J. ANDREW McCAMMON & STEPHEN C. HARVEY

Dynamics of proteins and nucleic acids

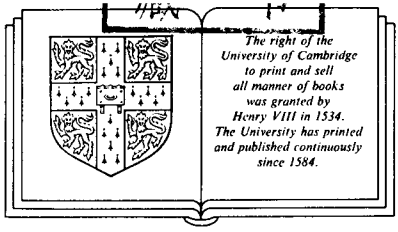
Dynamics of proteins and nucleic acids

J. ANDREW McCAMMON

M. D. Anderson Professor of Chemistry
University of Houston – University Park
Houston, Texas

STEPHEN C. HARVEY

Professor of Biochemistry
University of Alabama at Birmingham
Birmingham, Alabama



CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE

LONDON NEW YORK NEW ROCHELLE

MELBOURNE SYDNEY

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1987

Printed in Great Britain by the University Press, Cambridge

British Library cataloguing in publication data
McCammon, J. Andrew
Dynamics of proteins and nucleic acids.
1. Proteins 2. Molecular dynamics
3. Nucleic acids 4. Molecular dynamics
I. Title II. Harvey, Stephen C.
547.7'5045413 QD431

Library of Congress cataloguing in publication data
McCammon, J. Andrew
Dynamics of proteins and nucleic acids.
Bibliography
Includes index.
1. Proteins. 2. Nucleic acids. 3. Molecular
dynamics. I. Harvey, Stephen C. II. Title.
QD431.M4245 1987 574.19'245 86-17576

ISBN 0 521 30750 3

UP

Dedicated to
Anne W. McCammon and Marie A. Weaver

PREFACE

At macroscopic levels, the dynamic character of life is dramatically self-evident. Motion is no less important at the molecular level of biology. Indeed, the marked biochemical effects of temperature changes imply that the activity of biological molecules reflects their thermal mobility. An appreciation of molecular flexibility and dynamics is essential to the understanding of the activity of naturally occurring molecules and to the design of new molecules with specified activities.

Detailed studies of the atomic motion of proteins and nucleic acids are of recent origin. Nevertheless, far more has already been done than can be adequately described in a single volume. The aim of this book is accordingly modest. We attempt to provide the reader with a self-contained introduction to the theoretical aspects of protein and nucleic acid dynamics. The level of presentation is intended to be appropriate for graduate students as well as for research workers in biophysics, physical biochemistry, and molecular biotechnology. Our principal goals are (1) to outline the theoretical methods and their capabilities, (2) to provide a sense of the nature and biological significance of biomolecular dynamics by reference to representative theoretical studies, and (3) to indicate some prospects and directions for future work. Experimental work is covered incidentally in connection with theoretical results.

The book is organized generally to progress from fundamentals to applications and from short time scales to the longer time scales characteristic of most biological activity. Proteins and nucleic acids are treated in an integrated fashion, but mostly in separate sections that can be read selectively if the reader wishes. The first four chapters provide an introduction to conceptual foundations and methodology. Chapters five through eight present the results of selected applications. There, we attempt to describe the nature of the different types of molecular motion

that are found to occur in biological systems. The final chapter addresses current research and future prospects. This chapter shows that biomolecular dynamics is entering an exciting new phase, one that is concerned with the interpretation and prediction of biological activity as much as with physical properties. The fruits of this work will include useful tools for pharmacology, medicine and industry.

We hope that this book will give the reader a sense of the special challenges and rewards associated with theoretical studies of protein and nucleic acid dynamics. One of the challenges derives from the fact that these studies involve the fusion of three 'high technology' areas, namely, molecular biology, chemical physics and scientific computing. The rapid pace of development of each of these areas leads to more than the usual rate of obsolescence of research techniques. Such difficulties are, however, more than offset by the promise of aesthetic and pragmatic rewards. Among the former is the pleasure of bridging different scientific disciplines, e.g., using Newton's equations of motion to interpret basic events in biochemistry. The pragmatic rewards include the potential applicability of theoretical methods in the design of new drugs, enzymes, vaccines, etc.

