STRUCTURE AND STABILITY OF BIOLOGICAL MACROMOLECULES

Edited by

Serge N. Timasheff and Gerald D. Fasman

Graduate Department of Biochemistry
Brandeis University
Waltham, Massachusetts

1969

MARCEL DEKKER, INC., New York

MARCEL DEKKER, INC.
95 Madison Avenue, New York, New York 10016

LIBRARY OF CONGRESS CATALOG CARD NUMBER 70-76084

PRINTED IN THE UNITED STATES OF AMERICA

STRUCTURE AND STABILITY OF BIOLOGICAL MACROMOLEC

BIOLOGICAL MACROMOLECULES

A Series of Monographs

SERIES EDITORS

SERGE N. TIMASHEFF and GERALD D. FASMAN

Graduate Department of Biochemistry
Brandeis University
Waltham, Massachusetts

VOLUME 1

Poly-α-Amino Acids: Protein Models for Conformational Studies
edited by Gerald D. Fasman

VOLUME 2

Structure and Stability of Biological Macromolecules
edited by Serge N. Timasheff and Gerald D. Fasman

Other Volumes in Preparation

Fine Structure of Proteins and Nucleic Acids
edited by Gerald D. Fasman and Serge N. Timasheff

Subunits in Biological Systems
edited by Serge N. Timasheff and Gerald D. Fasman

INTRODUCTION TO THE SERIES

In any period marked by an explosion of scientific information, there is a tendency toward compartmentalization of knowledge into increasingly narrower and more specialized subdivisions. As the understanding of biological phenomena increases, there is an opposing tendency to cross these artificial demarcations between various disciplines and to blend them into a continuous spectrum. This movement is frequently accompanied by the realization, or "discovery," that what had been regarded as highly autonomous areas of knowledge are actually complementary and that phenomena previously considered independent are really manifestations of a general theory. The vast research effort on macromolecules, however, has led to its subdivision into many separate fields—proteins, nucleic acids, polysaccharides, and so forth. Recently, it has become evident that these are truly interdisciplinary fields with much to be learned from each other. It is with this realization in mind that we have undertaken the publication of the "Biological Macromolecules Series," devoted to structural and conformational aspects and their relations to function.

The purpose of this series is to bridge the artificial barriers between various disciplines dealing with many types of macromolecules. We propose in each volume to discuss a general problem of importance relating to all high molecular weight polymers of biological origin, or, conversely, to examine in depth a single type of molecule, utilizing a variety of approaches and showing how these complement each other. In doing this, we will draw on the expertise of colleagues, in various fields, to correlate and elucidate the common principles that form the basis of these various disciplines. The intention of this series is not the publication of a new set of annual advances or progress publications, but the periodic presentation of critical evaluations of various aspects of a general field. We hope that these analyses will not only help to unify present knowledge, but also stress that which is controversial.

Gerald D. Fasman
Serge N. Timasheff

CONTRIBUTORS TO THIS VOLUME

Rodney Biltonen, *Department of Physiological Chemistry, The Johns Hopkins University, School of Medicine, Baltimore, Maryland*

John F. Brandts, *Chemistry Department, University of Massachusetts, Amherst, Massachusetts*

Howard DeVoe,† *Section on Physical Chemistry, Laboratory of Neurochemistry, National Institute of Mental Health, Bethesda, Maryland*

Rufus Lumry, *Laboratory for Biophysical Chemistry, Chemistry Department, University of Minnesota, Minneapolis, Minnesota*

J. A. Rupley, *Department of Chemistry, University of Arizona, Tucson, Arizona*

Thomas Schleich,‡ *Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire*

A. L. Stone, *National Institute of Mental Health, N.I.H., Bethesda, Maryland*

H. Susi, *Eastern Regional Research Laboratory, U.S. Department of Agriculture, Philadelphia, Pennsylvania*

Peter H. von Hippel,‡ *Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire*

† Present address: Chemistry Department, University of Maryland, College Park, Maryland.

‡ Present address: Department of Chemistry and Institute of Molecular Biology, University of Oregon, Eugene, Oregon.

