, D.W. (1996). Failure of anterior cruciate ligament reconstruction: The biologic basis. *Clin Orthop, 325*, 42–49.

Diamant, J., Keller, A., Baer, E., Litt, M., Arridge, R.G. (1972). Collagen: Ultrastructure and its relations to mechanical properties as a function of aging. *Proc R Soc Lond (Biol), 180*, 293.

Duquette, J.J., Grigg, P., & Hoffman, A.H. (1996). The effect of diabetes on the viscoelastic properties of rat knee ligaments. *J Biomech Eng, 118*(4), 557–564.

Elliott, D.H. (1967). The biomechanical properties of tendon in relation to muscular strength. *Ann Phys Med, 9*, 1.

Fitton Jackson, S. (1965). Antecedent phases in matrix formation. In *Structure and Function of Connective and Skeletal Tissues* (p. 277). London: Butterworth.

Fung, Y.C.B. (1981). *Biomechanics: Mechanical Properties of Living Tissues* (p. 222). New York: Springer-Verlag.

Fung, Y.C.B. (1967). Elasticity of soft tissues in simple elongation. *Am J Physiol, 213*, 1532.

Fung, Y.C.B. (1972). Stress-strain-history relations of soft tissues in simple elongation. In Y.C. Fung, N. Perrone, & M. Anliker (Eds.), *Biomechanics: Its Foundations and Objectives* (pp. 181–208). Englewood Cliffs, NJ: Prentice-Hall.

Hirsch, G. (1974). Tensile properties during tendon healing. *Acta Orthop Scand, Suppl 153*.

Ikeda, S., Hara, M., Ogawa, K., Nara, H., Morimura, N. (1990). Beta 2 microglobulin amyloidosis in hemodialysis patients. An autopsy study of intervertebral disks and posterior ligaments. *Acta Pathol Jpn, 40*(11), 820–826.

Kasperczyk, W.J., Rosocha, S., Bosch, U., Oestern, H.J., Tscherne, H. (1991). Age, activity and strength of knee ligaments. *Unfallchirurg, 94*(7), 372–375.

Kasser, J. (Ed.) (1996). *Orthopaedic Knowledge Update 5: Home Study Syllabus*. Park Ridge, IL: American Academy of Orthopaedic Surgeons.

Kear, M., & Smith, R.N. (1975). A method for recording tendon strain in sheep during locomotion. *Acta Orthop Scand, 46*, 896.

Kennedy, J.C., Hawkins, R.J., Willis, R.B., et al. (1976). Tension studies of human knee ligaments: Yield point, ultimate failure, and disruption of the cruciate and tibial collateral ligaments. *J Bone Joint Surg, 58A*, 350.

Lancaster, R.L., Haut, R.C., DeCamp, C.E. (1994). Changes in the mechanical properties of patellar tendon preparations of spontaneously diabetic dogs under long-term insulin therapy. *J Biomech, 27*(8), 1105–1108.

Liu, S.H., Al-Shaikh, R.A., Panossian, V., Finerman, G.A., Lane, J.M. (1997). Estrogen affects the cellular metabolism of the anterior cruciate ligament: A potential explanation for female athletic injury. *Am J Sports Med, 25*(5), 704–709.

Morita, H., Shinzato, T., Cai, Z., David, G., Mizutani, A., Habuchi, H., [illegible] M., Asai, J., et al. (1995). Basic fibroblast growth factor-heparan sulphate complex in the human dialysis-related amyloidosis. *Virchows Arch, 427*(4), 395–400.

Nachemson, A.L., & Evans, J.H. (1968). Some mechanical properties of the third human lumbar interlaminar ligament (ligamentum flavum). *J Biomech, 1*, 211–220.

Newton, P.O., Woo, S.L-Y., MacKenna, D.A., Akeson, W.H. (1995). Immobilization of the knee joint alters the mechanical and ultrastructural properties of the rabbit anterior cruciate ligament. *J Orthop Res, 13*(2), 191–200.

Noyes, F.R., et al. (1977a). Functional properties of knee ligaments and alterations induced by immobilization. *Clin Orthop, 123*, 210–242.

Noyes, F.R., Grood, E.S., Nussbaum, N.S., Cooper, S.M. (1977b). Effect of intra-articular corticosteroids on ligament properties: A biomechanical and histological study in Rhesus knees. *Clin Orthop, 123*, 197–209.

Noyes, F.R., and Grood, E.S. (1976). The strength of the anterior cruciate ligament in humans and Rhesus monkeys: Age-related and species-related changes. *J Bone Joint Surg, 58A*, 1074–1082.

Ohkawa, S. (1982). Effects of orthodontic forces and anti-inflammatory drugs on the mechanical strength of the periodontium in the rat mandibular first molar. *Am J Orthod, 81*, 498–502.

Oxlund, H. (1980). The influence of a local injection of cortisol on the mechanical properties of tendons and ligaments and the indirect effect on skin. *Acta Orthop Scand, 51*(2), 231–238.

Parry, D.A.D., Barnes, G.R.G., and Craig, A.S. (1978). A comparison of the size distribution of collagen fibrils in connective tissues as a function of age and a possible relation between fibril size and mechanical properties. *Proc R Soc Lond, 203*, 305–321.

Prockop, D.J., & Guzman, N.A. (1977). Collagen diseases and the biosynthesis of collagen. *Hosp Pract*, Dec, 61–68.

Ramachandran, G.N. (1967). Molecular structure of collagen. *Int Rev Connect Tissue Res*, 1, 127.

Rich, A., Crick, F.H.C. (1961). The molecular structure of collagen. *J Mol Biol*, 3, 483.

Rillo, O.L., Babini, S.M., Basnak, A., Wainer, E., Balbachan, E., Cocco, J.A. (1994). Tendinous and ligamentous hyperlaxity in patients receiving long-term hemodialysis. *J Rheumatol*, 21(8), 1237–1239.

Rundgren, A. (1974). Physical properties of connective tissue as influenced by single and repeated pregnancies in the rat. *Acta Physiol Scand*, *Suppl*, 417.

Shino, K., Oakes, B.W., Horibe, S., Nakata, K., Nakamura, N. (1995). Collagen fibril populations in human anterior cruciate ligament allografts. Electron microscopic analysis. *Am J Sports Med*, 23(2), 203–208.

Simon, S.R. (1994). *Orthopaedic Basic Science*. Rosemont, IL: AAOS.

Snell, R.S. (1984). *Clinical and Functional Histology for Medical Students*. Boston: Little, Brown.

Strocchi, R., De Pasquale, V., Facchini, A., Raspanti, M., Zaffagnini, S., Marcacci, M. (1996). Age-related changes in human anterior cruciate ligament (ACL) collagen fibrils. *Ital J Anat Embryol*, 101(4), 213–220.

Strocchi, R., DePasquale, V., Gubellini, P., Marcacci, M., Ruggeri, A. (1992). Ultrastructural modifications of patellar tendon fibres used as anterior cruciate ligament (ACL) replacement. *Ital J Anat Embryol*, 97(4), 221–228.

Tipton, C.M., James, S.L., Mergner, W., et al. (1970). Influence of exercise on strength of medial collateral knee ligaments of dogs. *Am J Physiol*, 218, 894.

Tohyama, H., Beynnon, B.D., Johnson, R.J., Renstrom, P.A., Arms, S.W. (1996). The effect of anterior cruciate ligament graft elongation at the time of implantation on the biomechanical behavior of the graft and knee. *Am J Sports Med*, 24(5), 608–614.

Viidik, A. (1968). Elasticity and tensile strength of the anterior cruciate ligament in rabbits as influenced by training. *Acta Physiol Scand*, 74, 372.

Viidik, A. (1973). Functional properties of collagenous tissues. *Int Rev Connect Tissue Res*, 6, 127.

Viidik, A., Danielsen, C.C., Oxlund, H. (1982). Fourth International Congress of Biorheology Symposium on Mechanical Properties of Living Tissues: On fundamental and phenomenological models, structure and mechanical properties of collagen, elastic and glycosaminoglycan complexes. *Biorheology*, 19, 437.

Vogel, H.G. (1977). Mechanical and chemical properties of various connective tissue organs in rats as influenced by non-steroidal antirheumatic drugs. *Connect Tissue Res*, 5, 91.

Walsh, W.R., Wiggins, M.E., Fadale, P.D., Ehrlich, M.G. (1995). Effects of a delayed steroid injection on ligament healing using a rabbit medial collateral ligament model. *Biomaterials*, 16(12), 905–910.

Walsh, S., Frank, C., Shrive, N., Hart, D. (1993). Knee immobilization inhibits biomechanical maturation of the rabbit medial collateral ligament. *Clin Orthop*, 297, 253–261.

White, A., Handler, P., & Smith, E.L. (1964). *Principles of Biochemistry*. New York: McGraw-Hill.

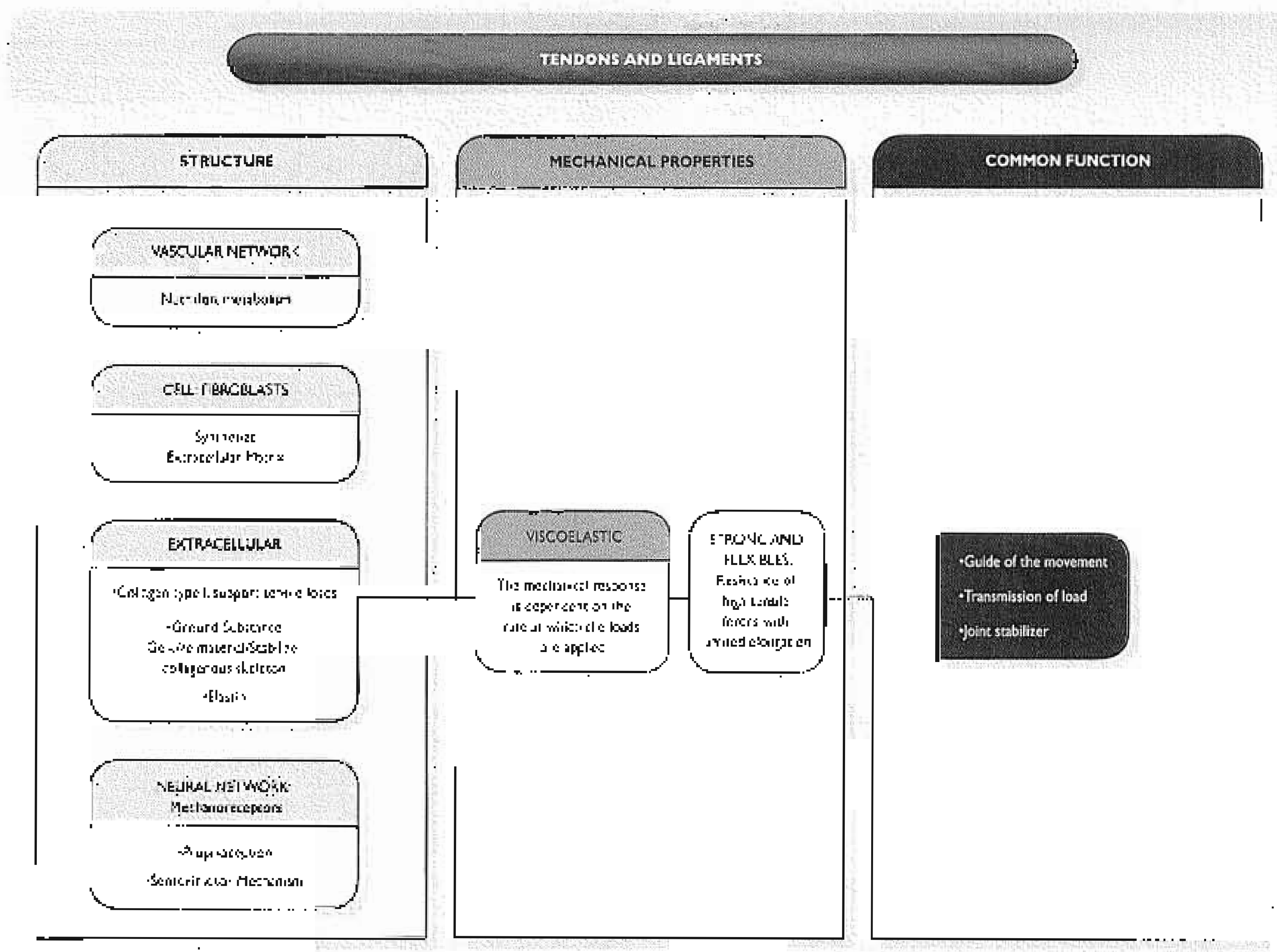
Wiggins, M.E., Fadale, P.D., Barrach, H., Ehrlich, M.G., Walsh, W.R. (1994). Healing characteristics of a type I collagenous structure treated with corticosteroids. *Am J Sports Med*, 22(2), 279–288.

Woo, S.L-Y., An, K.N., Arnoczky, D.V.M., Fithian, D., & Myers, B. (1994). Anatomy, biology and biomechanics of the tendon, ligament, and meniscus. In S.R. Simon (Ed.). *Orthopaedic Basic Science* (p. 52). Rosemont, IL: AAOS.

Woo, S.L.Y., Gomez, M.A., Sites, T.J., et al. (1987). The biomechanical and morphological changes in the medial collateral ligament of the rabbit after immobilization and remobilization. *J Bone Joint Surg*, 69A(8), 1200–1211.

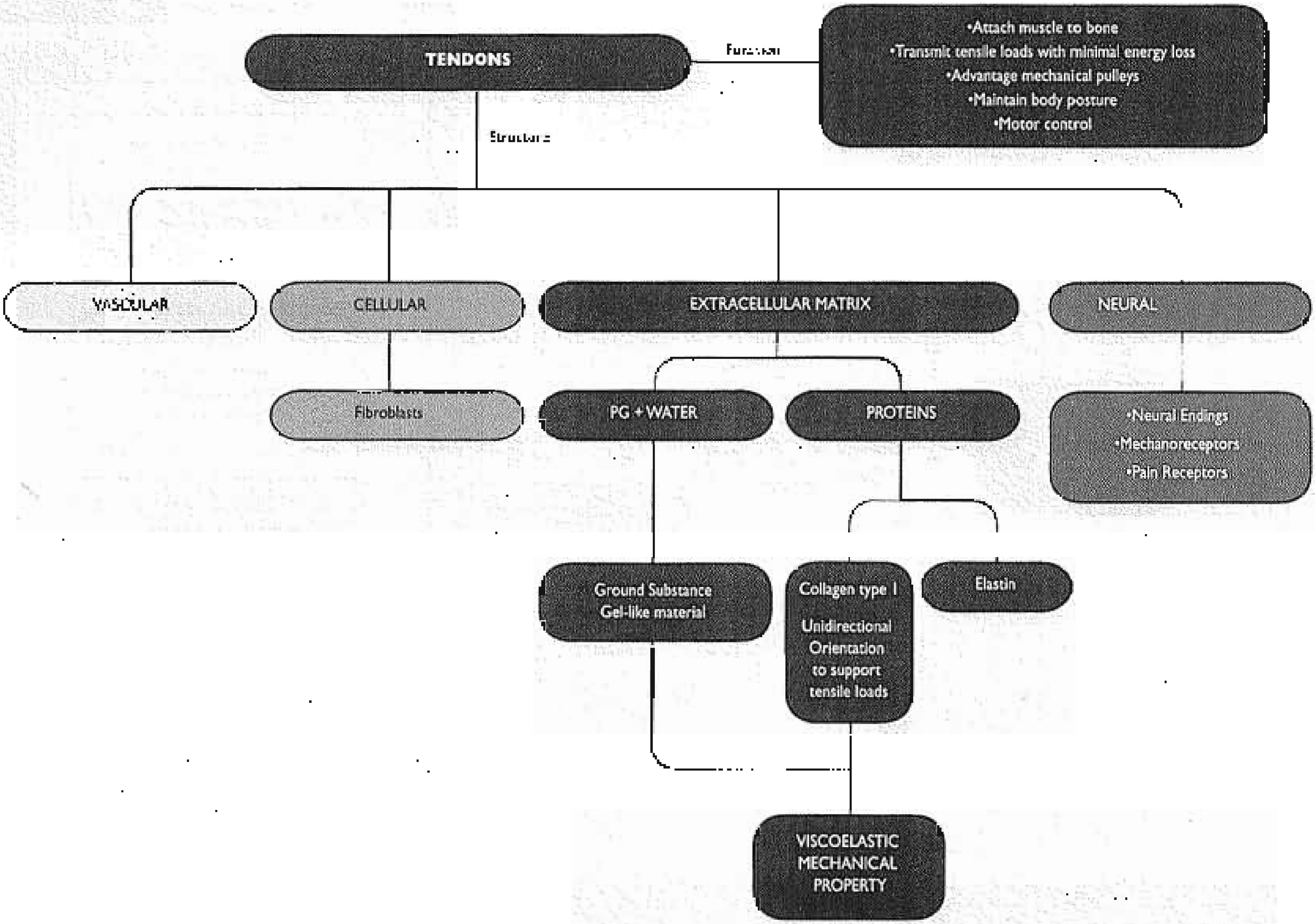
Woo, S.L.Y., Gomez, M.A., Amiel, D., Ritter, M.A., Gelberman, R.H., Akeson, W.H. (1981). The effects of exercise on the biomechanical and biochemical properties of swine digital flexor tendons. *J Biomech Eng*, 103, 51.

Woo, S.L.Y. (1988). Ligament, tendon, and joint capsule insertions to bone. In S.L.Y. Woo & J. Buckwalter (Eds.), *Injury and Repair of the Musculoskeletal Soft Tissues* (pp. 133–166). Park Ridge, IL: American Academy of Orthopaedic Surgeons.



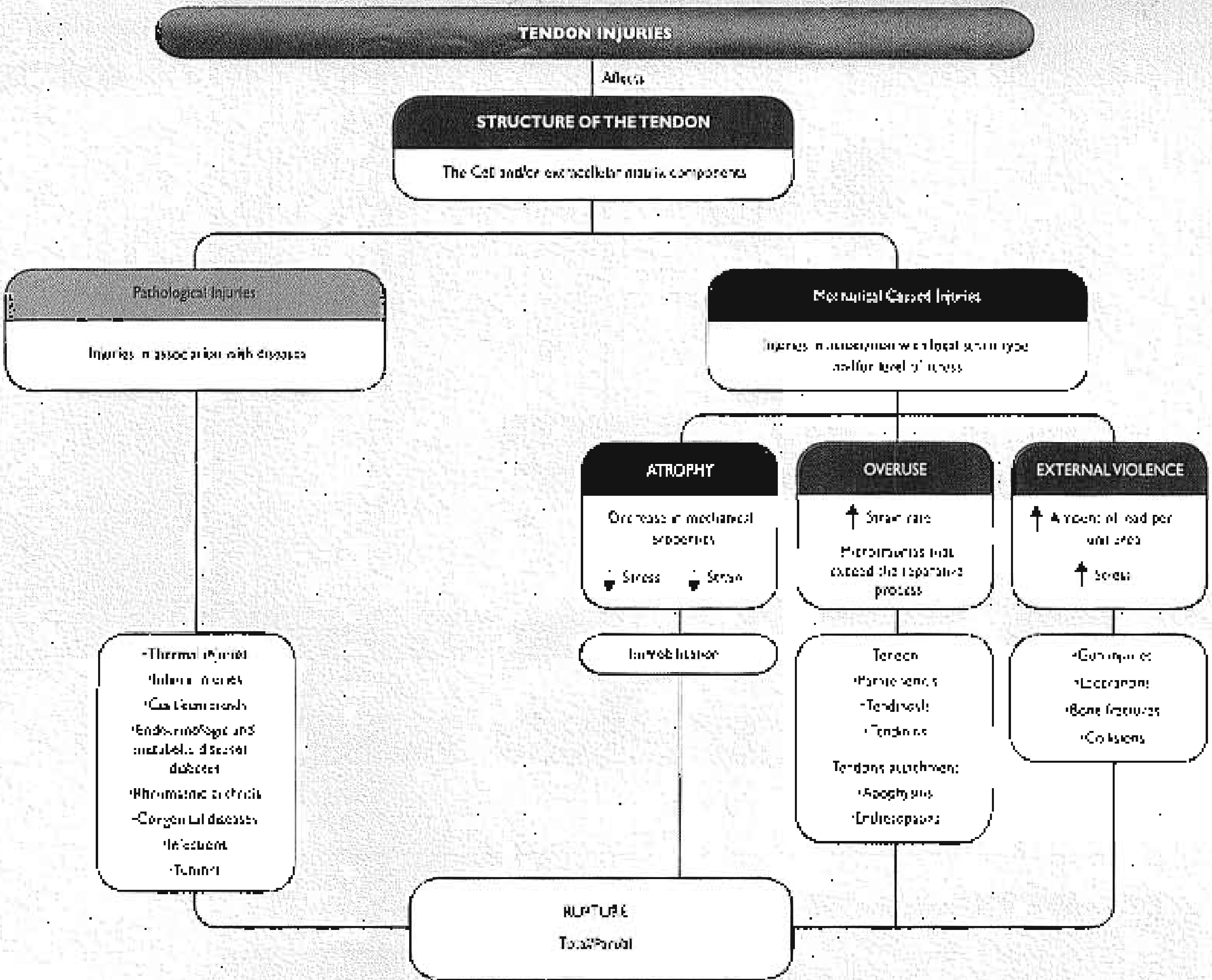
FLOW CHART 4-1 Common structure and mechanical properties of tendons and ligaments.*

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.

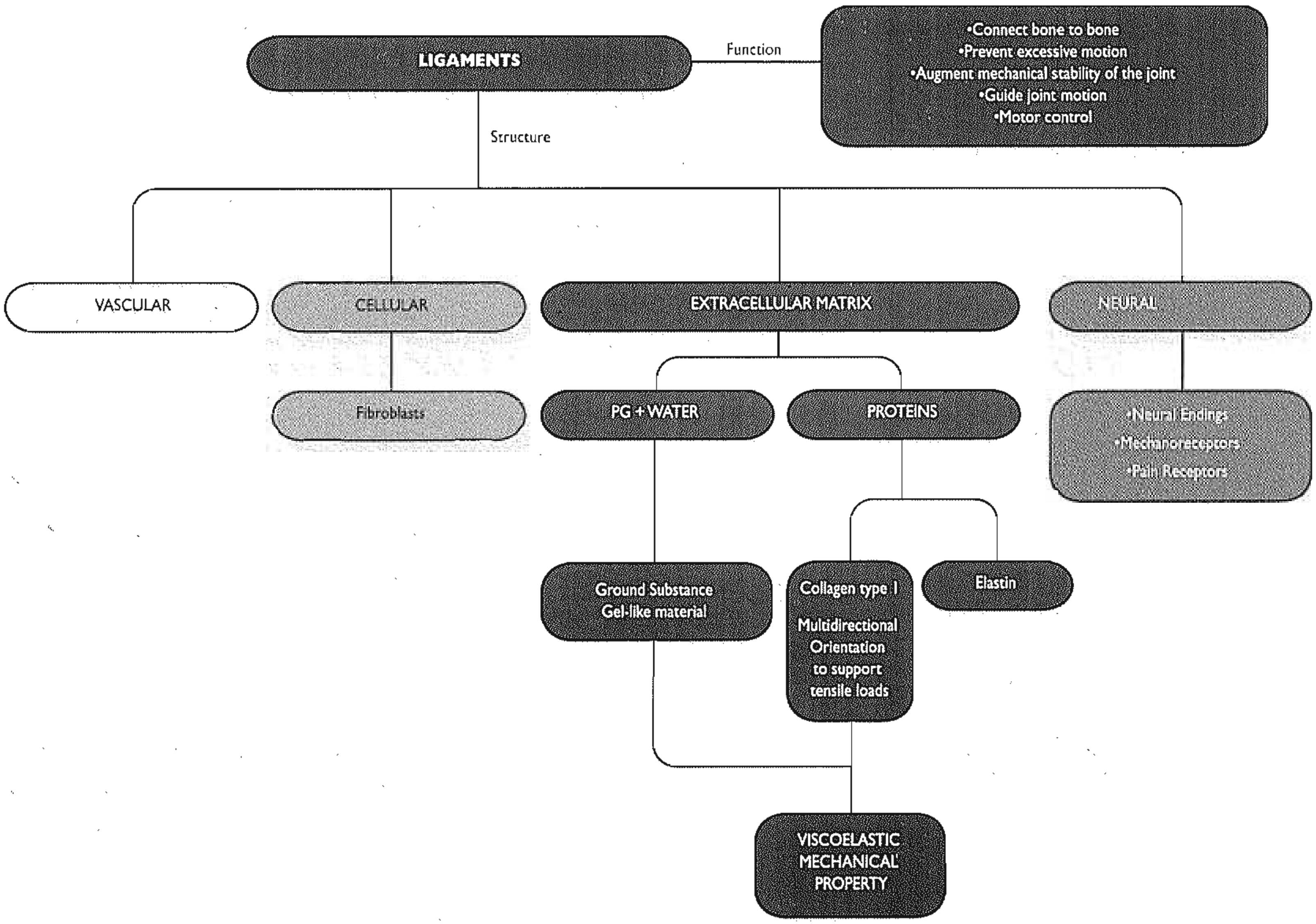


FLOW CHART 4-2 Tendons structure and mechanical properties.*

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.

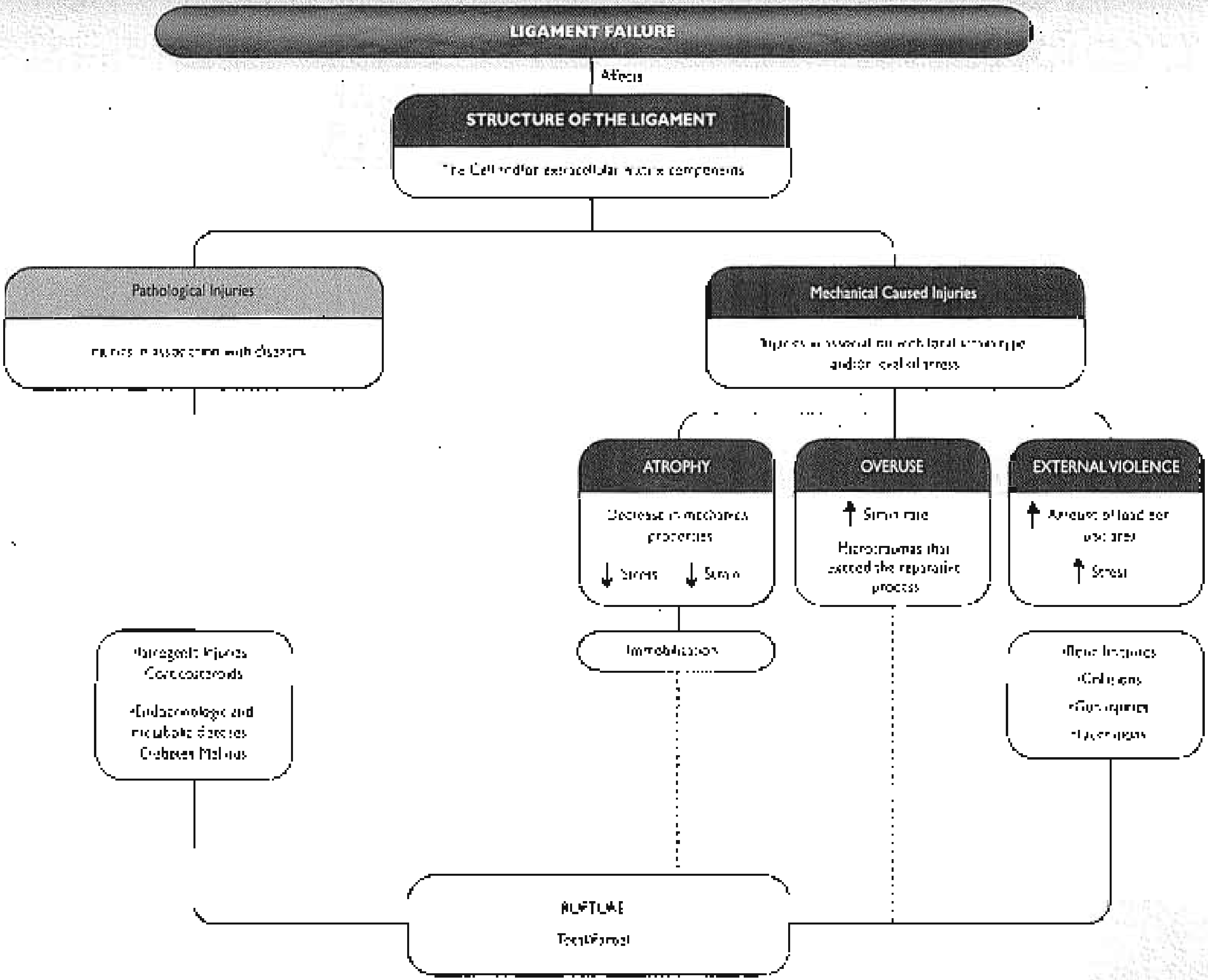


FLOW CHART 4-3 Tendon Injuries. Clinical examples.*

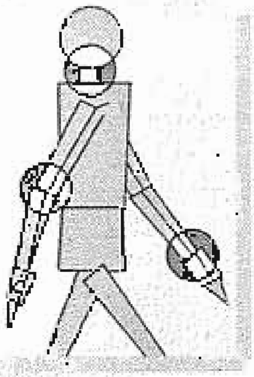


FLOW CHART 4-4 Ligaments structure and mechanical properties.* (*PG*, Proteoglycan)

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.



FLOW CHART 4-5 Ligament failure. Clinical examples.



Biomechanics of Peripheral Nerves and Spinal Nerve Roots

Bjorn Rydevik, Goran Lundborg, Kjell Olmarker, Robert R. Myers

Introduction

The nervous system serves as the body's control center and communications network. As such, it has three broad roles: it senses changes in the body and in the external environment, it interprets these changes, and it responds to this interpretation by initiating action in the form of muscle contraction or gland secretion.

For descriptive purposes, the nervous system can be divided into two parts: the central nervous system, consisting of the brain and spinal cord, and the peripheral nervous system, composed of the various nerve processes that extend from the brain and spinal cord. These peripheral nerve processes provide input to the central nervous system from sensory receptors in skin, joints, muscles, tendons, viscera, and sense organs and provide output from it to effectors (muscles and glands). The peripheral nervous system includes 12 pairs of cranial nerves and their branches and 31 pairs of spinal nerves and their branches (Fig. 5-1*A*). These branches are called peripheral nerves.

Each spinal nerve is connected to the spinal cord through a posterior (dorsal) root and an anterior (ventral) root, which unite to form the spinal nerve at the intervertebral foramen (Fig. 5-1, *B*, *D*). The posterior roots contain fibers of sensory neurons (those conducting sensory information from receptors in the skin, muscles, tendons, and joints to the central nervous system) and the anterior roots contain mainly fibers of motor neurons (those that conduct impulses from the central nervous system to distal targets such as muscle fibers).

Shortly after the spinal nerves leave their intervertebral foramina, they divide into two main branches: the dorsal rami, which innervate the muscles and skin of the head, neck, and back, and the generally larger and more important ventral rami, which innervate the ventral and lateral parts of these structures as well as the upper and lower extremities. Except in the thoracic region, the ventral rami do not run directly to the structures that they innervate but first form interlacing networks, or plexuses, with adjacent nerves (Fig. 5-1*A*).

This chapter focuses on both the peripheral nerves and spinal nerve roots, which contain not only nerve fibers but also connective tissue elements and vascular structures that encompass the nerve fibers. The nerves possess some special anatomical properties that may serve to protect the nerve from mechanical damage, for instance, stretching (tension) and compression. In this chapter, the basic microanatomy of the peripheral nerves and the spinal nerve roots are reviewed with special reference to these built-in mechanisms of protection. The mechanical behavior of peripheral nerves that are subjected to tension and compression is also described in some detail.

Anatomy and Physiology of Peripheral Nerves

The peripheral nerves are complex composite structures consisting of nerve fibers, connective tissue, and blood vessels. Because the three tissue elements that make up these nerves react to trauma in different ways and may each play distinct roles in the functional deterioration of the nerve after injury, each element is described separately.

THE NERVE FIBERS: STRUCTURE AND FUNCTION

The term nerve fiber refers to the elongated process (axon) extending from the nerve cell body along with its myelin sheath and Schwann cells (Figs. 5-2 and 5-3). The nerve fibers of sensory neurons conduct impulses from the skin, skeletal muscles, and joints to the central nervous system. The nerve fibers of the motor neurons convey impulses from the central nervous system to the skeletal muscles, causing muscle contraction. (A detailed description of the mechanics of muscle contraction is given in Chapter 6.)

The nerve fibers not only transmit impulses but also serve as an anatomical connection between the nerve cell body and its end organs. This connection is maintained by axonal transport systems through which various substances synthesized within the cell body (e.g., proteins) are transported from the cell body to the periphery and in the opposite direction. The axonal transport takes place at speeds that vary from approximately 1 to approximately 400 mm per day.

Most axons of the peripheral nervous system are surrounded by multilayered, segmented coverings known as myelin sheaths (Fig. 5-3). Fibers with this covering are said to be myelinated, whereas those without it (mainly small sensory fibers conducting impulses for pain from the skin) are unmyelinated. The myelin sheath of the axons of the peripheral nerves is produced by flattened cells called Schwann cells arranged along the axon (Fig. 5-3). A sheath is formed as the Schwann cell encircles the axon and

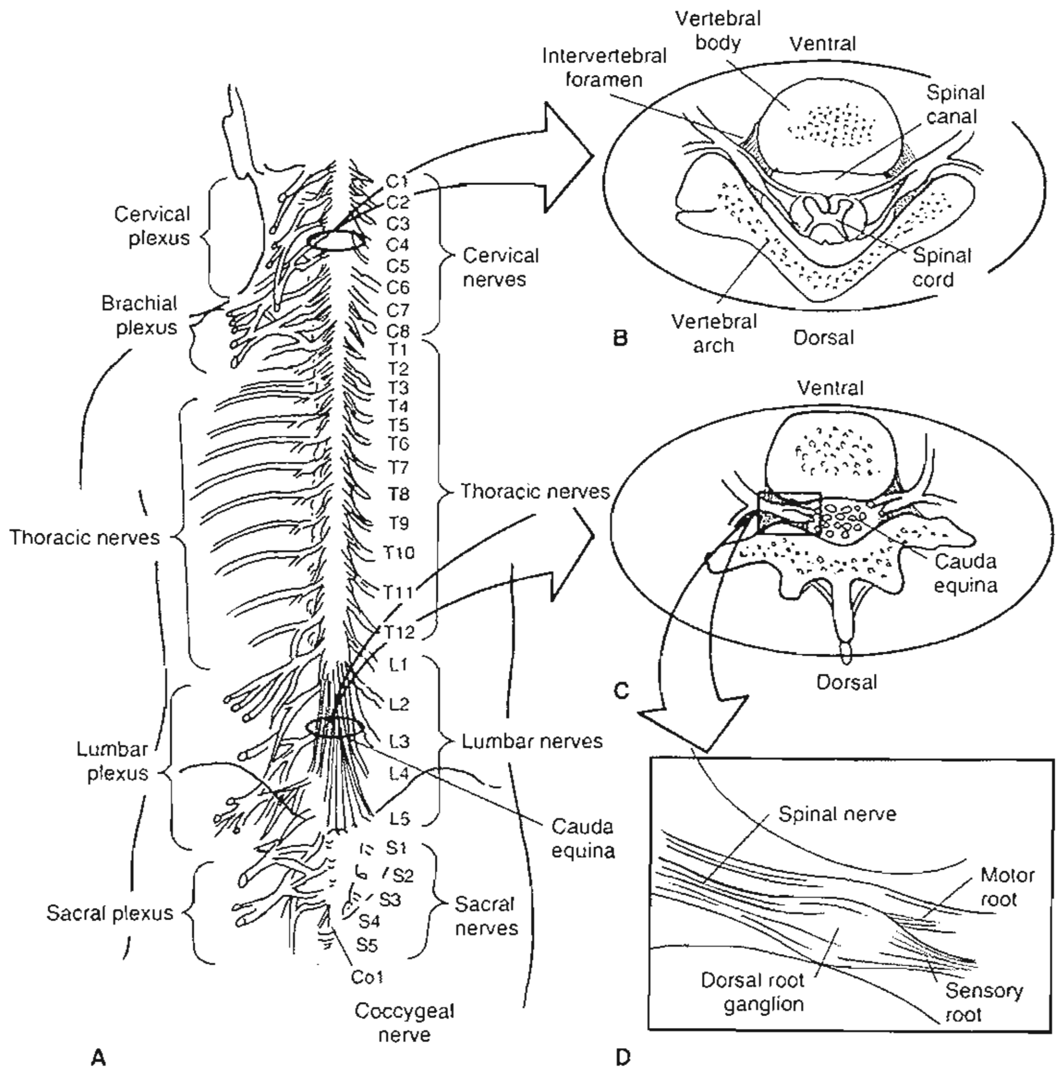


FIG. 5-1

A, Schematic drawing of the spinal cord and the spinal nerves (posterior view). The spinal nerves emerge from the spinal canal through the intervertebral foramina. There are 8 pairs of cervical nerves, 12 pairs of thoracic nerves, 5 pairs of lumbar nerves, 5 pairs of sacral nerves, and 1 pair of coccygeal nerves. Except in the region of the 2nd to the 11th thoracic vertebrae (T2–T11), the nerves form complex networks called plexuses after exiting the intervertebral foramina. Only the main branch of each nerve, the ventral ramus, is depicted. *Adapted from Tortora, G.J. & Anagnostakos, N.P. (1984).* Principles of Anatomy and Physiology *(4th ed.). New York: Harper & Row.* **B,** Cross-section of the cervical spine showing the spinal cord in the spinal canal and the nerve roots exiting through the intervertebral foramina. **C,** Cross-section of the lumbar spine showing the nerve roots of the cauda equina in the spinal canal. **D,** Each exiting nerve root complex in the intervertebral foramen consists of a motor root, a sensory root, and a dorsal root ganglion.

winds around it many times, pushing its cytoplasm and nucleus to the outside layer. Unmyelinated gaps called nodes of Ranvier lie between the segments of the myelin sheath at approximately 1 to 2 mm apart.

The myelin sheath increases the speed of the conduction of nerve impulses, and insulates and maintains the axon. Impulses are propagated along the unmyelinated nerve fibers in a slow, continuous way, whereas in the myelinated nerve fibers the impulses "jump" at a higher speed from one node of Ranvier to the next in a process called saltatory conduction. The conduction velocity of a myelinated nerve is directly proportional to the diameter of the fiber, which usually ranges from 2 to 20 μm. Motor fibers that innervate skeletal muscle have large diameters, as do sensory fibers that relay impulses associated with touch, pressure, heat, cold, and kinesthetic sense, such as skeletal muscle tension and joint position. Sensory fibers that conduct impulses for dull, diffuse pain (as opposed to sharp, immediate pain) have the smallest diameters. Nerve fibers are packed closely in fascicles, which are further arranged into bundles that make up the nerve itself. The fascicles are the functional subunits of the nerve.

INTRANEURAL CONNECTIVE TISSUE OF PERIPHERAL NERVES

Successive layers of connective tissue surround the nerve fibers—called the endoneurium, perineurium, and epineurium—and protect the fibers' continuity (Fig. 5-4). The protective function of these connective tissue layers is essential because nerve fibers are extremely susceptible to stretching and compression.

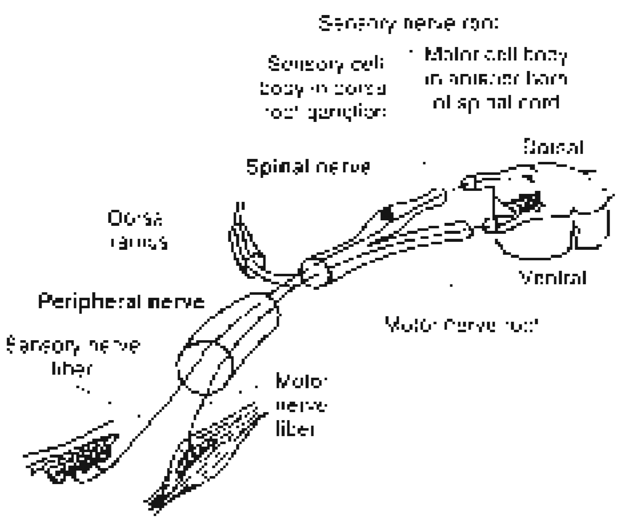


FIG. 5-2

Schematic representation of the arrangement of a typical spinal nerve as it emerges from its dorsal and ventral nerve roots. The peripheral nerve begins after the dorsal ramus branches off. (For the sake of simplicity, the nerve is not shown entering a plexus.) Spinal nerves and most peripheral nerves are mixed nerves; they contain both sensory (afferent) and motor (efferent) nerve fibers. The cell body and its nerve fibers make up the neuron. The cell bodies of the motor neurons are located in the anterior horn of the spinal cord, and those of the sensory neurons are found in the dorsal root ganglia. Here, a motor nerve fiber is shown innervating muscle and a sensory nerve fiber is depicted innervating skin. *Adapted from Rydevik, B., Brown, M.D., & Lundborg, G. (1984). Pathoanatomy and pathophysiology of nerve root compression. Spine, 9, 7.*

The outermost layer, the epineurium, is located between the fascicles and superficially in the nerve. This rather loose connective tissue layer serves as a cushion during movements of the nerve, protecting the fascicles from external trauma and maintaining the oxygen supply system via the epineural blood vessels. The amount of epineural connective tissue varies among nerves and at different levels within the same nerve. Where the nerves lie close to bone or pass joints, the epineurium is often more abundant than elsewhere, as the need for protection may be greater in these locations. The spinal nerve roots are devoid of both epineurium and perineurium, and the nerve fibers in the nerve root may therefore be more susceptible to trauma (Rydevik et al., 1984).

The perineurium is a lamellar sheath that encompasses each fascicle. This sheath has great mechanical strength as well as a specific biochemical barrier. Its strength is demonstrated by the fact that the fascicles can be inflated by fluid to a pressure of approximately 1000 mm of mercury (Hg) before the perineurium ruptures.

The barrier function of the perineurium chemically isolates the nerve fibers from their surroundings, thus preserving an ionic environment of the interior of the fascicles, a special milieu interieur. The endoneurium, the connective tissue inside the fascicles, is composed principally of fibroblasts and collagen.

The interstitial tissue pressure in the fascicles, the endoneurial fluid pressure, is normally slightly elevated (+1.5 ± 0.7 mm Hg [Myers & Powell, 1981]) compared with the pressure in surrounding tissues such as subcutaneous tissue (-4.7 ± 0.8 mm Hg) and muscle tissue (-2 ± 2 mm Hg). The elevated endoneurial fluid pressure is illustrated by the phenomenon whereby incision of the perineurium results in herniation of nerve fibers. The endoneurial fluid pressure may increase further as a result of trauma to the nerve, with subsequent edema. Such a pressure increase may affect the microcirculation and the function of the nerve.

THE MICROVASCULAR SYSTEM OF PERIPHERAL NERVES

The peripheral nerve is a well-vascularized structure containing vascular networks in the epineurium, the perineurium, and the endoneurium. Because both impulse propagation and axonal transport depend

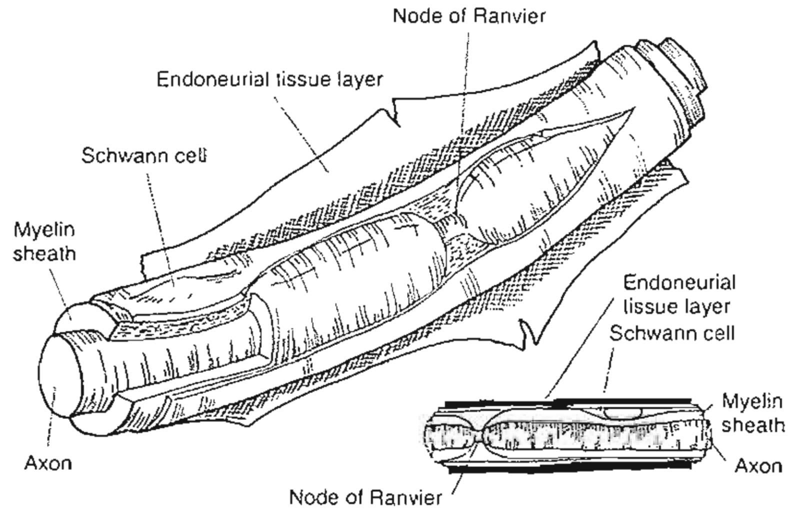


FIG. 5-3

Schematic drawings of the structural features of a myelinated nerve fiber. *Adapted from Sunderland, S. (1978).* Nerves and Nerve Injuries *(2nd ed.). Edinburgh: Churchill Livingstone.*

on a local oxygen supply, it is natural that the microvascular system has a large reserve capacity.

The blood supply to the peripheral nerve as a whole is provided by large vessels that approach the nerve segmentally along its course. When these local nutrient vessels reach the nerve, they divide into ascending and descending branches. These vessels run longitudinally and frequently anastomose with the vessels in the perineurium and endoneurium. Within the epineurium, large arterioles and venules, 50 to 100 μm in diameter, constitute a longitudinal vascular system (Fig. 5-4).

Within each fascicle lies a longitudinally oriented capillary plexus with loop formations at various levels. The capillary system is fed by arterioles 25 to 150 μm in diameter that penetrate the perineurial membrane. These vessels run an oblique course through the perineurium, and it is believed that because of this structural peculiarity, they are easily closed like valves in the event that tissue pressure inside the fascicles increases (Lundborg, 1975; Myers et al., 1986). This phenomenon may explain why even a limited increase in endoneurial fluid pressure is associated with a reduction in intrafascicular blood flow.

The built-in safety system of longitudinal anastomoses provides a wide margin of safety if the regional segmental vessels are transected. In an ex-

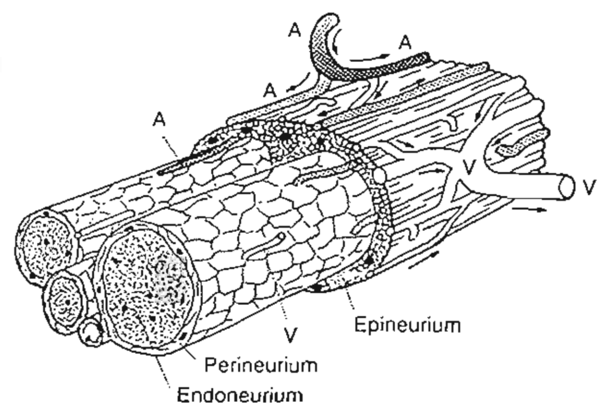


FIG. 5-4

Schematic drawing of a segment of a peripheral nerve. The individual nerve fibers are located within the endoneurium. They are closely packed in fascicles, each of which is surrounded by a strong sheath, the perineurium. A bundle of fascicles is embedded in a loose connective tissue, the epineurium. Blood vessels are present in all layer of the nerve. *A*, arterioles (*shaded*); *V*, venules (*unshaded*). The *arrows* indicated the direction of blood flow. *Adapted from Dahlin, L.B., Rydevik, B., & Lundborg, G. (1986). The pathophysiology of nerve entrapments and nerve compression injuries. In A.R. Hargens (Ed.).* Effects of Mechanical Stress on Tissue Viability. *New York: Springer-Verlag.*

perimental animal in vivo model, it is extremely difficult to induce complete ischemia to a nerve by local surgical procedures. For example, if the whole sciatic-tibial nerve complex of a rabbit (15 cm long) is surgically separated from its surrounding structures and the regional nutrient vessels are cut, there is no detectable reduction in the intrafascicular blood flow as studied by intravital microscopic techniques. Even if such a mobilized nerve is cut distally or proximally, the intraneural longitudinal vascular systems can maintain the microcirculation at least 7 to 8 cm from the cut end. If a nonmobilized nerve is cut, there is still perfect microcirculation even at the very tip of the nerve; this phenomenon demonstrates the sufficiency of the intraneural vascular collaterals. However, other studies in rats indicate that stripping the epineural circulation from nerve bundles causes demyelination of subperineural nerve fibers.

Anatomy and Physiology of Spinal Nerve Roots

In the early embryological developmental stages, the spinal cord has the same length as the spinal column. However, in the fully grown individual, the spinal cord ends as the conus medullaris, approximately at the level of the first lumbar vertebra. A nerve root that leaves the spinal canal through an intervertebral foramen in the lumbar or sacral spine therefore has to pass from the point where it leaves the spinal cord, which is in the lower thoracic spine, to the point of exit from the spine (Fig. 5-5). Because the spinal cord is not present below the first lumbar vertebra, the nervous content of the spinal canal is only comprised of the lumbosacral nerve roots. This "bundle" of nerve roots within the lumbar and sacral part of the spinal canal has been suggested to resemble the tail of a horse and is therefore often called the cauda equina, that is, tail of horse.

Two different types of nerve roots are found within the lumbosacral spine: ventral/motor roots and dorsal/sensory roots. The cell bodies of the motor axons are located in the anterior horns of the gray matter in the spinal cord, and because these nerve roots leave the spinal cord from the ventral aspect, they are also called ventral roots. The other type of nerve root is the sensory, or dorsal root. As the name suggests, these nerve roots mainly comprise sensory (i.e., afferent) axons and reach the spinal cord at the dorsal region of the spinal cord. The cell bodies of the sensory axons are located in a "swelling" of the most caudal part of the respective dorsal nerve root, called the dorsal root ganglion. The dorsal root ganglia are located in or close to the intervertebral foramen. Unlike the nerve roots, the dorsal root ganglia are not enclosed by cerebrospinal fluid and the meninges. Instead, they are enclosed by both a multilayered connective tissue sheath, similar to the perineurium of the peripheral

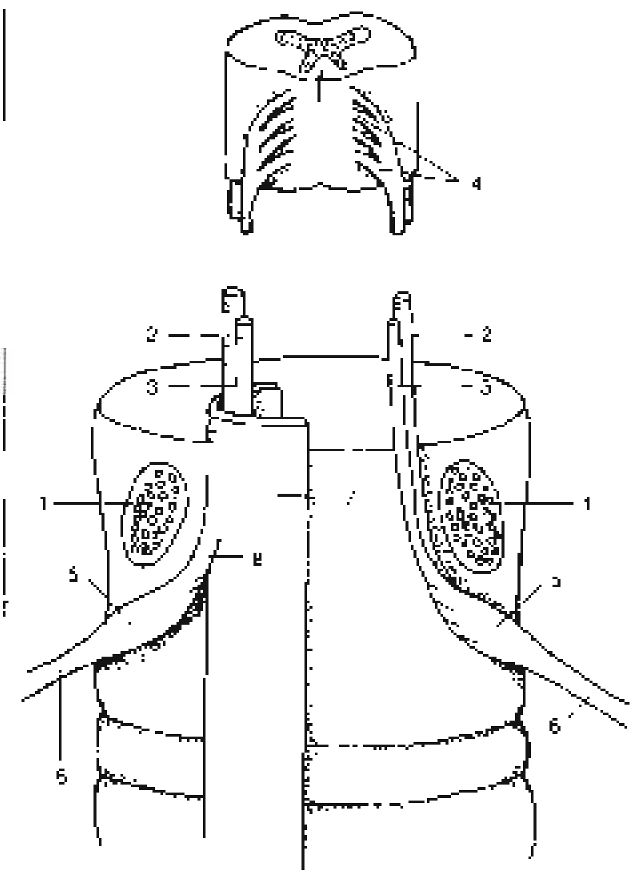


FIG. 5-5

The intraspinal nervous structures as seen from behind. The vertebral arches are removed by cutting the pedicles (1). A ventral (2) and a dorsal (3) nerve root leave the spinal cord as small rootlets (4). Before leaving the spinal canal, the dorsal root forms a swelling called the dorsal root ganglion (5), which contains the sensory cell bodies, before forming the spinal nerve (6) together with the ventral nerve root. The nerve roots are covered by a central dural sac (7) or with extensions of this sac called nerve root sleeves (8). *Reproduced with permission from Olmarker, K. (1991). Spinal nerve root compression. Acute compression of the cauda equina studied in pigs. Acta Orthop Scand, 62, Suppl 242.*

nerve, and a loose connective tissue layer called epineurium.