The present book grew out of a review article in *Reports on Progress in Physics* (McCammon, 1984). We are grateful to the Institute of Physics for permission to use material from that article here. Some of the material on nucleic acids is drawn from another recent review article (Harvey, 1986). We also wish to acknowledge our coworkers and colleagues, who have made invaluable contributions to our own understanding of the work described herein. Several colleagues read drafts of the manuscript and made helpful suggestions; particularly valuable comments were provided by Professors P. A. Kollman and B. M. Pettitt, and by Dr T. P. Lybrand, Dr M. Prabhakaran and Dr L. J. Ransom-Wright. Special thanks are due to Denise Marshall for her skillful assistance in word processing. The authors' research in protein and nucleic acid dynamics has been supported in part by the National Science Foundation, the National Institutes of Health, the Robert A. Welch Foundation, and the Texas Advanced Technology Research Program.

Houston and Birmingham J. A. MCC.
January 1987 S. C. H.

CONTENTS

1

Introduction

1.1 Function of proteins and nucleic acids

Proteins and nucleic acids are particularly prominent among the molecules essential to life. Their importance stems from the remarkable diversity of their functional roles. This diversity can be illustrated by listing a few of the major groups within each of these molecular families. Proteins are molecules that act to build the structural elements of organisms and to provide the energy necessary for life processes. Enzymes are proteins that catalyze biochemical reactions. Familiar examples include the digestive enzymes that degrade foodstuffs to simple, assimilable compounds; the biosynthetic enzymes that build complex molecules from simpler compounds; and muscle proteins that produce mechanical work from chemical reactions. Transport proteins such as hemoglobin facilitate the movement of molecular oxygen and other essential compounds to their sites of utilization. Antibodies are proteins that bind to and neutralize foreign materials that may be harmful to an organism. Other proteins are responsible for maintaining the structures of cells, organs, and organisms, while still others play essential roles in genetic expression, nerve conduction, and all other biological processes. Nucleic acids are the molecules that carry the information necessary for protein synthesis; they can be considered the 'blueprints' that contain the design of the living organism. In both procaryotes and eucaryotes, the genetic information of heredity is carried from one generation to the next in DNA, while various types of RNA's play vital roles in the translation of the DNA sequence of each gene into the amino acid sequence of the corresponding protein. The regulation of the expression of different genes, which is vital to the control of development, growth, repair, and reproduction, involves a wide range of interactions between proteins and nucleic acids.

The functional diversity of proteins and nucleic acids ultimately reflects

the large amount of information that is stored in these molecules and expressed in their interactions with other molecules in the course of biological activity. The information content and activity of a protein or nucleic acid can be varied over a large range by modifying its molecular structure. The diversity of naturally occurring molecules represents the cumulative result of structural modifications that have occurred during evolution. The rapid development of molecular biotechnology largely reflects recent discoveries of procedures for preparing systematically modified proteins and nucleic acids in the laboratory.

As suggested by the examples given in the first paragraph, proteins and nucleic acids are largely responsible for the expression and transmission of biological information, respectively, although this division is not a sharp one. As molecular 'machines', an important characteristic of proteins is their specificity of function. A particular enzyme will bind a specific substrate molecule and catalyze a specific chemical transformation of the substrate. A particular antibody molecule will bind specific antigens. For many proteins, the specificity of action is so narrowly defined that a small change in a ligand molecule that binds strongly to the protein (e.g., replacement of a hydrogen atom by a methyl group) leads to a dramatic reduction in binding. The activities of a number of proteins are regulated by interactions with other molecules. For example, their primary activity may be increased or decreased by the binding of specific auxiliary 'effector' ligands. Together with the spatial ordering of proteins imposed by the anatomy of an organism, the specificity and regulability of protein function are largely responsible for the required coherence of biochemical processes. Similarly, nucleic acids have highly specific interactions, both with one another and with other types of molecules, and these interactions are crucial for the control of many aspects of replication, transcription, translation, and recombination. Examples include the specificity of the recognition of the codon on messenger RNA (mRNA) by the anticodon on the cognate transfer RNA (tRNA), site-specific initiation and termination by RNA polymerase, and the remarkable specificity of the restriction endonucleases.