CONTENTS

CHAPTER 1

THEORY OF THE CONFORMATIONS OF BIOLOGICAL MACROMOLECULES IN SOLUTION

Howard DeVoe†

SECTION ON PHYSICAL CHEMISTRY
LABORATORY OF NEUROCHEMISTRY
NATIONAL INSTITUTE OF MENTAL HEALTH
BETHESDA, MARYLAND

† Present address: Chemistry Department, University of Maryland, College Park, Maryland.

I. INTRODUCTION

A wealth of experimental data exists about the effects of environmental conditions (temperature, solvent composition, ionic strength, and pH) on the conformations of many biological macromolecules. This chapter is concerned with theoretical principles for understanding and predicting these environmental effects.

The conformation of a macromolecule is its detailed three-dimensional structure, as specified by the positions of its atoms or by the values of its bond lengths and bond angles. There are, of course, restrictions imposed on the possible structures by the covalent bonding, by the normal ranges of covalent bond lengths and angles, and by the steric restraints of the atomic van der Waals radii. Actually a conformation or a conformational state refers to a range of possible structures specified, for instance, by limits for the value of each bond length and angle. These ranges are quite narrow for an almost rigid conformation and broad for disordered conformations. The infinite number of possible absolutely rigid structures (configurations) of a macromolecule are thus grouped in an arbitrary way into a finite number of discrete conformations.

The number and definitions of the conformations of a given biological macromolecule are arbitrary. The conformations may be chosen for convenience of definition or to best fit observed physical properties. The conformation of greatest interest is usually the *native* conformation—the compact, highly organized structure with biological activity: the compactly folded biologically functional structure in the case of a globular protein and the two-stranded Watson–Crick helix with complementary base pairs in the case of DNA.

In this chapter we shall consider the possibility that the native macromolecule may dissociate into two or more subunits without breaking covalent bonds. (We consider hydrogen bonds to be noncovalent.) A *subunit*, as used here, is a covalently bonded chemical unit which may join with other subunits by noncovalent forces to form a macromolecule. For example, an insulin subunit contains two polypeptide chains joined by covalent disulfide links, a native hemoglobin molecule contains eight subunits (four polypeptide chains and four heme groups), and a native DNA molecule contains two subunits (polynucleotide strands). Each possible molecular combination of subunits, including the combination found in the native conformation, will be called a distinct *species*. Each species has its own, arbitrary number of conformations.

The primary property of a macromolecule to be discussed theoretically is the equilibrium concentration of each conformation of each species under

given environmental conditions. When the theoretical framework for solving this problem is constructed, other properties of the macromolecule may be derived which in practice may be simpler to calculate or to interpret, such as the helix content, the degree of ionization, or the total concentration of all species (as measured by colligative properties). If a particular conformation predominates at equilibrium, it may be said to be the stable conformation under the given conditions. It may be simpler to predict changes in the concentrations of conformations as the environment is changed, rather than the absolute concentrations. Thus one may attempt to predict temperature effects, solvent effects, ionic strength effects, and pH effects.

The most difficult problem would be to calculate the three-dimensional structure of the native macromolecule; this would be the conformation calculated to predominate under environmental conditions known to stabilize the native state. (This stratagem would work in principle if the native macromolecules are in a state of minimum free energy with respect to other possible conformations, an assumption for which there is good evidence from the reversibility of conformation changes in the case of some nucleic acids and globular proteins.)

This chapter has two somewhat independent parts. The first part, Sections II through V, outlines a semiempirical theory for calculating conformational concentrations. Starting with classical statistical mechanical theory (Section II) as a conceptual framework, it is shown what assumptions are necessary to arrive at the basis of most theoretical discussions of conformational problems, namely, that the standard molar free-energy change between two conformations is a sum of independent contributions (Section III). These contributions can be evaluated theoretically (Section IV) or from model experiments (Section V). The second part of the chapter, Sections VI through VIII, uses helix–coil transitions of several regular macromolecules to illustrate free-energy change contributions, their relation to conformational concentrations, and their theoretical and experimental evaluation.