When the nerve root approaches the intervertebral foramen, the root sleeve gradually encloses the nerve tissue more tightly. The subarachnoid space and the amount of cerebrospinal fluid surrounding each nerve root pair will thus become gradually reduced in the caudal direction. Compression injury of a nerve root may induce an increase in the permeability of the endoneurial capillaries, resulting in edema formation (Olmarker et al., 1989b; Rydevik & Lundborg, 1977). This can lead to an increase of the intraneural fluid and subsequent impairment of the nutritional transport to the nerve (Myers & Powell, 1981; Myers, 1998). Such a mechanism might be particularly important at locations where the nerve roots are tightly enclosed by connective tissue. Thus there is a more pronounced risk for an "entrapment syndrome" within the nerve roots at the intervertebral foramen than more central in the cauda equina (Rydevik et al., 1984). The dorsal root ganglion, with its content of sensory nerve cell bodies tightly enclosed by meninges, might be particularly susceptible to edema formation.

MICROSCOPIC ANATOMY OF SPINAL NERVE ROOTS

There are two microscopically different regions of the nerve roots. Closest to the spinal cord is a central glial segment comprised of glial cells and therefore resembles the microscopic organization of central nervous structures of the spinal cord or the brain. This glial segment is transferred to a nonglial segment in a "cone-shaped" junction a few millimeters from the spinal cord. This nonglial segment is organized in the same manner as the endoneurium of the peripheral nerves, that is, with Schwann cells instead of glial cells. However, some small islets of glial cells also are found in this otherwise "peripherally" organized endoneurium.

MEMBRANOUS COVERINGS OF SPINAL NERVE ROOTS

The axons in the endoneurium are separated from the cerebrospinal fluid by a thin layer of connective tissue called the root sheath. This root sheath is the structural analogue to the pia mater that covers the spinal cord. There are usually 2 to 5 cellular layers in the root sheath, but as many as 12 layers have been identified. The cells of the proximal part of the outer layers of the root sheath are similar to the pia cells of the spinal cord, and the cells in the distal part are more similar to the arachnoid cells of the spinal dura. The inner layers of the root sheath are comprised of cells that show similarities to the cells of the perineurium of peripheral nerves. An interrupted basement membrane encloses these cells separately. The inner layers of the root sheath constitute a diffusion barrier between the endoneurium of the nerve roots and the cerebrospinal fluid. This barrier is considered to be relatively weak and may only prevent the passage of macromolecules.

The spinal dura encloses the nerve roots and the cerebrospinal fluid. When the two layers of the cranial dura enter the spinal canal, the outer layer blends with the periosteum of the part of the laminae of the cervical vertebrae facing the spinal canal. The inner layers join the arachnoid and become the spinal dura. In contrast to the root sheath, the spinal dura is an effective diffusion barrier. The barrier properties are located in a connective tissue sheath between the dura and the arachnoid called the neurothelium. Similar to the inner layer of the root sheath, the neurothelium resembles the perineurium of the peripheral nerves. It is suggested that these two layers in fact form the perineurium when the nerve root is transformed to a peripheral nerve upon leaving the spinal.

THE MICROVASCULAR SYSTEM OF SPINAL NERVE ROOTS

Information about the vascular anatomy of the nerve roots has mainly been derived from studies on the vascularization of the spinal cord. Therefore, the nomenclature of the various vessels has been somewhat confusing. A summary of the existing knowledge on nerve root vasculature will be presented below.

The segmental arteries generally divide into three branches when approaching the intervertebral foramen: (1) an anterior branch that supplies the posterior abdominal wall and lumbar plexus, (2) a posterior branch that supplies the paraspinal muscles and facet joints, and (3) an intermediate branch that supplies the contents of the spinal canal. A branch of the intermediate branch joins the nerve root at the level of the dorsal root ganglion. There are usually three branches from this vessel: one to the ventral root, one to the dorsal root, and one to the vasa corona of the spinal cord.

The branches to the vasa corona of the spinal cord, called medullary arteries, are inconsistent. Only 7 to 8 remain of the 128 from the embryological period of life, and each supplies more than one segment of the spinal cord. The main medullary artery in the thoracic region of the spine was discovered by Adamkiewicz in 1881 and still bears his name. The medullary arteries run parallel to the nerve roots (Fig. 5-6). In humans, there are no connections between these vessels and the vascular network of the nerve roots. Because the medullary feeder arteries only occasionally supply the nerve roots with blood, they have been referred to as the extrinsic vascular system of the cauda equina.

The vasculature of the nerve roots is formed by branches from the intermediate branch of the segmental artery distally and by branches from the vasa corona of the spinal cord proximally. As opposed to the medullary arteries, this vascular network has been named the intrinsic vascular system of the cauda equina. The distal branch to the dorsal root first forms the ganglionic plexus within the dorsal root ganglion. The vessels run within the outer layers of the root sheath, called epi-pial tissue. As there are vessels coming from both distal and proximal directions, the nerve roots are supplied by two separate vascular systems. The two systems anastomose at approximately two thirds of the nerve root length from the spinal cord. This location demonstrates a region of a less-developed vascular network and has been suggested to be a particularly vulnerable site of the nerve roots.

The arteries of the intrinsic system send branches down to the deeper parts of the nerve tissue in a T-like manner. To compensate for elongation of the nerve roots, the arteries are coiled both longitudinally and in the steep running branches between the different fascicles (Fig. 5-6). Unlike peripheral nerves, the venules do not course together with the arteries in the nerve roots but instead usually have a spiraling course in the deeper parts of the nerve.

There is a barrier of the endoneurial capillaries in peripheral nerves called the blood-nerve barrier, which is similar to the blood-brain barrier of the central nervous system (Lundborg, 1975; Rydevik & Lundborg, 1977). The presence of a corresponding barrier in nerve roots has been questioned. If present, a blood-nerve barrier in nerve roots does not seem to be as well developed as in endoneurial capillaries of peripheral nerve, which implies that edema may be formed more easily in nerve roots than in peripheral nerves (Rydevik et al., 1984).

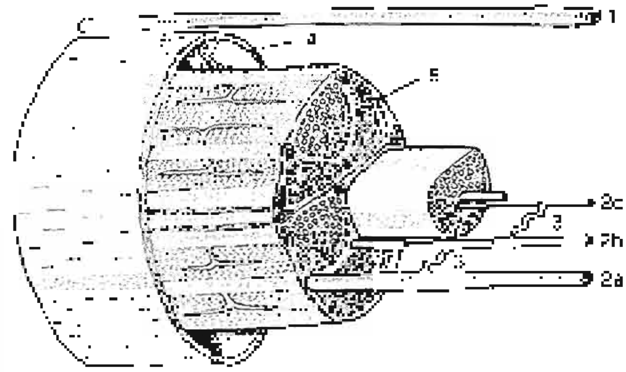


FIG. 5-6

Schematic presentation of some anatomical features of the intrinsic arteries of the spinal nerve roots. The arterioles within the cauda equina may be referred to either the extrinsic (1) or the intrinsic (2) vascular system. From the superficial intrinsic arterioles are branches that continue almost at right angles down between the fascicles. These vessels often run in a spiraling course, thus forming vascular "coils" (3). When reaching a specific fascicle they branch in a T-like manner, with one branch running cranially and one caudally, forming interfascicular arterioles (2b). From these interfascicular arterioles are small branches that enter the fascicles, where they supply the endoneurial capillary networks (2c). Arterioles of the extrinsic vascular system run outside the spinal dura (4) and have no connections with the intrinsic system by local vascular branches. The superficial intrinsic arterioles (2a) are located within the root sheath (5). *Reproduced with permission from Olmarker, K. (1991). Spinal nerve root compression. Acute compression of the cauda equina studied in pigs. Acta Orthop Scand, 62 (Suppl 242).*

Biomechanical Behavior of Peripheral Nerves

External trauma to the extremities and nerve entrapment may produce mechanical deformation of the peripheral nerves that results in the deterioration of nerve function. If the mechanical trauma exceeds a certain degree, the nerves' built-in mechanisms of protection may not be sufficient, resulting in changes in nerve structure and function. Common modes of nerve injury are stretching and compression, which may be inflicted, respectively, by rapid extension and crushing.

STRETCHING (TENSILE) INJURIES OF PERIPHERAL NERVES

Nerves are strong structures with considerable tensile strength. The maximal load that can be sustained by the median and ulnar nerves is in the range of 70 to 220 newtons (N) and 60 to 150 N, respectively. These figures are of academic interest only because severe intraneural tissue damage is produced by tension long before a nerve breaks.

A discussion of the elasticity and biomechanical properties of nerves is complicated by the fact that nerves are not homogeneous isotropic materials but, instead, composite structures, with each tissue component having its own biomechanical properties. The connective tissues of the epineurium and perineurium are primarily longitudinal structures.

When tension is applied to a nerve, initial elongation of the nerve under a very small load is followed by an interval in which stress and elongation show a linear relationship characteristic of an elastic material (Fig. 5-7). As the limit of the linear region is approached, the nerve fibers start to rupture inside the endoneurial tubes and inside the intact perineurium. The perineurial sheaths rupture at approximately 25 to 30% elongation (ultimate strain) above in vivo length (Rydevik et al., 1990). After this point, there is a disintegration of the elastic properties, and the nerve behaves more like a plastic material (i.e., its response to the release of loads is incomplete recovery).

Although variations exist in the tensile strength of various nerves, the maximal elongation at the elastic limit is approximately 20%, and complete structural failure seems to occur at a maximum elongation of approximately 25 to 30%. These values are for normal nerves; injury to a nerve may induce changes in its mechanical properties, namely increased stiffness and decreased elasticity.

Stretching, or tensile, injuries of peripheral nerves are usually associated with severe accidents, such as when high-energy tension is applied to the brachial plexus in association with a birth-related injury, as a result of high-speed vehicular collision, or after a fall from a height. Such plexus injuries may result in partial or total functional loss of some or all of the nerves in the upper extremity, and the consequent functional deficits represent a considerable disability in terms of sensory and motor loss. The outcome depends on which tissue components of the nerves are damaged as well as on the extent of the tissue injury. Of clinical importance is the observation that there can be considerable structural damage (perineurial sheath injuries) induced by stretching with no visible injury on the surface of the nerve (Case Study 5-1).

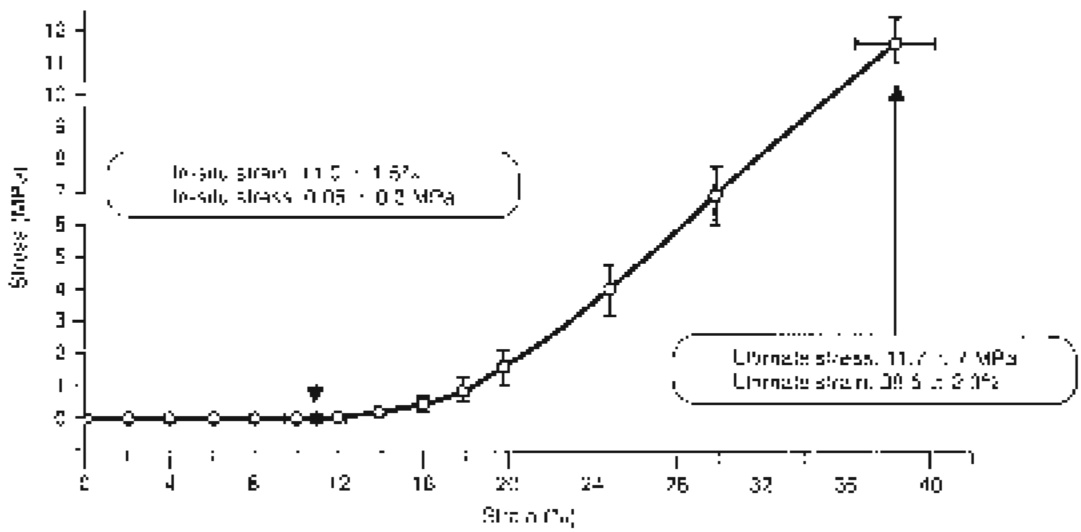


FIG. 5-7

The stress-strain behavior of a rabbit tibial nerve. The nerve exhibits a low-stiffness toe region of approximately 15% and begins to retain significant tension as the strain increases beyond 20%. *Reprinted with permission from Rydevik, B.L., Kwan, M.K., Myers, R.R., et al. (1990). An in vitro mechanical and histological study of acute stretching on rabbit tibial nerve. J Orthop Res, 8, 694–701.*

CASE STUDY 5-1

Brachial Plexus Palsy

During the birth process, a newborn suffered a traction injury of his left brachial plexus. A few months later, he presents with the upper left arm in a static position of adduction, internal rotation of the shoulder, extension of the elbow, pronation of the forearm, and flexion of the wrist. He does not respond to sensory stimulus in his shoulder and presents biceps and brachioradialis areflexia. A sudden deformation and high tensile stress injuries in the C5–C6 nerve roots affected the mixed (motor and sensory) neural functions, mainly the muscles responsible for the scapulohumeral rhythm (see Chapter 12).

An Erb's palsy is diagnosed. The sudden elongation suffered during the traction can lead to structural damage and reduction in the transverse fascicular cross sectional area, producing impairment of the intraneural vascular flow and impulse transmission.

In less severe cases, functional restoration may occur within weeks or months. In more severe cases, healing may take place during the first 2 to 3 years, but if the structural nerve injury is severe, considerable long-term functional disability can result. If structural derangement of the nerve trunk has taken place, nerve grafting may be required.

High-energy plexus injuries represent an extreme type of stretching lesion caused by sudden violent trauma. A different stretching situation of considerable clinical interest is the suturing of the two ends of a cut nerve under moderate tension. This situation occurs when a substantial gap exists in the continuity of a nerve trunk and the restoration of the continuity requires the application of tension to bring the nerve ends back together. The moderate, gradual tension applied to the nerve in these cases may stretch and angulate local feeding vessels. It may also be sufficient to reduce the transverse fascicular cross-sectional area and impair the intraneural nutritive capillary flow (Fig. 5-8).

As the sutured nerve is stretched, the perineurium tightens; as a result, the endoneurial fluid pressure is increased and the intrafascicular capillaries may be obliterated. Also, the flow is impaired in the segmental, feeding, and draining vessels, as it is in larger vessels in the epineurium, and at a certain stage the intraneural microcirculation ceases. Intravital observations of intraneural blood flow in rabbit tibial nerves (Lundborg & Rydevik, 1973) showed that an elongation of 8% induced impaired venular flow and that even greater tension produced continuous impairment of capillary and arteriolar flow until, at 15% elongation, all intraneural microcirculation ceased completely. For the same nerve, an elongation (strain) of 6% induced a reduction of nerve action potential amplitude by 70% at 1 hour with recovery to normal values during 1-hour restitution. At 12% elongation, conduction was completely blocked by 1 hour and showed minimal recovery (Wall et al., 1992). Such data have clinical implication in nerve repair, limb trauma, and limb lengthening.

A situation of even more gradual stretching, applied over a long time, is the growth of intraneural tumors such as schwannomas. In this situation, the nerve fibers are forced into a circumferential course around the gradually expanding tumor. Functional changes in cases of such very gradual stretching are often minimal or nonexistent.

COMPRESSION INJURIES OF PERIPHERAL NERVES

It has long been known that compression of a nerve can induce symptoms such as numbness, pain, and muscle weakness. The biological basis for the functional changes has been investigated extensively (Rydevik & Lundborg, 1977; Rydevik et al., 1981). In these investigations (Fig. 5-9), even mild compression was observed to induce structural and functional changes, and the significance of mechanical

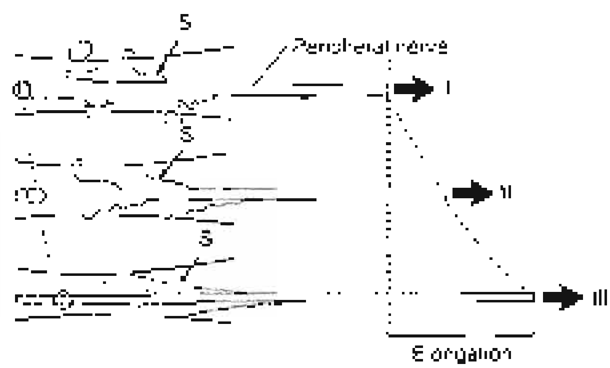


FIG. 5-8

Schematic representation of a peripheral nerve and its blood supply at three stages during stretching. Stage I: The segmental blood vessels (S) are normally coiled to allow for the physiological movements of the nerve. Stage II: Under gradually increasing elongation, these regional vessels become stretched and the blood flow in them is impaired. Stage III: The cross sectional area of the nerve (represented within the circle) is reduced during stretching and the intraneural blood flow is further impaired. Complete cessation of all blood flow in the nerve usually occurs at approximately 15% elongation. *Adapted from Lundborg, G. & Rydevik, B. (1973). Effects of stretching the tibial nerve of the rabbit. A preliminary study of the intraneural circulation and the barrier function of the perineurium. J Bone Joint Surg, 55B, 390.*

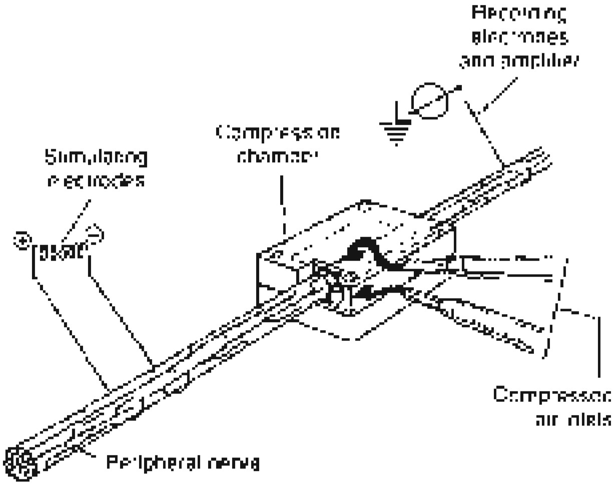


FIG. 5-9

Schematic drawing of an experimental setup for studying deterioration of nerve function during compression. *Adapted from Dahlin, L.B., Rydevik, B., & Lundborg, G. (1986). The pathophysiology of nerve entrapments and nerve compression injuries. In A.R. Hargens (Ed.). Effects of Mechanical Stress on Tissue Viability. New York: Springer Verlag.*

factors such as pressure level and mode of compression became apparent.

Critical Pressure Levels

Experimental and clinical observations have revealed some data on the critical pressure levels at which disturbances occur in intraneural blood flow, axonal transport, and nerve function. Certain pressure levels seem to be well defined with respect to structural and functional changes induced in the nerve. The duration of the compression also influences the development of these changes.

At 30 mm Hg of local compression, functional changes may occur in the nerve, and its viability may be jeopardized during prolonged compression (4 to 6 hours) at this pressure level (Lundborg et al., 1982). Such changes appear to be caused by impairment of the blood flow in the compressed part of the nerve (Rydevik et al., 1981). Corresponding pressure levels (approximately 32 mm Hg) were recorded close to the median nerve in the carpal tunnel in patients with carpal tunnel syndrome, while in a group of control subjects the pressure in the carpal tunnel averaged only 2 mm Hg. Long-standing or intermittent compression at low pressure levels (approximately 30 to 80 mm Hg) may induce intraneural edema, which in turn may become organized into a fibrotic scar in the nerve (Rydevik & Lundborg, 1977).

Compression at approximately 30 mm Hg also brings about changes in the axonal transport systems, and long-standing compression may thus lead to depletion of axonally transported proteins distal to the compression site. Such blockage of axonal transport induced by local compression (pinching) may cause the axons to be more susceptible to additional compression distally, the so-called double crush syndrome.

Slightly higher pressure (80 mm Hg, for example) causes complete cessation of intraneural blood flow; the nerve in the locally compressed segment becomes completely ischemic. Yet, even after 2 hours or more of compression, blood flow is rapidly restored when the pressure is released (Rydevik et al., 1981). Even higher levels of pressure (200 to 400 mm Hg, for example) applied directly to a nerve can induce structural nerve fiber damage and rapid deterioration of nerve function, with incomplete recovery after even shorter periods of compression. Hence, the magnitude of the applied pressure and the severity of the induced compression lesion appear to be correlated.

Mode of Pressure Application

The pressure level is not the only factor that influences the severity of nerve injury brought about by compression. Experimental and clinical evidence indicates that the mode of pressure application is also of major significance. Its importance is illustrated by the fact that direct compression of a nerve at 400 mm Hg by means of a small inflatable cuff around the nerve induces a more severe nerve injury than does indirect compression of the nerve at 1000 mm Hg via a tourniquet applied around the extremity. Even though the hydrostatic pressure acting on the nerve in the former situation is less than half that in the latter, the nerve lesion is more severe, probably because direct compression causes a more pronounced deformation of the nerve (especially at its edges) than does indirect compression, in which the tissue layers between the compression device and the nerve "bolster" the nerve. One may also conclude that the nerve injury caused by compression is not directly related to the high hydrostatic pressure in the center of the compressed nerve segment but instead is more dependent on the specific mechanical deformation induced by the applied pressure.

Mechanical Aspects of Nerve Compression

Electron microscopic analysis of the deformation of the nerve fibers in the peroneal nerve of the baboon hind limb induced by tourniquet compression demonstrated the so-called edge effect; that is, a specific lesion was induced in the nerve fibers at both edges of the compressed nerve segment: the nodes of Ranvier were displaced toward the noncompressed parts of the nerve. The nerve fibers in the center of the compressed segment, where the hydrostatic pressure is highest, generally were not affected acutely. The large-diameter nerve fibers were usually affected, but the thinner fibers were spared. This finding confirms theoretical calculations that indicate larger nerve fibers undergo a relatively greater deformation than do thinner fibers at a given pressure. It is also known clinically that a compression lesion of a nerve first affects the large fibers (e.g., those that carry motor function), while the thin fibers (e.g., those that mediate pain sensation) are often preserved. The intraneural blood vessels have also been shown to be injured at the edges of the compressed segment (Rydevik & Lundborg, 1977). Basically, the lesions of nerve fibers and blood vessels seem to be consequences of the pressure gradient, which is maximal just at the edges of the compressed segment.

In considering the mechanical effects on nerve compression, keep in mind that the effect of a given pressure depends on the way in which it is applied, its magnitude and duration. Although pressure may be applied with a variety of spatial distributions, two basic types of pressure applications are generally encountered in experimental settings and in pathological conditions. One type is uniform pressure applied around the entire circumference of a longitudinal segment of a nerve or extremity. This is the kind of purely radial pressure that is applied by the common pneumatic tourniquet. It has also been used in miniature apparatus to produce controlled compression of individual nerves (Rydevik & Lundborg, 1977) (Fig. 5-9). Clinically, this type of loading on a nerve probably occurs when the pressure on the median nerve is elevated in the carpal tunnel, producing a characteristic syndrome.

Another type of mechanical action takes place when the nerve is compressed laterally. This is the kind of deformation that occurs if a nerve or extremity is placed between two parallel flat rigid surfaces that are then moved toward each other, squeezing the nerve or extremity. This type of deformation occurs if a sudden blow by a rigid object squeezes a nerve against the surface of an underlying bone. It may also occur when a spinal nerve is compressed by a herniated disc (Case Study 5-2).

The details of the deformation of a nerve may be quite different in these two cases of loading. In uniform circumferential compression like that applied

CASE STUDY 5-2

Sciatic Pain

A 35-year-old male construction worker has chronic low back pain radiating below the left knee that is more severe with lifting activities and prolonged postures. After a careful examination, distinct neurologic signs were found. Positive straight leg raising and L5 motor and sensory function were affected.

An MRI shows a herniated disc at level L4-L5 with posterolateral protrusion, which laterally compresses the left L5 nerve root. Compression of the nerve deforms it toward a more elliptical shape, increasing strain and stress loads. The effects of the pressure and mechanical deformation resultant from the load affects the nerve tissue, its nutrition, and the transmission function. Inflammation of the nerve root, induced by the nucleus pulposus, may sensitize the nerve root so that mechanical nerve root deformation causes sciatic pain (Case Study Fig 5-2-1).

Case Study Figure 5-2-1.

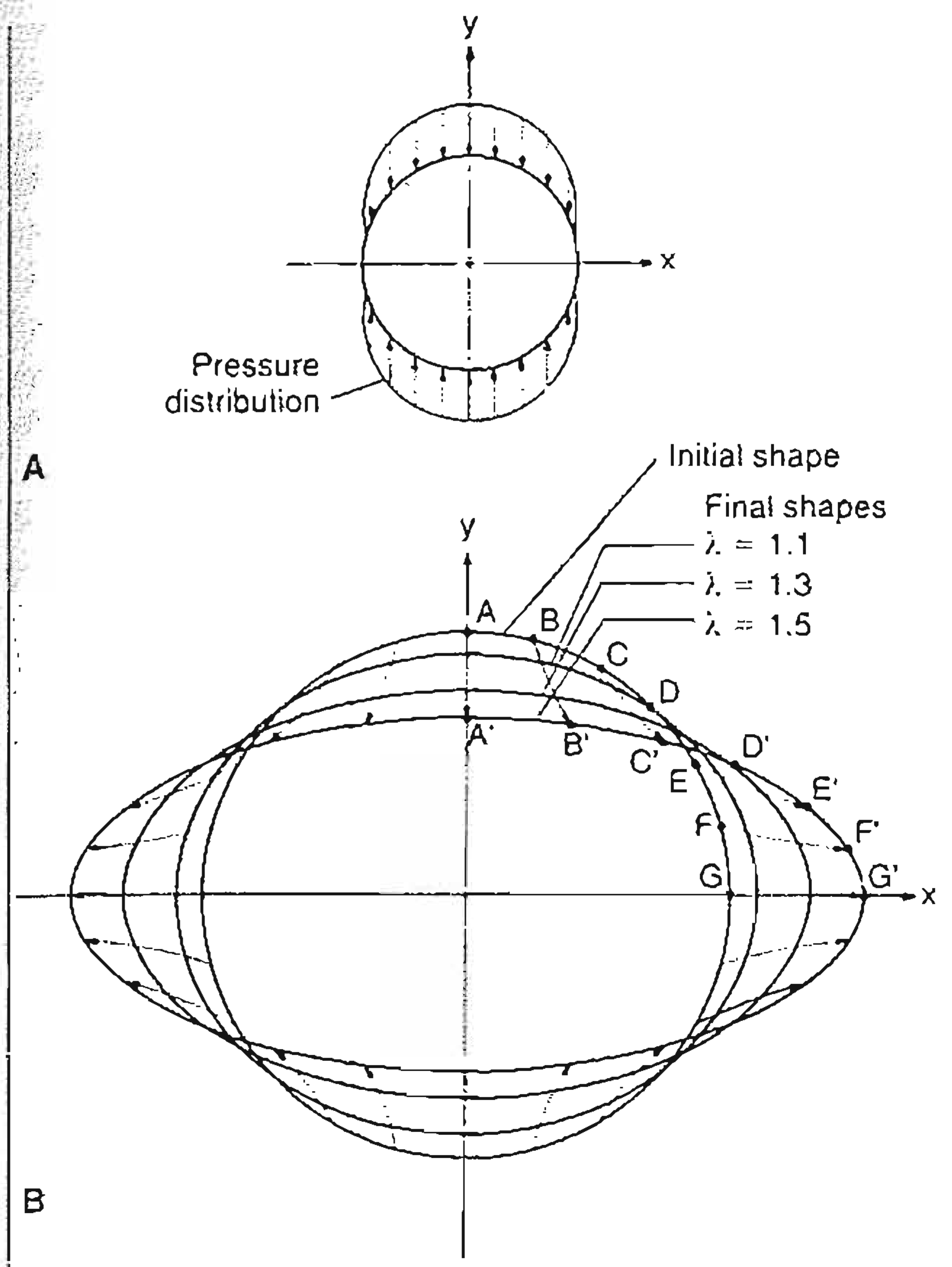


FIG. 5-11

A, Theoretical displacement field under lateral compression as a result of uniform clamping pressure. B, The original and deformed cross-sections are shown for maximum elongation in the x direction of 10, 30, and 50%. The vectors shown from A to A', B to B', and so forth, indicate the paths followed by the particular points A, B, and so forth during the deformation.

that this kind of deformation can trigger firing of nerves, resulting in a sensation of pain when the nerve fibers are laterally compressed. The details of such deformation of nerves and their functional consequences have not been studied extensively and require further research for their elucidation.

Duration of Pressure Versus Pressure Level

Knowledge is limited regarding the relative importance of pressure and time, respectively, in the production of nerve compression lesions. Mechanical factors seem to be relatively more important at higher than at lower pressures. Time is a significant factor at both high and low pressures, but ischemia plays a dominant role in longer-duration compression. This phenomenon is illustrated by the fact that direct nerve compression at 30 mm Hg for 2 to 4 hours produces reversible changes, whereas prolonged compression above this time period at this pressure level may cause irreversible damage to the nerve (Lundborg et al., 1982; Rydevik et al., 1981). Compression at 400 mm Hg causes a much more severe nerve injury after 2 hours than after 15 minutes. Such information indicates that even high pressure has to "act" for a certain period of time for injury to occur. These data also give some information about the viscoelastic (time-dependent) properties of peripheral nerve tissue. Sufficient time must elapse for permanent deformation to develop.

Biomechanical Behavior of Spinal Nerve Roots

The nerve roots in the thecal sac lack epineurium and perineurium, but under tensile loading they exhibit both elasticity and tensile strength. The ultimate load for ventral spinal nerve roots from the thecal sac is between 2 and 22 N, and for dorsal nerve roots from the thecal sac the load is between 5 and 33 N. The length of the nerve roots from the spinal cord to the foramina varies from approximately 60 mm at the L1 level to approximately 170 mm at the S1 level. The mechanical properties of human spinal nerve roots are different for any given nerve root at its location in the central spinal canal and in the lateral intervertebral foramina. The ultimate load for the intrathecal portion of human S1 nerve roots at the S1 level is approximately 13 N, and that for the foraminal portion is approximately 73 N. For human nerve roots at the L5 level, the corresponding values are 16 N and 71 N, respectively (Fig. 5-12). Thus, the values for ultimate load are approximately five times higher for the foraminal segment of the spinal nerve roots than for the intrathecal portion of the same nerve roots under tensile loading. However, the cross-sectional area of the nerve root in the intervertebral foramen is significantly larger than that of the same nerve root in the thecal sac; thus, the ultimate tensile stress was more comparable for the two locations. The ultimate strain under tensile loading is 13 to 19% for the human nerve root at the L5–S1 level (Fig. 5-13).

The nerve roots in the spine are not static structures; they move relative to the surrounding tissues

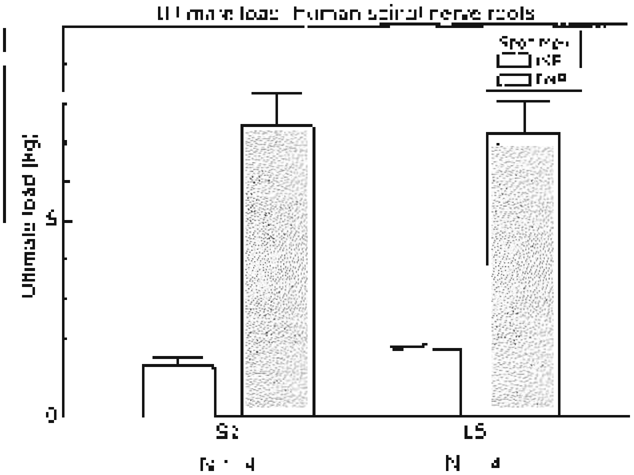


FIG. 5-12

Diagram illustrating values for ultimate load obtained for human spinal nerve roots under tensile loading. INR, intrathecal nerve roots; FNR, foraminal nerve root. Note the marked difference in ultimate load for the intrathecal and the foraminal portions of the nerve roots. Error bars indicate standard deviation. *Reproduced with permission from Weinstein, J.N., LaMotte, R., Rydevik, B., et al. (1989). Nerve. In J.W. Frymoyer & S.L. Gordon (Eds.), New Perspectives on Low Back Pain (Chapter 4, pp. 35–130). Park Ridge, IL: AAOS. (Based on a workshop arranged by the National Institutes of Health (NIH) in Airlie, Virginia, USA, May 1988.)*

with every spinal motion. To allow for such motion the nerve roots in the intervertebral foramina, for example, must have the capacity to glide. Chronic irritation with subsequent fibrosis around the nerve roots, in association with conditions such as disc herniation and/or foraminal stenosis, can thus impair the gliding capacity of the nerve roots. This produces repeated "microstretching" injuries of the nerve roots even during normal spinal movements, which might be speculated to induce yet further tissue irritation in the nerve root components. The normal range of movements of nerve roots in the human lumbar spine has been measured in cadaver experiments. It was found that straight leg raising moved the nerve roots at the level of the intervertebral foramina approximately 2 to 5 mm.

Certain biomechanical factors are obviously involved in the pathogenesis of various symptoms induced by nerve root deformation in association with disc herniation and spinal stenosis and resulting in radiating pain. In disc herniation, only one nerve root is usually compressed. Because individual nerve roots normally adhere to the surrounding tissues above and below the intervertebral disc they traverse, compression may give rise to intraneural tension. Spencer and associates (1984) measured the contact force between a simulated disc herniation and a deformed nerve root in cadavers. Taking the area of contact into account, they assumed a contact pressure of approximately 400 mm Hg. With reduced disc height, the contact force and pressure between the experimental disc herniation and the nerve root was reduced. They suggested that these findings may explain in part why sciatic pain is relieved after chemonucleolysis, and as disc degeneration progresses over time and the disc height thereby decreases.

In central spinal stenosis, the mechanics of nerve root compression are completely different. Under these conditions, the pressure is applied circumferentially around the nerve roots in the cauda equina at a slow, gradual rate. These different deformation factors, together with the fact that the nerve roots centrally within the cauda equina differ completely from the nerve roots located more laterally, close to the discs, may explain some of the different symptoms found in spinal stenosis and disc herniation.

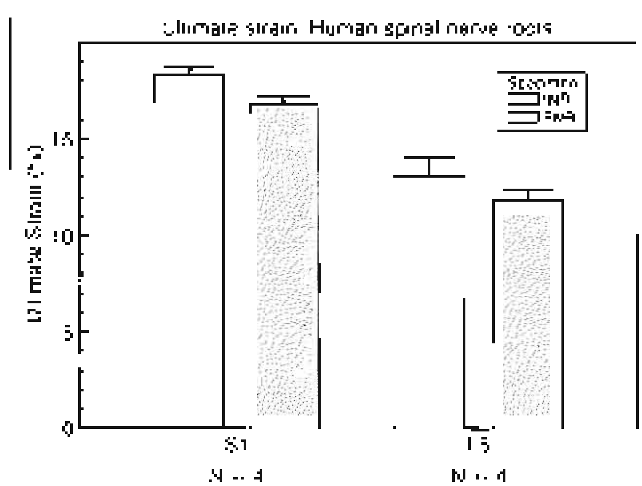


FIG. 5-13

Ultimate strain for human spinal nerve roots under tensile loading. INR, intrathecal nerve root; FNR, foraminal nerve root. *Reproduced with permission from Weinstein, J.N., LaMotte, R., Rydevik, B., et al. (1989). Nerve. In J.W. Frymoyer & S.L. Gordon (Eds.), New Perspectives on Low Back Pain (Chapter 4, pp. 35–130). Park Ridge, IL: AAOS. (Based on a workshop arranged by the National Institutes of Health (NIH) in Airlie, Virginia, USA, May 1988.)*

EXPERIMENTAL COMPRESSION OF SPINAL NERVE ROOTS

There has been moderate interest in the past to study nerve root compression in experimental models. Early studies in the 1950s and 1970s found that nerve roots seemed to be more susceptible to compression than did peripheral nerves. During recent years, however, the interest in nerve root pathophysiology has increased considerably and a number of studies have been performed that are reviewed below.

Some years ago, a model was presented to evaluate the effects of compression of the cauda equina in pigs, which for the first time allowed for experimental, graded compression of cauda equina nerve roots at known pressure levels (Olmarker, 1991) (Fig. 5-14). In this model, the cauda equina was compressed by an inflatable balloon that was fixed to the spine. The cauda equina could also be observed through the translucent balloon. This model made it possible to study the flow in the intrinsic nerve root blood vessels at various pressure levels (Olmarker et al., 1989a). The experiment was designed in a way that the pressure in the compression balloon was increased by 5 mm Hg every 20 seconds. Blood flow and vessel diameters of the intrinsic vessels could simultaneously be observed through the balloon using a vital microscope. The average occlusion pressure for the arterioles was found to be slightly below and directly related to the systolic blood pressure, and the blood flow in the capillary networks was intimately dependent on the blood flow of the adjacent venules. This corroborates the assumption that venular stasis may induce capillary stasis and thus changes in the microcirculation of the nerve tissue and is in accordance with previous studies in which such a mechanism has been suggested as involved in carpal tunnel syndrome. The mean occlusion pressures for the venules demonstrated large variations. However, a pressure of 5 to 10 mm Hg was found to be sufficient for inducing venular occlusion. Because of retrograde stasis, it is not unlikely to assume that the capillary blood flow will be affected as well in such situations.

In the same experimental set-up, the effects of gradual decompression, after initial acute compression was maintained for only a short while, were studied. The average pressure for starting the blood flow was slightly lower at decompression than at compression for arterioles, capillaries, and venules.

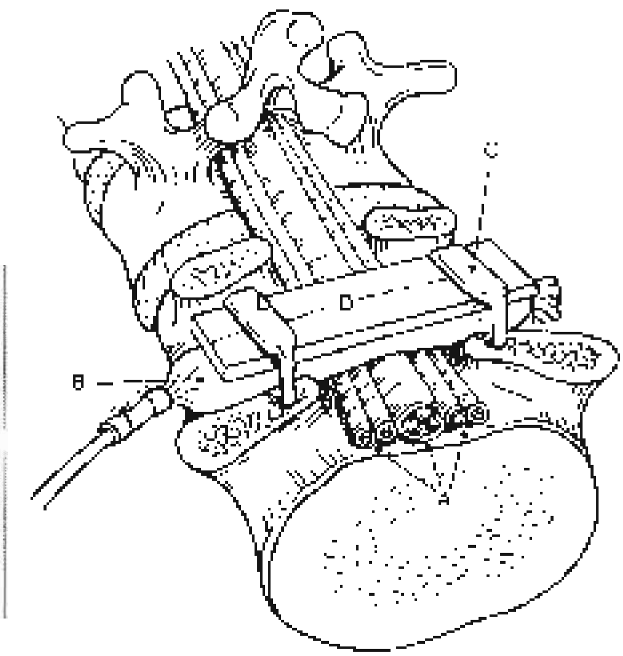


FIG. 5-14

Schematic drawing of an experimental model. The cauda equina (A) is compressed by an inflatable balloon (B) that is fixed to the spine by two L-shaped pins (C) and a plexiglass plate (D). *Reproduced with permission from Olmarker, K., Rydevik, B., & Holm, S. (1989a). Edema formation in spinal nerve roots induced by experimental, graded compression. An experimental study on the pig cauda equina with special reference to differences in effects between rapid and slow onset of compression. Spine, 14, 579.*

However, with this protocol a full restoration of the blood flow did not occur until the compression was lowered from 5 to 0 mm Hg. This observation further supports the previous hypothesis that vascular impairment is present even at low pressure levels.

A compression-induced impairment of the vasculature may thus be one mechanism for nerve root dysfunction because the nutrition of the nerve root will be affected. However, the nerve roots will also derive a considerable nutritional supply via diffusion from the cerebrospinal fluid. To assess the compression-induced effects on the total contribution to the nerve roots, an experiment was designed in which 3H-labeled methyl-glucose was allowed to be transported to the nerve tissue in the compressed segment via both the blood vessels and the cerebrospinal fluid diffusion after systemic injection. The results showed that no compensatory mechanism from cerebrospinal fluid diffusion could be expected at the low pressure levels. On the contrary, 10 mm Hg compression was sufficient to induce a 20 to 30% reduction of the transport of methyl-glucose to the nerve roots, as compared with the control.

We know from experimental studies on peripheral nerves that compression may also induce an increase in the vascular permeability, leading to an intraneural edema formation. Such edema may increase the endoneurial fluid pressure, which in turn may impair the endoneurial capillary blood flow and jeopardize the nutrition of the nerve roots. Because the edema usually persists for some time after the removal of a compressive agent, edema may negatively affect the nerve root for a longer period than the compression itself. The presence of intraneural edema is also related to the subsequent formation of intraneural fibrosis and may therefore contribute to the slow recovery seen in some patients with nerve compression disorders. To assess if intraneural edema also may form in nerve roots as the result of compression, the distribution of Evan's blue-labeled albumin in the nerve tissue was analyzed after compression at various pressures and at various durations (Olmarker et al., 1989b). The study showed that edema was formed even at low pressure levels. The predominant location was at the edges of the compression zone.

The function of the nerve roots has been studied by direct electrical stimulation and recordings either on the nerve itself or in the corresponding muscular segments. During a 2-hour compression period, a critical pressure level for inducing a reduction of MAP-amplitude seems to be located between 50 and 75 mm Hg. Higher pressure levels (100–200 mm Hg) may induce a total conduction block with varying degrees of recovery after compression release. To study the effects of compression on sensory nerve fibers, electrodes in the sacrum were used to record a compound nerve action potential after stimulating the sensory nerves in the tail, that is, distal to the compression zone. The results showed that the sensory fibers are slightly more susceptible to compression than are the motor fibers. Also, the nerve roots are more susceptible to compression injury if the blood pressure is lowered pharmacologically. This further indicates the importance of the blood supply to maintain the functional properties of the nerve roots.

ONSET RATE OF COMPRESSION

One factor that has not been fully recognized in compression trauma of nerve tissue is the onset rate of the compression. The onset rate, that is, the time from start to full compression, may vary clinically from fractions of seconds in traumatic conditions to months or years in association with degenerative processes. Even in the clinically rapid onset rates, there may be a wide variation of onset rates. With the presented model, it was possible to vary the onset time of the applied compression. Two onset rates have been investigated. Either the pressure is preset and compression is started by flipping the switch of the compressed-air system used to inflate the balloon or the compression pressure level is slowly increased during 20 seconds. The first onset rate was measured at 0.05 to 0.1 seconds, thus providing a rapid inflation of the balloon and a rapid compression onset.

Such a rapid-onset rate has been found to induce more pronounced effects on edema formation, methyl-glucose transport, and impulse propagation than the slow-onset rate (Olmarker, 1991). Regarding methyl-glucose transport, the results show that the levels within the compression zone are more pronounced at the rapid than at the slow onset rate at corresponding pressure levels. There was also a striking difference between the two onset rates when considering the segments outside the compression zones. In the slow onset series, the levels approached baseline values closer to the compression zone than in the rapid onset series. This may indicate the presence of a more pronounced edge-zone edema in the rapid-onset series, with a subsequent reduction of the nutritional transport in the nerve tissue adjacent to the compression zone.

For the rapid-onset compression, which is likely to be more closely related to spine trauma or disc herniation than to spinal stenosis, a pressure of 600 mm Hg maintained for only 1 second is sufficient to induce a gradual impairment of nerve conduction during the 2 hours studied after the compression was ended. Overall, the mechanisms for these pronounced differences between the different onset rates are not clear but may be related to differences in the displacement rates of the compressed nerve tissue toward the uncompressed parts, as a result of the viscoelastic properties of the nerve tissue. Such phenomena may lead not only to structural damage to the nerve fibers but also to structural changes in the blood vessels with subsequent edema formation.

The gradual formation of intraneural edema may also be closely related to observations of a gradually increasing difference in nerve conduction impairment between the two onset rates (Olmarker et al., 1989b).

MULTIPLE LEVELS OF SPINAL NERVE ROOT COMPRESSION

Patients with double or multiple levels of spinal stenosis seem to have more pronounced symptoms than do patients with stenosis only at one level. The presented model was modified to address this interesting clinical question. Using two balloons at two adjacent disc levels, which resulted in a 10-mm uncompressed nerve segment between the balloons, induced a much more pronounced impairment of nerve impulse conduction than previously had been found at corresponding pressure levels (Olmarker & Rydevik, 1992). For instance, a pressure of 10 mm Hg in two balloons induced a 60% reduction of nerve impulse amplitude during 2 hours of compression, whereas 50 mm Hg in one balloon showed no reduction.

The mechanism for the difference between single and double compression may not simply be based on the fact that the nerve impulses have to pass more than one compression zone at double-level compression. There may also be a mechanism based on the local vascular anatomy of the nerve roots. Unlike for peripheral nerves, there are no regional nutritive arteries from the surrounding structures to the intraneural vascular system in spinal nerve roots. Compression at two levels might therefore induce a nutritionally impaired region between the two compression sites. In this way, the segment affected by the compression would be widened from one balloon diameter (10 mm) to two balloon diameters including the interjacent nerve segment (30 mm). This hypothesis was partly confirmed in an experiment on continuous analyses of the total blood flow in the uncompressed nerve segment located between two compression balloons (Takahashi et al., 1993). The results showed that a 64% reduction of total blood flow in the uncompressed segment was induced when both balloons were inflated to 10 mm Hg. At a pressure close to the systemic blood pressure there was complete ischemia in the nerve segment. Thus, experimental evidence shows that the blood supply to the nerve segment located between two compression sites in nerve roots is severely impaired although this nerve segment itself is uncompressed. Regarding nerve conduction, the effects were much enhanced if the distance between the compression balloons was increased from one vertebral segment to two vertebral segments (Olmarker & Rydevik, 1992). This indicates that the functional impairment may be directly related to the distance between the two compression sites.

CHRONIC NERVE ROOT COMPRESSION IN EXPERIMENTAL MODELS

The discussion of compression-induced effects on nerve roots has dealt primarily with acute compression, that is, compression that lasts for some hours and with no survival of the animal. To better mimic various clinical situations, compression must be applied over longer periods of time. There are probably many changes in the nerve tissue, such as adaptation of axons and vasculature, that will occur in patients but cannot be studied in experimental models using only 1 to 6 hours of compression. Another important factor in this context is the onset rate that was discussed previously. In clinical syndromes with nerve root compression, the onset time may in many cases be quite slow. For instance, a gradual remodeling of the vertebrae to induce a spinal stenosis probably leads to an onset time of many years. It will of course be difficult to mimic such a situation in an experimental model. It will also be impossible to have control over the pressure acting on the nerve roots in chronic models because of the remodeling and adaptation of the nerve tissue to the applied pressure. However, knowledge of the exact pressures is probably of less importance in chronic than in acute compression situations. Instead, chronic models should induce a controlled compression with a slow onset time that is easily reproducible. Such models may be well suited for studies on pathophysiological events as well as intervention by surgery or drugs. Some attempts have been made to induce such compression.

Delamarter and collaborators (1990) presented a model on dog cauda equina in which they applied a constricting plastic band. The band was tightened around the thecal sac to induce a 25, 50, or 75% reduction of the cross-sectional area. The band was left in place for various times. Analyses were performed and showed both structural and functional changes that were proportional to the degree of constriction.

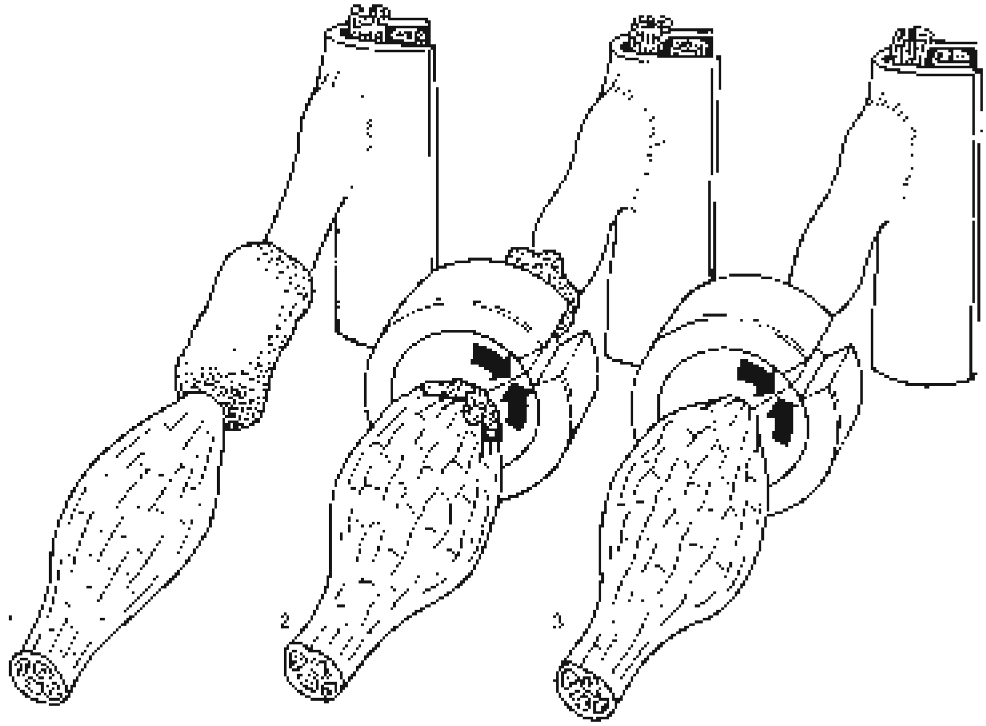


FIG. 5-15

Experimental study to analyze the effects on nerve conduction velocity of nucleus pulposus (1), the combination of nucleus pulposus and compression (2), and compression only (3). The nucleus pulposus and the constrictor were applied to the first sacral nerve root in pigs. The contralateral nerve root served as a control. *Reproduced with permission from Cornefjord, M., Sato, K., Olmarker, K., et al. (1997). A model for chronic nerve root compression studies. Presentation of a porcine model for controlled, slow-onset compression with analyses of anatomic aspects, compression onset rate, and morphologic and neurophysiologic effects. Spine, 22, 946–957.*

To induce a slower onset and more controlled compression, Cornefjord and collaborators (1997) used a constrictor to compress the nerve roots in the pig (Fig. 5-15). The constrictor was initially intended for inducing vascular occlusion in experimental ischemic conditions in dogs. The constrictor consists of an outer metal shell that on the inside is covered with a material called ameroid that expands when in contact with fluids. Because of the metal shell, the ameroid expands inwards with a maximum expansion after 2 weeks, resulting in compression of a nerve root placed in the central opening of the constrictor. Compression of the first sacral nerve root in the pig resulted in a significant reduction of nerve conduction velocity and axonal injuries using a constrictor with a defined original diameter. An increase in substance P in the nerve root and the dorsal root ganglion following such compression also has been found. Substance P is a neurotransmitter that is related to pain transmission. The study may thus provide experimental evidence that compression of nerve roots produces pain.

The constrictor model has also been used to study blood flow changes in the nerve root vasculature. It could then be observed that the blood flow is not reduced just outside the compression zone but significantly reduced in parts of the nerve roots located inside the constrictor. In this context, note that in case of disc herniation, the nerve root may become sensitized by substances from the disc tissue (nucleus pulposus) so that mechanical root deformation can induce pronounced sciatic pain.