1.2 Structure and dynamics

Given the functional richness of proteins and nucleic acids, one would expect to observe a corresponding complexity in the detailed structure of these molecules. This expectation has been confirmed by X-ray diffraction studies, which have provided the crystal structures of more than 100 proteins and nucleic acids during the past 25 years (Bernstein *et al.*, 1977; Richardson, 1981; Dickerson *et al.*, 1982).

Proteins are very large molecules; their molecular weights are often in the tens of thousands. The basic component of these molecules is the polypeptide chain, an unbranched polymer consisting of a sequence of amino acid residues. There are 20 commonly occurring amino acids, and a typical chain will contain a few hundred of these elementary structural units. Protein molecules consist of one or a small number of such polypeptide chains, complemented in some cases by one or more prosthetic groups (e.g., metal ions or special organic molecules). For a given protein, the polypeptide chain of each molecule is folded compactly into a characteristic three dimensional structure. Although the resulting structures are complicated, it is commonly observed that the packing density of the protein components is nearly maximized, subject to the requirement that those amino acid residues which have a favorable free energy of interaction with water tend to remain near the protein surface. In many cases, it has been possible to carry out X-ray diffraction studies of globular proteins with bound ligands (e.g., substrate analogs). These studies show that the folding of a protein is such that key amino acids with chemically active groups are strategically located in a well-defined 'active site', where the groups can interact in a coordinated fashion with the ligand. Such studies have been invaluable in the development of structural interpretations of protein function.

Nucleic acids are linear polymers whose monomeric units are nucleotides. The size of these molecules covers several orders of magnitude, from tRNA's (with roughly 75 nucleotides, molecular weights around 25000 and end-to-end lengths of less than 10 nm) to eucaryotic DNA's. For the latter, the single DNA molecule of a large chromosome may contain billions of nucleotides and have a molecular weight of more than 10^{12}. If extended, such a molecule would be several centimeters in length. Although tRNA's do have relatively compact structures, large DNA's have extended, wormlike coil configurations at physiological temperature, pH, and ionic strength. Their folding into stable, compact structures *in vivo* is determined by their association with a variety of proteins and, for closed circular DNA, by supercoiling. X-ray diffraction studies on DNA fibers revealed the now classic right-hand double helical structures of A- and B-DNA (Watson & Crick, 1953; Arnott, Smith & Chandrasekaran, 1976). Crystallographic studies have uncovered subtle sequence-dependent variations about these ideal average structures (Fratini, Kopka, Drew & Dickerson, 1982; Shakked *et al.*, 1983), and they have also revealed that alternating purine–pyrimidine sequences can, under some conditions, form left-handed double helices (Wang *et al.*, 1979). A variety of nucleic acid structures other than simple double helices do exist, some transient and

others being quite stable. These include single strands, bulged loops, hairpin loops, cruciforms, catenanes, knots, and branches in the double helical structure.

During the past ten years, increasing attention has been focused on the dynamic aspects of protein and nucleic acid structure and function. It has long been inferred from a variety of experimental studies that substantial structural fluctuations occur in these molecules, and that these fluctuations are essential to biological activity (Linderstrom-Lang & Schellman, 1959; Koshland, 1963; Edsall, 1968). Until recently, the exact nature of the structural fluctuations has proved elusive. The recent surge of interest in biomolecular dynamics has largely been stimulated by theoretical studies that have provided a detailed picture of the atomic motion in proteins and, more recently, nucleic acids. These theoretical studies are the primary subject of the present book. The theoretical work involves a combination of methods from theoretical chemical physics and biomolecular structure theory. The methods from chemical physics include techniques that have previously been used successfully to study the structure and atomic motion in dense materials such as liquids and solids. These methods are appropriate in view of the high density and large size of proteins and nucleic acids. Along with the theoretical developments, new experimental techniques that provide detailed insights to biomolecular dynamics have become available. Indeed, the present robustness of this field is largely a result of the interplay of modern theoretical and experimental work. Theory has successfully predicted a number of fundamental properties such as the average magnitude of atomic thermal displacements, the variation of these magnitudes throughout a molecule, and the time scales of certain displacements. Recent experiments have presented new challenges that are stimulating further theoretical work. The results achieved during the past few years and the history of corresponding efforts for systems such as liquids both suggest that the theoretical work on proteins and nucleic acids will become increasingly sophisticated and useful in the coming years.