II. STATISTICAL MECHANICS THEORY

A. Equilibrium Constants

The problem to be discussed theoretically is the calculation of the equilibrium concentrations of various macromolecular species and conformations in a given solvent at a given temperature and pressure. The equilibrium situation can be imagined to come about as follows. We start with a closed macroscopic chemical system consisting of the solvent molecules, in general of more than one species, and identical macromolecules in the native conformation. The macromolecules are initially identical in the sense of having

the same atomic composition, covalent structure, residue sequence, and number of subunits. Now the system is allowed to come to equilibrium at the given temperature and pressure by means of a number of discrete processes analogous to chemical reactions. If the native macromolecule contains more than one subunit, new species may form by dissociation reactions. The molecules of each species change from one conformation to another. An equilibrium constant analogous to that for a chemical reaction may be defined for each of these processes. From the values of these equilibrium constants, and the known initial concentration of the native macromolecules, the evaluation of the desired equilibrium concentrations is straightforward.

Two kinds of equilibrium constants are defined here. Let $[k_a]$ be the equilibrium concentration of molecules of species k in conformation a and $[k_b]$ the same for conformation b of this species. Likewise, let $[k]$ be the equilibrium total concentration of species k, including all conformations. The concentrations are measured in moles per unit volume.

The equilibrium constant for the interconversion of two conformations, a and b, of species k is denoted by $K_{k,ab}$ and defined in the usual way as a ratio of product and reactant concentrations:

$$K_{k,ab} = [k_b]/[k_a] \tag{1}$$

The equilibrium fraction $\theta_{k,a}$ of macromolecules of this species which are in conformation a is

$$\theta_{k,a} = \left(1 + \sum_{b \neq a} K_{k,ab}\right)^{-1} \tag{2}$$

This fraction is independent of the total concentration $[k]$.

The equilibrium constant for any dissociation reaction between different species is defined in general by

$$K = \prod_k [k]^{\nu_k} \tag{3}$$

in which ν_k is the stoichiometric coefficient appearing in the chemical equation for the reaction (a small integer which is negative for a reactant and positive for a product).

We shall assume that the solution is sufficiently dilute to allow us to set the activity coefficients of the solute species equal to unity, so that these equilibrium constants have values which are independent of concentrations. The values do, however, depend on the temperature and on the solvent composition, ionic strength, and pH.

The equilibrium constants may be evaluated through statistical mechanical theory from partition functions, phase integrals, or configuration integrals for

the macromolecules. The simplest and most direct method is to use the configuration integrals of classical statistical mechanics (*1–3*). Starting with the potential energy of the chemical system (Section II.B), one may define the potential-energy function of a smaller macromolecule–solvent system from which the configuration integrals may in principle be evaluated (Section II.C). Section II.D describes successive simplifications of the form of the configuration integrals.

B. System Potential Energy

Classical statistical mechanics starts with the premise that a total potential energy of the macroscopic system, V_{tot}, may be defined whenever the position of each atomic nucleus is specified. The value of V_{tot} at any instant in time depends only on the atomic positions at that instant and is independent of the temperature.

The zero of energy is arbitrary; for instance, V_{tot} may be set equal to zero when all the atoms are separated at very large distances from one another. V_{tot} may in principle be calculated for any given atom positions as follows. The atoms are numbered to distinguish them and are brought together into rigid molecules. (The atoms are always brought together into the same bonding arrangement; atom exchange or rearrangement of covalent bonds is not allowed.) The potential energy of each isolated, rigid molecule is evaluated as the lowest eigenvalue of the Schrödinger differential equation for the electrons, that is, the energy of the ground electronic state for the given atom positions. Next the molecules are brought together into the final positions. V_{tot} for this configuration of the system is the sum of the energies of the isolated molecules plus perturbation terms for the intermolecular van der Waals interactions (*4,5*).

There are three kinds of intermolecular terms: (1) for the electrostatic interaction between the static (time-average) charge distributions of two molecules, (2) for the polarization interaction between the static charge distribution of one molecule and the induced charge polarization of another molecule, and (3) for the London dispersion interaction between two molecules. The first kind of interaction is pairwise additive; i.e., the total electrostatic interaction of many molecules is the sum of the terms for each pair of molecules considered separately. The polarization interactions, however, are nonadditive; using a sum of pairwise terms is a first-order approximation. The dispersion interactions are usually calculated from second-order quantum-mechanical perturbation theory and in this order (but not in higher orders) are pairwise additive. Further discussion will be given in Section IV.B.