Summary

1 The peripheral nerves are composed of nerve fibers, layers of connective tissue, and blood vessels.

2 The nerve fibers are extremely susceptible to trauma but because they are surrounded by successive layers of connective tissue (the epineurium and perineurium), they are mechanically protected.

3 Stretching induces changes in intraneural blood flow and nerve fiber structure before the nerve trunk ruptures.

4 Compression of a nerve can cause injury to both nerve fibers and blood vessels in the nerve, mainly at the edges of the compressed nerve segment, but also by ischemic mechanisms.

5 Pressure level, duration of compression, and mode of pressure application are significant variables in the development of nerve injury.

6 Spinal nerve roots are anatomically different from peripheral nerves and therefore react differently to mechanical deformation.

7 Spinal nerve roots are more susceptible than peripheral nerves to mechanical deformation, mainly because of the lack of protective connective tissue layers in nerve roots.

REFERENCES

Cornefjord, M., Sato, K., Olmarker, K., et al. (1997). A model for chronic nerve root compression studies. Presentation of a porcine model for controlled slow-onset compression with analyses of anatomic aspects, compression onset rate, and morphologic and neurophysiologic effects. *Spine, 22*, 946–957.

Dahlin, L.B., Rydevik, B., & Lundborg, G. (1986). The pathophysiology of nerve entrapments and nerve compression injuries. In A.R. Hargens (Ed.), *Effects of Mechanical Stress on Tissue Viability*. New York: Springer-Verlag.

Delamarter, R.B., Bohlman, H.H., Dodge, L.D., et al. (1990). Experimental lumbar spinal stenosis. Analysis of the cortical evoked potentials, microvasculature and histopathology. *J Bone Joint Surg, 72A*, 110–120.

Lundborg, G., & Rydevik, B. (1973). Effects of stretching the tibial nerve of the rabbit. A preliminary study of the intraneural circulation and the barrier function of the perineurium. *J Bone Joint Surg, 55B*, 390.

Lundborg, G. (1975). Structure and function of the intraneural microvessels as related to trauma, edema formation and nerve function. *J Bone Joint Surg, 57A*, 938.

Lundborg, G., et al. (1982). Median nerve compression in the carpal tunnel: The functional response to experimentally induced controlled pressure. *J Hand Surg, 7*, 252.

Myers, R.R. (1998). Morphology of the peripheral nervous system and its relationship to neuropathic pain. In T.L. Yaksh, C. Lynch III, W.M. Zapol, M. Maze, J.F. Biebuyck, & L.J. Saidman (Eds.), *Anesthesia: Biologic Foundations* (pp. 483–514). Philadelphia: Lippincott-Raven.

Myers, R.R., Murakami, H., & Powell, H.C. (1986). Reduced nerve blood flow in edematous neuropathies—A biomechanical mechanism. *Microvasc Res, 32*, 145–151.

Myers, R.R., & Powell, H.C. (1981). Endoneurial fluid pressure in peripheral neuropathies. In A.R. Hargens (Ed.), *Tissue Fluid Pressure and Composition* (p. 193). Baltimore: Williams & Wilkins.

Olmarker, K., Rydevik, B., & Holm, S. (1989a). Edema formation in spinal nerve roots induced by experimental, graded compression. An experimental study on the pig cauda equina with special reference to differences in effects between rapid and slow onset of compression. *Spine, 14*, 569.

Olmarker, K., Rydevik, B., Holm, S., et al. (1989b). Effects of experimental graded compression on blood flow in spinal nerve roots. A vital microscopic study on the porcine cauda equina. *J Orthop Res, 7*, 817.

Olmarker, K. (1991). Spinal nerve root compression. Acute compression of the cauda equina studied in pigs. *Acta Orthop Scand, 62* (Suppl 242).

Olmarker, K., & Rydevik, B. (1992). Single- versus double-level compression. An experimental study on the porcine cauda equina with analyses of nerve impulse conduction properties. *Clin Orthop, 279*, 35–39.

Olmarker, K., & Hasue, M. (1995). Classification and pathophysiology of spinal pain syndromes. In J.N. Weinstein & B. Rydevik (Eds.), *Essentials of the Spine*. Raven Press, New York, NY.

Rydevik, B.L., Kwan, M.K., Myers, R.R., et al. (1990). An in vitro mechanical and histological study of acute stretching on rabbit tibial nerve. *J Orthop Res, 8*, 694–701.

Rydevik, B., & Lundborg, G. (1977). Permeability of intraneural microvessels and perineurium following acute, graded experimental nerve compression. *Scand J Plast Reconstr Surg, 11*, 179.

Rydevik, B., Lundborg, G., & Bagge, U. (1981). Effects of graded compression on intraneural blood flow. An in vivo study on rabbit tibial nerve. *J Hand Surg, 6*, 3.

Rydevik, B., Brown, M.D., & Lundborg, G. (1984). Pathoanatomy and pathophysiology of nerve root compression. *Spine, 9*, 7.

Rydevik, B.L., Kwan, M.K., Myers, R.R., et al. (1990). An in vitro mechanical and histological study of acute stretching on rabbit tibial nerve. *J Orthop Res, 8*, 694.

Spencer, D.L., Miller, J.A., & Bertolini, J.E. (1984). The effect of intervertebral disc space narrowing on the contact force between the nerve root and a simulated disc protrusion. *Spine, 9*, 422.

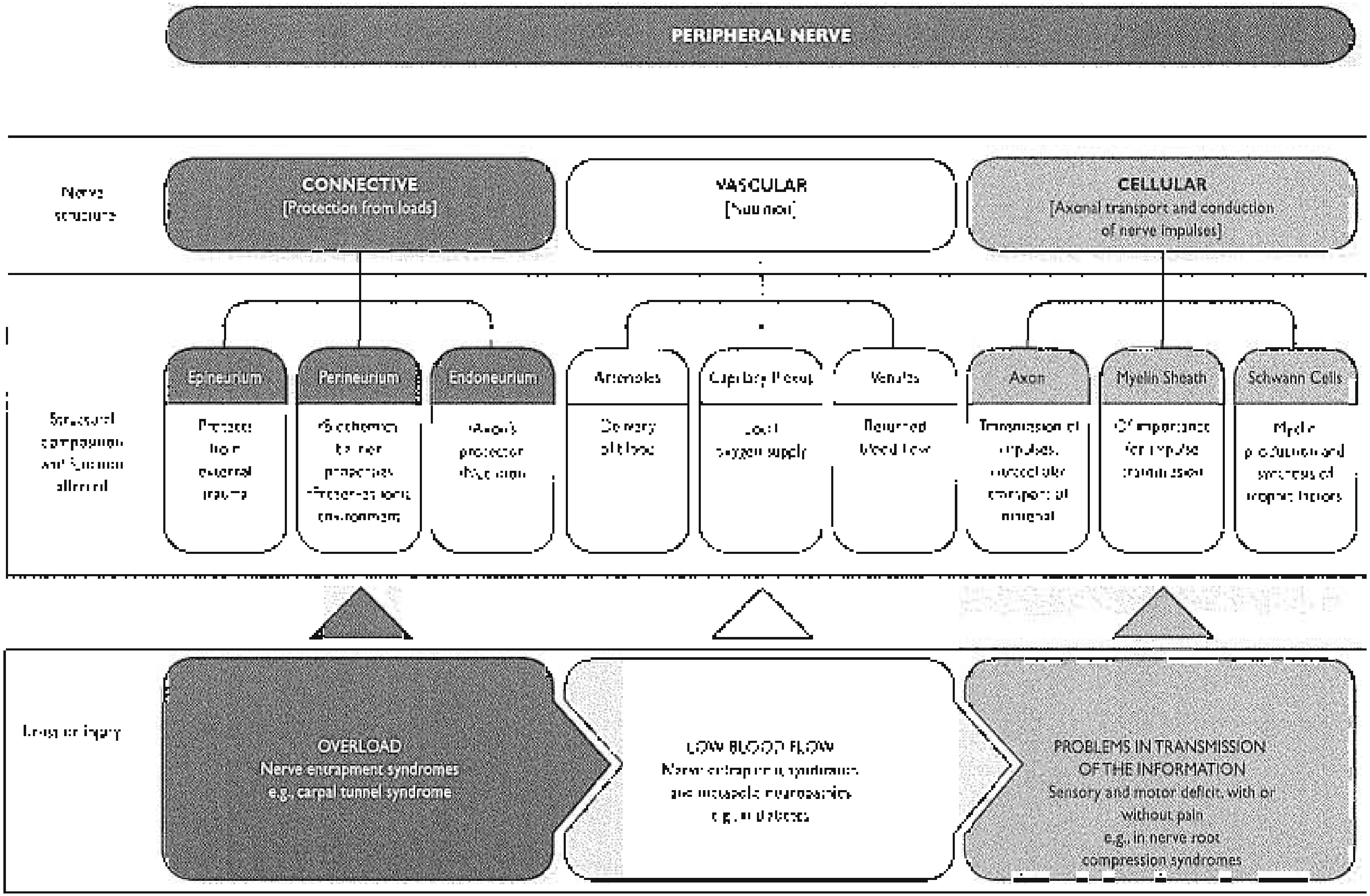
Sunderland, S. (1978). *Nerves and Nerve Injuries* (2nd ed.). Edinburgh: Churchill Livingstone.

Takahashi, K., Olmarker, K., Holm, S., et al. (1993). Double-level cauda equina compression. An experimental study with continuous monitoring of intraneural blood flow. *J Orthop Res, 11*, 104.

Tortora, G.J., & Anagnostakos, N.P. (1984). *Principles of Anatomy and Physiology* (4th ed.). New York: Harper & Row.

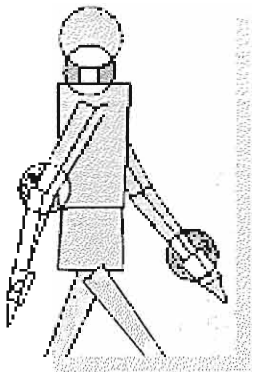
Wall, E.J., Massie, J.B., Kwan, M.K., et al. (1992). Experimental stretch neuropathy. Changes in nerve conduction under tension. *J Bone Joint Surg, 74B*, 126.

Weinstein, J.N., LaMotte, R., Rydevik, B., et al. (1989). Nerve. In J.W. Frymoyer & S.L. Gordon (Eds.), *New Perspectives on Low Back Pain* (Chapter 3, pp. 35–130). Park Ridge, IL: AAOS. [Based on a workshop arranged by the National Institutes of Health (NIH) in Airlie, Virginia, USA, May 1988.]



FLOW CHART 5-1. Peripheral nerve's structure and alteration. Clinical examples.*

* This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.



CHAPTER 6

Biomechanics of Skeletal Muscle

Tobias Lorenz, Marco Campello adapted from Mark I. Pitman, Lars Peterson

Introduction

The muscular system consists of three muscle types: the cardiac muscle, which composes the heart; the smooth (nonstriated or involuntary) muscle, which lines the hollow internal organs; and the skeletal (striated or voluntary) muscle, which attaches to the skeleton via the tendons. The focus of this chapter is the role and function of skeletal muscle.

Skeletal muscle is the most abundant tissue in the human body, accounting for 40 to 45% of the total body weight. The human body has more than 430 skeletal muscles, found in pairs on the right and left sides of the body. The most vigorous movements are produced by fewer than 80 pairs. The muscles provide strength and protection to the skeleton by distributing loads and absorbing shock. They enable the bones to move at the joints and provide the maintenance of body posture against force. Such abilities usually represent the action of muscle groups, not of individual muscles.

The skeletal muscles perform both dynamic and static work. Dynamic work permits locomotion and the positioning of the body segments in space. Static work maintains body posture or position. In this chapter we describe the composition and structure of skeletal muscle, the mechanics of muscle contraction, force production in muscle, muscle fiber differentiation, and muscle remodeling.

Composition and Structure of Skeletal Muscle

An understanding of the biomechanics of muscle function requires knowledge of the gross anatomical structure and function of the musculotendinous unit and the basic microscopic structure and chemical composition of the muscle fiber.

STRUCTURE AND ORGANIZATION OF MUSCLE

The structural unit of skeletal muscle is the muscle fiber, a long cylindrical cell with many hundreds of nuclei. Muscle fibers range in thickness from approximately 10 to 100 μm and in length from approximately 1 to 30 cm. A muscle fiber consists of many myofibrils, which are invested by a delicate plasma membrane called the sarcolemma. The sarcolemma is connected via vinculin- and dystrophin-rich costameres with the sarcomeric Z lines, which represent a part of the extramyofibrillar cytoskeleton. The myofibril is made up of several sarcomeres that contain thin (actin), thick (myosin), elastic (titin), and inelastic (nebulin) filaments. Actin and myosin are the contractile part of the myofibrils, whereas titin and nebulin are part of the intramyofibrillar cytoskeleton (Stromer et al., 1998). The myofibrils are the basic unit of contraction.

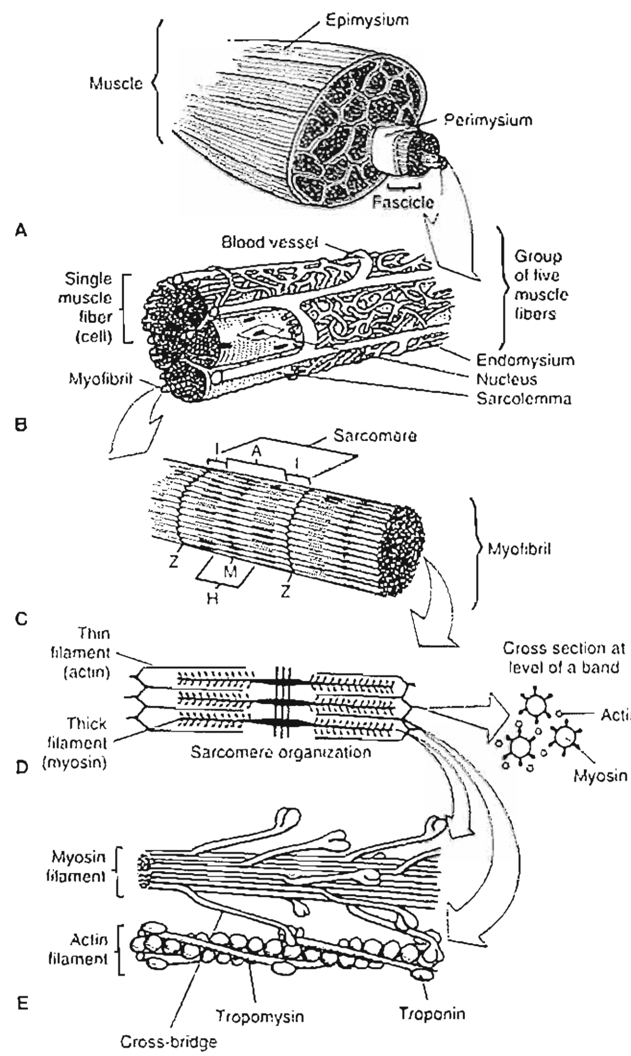
Each fiber is encompassed by a loose connective tissue called the endomysium and the fibers are organized into various-sized bundles, or fascicles (Fig. 6-1, *A* & *B*), which are in turn encased in a dense connective tissue sheath known as the perimysium. The muscle is composed of several fascicles surrounded by a fascia of fibrous connective tissue called the epimysium.

In general, each end of a muscle is attached to bone by tendons, which have no active contractile properties. The muscles form the contractile component and the tendons the series elastic component. The collagen fibers in the perimysium and epimysium are continuous with those in the tendons; together these fibers act as a structural framework for the attachment of bones and muscle fibers. The perimysium, endomysium, epimysium and sarcolemma act as parallel elastic components. The forces produced by the contracting muscles are transmitted to bone through these connective tissues and tendons (Kasser, 1996).

Each muscle fiber is composed of a large number of delicate strands, the myofibrils. These are the contractile elements of muscle. Their structure and function have been studied exhaustively by light and electron microscopy, and their histochemistry and biochemistry have been explained elsewhere (Arvidson et al., 1984; Guyton, 1986). Approximately 1 μm in diameter, the myofibrils lie parallel to each other within the cytoplasm (sarcoplasm) of the muscle fiber and extend throughout its length. They vary in number from a few to several thousand depending on the diameter of the muscle fiber, which depends in turn on the type of muscle fiber.

The transverse banding pattern in striated muscles repeats itself along the length of the muscle fiber, each repeat being known as a sarcomere (Fig. 6-1C). These striations are caused by the individual myofibrils, which are aligned continuously throughout the muscle fiber. The sarcomere is the functional unit of the contractile system in muscle, and the events that take place in one sarcomere are duplicated in the others. Various sarcomere build a myofibril, various myofibrils build the muscle fiber, and various muscle fibers build the muscle.

FIG. 6-1



Schematic drawings of the structural organization of muscle. A, A fibrous connective tissue fascia, the epimysium, surrounds the muscle, which is composed of many bundles, or fascicles. The fascicles are encased in a dense connective tissue sheath, the perimysium. B, The fascicles are composed of muscle fibers, which are long, cylindrical, multinucleated cells. Between the individual muscle fibers are capillary blood vessels. Each muscle fiber is surrounded by a loose connective tissue called the endomysium. Just beneath the endomysium lies the sarcolemma, a thin elastic sheath with infoldings that invaginate the fiber interior. Each muscle fiber is composed of numerous delicate strands—myofibrils, the contractile elements of muscle. C, Myofibrils consist of smaller filaments that form a repeating banding pattern along the length of the myofibril. One unit of this serially repeating pattern is called a sarcomere. The sarcomere is the functional unit of the contractile system of muscle. D, The banding pattern of the sarcomere is formed by the organization of thick and thin filaments, composed of the proteins myosin and actin, respectively. The actin filaments are attached at one end but are free along their length to interdigitate with the myosin filaments. The thick filaments are arranged in a hexagonal fashion. A cross-section through the area of overlap shows the thick filaments surrounded by six equally spaced thin filaments. E, The lollipop-shaped molecules of each myosin filament are arranged so that the long tails form a sheaf with the heads, or cross-bridges, projecting from it. The cross-bridges point in one direction along half of the filament and in the other direction along the other half. Only a portion of one half of a filament is shown here. The cross-bridges are an essential element in the mechanism of muscle contraction, extending outward to interdigitate with receptor sites on the actin filaments. Each actin filament is a double helix, appearing as two strands of beads spiraling around each other. Two additional proteins, tropomyosin and troponin, are associated with the actin helix and play an important role in regulating the interdigitation of the actin and myosin filaments. Tropomyosin is a long polypeptide chain that lies in the grooves between the helices of actin. Troponin is a globular molecule attached at regular intervals to the tropomysin. *Adapted from Williams, P. & Warwick, R. (1980). Gray's Anatomy (36th ed., pp. 506–515). Edinburgh: Churchill Livingstone*

Each sarcomere is composed of the following:

1. The thin filaments (approximately 5 nm in diameter) composed of the protein actin
2. The thick filaments (approximately 15 nm in diameter) composed of the protein myosin (Fig. 6-1, D & E)
3. The elastic filaments composed of the protein titin (Fig. 6-2)
4. The inelastic filaments composed of the proteins nebulin and titin

Actin, the chief component of the thin filament, has the shape of a double helix and appears as two strands of beads spiraling around each other. Two additional proteins, troponin and tropomyosin, are important constituents of the actin helix because they appear to regulate the making and breaking of contacts between the actin and myosin filaments during contraction. Tropomyosin is a long polypeptide chain that lies in the grooves between the helices of actin. Troponin is a globular molecule attached at regular intervals to the tropomyosin (Fig. 6-1, *D* & *E*).

The thick filaments are located in the central region of the sarcomere, where their orderly, parallel arrangement gives rise to dark bands known as A bands because they are strongly anisotropic. The thin filaments are attached at either end of the sarcomere to a structure known as the Z line, which consists of short elements that link the thin filaments of adjacent sarcomeres, defining the limits of each sarcomere. The thin filaments extend from the Z line toward the center of the sarcomere, where they overlap with the thick filaments. Recently it was shown that there is a third set of myofibril filaments in the vertebrate striated muscles. This connecting filament, named titin, links the thick filaments with the Z line (elastic I band region of titin) and is part of the thick filaments (A band region of titin). This filament maintains the central position of the A band throughout contraction and relaxation and might act as a template during myosin assembly.

Myosin, the thicker filament, is composed of individual molecules, each of which resembles a lollipop with a globular "head" projecting from a long shaft or "tail." Several hundred such molecules are packed tail to tail in a sheaf with their heads pointed in one direction along half of the filament and in the opposite direction along the other half, leaving a head-free region (the H zone) in between. The globular heads spiral about the myosin filament in the region where actin and myosin overlap (the A band) and extend as cross-bridges to interdigitate with sites on the actin filaments, thus forming the structural and functional link between the two filament types.

The intramyofibrillar cytoskeleton includes inelastic nebulin filaments, which span from the Z line to the actin filaments. Nebulin might also act as a template for the thin filament assembly.

Titin is 1 μm long. It is the largest polypeptide and spans from the Z line to the M line. Titin is an elastic filament. The part between the Z line and myosin has a string-like appearance. Titin has been suggested to contribute greatly to the passive force development of muscle during stretch (Fig. 6-2). It also might act as a template for the thick filament assembly (Linke et al., 1998; Simmons et al., 1997; Stroms et al., 1998).

The I band is bisected by the Z lines, which contain the portion of the thin filaments that does not overlap with the thick filaments and the elastic part of titin. In the center of the A band, in the gap between the ends of the thin filaments, is the H zone, a light band containing only thick filaments and that part of titin that is integrated in the thick filaments. A narrow, dark area in the center of the H zone is the M line, pro-

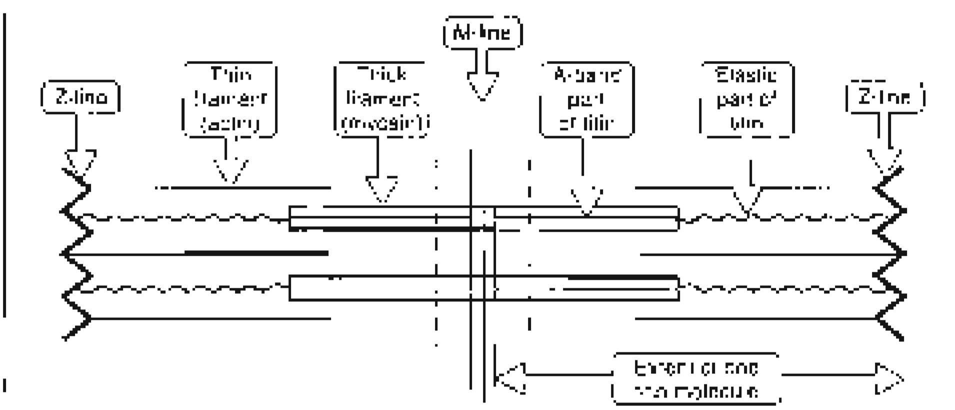


FIG. 6-2

The arrangement of titin molecules within the sarcomere. *Adapted from Craig, R. (1994). The structure of the contractile filaments. In A.G. Engel & Franzini-Armstrong (Eds.), Myology (2nd ed., p. 152). New York: McGraw-Hill, Inc.*

duced by transversely and longitudinally oriented proteins that link adjacent thick filaments, maintaining their parallel arrangement. The various areas of the banding pattern are apparent in the photomicrograph of human skeletal muscle shown in Figure 6-3.

Closely correlated with the repeating pattern of the sarcomeres is an organized network of tubules and sacs known as the sarcoplasmic reticulum. The tubules of the sarcoplasmic reticulum lie parallel to the myofibrils and tend to enlarge and fuse at the level of the junctions between the A and I bands, forming transverse sacs, or the terminal cisternae, that surround the individual myofibril completely.

The terminal cisternae enclose a smaller tubule that is separated from them by its own membrane. The smaller tubule and the terminal cisternae above and below it are known as a triad. The enclosed

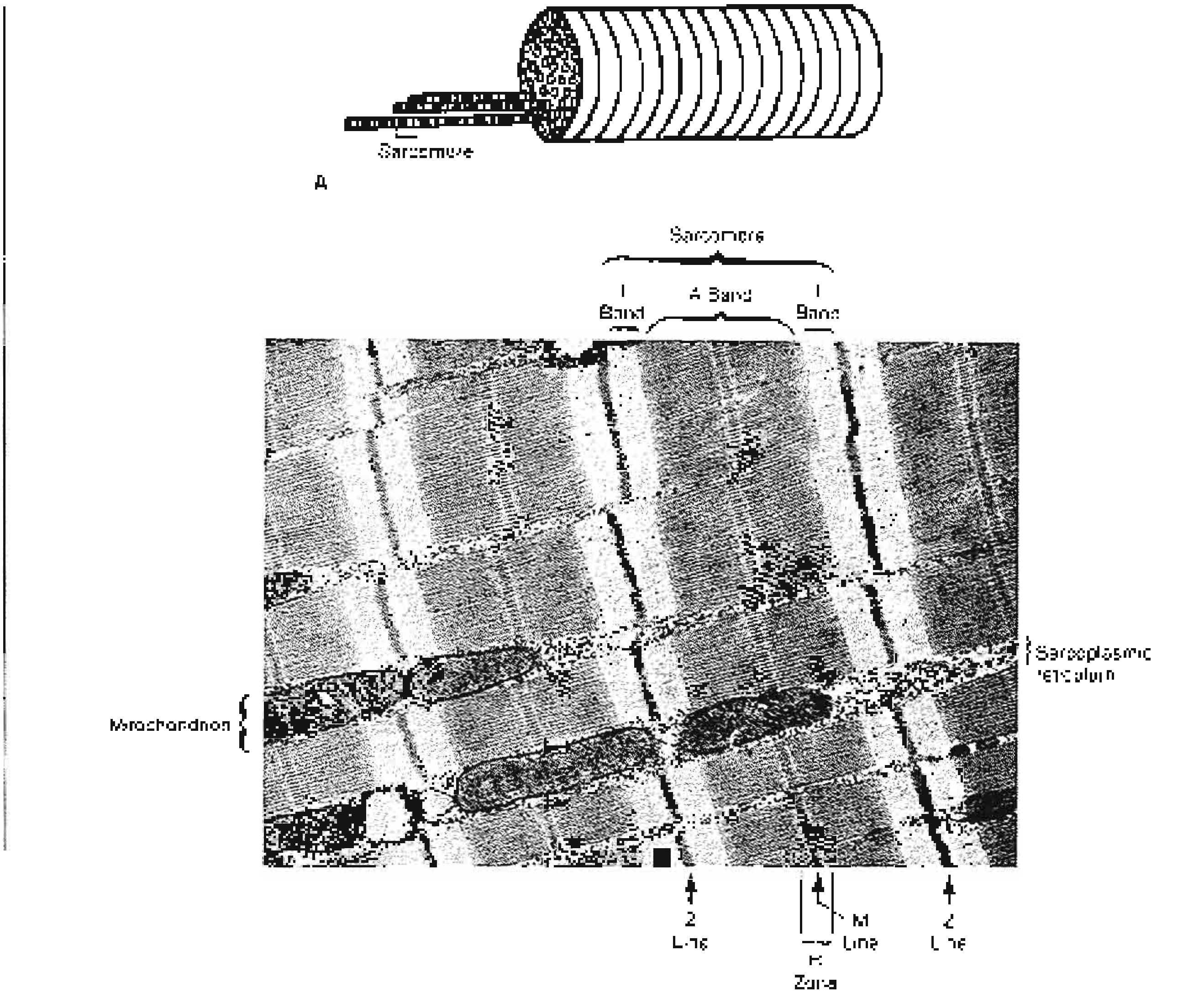


FIG. 6-3

A. Single muscle fiber with three protruding myofibrils. B. Electron photomicrograph of a cross section of human skeletal muscle. The sarcomeres are apparent along the myofibrils. Characteristic regions of the sarcomere are indicated.

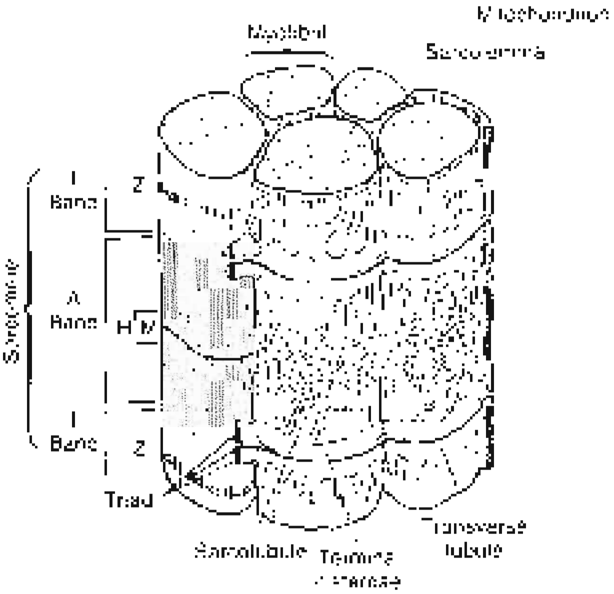


FIG. 6-4

Diagram of a portion of a skeletal muscle fiber illustrating the sarcoplasmic reticulum that surrounds each myofibril. The various regions of the sarcomere are indicated on the left myofibril to show the correlation of these regions with the sarcoplasmic reticulum, shown surrounding the middle and right myofibrils. The transverse tubules represent an infolding of the sarcolemma, the plasma membrane that encompasses the entire muscle fiber. Two transverse tubules supply each sarcomere at the level of the junctions of the A band and I bands. Terminal cisternae are located on each side of the transverse tubule, and together these structures constitute a triad. The terminal cisternae connect with a longitudinal network of sarcotubules spanning the region of the A band. *Adapted from Bloom, W. & Fawcett, D.W. (1975). Histology (10th ed.). Philadelphia: W.B. Saunders.*

tubule is part of the transverse tubule system, or T system, which are invaginations of the surface membrane of the fiber. This membrane, the sarcolemma, is a plasma membrane that invests every striated muscle (Fig. 6-4).

Molecular Basis of Muscle Contraction

The most widely held theory of muscle contraction is the sliding filament theory, proposed simultaneously by A.F. Huxley and H.E. Huxley in 1964 and subsequently refined (Huxley, 1974). According to this theory, active shortening of the sarcomere, and hence of the muscle, results from the relative movement of the actin and myosin filaments past one another while each retains its original length. The force of contraction is developed by the myosin heads, or cross-bridges, in the region of overlap between actin and myosin (the A band). These cross-bridges swivel in an arc around their fixed positions on the surface of the myosin filament, much like the oars of a boat. This movement of the cross-bridges in contact with the actin filaments produces the sliding of the actin filaments toward the center of the sarcomere. A muscle fiber contracts when all sarcomere shorten simultaneously in an all-or-nothing fashion, which is called a twitch.

Because a single movement of a cross-bridge produces only a small displacement of the actin filament relative to the myosin filament, each individual cross-bridge detaches itself from one receptor site on the actin filament and reattaches itself to another site further along, repeating the process five or six times, "with an action similar to a man pulling on a rope hand over hand" (Wilkie, 1968). The cross-bridges do not act in a synchronized manner; each acts independently. Thus, at any given moment only approximately half of the cross-bridges actively generate force and displacement, and when these detach, others take up the task so that shortening is maintained. The shortening is reflected in the sarcomere as a decrease in the I band and a decrease in the H zone as the Z lines move closer together; the width of the A band remains constant.

A key to the sliding mechanism is the calcium ion (Ca^{++}), which turns the contractile activity on and off. Muscle contraction is initiated when calcium is made available to the contractile elements and ceases when calcium is removed. The mechanisms that regulate the availability of calcium ions to the contractile machinery are coupled to electric events occurring in the muscle membrane (sarcolemma). An action potential in the sarcolemma provides the electric signal for the initiation of contractile activity. The mechanism by which the electric signal triggers the chemical events of contraction is known as excitation-contraction coupling.

When the motor neuron stimulates the muscle at the neuromuscular junction (Fig. 6-5A) and the propagated action potential depolarizes the muscle cell membrane (sarcolemma), there is an inward spread of the action potential along the T system. (Details of this process are given in Figure 6-5, A–C and in Box 6-1, which summarizes the events during the excitation, contraction, and relaxation of muscle. Figure 6-5D shows the struc-

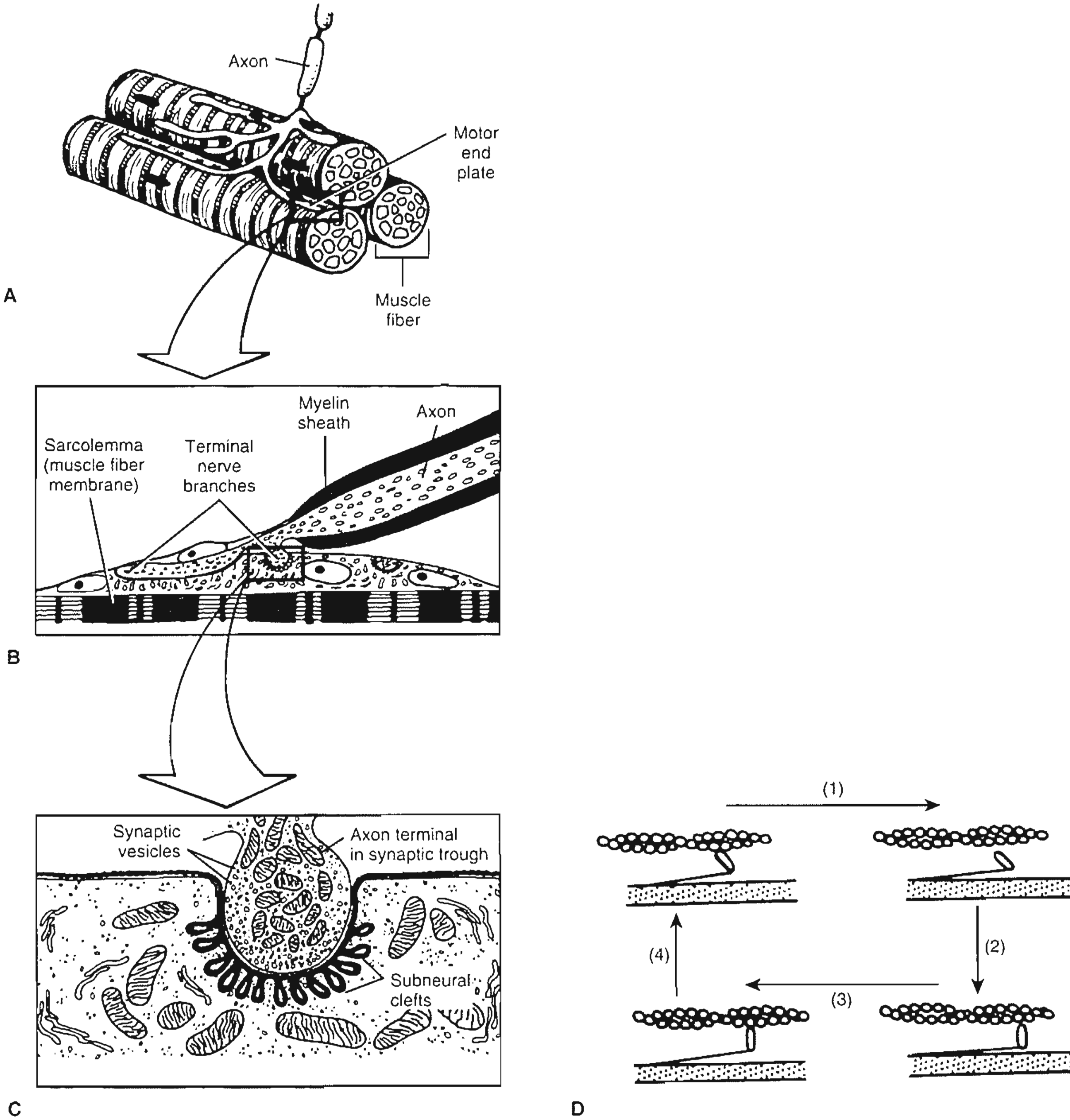


FIG. 6-5

Schematic representation of the innervation of muscle fibers. **A**, An axon of a motor neuron (originating from the cell body in the anterior horn of the spinal cord) branches near its end to innervate several skeletal muscle fibers, forming a neuromuscular junction with each fiber. The region of the muscle membrane (sarcolemma) lying directly under the terminal branches of the axon has special properties and is known as the motor end plate, or motor end plate membrane. The rectangular area is shown in detail in **B**. **B**, The fine terminal branches of the nerve (axon terminals), devoid of myelin sheaths, lie in grooves on the sarcolemma. The rectangular area in this section is shown in detail in **C**. **C**, Ultrastructure of the junction of an axon terminal and the sarcolemma. The invagination of the sarcolemma forms the synaptic trough into which the axon terminal protrudes. The invaginated sarcolemma has many folds, or subneural clefts, which greatly increase its surface area. Acetylcholine is stored in synaptic vesicles in the axon terminal. **B** and **C**, *adapted from Brobeck, J.R. (Ed.) (1979).* Best and Taylor's Physiological Basis of Medical Practice *(10th ed., pp. 59–113). Baltimore: Williams & Wilkins.* **D**, Cross-bridge cycle of muscle contraction.

BOX 6-1 Events During Excitation, Contraction, and Relaxation of Muscle Fiber

1. An action potential is initiated and propagated in a motor axon.
2. This action potential causes the release of acetylcholine from the axon terminals at the neuromuscular junction.
3. Acetylcholine is bound to receptor sites on the motor end-plate membrane.
4. Acetylcholine increases the permeability of the motor end-plate to sodium and potassium ions, producing an end-plate potential.
5. The end-plate potential depolarizes the muscle membrane (sarcolemma), generating a muscle action potential that is propagated over the membrane surface.
6. Acetylcholine is rapidly destroyed by acetylcholinesterase on the end-plate membrane.
7. The muscle action potential depolarizes the transverse tubules.
8. Depolarization of the transverse tubules leads to the release of calcium ions from the terminal cisternae of the sarcoplasmic reticulum surrounding the myofibrils. These ions are released into the sarcoplasm in the direct vicinity of the regulatory proteins tropomyosin and troponin.
9. Calcium ions bind to troponin, allowing movement of the tropomyosin molecule away from the myosin receptor sites on the actin filament that it had been blocking and releasing the inhibition that had prevented actin from combining with myosin.
10. Actin (A) combines with myosin ATP (M-ATP). (In this state, ATP has been hydrolyzed to ADP and phosphate, but the products are still attached to myosin. Receptor sites on the myosin cross-bridges bind to receptor sites on the actin chain.)

$$A + M \cdot ATP \rightarrow A \cdot M \cdot ATP$$

11. Actin activates the myosin ATPase found on the myosin cross-bridge, enabling ATP to be split (hydrolyzed). This process releases energy used to produce movement of the myosin cross-bridges.

$$A \cdot M \cdot ATP \rightarrow A \cdot M + ADP + P_i$$

12. Oarlike movements of the cross-bridges produce relative sliding of the thick and thin filaments past each other.
13. Fresh ATP binds to the myosin cross-bridge, breaking the actin-myosin bond and allowing the cross-bridge to dissociate from actin.

$$A \cdot M + ATP \rightarrow A + M \cdot ATP$$

14. The ATPase hydrolyzes the myosin ATP complex to the M · ATP complex, which represents the relaxed state of the sarcomere.

$$M \cdot ATP \rightarrow M \cdot ATP$$

15. Cycles of binding and unbinding of actin with the myosin cross-bridges at successive sites along the actin filament (steps 11, 12, 13, and 14) continue as long as the concentration of calcium remains high enough to inhibit the action of the troponin-tropomyosin system.
16. Concentration of calcium ions falls as they are pumped into the terminal cisternae of the sarcoplasmic reticulum by an energy-requiring process that splits ATP.
17. Calcium dissociates from troponin, restoring the inhibitory action of troponin-tropomyosin. The actin filament slides back and the muscle lengthens. In the presence of ATP, actin and myosin remain in the dissociated, relaxed state.

[illegible]

[illegible]

...ural features between actin and the cross-bridges of myosin.)

THE MOTOR UNIT

The functional unit of skeletal muscle is the motor unit, which includes a single motor neuron and all of the muscle fibers innervated by it. This unit is the smallest part of the muscle that can be made to contract independently. When stimulated, all muscle fibers in the motor unit respond as one. The fibers of a motor unit are said to show an all-or-none response to stimulation: they contract either maximally or not at all.

The number of muscle fibers forming a motor unit is closely related to the degree of control required of the muscle. In small muscles that perform very fine movements, such as the extraocular muscles, each motor unit may contain less than a dozen muscle fibers; in large muscles that perform coarse movements, such as the gastrocnemius, the motor unit may contain 1,000 to 2,000 muscle fibers.

The fibers of each motor unit are not contiguous but dispersed throughout the muscle with fibers of other units. Thus, if a single motor unit is stimulated, a large portion of the muscle appears to contract. If additional motor units of the nerve innervating the muscle are stimulated, the muscle contracts with greater force. The calling in of additional motor units in response to greater stimulation of the motor nerve is called recruitment.

THE MUSCULOTENDINOUS UNIT

The tendons and the connective tissues in and around the muscle belly are viscoelastic structures that help determine the mechanical characteristics of whole muscle during contraction and passive extension. Hill (1970) showed that the tendons represent a spring-like elastic component located in series with the contractile component (the contractile proteins of the myofibril, actin, and myosin), while the epimysium, perimysium, endomysium, and sarcolemma represent a second elastic component located in parallel with the contractile component (Fig. 6-6).

When the parallel and series elastic components stretch during active contraction or passive extension of a muscle, tension is produced and energy is stored; when they recoil with muscle relaxation, this energy is released. The series elastic fibers are more important in the production of tension than are the parallel elastic fibers (Wilkie, 1956). Several investigators have suggested that the cross-bridges of the myosin filaments have a spring-like property and also contribute to the elastic properties of muscle (Hill, 1968).

The distensibility and elasticity of the elastic components are valuable to the muscle in several ways:

1. They tend to keep the muscle in readiness for contraction and assure that muscle tension is produced and transmitted smoothly during contraction.
2. They assure that the contractile elements return to their original (resting) positions when contraction is terminated.
3. They may help prevent the passive overstretch of the contractile elements when these elements are relaxed, thereby lessening the danger of muscle injury.

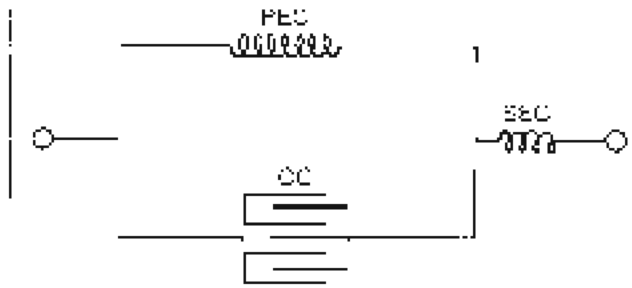


FIG. 6-6

The musculotendinous unit may be depicted as consisting of a contractile component (CC) in parallel with an elastic component (PEC) and in series with another elastic component (SEC). The contractile component is represented by the contractile proteins of the myofibril, actin, and myosin. (The myosin cross-bridges may also exhibit some elasticity.) The parallel elastic component comprises the connective tissue surrounding the muscle fibers (the epimysium, perimysium, and endomysium) and the sarcolemma. The series elastic component is represented by the tendons. Adapted from [illegible]

4. The viscous property of the series and parallel elastic components allows them to absorb energy proportional to the rate of force application and to dissipate energy in a time-dependent manner. (For a discussion of viscoelasticity, see Chapter 4.)

This viscous property, combined with the elastic properties of the musculotendinous unit, is demonstrated in everyday activities. For example, when a person attempts to stretch and touch the toes, the stretch is initially elastic. As the stretch is held, however, further elongation of the muscle results from the viscosity of the muscle-tendon structure, and the fingers slowly reach closer to the floor.

Mechanics of Muscle Contraction

Electromyography provides a mechanism for evaluating and comparing neural effects on muscle and the contractile activity of the muscle itself in vivo and in vitro. Much has been learned by using elec electromyography to study various aspects of the contractile process, particularly the time relationship between the onset of electrical activity in the muscle and actual contraction of the muscle or muscle fiber. The following sections discuss the mechanical response of a muscle to electrical (neural) stimulation and the various ways in which the muscle contracts to move a joint, control its motion, or maintain its position.

SUMMATION AND TETANIC CONTRACTION

The mechanical response of a muscle to a single stimulus of its motor nerve is known as a twitch, which is the fundamental unit of recordable muscle activity. Following stimulation there is an interval of a few milliseconds known as the latency period before the tension in the muscle fibers begins to rise. This period represents the time required for the "slack" in the elastic components to be taken up. The time from the start of tension development to peak tension is the contraction time, and the time from peak tension until the tension drops to zero is the relaxation time. The contraction time and relaxation time vary among muscles, depending largely on the muscle fiber makeup (described below). Some muscle fibers contract with a speed of only 10 msec; others may take 100 msec or longer.

An action potential lasts only approximately 1 to 2 msec. This is a small fraction of the time taken for the subsequent mechanical response, or twitch, even in muscles that contract quickly; thus it is possible for a series of action potentials to be initiated before the first twitch is completed if the activity of the motor axon is maintained. When mechanical responses to successive stimuli are added to an initial response, the result is known as summation (Fig. 6-7). If a second stimulus occurs during the latency period of the first muscle twitch, it produces no additional response and the muscle is said to be completely refractory.

The frequency of stimulation is variable and is modulated by individual motor units. The greater the frequency of stimulation of the muscle fibers, the greater the tension produced in the muscle as a whole. However, a maximal frequency will be reached beyond which the tension of the muscle no longer increases. When this maximal tension is sustained as a result of summation, the muscle is said to contract tetanically. In this case, the rapidity of stimulation outstrips the contraction-relaxation time of the muscle so that little or no relaxation can occur before the next contraction is initiated (Fig. 6-8).

The considerable gradation of contraction exhibited by whole muscles is achieved by the differential

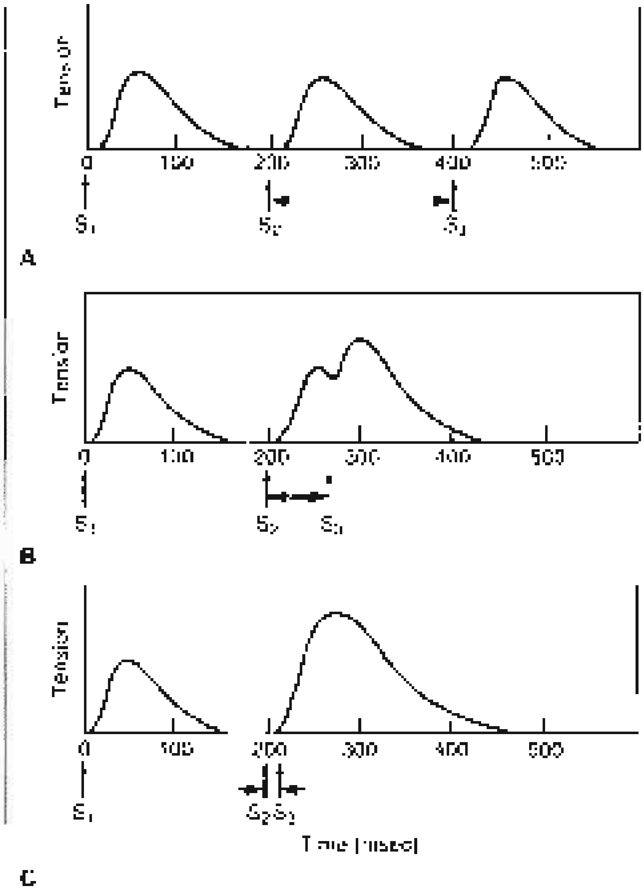


FIG. 6-7.

Summation of contractions in a muscle held at a constant length. A, An initial stimulus (S_1) is applied to the muscle, and the resulting twitch lasts 150 msec. The second (S_2) and third (S_3) stimuli are applied to the muscle after 200-msec intervals when the muscle has relaxed completely; thus no summation occurs. B, S_2 is applied 60 msec after S_1, when the mechanical response from S_1 is beginning to decrease. The resulting peak tension is greater than that of the single twitch. C, The interval between S_2 and S_3 is further reduced to 10 msec. The resulting peak tension is even greater than in B, and the increase in tension produces a smooth curve. The mechanical response evoked by S_3 appears as a continuation of that evoked by S_2. Adapted from Luciano, D.S., Vander, A.J., & Sherman, J.H. (1978). Human Function and Structure (pp. 113–136). New York: McGraw-Hill.

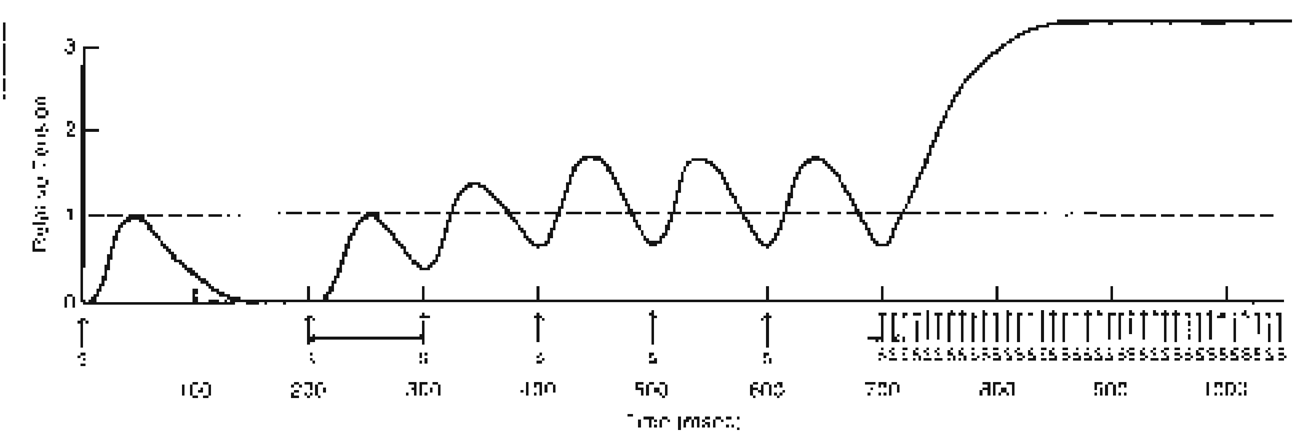


FIG. 6-8

Generation of muscle tetanus. As the frequency of stimulation (S) increases (i.e., the intervals shorten from 200 to 100 msec), the muscle tension rises as a result of summation. When the frequency is increased to 100/second, summation becomes maximal and the muscle contracts tetanically, exerting sustained peak tension. (Adapted from [illegible])

activity of their motor units, in both stimulation frequency and the number of units activated. The repetitive twitching of all recruited motor units of a muscle in an asynchronous manner results in brief summations or more prolonged subtetanic or tetanic contractions of the muscle as a whole and is a principal factor responsible for the smooth movements produced by the skeletal muscles.

TYPES OF MUSCLE CONTRACTION

During contraction, the force exerted by a contracting muscle on the bony lever(s) to which it is attached is known as the muscle tension, and the external force exerted on the muscle is known as the resistance, or load. As the muscle exerts its force, it generates a turning effect, or moment (torque), on the involved joint because the line of application of the muscle force usually lies at a distance from the center of motion of the joint. The moment is calculated as the product of the muscle force and the perpendicular distance between its point of application and the center of motion (this distance is known as the lever arm, or moment arm, of the force).

Muscle contractions and the resulting muscle work can be classified according to the relationship between either the muscle tension and the resistance to be overcome or the muscle moment generated and the resistance to be overcome, as shown in Box 6-2 (Kroemer et al., 1990).