1.3 Scope of this book

The number of publications on dynamic aspects of biomolecular structure and function is growing at an extraordinary rate. As stated in the preface, this book is not intended to provide an all-inclusive catalogue of this activity. It is, rather, intended to provide a reasonably self-contained introduction to the theoretical foundations of the subject, and to highlight a representative selection of important theoretical results within an integrated framework. The reader may wish to consult recent review articles for additional material (Careri, Fasella & Gratton, 1979;

McCammon & Karplus, 1980*a*, 1983; Karplus & McCammon, 1981*a*, 1983; Levitt, 1982; van Gunsteren & Berendsen, 1982; Cooper, 1984; Edholm *et al.*, 1984; McCammon, 1984; Berg & von Hippel, 1985; Allison, Northrup & McCammon, 1986; Friesner & Levy, 1986; Harvey, 1986; Karplus & McCammon, 1986; Levy, 1986; Levy & Keepers, 1986; McCammon, Northrup & Allison, 1986*b*; Pettitt & Karplus, 1986). A number of experimental results are also described to illustrate the types of data available and the degree of overlap with theoretical findings. Again, excellent reviews that focus on experimental work have recently been published (Gurd & Rothgeb, 1979; Peticolas, 1979; Woodward & Hilton, 1979; Williams, 1980; Jardetzky, 1981; Karplus & McCammon, 1981*a*; Phillips, 1981; Debrunner & Frauenfelder, 1982; Hilinski & Rentzepis, 1983; Huber & Bennett, 1983; Janin & Wodak, 1983; Rigler & Wintermeyer, 1983; Shank, 1983; Wagner, 1983; Bennett & Huber, 1984; Cooper, 1984; Edholm *et al.*, 1984; Englander & Kallenbach, 1984; Middendorf, 1984; Petsko & Ringe, 1984; Torchia, 1984; Friedman, 1985*b*; Ringe & Petsko, 1985; Terner & El-Sayed, 1985).

2

Structure of proteins, nucleic acids, and their solvent surroundings

Before considering the dynamics of proteins and nucleic acids, it is necessary to review some of the structural and energetic properties of these molecules and their solvent surroundings. In the following sections, we sketch some important structural characteristics in simple physical terms. We also describe the interatomic forces that govern molecular structure and flexibility. This general discussion is far from complete. Some additional details are reviewed as necessary in discussing specific dynamic processes in later chapters. Fortunately, many comprehensive reviews of these structural topics are available; references to some of these are given in the following sections.

In considering structure at the level of atomic detail, it is essential to keep in mind the time scale of observation. In macromolecules, there will be subtle differences between any particular instantaneous structure (observed on a time scale that is shorter than the period of vibration of bond lengths) and the average structure that is seen by an X-ray diffraction study which observes average positions over many hours. This is an even more important issue in liquids. An instantaneous structure of liquid water, such as may be obtained from a Monte Carlo or molecular dynamics simulation, will be characterized by well-defined atomic positions. Discussions about this structure in terms of the extent of hydrogen bonding and the similarities to and differences from the structure of ice are possible. Experimentally, however, most methods look at properties averaged over times that are much longer than the characteristic times of rotational and translational diffusion. In this case, only average properties, such as bulk thermodynamic properties, can be determined precisely. In favorable cases, experiment can provide information on the time scales and nature of particular motions, distinguishing, for example, differences

in diffusive behavior of water molecules in the bulk solvent and those in hydration layers near macromolecules.

2.1 Water and aqueous solutions

In liquid water, as in other dense molecular systems, an important structural determinant is the size and shape of the individual molecules. When two water molecules approach quite closely, a short range but strongly repulsive force arises from the overlap of their electron clouds. To a good approximation, each water molecule can be thought to have a repulsive, spherical core centered at the oxygen nucleus. These cores do not allow the oxygen nuclei of two water molecules to approach more closely than about 0.24 nm. In simple liquids, such repulsive cores are essentially the only structural determinant. For example, the instantaneous structure of liquid ethane is closely similar to a random dense packing of pairs of overlapping hard spheres, where the overlapping spheres correspond to bonded methyl groups.