C. Configuration Integrals from the System Potential Energy

1. Independent Macromolecule–Solvent Systems

The assumed unity activity coefficient of each solute species implies that the system may be divided into smaller independent systems, each containing one macromolecule. We shall call each of these small systems a macromolecule–solvent system or small system. It is independent in the sense of making a contribution to the total potential energy, V_{tot}, which depends only on the positions of the atoms within the small system, independent of the other small systems.

The simplest macromolecule–solvent system to consider is a spherical one with the macromolecule at the center. It should be large enough to contain all the solvent molecules which are perturbed from bulk behavior by the presence of the macromolecule. The volume of the small system is the sum of the partial molecular volumes of the constituent molecules. The bulk solvent not included in any of the macromolecule–solvent systems comprises an additional independent system.

The numbers of solvent molecules included in the various macromolecule–solvent systems depend on the macromolecular species but are arranged so that reactions between species do not change the total number of solvent molecules in the bulk solvent system. Therefore, the bulk solvent does not enter into the calculation of reaction equilibrium constants. Since individual solvent molecules are free to move between the macromolecule–solvent systems and the bulk solvent, the identity of the solvent molecules included in one of these systems (but not their number) is constantly changing.

2. Configuration Integrals

Two types of configuration integrals will be defined for each of the small independent macromolecule–solvent systems. A configuration integral of one of these systems is a quantity which depends on the kind of system, on the temperature, and on the volume of the macroscopic system but is independent of the number or concentration of the small systems.

The *conformational configuration integral* $Z_{k,a}$ of a macromolecule–solvent system in which the macromolecule is of species k and conformation a is defined by

$$Z_{k,a} = \left(\prod_s N_s!\right)^{-1} \int \cdots \int \exp\left(-V_k/kT\right) \prod_i dx_i\, dy_i\, dz_i \prod_I dx_I\, dy_I\, dz_I \quad (4)$$

where N_s is the number of molecules of a chemically distinct kind of solvent species s, k the Boltzmann constant, and T the absolute temperature. The potential energy V_k of the macromolecule–solvent system is calculated by

treating the constituent solvent molecules as distinguishable molecules remaining inside the small system. V_k is a function of the Cartesian coordinates x_i, y_i, z_i of each atom i in the macromolecule and the Cartesian coordinates x_I, y_I, z_I of each atom I in the solvent molecules. (Besides the interactions among these atoms within the small system, V_k may include the interaction of the small system with the surrounding bulk solvent considered as a continuum.) These coordinates determine the configuration of the small system. One atom in the macromolecule may be chosen as the center of the sphere which encloses the small system; the coordinates of this atom may range over the volume of the total macroscopic system, and the coordinates of all the other atoms may range over the volume of the small sphere. The integration indicated in Eq. (4) is to be performed over all the values of the coordinates within these ranges, which give configurations consistent with the covalent bonding of the atoms and with the particular conformation a of the macromolecule. The resulting configuration integral has the dimensions of volume to the power of the number of atoms in the macromolecule–solvent system.

In the integration of Eq. (4), the atoms are treated as distinguishable with a fixed arrangement of covalent bonding. Thus a configuration in which two atoms of the same isotope have exchanged positions on the same or different molecules, the coordinates of all other atoms remaining the same, is unacceptable even though the new configuration would be physically indistinguishable from the original one. The factor $1/N_s!$ for each solvent species s corrects the configuration integral for the number of physically indistinguishable configurations $N_s!$ which are obtained during the integration by interchanging solvent molecules of this species. This correction is necessary in order to avoid calculating a spurious entropy of solvent mixing for a reaction, due to the fact that in a reaction solvent molecules leave one set of macromolecule–solvent systems and enter a new set.

The *species configuration integral* Z_k of a macromolecule–solvent system in which the macromolecule belongs to species k but is allowed to assume all conformations is defined in the same way, except that the integration includes the configurations for all conformations of the species. We may write simply

$$Z_k = \sum_a Z_{k,a} \tag{5}$$

3. Equilibrium Constants from Configuration Integrals

According to classical statistical mechanics, the equilibrium constants which were defined in Eqs. (1) and (2) are given, in terms of the above configuration integrals, by the relations

$$K_{k,ab} = Z_{k,b}/Z_{k,a} \tag{6}$$

and

$$K = \prod_k (Z_k/\sigma_k N_0 V)^{\nu_k} \tag{7}$$