Although no motion is accomplished and no mechanical work is performed during an isometric contraction, muscle work (physiological work) is performed; energy is expended and is mostly dissipated as heat, which is also called the isometric heat production. All dynamic contractions involve what may be considered an initial static (isometric) phase as the muscle first develops tension equal to the load it is expected to overcome.

The tension in a muscle varies with the type of contraction. Isometric contractions produce greater tension than do concentric contractions. Studies suggest that the tension developed in an eccentric contraction may even exceed that developed during an isometric contraction. These differences are thought to be due in large part to the varying amounts of supplemental tension produced in the series elastic component of the muscle and to differences in contraction time. The longer contraction time of the isometric and eccentric contractions allows greater cross-bridge formation by the contractile components, thus permitting greater tension to be generated (Kroll, 1987). More time is also available for this tension to be transmitted to the series elastic component as the muscle-tendon unit is stretched. The longer contraction time allows the recruitment of additional motor units.

Komi (1986) has pointed out that concentric, isometric, and eccentric muscle contractions seldom

BOX 6-2 Types of Muscle Work and Contraction

Dynamic work: Mechanical work is performed and joint motion is produced through the following forms of muscle contraction.

1. Concentric (con, together; centrum, center) contraction. When muscles develop sufficient tension to overcome the resistance of the body segment, the muscles shorten and cause joint movement. The net moment generated by the muscle is in the same direction as the change in joint angle. An example of a concentric contraction is the action of the quadriceps in extending the knee when ascending stairs.
2. Eccentric (e or ex, out of; centrum, center) contraction. When a muscle cannot develop sufficient tension and is overcome by the external load, it progressively lengthens instead of shortening. The net muscle moment is in the opposite direction from the change in joint angle. One purpose of eccentric contraction is to decelerate the motion of a joint. For example, when one descends stairs, the quadriceps works eccentrically to decelerate flexion of the knee, thus decelerating the limb. The tension that it applies is less than the force of gravity pulling the body downward, but it is sufficient to allow controlled lowering of the body.
3. Isokinetic (iso, constant; kinetic, motion) contraction. This is a type of dynamic muscle work in which movement of the joint is kept at a constant velocity, and hence the velocity of shortening or lengthening of the muscle is constant. Because velocity is held constant, muscle energy cannot be dissipated through acceleration of the body part and is entirely converted to a resisting moment. The muscle force varies with changes in its lever arm throughout the range of joint motion (Hislop & Perrine, 1967). The muscle contracts concentrically and eccentrically with different directions of joint motion. For example, the flexor muscles of a joint contract concentrically during flexion and eccentrically during extension, acting as decelerators during the latter.
4. Isoinertial (iso, constant; inertial, resistance) contraction. This is a type of dynamic muscle work wherein the resistance against which the muscle must contract remains constant. If the moment (torque) produced by the muscle is equal to or less than the resistance to be overcome, the muscle length remains unchanged and the muscle contracts isometrically. If the moment is greater than the resistance, the muscle shortens (contracts concentrically) and causes acceleration of the body part. Isoinertial contraction occurs, for example, when a constant external load is lifted. At the extremes of motion, the inertia of the load must be overcome; the involved muscles contract isometrically and muscle torque is maximal. In the mid-range of the motion, with the inertia overcome, the muscles contract concentrically and the torque is submaximal.
5. Isotonic (iso, constant; tonic, force) contraction. This term is commonly used to define muscle contraction in which the tension is constant throughout a range of joint motion. This term does not take into account the leverage effects at the joint. However, because the muscle force moment arm changes throughout the range of joint motion, the muscle tension must also change. Thus, isotonic muscle contraction in the truest sense does not exist in the production of joint motion (Kroll, 1987).

Static work: No mechanical work is performed and posture or joint position is maintained through the following form of muscle contraction.

1. Isometric (iso, constant; metric, length) contraction. Muscles are not always directly involved in the production of joint movements. They may exercise either a restraining or a holding action, such as that needed to maintain the body in an upright position in opposing the force of gravity. In this case the muscle attempts to shorten (i.e., the myofibrils shorten and in doing so stretch the series elastic component, thereby producing tension), but it does not overcome the load and cause movement; instead, it produces a moment that supports the load in a fixed position (e.g., maintains posture) because no change takes place in the distance between the muscle's points of attachment.

order alone in normal human movement. Instead, one type of contraction or load is preceded by a different type. An example is the eccentric loading prior to the concentric contraction that occurs at the ankle from midstance to toe-off during gait.

Because muscles normally shorten or lengthen at varying velocities and with varying amounts of tension, performance and measurement of isokinetic work require the use of an isokinetic dynamometer. This device provides constant velocity of joint motion and maximum external resistance throughout the range of motion of the involved joint, thereby requiring maximal muscle torque. The use of the isokinetic dynamometer provides a method of selective training and measurement, but physiological movement is not simulated.

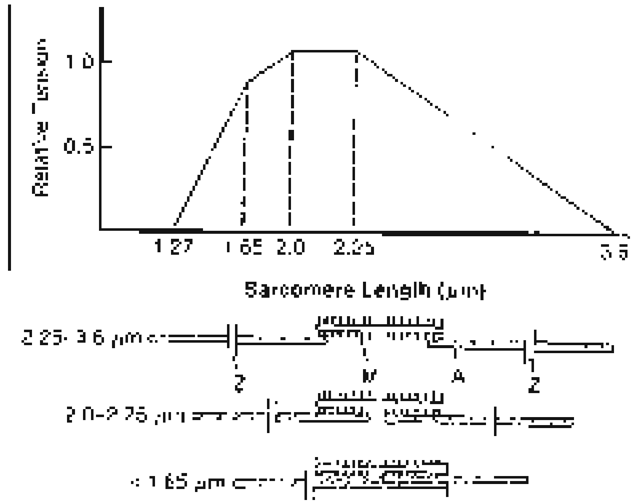


FIG. 6-9

Tension-length curve from part of an isolated muscle fiber stimulated at different lengths. The isometric tetanic tension is closely related to the number of cross-bridges on the myosin filament overlapped by the actin filament. The tension is maximal at the slack length, or resting length, of the sarcomere (2 μm), where overlap is greatest, and falls to zero at the length where overlap no longer occurs (3.6 μm). The tension also decreases when the sarcomere length is reduced below the resting length, falling sharply at 1.65 μm and reaching zero at 1.27 μm as the extensive overlap interferes with cross-bridge formation. The structural relationship of the actin and myosin filaments at various stages of sarcomere shortening and lengthening is portrayed below the curve. A, actin filaments; M, myosin filaments; Z, Z lines. *Adapted from Crawford, C.N.C., & James, N.T. (1980). The design of muscles. In R. Owen, J. Goodfellow, & P. Bullough (Eds.), Scientific Foundations of Orthopaedics and Traumatology (pp. 67–74). London: William Heinemann; as modified from Gordon, A.M., Huxley, A.F., & Julian, F.J. (1966). The variation in isometric tension with sarcomere length in vertebrate muscle fibers. J Physiol, 184, 170.*

Force Production in Muscle

The total force that a muscle can produce is influenced by its mechanical properties, which can be described by examining the length-tension, load-velocity, and force-time relationships of the muscle and the skeletal muscle architecture. Other principal factors in force production are muscle temperature, muscle fatigue, and prestretching.

LENGTH-TENSION RELATIONSHIP

The force, or tension, that a muscle exerts varies with the length at which it is held when stimulated. This relationship can be observed in a single fiber contracting isometrically and tetanically, as illustrated by the length-tension curve in Figure 6-9. Maximal tension is produced when the muscle fiber is approximately at its "slack," or resting, length. If the fiber is held at shorter lengths, the tension falls off slowly at first and then rapidly. If the fiber is lengthened beyond the resting length, tension progressively decreases.

The changes in tension when the fiber is stretched or shortened primarily are caused by structural alterations in the sarcomere. Maximal isometric tension can be exerted when the sarcomeres are at their resting length (2.0–2.25 μm) because the actin and myosin filaments overlap along their entire length and the number of cross-bridges is maximal. If the sarcomeres are lengthened, there are fewer junctions between the filaments and the active tension decreases. At a sarcomere length of approximately 3.6 μm, there is no overlap and hence no active tension. Sarcomere shortening to less than its resting length decreases the active tension because it allows overlapping of the thin filaments at opposite ends of the sarcomere, which are functionally polarized in opposite directions. At a sarcomere length of less than 1.65 μm, the thick filaments abut on the Z line and the tension diminishes sharply.

The length-tension relationship illustrated in Figure 6-9 is for an individual muscle fiber. If this relationship is measured in a whole muscle contracting isometrically and tetanically, the tension produced by both active components and passive components must be taken into account (Fig. 6-10).

The curve labeled "active tension" in Figure 6-10 represents the tension developed by the contractile elements of the muscle, and it resembles the curve for the individual fiber. The curve labeled "passive tension" reflects the tension developed when the muscle surpasses its resting length and the noncontractile muscle belly is stretched. This passive tension is mainly developed in the parallel and series elastic components (Fig. 6-6). When the belly contracts, the combined active and passive tensions produce the total tension exerted. The curve demonstrates that as a muscle is progressively stretched beyond its resting length, the passive tension rises and the active tension decreases.

Most muscles that cross only one joint normally are not stretched enough for the passive tension to play an important role, but the case is different for two-joint muscles, in which the extremes of the length-tension relationship may be functioning (Crawford & James, 1980). For example, the hamstrings shorten so much when the knee is fully flexed that the tension they can exert decreases considerably. Conversely, when the hip is flexed and the knee extended, the muscles are so stretched that it is the magnitude of their passive tension that prevents further elongation and causes the knee to flex if hip flexion is increased.

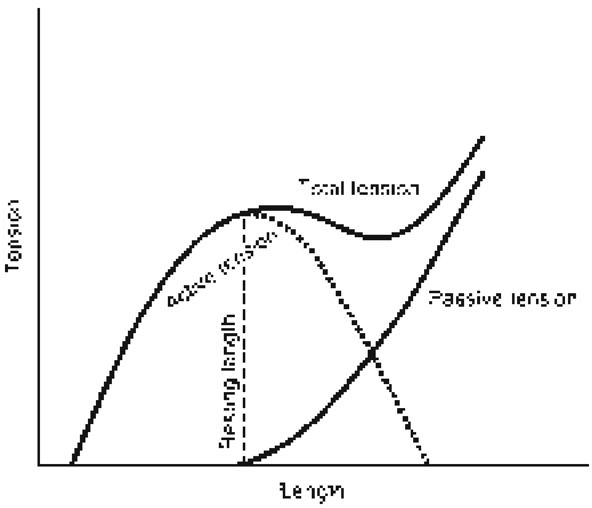


FIG. 6-10

The active and passive tension exerted by a whole muscle contracting isometrically and tetanically is plotted against the muscle's length. The active tension is produced by the contractile muscle components and the passive tension by the series and parallel elastic components, which develop stress when the muscle is stretched beyond its resting length. The greater the amount of stretching, the larger the contribution of the elastic component to the total tension. The shape of the active curve is generally the same in different muscles, but the passive curve, and hence the total curve, varies depending on how much connective tissue (elastic component) the muscle contains. *Adapted from Crawford, C.N.C. & James, N.T. (1980). The design of muscles. In R. Owen, J. Goodfellow, & P. Bullough (Eds.), Scientific Foundations of Orthopaedics and Traumatology (pp. 67-74). London: William Heinemann.*

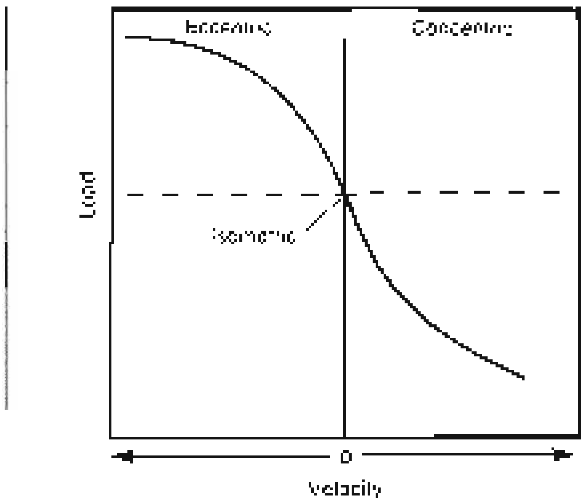


FIG. 6-11

Load-velocity curve generated by plotting the velocity of motion of the muscle lever arm against the external load. When the external load imposed on the muscle is negligible, the muscle contracts concentrically with maximal speed. With increasing loads the muscle shortens more slowly. When the external load equals the maximum force that the muscle can exert, the muscle fails to shorten (i.e., has zero velocity) and contracts isometrically. When the load is increased further, the muscle lengthens eccentrically. This lengthening is more rapid with greater load.

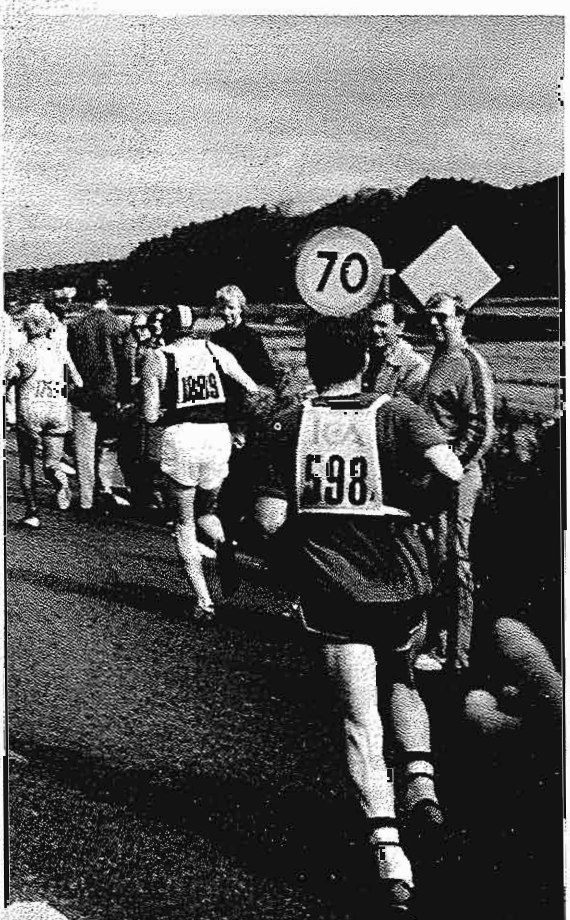
LOAD-VELOCITY RELATIONSHIP

The relationship between the velocity of shortening or eccentric lengthening of a muscle and different constant loads can be determined by plotting the velocity of motion of the muscle lever arm at various external loads, thereby generating a load-velocity curve (Fig. 6-11). The velocity of shortening of a muscle contracting concentrically is inversely related to the external load applied (Guyton, 1986). The velocity of shortening is greatest when the ex-

CASE STUDY 6-1

Gastrocnemius Muscle Tear

A 22-year-old male professional athlete tears his gastrocnemius during a race (Fig. Case Study 6-1-1). The muscle overload that happens during strenuous eccentric and concentric contractions increases the risk of injury, especially when the forces involve biarticular muscles such as the gastrocnemius. This indirect trauma is associated with high tensile forces during fast contraction (high velocity) and continued changes in muscle length. The status of muscle contraction at the time of overload is usually eccentric, and failure most often occurs at or near the myotendinous junction unless the muscle has been previously injured (Kasser, 1996). Swelling from hemorrhage occurs initially in the inflammatory phase. The cellular response is more rapid and repair is more complete if the vascular channels are not disrupted and the nutrition of the tissue is not disturbed. The degree of injury from a tensile overload will dictate the potential host response and the time needed for repair.



Case Study Figure 6-1-1.

ternal load is zero, but as the load increases the muscle shortens more and more slowly. When the external load equals the maximal force that the muscle can exert, the velocity of shortening becomes zero and the muscle contracts isometrically. When the load is increased still further, the muscle contracts eccentrically; it elongates during contraction. The load-velocity relationship is reversed from that of the concentrically contracting muscle; the muscle eccentrically lengthens more quickly with increasing load (Kroll, 1987) (Case Study 6-1).

FORCE-TIME RELATIONSHIP

The force, or tension, generated by a muscle is proportional to the contraction time: the longer the contraction time, the greater is the force developed, up to the point of maximum tension. In Figure 6-12, this relationship is illustrated by a force-time curve for a whole muscle contracting isometrically. Slower contraction leads to greater force production because time is allowed for the tension produced by the contractile elements to be transmitted through the parallel elastic components to the tendon. Although tension production in the contractile component can reach a maximum in as little as 10 msec, up to 300 msec may be needed for that tension to be transferred to the elastic components. The tension in the tendon will reach the maximum tension developed by the contractile element only if the active contraction process is of sufficient duration (Ottoson, 1983).

EFFECT OF SKELETAL MUSCLE ARCHITECTURE

The muscles consist of the contractile component, the sarcomere, which produces active tension. The arrangement of the contractile components affects the contractile properties of the muscle dramatically. The more sarcomeres lie in series, the longer the myofibril will be; the more sarcomeres lie parallel, the larger the cross-sectional area of the myofibril will be. These two basic architectural patterns of myofibrils (long or thick) affect the contractile properties of the muscles in the following ways:

1. The force the muscle can produce is proportional to the cross section of the myofibril (Fig. 6-13A).
2. The velocity and the excursion (working range) that the muscle can produce are proportional to the length of the myofibril (Fig. 6-13B).

Muscles with shorter fibers and a larger cross-sectional area are designed to produce force, whereas muscles with long fibers are designed for excursion and velocity. The quadriceps muscle contains shorter myofibrils and appears to be specialized for force production. The sartorius muscle has longer fibers and a smaller cross-sectional area and is better suited for high excursion (Baratta et al. 1998; Lieber & Bodine-Fowler 1993).

EFFECT OF PRESTRETCHING

It has been demonstrated in amphibians and in humans (Ciullo & Zarins, 1983) that a muscle performs more work when it shortens immediately after being stretched in the concentrically contracted state than when it shortens from a state of isometric contraction. This phenomenon is not entirely accounted for by the elastic energy stored in the series elastic component during stretching but must also be caused by energy stored in the contractile component. It has been suggested that changes in the intrinsic mechanical properties of myofibrils are important in the stretch-induced enhancement of work production (Takarada et al. 1997).

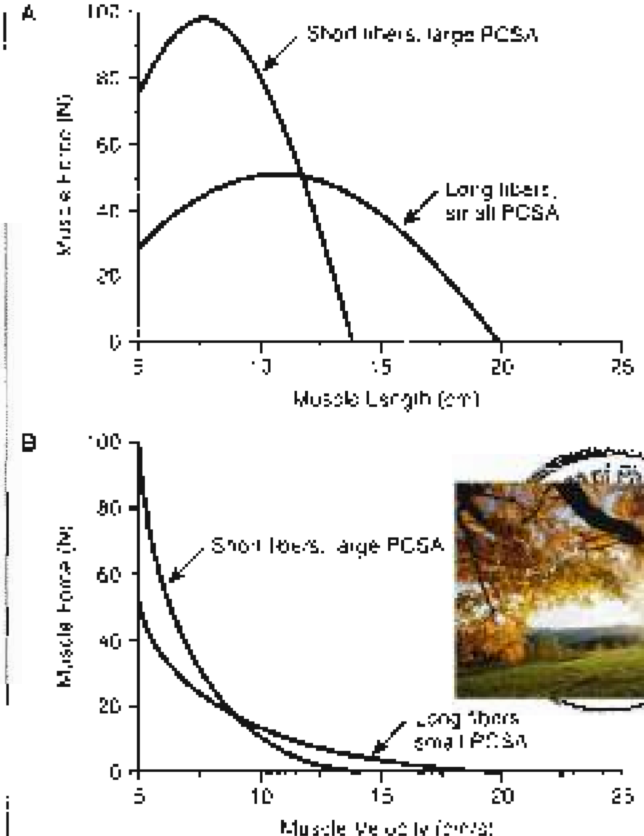


FIG. 6-13

Isometric and isotonic properties of muscles with different architecture. A. Force-length relationship. B. Force-velocity relationship. PSCA, physiological cross-sectional area. *Reprinted with permission of the American Physical Therapy Association. From Lieber, R.L. (1993). Skeletal muscle architecture: Implications for rehabilitation. Physical Therapy, 73(12), 852.*

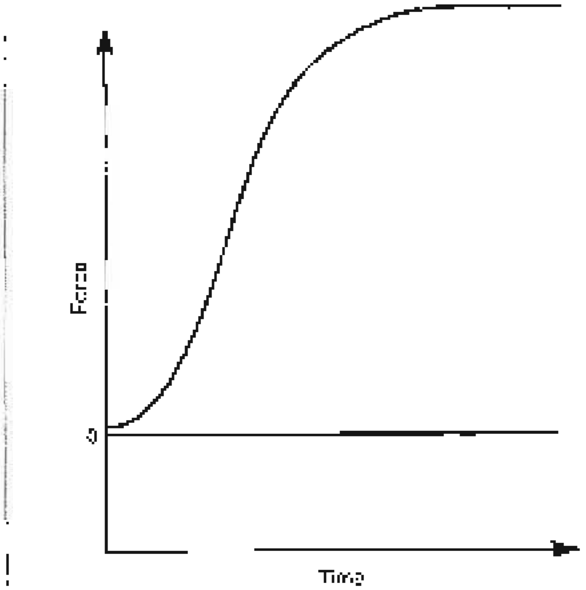


FIG. 6-12

Force-time curve for a whole muscle contracting isometrically. The force exerted by the muscle is greater when the contraction time is longer because time is required for the tension created by the contractile components to be transferred to the parallel elastic component and then to series elastic component as the musculotendinous unit is stretched.

EFFECT OF TEMPERATURE

A rise in muscle temperature causes an increase in conduction velocity across the sarcolemma (Phillips & Petrofsky, 1983), increasing the frequency of stimulation and hence the production of muscle force. Rising of the muscle temperature from 6 to 34°C results in an almost linear increase of the tension/stiffness ratio (Galler et al., 1998).

A rise in temperature also causes greater enzymatic activity of muscle metabolism, thus increasing the efficiency of muscle contraction. A further effect of a rise in temperature is the increased elasticity of the collagen in the series and parallel elastic components, which enhances the extensibility of the muscle-tendon unit. This increased prestretch increases the force production of the muscle.

Muscle temperature increases by means of two mechanisms:

1. Increase in blood flow, which occurs when an athlete "warms up" his or her muscles
2. Production of the heat of reaction generated by metabolism, by the release of the energy of contraction, and by friction as the contractile components slide over each other

However, at low temperature (10°C), it has been shown that the maximum shortening velocity and the isometric tension are inhibited significantly. This is caused by decreased pH (acidosis) in the muscle. The pH plays a much less important role at temperatures close to the physiological level (Pate et al., 1995).

EFFECT OF FATIGUE

The ability of a muscle to contract and relax is dependent on the availability of adenosine triphosphate (ATP) (Box 6-1). If a muscle has an adequate supply of oxygen and nutrients that can be broken down to provide ATP, it can sustain a series of low-frequency twitch responses for a long time. The frequency must be low enough to allow the muscle to synthesize ATP at a rate sufficient to keep up with the rate of ATP breakdown during contraction. If the frequency of stimulation increases and outstrips the rate of replacement of ATP, the twitch responses soon grow progressively weaker and eventually fall to zero (Fig. 6-14). This drop in tension following prolonged stimulation is muscle fatigue. If the frequency is high enough to produce tetanic contractions, fatigue occurs even sooner. If a period of rest is allowed before stimulation is continued, the ATP concentration rises and the muscle briefly recovers its contractile ability before again undergoing fatigue.

Three sources supply ATP to muscle: creatine phosphate, oxidative phosphorylation in the mitochondria, and substrate phosphorylation during anaerobic glycolysis. When contraction begins, the myosin ATPase rapidly breaks down ATP. The increase in adenosine diphosphate (ADP) and phosphate (Pi) concentrations resulting from this breakdown ultimately leads to increased rates of oxidative phosphorylation and glycolysis. After a short lapse, however, these metabolic pathways begin to deliver ATP at a high rate. During this interval, the energy for ATP formation is provided by creatine phosphate, which offers the most rapid means of forming ATP in the muscle cell.

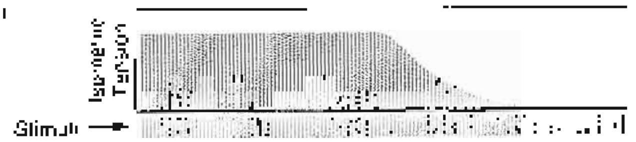


FIG. 6-14

Fatigue in a muscle contracting isometrically. Prolonged stimulation occurs at a frequency that outstrips the muscle's ability to produce sufficient ATP for contraction. As a result, tension production declines and eventually ceases. *Adapted from Luciano, D.S., Vander, A.J., & Sherman, J.H. (1978). Human Function and Structure (pp. 113–136). New York: McGraw-Hill.*

At moderate rates of muscle activity, most of the required ATP can be formed by the process of oxidative phosphorylation. During intense exercise, when ATP is being broken down rapidly, the cell's ability to replace ATP by oxidative phosphorylation may be limited, primarily by inadequate delivery of oxygen to the muscle by the circulatory system.

Even when oxygen delivery is adequate, the rate at which oxidative phosphorylation can produce ATP may be insufficient to sustain intense exercise because the enzymatic machinery of this pathway is relatively slow. Anaerobic glycolysis then begins to contribute an increasing portion of the ATP. The glycolytic pathway, although it produces much smaller amounts of ATP from the breakdown of glucose, operates at a much faster rate. It can also proceed in the absence of oxygen, with the formation of lactic acid as its end product. Thus, during intense exercise, anaerobic glycolysis becomes an additional source for rapidly supplying the muscle with ATP.

The glycolytic pathway has the disadvantage of requiring large amounts of glucose for the production of small amounts of ATP. Thus, even though muscle stores glucose in the form of glycogen, existing glycogen supplies may be depleted quickly when muscle activity is intense. Finally, myosin ATPase may break down ATP faster than even glycolysis can replace it, and fatigue occurs rapidly as ATP concentrations drop.

After a period of intense exercise, creatine phosphate levels have become low and much of the muscle glycogen may have been converted to lactic acid. For the muscle to be returned to its original state, creatine phosphate must be resynthesized and the glycogen stores must be replaced. Because both processes require energy, the muscle will continue to consume oxygen at a rapid rate even though it has stopped contracting. This sustained high oxygen uptake is demonstrated by the fact that a person continues to breathe heavily and rapidly after a period of strenuous exercise.

When the energy necessary to return glycogen and creatine phosphate to their original levels is taken into account, the efficiency with which muscle converts chemical energy to work (movement) is usually no more than 20 to 25%, the majority of the energy being dissipated as heat. Even when muscle is operating in its most efficient state, a maximum of only approximately 40% of the energy is used for contraction (Andersson et al., 1984; Guyton, 1986).

In growth biomechanics, muscle fatigue is first observed by the lack of coordination of movement and its effect on the increasing of loads in tissue. Researchers including Bates et al. (1977) have indicated that the skill of the person in performing a given action is affected by fatigue. They studied the fatigue effect on runners and observed that runners decrease their knee extension when fatigue occurs (Bates et al., 1977). Parnianpour (1988) studied the motion coupling of the spine at exhaustive extension flexion. This study showed that when an individual became fatigued, the coupled motion increased and therefore the spinal torque increased. The most deleterious component of the neuromuscular adaptation to the fatigue state was the reduction in accuracy control and speed of contraction, which may predispose an individual to injury if muscle fatigue occurs.

Muscle Fiber Differentiation

In the preceding section, we described the major factors that determine the total tension developed by the whole muscle when it contracts. Individual muscle fibers also display distinct differences in their rates of contraction, development of tension, and susceptibility to fatigue.

Many methods of classifying muscle fibers have been devised. As early as 1678, Lorenzini observed anatomically the gross difference between red and white muscle, and in 1873 Ranvier typed muscle on the basis of speed of contractility and fatigability. Although considerable confusion has existed concerning the method and terminology for classifying skeletal muscle, recent histological and histochemical observations have led to the identification of three distinct types of muscle fibers on the basis of differing contractile and metabolic properties (Brandstater & Lambert, 1969; Buchthal & Schmalbruch, 1980) (Table 6-1).

The fiber types are distinguished mainly by the metabolic pathways by which they can generate ATP and the rate at which its energy is made available to the contractile system of the sarcomere, which determines the speed of contraction. The three fiber types are termed type I, slow-twitch oxidative (SO) fibers; type IIA, fast-twitch oxidative-glycolytic (FOG) fibers; and type IIB, fast-twitch glycolytic (FG) fibers.

Type I (SO) fibers are characterized by a low activity of myosin ATPase in the muscle fiber and, therefore, a relatively slow contraction time. The glycolytic (anaerobic) activity is low in this fiber type, but a high content of mitochondria produces a high potential for oxidative (aerobic) activity. Type I fibers are difficult to fatigue because the high rate of blood flow to these fibers delivers oxygen and nutrients at a sufficient rate to keep up with the relatively slow rate of ATP breakdown by myosin ATPase. Thus, the fibers are well suited for prolonged, low-intensity work. These fibers are relatively small in diameter and so produce relatively little tension. The high myoglobin content of type I fibers gives the muscle a red color.

Type II muscle fibers are divided into two main subgroups, IIA and IIB, on the basis of differing susceptibility to treatment with different buffers prior to incubation (Brooke & Kaiser, 1970). A third subgroup, the type IIC fibers, are rare, undifferentiated fibers, which are usually seen before the 30th week of gestation. This fiber type is infrequent in human muscle (Banker, 1994). Type IIA and IIB fibers are characterized by a high activity of myosin ATPase, which results in relatively fast contraction.

Type IIA (FOG) fibers are considered intermediate between type I and type IIB because their fast contraction time is combined with a moderately well-developed capacity for both aerobic (oxidative) and anaerobic (glycolytic) activity. These

TABLE 6-1

Properties of Three Types of Skeletal Muscle Fibers

	TYPE I Slow-Twitch Oxidative (SO)	TYPE IIA Fast-Twitch Oxidative-Glycolytic (FOG)	TYPE IIB Fast-Twitch Glycolytic (FG)
Speed of contraction	Slow	Fast	Fast
Primary source of ATP production	Oxidative phosphorylation	Oxidative phosphorylation	Anaerobic glycolysis
Glycolytic enzyme activity	Low	Intermediate	High
Capillaries	Many	Many	Few
Myoglobin content	High	High	Low
Glycogen content	Low	Intermediate	High
Fiber diameter	Small	Intermediate	Large
Rate of fatigue	Slow	Intermediate	Fast

fibers also have a well-developed blood supply. They can maintain their contractile activity for relatively long periods; however, at high rates of activity, the high rate of ATP splitting exceeds the capacity of both oxidative phosphorylation and glycolysis to supply ATP, and these fibers thus eventually fatigue. Because the myoglobin content of this muscle type is high, the muscle is often categorized as red muscle.

Type IIB (FG) fibers rely primarily on glycolytic (anaerobic) activity for ATP production. Few capillaries are found in the vicinity of these fibers and because they contain little myoglobin they are often referred to as white muscle. Although type IIB fibers are able to produce ATP rapidly, they fatigue easily because their high rate of ATP splitting quickly depletes the glycogen needed for glycolysis. These fibers generally are of large diameter and are thus able to produce great tension, but only for short periods before they fatigue.

It has been well demonstrated that the nerve innervating the muscle fiber determines its type (Burke et al., 1971); thus, the muscle fibers of each motor unit are of a single type. In humans and other species, electrical stimulation was found to change the fiber type (Munsat, McNeal, & Waters, 1976). In animal studies, transecting the nerves that innervate slow-twitch and fast-twitch muscle fibers and then crossing these nerves was noted to reverse the fiber types. After recovery from the cross-innervation, the slow-twitch fibers became fast in their contractile and biochemical properties and the fast-twitch fibers became slow.

The fiber composition of a given muscle depends on the function of that muscle. Some muscles perform predominantly one form of contractile activity and are often composed mostly of one muscle fiber type. An example is the soleus muscle in the calf, which primarily maintains posture and is composed of a high percentage of type I fibers. More commonly, however, a muscle is required to perform endurance-type activity under some circumstances and high-intensity strength activity under others. These muscles generally contain a mixture of the three muscle fiber types.

In a typical mixed muscle exerting low tension, some of the small motor units, composed of type I fibers, contract. As the muscle force increases, more motor units are recruited and their frequency of stimulation increases. As the frequency becomes maximal, greater muscle force is achieved by recruitment of larger motor units composed of type IIA (FOG) fibers and eventually type IIB (FG) fibers. As the peak muscle force decreases, the larger units are the first to cease activity (Guyton, 1986; Luciano, Vander, & Sherman, 1978).

It is generally, but not universally, accepted that fiber types are genetically determined (Costill et al., 1976; Gollnick, 1983). In the average population, approximately 50 to 55% of muscle fibers are type I, approximately 30 to 35% are type IIA, and approximately 15% are type IIB, but these percentages vary greatly among individuals.

In elite athletes, the relative percentage of fiber types differs from that in the general population and appears to depend on whether the athlete's principal activity requires a short, explosive, maximal effort or involves submaximal endurance. Sprinters and shot putters, for example, have a high percentage of type II fibers, whereas distance runners and cross-country skiers have a higher percentage of type I fibers. Endurance athletes may have as many as 80% type I fibers, and those engaged in short, explosive efforts as few as 30% of these fibers (Saltin et al., 1977).

The genetically determined fiber typing may be responsible for the natural selective process by which athletes are drawn to the type of sport for which they are most suited. Because fiber types are determined by the nerve that innervates the muscle fiber, there may be some cortical control of this innervation that influences an athlete to choose the sport in which he or she is genetically able to excel.

Muscle Injuries

Muscle injuries comprise contusion, laceration, ruptures, ischemia, compartment syndromes, and denervation. These injuries weaken the muscles and can cause significant disability. Blunt trauma can diminish muscle strength, limit joint motion, and finally lead to myositis ossificans. Muscle laceration, surgical incisions, and traumatic lesion to muscle tissue and denervation weaken the muscles, sometimes significantly. Ruptures in muscles also can cause weakness. Like the other injuries, they may result from direct trauma, but muscle contractions against resistance also can lead to tears in muscle tissue.

Acute muscle ischemia and compartment syndromes can cause extensive muscle necrosis. The many potential causes of compartment syndrome all result in increased pressure within a confined muscle compartment. In this case, failure to relieve the pressure rapidly may cause complications that range from weakness and decreased motion to loss of an entire limb.

Studies have shown that healthy skeletal muscle has a substantial capacity to repair itself. This repair process following a specific injury is inferred by the prior innervation pattern, vascularization, physical constraint of the surrounding tissues, the extent and condition of extracellular matrices, and the development of repair cells. Muscle injuries are important but the topic is not within the scope of this chapter. Injuries should be investigated carefully if suspicion arises that a patient has muscle damage.

Muscle Remodeling

The remodeling of muscle tissue is similar to that of other skeletal tissues such as bone, articular cartilage, and ligaments. As in these other tissues, muscle atrophies in response to disuse and immobilization and hypertrophies when subjected to greater use than usual.

EFFECTS OF DISUSE AND IMMOBILIZATION

Disuse and immobilization have detrimental effects on muscle fibers. These effects include loss of endurance and strength and muscle atrophies on a microstructural and macrostructural level, such as decreased numbers and size of fiber. Biochemical changes occur and affect aerobic and anaerobic energy production. These effects are dependent on fiber type and muscle length during immobilization. Immobilization in a lengthened position has a less deleterious effect (Appell, 1997; Kasser, 1996; Ohira et al., 1997; Sandmann, et al., 1998).

Clinical and laboratory studies of human and animal muscle tissue suggest that a program of immediate or early motion may prevent muscle atrophy after injury or surgery. In a study of crush injuries to rat muscle, the effect of immobilization of the crushed limb was compared with that of immediate motion. The muscle fibers were found to regenerate in a more parallel orientation in the mobilized animal than in the immobilized animal, capillarization occurred more rapidly, and tensile strength returned more quickly. Similar results were found in a later study on the effect of immobilization on the morphology of rat calf muscles (Kannus et al., 1998a).

It has been found clinically that atrophy of the quadriceps muscle that develops while the limb is immobilized in a rigid plaster cast cannot be reversed through the use of isometric exercises. Atrophy may be limited by allowing early motion such as that permitted by a partly mobile cast brace. In this case, dynamic exercises can be performed.

Human muscle biopsy studies have shown that it is mainly the type I fibers that atrophy with immobilization; their cross-sectional area decreases and their potential for oxidative enzyme activity is reduced (Kannus et al., 1998b). Early motion may prevent this atrophy. It appears that if the muscle is placed under tension when the body segment moves, afferent (sensory) impulses from the intrafusal muscle spindles will increase, leading to increased stimulation of the type I fiber. Although intermittent isometric exercise may be sufficient to maintain the metabolic capacity of the type II fiber, the type I fiber (the postural fiber) requires a more continuous impulse. Evidence also suggests that electric stimulation may prevent the decrease in type I fiber size and the decline in its oxidative enzyme activity caused by immobilization (Eriksson et al., 1981).

In elite athletes, inactivity following injury, surgery, or immobilization rapidly decreases the size and aerobic capability of muscle fibers, particularly in the fiber type affected by the chosen sport. In endurance athletes, type I fibers are affected, while in athletes engaged in an explosive activity such as sprinting, type II fibers are affected.

EFFECTS OF PHYSICAL TRAINING

Physical training increases the cross-sectional area of all muscle fibers, accounting for the increase in muscle bulk and strength. Some evidence suggests that the relative percentage of fiber types composing a person's muscles may also change with physical training (Arvidson, Eriksson, & Pitman, 1984). The cross-sectional area of the fibers affected by the athlete's principal activity increases. For example, in endurance athletes, the area of muscle taken up by type I and type IIA fibers increases at the expense of the total area of type IIB fibers (Case Study 6-2).

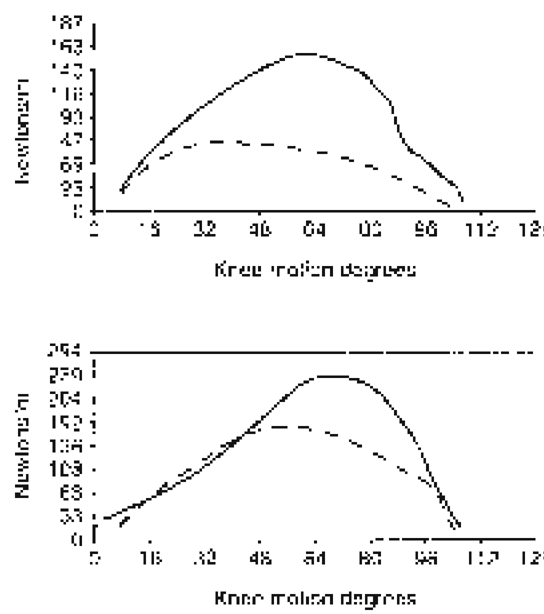
Stretching increases muscle flexibility, maintains and augments the range of joint motion, and increases the elasticity and length of the musculotendinous unit (Brobeck, 1979; Cuillo & Zarins, 1983). It also permits the musculotendinous unit to store more energy in its viscoelastic and contractile components.

The events that take place during muscle stretching are complex and incompletely understood (Kollrack, 1987; Guyton, 1986). It appears that these events are controlled or modified by both the intrafusal muscle spindles, located in parallel with the extrafusal fibers of the muscle belly, and the Golgi tendon organs, located in series with

CASE STUDY 6-2

Ruptured Left Anterior Cruciate Ligament

A 25-year-old male, semi-professional soccer player, ruptured his left anterior cruciate ligament. Isokinetic measurements taken from the involved and uninvolved limb 10 weeks after the surgical procedure (Fig. 6-2-1A) are repeated 6 weeks after the training began (Fig. 6-2-1B). An increase of muscle torque is shown in the repeated isokinetic test. The initial deficit of the involved side was approximately 63% when compared with the uninvolved side. After 6 weeks of training, the deficit of the involved side compared with the uninvolved side decreased to 43%.



A

B

Case Study Figure 6-2-1 Isokinetic test at 180°/sec. **A.** Measurement of the quadriceps femoris torque production at 10 weeks postsurgical procedure. The *dashed line* represents torque output by the involved limb. The *solid line* represents torque output by the noninvolved limb. **B.** Measurements of the quadriceps femoris torque production at 16 weeks postsurgical procedure and 6 weeks after training sessions. The *dashed line* represents torque output by the involved limb. The *solid line* represents torque output by the noninvolved limb.

these fibers. The spindles respond to an increase in muscle length and the Golgi apparatus to an increase in muscle tension. The resulting spindle reflex increases muscle contraction, while the Golgi reflex inhibits contraction and enhances muscle relaxation.

The intrafusal muscle spindles are of two types: primary and secondary. The primary spindles respond to changes in the rate of muscle lengthening (dynamic response) and the actual amount of lengthening. The secondary spindles respond only to the actual length change (static response). The static response is weak and the dynamic response is strong; therefore, keeping the rate of stretch low may allow the dynamic response to be bypassed, essentially negating the effect of the spindles. Conversely, the increase in muscle tension during stretching may activate the relaxing effect of the Golgi apparatus and thus enhance further stretching. The various methods and theories of stretching all have as a common goal inhibition of the spindle effect and enhancement of the Golgi effect to relax the muscle and promote further lengthening.

Summary

1 The structural unit of skeletal muscle is the fiber, which is encompassed by the endomysium and organized into fascicles encased in the perimysium. The epimysium surrounds the entire muscle.

2 The fibers are composed of myofibrils, aligned so as to create a band pattern. Each repeat of this pattern is a sarcomere, the functional unit of the contractile system.

3 The myofibrils are composed of thin filaments of the protein actin and thick filaments of the protein myosin, and the intramyofibrillar cytoskeleton is composed of the elastic filaments titin and the inelastic filaments nebulin.

4 According to the sliding filament theory, active shortening of the muscle results from the relative movement of the actin and myosin filaments past one another. The force of contraction is developed by movements of the myosin heads, or cross-bridges, in contact with the actin filaments. Troponin and tropomyosin, two proteins in the actin helix, regulate the making and breaking of the contacts between filaments.

5 A key to the sliding mechanism is the calcium ion, which turns the contractile activity on and off.

6 The motor unit, a single motor neuron and all muscle fibers innervated by it, is the smallest part of the muscle that can contract independently. The calling in of additional motor units in response to greater stimulation of the motor nerve is known as recruitment.

7 The tendons and the endomysium, perimysium, sarcolemma, and epimysium represent parallel and series elastic components that stretch with active contraction or passive muscle extension and recoil with muscle relaxation.

8 Summation occurs when mechanical responses of the muscle to successive stimuli are added to an initial response. When maximal tension is sustained as a result of summation, the muscle contracts tetanically. The muscle fiber contracts in an all-or-nothing fashion.

9 Muscles may contract concentrically, eccentrically, or isometrically depending on the relationship between the muscle tension and the resistance to be overcome. Concentric and eccentric contractions involve dynamic work, in which the muscle moves a joint or controls its movement.

10 Force production in muscle is influenced by the length-tension, load-velocity, and force-time relationships of the muscle. The length-tension relationship in a whole muscle is influenced by both active (contractile) and passive (series and parallel elastic) components.

11 Two other factors that increase force production are prestretching of the muscle and a rise in muscle temperature.

12 The energy for muscle contraction and its release is provided by the hydrolytic splitting of ATP. Muscle fatigue occurs when the ability of the muscle to synthesize ATP is insufficient to keep up with the rate of ATP breakdown during contraction.

13 Three main fiber types have been identified: type I, slow-twitch oxidative; type IIA, fast-twitch oxidative glycolytic; and type IIB, fast-twitch glycolytic fibers. Most muscles contain a mixture of these types.

14 Muscle atrophies occur under disuse and immobilization; muscle trophism can be restored through early and active rehabilitation.

REFERENCES

Appell, H.J. (1997). The muscle in the rehabilitation process. *Orthopade, 26*(11), 930–934.

Arvidsson, I., Eriksson, E., & Pitman, M. (1984). Neuromuscular basis of rehabilitation. In E. Hunter & J. Funk (Eds.), *Rehabilitation of the Injured Knee* (pp. 210–234). St. Louis: C.V. Mosby.

Banker, B.Q. (1994). Basic reaction of muscle. In A.G. Engel & C. Franzini-Armstrong (Eds.), *Myology* (2nd ed.). New York: McGraw-Hill, Inc.

Baratta, R.V., Solomonow, M., Zhou, B.H. (1998). Frequency domain-based models of skeletal muscle. *J Electromyogr Kinesiol, 8*(2), 79–91.

Bates, B.T., Osternig, L.R., James, S.L. (1977). Fatigue effects in running. *J Motor Behav, 9*, 203–207.

Brobeck, J.R. (Ed.) (1979). *Best and Taylor's Physiological Basis of Medical Practice* (10th ed.) (pp. 59–113). Baltimore: Williams & Wilkins.

Brooke, M.H., & Kaiser, K.K. (1970). Three myosin adenosine triphosphatase systems: The nature of their pH lability and sulfhydryl dependence. *J Histochem Cytochem, 18*, 670.

Buchthal, F., & Schmalbruch, H. (1980). Motor units of mammalian muscle. *Physiol Rev, 60*, 90.

Burke, R.E., Levine, D.N., Zajac, F.E. (1971). Mammalian motor units: Physiological-histochemical correlation in three types of motor units in cat gastrocnemius. *Science, 174*, 709.

Costill, D.L., Coyle, E.F., Fink, W.F., Lesmes, G.R., Witzmann, F.A. (1979). Adaptations in skeletal muscles following strength training. *J Appl Physiol, 46*, 96.

Craig, R. (1994). The structure of the contractile filaments. In A.G. Engel & C. Franzini-Armstrong (Eds.), *Myology* (2nd ed.). New York: McGraw-Hill, Inc.

Crawford, G.N.C., & James, N.T. (1980). The design of muscles. In R. Owen, J. Goodfellow, & P. Bullough (Eds.), *Scientific Foundations of Orthopaedics and Traumatology* (pp. 67–74). London: William Heinemann.

Ciullo, J.V., & Zarins, B. (1983). Biomechanics of the musculotendinous unit: Relation to athletic performance and injury. *Clin Sports Med, 2*, 71.

Eriksson, E., Haggmark, T., Kiessling, K.H., et al. (1981). Effect of electrical stimulation on human skeletal muscle. *Int J Sports Med, 2*, 18.

Galler, S., Hilber, K. (1998). Tension/stiffness ratio of skinned rat skeletal muscle fibre types at various temperatures. *Acta Physiol Scand, 162*(3), 119–126.

Gollnick, P.D. (1982). Relationship of strength and endurance with skeletal muscle structure and metabolic potential. *Int J Sports Med (Suppl), 3*, 26.

Gordon, A.M., Huxley, A.F., & Julian, F.J. (1966). The variation in isometric tension with sarcomere length in vertebrate muscle fibres. *J Physiol, 184*, 170.

Guyton, A.C. (1986). *Textbook of Medical Physiology* (7th ed.). Philadelphia: W.B. Saunders.

Ham, A.W., & Cormack, D.H. (1979). *Histology* (8th ed.). Philadelphia: J.B. Lippincott.

Hill, A.V. (1970). *First and Last Experiments in Muscle Mechanics*. Cambridge: Cambridge University Press.

Hill, D.K. (1968). Tension due to interaction between the sliding filaments of resting striated muscle. The effect of stimulation. *J Physiol (Lond), 199*, 637.

Hislop, H.J., & Perrine, J.J. (1967). The isokinetic concept of exercise. *Phys Ther, 47*, 114.

Huxley, A.F. (1974). Muscular contraction. *J Physiol, 243*, 1.

Huxley, A.F., & Huxley, H.E. (1964). Organizers of a Discussion on the physical and chemical basis of muscular contraction. *Proc R Soc, B160*, 433.

Kannus, P., Jozsa, L., Kvist, M., Järvinen, T., Järvinen, M. (1998a). Effects of immobilization and subsequent low- and high-intensity exercise on morphology of rat calf muscles. *Scand J Med Sci Sports, 8*(3), 160–171.

Kannus, P., Jozsa, L., Järvinen, T.L., Kvist, M., Vieno, T., Järvinen, M. (1998b). Free mobilization and low- to high-intensity exercise in immobilization-induced muscle atrophy. *J Appl Physiol, 84*(4), 1418–1424.

Kasser, J.R. (1996). General Knowledge. In J.R. Kasser (Ed.), *Orthopaedic Knowledge Update 5: Home Study Syllabus*. Illinois: American Academy of Orthopaedic Surgeons.

Keele, C.A., Neil, E., & Joels, N. (1982). Muscle and the nervous system. In *Samson Wright's Applied Physiology* (13th ed.) (pp. 246–259). Oxford: Oxford University Press.

Komi, P.V. (1986). The stretch-shortening cycle and human power output. In N.L. Jones, N. McCartney, & A.J. McComas (Eds.), *Human Muscle Power* (pp. 27–39). Champaign, IL: Human Kinetics Publishers.

Kroemer, K.H.E., Marras, W.S., McGlothlin, J.D., et al. (1990). On the measurement of human strength. *Int J Industr Ergonomics, 6*, 199–210.

Kroll, P.G. (1987). *The effect of previous contraction condition on subsequent eccentric power production in elbow flexor muscles.* (Unpublished doctoral dissertation, New York University, New York.

Lieber, R.L., & Bodine-Fowler, S.C. (1993). Skeletal muscle mechanics: Implications for rehabilitation. *Phys Ther, 73*(12), 844–856.

Linke, W.A., Ivemeyer, M., Mundel, P., Stockmeier, M.R., Kolmerer, B. (1998). Nature of PEVK-titin elasticity in skeletal muscle. *Proc Natl Acad Sci USA, 95*(14), 8052–8057.

Luciano, D.S., Vander, A.J., & Sherman, J.H. (1978). *Human Function and Structure* (pp. 113–136). New York: McGraw-Hill.

Munsat, T.L., McNeal, D., & Waters, R. (1976). Effects of nerve stimulation on human muscle. *Arch Neurol, 33*, 608.

Ohira, Y., Yasui, W., Roy, R.R., Edgerton, V.R. (1997). Effects of muscle length on the response to unloading. *Acta Anat (Basel), 159*(2–3), 90–98.

Ottoson, D. (1983). *Physiology of the Nervous System* (pp. 78–116). New York: Oxford University Press.

Parnianpour, M., Nordin, M., Kahanovitz, N., et al. (1988). The triaxial coupling of torque generation of trunk muscles during isometric exertions and the effect of fatiguing isoinertial movements on the motor output and movement patterns. *Spine, 13*(9).

Pate, E., Bhimani, M., Franks-Skiba, K., Cooke, R. (1995). Reduced effect of pH on skinned rabbit psoas muscle mechanics at high temperatures: Implications for fatigue. *J Physiol (Lond), 486*(Pt 3), 689–694.

Phillips, C.A., & Petrofsky, J.S. (1983). *Mechanics of Skeletal and Cardiac Muscle*. Springfield: Charles C. Thomas.

Saltin, B., et al. (1977). Fiber types and metabolic potentials of skeletal muscles in sedentary man and endurance runners. *Ann NY Acad Sci, 301*, 3.

Sandmann, M.E., Shoeman, J.A., Thompson, L.V. (1998). The fiber-type-specific effect of inactivity and intermittent weight-bearing on the gastrocnemius of 30-month-old rats. *Arch Phys Med Rehabil, 79*(6), 658–662.

Squire, J.M. (1997). Architecture and function in the muscle sarcomere. *Curr Opin Struct Biol, 7*(2), 247–257.