Water is not a simple liquid, however. Water molecules have directional attractive interactions that are strong enough to compete with the repulsion of the molecular cores. The most stable molecular configurations still avoid overlapping cores, but have an expanded structure that optimizes the attractions. These attractive forces are of electrostatic origin. Particularly important are hydrogen bonding interactions. When a hydrogen atom is covalently bonded to oxygen, nitrogen, or certain other electronegative atoms, the bonding electron density is partly shifted onto the heavier atom. The hydrogen is left with a significant partial positive charge, and can approach other atoms relatively closely because of the drawn-in character of its electron cloud. Such a hydrogen therefore has a relatively strong electrostatic attraction for oxygen or nitrogen atoms with partial negative charges. These attractive interactions are termed hydrogen bonds; the molecule or group with the hydrogen is referred to as a hydrogen bond donor and the partner molecule or group is referred to as a hydrogen bond acceptor. Hydrogen bonding interactions play an important role in shaping the structure of aqueous systems because these interactions operate only within narrow geometric ranges. The donor and acceptor must be quite close, and the hydrogen can not be too far off the line between the electronegative atoms that it links. For a pair of water molecules with a linear hydrogen bond, the maximum stabilization is about 20 kJ/mol when the oxygens are separated by 0.28 nm; the interaction approaches zero for oxygen separations greater than 0.4 nm.

The electronic configuration of the oxygen atom in H_2O has sp^3

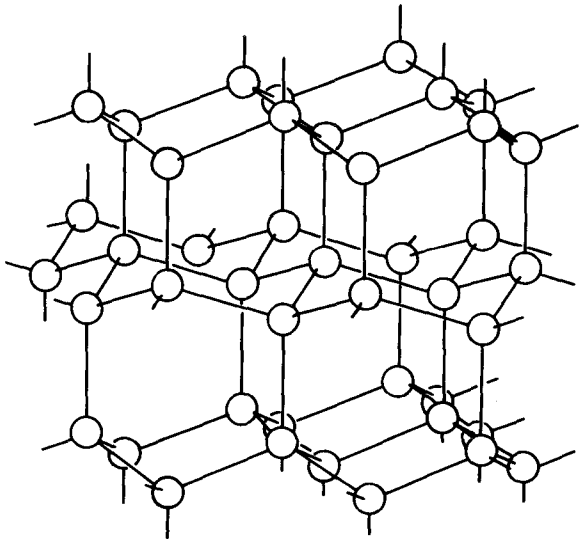


Fig. 2.1. Schematic illustration of the structure of normal ice, I_h. The circles represent oxygen atoms. The solid lines represent hydrogen bonds. In liquid water, thermal energy disrupts the regular lattice structure shown here.

hybridization, with a near tetrahedral geometry that has hydrogen atoms at two of the vertices of the tetrahedron and lone-pair electrons at the other two vertices. Consequently, a water molecule can simultaneously act as a hydrogen bond donor to two other water molecules and an acceptor from two more water molecules. This particularly stable arrangement, in which the central molecule has an approximately tetrahedral set of hydrogen bonds, is replicated in the three dimensional structure of ordinary ice, ice I_h (figure 2.1). In liquid water at 300 K, the average translational and rotational kinetic energy of a water molecule is about 7 kJ/mol, or only about one-third the energy required to break a hydrogen bond. Thus, liquid water retains significant vestiges of the ice I_h structure. Tetrahedral coordination is prominent on a local scale, but broken and defective hydrogen bonds occur frequently enough to destroy the large scale order and rigidity of the crystalline state.

The principal structural features of aqueous solutions of nonpolar molecules can be understood by reference to the local structure of water. When small nonpolar molecules are dissolved in water, the solvent will distribute itself around the solute so as to minimize the breaking of hydrogen bonds. For solutes such as methane, the water molecules in the first solvation shell retain their tetrahedral hydrogen bonding pattern by