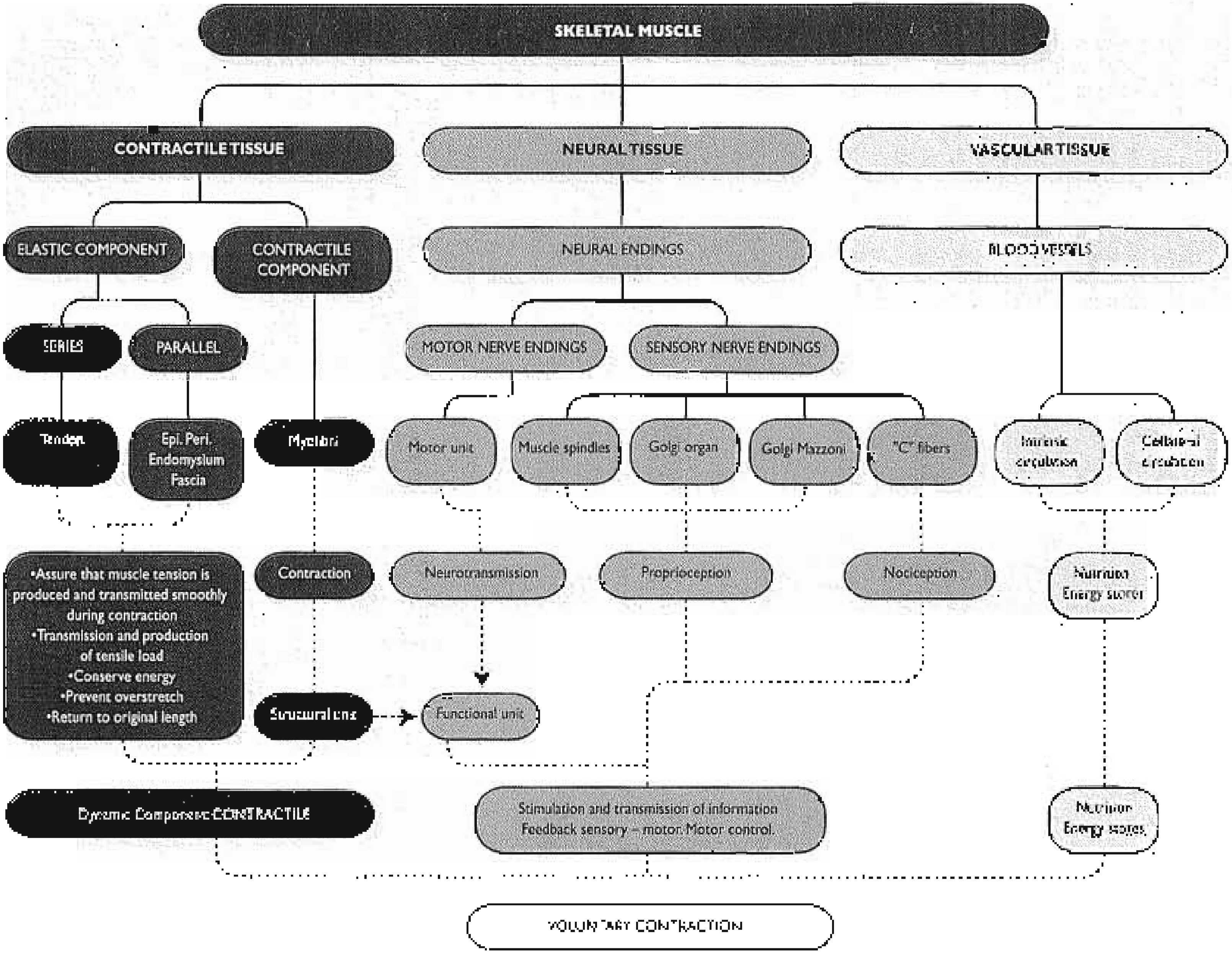
Stromer, M.H. (1998). The cytoskeleton in skeletal, cardiac and smooth muscle cells. *Histol Histopathol, 13*(1), 283–291.

Takarada, Y., Iwamoto, H., Sugi, H., Hirano, Y., Ishii, N. (1997). Stretch-induced enhancement of mechanical work production in frog single fibers and human muscle. *J Appl Physiol, 83*(5), 1741–1748.

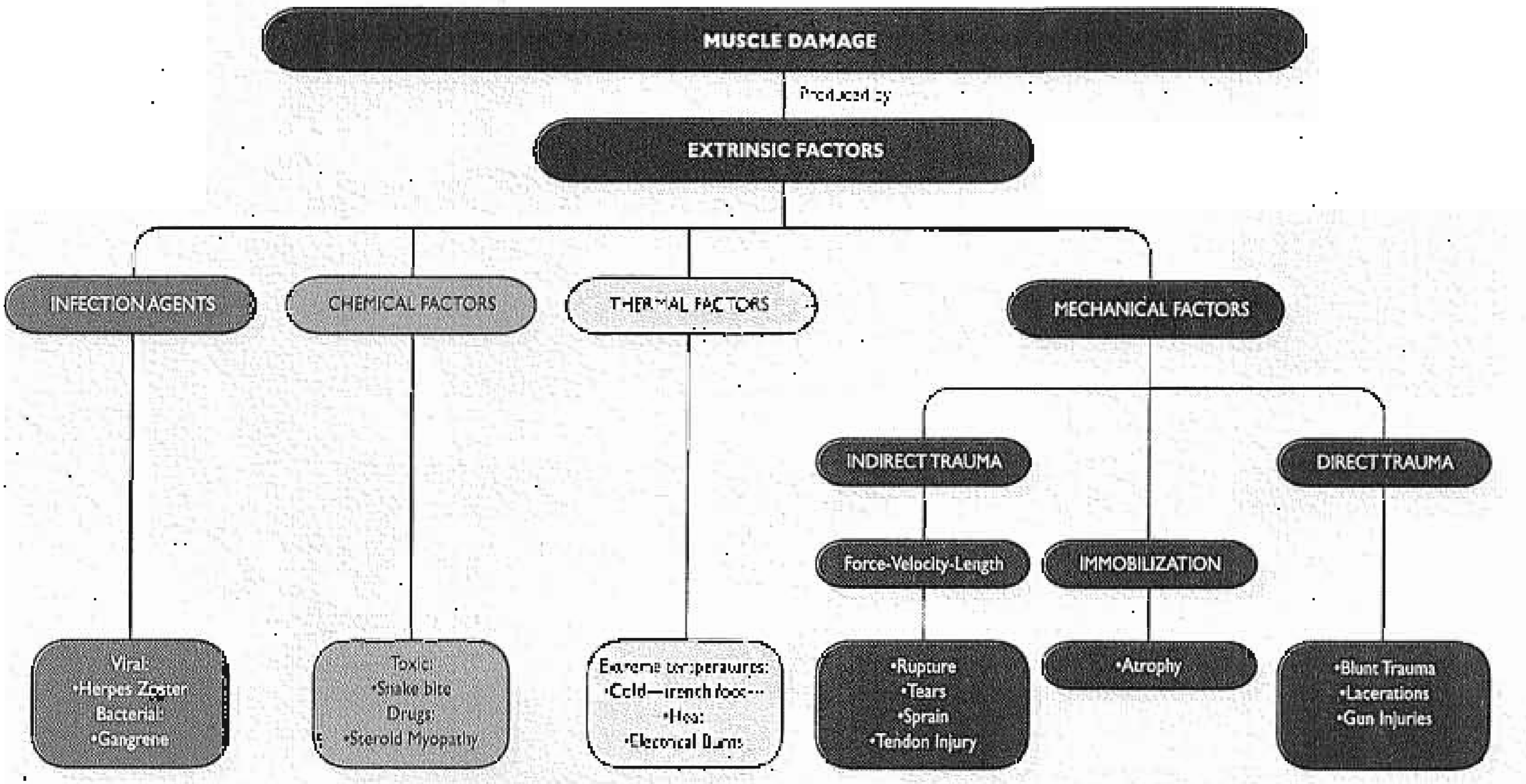
Wilkie, D.R. (1956). The mechanical properties of muscle. *Br Med Bull, 12*, 177.

Wilkie, D.R. (1968). *Muscle*. London: Edward Arnold.

Williams, P. & Warwick, R. (1980). *Gray's Anatomy* (36th ed., pp. 505–515). Edinburgh: Churchill Livingstone.

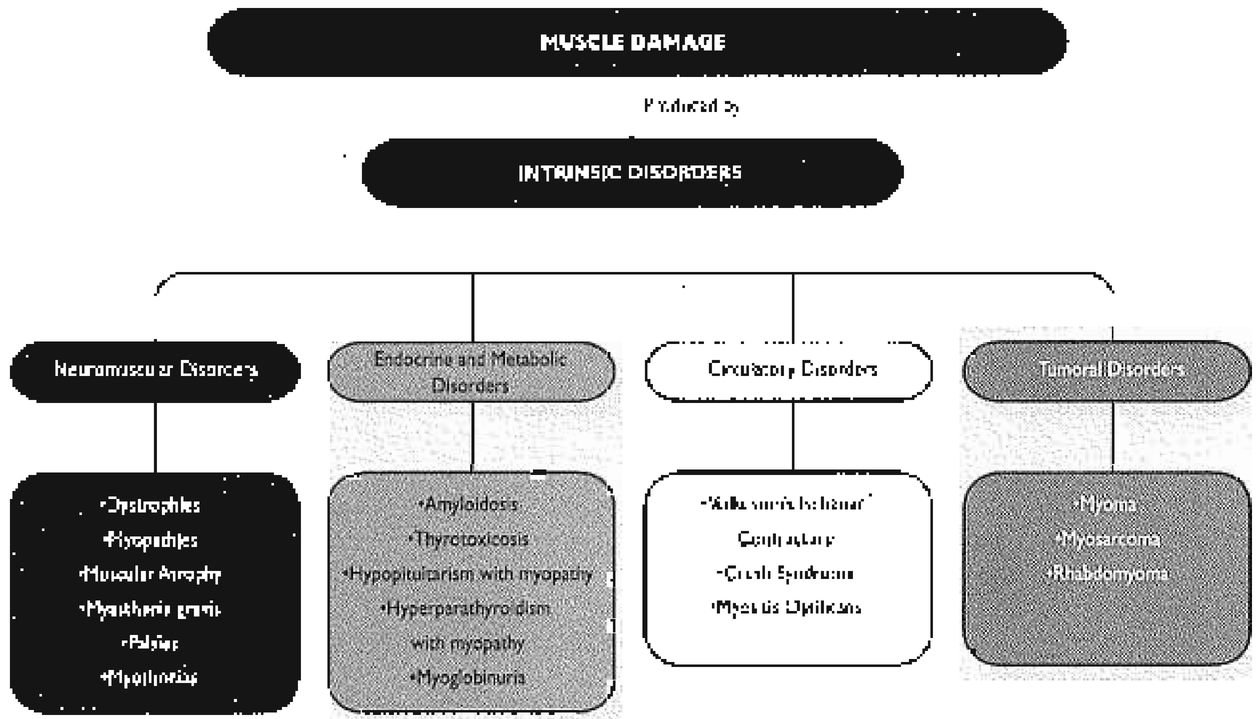


FLOW CHART 6-1 Structure and organization of the skeletal muscle.



FLOW CHART 6-2 Extrinsic factors associated with muscle damage. Clinical examples.*

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.

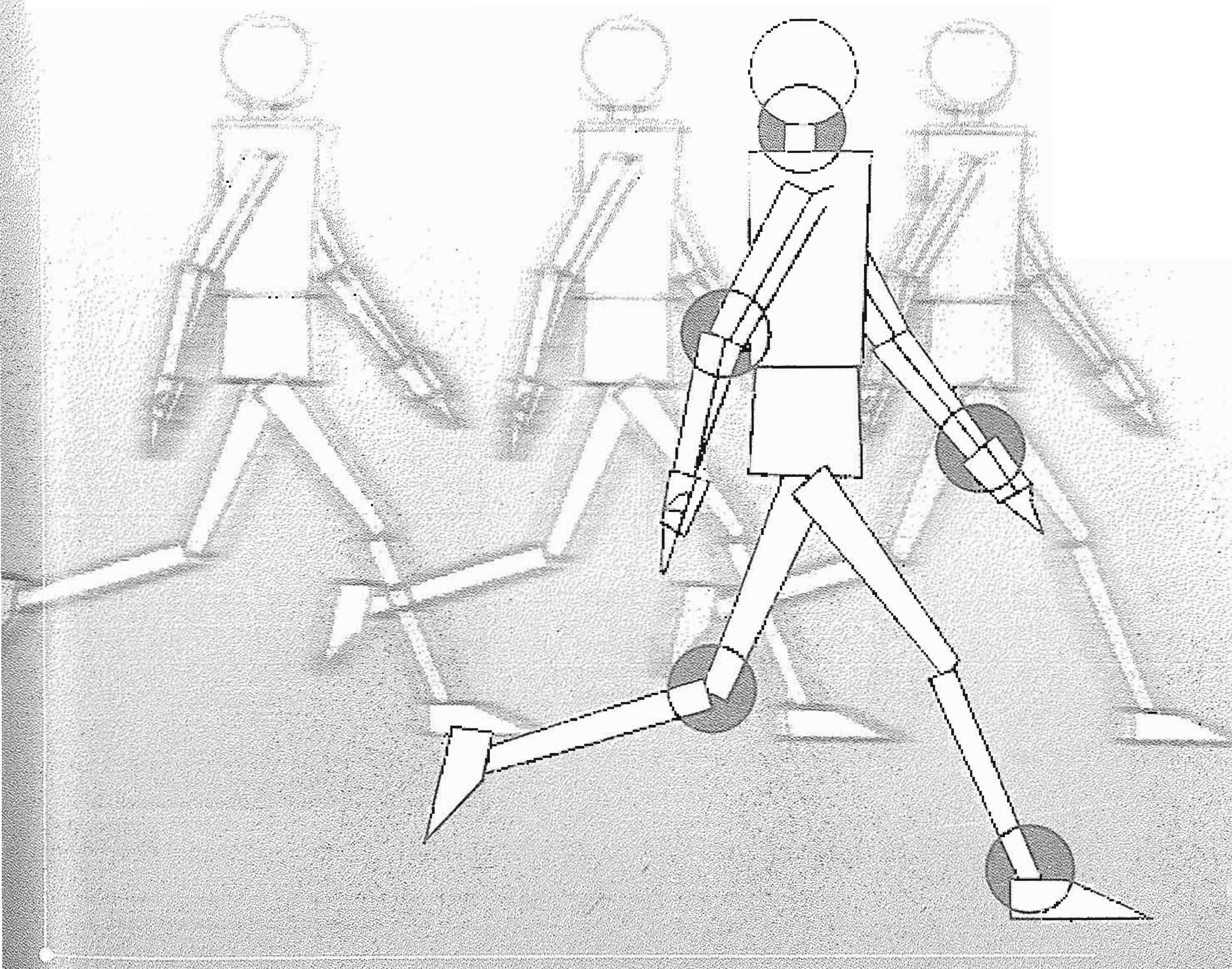


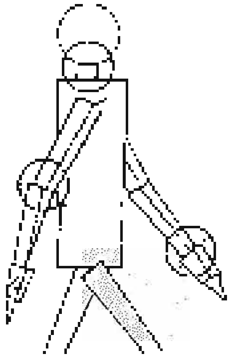
FLOW CHART 6-3 Intrinsic disorders associated with muscle damage. Clinical examples.*

*This flow chart is designed for classroom or group discussion. Flow chart is not meant to be exhaustive.

PART II

Biomechanics of Joints





CHAPTER 7

Biomechanics of the Knee

Margareta Nordin, Victor H. Frankel

Introduction

The knee transmits loads, participates in motion, aids in conservation of momentum, and provides a force couple for activities involving the leg. The human knee, the largest and perhaps most complex joint in the body, is a two-joint structure composed of the tibiofemoral joint and the patellofemoral joint (Fig. 7-1). The knee sustains high forces and moments and is situated between the body's two longest lever arms (the femur and the tibia), making it particularly susceptible to injury. This chapter utilizes the knee to introduce the basic terms, explain the methods, and demonstrate the calculations necessary for analyzing joint motion and the forces and moments acting on a joint. This information is applied to other joints in subsequent chapters.

The knee is particularly well suited for demonstrating biomechanical analyses of joints because these analyses can be simplified in the knee and still yield useful data. Although knee motion occurs simultaneously in three planes, the motion in one plane is so great that it accounts for nearly all of the motion. Also, although many muscles produce forces on the knee, at any particular instant one muscle group predominates, generating a force so great that it accounts for most of the muscle force acting on the knee. Thus, basic biomechanical analyses can be limited to motion in one plane and to the force produced by a single muscle group and still give an understanding of knee motion and an estimation of the magnitude of the principal forces and moments on the knee. Advanced biomechanical dynamic analyses of the knee joint that include all

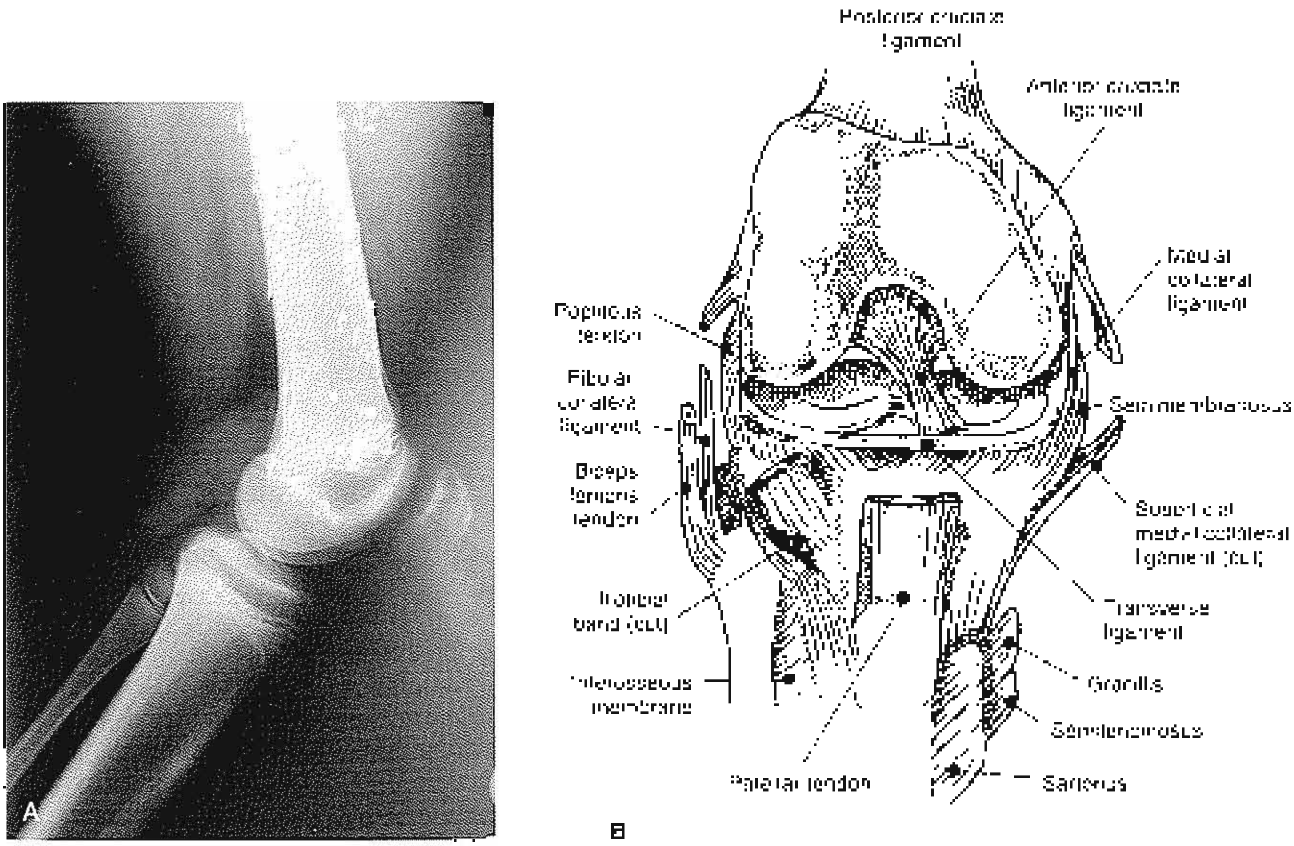


FIG. 7-1

Two-joint structure of the knee. A, Lateral view of a knee joint with open growth plates. B, Anterior view without patella.

soft tissue structures are complex and still under investigation.

Analysis of motion in any joint requires the use of kinematic data. Kinematics is the branch of mechanics that deals with motion of a body without reference to force or mass. Analysis of the forces and moments acting on a joint necessitates the use of both kinematic and kinetic data. Kinetics is the branch of mechanics that deals with the motion of a body under the action of given forces and/or moments.

Kinematics

Kinematics defines the range of motion and describes the surface motion of a joint in three planes: frontal (coronal or longitudinal), sagittal, and transverse (horizontal) (Fig. 7-2, *A* & *B*). Clinical measurements of joint range of motion define the anatomical position as a zero position for measurement. This taxonomy will be used for joint motion throughout this book. Other taxonomies and reference systems exist (Andriacchi et al., 1979; Grood &

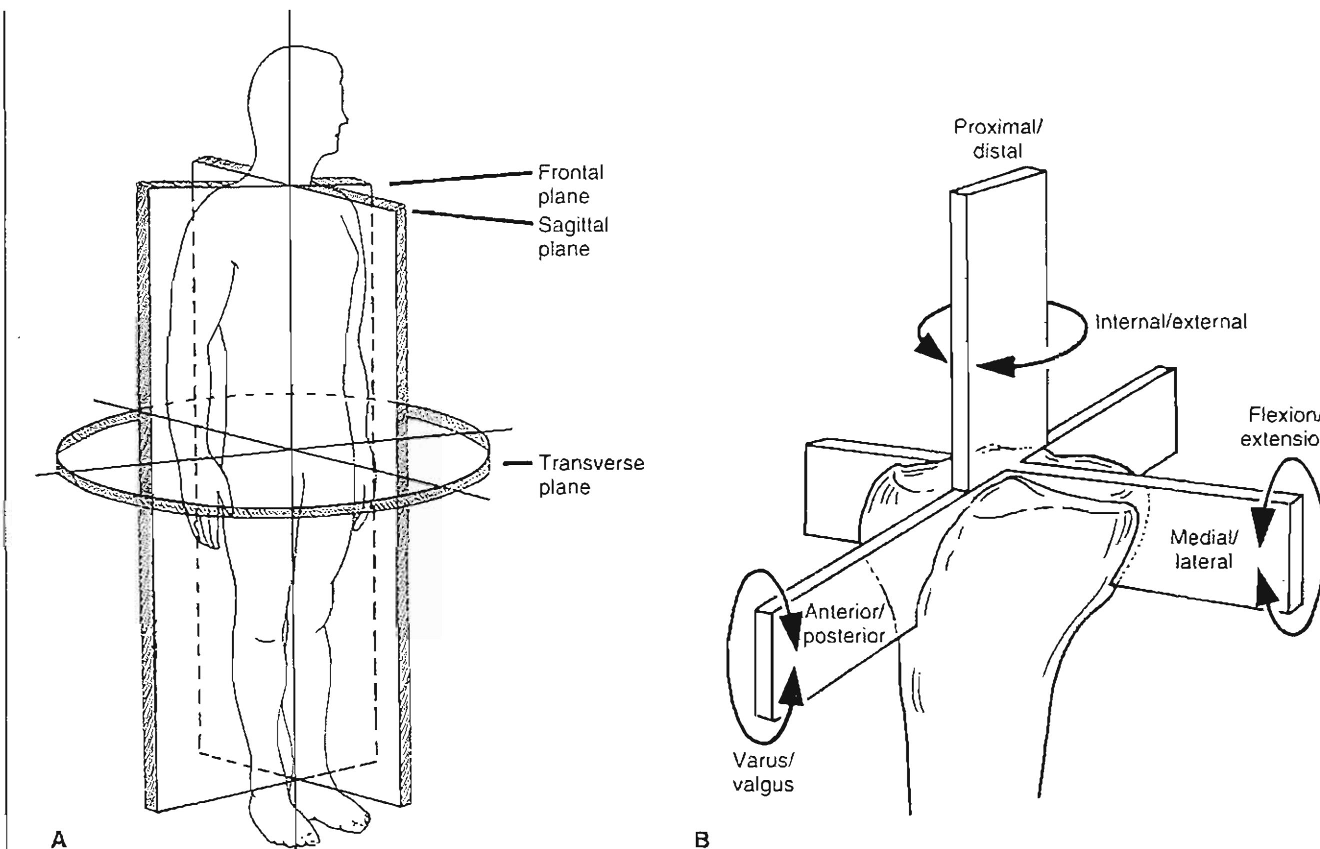


FIG. 7-2

A, Frontal (coronal or longitudinal), sagittal, and transverse (horizontal) planes in the human body performed easily for both the tibiofemoral and the patellofemoral joint. B, Depiction and nomenclature of the six degrees of freedom of knee motion: anterior posterior translation, medial/lateral translation and proximal distal translation, flexion-extension rotation, internal-external rotation, varus-valgus rotation. Adapted from Wilson, S.A., Vigorita, V.J., & Scott, W.N. (1994). *Anatomy. In N. Scott (Ed.), The Knee (p. 17). Philadelphia: Mosby-Year Book.*

Surlas, 1983; Kroemer et al., 1990; Özkaya & Nordin, 1999), but the anatomical reference system by far is the most commonly used among clinicians. Of the two joints composing the knee, the tibiofemoral joint lends itself particularly well to an analysis of range of joint motion. Analysis of surface joint motion can be performed easily for both the tibiofemoral and the patellofemoral joint. Any impediment of range of motion or surface joint motion will disturb the normal loading pattern of a joint and bear consequences.

RANGE OF MOTION

The range of motion of any joint can be measured in any plane. Gross measurements can be made with a goniometer, but more specific measurements require the use of more precise methods such as electrogoniometry, roentgenography, stereophotogrammetry, or photographic and video techniques using skeletal pins.

In the tibiofemoral joint, motion takes place in all three planes, but the range of motion is greatest by far in the sagittal plane. Motion in this plane from full extension to full flexion of the knee is from 0° to approximately 140°.

Motion in the transverse plane, internal and external rotation, is influenced by the position of the joint in the sagittal plane. With the knee in full extension, rotation is almost completely restricted by the interlocking of the femoral and tibial condyles, which occurs mainly because the medial femoral condyle is longer than the lateral condyle. The range of rotation increases as the knee is flexed, reaching a maximum at 90° of flexion; with the knee in this position, external rotation ranges from 0° to approximately 45° and internal rotation ranges from 0° to approximately 30°. Beyond 90° of flexion, the range of internal and external rotation decreases, primarily because the soft tissues restrict rotation.

Motion in the frontal plane, abduction and adduction, is similarly affected by the amount of joint flexion. Full extension of the knee precludes almost all motion in the frontal plane. Passive abduction and adduction increase with knee flexion up to 30°, but each reaches a maximum of only a few degrees. With the knee flexed beyond 30°, motion in the frontal plane again decreases because of the limiting function of the soft tissues.

The range of tibiofemoral joint motion required for the performance of various physical activities can be determined from kinematic analysis. Motion

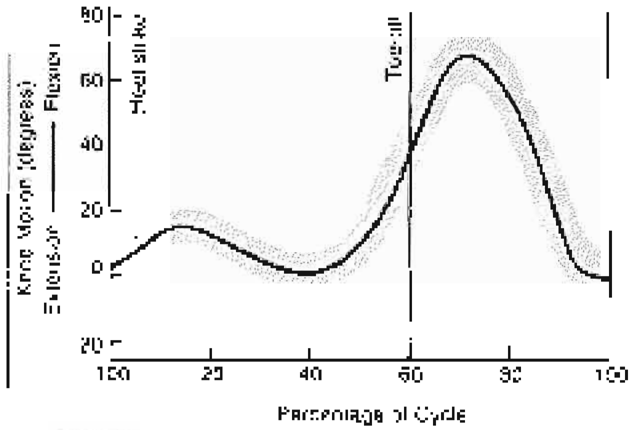


FIG. 7-3

Range of motion of the tibiofemoral joint in the sagittal plane during level walking in one gait cycle. The shaded area indicates variation among 60 subjects (age range 20 to 65 years). *Adapted from Murray, M.P., Drought, A.B., Kory, R.C. (1964). Walking patterns of normal men. J Bone Joint Surg, 46A, 335.*

in this joint during walking has been measured in all planes. The range of motion in the sagittal plane during level walking was measured with an electrogoniometer by Lamoreux (1971) and Murray et al. (1964). Full or nearly full extension was noted at the beginning of the stance phase (0% of cycle) at heel strike, and at the end of the stance phase before toe-off (around 60% of cycle) (Fig. 7-3). Maximum flexion (approximately 60°) was observed during the middle of the swing phase (see Chapter 18, Biomechanics of Gait, for more detailed information). These measurements are velocity-dependent and must be interpreted with caution.

Motion in the transverse plane during walking has been measured by several investigators. Using a photographic technique involving the placement of skeletal pins through the femur and tibia, Levens and associates (1948) found that total rotation of the tibia with respect to the femur ranged from approximately 4 to 13° in 12 subjects (mean 8.6°). Greater rotation (mean 13.3°) was noted by Kettelkamp and coworkers (1970), who used electrogoniometry on 22 subjects. In both studies, external rotation began during knee extension in the stance phase and reached a peak value at the end of the swing phase just before heel strike. Internal rotation was noted during flexion in the swing phase.

TABLE 7-1

Range of Tibiofemoral Joint Motion in the Sagittal Plane During Common Activities

Activity	Range of Motion from Knee Extension to Knee Flexion (Degrees)
Walking	0–67[a]
Climbing stairs	0–83[b]
Descending stairs	0–90
Sitting down	0–93
Tying a shoe	0–106
Lifting an object	0–117

[a]Data from Kettelkamp et al. (1970). Mean for 22 subjects. A slight difference was found between right and left knees (mean for right knee, 68.1°; mean for left, 63.7°).

[b]These and subsequent data from Laubenthal et al. (1972). Mean for 30 subjects.

Kettelkamp's group (1970) also measured motion in the frontal plane during walking. In nearly all of the 22 subjects, maximal abduction of the tibia was observed during extension at heel strike and at the beginning of the stance phase; maximal adduction occurred as the knee was flexed during the swing phase. The total amount of abduction and adduction averaged 11°.

Values for the range of motion of the tibiofemoral joint in the sagittal plane during several common activities are presented in Table 7-1. Maximal knee flexion occurs during lifting. A range of motion from full extension to at least 117° of flexion appears to be required for an individual to carry out the activities of daily living in a normal manner. Any restriction of knee motion can be compensated for by increased motion in other joints. In studying the range of tibiofemoral joint motion during various activities, researchers found that an increased speed of movement requires a greater range of motion in the tibiofemoral joint (Holden et al., 1997; Perry et al., 1977). As the pace accelerates from walking slowly to running, progressively more knee flexion is needed during the stance phase (Table 7-2).

SURFACE JOINT MOTION

Surface joint motion, which is the motion between the articulating surfaces of a joint, can be described for any joint in any plane with the use of stereophotogrammetric methods (Selvik, 1978, 1983). Because these methods are highly technical and complex, a simpler method evolved in the nineteenth century is still used (Reuleaux, 1876). This method, called the instant center technique, allows surface joint motion to be analyzed in the sagittal and frontal planes but not in the transverse plane. The instant center technique provides a description of the relative uniplanar motion of two adjacent segments of a body and the direction of displacement of the contact points between these segments.

The skeletal portion of a body segment is called a link. As one link rotates about the other, at any instant there is a point that does not move, that is, a point that has zero velocity. This point constitutes an instantaneous center of motion, or instant center. The instant center is found by identifying the displacement of two points on a link as the link moves from one position to another in relation to an adjacent link, which is considered to be stationary. The points on the moving link in its original position and in its displaced position are designated on a graph and lines are drawn connecting the two sets of points. The perpendicular bisectors of these two lines are then drawn. The intersection of the perpendicular bisectors is the instant center.

Clinically, a pathway of the instant center for a joint can be determined by taking successive roentgenograms of the joint in different positions (usually 10° apart) throughout the range of motion in one plane and applying the Reuleaux method for locating the instant center for each interval of motion.

When the instant center pathway has been determined for joint motion in one plane, the surface joint motion can be described. For each interval of motion, the point at which the joint surfaces make contact is located on the roentgenograms used for the instant center analysis, and a line is drawn from

TABLE 7-2

Amount of Knee Flexion During Stance Phase of Walking and Running

Activity	Range in Amount of Knee Flexion During Stance Phase (Degrees)
Walking	
Slow	0–6
Free	6–12
Fast	12–18
Running	18–30

Data from Perry et al. (1977). Mean for seven subjects.

the instant center to the contact point. A second line drawn at right angles to this line indicates the direction of displacement of the contact points. The direction of displacement of these points throughout the range of motion describes the surface motion in the joint. In most joints, the instant centers lie at a distance from the joint surface, and the line indicating the direction of displacement of the contact points is tangential to the load-bearing surface, demonstrating that one joint surface is gliding on the other (load-bearing) surface. In the case in which the instant center is found on the surface, the joint has a rolling motion and there is no gliding function. Because the instant center technique allows a description of motion in one plane only, it is not useful for describing the surface joint motion if more than 15° of motion takes place in any plane other than the one being measured.

In the knee, surface joint motion occurs between the tibial and femoral condyles and between the femoral condyles and the patella. In the tibiofemoral joint, surface motion takes place in all three planes simultaneously but is considerably less in the transverse and frontal planes. Surface motion in the patellofemoral joint takes place in two planes simultaneously, the frontal and transverse, but is far greater in the frontal plane.

Tibiofemoral Joint

An example will illustrate the use of the instant center technique to describe the surface motion of the tibiofemoral joint in the sagittal plane. To determine the pathway of the instant center of this joint during flexion, a lateral roentgenogram is taken of the knee in full extension and successive films are taken at 10° intervals of increased flexion. Care is taken to keep the tibia parallel to the x-ray table and to prevent rotation about the femur. When a patient has limited knee motion, the knee is flexed or extended only as far as the patient can tolerate.

Two points on the femur that are easily identified on all roentgenograms are selected and desig-

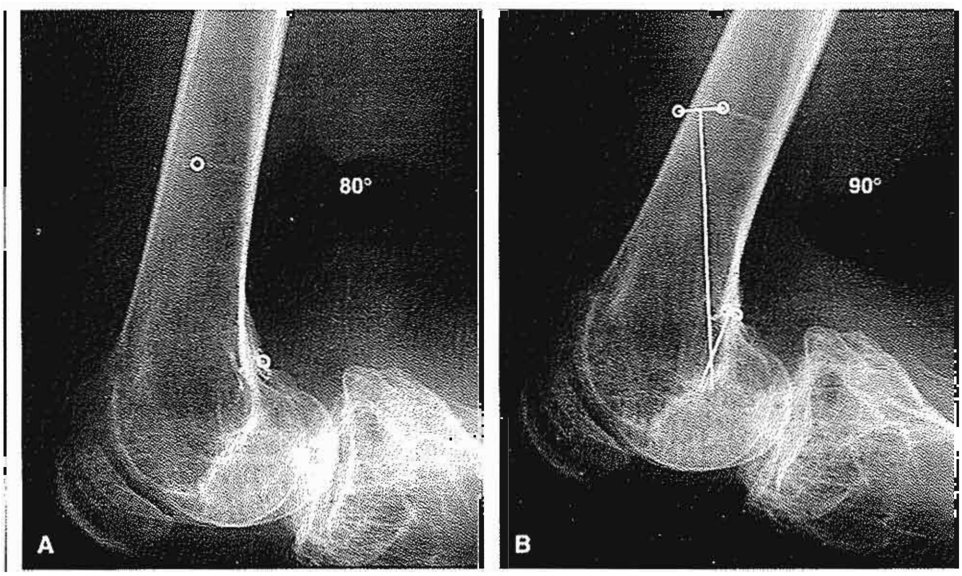


FIG. 7-4

Locating the instant center. A, Two easily identifiable points on the femur are designated on a roentgenogram of a knee flexed 80°. B, This roentgenogram is compared with a roentgenogram of the knee flexed 90°, on which the same two points have been indicated. The images of the tibiae are superimposed, and lines are drawn connecting each set of points. The perpendicular bisectors of these two lines are then drawn. The point at which these perpendicular bisectors intersect is the instant center of the tibiofemoral joint for the motion between 80 and 90° of flexion. *Courtesy of Ian Goldie, M.D., University of Gothenburg, Gothenburg, Sweden.*

fixed on each roentgenogram (Fig. 7-4A). The films are then compared in pairs, with the images of the tibiae superimposed on each other. Roentgenograms with marked differences in tibial alignment are not used. Lines are drawn between the points on the femur in the two positions, and the perpendicular bisectors of these lines are then drawn. The point at which these perpendicular bisectors intersect is the instant center of the tibiofemoral joint for each 10° interval of motion (Fig. 7-4B). The instant center pathway throughout the entire range of knee flexion and extension can then be plotted. In a normal knee, the instant center pathway for the tibiofemoral joint is semicircular (Fig. 7-5).

After the instant center pathway has been determined for the tibiofemoral joint, the surface motion can be described. On each set of superimposed roentgenograms the point of contact of the tibiofemoral joint surfaces (the narrowest point in the joint space) is determined and a line is drawn connecting this point with the instant center. A second line drawn at right angles to this line indicates the direction of displacement of the contact points. In a normal knee, this line is tangential to the surface of the tibia for each interval of motion from full extension to full flexion, demonstrating that the femur is gliding on the tibial condyles (Frankel et al., 1971) (Fig. 7-6). During normal knee motion in the sagittal plane from full extension to full flexion, the instant center pathway moves posteriorly, forcing a combination of rolling and sliding to occur between the articular surface (Fig 7-6, *A & B*). The unique mechanism prevents the femur from rolling off the posterior aspect of the tibia plateau as the knee goes into increased flexion (Draganich et al., 1987; Fu et al., 1994; Kapandji, 1970). The mechanism that prevents this roll-off is the link formed between the tibial and femoral attachment sites of the anterior and posterior cruciate ligaments and the osseous geometry of the femoral condyles (Fu et al., 1994) (Fig. 7-6, *B-D*).

Frankel and associates (1971) determined the instant center pathway and analyzed the surface motion of the tibiofemoral joint from 90° of flexion to full extension in 25 normal knees; tangential gliding was noted in all cases. They also determined the instant center pathway for the tibiofemoral joint in 30 knees with internal derangement and found that, in all cases, the instant center was displaced from the normal position during some portion of the motion examined. The abnormal instant center pathway for one subject, a 35-year-old man with a bucket-handle derangement, is shown in Figure 7-7.

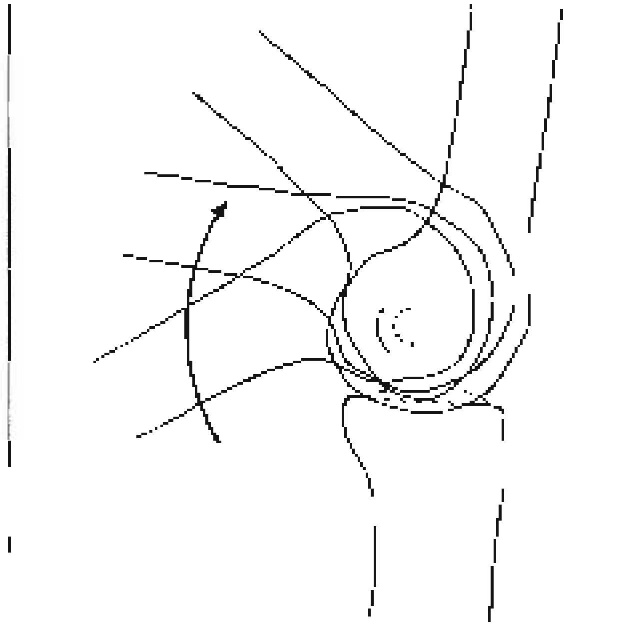


FIG. 7-5

Semicircular instant center pathway for the tibiofemoral joint in a 19-year-old man with a normal knee.

If the knee is extended and flexed about an abnormal instant center pathway, the tibiofemoral joint surfaces do not glide tangentially throughout the range of motion but become either distracted or compressed (Fig. 7-8). Such a knee is analogous to a door with a bent hinge that no longer fits into the door jamb. If the knee is continually forced to move about a displaced instant center, a gradual adjustment to the situation will be reflected either by stretching of the ligaments and other supporting soft tissues or by the imposition of abnormally high pressure on the articular surfaces.

Internal derangements of the tibiofemoral joint may interfere with the so-called screw-home mechanism, which is external rotation during extension of the tibia (Fig. 7-9). The tibiofemoral joint is not a simple hinge joint; it has a spiral, or helicoid, motion. The spiral motion of the tibia about the femur during flexion and extension results from the anatomical configuration of the

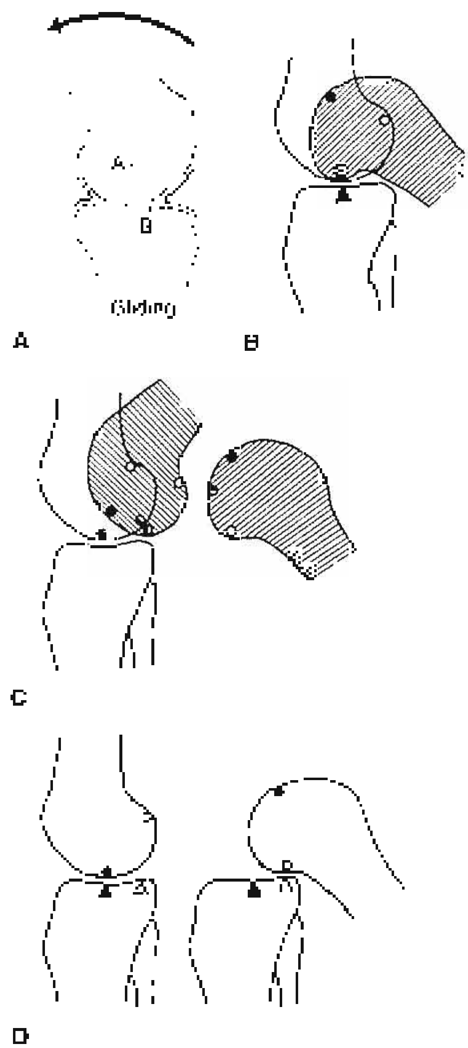


FIG. 7-6

A, In a normal knee, a line drawn from the instant center of the tibiofemoral joint to the tibiofemoral contact point (line A) forms a right angle with a line tangential to the tibial surface (line B). The arrow indicates the direction of displacement of the contact points. Line B is tangential to the tibial surface, indicating that the femur glides on the tibial condyles during the measured interval of motion. B, Pure sliding of the femur on the tibia with knee extension. Note that the contact point of the tibia does not change as the femur slides over it. Eventually impingement would occur if all surface motion was restricted to sliding. Round points delineate contact points at the femur and triangles delineate contact points at the tibia. C, Pure rolling of the femur on the tibia with knee flexion. Note that both the tibia and the femoral contact points change as the femur rolls on the tibia. Also note that with moderate flexion, the femur will begin to roll off the tibia if surface motion was restricted to rolling. D, Actual knee motion including both sliding and rolling

medial femoral condyle in a normal knee. This condyle is approximately 1.7 cm longer than the lateral condyle. As the tibia moves on the femur from the fully flexed to the fully extended position, it descends and then ascends the curves of the medial femoral condyle and simultaneously rotates externally. This motion is reversed as the tibia moves back into the fully flexed position. This screw-home mechanism (rotation around the longitudinal axis of the tibia) provides more stability to the knee in any position than would a simple hinge configuration of the tibiofemoral joint.

Matsumoto et al. (2000) investigated the axis of tibia axial rotation and its change with knee flexion angle in 24 fresh-frozen normal knee cadaver specimens ranging in age from 22 to 67 years. The magnitude and location of the longitudinal axis of tibia rotation were measured at 15° increments between 0 and 90° of knee flexion. The magnitude of tibia rotation was 8° at 0°

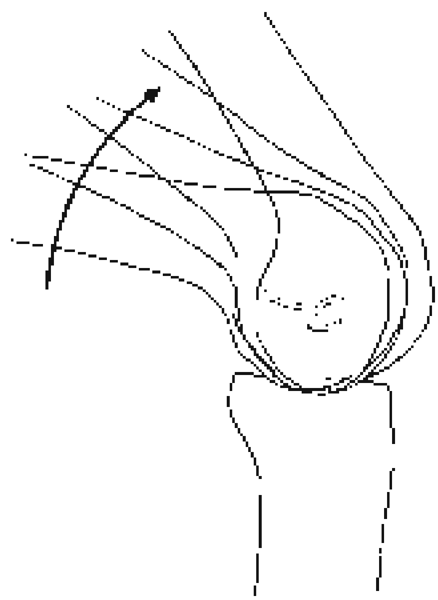


FIG. 7-7

Abnormal instant center pathway for a 35 year old man with a bucket-handle derangement. The instant center jumps at full extension of the knee. *Adapted from Frankel, V.H., Burstein, A.H., & Brooks, D.B. (1971). Biomechanics of internal derangement of the knee. Pathomechanics as determined by analysis of the instant centers of motion. J Bone Joint Surg, 53A, 945.*

of knee flexion. The tibial rotation increased rapidly as the knee flexion angle increased and reached a maximum of 31° at 30° of knee flexion. It then decreased again with additional flexion (Fig. 7-10). The location of the longitudinal rotational axis was close to the insertion of the anterior cruciate ligament (ACL) at 0° of flexion. At continuous flexion up to 60°, the rotational axis moved toward the insertion of the posterior cruciate ligament. Between 60 and 90° of flexion, the rotational axis moved anteriorly again (Fig. 7-11). This study showed that the rotational axis remains approximately in the area between the two cruciate ligaments. Any change of direction and tension of the cruciate ligaments and surrounding soft tissue may affect the movement and the location of the longitudinal tibial axis of rotation and thereby affect joint load distribution.

A clinical test, the Helfet test, is often used to determine whether external rotation of the tibiofemoral joint takes place during knee extension, thereby indicating whether the screw-home mechanism is intact. This clinical test is performed with the patient sitting with the knee and hip flexed 90° and the leg hanging free. The medial and lateral borders of the patella are marked on the skin. The tibial tuberosity and the midline of the patella are then designated, and the alignment of the tibial tuberosity with the patella is checked. In a normal knee flexed 90°, the tibial tuberosity aligns with the medial half of the patella (Fig. 7-12A). The knee is then extended fully and the movement of the tibial tuberosity is observed. In a normal knee, the tibial tuberosity moves laterally during extension and aligns with the lateral half of the patella at full extension (Fig. 7-12B). Rotatory motion in a normal knee may be as great as half the width of the patella. In a deranged knee, the tibia may not rotate externally during extension. Because of the altered surface motion in such a knee, the tibiofemoral joint will be abnormally compressed if the knee is forced into extension, and the joint surfaces may be damaged.

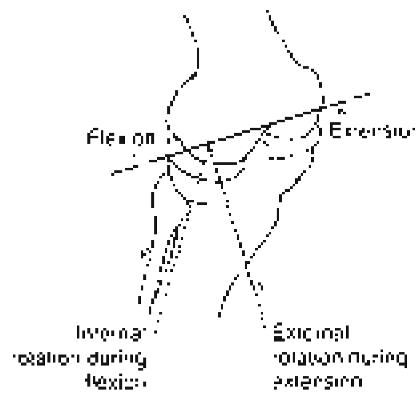


FIG. 7-9

Screw-home mechanism of the tibiofemoral joint. During knee extension, the tibia rotates externally. This motion is reversed as the knee is flexed. A, Oblique view of the femur and tibia. The shaded area indicates the tibial plateau; solid line, axis for knee flexion and extension; dotted lines, internal and external rotation axis of the tibia during flexion and extension. *Adapted from Helfet, A.J. (1974). Anatomy and mechanics of movement of the knee joint. In A. Helfet (Ed.). Disorders of the Knee (pp. 1–17). Philadelphia: J.B. Lippincott.*

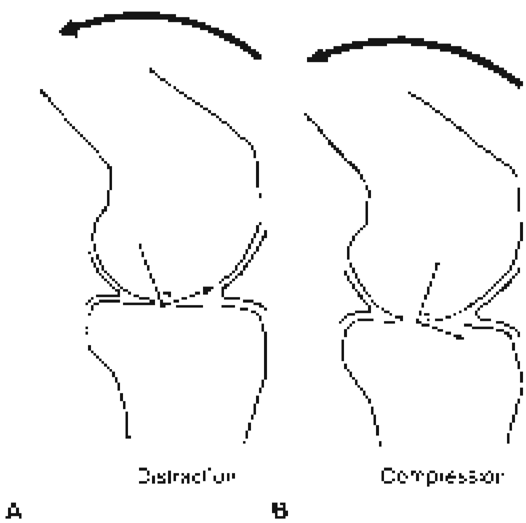


FIG. 7-8

Surface motion in two tibiofemoral joints with displaced instant centers. In both joints, the arrowed line at right angles to the line between the instant center and the tibiofemoral contact point indicates the direction of displacement of the contact points. A, The small arrow indicates that with further flexion, the tibiofemoral joint will be distracted. B, With increased flexion, this joint will be compressed.

Patellofemoral Joint

The surface motion of the patellofemoral joint in the frontal plane may also be described by means of the instant center technique. This joint is shown to have a gliding motion (Fig. 7-13). From full extension to full flexion of the knee, the patella glides caudally approximately 7 cm on the femoral condyles. Both the medial and lateral facets of the femur articulate with the patella from full extension to 140° of flexion (Hehne, 1990) (Fig. 7-14). Beyond 90° of flexion, the patella rotates externally, and only the medial femoral facet articulates with the patella (Fig. 7-14B). At full flexion, the patella sinks into the intercondylar groove (Goodfellow et al., 1976). The contact area of the lateral facet joint of the patella is larger than the medial contact areas and ranges from 0.5 to 2.5 cm² and less than 0.5 to 2 cm², respectively. Contact areas increase with an increased amount of flexion of the knee joint and increased pulling force of the quadriceps muscle (Hehne, 1990).

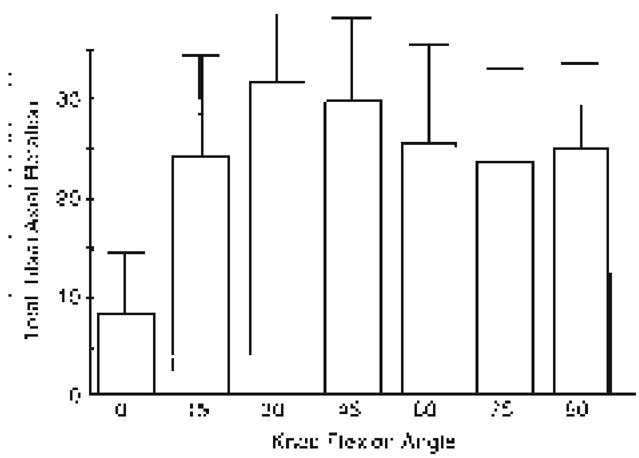


FIG. 7-10

The total tibial axial rotation (y-axis) plotted against the magnitude of knee flexion (x-axis) in fresh-frozen specimens tested under unloaded axial conditions. The magnitude of tibial rotation was below 9° at 0° of knee flexion. Maximum rotation (31.7°) was measured between 30 to 45° of knee flexion. *Reprinted with permission from Matsumoto, H., Seedhom, B.B., Suda, Y., et al. (2000). Axis of tibial rotation and its change with flexion angle. Clin Orthop, 371, 178–182.*

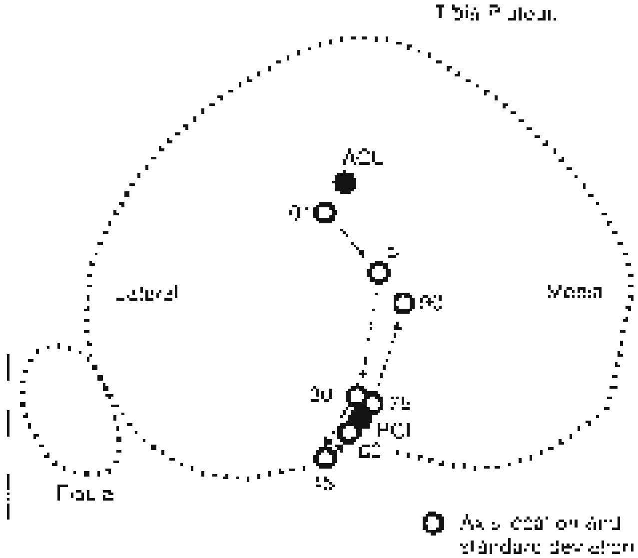
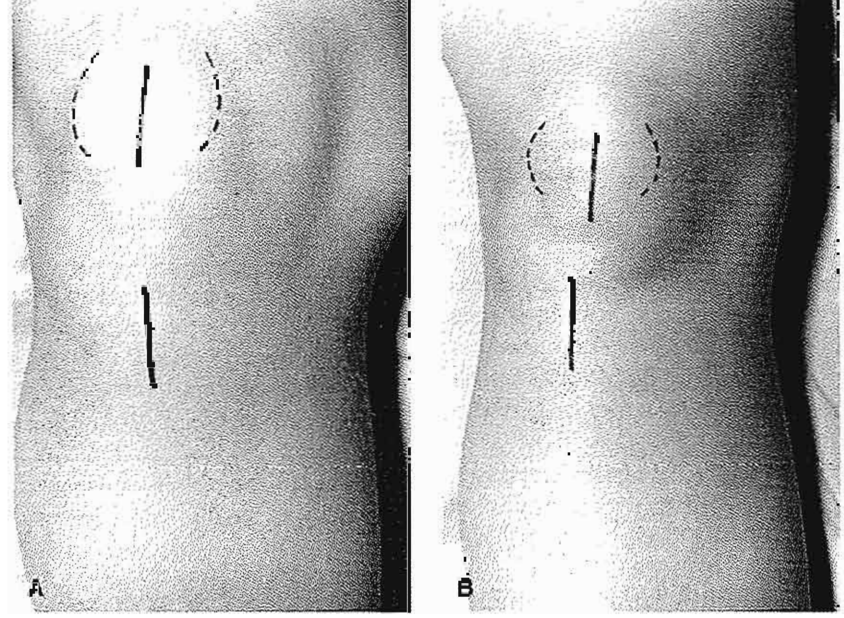


FIG. 7-11

Location of the axis of tibial axial rotation. The location of the axis was close to the tibial insertion of the anterior cruciate ligament (ACL) at 0° of flexion and gradually moved toward that of the posterior cruciate ligament at 45 and 90° of knee flexion. The axis then moved anteriorly again and was approximately at equal distance from the two insertions of the cruciate ligaments at 90° of knee flexion. ACL, insertion of the anterior cruciate ligament; PCL, tibial insertion of the posterior cruciate ligament. *Reprinted with permission from Matsumoto, H., Seedhom, B.B., Suda, Y., et al. (2000). Axis of tibial rotation and its change with flexion angle. Clin Orthop, 371, 178–182.*

Kinetics

Kinetics involves both static and dynamic analysis of the forces and moments acting on a joint. Statics is the study of the forces and moments acting on a body in equilibrium (a body at rest or moving at a constant speed). For a body to be in equilibrium, two equilibrium conditions must be met: force (translatory) equilibrium, in which the sum of the forces is zero, and moment (rotatory) equilibrium, in which the sum of the moments is zero. Dynamics is the study of the moments and forces acting on a body in motion (an accelerating or decelerating body). In this case, the forces do not add up to zero and the body displaces and/or the moments do not add up to zero and the body



Knee flexed 90°   Knee fully extended

FIG. 7-12.

Helfet test. A, In a normal knee flexed 90°, the tibial tuberosity aligns with the medial half of the patella. B, When the knee is fully extended, the tibial tuberosity aligns with the lateral half of the patella.

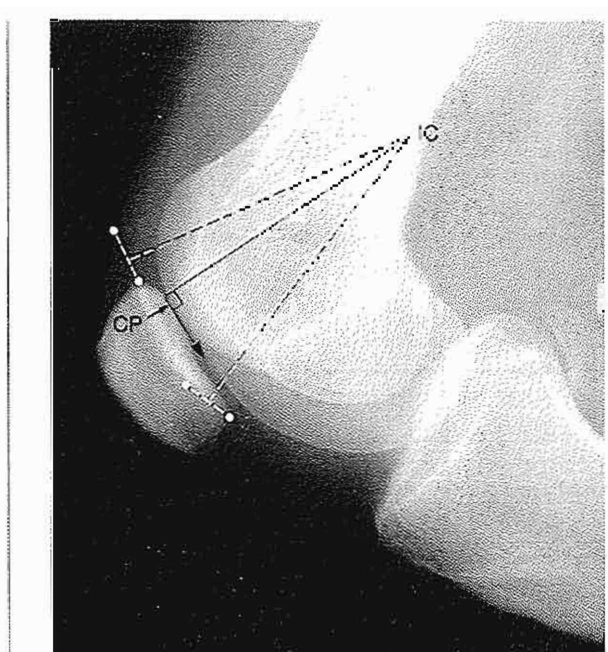


FIG. 7-13.

After the instant center (IC) is determined for the patellofemoral joint for the motion from 75 to 90° of knee flexion, a line is drawn from the instant center to the contact point (CP) between the patella and the femoral condyle. A line drawn at right angles to this line is tangential to the surface of the patella, indicating gliding.

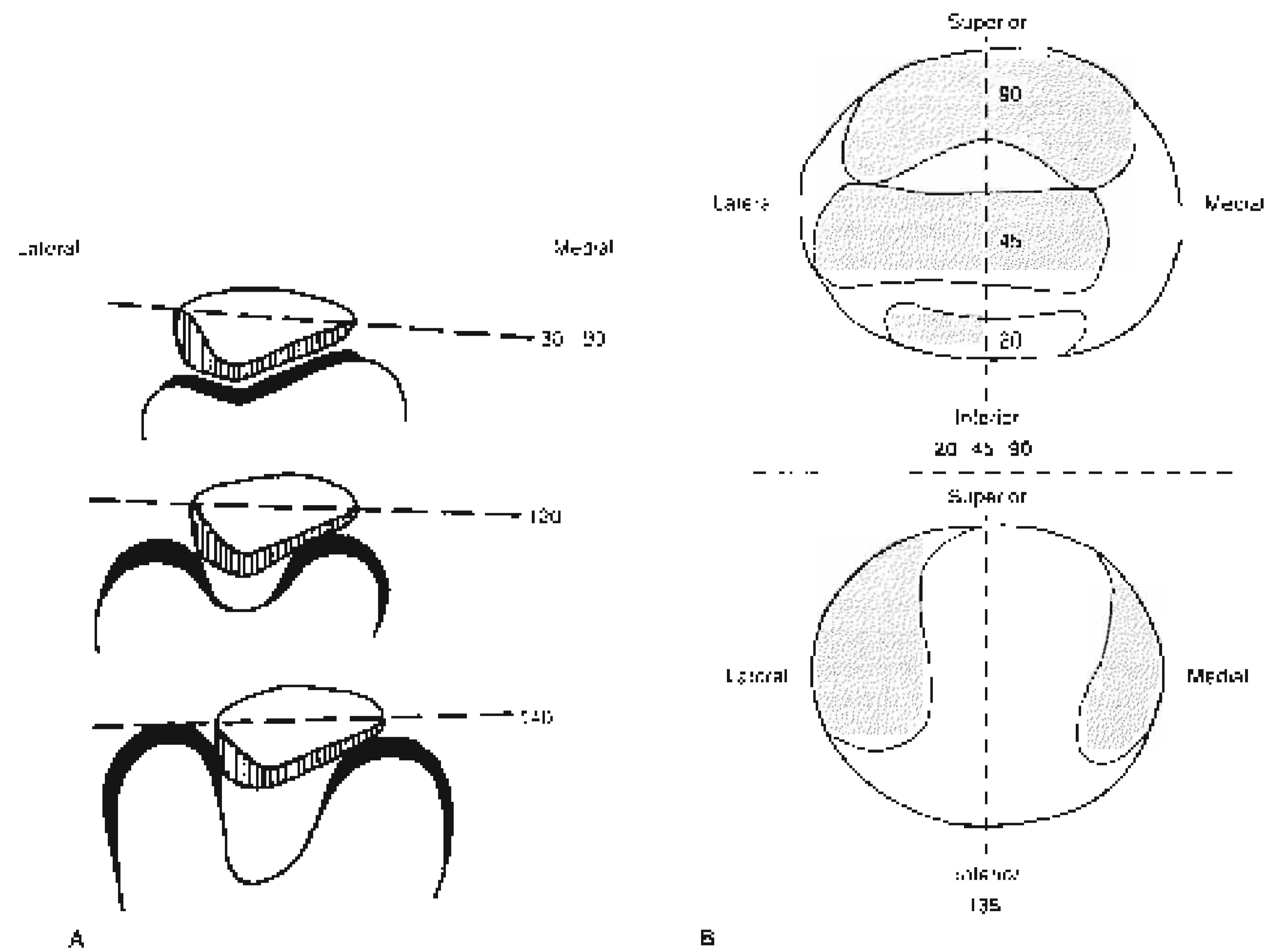


FIG. 7-14

A, The position of the patella at different ranges of knee flexion motion. B, Contact areas during different degrees of flexion. Beyond 90° of flexion, the patella rotates slightly outwards. *Adapted from Goodfellow, J., Hungerford, D.S., & Zindel, M. (1976). Patellofemoral joint mechanics and pathology. 1. Functional anatomy of the patellofemoral joint. J Bone Joint Surg, 58B, 287, and from Hehne, H.J. (1990). Biomechanics of the patellofemoral joint and its clinical relevance. Clin Orthop, 258, 73–85.*

rotates around an axis perpendicular to the plane of the forces producing the moments. Kinetic analysis allows one to determine the magnitude of the moments and forces on a joint produced by body weight, muscle action, soft tissue resistance, and externally applied loads in any situation, either static or dynamic, and to identify those situations that produce excessively high moments or forces.

In this and subsequent chapters, the discussion of statics and dynamics of the joints of the skeletal system concerns the magnitude of the forces and moments acting to move a joint about an axis or to maintain its position. It does not take into account the deforming effect of these forces and moments on the joint structures. This effect is indeed present, but the discussion is not within the scope of this text.

STATICS OF THE TIBIOFEMORAL JOINT

Static analysis may be used to determine the forces and moments acting on a joint when no motion takes place or at one instant in time during a dynamic activity such as walking, running, or lifting an object. It can be performed for any joint in any position and under any loading configuration. In such analyses, either graphic or mathematical methods may be used to solve for the unknown forces or moments.

A complete static analysis involving all moments and all forces imposed on a joint in three dimensions is complicated. For this reason, a simplified technique is often used. The technique utilizes a free-body diagram and limits the analysis to one plane, to the three main coplanar forces acting on the free-body, and to the main moments acting about the joint under consideration. The minimum magnitudes of the forces and moments are obtained.

When the simplified free-body technique is used to analyze coplanar forces, one portion of the body is considered as distinct from the entire body, and all forces acting on this free-body are identified. A diagram is drawn of the free-body in the loading situation to be analyzed. The three principal coplanar forces acting on the free-body are identified and designated on the free-body diagram.

These forces are designated as vectors if four characteristics are known: magnitude, sense, line of application, and point of application. (The term "direction" includes line of application and sense.) If the points of application for all three forces and the directions for two forces are known, all remaining characteristics can be obtained for a force equilibrium situation. When the free-body is in equilibrium, the three principal coplanar forces are concurrent; that is, they intersect at a common point. In other words, these forces form a closed system with no resultant (i.e., their vector sum is zero). For this reason, the line of application for one force can be determined if the lines of application for the other two forces are known. Once the lines of application for all three forces are known, a triangle of forces can be constructed and the magnitudes of all three forces can be scaled from this triangle.

An example will illustrate the application of the simplified free-body technique for coplanar forces to the knee. In this case, the technique is used to estimate the minimum magnitude of the joint reaction force acting on the tibiofemoral joint of the weight-bearing leg when the other leg is lifted during stair climbing. The lower leg is considered as a free-body, distinct from the rest of the body, and a diagram of this free-body in the stair-climbing situation is drawn (Calculation Box 7-1). From all forces acting on the free-body, the three main coplanar forces are identified as the ground reaction force (equal to body weight*), the tensile force through the patellar tendon exerted by the quadriceps muscle, and the joint reaction force on the tibial plateau. The ground reaction force (W) has a known magnitude (equal to body weight), sense, line of application, and point of application (point of contact between the foot and the ground). The patellar tendon force (P) has a known sense (away from the knee joint), line of application (along the patellar tendon), and point of application (point of insertion of the patellar tendon on the tibial tuberosity), but an unknown magnitude. The joint reaction force (J) has a known point of application on the surface of the tibia (the contact point of the joint surfaces between the tibial and femoral condyles, estimated from a roentgenogram of the joint in the proper loading configuration), but an unknown magnitude, sense, and line of application. Using vectors calculations and triangles laws the joint reaction force (J) and the patellar tendon force (P) can be calculated (Calculation Box 7-1).

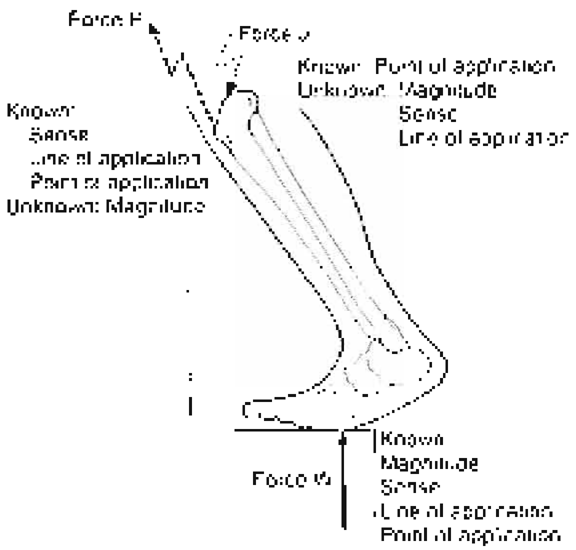
It can be seen that the main muscle force has a much greater influence on the magnitude of the joint reaction force than does the ground reaction force produced by body weight. Note that, in this example (Calculation Box 7-1), only the minimum magnitude of the joint reaction force has been calculated. If other muscle forces are considered, such as the force produced by the contraction of the hamstring muscles in stabilizing the knee, the joint reaction force increases.

The next step in the static analysis is analysis of the moments acting around the center of motion of the tibiofemoral joint with the knee in the same position and the loading configuration shown in Calculation Box Figure 7-1. The moment analysis is used to estimate the minimum magnitude of the moment produced through the patellar tendon, which counterbalances the moment on the lower leg produced by the weight of the body as the subject ascends stairs (Calculation Box 7-2).

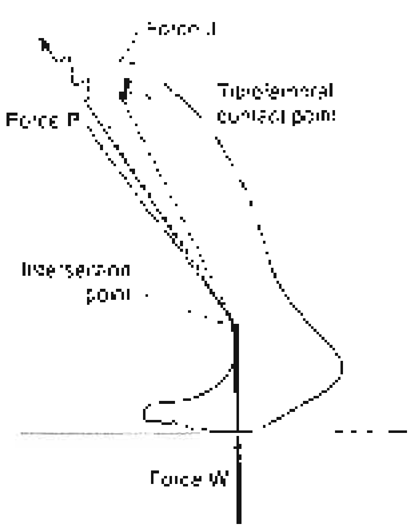
*In this case, the ground reaction force is actually equal to body weight minus the weight of the lower leg. Because the weight of the lower leg is minimal (less than one-tenth the body weight), it can be disregarded and the figure for total body weight can be used in the calculation.

CALCULATION BOX 7-1

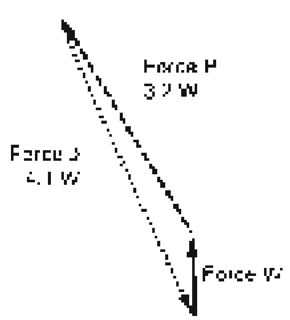
Free Body Diagram of the Knee Joint



The three main coplanar forces acting on the lower leg (ground reaction force (W), patellar tendon force (P), and joint reaction force (J)) are designated on a free body diagram of the lower leg while climbing stairs (Calculation Box Fig. 7-1-1).



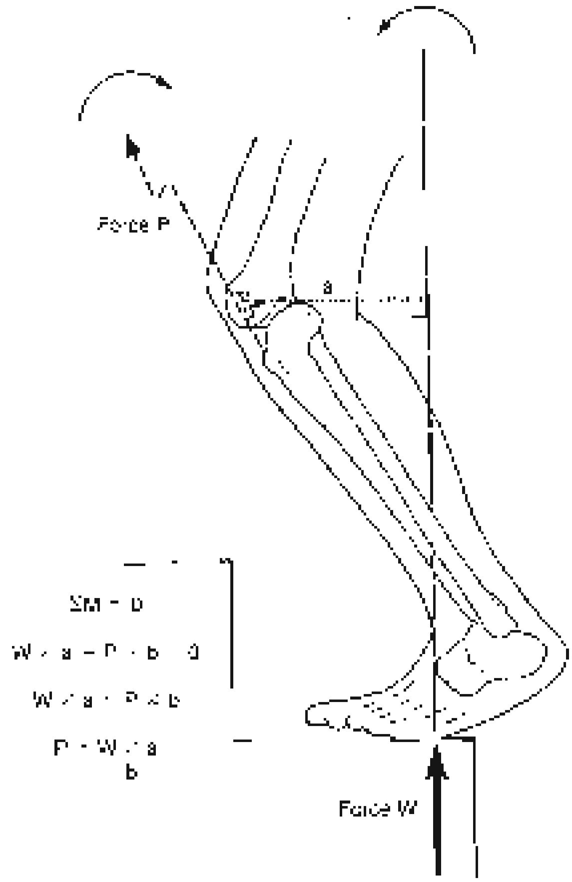
Because the lower leg is in equilibrium, the lines of application for all three forces intersect at one point. Because the lines of application for two forces (W and P) are known, the line of application for the third force (J) can be determined. The lines of application for forces W and P are extended until they intersect. The line of application for J can then be drawn from its point of application on the tibial surface through the intersection point (Calculation Box Fig. 7-1-2).



Now that the line of application for J has been determined, it is possible to construct a triangle of forces (Calculation Box Fig. 7-1-3). First, a vector representing W is drawn. Next, P is drawn from the head of vector W. Then, to close the triangle, force J is drawn from the head of vector W. The point at which forces P and J intersect defines the length of these vectors. Now that the length of all three vectors is known, the magnitude of forces P and J can be scaled from force W, which is equal to body weight, this determining the number of times the length of force W can be aligned along the force P and J, respectively. In this case, force P is 3.2 times body weight, and force J is 4.1 times body weight.

CALCULATION BOX 7-2

Free-Body Diagram of the Lower Leg During Stair Climbing



Calculation Box Figure 7-2-1.

The two main moments acting around the center of motion of the tibiofemoral joint (solid dot) are designated on the free-body diagram of the lower leg during stair climbing (Calculation Box Fig. 7-2-1).

The flexing moment on the lower leg is the product of the weight of the body (W; the ground reaction force) and its lever arm (a), which is the perpendicular distance of the force W to the center of motion of the tibiofemoral joint. The counterbalancing extending moment is the product of the quadriceps muscle force through the patellar tendon (P) and its lever arm (b). Because the lower leg is in equilibrium, the sum of these two moments must equal zero ($\Sigma M = 0$).

In this example, the counterclockwise moment is arbitrarily designated as positive ($W \times a - P \times b = 0$). Values for lever arms a and b can be measured from anatomic specimens or on soft tissue imaging or fluoroscopy (Wells & Bahrypoulos, 1999; Wretenberg et al., 1996), and the magnitude of W can be determined from the body weight of the individual. The magnitude of P can then be found from the moment equilibrium equation:

$$P = \frac{W \times a}{b}$$

*Again the weight of the lower leg is disregarded because it is less than one tenth of body weight.

DYNAMICS OF THE TIBIOFEMORAL JOINT

Although estimations of the magnitude of the forces and moments imposed on a joint in static situations are useful, most of our activities are of a dynamic nature. Analysis of the forces and moments acting on a joint during motion requires the use of a different technique for solving dynamic problems.

As in static analysis, the main forces considered in dynamic analysis are those produced by body weight, muscles, other soft tissues, and externally applied loads. Friction forces are negligible in a normal joint and thus not considered here. In dynamic analysis, two factors in addition to those in static analysis must be taken into account: the acceleration of the body part under consideration and the mass moment of inertia of the body part. (The mass moment of inertia is the unit used to express the amount of torque needed to accelerate a body and depends on the shape of the body.) (For more in-depth studies of dynamics, see Özkaya & Nordin, 1999.)

The steps for calculating the minimum magnitudes of the forces acting on a joint at a particular instant in time during a dynamic activity are as follows.

1. The anatomical structures are identified: definitions of structures, anatomical landmarks, point of contact of articular surface, and lever

parts involved in the production of forces for the biomechanical analyses.
2. The angular acceleration of the moving body part is determined.
3. The mass moment of inertia of the moving body part is determined.
4. The torque (moment) acting about the joint is calculated.
5. The magnitude of the main muscle force accelerating the body part is calculated.
6. The magnitude of the joint reaction force at a particular instant in time is calculated by static analysis.

In the first step, the structures of the body involved in producing forces on the joint are identified. These are the moving body part and the main muscles in that body part that are involved in the production of the motion. Great care must be taken in applying this first step. For example, the lever arms for all major knee muscles change according to the degree of knee flexion and gender (Wretenberg et al., 1996).

In joints of the extremities, acceleration of the body part involves a change in joint angle. To determine this angular acceleration of the moving body part, the entire movement of the body part is recorded photographically. Recording can be done with a stroboscopic light and movie camera, with video photogrammetry, with Selspot systems, with stereophotogrammetry, or with other methods (Gardner et al., 1994; Ramsey & Wretenberg, 1999; Winter, 1990). The maximal angular acceleration for a particular motion is calculated.

Next, the mass moment of inertia for the moving body part is determined. Anthropometric data on the body part can be used for this determination. As calculating these data is a complicated procedure, tables are commonly used (Drillis et al., 1964).

The torque about the joint can now be calculated using Newton's second law of motion, which states that when motion is angular, the torque is a product of the mass moment of inertia of the body part and the angular acceleration of that part:

$$T = I\alpha$$

where

T is the torque expressed in newton meters (Nm)
I is the mass moment of inertia expressed in newton meters × seconds squared (Nm sec²)
α is the angular acceleration expressed in radians per second squared (r/sec²)

The torque is not only a product of the mass moment of inertia and the angular acceleration of the body part but also a product of the main muscle force accelerating the body part and the perpendicular distance of the force from the center of motion of the joint (lever arm). Thus,

$$T = Fd$$

where

F is the force expressed in newtons (N)
d is the perpendicular distance expressed in meters (m)

Because T is known and d can be measured on the body part from the line of application of the force to the center of motion of the joint, the equation can be solved for F. When F has been calculated, the remaining problem can be solved like a static problem using the simplified free-body technique to determine the minimum magnitude of the joint reaction force acting on the joint at a certain instant in time.

A classic example will illustrate the use of dynamic analysis in calculating the joint reaction force on the tibiofemoral joint at a particular instant during a dynamic activity (e.g., kicking a football) (Frankel & Burstein, 1970). A stroboscopic film of the knee and lower leg was taken, and the angular acceleration was found to be maximal at the instant the foot struck the ball; the lower leg was almost vertical at this instant. From the film, the maximal angular acceleration was computed to be 453 r/sec². From anthropometric data tables (Drillis et al., 1964), the mass moment of inertia for the lower leg was determined to be 0.35 Nm sec². The torque about the tibiofemoral joint was calculated according to the equation: torque equals mass moment of inertia times angular acceleration ($T = I\alpha$):

$$0.35 \text{ Nm sec}^2 \times 453 \text{ r/sec}^2 = 158.5 \text{ Nm}$$

After the torque had been determined to be 158.5 Nm and the perpendicular distance from the subject's patellar tendon to the instant center for the tibiofemoral joint had been found to be 0.05 m, the muscle force acting on the joint through the patellar tendon was calculated using the equation: torque equals force times distance ($T = Fd$):

$$158.5 \text{ Nm} = F \times 0.05 \text{ m}$$

$$F = \frac{158.5 \text{ Nm}}{0.05 \text{ m}}$$

$$F = 3170 \text{ N}$$

Thus, 3,170 N was the maximal force exerted by the quadriceps muscle during the kicking motion.

Static analysis can now be performed to determine the minimum magnitude of the joint reaction force on the tibiofemoral joint. The main forces on this joint are identified as the patellar tendon force (P), the gravitational force of the lower leg (T), and the joint reaction force (J). P and T are known vectors. J has an unknown magnitude, sense, and line of application. The free-body technique for three coplanar forces is used to solve for J, which is found to be only slightly lower than P.

As is evident from the calculations, the two main factors that influence the magnitude of the forces on a joint in dynamic situations are the acceleration of the body part and its mass moment of inertia. An increase in angular acceleration of the body part will produce a proportional increase in the torque about the joint. Although in the body the mass moment of inertia is anatomically set, it can be manipulated externally. For example, it is increased when a weight boot is applied to the foot during rehabilitative exercises of the extensor muscles of the knee. Normally, a joint reaction force of approximately 50% of body weight results when the knee is slowly (with no acceleration forces) extended from 90° of flexion to full extension. In a person weighing 70 kg, this force is approximately 350 N. If a 10-kg weight boot is placed on the foot, it will exert a gravitational force of 100 N. This will increase the joint reaction force by 1,000 N, making this force almost four times greater than it would be without the boot.

Dynamic analysis has been used to investigate the peak magnitudes of the joint reaction forces, muscle forces, and ligament forces on the tibiofemoral joint during walking. Morrison (1970) calculated the magnitude of the joint reaction force transmitted through the tibial plateau in male and female subjects during level walking. He simultaneously recorded muscle activity electromyographically to determine which muscles produced the peak magnitudes of this force on the tibial plateau during various stages of the gait cycle (Fig. 7-15).

Just after heel strike, the joint reaction force ranged from two to three times body weight and was associated with contraction of the hamstring muscles, which have a decelerating and stabilizing effect on the knee. During knee flexion in the beginning of the stance phase, the joint reaction force was approximately two times body weight and was associated with the contraction of the quadriceps muscle, which acts to prevent buckling of the knee. The peak joint reaction force occurred during the late stance phase just before toe-off. This force ranged from two to four times body weight, varying among the subjects tested, and was associated with contraction of the gastrocnemius muscle. In the late swing phase, contraction of the hamstring muscles resulted in a joint reaction force approximately equal to body weight. No significant difference was found between the joint reaction force magnitudes for men and women when the values were normalized by dividing them by body weight.

Andriacchi & Strickland (1985) studied the normal moment patterns around the knee joint during level walking for 29 healthy volunteers (15 women and 14 men with an average age of 39 years). Figure 7-16 depicts the flexion-extension, abduction-adduction, and internal-external moments during the stance and swing phase of level walking. The moments are normalized to the individual's body weight and height and are presented as a percentage. The flexion-

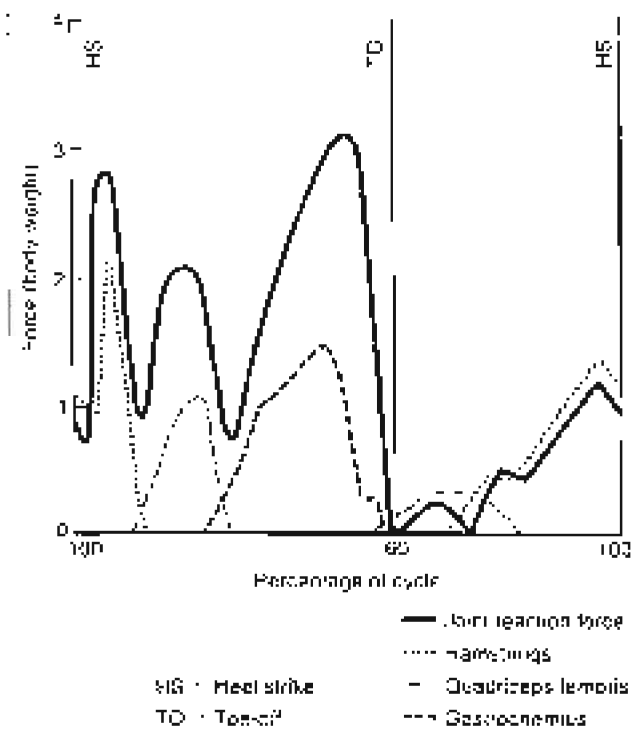


FIG. 7-15

Joint reaction forces expressed as body weight transmitted through the tibial plateau during walking, one gait cycle (12 subjects). The muscle forces producing the peak magnitudes of this force are also designated. Adapted from Morrison, J.B. (1970). The mechanics of the knee joint in relation to normal walking. J Biomech, 3, 51.

extension moments during the stance phase are approximately 20 to 30 times larger than the moment produced in the frontal (abduction-adduction) and transverse (internal-external) planes.

An increase in knee joint flexion-extension moment amplitude has been reported at increased walking speeds (Andriacchi & Strickland, 1985; Holden et al., 1997). An increase in the production of adduction knee joint moment during stair climbing compared with level walking was reported by Yu et al. (1997).

During the gait cycle, the joint reaction force shifts from the medial to the lateral tibial plateau. In the stance phase, when the force reaches its peak value, it is sustained mainly by the medial plateau (adduction moment); in the swing phase, when the force is minimal, it is sustained primarily by the lateral plateau. The contact area of the medial tibial plateau is approximately 50% larger than that of the lateral tibial plateau (Kettelkamp & Jacobs, 1972). Also, the cartilage on this plateau is approximately three times thicker than that on the lateral plateau. The larger surface area and the greater thickness of the medial plateau allow it to more easily sustain the higher forces imposed on it.

In a normal knee, joint reaction forces are sustained by the menisci as well as by the articular cartilage. The function of the menisci was investigated by Seedhom and coworkers (1974), who examined the distribution of stresses in knees of human autopsy subjects with and without menisci. Their results suggest that in load-bearing situations, the magnitude of the stresses on the tibiofemoral joint when the menisci have been removed may be as much as three times greater than when these structures are intact. Fukuda et al. (2000) studied in vitro the load-compressive transmission of the knee joint and the role of menisci and articular cartilage. The load simulated was static and dynamic impact loading. The testing was done in neutral, varus, and valgus alignment of the knee joints in 40 fresh-frozen pig knee specimens. The compressive stress on the medial subchondral bone was up to five times higher with the menisci removed. This study points to the importance of the menisci as a structure to absorb load and protect the cartilage and subchondral bone under dynamic conditions.

In a normal human knee, stresses are distributed over a wide area of the tibial plateau. If the menisci are removed, the stresses are no longer distributed over such a wide area but instead are limited to a contact area in the center of the plateau (Fig. 7-17). Thus, removal of the menisci not only increases the

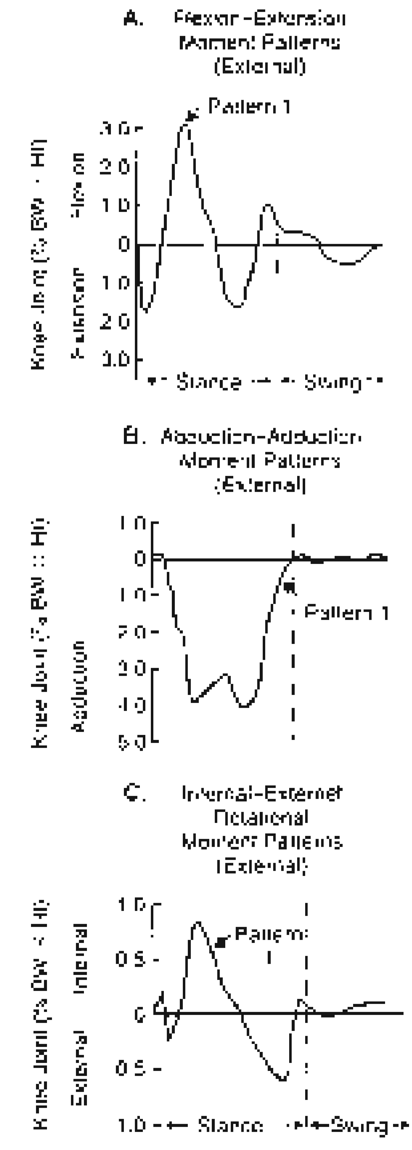


FIG. 7-16

Flexion-extension (A), abduction-adduction (B), and internal-external rotation (C) moments produced during one gait cycle in normal subjects. The moments are normalized to each individual body weight × height and expressed as a percentage. *Reprinted with permission from Andriacchi, T.P. & Strickland, A.B. (1985). Gait analysis as a tool to assess joint kinetics. In N. Berme, A.E. Engin, et al. (Eds.). Biomechanics of Normal and Pathological Human Articulating Joints. NATO ASI Series, Vol 93, pp. 83–102). Dordrecht, Netherlands: Martinus Nijhoff.*

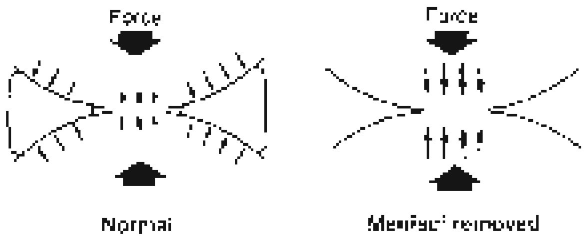


FIG. 7-17

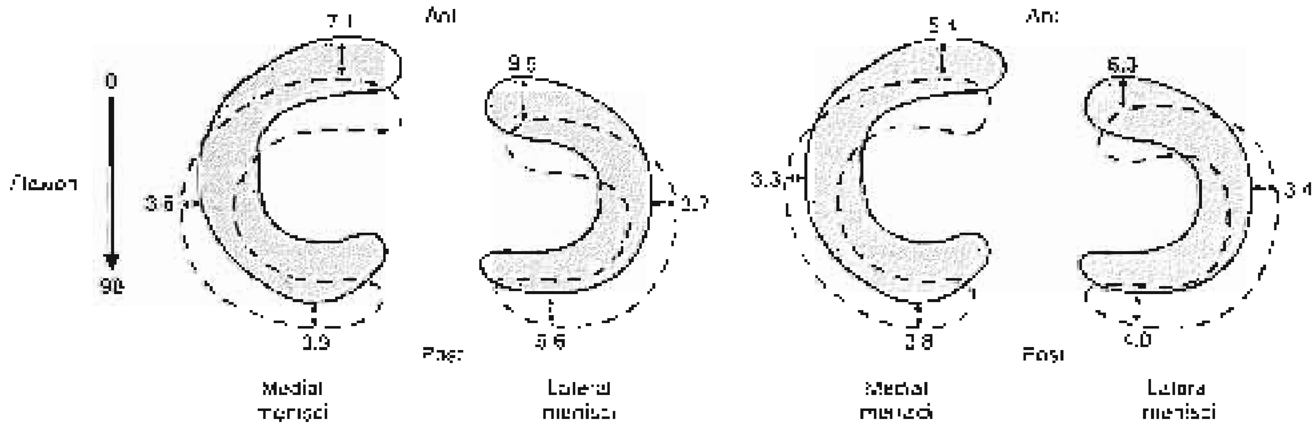
Stress distribution in a normal knee and in a knee with the menisci removed. Removal of the menisci increases the magnitude of stresses on the cartilage of the tibial plateau and changes the size and location of the tibiofemoral contact area. With the menisci intact, the contact area encompasses nearly the entire surface of the tibial plateau. With the menisci removed, the contact area is limited to the center of the tibial plateau.

magnitude of the stresses on the cartilage and subchondral bone at the center of the tibial plateau but also diminishes the size and changes the location of the contact area. Over the long term, the high stresses placed on this smaller contact area may be harmful to the exposed cartilage, which is usually soft and fibrillated in that area. The menisci are thought to carry up to 70% of the load across the knee. Movement during knee flexion of the menisci would therefore protect the articulating surfaces while avoiding injury to it.

Vedi et al. (1999) studied meniscal movement in 16 young football players with normal knees with MRI. The knee flexion movement was scanned from full knee extension to 90° of knee flexion. The imaging technique allowed for both standing (weight-bearing) and sitting (non-weight-bearing) and was performed simultaneously in the sagittal and transverse plane. Figure 7-18 shows the movements in the transverse plane of the medial and lateral menisci expressed in millimeters (mean) from full extension to 90° flexion of knee joint motion. Movement was significantly greater in weight-bearing than in non-weight-bearing for both lateral and medial menisci. The contributions of the menisci are therefore not only to protect the articular cartilage and subchondral bone but also to contribute actively to knee joint stability.

STABILITY OF THE KNEE JOINT

The key to a healthy knee joint is joint stability. The osseous configuration, the menisci, the ligaments, the capsule, and the muscles surrounding the knee joint produce joint stability (Fig. 7-7, A & B). If any



A

B

FIG. 7-18

Simplified diagrams showing the mean movements of the medial and lateral menisci from full knee extension to 90° knee flexion during two conditions. A. Erect and weight-bearing. B. Sitting, relaxed, and bearing no weight. *Adapted with permission from Vedi, V., Williams, A., Tennant, S.J., et al. (1999). Meniscal movement: An in-vivo study using dynamic MRI. J Bone Joint Surg, 81B(1), 37–41.*

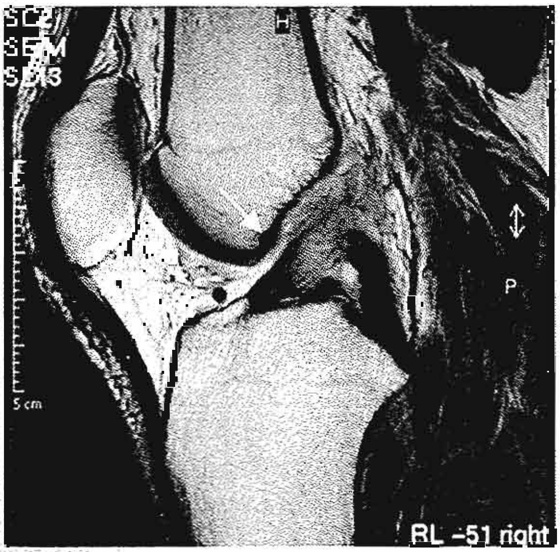
CASE STUDY 7-1

ACL Injury

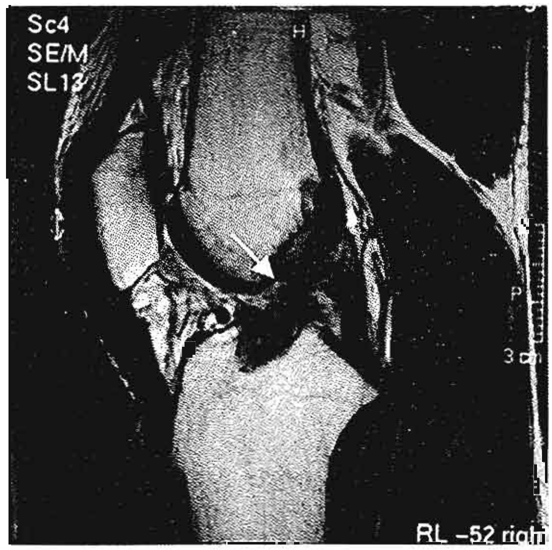
A 30-year-old male suffered an external rotation trauma in his right knee while downhill skiing. Following the trauma, he experienced sharp pain, progressive joint effusion, and subjective instability. During careful examination by a specialist, an anterior positive drawer test was diagnosed, and the Lachman test and pivot shift test were found positive. An MRI confirmed the ACL rupture (Case Study Fig. 7-1-1).

The rupture of the primary stabilizer of the knee joint (ACL) leads to a progressive structural alteration of the knee. A primary objective of the treatment is the prevention of reinjury of the knee in the hope of preventing additional ligamentous injuries, meniscal injuries, and possible cartilage degeneration. In this case, the patient first completed a course of conservative treatment with physical therapy. After 6 months, the subjective instability was present during sports and daily activities such as gait and stair climbing. To compensate for the ACL deficiency, the patient altered his gait patterns, presenting quadriceps avoidance gait to prevent the anterior translation of the tibia when the quadriceps contracts at the midstance phase of the gait (Andriacchi & Birac, 1993; Berchuck et al., 1990).

The patient went for surgical treatment. The MRI below (Case Study Fig. 7-1-2) shows the ACL status after patella bone-tendon-bone autograft was performed 10 months post trauma.



Case Study Figure 7-1-1



Case Study Figure 7-1-2

of these structures are malfunctioning or disturbed, knee joint instability will occur. The ligaments are the primary stabilizer for anterior and posterior translation, varus and valgus angulation, and internal and external rotation of the knee joint (Case Study 7-1).

Fu et al. (1994) summarized the functions of the knee ligaments. The ACL is the predominant restraint to anterior tibial displacement. The ligament accepts 75% of the anterior force at full extension and an additional 10% (up to 90°) of knee flexion. The posterior cruciate ligament is the primary restraint to posterior tibial translation; it sustains 85 to 100% of the posterior force at both 30 and 90° of knee flexion. The lateral collateral ligament is the primary restraint to varus angulation and it resists approximately 55% of the applied load at full extension. The role of the lateral collateral ligament increases with joint flexion as the posterior structures become lax. The medial collateral ligament (superficial portion) is the primary restraint to valgus (abduction) angulation and resists 50% of the external valgus load. The capsule, the anterior and posterior cruciate ligaments, share the remaining valgus load. Internal rotational laxity seen in the 20 to 40° range of knee flexion is restrained by the medial collateral ligament and the ACL. Finally, external rotation laxity seen in the 30 to 40° range of knee flexion is restrained by the posterior cruciate ligament at 90° of knee flexion.

In vivo measurement of the normal ACL has been performed by Beynnon et al. (1992). They placed a strain transducer arthroscopically in the ACL. The results showed that strain in the ACL was related to knee flexion (with the most strain occurring near full extension) and increased with quadriceps contraction. Less strain occurred with co-contractions of both the quadriceps and the hamstring muscle groups and at greater degrees of knee flexion. This indicates that muscle contraction and co-contraction contribute to the stability of the knee joint by increasing the stiffness of the joint. Kwak et al. (2000) studied in vitro the effect of hamstrings and iliotibial band forces on the kinematics of the knee. At various knee flexion angles, human knee specimens were tested with different muscle-loading patterns. The quadriceps muscle force was always present, and the test was performed with and without hamstring muscle force and with and without iliotibial band force. With loading of simultaneous quadriceps and hamstring muscle force, the tibia translated posteriorly and rotated externally. The effect was similar for the iliotibial band simulated forces but the effect was smaller.

Many in vitro studies suggest that the hamstrings are important anterior and rotational stabilizers of the tibia. In vivo studies have shown that co-contractions of the quadriceps and hamstring muscles are highly present in normal knee joints and daily activities (Baratta et al., 1988; Solomonow & D'Ambrosia, 1994). The co-contraction mechanisms also increase the knee joint stability in vivo (Aagaard et al., 2000; Markholf et al., 1978; Solomonow & D'Ambrosia, 1994). However, the complex mechanism in vivo of muscle activity as a knee stabilizer, the extent of protection, and the biomechanical and clinical importance needs further research (Grabiner & Weiker, 1993).

FUNCTION OF THE PATELLA

The patella serves two important biomechanical functions in the knee. First, it aids knee extension by producing anterior displacement of the quadriceps tendon throughout the entire range of motion, thereby lengthening the lever arm of the quadriceps muscle force. Second, it allows a wider distribution of compressive stress on the femur by increasing the area of contact between the patellar tendon and the femur. The contribution of the patella to the length of the quadriceps muscle force lever arm varies from full flexion to full extension of the knee (Lindahl & Movin, 1967; Smidt, 1973). At full flexion, when the patella is in the intercondylar groove, it produces little anterior displacement of the quadriceps tendon and it contributes the least to the length of the quadriceps muscle force lever arm (approximately 10% of the total length). As the knee is extended, the patella rises from the intercondylar groove, producing significant anterior displacement of the tendon. The length of the quadriceps force lever arm rapidly increases with extension up to 45°, at which point the patella lengthens the lever arm by approximately 30%.

With knee extension beyond 45°, the length of the lever arm is diminished slightly. With this decrease in its lever arm, the quadriceps muscle force must increase for the torque about the knee to remain the same. In an in vitro study of normal knees, Lieb and Perry (1968) showed that the quadriceps muscle force required to extend the knee the last 15° increased by approximately 60% (Fig. 7-19).

If the patella is removed from a knee, the patellar tendon lies closer to the center of motion of the tibiofemoral joint (Fig. 7-20). Acting with a shorter lever arm, the quadriceps muscle must produce even more force than is normally required for a certain torque about the knee to be maintained during

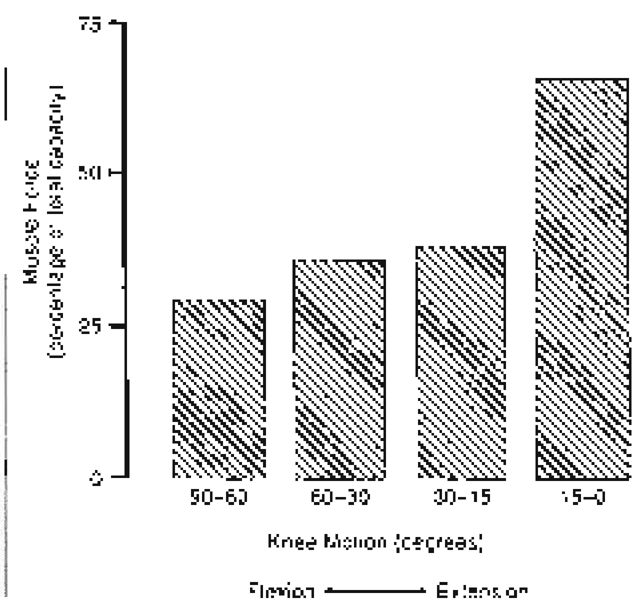


FIG. 7-19

Quadriceps muscle force required during knee extension from 90° of flexion to full extension. Adapted from Lieb, F.J., & Perry, J. (1968). Quadriceps function: An anatomical and mechanical study using amputated limbs. J Bone Joint Surg, 50A, 1535.

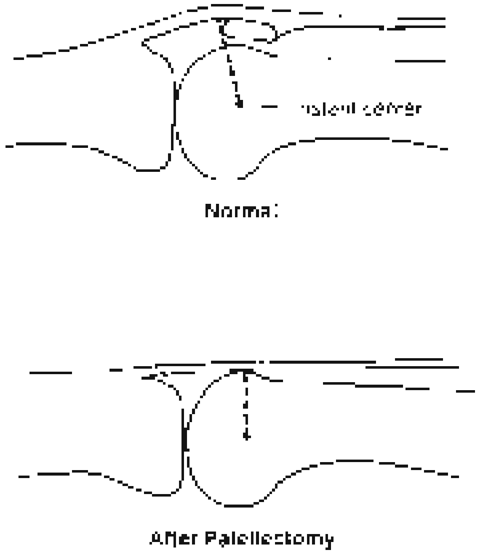


FIG. 7-20

Quadriceps muscle lever arm (represented by the broken line) in a normal knee and in a knee in which the patella has been removed. The lever arm is the perpendicular distance between the force exerted by the quadriceps muscle through the patellar tendon and the instant center of the tibiofemoral joint for the last two degrees of extension. The patellar tendon lies closer to the instant center in the knee without a patella. Adapted from Kaufer, H. (1971). Mechanical function of the patella. J Bone Joint Surg, 53A, 1551.

the last 45° of extension. Full active extension of such a knee may require as much as 30% more quadriceps force than is normally required (Kaufer 1971). This increase in force may be beyond the capacity of the quadriceps muscle in some patients, particularly those who have intra-articular disease or are advanced in age.

STATICS AND DYNAMICS OF THE PATELLOFEMORAL JOINT

During dynamic activities, the magnitude of the muscle forces acting on a joint directly affects the magnitude of the joint reaction force. In general, the greater the muscle forces, the greater the joint reaction force.

In the patellofemoral joint, the quadriceps muscle force increases with knee flexion. During relaxed upright standing, minimal quadriceps muscle forces are required to counterbalance the small flexion moments about the patellofemoral joint because the center of gravity of the body above the knee is almost directly above the center of rotation of the patellofemoral joint. As knee flexion increases, the center of gravity shifts further away from the center of rotation, thereby greatly increasing the flexion moments to be counterbalanced by the quadriceps muscle force. As the quadriceps muscle force rises, so does the patellofemoral joint reaction force (Hungerford & Barry, 1979; Reilly & Martens, 1972).

Knee flexion also influences the patellofemoral joint reaction force by affecting the angle between the patellar tendon force and the quadriceps tendon force. The angle of these two force components becomes more acute with knee flexion, increasing the magnitude of the patellofemoral joint reaction force (Calculation Box 7-3). Reilly and Martens (1972) determined the magnitude of the patellofemoral joint reaction force during several dynamic activities involving varying amounts of knee flexion. During level walking, which requires relatively little knee flexion, the reaction force was low. The peak value, in the middle of the stance phase when flexion was greatest, was one half body weight.

The joint reaction force was much greater during activities that require greater flexion. During knee bends to 90°, this force reached 2.5 to 3 times body weight with the knee flexed 90° (Fig. 7-21). Through-

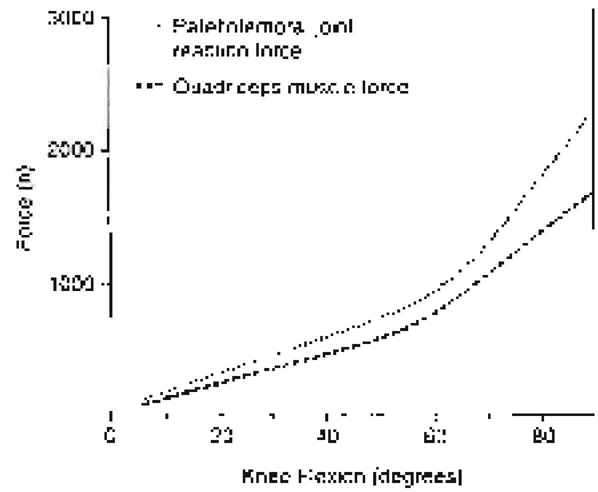


FIG. 7-21

Patellofemoral joint reaction force and quadriceps muscle force during knee bend to 90° (three subjects). Adapted from Reilly, D.T. & Martens, M. (1972). Experimental analysis of the quadriceps muscle force and patellofemoral joint reaction force for various activities. Acta Orthop Scand, 43, 126.

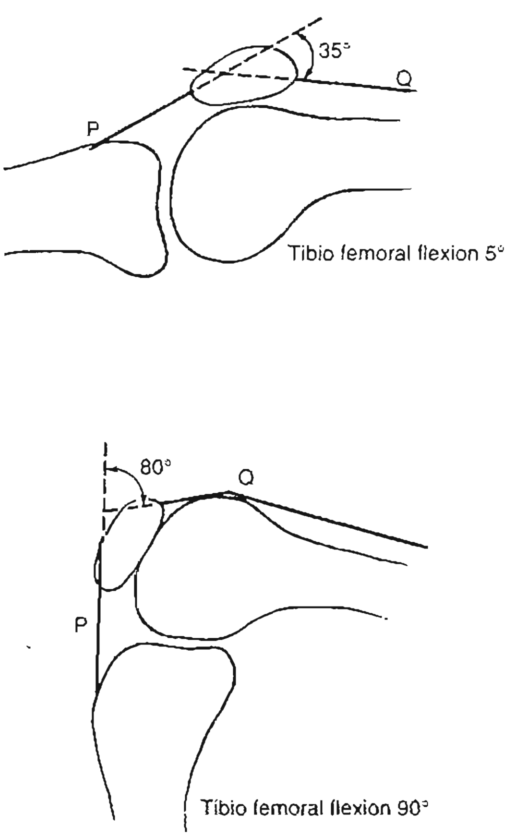
CALCULATION BOX 7-3

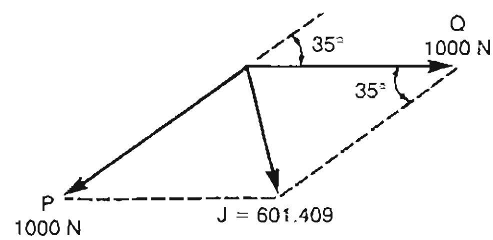
Joint Reaction Forces at the Knee in Flexion

Knee flexion influences the patellofemoral joint reaction force by changing the angle between the patellar tendon and the quadriceps tendon (Calculation Box Fig. 7-3-1, *A & B*).

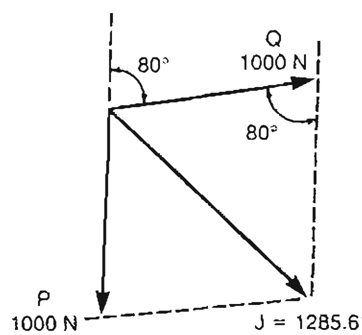
The angle between the patellar tendon (P) and the quadriceps tendon (Q) is 35° with the knee flexed 5° (left top) and 80° with the knee flexed 90° (left bottom). Values for the tendon angles are from Matthews and Associates (1977), who determined the angle roentgenographically after placing two metal wires along each of these tendons.

The patellofemoral joint reaction force with the knee in 5 and 90° of flexion is obtained by constructing a parallelogram of forces for each situation and using trigonometric calculations. The patellofemoral joint reaction force (J) is the resultant of the two equal force components through the patellar tendon (P) and the quadriceps tendon (Q). As the angle between these force components becomes more acute with greater knee flexion, the resultant joint reaction force (J) becomes larger. *Adapted from Wiktorin, C.v.H. & Nordin, M. (1986).* Introduction to Problem Solving in Biomechanics *(pp. 87–129). Philadelphia: Lea & Febiger.*





$$J^2 = Q^2 + P^2 - 2PQ \cdot \cos 35°$$
$$J = \sqrt{361695}$$
$$J = 601.41$$



$$J^2 = Q^2 + P^2 - 2PQ \cdot \cos 80°$$
$$J = \sqrt{1652703}$$
$$J = 1285.57$$

Calculation Box Figure 7-3-1A.

Calculation Box Figure 7-3-1B.

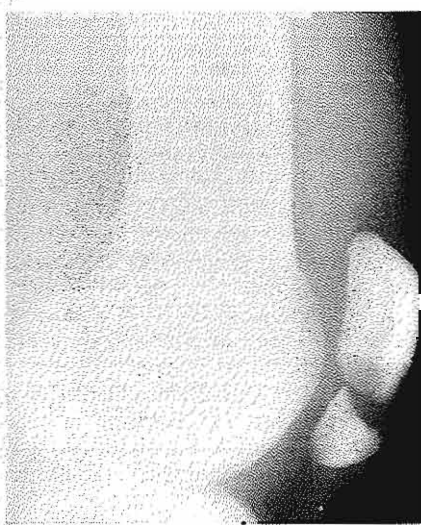
CASE STUDY 7-2

Extensor Mechanism Injury

A 30-year-old basketball player had a forceful knee flexion while coming down from a jump. A strong eccentric contraction of the quadriceps produced abnormally high tensile loads in the patella, leading to a fracture at the inferior pole. In this case, the patella fracture occurred because the muscle forces of the quadriceps exceeded the osseous strength of the patella. The weakest link was the patella.

The picture shows a fracture at the patella accompanied by a significant fracture separation that resulted from the quadriceps traction force.

Because of the fracture, the extensor mechanism is unable to function and extend the knee. It will directly affect the stability of the patellofemoral joint and the distribution of the compressive stresses on the femur. At the same time, the impaired function of the quadriceps decreases the dynamic stability at the knee joint (patellofemoral and tibiofemoral joints) that is necessary for daily activities such as gait and stair climbing.



Case Study Figure 7-2-1

out knee bend, the patellofemoral joint reaction force remained higher than the quadriceps muscle force. During stair climbing and descent, at the point when knee flexion reached a maximum of approximately 60°, the peak value equaled 3.3 times body weight.

When the knee is extended, the lower part of the patella rests against the femur. As the knee is flexed to 90°, the contact surface between the patella and femur shifts cranially in size and under weight-bearing conditions (Komistek et al., 2000). The contact surface area increases in size somewhat (Goodfellow et al., 1976). To some extent, this increase in the contact surface with knee flexion compensates for the larger patellofemoral joint reaction force. If a tight iliotibial band is present, the patellofemoral joint force may shift laterally causing abnormal patellar kinematics and load-bearing (Kwak et al., 2000).

The quadriceps muscle force and the torque around the patellofemoral joint can be extremely high under certain circumstances, particularly when the knee is flexed—for instance, when a basketball player suffers a patella fracture as a result of indirect forces played by an eccentric contraction of the quadriceps (Case Study 7-2). Another extreme situation was observed during a study of the external torque on the knee produced by weight lifting; one subject ruptured his patellar tendon when he lifted a barbell weighing 175 kg (Zernicke et al., 1977). At the instant of tendon rupture, the knee was flexed 90°, the torque on the knee joint was 550 Nm and the quadriceps muscle force was approximately 10,330 N.

Because of the high magnitude of quadriceps muscle force and joint reaction force during activities requiring a large amount of knee flexion, patients with patellofemoral joint derangements experience increased pain when performing these activities. An effective mechanism for reducing these forces is to limit the amount of knee flexion.

Summary

1 The knee is a two-joint structure that is composed of the tibiofemoral joint and the patellofemoral joint.

2 In the tibiofemoral joint, surface motion occurs in three planes simultaneously but is greatest by far in the sagittal plane. In the patellofemoral joint, surface motion occurs simultaneously in two planes, the frontal and the transverse, and is greater in the frontal plane.

3 The surface joint motion can be described with the use of an instant center technique. When performed on a normal knee, the technique reveals the following: the instant center for successive intervals of motion of the tibiofemoral joint in the sagittal plane follows a semicircular pathway, and the direction of displacement of the tibiofemoral contact points is tangential to the surface of the tibia, indicating gliding throughout the range of motion.

4 The screw-home mechanism of the tibiofemoral joint adds stability to the joint in full extension. Additional passive stability to the knee is given by the ligamentous structure and menisci and the dynamic stability by the muscles around the knee.

5 The tibiofemoral and patellofemoral joints are subjected to great forces. Muscle forces have the greatest influence on the magnitude of the joint reaction force, which can reach several times body weight in both joints. In the patellofemoral joint, knee flexion also affects the joint reaction force, with greater knee flexion resulting in a higher joint reaction force.

6 Although the tibial plateaus are the main load-bearing structures in the knee, the cartilage, menisci, and ligaments also bear loads. The menisci aid in distributing the stresses and reducing the load imposed on the tibial plateaus.

7 The patella aids knee extension by lengthening the lever arm of the quadriceps muscle force throughout the entire range of motion and allows a wider distribution of compressive stress on the femur.

REFERENCES

Aagaard, P., Simonsen, E.B., Andersen, J.L., et al. (2000). Antagonist muscle coactivation during isokinetic extension. *Scand J Med Sci Sports, 10*(2), 58–67.

Andriacchi, T.P., & Birac, D. (1993). Functional testing in the anterior cruciate ligament-deficient knee. *Clin Orthop Rel Res, 288* (Suppl), 40–47.

Andriacchi, T.P., Kramer, G.M., & Landon, G.C. (1979). Three-dimensional coordinate data processing in human motion analysis. *J Biomech Eng, 101*, 279–283.

Andriacchi, T.P., & Strickland, A.B. (1985). Gait analysis as a tool to assess joint kinetics. In N. Berme, A.E. Engin, D.A. Correia, et al. (Eds.), *Biomechanics of Normal and Pathological Human Articulating Joints*. NATO ASI series, No. 93, pp. 83–102. Dordrecht, Netherlands: Martinus Nijhoff.

Baratta, R., Solomonow, M., Zhou, B.H., Letson, D., Chuinard, R., D'Ambrosia, R. (1988). Muscular coactivation: The role of the antagonist musculature in maintaining knee stability. *Am J Sports Med, 16*, 113–122.

Berchuck, M., Andriacchi, T.P., Bach, B.R., Reider, B. (1990). Gait adaptations by patients who have a deficient anterior cruciate ligament. *J Bone Joint Surg, 72A*, 871–877.

Beynnon, B., Howe, J.G., & Pope, M.H. (1992). The measurement of anterior cruciate ligament strain in vivo. *Int Orthop, 16*, 1–12.

Draganich, L.F., Andriacchi, T.P., & Andersson, G.B.J. (1987). Interaction between intrinsic knee mechanics and the knee extensor mechanism. *J Orthop Res, 5*, 539–547.

Drillis, R., Contini, R., & Bluestein, M. (1964). Body segment parameters: A survey of measurement techniques. *Artif Limbs, 8*, 44.

Frankel, V.H., & Burstein, A.H. (1970). *Orthopaedic Biomechanics*. Philadelphia: Lea & Febiger.

Frankel, V.H., Burstein, A.H., & Brooks, D.B. (1971). Biomechanics of internal derangement of the knee. Pathomechanics as determined by analysis of the instant centers of motion. *J Bone Joint Surg, 53A*, 945.

Fu, F.H., Harner, C.D., Johnson, D.L., et al. (1994). Biomechanics of the knee ligaments: Basic concepts and clinical application. *Instr Course Lect, 43*, 137–148.

Fukuda, Y., Takai, S., Yoshino, N., et al. (2000). Impact load transmission of the knee joint—influence of leg alignment and the role of meniscus and articular cartilage. *Clinical Biomech, 15*, 516–521.

Gardner, J.R., Ateshian, G.A., Grelsamer, R.P., et al. (1994). A 6 DOF knee testing device to determine patellar tracking and patellofemoral joint contact area via stereophotogrammetry. *J Biomech Eng (ASME), 116*, 256–260.

Goodfellow, J., Hungerford, D.S., & Zindel, M. (1976). Patello-femoral joint mechanics and pathology. 1. Functional anatomy of the patello-femoral joint. *J Bone Joint Surg, 58B*, 287.

Grabiner, M.D., & Weiker, G.G. (1993). Anterior cruciate ligament injury and bracing considerations. *Clinical Biomech, 8*, 215–219.

Grood, E.S., & Suntay, W.J. (1983). A joint coordinate system for the clinical description of three-dimensional motions: Application to the knee. *J Biomech Eng, 105*, 136–144.

Hehne, H.J. (1990). Biomechanics of the patellofemoral joint and its clinical relevance. *Clin Orthop, 258*, 73–85.

Helfet, A.J. (1974). Anatomy and mechanics of movement of the knee joint. In A.J. Helfet (Ed.), *Disorders of the Knee* (pp. 1–17). Philadelphia: J.B. Lippincott.

Holden, J.P., Chou, G., & Stanhope, S.J. (1997). Changes in knee joint function over a wide range of walking speeds. *Clinical Biomech, 12*(6), 375–382.

Hungerford, D.S., & Barry, M. (1979). Biomechanics of the patellofemoral joint. *Clin Orthop, 144*, 9–15.

Kapandji, I.A. (1970). The knee. In I.A. Kapandji (Ed.), *The Physiology of the Joints* (Vol 2, pp. 72–135). Paris: Editions Maloine.

Kaufer, H. (1971). Mechanical function of the patella. *J Bone Joint Surg, 53A*, 1551.

Kellis, E., & Baltzopoulos, V. (1999). In vivo determination of the patella tendon and hamstrings moment arms in adult males using videofluoroscopy during submaximal knee extension and flexion. *Clinical Biomech, 14*, 118–124.

Kettelkamp, D.B., Johnson, R.J., Smidt, G.L., Chao, E.Y., Walker, M. (1970). An electrogoniometric study of knee motion in normal gait. *J Bone Joint Surg, 52A*, 775.

Kettelkamp, D.B., & Jacobs, A.W. (1972). Tibiofemoral contact area—determination and implications. *J Bone Joint Surg, 54A*, 349.

Komistek, R.D., Dennis, D.A., Mabe, J.A., et al. (2000). An in vivo determination of patellofemoral contact positions. *Clinical Biomech, 15*, 29–36.

Kroemer, K.H., Marras, W.S., McGlothlin, J.D., et al. (1990). On the measurement of human strength. *Int J Industrial Ergonomics, 6*, 199–210.

Kwak, S.D., Ahmad, C.S., Gardner, T.R., et al. (2000). Hamstrings and iliotibial band forces affect knee kinematics and contact pattern. *J Orthop Res, 18*, 101–108.

Lamoreux, L. (1971). Kinematic measurements in the study of human walking. Biomechanics Lab, University of California, San Francisco. *Bull Prosthetic Res, Sp 1971*, 10–15.

Laubenthal, K.N., Smidt, G.L., & Kettelkamp, D.B. (1972). A quantitative analysis of knee motion during activities of daily living. *Phys Ther, 52*, 34.

Levens, A.S., Inman, V.T., & Blosser, J.A. (1948). Transverse rotation of the segments of the lower extremity in locomotion. *J Bone Joint Surg, 30A*, 859.

Lieb, F.J., & Perry, J. (1968). Quadriceps function. An anatomical and mechanical study using amputated limbs. *J Bone Joint Surg, 50A*, 1535.

Lindahl, O., & Movin, A. (1967). The mechanics of extension of the knee joint. *Acta Orthop Scand, 38*, 226.

Markolf, K., Graff-Radford, A., & Amstutz, H. (1978). In vivo knee stability. *J Bone Joint Surg, 60A*, 664–674.

Matsumoto, H., Seedhom, B.B., Suda, Y., et al. (2000). Axis of tibial rotation and its change with flexion angle. *Clin Orthop, 371*, 178–182.

Matthews, L.S., Sonstegard, D.A., & Henke, J.A. (1977). Load bearing characteristics of the patellofemoral joint. *Acta Orthop Scand, 48*, 511.

Morrison, J.B. (1970). The mechanics of the knee joint in relation to normal walking. *J Biomech, 3*, 51.

Murray, M.P., Drought, A.B., & Kory, R.C. (1964). Walking patterns of normal men. *J Bone Joint Surg, 46A*, 335.

Özkaya, N., & Nordin, M. (1999). *Fundamentals of Biomechanics: Equilibrium, Motion, and Deformation* (2nd ed.). New York: Springer-Verlag.

Perry, J., Norwood, L., & House, K. (1977). Knee posture and biceps and semimembranosus muscle action in running and cutting (an EMG study). *Trans Orthop Res Soc, 2*, 258.

Ramsey, D.K., & Wretenberg, P.F. (1999). Biomechanics of the knee: Methodological considerations in the in vivo kinematic analysis of the tibiofemoral and patellofemoral joint. Review paper. *Clinical Biomech, 14*, 595–611.

Reilly, D.T., & Martens, M. (1972). Experimental analysis of the quadriceps muscle force and patellofemoral joint reaction force for various activities. *Acta Orthop Scand, 43*, 126.

Reuleaux, F. (1876). In *The Kinematics of Machinery: Outline of a Theory of Machines*. London: Macmillan.

Seedhom, B.B., Dowson, D., & Wright, V. (1974). The load-bearing function of the menisci: A preliminary study. In O.S. Ingwersen et al. (Eds.), *The Knee Joint: Recent Advances in Basic Research and Clinical Aspects* (pp. 37–42). Amsterdam: Excerpta Medica.

Selvik, G. (1978). Roentgen stereophotogrammetry in Lund, Sweden. In A.M. Coblentz & R.E. Herron (Eds.), *Applications of Human Biostereometrics. Proc. SPIE, 166* (pp. 184–189).

Selvik, G. (1983). Roentgen stereophotogrammetry in orthopedics. In R.E. Herron (Ed.), *Biostereometrics '82. Proc SPIE, 361* (pp. 178–185).

Smidt, G.L. (1973). Biomechanical analysis of knee flexion and extension. *J Biomech, 6*, 79.

Solomonow, M., & D'Ambrosia, R. (1994). Neural reflex arcs and muscle control of knee stability and motion. In W.N. Scott (Ed.), *The Knee* (pp. 107–120). New York: Mosby.

Vedi, V., Williams, A., Tennant, S.J., et al. (1999). Meniscal movement. An in-vivo study using dynamic MRI. *J Bone Joint Surg, 81B*(1), 37–41.

Winter, D.A. (1990). In *Biomechanics and Motor Control of Human Behavior* (2nd ed.). New York: John Wiley & Sons.

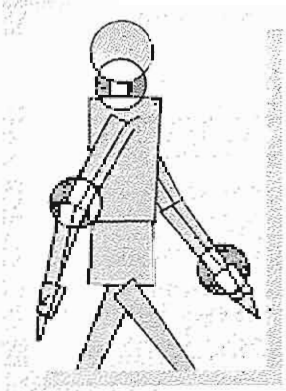
Wiktorin, C.v.H., & Nordin, M. (1986). *Introduction to Problem Solving in Biomechanics* (pp. 67–129). Philadelphia: Lea & Febiger.

Wilson, S.A., Vigorita, V.J., & Scott, W.N. (1994). Anatomy. In W.N. Scott (Ed.), *The Knee* (p. 17). Philadelphia: Mosby.

Wretenberg, P., Németh, G., Lamontagne, M., et al. (1996). Passive knee muscle moment arms measured in vivo with MRI. *Clinical Biomech, 11*(8), 439–446.

Yu, B., Stuart, M.J., Kienbacher, T., et al. (1997). Valgus-varus motion of the knee in normal level walking and stair climbing. *Clinical Biomech, 12*(5), 286–293.

Zernicke, R.F., Garhammer, J., & Jobe, F.W. (1977). Human patellar-tendon rupture. *J Bone Joint Surg, 59A*, 179–183.



CHAPTER 8

Biomechanics of the Hip

Margareta Nordin, Victor H. Frankel

Introduction

The hip joint is one of the largest and most stable joints in the body. In contrast to the knee, the hip joint has intrinsic stability provided by its relatively rigid ball-and-socket configuration. It also has a great deal of mobility, which allows normal locomotion in the performance of daily activities. Derangements of the hip can produce altered stress distributions in the joint cartilage and bone, leading to degenerative arthritis. Such damage is further potentiated by the large forces borne by the joint.

Anatomical Considerations

The hip joint is composed of the head of the femur and the acetabulum of the pelvis (Fig. 8-1). This articulation has a loose joint capsule and is surrounded by large, strong muscles. The construction of this stable joint allows for the wide range of motion required for normal daily activities such as walking, sitting, and squatting. Such a joint must be precisely aligned and controlled.

THE ACETABULUM

The acetabulum is the concave component of the ball-and-socket configuration of the hip joint. The acetabular surface is covered with articular cartilage that thickens peripherally (Kempson et al. 1971) and predominantly laterally (Rushfeldt et al. 1979). The cavity of the acetabulum faces obliquely forward, outward, and downward. The osseous acetabulum in the hip is deep and provides substantial static stability to the hip. A plane through the circumference of the acetabulum at its opening would intersect with the sagittal plane at an angle of 40° opening posteriorly and with the transverse plane at an angle of 60° opening laterally. The acetabular cavity is deepened by the labrum, a flat rim of fibrocartilage, and the transverse acetabular ligament (Fig. 8-2). The labrum contains free nerve endings and sensory end organs in its superficial layer, which may participate in nociceptive and proprioceptive mechanisms (Kim & Azuma, 1995).

The unloaded acetabulum has a smaller diameter than the femoral head (Greenwald & Haynes, 1972) (Fig. 8-3). The acetabulum deforms about the femoral head when the hip joint is loaded. It undergoes elastic deformation to become congruous with the femoral head, and contact is made about the periphery of the anterior, superior, and posterior

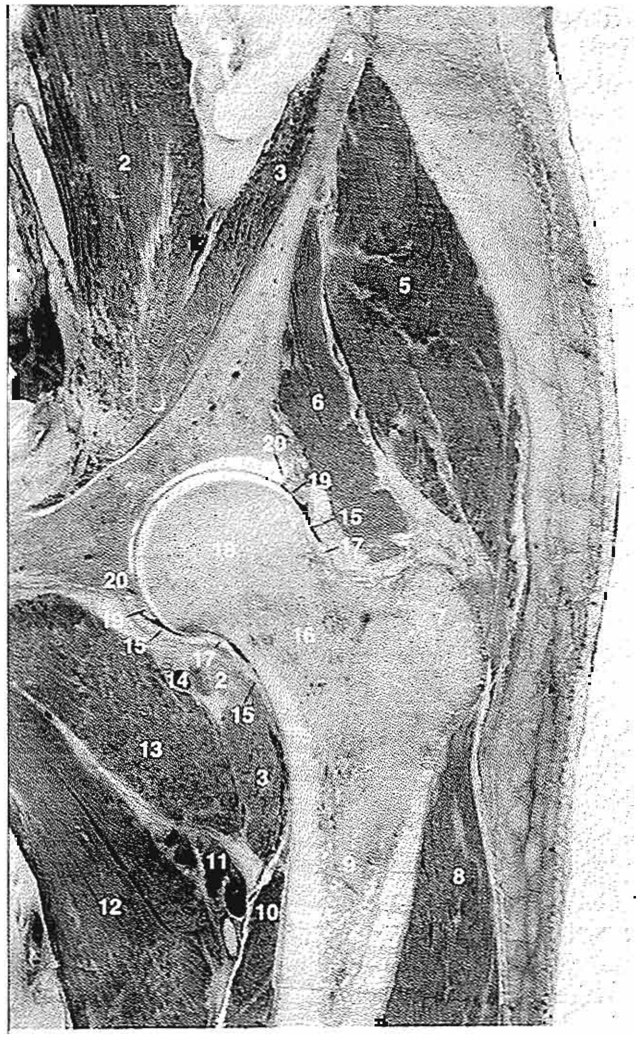


FIG. 8-1

The hip joint (front view). 1. External iliac artery. 2. Psoas major muscle. 3. Iliacus muscle. 4. Iliac crest. 5. Gluteus medius muscle. 6. Gluteus minimus muscle. 7. Greater trochanter. 8. Vastus lateralis muscle. 9. Shaft of femur. 10. Vastus medialis muscle. 11. Profunda femoris vessels. 12. Adductor longus muscle. 13. Pectineus muscle. 14. Medial circumflex femoral vessels. 15. Capsule of the hip joint. 16. Neck of femur. 17. Zona orbicularis of capsule. 18. Head of femur. 19. Acetabular labrum. 20. Rim of acetabulum. *Reprinted with permission from McMinn, R.M. & Hutchings, R.T. (1988). Color Atlas of Human Anatomy (2nd ed., p. 299). Chicago: Year Book Medical Publishers, Inc.*

articular surface of the acetabulum (Konrath et al., 1998). Load distribution of the acetabulum was studied in vitro in human specimens (Greenwald & Haynes, 1972; Konrath et al., 1998). Joint reaction forces were simulated to physiological levels. The loading pattern of the acetabulum is shown in Figure 8-3. Removal of the transverse acetabular ligament and labrum sequentially did not affect the loading pattern of the acetabulum significantly (Konrath et al., 1998).

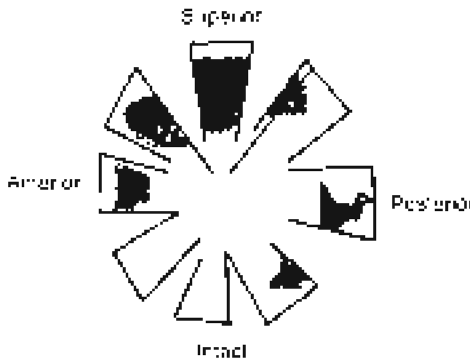


FIG. 8-3

Loading pattern of a human acetabulum in vitro with intact labrum and transverse acetabular ligament. Note: This pattern was grossly unchanged after removal of the transverse acetabular ligament, or the labrum, or both, and therefore these patterns are not displayed. *Adapted from Konrath, G.A., Hamel, A.J., Olson, S.A., et al. (1998). The role of the acetabular labrum and the transverse acetabular ligament in load transmission of the hip. J Bone Joint Surg, 80A(12), 1781–1788.*

THE FEMORAL HEAD

The femoral head is the convex component of the ball-and-socket configuration of the hip joint and forms two-thirds of a sphere. The articular cartilage covering the femoral head is thickest on the medial-central surface and thinnest toward the periphery. The variations in the cartilage thickness result in a different strength and stiffness in various regions of the femoral head (Kempson et al., 1971). Rydell (1965) suggested that most load was transmitted to the femoral head through the superior quadrant. Von Eisenhart-Rothe et al. (1997) demonstrated in an in vitro study that the magnitude of load influences the loading pattern on the femoral head. At smaller loads, the load-bearing area was concentrated at the periphery of the lunate surface of the femoral head, and at higher loads to the center of the lunate and the anterior and posterior horns. It is still not known exactly how the stresses in vivo on the normal femoral head are distributed, but indications from in vivo measurements with an instrumented prosthetic head show that the anterior and medial lunate is transmitting most of the load during daily activities (Bergmann et al., 1993, 1995).

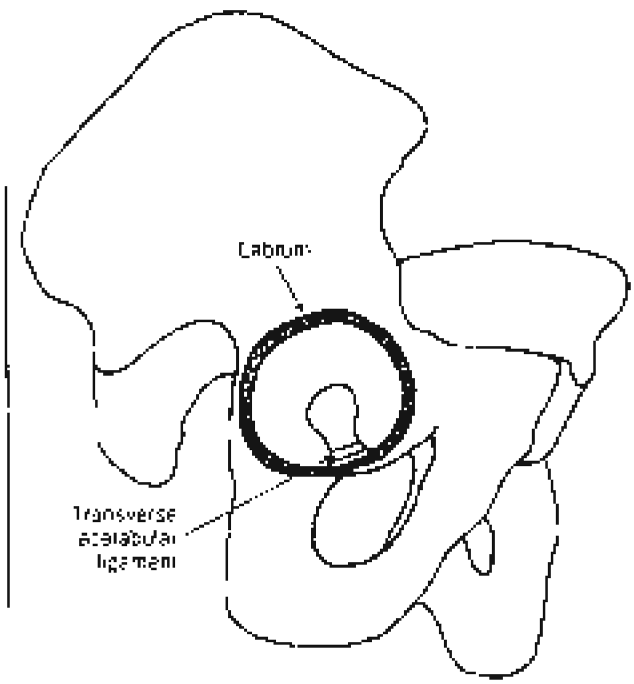


FIG. 8-2

Schematic drawing showing the lateral view of the acetabulum with the labrum and the transverse acetabular ligament intact. *Adapted from Konrath, G.A., Hamel, A.J., Olson, S.A., et al. (1998). The role of the acetabular labrum and the transverse acetabular ligament in load transmission of the hip. J Bone Joint Surg, 80A(12), 1781–1788.*

THE FEMORAL NECK

The femoral neck has two angular relationships with the femoral shaft that are important to hip joint function: the angle of inclination of the neck to the shaft in the frontal plane (the neck-to-shaft angle) and the angle of inclination in the transverse plane (the angle of anteversion). Freedom of motion of the hip joint is facilitated by the neck-to-shaft angle, which offsets the femoral shaft from the pelvis laterally. In most adults, this angle is approximately

125°, but it can vary from 90° to 135°. An angle exceeding 125° produces a condition known as coxa valga; an angle less than 125° results in coxa vara (Fig. 8-4). Deviation of the femoral shaft in either way alters the force relationships about the hip joint and has a mechanical effect on the lever arms for muscle force and line of gravity.

The angle of anteversion is formed as a projection of the long axis of the femoral head and the transverse axis of the femoral condyles (Fig. 8-5). In adults, this angle averages approximately 12°, but it varies a great deal. Anteversion of more than 12° causes a portion of the femoral head to be uncovered and creates a tendency toward internal rotation of the leg during gait to keep the femoral head in the acetabular cavity. An angle of less than 12° (retroversion) produces a tendency toward external rotation of the leg during gait. Both anteversion and retroversion are fairly common in children but are usually outgrown.

The interior of the femoral neck is composed of cancellous bone with trabeculae organized into medial and lateral trabecular systems (Fig. 8-1, 8-6). The fact that the joint reaction force on the femoral head parallels the trabeculae of the medial

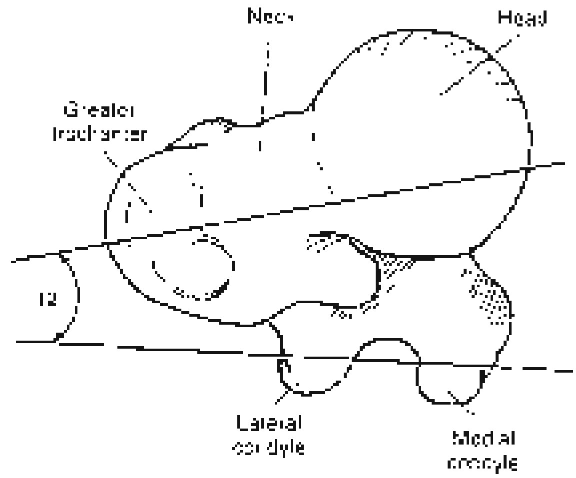


FIG. 8-5

Top view of the proximal end of the left femur showing the angle of anteversion, formed by the intersection of the long axis of the femoral head and the transverse axis of the femoral condyles. This angle averages approximately 12° in adults.

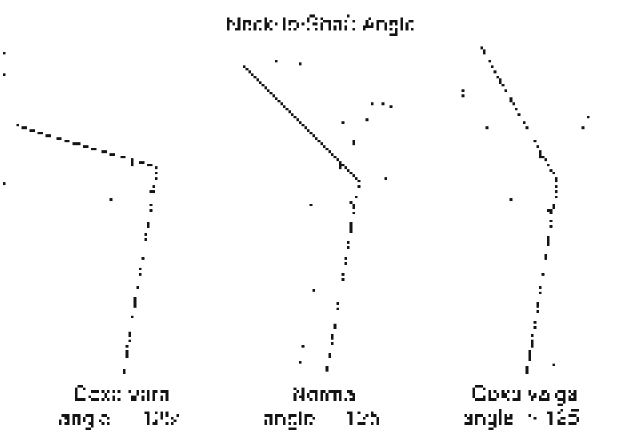


FIG. 8-4

The normal neck-to-shaft angle (angle of inclination of the femoral neck to the shaft in the frontal plane) is approximately 125°. The condition in which this angle is less than 125° is called coxa vara. If the angle is greater than 125°, the condition is called coxa valga.

system (Frankel, 1960) indicates the importance of this system for supporting this force. The epiphyseal plates are at right angles to the trabeculae of the medial system and are thought to be perpendicular to the joint reaction force on the femoral head (Inman, 1947). It is likely that the lateral trabecular system resists the compressive force on the femoral head produced by contraction of the abductor muscles—the gluteus medius, the gluteus minimus, and the tensor fasciae latae. The thin shell of cortical bone around the superior femoral neck progressively thickens in the inferior region.

With aging, the femoral neck gradually undergoes degenerative changes: the cortical bone is thinned and cancellated and the trabeculae are gradually resorbed (see Fig. 2-50). These changes may predispose the femoral neck to fracture. It is noteworthy that the femoral neck is the most common fracture site in elderly persons (Case Study 8-1).

Kinematics

In considering the kinematics of the hip joint, it is useful to view the joint as a stable ball and socket configuration wherein the femoral head and acetabulum can move in all directions.

FIG. 8-6

Roentgenogram of a femoral neck showing the medial and lateral trabecular systems. The thin shell of cortical bone around the superior femoral neck progressively thickens in the inferior region.

RANGE OF MOTION

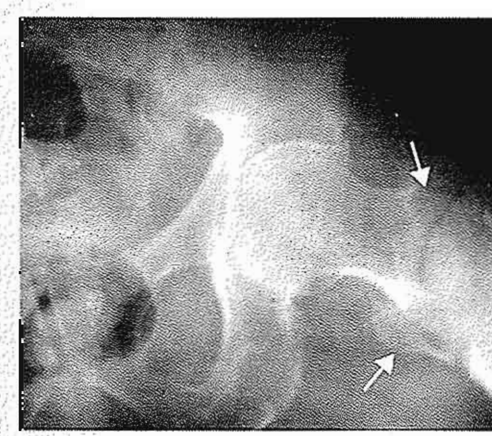
Hip motion takes place in all three planes: sagittal (flexion-extension), frontal (abduction-adduction), and transverse (internal-external rotation) (Fig. 8-7). Motion is greatest in the sagittal plane, where the range of flexion is from 0 to approximately 140° and the range of extension is from 0 to 15°. The range of abduction is from 0 to 30°, whereas that of adduction is somewhat less, from 0 to 25°. External rotation ranges from 0 to 90° degrees and internal rotation from 0 to 70° when the hip joint is flexed. Less rotation occurs when the hip joint is extended because of the restricting function of the soft tissues.

The range of motion of the hip joint during gait has been measured electrogoniometrically in all three planes. Measurements in the sagittal plane during level walking (Murray, 1967) showed that the joint was maximally flexed during the late swing phase of gait, as the limb moved forward for heel strike. The joint extended as the body moved forward at the beginning of stance phase. Maximum extension was reached at heel-off. The joint reversed into flexion during swing phase and again reached maximal flexion, 35 to 40°, prior to heel strike. Figure 8-8A shows the pattern of hip joint motion in the sagittal plane during a gait cycle and

CASE STUDY 8-1

Femoral Intertrochanteric Fractures

An 80-year-old woman falls from her own height after losing her balance. She presented with sharp pain in her hip and an inability to stand and walk by herself. She is transported to the E.R. and after a careful examination and x-ray evaluation, a right intertrochanteric fracture is diagnosed.



Case Study Figure 8-1-1.

The radiograph illustrates a right femoral intertrochanteric unstable fracture with separation of the lesser trochanter. The image shows osteoporotic changes characteristic of the aging process. The decrease in the bone mass at the femoral neck leads to reduced bone strength and stiffness as a result of diminution in the amount of cancelous bone and thinning of cortical bone. It increases the likelihood of a fracture at the weakest level.

During the fall, the magnitude of the compressive forces at the femoral neck overcame its stiffness and strength. In addition, the tensile forces produced by protective contraction of muscles such as the iliopsoas generated a traction fracture at the lesser trochanter level.

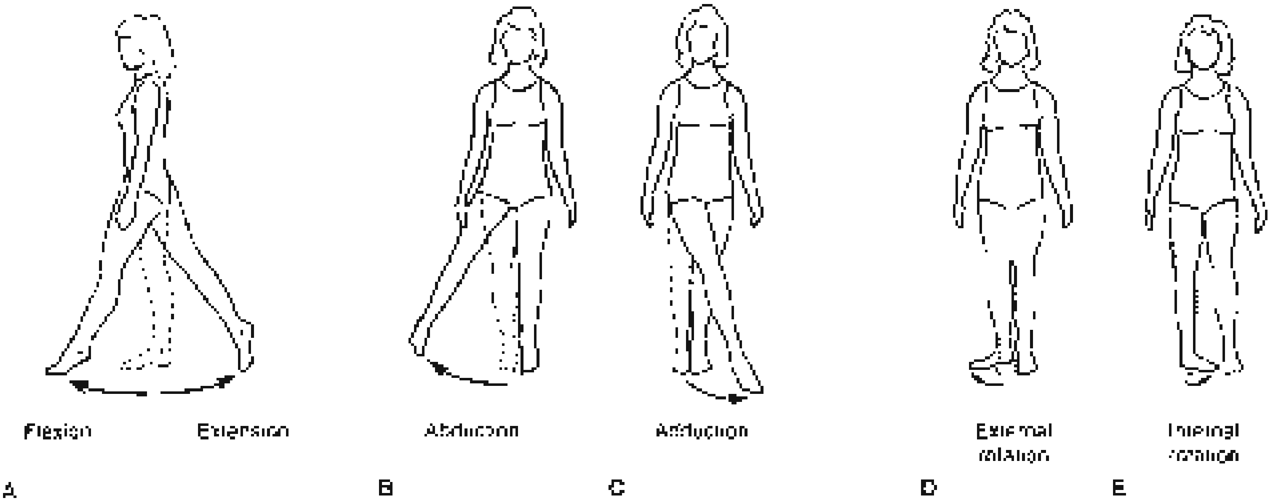


FIG. 8-7

Movements of the hip joint. A, Flexion-extension. B, Abduction. C, Adduction. D, External rotation. E, Internal rotation.

allows a comparison of this motion with that of the knee and ankle.

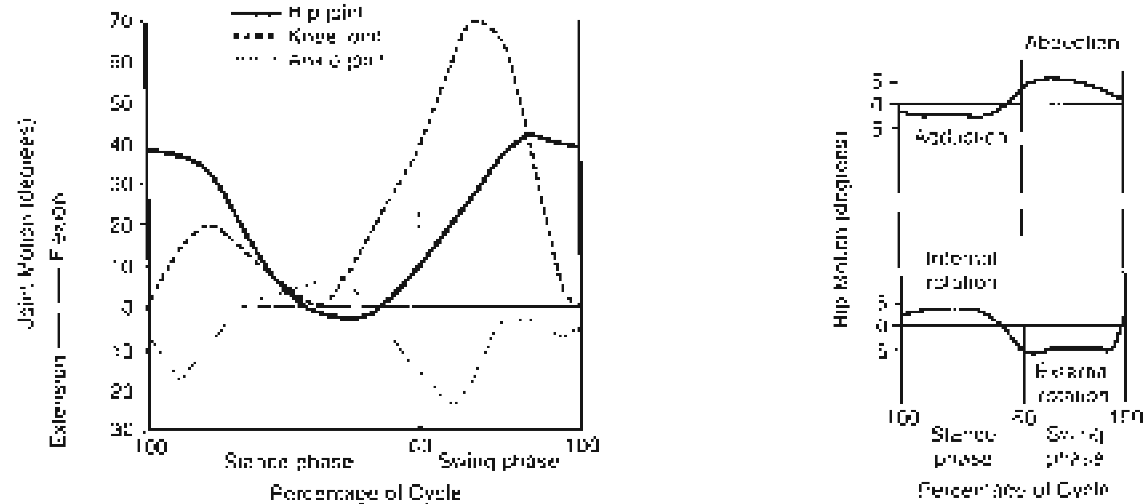
Motion in the frontal plane (abduction-adduction) and transverse plane (internal-external rotation) during gait (Johnston & Smidt, 1969) is illustrated in Figure 8-8B. Abduction occurred during swing phase, reaching its maximum just after toe-off; at heel strike, the hip joint reversed into adduction, which continued until late stance phase. The hip joint was externally rotated throughout the swing phase, rotating internally just before heel strike. The joint remained internally rotated until late stance phase, when it again rotated externally. The average ranges of motion recorded for the 33 normal men in this study were 12° for the frontal plane and 13° for the transverse plane.

As people age, they use a progressively smaller portion of the range of motion of the lower extremity joints during ambulation. Murray and coworkers (1969) studied the walking patterns of 67 normal men of similar weight and height ranging in age from 20 to 87 years and compared the gait patterns of older and younger men. The differences in the sagittal body positions of the two groups at the instant of heel strike are illustrated in Figure 8-9. The older men had shorter strides, a decreased range of hip flexion and extension, decreased plantar flexion of the ankle, and a decreased heel-to-floor angle of the tracking limb; they also showed reduced dorsiflexion of the ankle and diminished elevation of the toe of the forward limb.

The range of motion in three planes during common daily activities such as tying a shoe, sitting down on a chair, rising from it, picking up an object from the floor, and climbing stairs was measured electrogoniometrically in 33 normal men by Johnston and Smidt (1970). The mean motion during these activities is shown in Table 8-1. Maximal motion in the sagittal plane (hip flexion) was needed for tying the shoe and bending down to squat to pick up an object from the floor. The greatest motion in the frontal and transverse planes was recorded during squatting and during shoe tying with the foot across the opposite thigh. The values obtained for these common activities indicate that hip flexion of at least 120°, abduction and external rotation of at least 20° are necessary for carrying out daily activities in a normal manner.

SURFACE JOINT MOTION

Surface motion in the hip joint can be considered as gliding of the femoral head on the acetabulum. The pivoting of the ball and socket in three planes around the center of rotation in the femoral head (estimated as the center of the femoral head) produces this glid



A

B

FIG. 8-8

A, Range of hip joint motion in the sagittal plane for 30 normal men during level walking, one gait cycle. The ranges of motion for the knee and ankle joints are shown for comparison. Adapted from Murray, M.P. (1967). Gait as a total pattern of movement. *Am J Phys Med, 46,* 290. B, A typical pattern for range of motion in the frontal plane (top) and transverse plane (bottom) during level walking, one gait cycle. Adapted from Johnston, R.C., & Smidt, G.L. (1969). Measurement of hip joint motion during walking. Evaluation of an electrogoniometric method. *J Bone Joint Surg, 51A,* 1083.

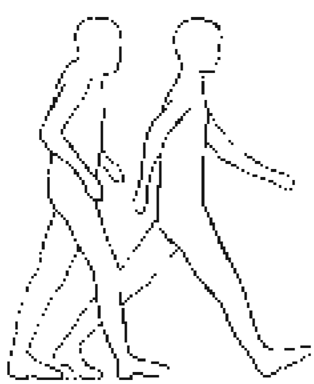


FIG. 8-9

Differences in the sagittal body positions of older men (left) and younger men (right) at the instant of heel strike. The older men showed shorter strides, a decreased range of hip flexion and extension, decreased plantar flexion of the ankle, and a decreased heel to floor angle of the tracking limb; they also showed less dorsiflexion of the ankle and less elevation of the toe of the forward limb. Reprinted with permission from Murray, M.P., Kory, R.C., & Clarkson, B.H. (1969). Walking patterns in healthy old men. *J Gerontol, 24,* 169–178.

ing of the joint surfaces. If incongruity is present in the femoral head, gliding may not be parallel or tangential to the joint surface and the articular cartilage may be abnormally compressed or distracted. Instant center analysis by means of the Reuleaux method cannot be performed accurately in the hip joint because motion takes place in three planes simultaneously. Locating the center of rotation of the hip joint is essential for prosthetic surgery of the hip to reconstruct an optimal lever arm of the gluteus medius muscle (Fuss et al., 1999).

Kinetics

Kinetic studies have demonstrated that substantial forces act on the hip joint during simple activities (Hurwitz & Andriacchi, 1997, 1998). The factors involved in producing these forces must be understood if rational rehabilitation programs are to be

TABLE 8-1

Mean Values for Maximum Hip Motion in Three Planes During Common Activities

Activity	Plane of Motion	Recorded Value (Degrees)
Tying shoe with foot on floor	Sagittal	124
	Frontal	19
	Transverse	15
Tying shoe with foot across opposite thigh	Sagittal	110
	Frontal	23
	Transverse	33
Sitting down on chair and rising from sitting	Sagittal	104
	Frontal	20
	Transverse	17
Stooping to obtain object from floor	Sagittal	117
	Frontal	21
	Transverse	18
Squatting	Sagittal	122
	Frontal	28
	Transverse	26
Ascending stairs	Sagittal	67
	Frontal	16
	Transverse	18
Descending stairs	Sagittal	36

Data from Johnston, R.C., & Smidt, G.L. (1970). Hip motion measurements for selected activities of daily living. Clin Orthop, 72, 205.

developed for patients with pathological conditions of the hip. The abductor muscle group (the gluteus medius and minimus muscles) is the main stabilizer during one-legged stance (Kumagai et al., 1997; University of California, 1953).

STATICS

During a two-leg stance, the line of gravity of the superincumbent body passes posterior to the pubic symphysis, and, because the hip joint is stable, an erect stance can be achieved without muscle contraction through the stabilizing effect of the joint capsule and capsular ligaments. With no muscle activity to produce moments around the hip joint, calculation of the joint reaction force becomes simple: the magnitude of this force on each femoral head during upright two-legged stance is one half the weight of the superincumbent body. Because each lower extremity is one sixth of body weight, the reaction force on each hip joint will be one half of the remaining two thirds, or one third of body weight; however, if the muscles surrounding the hip joint contract to prevent swaying and to maintain an upright position of the body (e.g., during prolonged standing), this force increases in proportion to the amount of muscle activity.

When a person changes from a two-leg to a single-leg stance, the line of gravity of the superincumbent body shifts in all three planes, producing moments around the hip joint that must be counteracted by muscle forces and thus increasing the joint reaction force. The magnitude of the moments, and hence the magnitude of the joint reaction force, depends on the posture of the spine, the position of the non–weight-bearing leg and upper extremities, and the inclination of the pelvis (McLeish & Charnley, 1970). Figure 8-10 demonstrates how the line of gravity in the frontal plane shifts with four different positions of the upper body and inclinations of the pelvis: standing with the pelvis in a neutral position, standing with a maximum tilt of the upper body over the supporting hip joint, standing with the upper body tilting away from the supporting hip joint, and standing with the pelvis sagging away from the supporting hip joint (Trendelenburg's test). The shift in the gravity line, and hence in the length of the lever arm of the gravitational force (the perpendicular distance between the gravity line and the center of rotation in the femoral head), influences the magnitude of the moments about the hip joint and, consequently, the joint reaction force. The gravitational force lever arm and the joint reaction force are minimized when the trunk is tilted over the hip joint (Fig. 8-10B).

Two methods are used for deriving the magnitude of the joint reaction force acting on the femoral head: the simplified free-body technique for coplanar forces and a mathematical method utilizing equilibrium equations. The simplified free-body technique for coplanar forces was described in detail in Chapter 7, in Calculation Box 7-1. This technique is used in the hip to estimate the joint reaction force in the frontal plane acting on the femoral head during a single-leg stance with the pelvis in a neutral position (Calculation Box 8-1). The second method is a mathematical calculation of the joint reaction force on the femoral head using equilibrium equations for a single-leg stance with the pelvis level (Calculation Box 8-2).

To understand and solve the equations it is necessary to indicate first the location of the external forces acting on the body during the single-leg

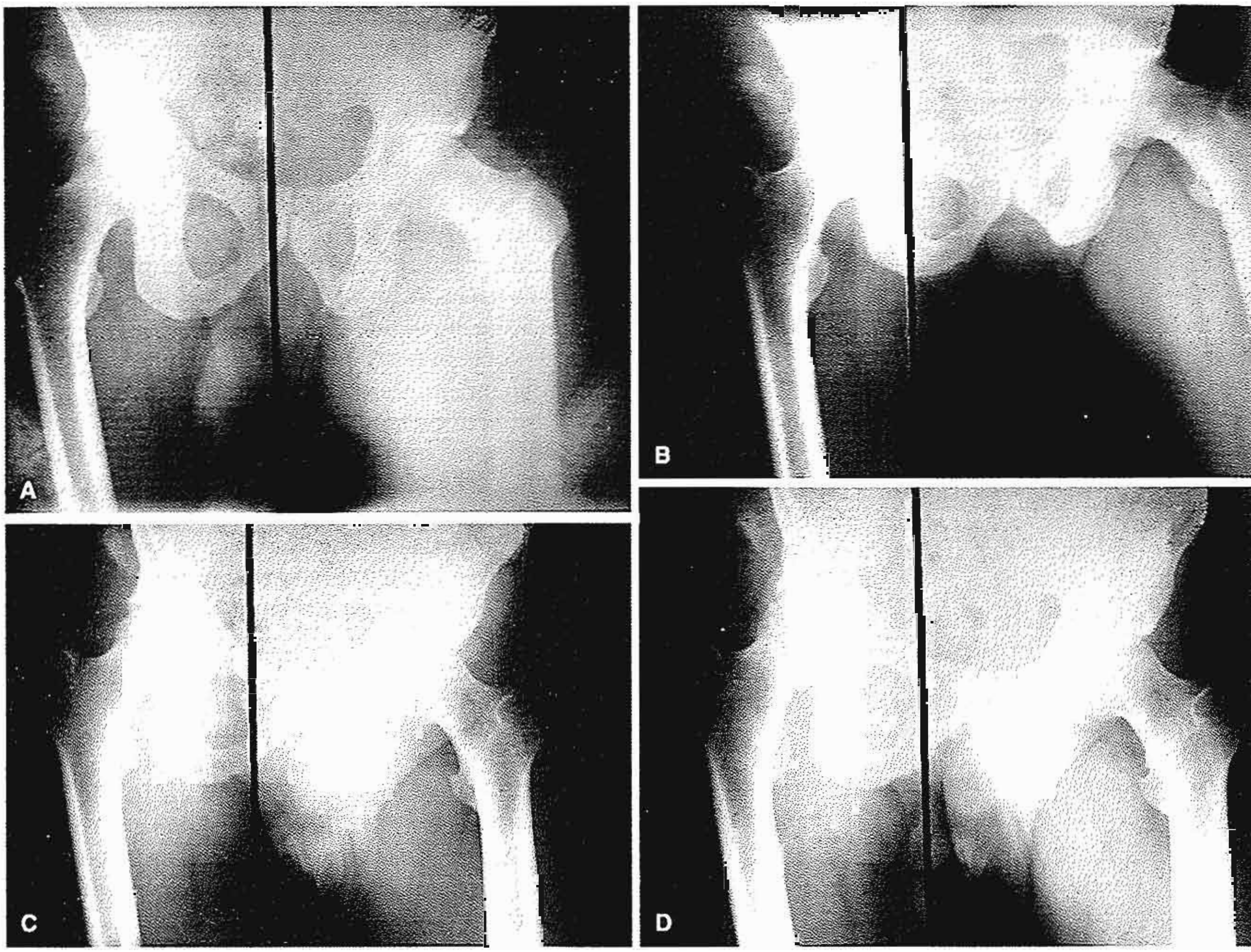


FIG. 8-10

Roentgenograms utilizing a plumb line (*black line*) show that the line of gravity shifts in the frontal plane with different positions of the upper body and inclinations of the pelvis. **A,** The pelvis is in a neutral position. The gravity line falls approximately through the pubic symphysis. The lever arm for the force produced by body weight (the perpendicular distance between the gravity line and the center of rotation in the femoral head) influences the moment about the hip joint and hence the joint reaction force. **B,** The shoulders are maximally tilted over the supporting hip joint. The gravity line has shifted and is now nearest the supporting hip. Because this shift minimizes the lever arm, the moment about the hip joint and the joint reaction force are also minimized. **C,** The shoulders are maximally tilted away from the supporting hip joint. Again the gravity line has shifted toward the supporting hip, thus decreasing the joint reaction force. **D,** The pelvis sags away from the supporting hip joint (Trendelenburg's test). The shift in the gravity line is similar to that in C. (*Courtesy of* [illegible], *M.D., Case Western Reserve University, Cleveland, Ohio.*) Note: In B, the antalgic gait is illustrated, which lowers the load on the head of the femur but alters the load line to a more vertical position. Following arthroplasty for arthritis, the abductor muscles are weak and atrophic as a result of the disease process and the surgery. External support such as a cane should be used until the abductor muscles are rehabilitated. The best indication for a rehabilitated abductor muscle is the lack of limping.

stance on a free-body diagram (Calculation Box Fig 8-2-1). Because the body is in equilibrium (i.e., the sum of the moments and the sum of the forces both equal zero), the ground reaction force is equal to the gravitational force of the body, which can be divided into two components, the gravitational force of the stance leg (equal to one-sixth body weight) and the remaining force (equal to five-sixths body weight).

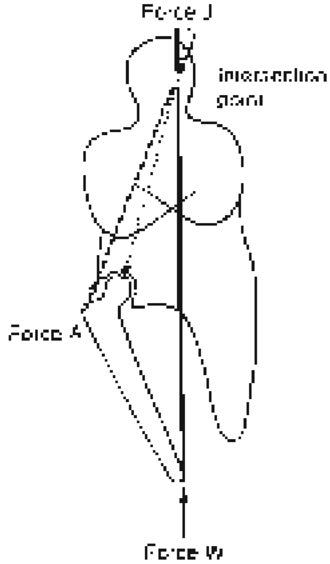
Next, the body is divided at the hip joint into two free-bodies. The main coplanar forces and moments acting on these free-bodies must be determined. The upper free body is considered first (Calculation Box Fig. 8-2-2). In this free-body, two moments are required for stability. The moment arising from the superincumbent body weight (equal to ⅚ W) must be balanced by a moment arising from the force of the abductor muscles. The force produced by the superincumbent body weight (⅚ W) acts at a distance of b from the center of rotation of the hip (O), thus producing a moment of ⅚ W times b. The force produced by the principal abductor, the gluteus medius, designated as A, acts at a distance of c from the center of rotation, producing a counterbalancing moment of A times c. For the body to remain in moments equilibrium, the sum of the moments must equal zero. In this example, the moments acting

CALCULATION BOX 8-1

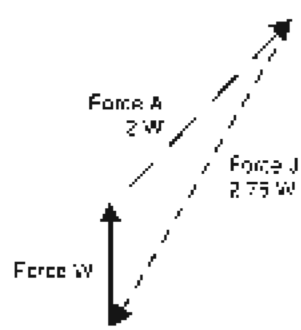
Simplified Free-Body Technique for Coplanar Forces

The stance limb is considered as a free-body, and a free-body diagram is drawn. From all of the forces acting on the free-body, the three main coplanar forces are identified as the force of gravity against the foot (the ground reaction force transmitted through the tibia to the femoral condyles), the force produced by contraction of the abductor muscles, and the joint reaction force on the femoral head. The ground reaction force (W) has a known magnitude equal to five-sixths of body weight and a known sense, line of application, and point of application. The abductor muscle force (A) has a known sense, a known line of application, and point of application estimated from the muscle origin and insertion on a roentgenogram but an unknown magnitude. Because several muscles are involved in the action of hip abduction, simplifying assumptions are made in determining the direction of this force (McLeish & Charnley, 1970). Furthermore, forces produced by other muscles active in stabilizing the hip joint are not taken into account. The joint reaction force (J) has a known point of application on the surface (outside) of the femoral head but an unknown magnitude, sense, and line of application.

The magnitudes of the abductor muscle force and the joint reaction force can be derived by designating all three forces on the free-body diagram (Calculation Box Fig. 8-1-1) and constructing a triangle of forces (Calculation Box Fig. 8-1-2). The muscle force is found to be approximately two times body weight, whereas the joint reaction force is somewhat greater.



Calculation Box Figure 8-1-1

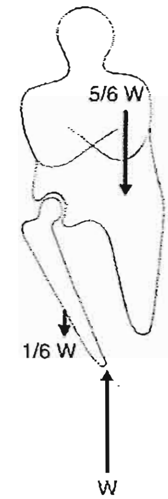


Calculation Box Figure 8-1-2

CALCULATION BOX 8-2

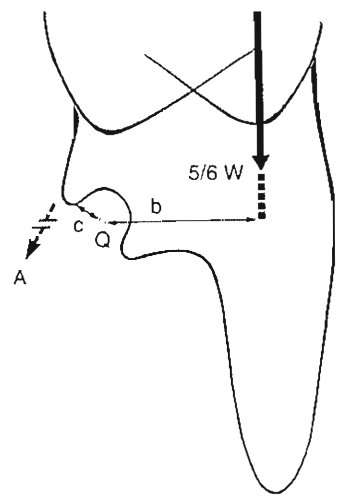
External Forces Acting on the Body in Equilibrium During a Single-Leg Stance

Calculation Box Figure 8-2-1 shows external forces acting on the body in equilibrium during a single-leg stance. The ground reaction force is equal to body weight (W). The gravitational force of the stance leg is equal to one-sixth of body weight; the remaining force is equal to five-sixths of body weight.



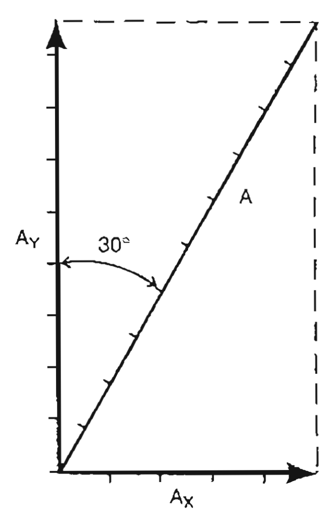
Calculation Box Figure 8-2-1.

The internal forces acting on the hip joint are found by separating the joint into an upper and lower free-body; the upper free-body is considered first. In this free-body, two moments are required for stability. Moment equilibrium is attained by the production of two equal moments. A moment arising from the force of the abductor muscle (A) times abductor force lever arm (c) counterbalances the moment arising from the gravitational force of the superincumbent body (5/6 W) times gravitational force lever arm (b), which tends to tilt the pelvis away from the supporting lower extremity. Q, center of rotation of the hip joint.



Calculation Box Figure 8-2-2.

Force A is equal to two times body weight and has a direction of 30° from the vertical. The magnitudes of its horizontal (A_x) and vertical (A_y) components are found by vector analysis. Perpendiculars are drawn from the tail of A to a horizontal and a vertical line representing A_x and A_y, respectively. A_x and A_y can then be scaled off. Alternatively, trigonometry is used to find the magnitudes of the components.



$A = 2\,W$
$A_X = A \cdot \sin 30°$
$A_X = 0.5\,A = W$

$A_Y = A \cdot \cos 30°$
$A_Y = 0.8\,A = 1.7\,W$

Calculation Box Figure 8-2-3.

clockwise are arbitrarily considered to be positive and the counterclockwise moments are considered to be negative. Thus,

$$(\tfrac{5}{6}W \times b) - (A \times c) = 0$$

$$A = \frac{\tfrac{5}{6}W \times b}{c}$$

To solve for A it is necessary to find the values of b and c. The gravitational force lever arm (b) is found roentgenographically. Because the center of gravity must be over the base of support, a plumb line intersecting the heel can be extended upward; a perpendicular line drawn from the center of rotation in the femoral head (O) to the line represents distance b. The muscle force lever arm (c) is similarly found by identifying the gluteus medius muscle on a roentgenogram (Nemeth & Ohlsen, 1985, 1989) and drawing a perpendicular line from the center of rotation of the femoral head to a line approximating the gluteus medius muscle tendon.

In this example, a value for A of two times body weight was derived from the static free-body diagram and confirmed by instrumented in vivo measurements (English & Kilvington, 1979; Rydell, 1966). The direction of force A is found from a roentgenogram to be 30° from the vertical. The horizontal and vertical components of this force are found by vector analysis (Calculation Box Fig. 8-2-3). The horizontal component (A_x) is equal to body weight; the vertical component (A_y) is approximately 1.7 times body weight.

Attention is then directed to the lower free-body (Calculation Box Fig 8-3-1). The gravitational forces (W and ⅚ W) are known. The joint reaction force (force J) has an unknown magnitude and direction but originates from the most narrow joint space in the radiograph and must pass through the estimated center of rotation in the femoral head. The magnitude of force J is determined by finding the horizontal and vertical force components and adding them (Calculation Box Fig. 8-3-2).

The value of J is found by vector addition (Calculation Box Fig. 8-3-3), and its direction is measured on the parallelogram of forces. The joint reaction force on the femoral head in a single-leg stance with the pelvis leveled in the horizontal plane is found to be approximately 2.7 times body weight, and its direction is 69° from the horizontal (Calculation Box Fig. 8-3-4).

A key factor influencing the magnitude of the joint reaction force on the femoral head is the ratio of the abductor muscle force lever arm (c) to the gravitational force lever arm (b) (Calculation Box Fig. 8-2-2). This is particularly important in prosthetic replacements of the hip joint (Delp & Maloney, 1993; Free & Delp, 1996; Lim et al., 1999; Sutherland et al., 1999; Vasavada et al., 1994). The center of motion can be altered by the prosthetic design and the lever arm for the abductor muscles can be slightly changed by surgery techniques. A change of the center of location of the hip joint can decrease the abduction force by more than 40% and thereby the generated abductor moment by almost 50% (Delp & Maloney, 1993). Figure 8-11 illustrates the relationship of this ratio to the joint reaction force. A low ratio (i.e., a small muscle force lever arm and a large gravitational force lever arm) yields a greater joint reaction force than does a high ratio.

A short lever arm of the abductor muscle force, as in coxa valga (Fig. 8-5), results in a small ratio and thus a somewhat elevated joint reaction force. Moving the greater trochanter laterally during total hip replacement lowers the joint reaction force, as it increases the lever arm ratio by increasing the muscle force lever arm (Free & Delp, 1996). Inserting a prosthetic cup deeper in the acetabulum reduces the gravitational force lever arm, thereby increasing the ratio and decreasing the joint reaction force. It is difficult, however, to change the lever arm ratio in such a way as to reduce the joint reaction force significantly because the curve formed from plotting the ratios becomes asymptotic when the ratio of c to b approaches 0.8 (Fig. 8-11).

DYNAMICS

The loads on the hip joint during dynamic activities have been studied by several investigators (Andriacchi et al., 1980; Draganich et al., 1980; English & Kilvington, 1979; Rydell, 1966). Using a force plate system and kinematic data for the normal hip, J. P. Paul, 1967, (Forces at the human hip joint. Unpublished doctoral theses. University of Chicago) examined the joint reaction force on the femoral head in normal men and women during gait and correlated the peak magnitudes with specific muscle activity recorded electromyographically. In the men, two peak forces were produced during the stance phase when the abductor muscles contracted to stabilize the pelvis. One peak of approximately four times body weight occurred just after heel strike, and a large peak of approximately seven times body weight was reached just before toe-off (Fig. 8-12A). During foot flat, the joint reaction force decreased to approximately

CALCULATION BOX 8-3

Free-Body Diagram of the Lower Extremity

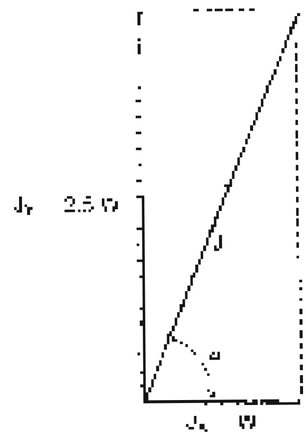
In Calculation Box Figure 8-3-1, the supporting lower extremity is considered as a free body and the forces acting on the free body are identified: [illegible] force J, joint reaction force, 1/6 W, gravitational force of the limb, W, ground reaction force, O, center of rotation of the hip joint.

In Calculation Box Figure 8-3-2, the forces acting on the lower extremity are resolved into horizontal and vertical components. Because the body is in equilibrium, the sum of the forces in the horizontal direction must equal zero and so must the forces in the vertical direction. The horizontal and vertical forces are added and the magnitudes of J and [illegible] are found from force equilibrium equations.

[illegible]

From the Calculation Box 8-2, A = [illegible]

$J_y = 2.5W$

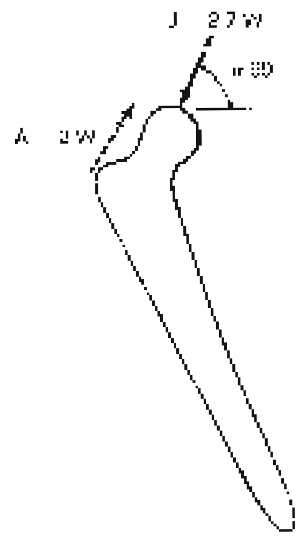


Calculation Box Figure 8-3-3.

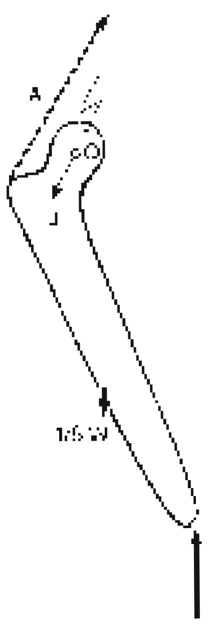
$\text{Tan}\,\alpha = \frac{J_y}{J_x}$

$\text{Tan}\,\alpha = 2.5$

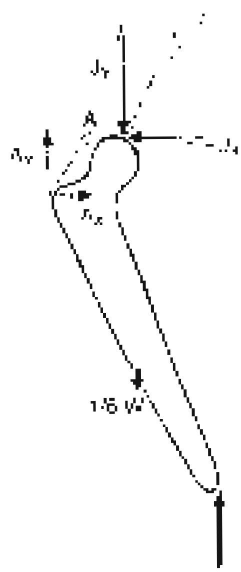
$\alpha = 69$



Calculation Box Figure 8-3-4.



Calculation Box Figure 8-3-1.



Calculation Box Figure 8-3-2.

Addition of the horizontal and vertical components [illegible] and [illegible] is performed graphically in Figure 8-3-3, and the joint reaction force [illegible] parallelogram is constructed and the diagonal of the parallelogram indicates the magnitude of the force J. Its inclination in relation to the horizontal plane is measured from the force diagram. Alternately, trigonometry is used to find the direction of J using tangent equations (Fig. 8-3-4).

The joint reaction force has a magnitude of approximately 2.7 times body weight and acts at an angle 69 from the horizontal.

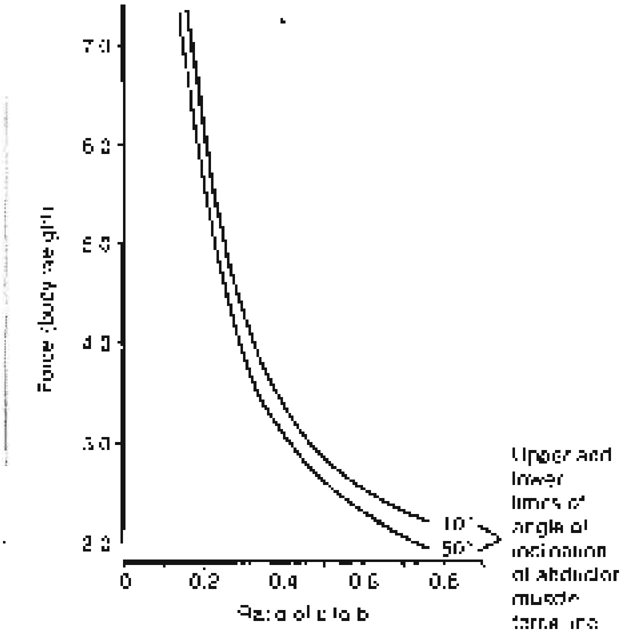


FIG. 8-11

The value of the ratio of the abductor muscle force lever arm (c) to the gravitational force lever arm (b) is plotted against the joint reaction force on the femoral head in units of body weight. Because the line of application of the abductor muscle force (its angle of inclination in the frontal plane) has finite upper and lower limits (10 and 50°), the force envelope is plotted. The curve can be utilized to determine the minimal force acting on the femoral head during a one-leg stance if the ratio of c to b is known. *Adapted from Frankel, V.H. (1960). The Femoral Neck: Function, Fracture Mechanisms, Internal Fixation. Springfield: Charles C Thomas.*

body weight because of the rapid deceleration of the body's center of gravity. During the swing phase, the joint reaction force was influenced by contraction of the extensor muscles in decelerating the thigh, and the magnitude remained relatively low, approximately equal to body weight.

In the women, the force pattern was the same but the magnitude was somewhat lower, reaching a maximum of only approximately four times body weight at late stance phase (Fig. 8-12B). The lower magnitude of the joint reaction force in the women may have been the result of several factors: a wider female pelvis, a difference in the inclination of the femoral neck-to-shaft angle, a difference in footwear, and differences in the general pattern of gait.

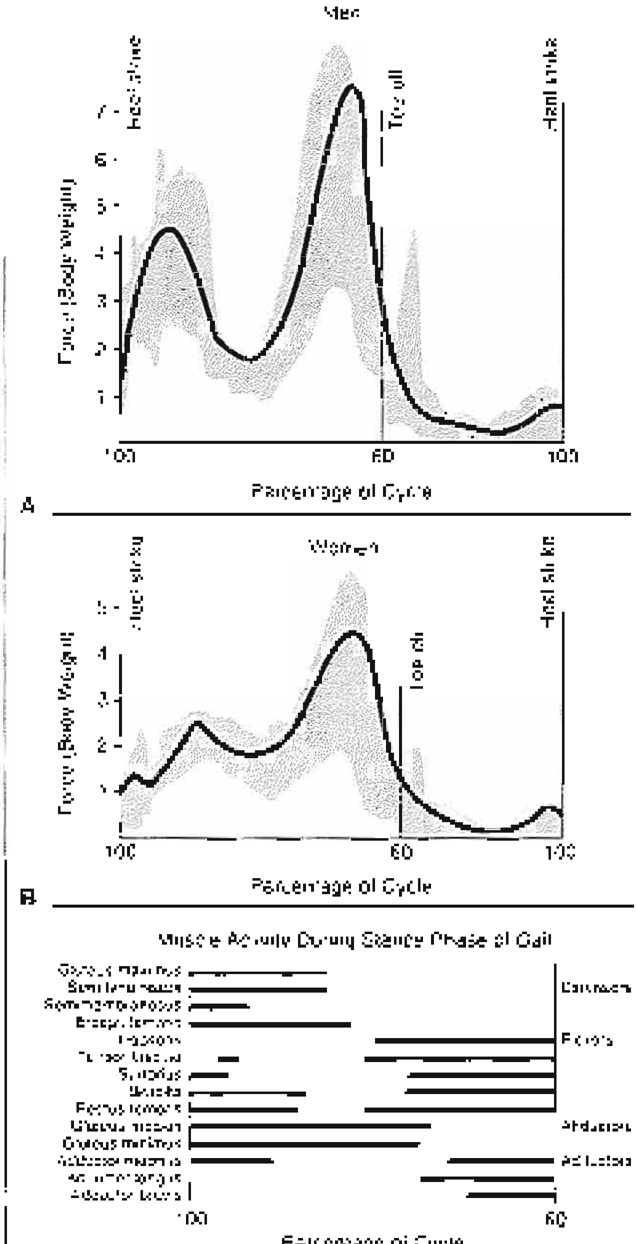


FIG. 8-12

Hip joint reaction force in units of body weight during walking, one gait cycle. The shaded area indicates variations among subjects. **A**, Force pattern for normal men. **B**, Force pattern for normal women. *Adapted from Paul, J.P. (1967). Forces at the human hip joint. Unpublished doctoral thesis, University of Glasgow.* **C**, Muscle activity during stance phase of gait. The first peak corresponds mainly to the extensor and abductor muscles. The last peak is for the flexors and adductor muscles. *Adapted from the University of California (1953). The pattern of muscular activity in the lower extremity during walking. Univ Cal Prosthet Dev Res Rep 2(25), 1–41.*

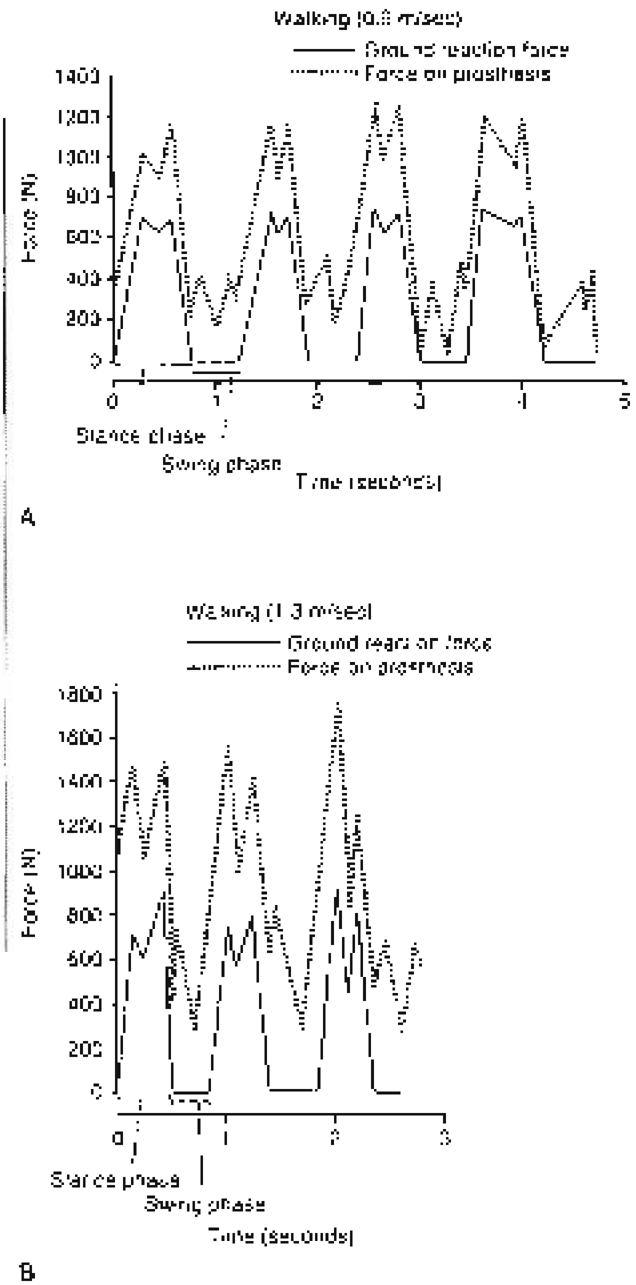


FIG. 8-13

Forces on an instrumented hip prosthesis during walking. The *broken line* represents the force on the prosthesis, and the *solid line* represents the ground reaction force. **A**, Walking speed 0.9 m per second. **B**, Walking speed 1.3 m per second. An increase in muscle activity at the faster cadence resulted in higher forces on the prosthesis. *Adapted from Rydell, N. (1965). Forces in the hip joint. Part II. Intravital measurements. In R.M. Kenedi (Ed.): Biomechanics and Related Bio-Engineering Topics (pp. 351–357). Oxford: Pergamon Press.*

In vivo measurements of the forces acting on an instrumented hip joint prosthesis demonstrate the lower joint reaction force of the femoral head during the stance phase of gait compared with external measurements and calculations (Rydell, 1965) (Fig. 8-13A). At a faster cadence, the forces acting on the prosthesis greatly increased because of an increase in muscle activity (Fig. 8-13B). At both cadences, the magnitude of the forces during swing phase was approximately half that during stance phase.

Table 8-2 summarizes the typical peak joint forces on the hip joint load expressed as body weight from different studies and with different methods. The pattern of loading for walking is similar for all studies, but the magnitude of joint peak load differs. External measurements generally yield higher calculated peak force on the hip joint while instrumented implant in vivo measurements yield lower peak forces. There are many reasons for the difference, for example, the method and instrumentation, the normal hip versus the "abnormal" instrumented implant, the gait velocity, and age. Activities other than walking, such as stair ascending/descending, yield loads of around 2.6 to 5.5 body weight measured

TABLE 8-2

Range of Typical Reported Peak Hip Joint Forces From Selected Studies

Activity	Reported Peak Force BW	Instrumentation	Reference
Walking	2.7–4.3	Instrumented implants	Bergmann et al., 1993, 1995
	2.7–3.5	Instrumented implants	Kotzar et al., 1991
	2.7	Instrumented implants	English et al., 1979
	1.9–3.3	Instrumented implants	Rydell et al., 1966
Walking	4.9–7.0	EMG Force plate	Paul, 1967
	4.5–7.5	EMG Force plate	Crowninshield et al., 1978
	5.0–8.0	EMG Force plate	Röhrle et al., 1984
	2.2–2.8	Accelerometers	van den Bogert et al., 1999

BW, body weight; EMG, electromyography.

CASE STUDY 8-2

Fatigue Fracture of the Hip

A 64-year-old, very active retired man experienced a femoral neck fracture after changing his training regime to prepare for a marathon. The fracture was classified as a fatigue fracture caused by overload of the hip joint.

Case Study Figure 8-2-1.

The figure shows an MR (frontal view) of the pelvis and both hip joints. The fracture is seen in the left femoral neck distal to the femoral head. The fracture is believed to have occurred during running and after an extensive change of training program. Because of the high repetitive loading, muscle fatigue, and the change in the load pattern on the hip joint and femoral neck, the bone fractured.

with an instrumented hip implant (Bergmann et al., 1993; Kotzar et al., 1991). The highest magnitudes of load during daily activities are measured during stair climbing and getting up from a low chair when the hip is flexed more than 100° (Catani et al., 1995; Johnston et al., 1979). Co-contraction of the biarticular muscles was evident during these activities. Running and skiing using accelerometers yielded calculated forces up to eight times body weight in middle aged and older people (van den Bogert et al., 1999) (Case Study 8-2).

Insertion of an instrumented nail plate in the proximal femur after osteotomy or during fixation of a femoral neck fracture allowed a subsequent determination of the forces acting on the implant during activities of daily living (Fig. 8-14) (Frankel et al., 1971). Although the device measured forces on the implant and not on the hip joint, it was possible to determine the proportion of the load transmitted through the device and to calculate the total load acting on the hip joint by means of static analysis. In the case illustrated in Figure 8-14, the nail plate transmitted one fourth of the total load.

Strong forces acting on the nail plate were encountered during such diverse activities as moving onto a bedpan, transferring to a wheelchair, and walking. The magnitude of the forces was greatly modified by skillful assistance from the nurse or therapist to control the patient's movement. Forces of up to four times body weight acted on the hip joint when the patient used the elbows and heels to elevate the hips while being placed on a bedpan (Fig. 8-15A), but these forces were greatly reduced through the use of a trapeze and assistance from an attendant (Fig. 8-15B). A 5-kg extension traction on the hip had little effect in modifying the forces acting on the hip joint. Exercises of the foot and ankle increased these forces.

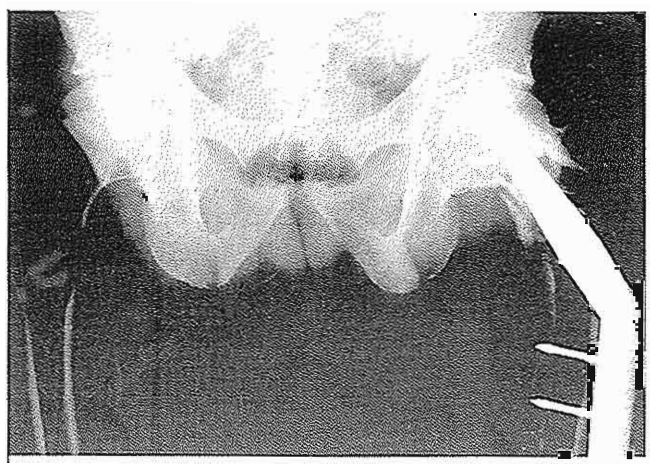


FIG. 8-14

An instrumented nail plate in the proximal end of the femur was used to determine the forces acting on the implant during the activities of daily living following fracture of the femoral neck. In this case, the nail plate was found to transmit one fourth of the total load on the hip joint.

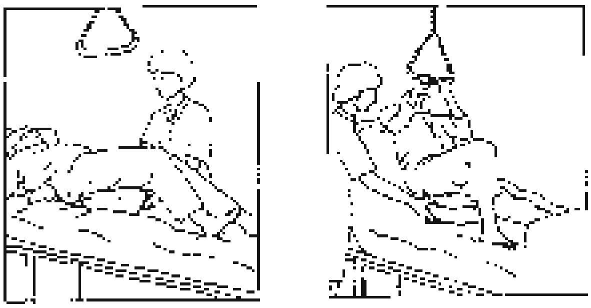


FIG. 8-15

A. When the patient used elbows and heels to elevate the hips while being placed on a bedpan, the force on the tip of the instrumented nail was 870 N. With a spica cast, the force on the tip of the nail was 190 N. B. The use of a trapeze and assistance from an attendant reduced the force to 190 N without a cast and to 70 N with a spica cast. Reprinted with permission from Frankel, V.H. (1971). Biomechanics of the Hip. In A.S. Tronzo (Ed.). Surgery of the Hip Joint (pp. 105–125). Philadelphia: Lea & Febiger.

Use of the instrumented nail plate demonstrated that, for a bedridden patient with a fractured femoral neck, the forces on the femoral head during activities of daily living approached those during walking with external supports. These studies support clinical protocols for early mobilization of patients and decreased bed rest for patients with hip fractures. The magnitude (approximately 8 Nm) of the moments acting on the nail-plate junction in the transverse plane (i.e., during internal and external rotation) was only approximately one half the magnitude (18 Nm) of the moments acting in the frontal plane (i.e., during abduction) for many activities.

EFFECT OF EXTERNAL SUPPORT ON THE HIP JOINT REACTION FORCE

Static analysis of the joint reaction force on the femoral head during walking with a cane demonstrates that the cane should be used on the side opposite the painful or operated hip. Neumann (1998)

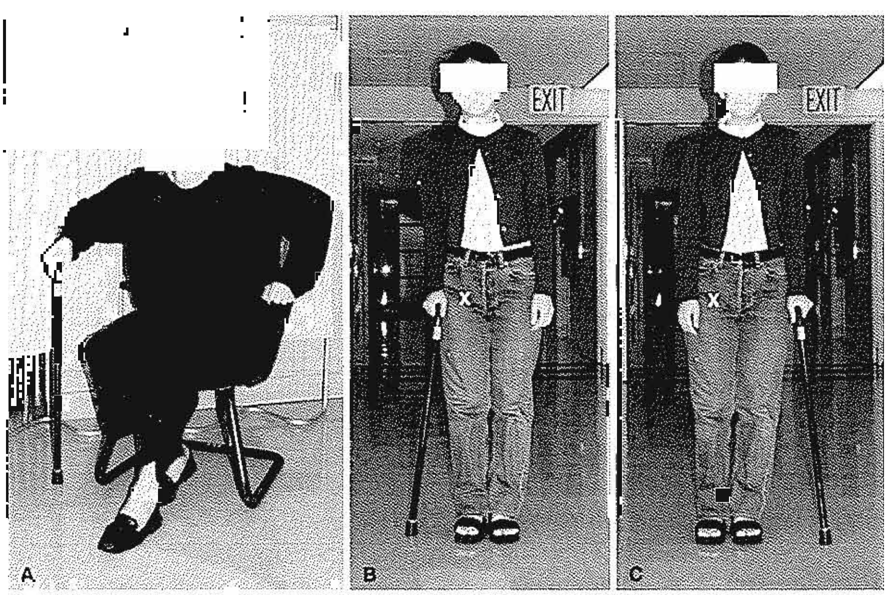


FIG. 8-16

High and low load on the hip joint during daily activities. Raising from a low chair produces approximately 8 times body weight (A). Walking with a cane on the ipsilateral side of the affected hip produces approximately 3.4 times body weight (B), and walking with a cane on the contralateral side of the affected hip reduces hip joint load substantially to 2.2 times body weight (C). This figure illustrates how load on the hip joint can be manipulated by simple means (X denotes affected hip).

CALCULATION BOX 8-4

Effect of External Support on the Hip Joint Reaction Force

Kinematic data were used to determine the joint reaction force acting on the femoral head in the late swing phase of the gait cycle for an 8-year-old boy weighing 24 kg and wearing a long-leg brace. The main muscle force was produced by contraction of the gluteus maximus muscle and identified through electromyography. The torque about the hip joint was calculated according to the formula

$$T = I\alpha$$

where

T is the torque expressed in newton meters (Nm)

I is the mass moment of inertia expressed in newton meters times seconds squared (Nm sec²)

α is the angular acceleration in late swing phase, expressed in radians per second squared (r/sec²).

In the case of the braced side, $I = I_L + I_B$, where

I_L is the mass moment of inertia of the leg

I_B is the mass moment of inertia of the brace

On the normal side,

$I = 0.45$ Nm sec²

$\alpha = 24$ r/sec²

On the braced side,

$I = 0.45$ Nm sec² + 0.35 Nm sec²

$\alpha = 24$ r/sec²

Thus,

$T = 0.45$ Nm sec² × 24 r/sec²

$T = 10.8$ Nm

Thus,

$T = (0.45$ Nm sec² + 0.35 Nm sec²$)$ × 24 r/sec²

$T = 19.2$ Nm

The extensor muscle force (E) was then found from the moment relationship

$$T = Ed$$

where E is the extensor muscle force and d is the perpendicular distance from the center of rotation of the femur to the midline of the gluteus maximus muscle. Distance d was measured from a roentgenogram and found to be 3.2 cm. From the equation $E = T/d$, the muscle force on the normal side was calculated to be 338 N and on the braced side, 600 N. The joint reaction force on the femoral head (J) is equal to the muscle force (E) minus the gravitational force produced by the weight of the limb (W_L). In this example, W_L was estimated to be 40 N.

On the normal side,

$J = E - W_L$

$J = 338 \text{ N} - 40 \text{ N}$

$J = 298$ N.

On the braced side,

$J = E - W_L$

$J = 600 \text{ N} - 40 \text{ N}$

$J = 560$ N

Thus, the joint reaction force on the femoral head in the braced limb was over 80% greater than the force on the non-braced limb, reaching more than two times body weight.

studied the effects of cane use in 24 subjects with a mean age of 63 years. During walking, the electromyographic activity of the hip abductor muscles was measured. Neumann found that use of a cane on the contralateral side of the affected hip joint, with careful instructions to use with near maximal effort, could reduce the muscle activity by 42% (Fig. 8-16). This calculates to a reduction of approximately one times body weight from 2.2 body weight with a cane, compared with 3.4 body weight without a cane. These studies give important information to the clinician about ways to moderate the load for the patient with hip problems.

Such use reduces the force on the femoral head of the painful joint without necessitating an analgesic body position. A cane used on the side of the painful hip works through a shorter lever arm and thus an even greater push on the cane is needed to decrease the joint reaction force. For the older patient, such a large push may not be possible because of a lack of strength in the upper extremities.

The use of a brace on the leg may alter the forces on the hip joint but may not always reduce the joint reaction force on the femoral head. An ischial long-leg brace used in the treatment of Perthes' disease raises the joint reaction force during late swing phase because the large mass moment of inertia of the brace results in a higher extensor muscle force during this part of the gait cycle (Calculation Box 8-4).

Summary

1 The hip joint is a ball-and-socket joint composed of the acetabulum and femoral head.

2 The thickness and mechanical properties of the cartilage on the femoral head and acetabulum vary from point to point.

3 Hip flexion of at least 120°, abduction of at least 20°, and external rotation of at least 20° are necessary for carrying out daily activities in a normal manner.

4 A joint reaction force of approximately three times body weight acts on the hip joint during a single leg stance with the pelvis in a neutral position; its magnitude varies as the position of the upper body changes.

5 The magnitude of the hip joint reaction force is influenced by the ratio of the abductor muscle force and gravitational force lever arms. A low ratio yields a greater joint reaction force than does a high ratio.

6 The hip joint reaction force during gait reaches levels of three to six times body weight or more in stance phase and is approximately equal to body weight during swing phase.

7 An increase in gait velocity increases the magnitude of the hip joint reaction force in both swing and stance phase.

8 The forces acting on an internal fixation device during the activities of daily living vary greatly depending on the nursing care and the therapeutic activities undertaken by the patient.

9 The use of a cane or a brace on the leg can alter the magnitude of the hip joint reaction force.

REFERENCES

Andriacchi, T.P., Andersson, G.B., Fermier, R.W., Stern, D., & Galante, J.O. (1980). A study of lower limb mechanics during stair-climbing. *J Bone Joint Surg, 62A*, 749.

Bergmann, G., Graichen, F., & Rohlmann, A. (1993). Hip joint loading during walking and running, measured in two patients. *J Biomech, 26*(8), 969–990.

Bergmann, G., Graichen, F., & Rohlmann, A. (1995). Is staircase walking a risk for the fixation of hip implants? *J Biomech, 28*(5), 535–553.

Carlsen, F., Hodge, A., Mann, R.W., et al. (1993). The role of muscular co-contraction of the hip during movement. *Clin Orthop Rel Res, 68*(1), [illegible]–235.

Crowninshield, R.D., Johnston, R.C., Brand, R.A. (1978). The effects of walking velocity and age on hip kinematics and kinetics. *Clin Orthop Rel Res, 132*, 140–144.

Delp, S.L., & Maloney, W. (1993). Effects of hip center location on the moment-generating capacity of the muscles. *J Biomech, 26*(5), 485–499.

Draganich, L.F., Andriacchi, T.P., Strongwater, A.M., et al. (1980). Electronic measurement of instantaneous foot-floor contact patterns during gait. *J Biomech, 13*, 875.

English, T.A., & Kilvington, M. (1979). In vivo records of hip loads using a femoral implant with telemetric output (a preliminary report). *J Biomed Eng, 1*(2), 111.

Fessy, M.H., N'Diaye, A., Carret, J.P., et al. (1999). Locating the center of rotation of the hip. *Surg Radiol Anat, 21*(4), 247–250.

Frankel, V.H. (1973). Biomechanics of the hip. In R.G. Tronzo (Ed.), *Surgery of the Hip Joint* (pp. 105–125). Philadelphia: Lea & Febiger.

Frankel, V.H. (1960). *The Femoral Neck: Function, Fracture Mechanism, Internal Fixation.* Springfield: Charles C Thomas.

Frankel, V.H., Burstein, A.H., Lygre, L., et al. (1971). The telltale nail. *J Bone Joint Surg, 53A*, 1232.

Free, S.A., & Delp, S.L. (1996). Trochanteric transfer in total hip replacement: Effects on the moment arms and force-generating capacities of the hip abductors. *J Orthop Res, 14*(2), 245–250.

Greenwald, A.S., & Haynes, D.W. (1972). Weight-bearing areas in the human hip joint. *J Bone Joint Surg, 54B*(1), 157–163.

Hurwitz, D.E., & Andriacchi, T.P. (1998). Biomechanics of the hip. In J.J. Callaghan, A.G. Rosenberg, & H.E. Rubash (Eds.), *The Adult Hip* (pp. 75–85). Philadelphia: Lippincott-Raven Publishers.

Hurwitz, D.E., & Andriacchi, T.P. (1997). Biomechanics of the hip and the knee. In M. Nordin, G.B.J. Andersson, & M.H. Pope (Eds.), *Musculoskeletal Disorders in the Workplace: Principles and Practice* (pp. 486–496). Philadelphia: Mosby-Year Book.

Inman, V.T. (1947). Functional aspects of the abductor muscles of the hip. *J Bone Joint Surg, 29A*, 607.

Johnston, R.C., Brand, R.A., & Crowninshield, R.D. (1979). Reconstruction of the hip. *J Bone Joint Surg, 61A*(5), 639–652.

Johnston, R.C., & Smidt, G.L. (1969). Measurement of hip-joint motion during walking: Evaluation of an electrogoniometric method. *J Bone Joint Surg, 51A*, 1083.

Johnston, R.C., & Smidt, G.L. (1970). Hip motion measurements for selected activities of daily living. *Clin Orthop, 72*, 205.

Kempson, G.E., Spivey, C.J., Swanson, S.A.V., et al. (1971). Patterns of cartilage stiffness on normal and degenerate human femoral heads. *J Biomech, 4*, 597.

Kim, Y.T., & Azuma, H. (1995). The nerve endings of the acetabular labrum. *Clin Orthop, 320*, 176–181.

Konrath, G.A., Hamel, A.J., Olson, S.A., et al. (1998). The role of the acetabular labrum and the transverse acetabular ligament in load transmission of the hip. *J Bone Joint Surg, 80A*(12), 1781–1788.

Kotzar, G.M., Davy, D.T., Goldberg, V.M., et al. (1991). Telemeterized in vivo hip joint force data: A report on two patients after total hip surgery. *J Orthop Res, 9*, 621–633.

Kumagai, M., Shiba, N., Higuchi, F., et al. (1997). Functional evaluation of hip abductor muscles with use of magnetic resonance imaging. *J Orthop Res, 15*(6), 888–893.

Lim, L.A., Carmichael, S.W., & Cabanela, M.E. (1999). Biomechanics of total hip arthroplasty. *Anat Rec, 257*(3), 110–116.

McLeish, R.D., & Charnley, J. (1970). Abduction forces in the one-legged stance. *J Biomech, 3*, 191.

McMinn, R.H., & Hutchings, R.T. (1988). In *Color Atlas of Human Anatomy* (2nd ed., p. 302). Chicago: Year Book Medical Publishers, Inc.

Murray, M.P. (1967). Gait as a total pattern of movement. *Am J Phys Med, 46*, 290.

Murray, M.P., Kory, R.C., & Clarkson, B.H. (1969). Walking patterns in healthy old men. *J Gerontol, 24,* 169–178.

Németh, G., & Ohlsén, H. (1985). In vivo moment arm lengths for hip extensor muscles at different angles of hip flexion. *J Biomech, 18,* 129–140.

Németh, G., & Ohlsén, H. (1989). Moment arms of hip abductor and adductor muscles measured in vivo by computed tomography. *Clin Biomech, 4,* 133–136.

Neumann, D.A. (1998). Hip abductor muscle activity as subjects with hip prostheses walk with different methods of using a cane. *Phys Ther, 78*(5), 490–501.

Rohrle, H., Scholten, R., Sigolotto, C., et al. (1984). Joint forces in the human pelvis-leg skeleton during walking. *J Biomech, 17,* 409–424.

Rushfeldt, P.D., Mann, R.W., & Harris, W.H. (1979). Influence of cartilage geometry on the pressure distribution in the human hip joint. *Science, 204*(4391), 413–415.

Rydell, N.W. (1966). Forces acting on the femoral head prosthesis: A study on strain gauge supplied prostheses in living persons. *Acta Orthop Scand Suppl 88,* 1–132.

Rydell, N. (1965). Forces in the hip joint. Part II: Intravital measurements. In R.M. Kenedi (Ed.), *Biomechanics and Related Bio-Engineering Topics* (pp. 351–357). Oxford: Pergamon Press.

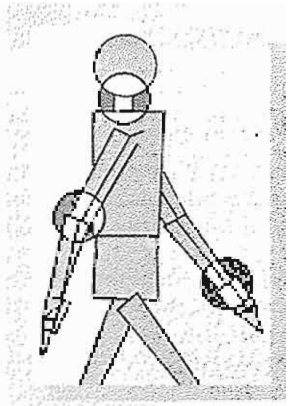
Sutherland, A.G., D'Arcy, S., Smart, D., et al. (2000). Abductor weakness and stress around acetabular components of total hip arthroplasty: A finite element analysis. *Int Orthop, 24*(5), 275–278.

University of California (1953). The pattern of muscular activity in the lower extremity during walking. *Prosthet Devices Res Rep, 2*(25), 1–41.

van den Bogert, A.J., Read, L., & Nigg, B.M. (1999). An analysis of hip joint loading during walking, running and skiing. *Med Sci Sports Exerc, 31*(1), 131–142.

Vasavada, A.N., Delp, S.L., Maloney, W.J., et al. (1994). Compensating for changes in muscle length in total hip arthroplasty. Effects on the moment generating capacity of the muscles. *Clin Orthop, 302,* 121–133.

Von Eisenhart-Rothe, R., Eckstein, F., Muller-Gerbl, M., et al. (1997). Direct comparison of contact areas, contact stress and subchondral mineralization in human hip joint specimens. *Anat Embryol (Berl), 195*(3), 279–288.



Biomechanics of the Foot and Ankle

G. James Sammarco, Ross Todd Hockenbury

Introduction

The biomechanics of the foot and ankle are complex and intricately associated with each other. The foot is an integral mechanical part of the lower extremity necessary for a smooth and stable gait. The ankle transfers load from the lower extremity to the foot and closely influences foot orientation with the ground.

The foot is composed of 28 bones (including sesamoids) whose motions are closely interrelated (Fig. 9-1). Besides acting as a structural supporting platform capable of withstanding repetitive loads of multiples of body weight, the foot/ankle complex also must be able to adjust to different ground surfaces and varying speeds of locomotion. The unique qualities of the foot allow it to be rigid when necessary, as in ballet dancing on point, or quite flexible, as in walking barefoot on the sand. The transition from shock-absorbing platform to rigid lever capable of forward propulsion occurs with each step of the gait cycle.

The ankle is composed of three bones that form the ankle mortise. This joint complex consists of the tibiotalar, fibulotalar, and tibiofibular joints (Fig. 9-2). The ankle is a hinge joint whose stability depends on joint congruency and the medial, lateral, and syndesmotic ligaments.

This chapter discusses the motions that occur in the foot and ankle during the various phases of gait as well as during extremes of motion. The close interplay between lower extremity rotation and forefoot orientation is explained. The ground (foot-to-floor) reaction force and distribution of forces on the plantar aspect of the foot are explored. The location of forces as they pass from the tibiofibular complex into the dome of the talus and then into the foot is discussed. We also discuss the roles of ligaments and muscles in the support of the medial longitudinal arch. Finally, ankle motion and ligamentous stability is outlined.

A discussion of sophisticated electromyographic activity during walking is not within the scope of this chapter; however, the activity of certain extrinsic and intrinsic muscles is by necessity presented to allow a better understanding of foot and ankle control during gait. The moments produced about joints by muscle action and resultant effects on foot and ankle position are detailed. Joint axes and instant centers of joint motion are described. Refer to Chapter 18 for information about the application of biomechanics to gait.

Any pathological change in foot or ankle structure or motion, however subtle, may have a profound impact on the foot and ankle's shock-absorbing, propulsive, and stabilizing roles. Clinical correlation of alterations in biomechanical function is presented in case studies. Footwear in Western society may vary from a rigid ski boot to a soft moccasin. These externally restrictive materials may alter normal foot and ankle biomechanics and ultimately influence the development of some pathological conditions, such as hallux valgus.

Growth of the Foot

The foot is formed when the limb buds develop during the eighth week of gestation. Foot length and width increases linearly from age 3 to 12 in girls and age 3 to 15 in boys at an average of 8 to 10 mm per year, followed by a plateau in growth (Cheng et al., 1997). Blais and associates (1956) showed that the foot appears to be closer to the adult size at all times during normal development of the child than are other parts of the limb. On average, at age 1 year in girls and 18 months in boys, the length of the foot is one half the length of the respective adult foot (Fig. 9-3). This situation contrasts with that in the femur and tibia, which do not attain their mature length until 3 years later in both boys and girls. The relatively large size of the foot, then, is important for providing a broad base on which the child's body is supported, and this base may at times compensate for the child's lack of muscle strength and coordination.

Kinematics of the Foot

Gross motion of the foot is complex and occurs around three axes and in three planes (Fig. 9-4). Flexion-extension occurs in the sagittal plane, abduction-adduction occurs in the horizontal or transverse plane, and inversion-eversion occurs in the coronal or frontal plane. Supination and pronation are terms commonly used to describe positioning of the plantar surface of the foot and occur primarily at the subtalar (talocalcaneal) joint. During supination the sole faces medially, and during pronation the sole faces laterally. Supination is a combination of inversion, flexion, and adduction. Pronation is a combination of eversion, extension, and abduction (Fig. 9-5). Toe motion includes flexion, extension, adduction, and abduction.

For practical purposes, foot motion can be considered to be of two distinct types: non-weight

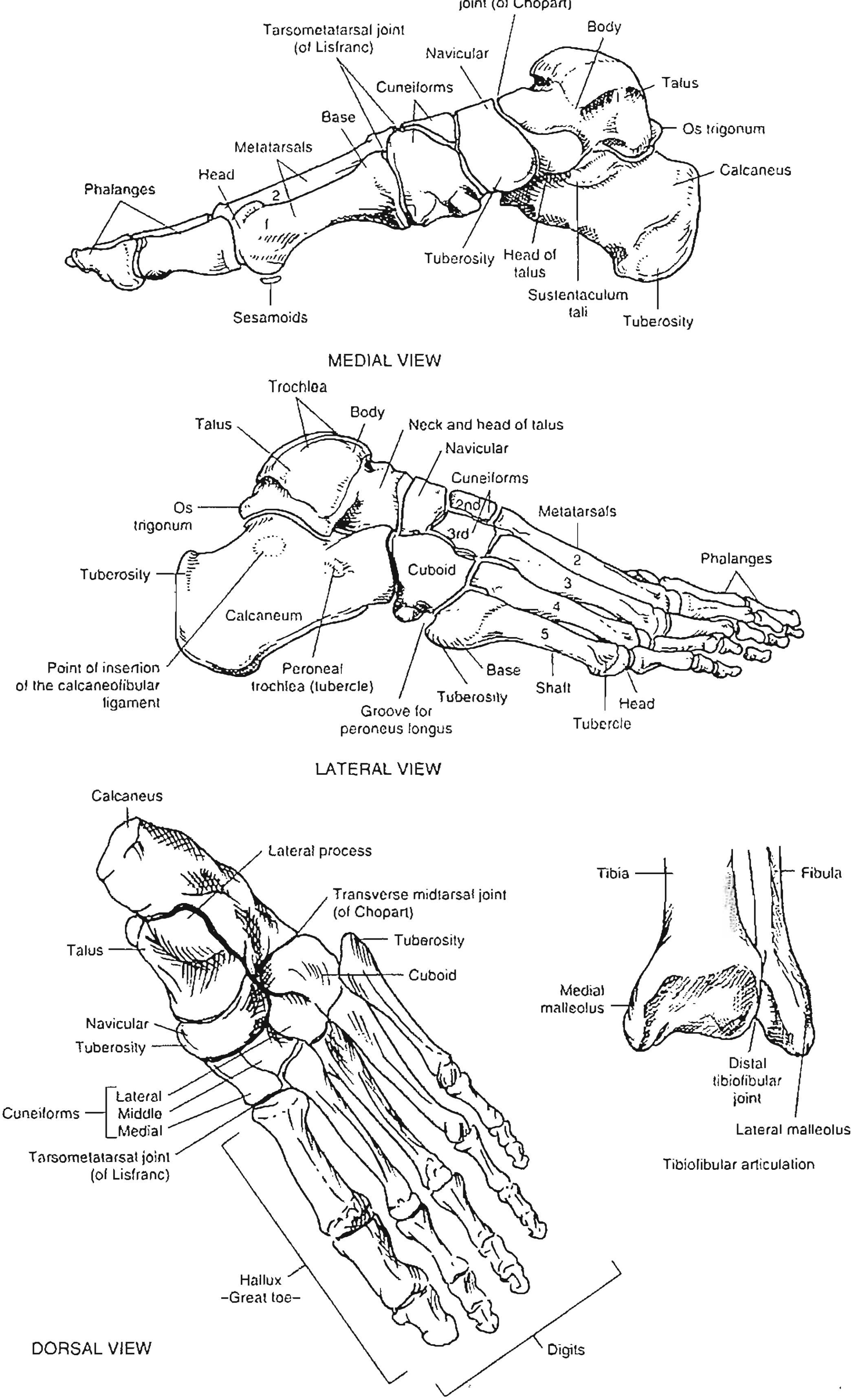


FIG. 9-1

Top, View of the medial aspect of the foot. *Middle*, View of the lateral aspect of the foot. *Bottom left*, Superior view of the foot. *Bottom right*, Anterior view of the ankle mortise.