Principles of Physical Biochemistry

Second Edition

Kensal E. van Holde
Professor Emeritus of Biochemistry and Biophysics
Department of Biochemistry and Biophysics
Oregon State University

W. Curtis Johnson
Professor Emeritus of Biochemistry and Biophysics
Department of Biochemistry and Biophysics
Oregon State University

P. Shing Ho
Professor and Chair, Biochemistry and Biophysics
Department of Biochemistry and Biophysics
Oregon State University

Upper Saddle River, New Jersey 07458

CHAPTER 

Biological Macromolecules

1.1 GENERAL PRINCIPLES

In physical biochemistry, we are interested in studying the physical properties of biological macromolecules, including proteins, RNA and DNA, and other biological polymers (or *biopolymers*). These physical properties provide a description of their structures at various levels, from the atomic level to large multisubunit assemblies. To measure these properties, the physical biochemist will study the interaction of molecules with different kinds of radiation, and their behavior in electric, magnetic, or centrifugal fields. This text emphasizes the basic principles that underlie these methodologies.

In this introductory chapter, we briefly review some of the basic principles of structure and structural complexity found in biological macromolecules. Most readers will have already learned about the structure of biological macromolecules in great detail from a course in general biochemistry. We take a different point of view; the discussion here focuses on familiarizing students with the quantitative aspects of structure. In addition, this discussion includes the symmetry found at nearly all levels of macromolecular structure. This approach accomplishes two specific goals: to illustrate that the structures of macromolecules are very well defined and, in many ways, are highly regular (and therefore can be generated mathematically); and to introduce the concepts of symmetry that help to simplify the study and determination of molecular structure, particularly by diffraction methods (Chapters 6 and 7). This discussion focuses primarily on the structures of proteins and nucleic acids, but the general principles presented apply to other macromolecules as well, including polysaccharides and membrane systems.

1.1.1 Macromolecules

As a basic review of molecular structure, perhaps the place to start is to ask the question, What is a *molecule*? Here, the definition of a biological molecule differs slightly from the definition learned in chemistry. In organic chemistry, a molecule consists of two or more atoms that are covalently bonded in specific proportions according to weight or stoichiometry, and with a unique geometry. Both stoichiometry (the chemical formula) and geometry (the chemical structure) are important. Dichloroethylene, for example, has the specific chemical formula $C_2H_2Cl_2$. This, however, does not describe a unique molecule, but rather three different molecules. The geometry for one such molecule is defined by the arrangements of the chlorine atoms, as in *cis*-1,2-dichloroethylene (Figure 1.1). Now, the identity of the molecule is unambiguous.

In biochemistry, a single molecule is considered to be a component that has well-defined stoichiometry and geometry, and is not readily dissociated. Thus, to a biochemist, a molecule may not necessarily have all the parts covalently bonded, but may be an assembly of noncovalently associated polymers. An obvious example of this is hemoglobin. This is considered to be a single molecule, but it consists of four distinct polypeptides, each with its own heme group for oxygen binding. One of these polypeptide-heme complexes is a *subunit* of the molecule. The heme groups are noncovalently attached to the polypeptide of the subunit, and the subunits are noncovalently interacting with each other. The stoichiometry of the molecule can also be described by a chemical formula, but is more conveniently expressed as the

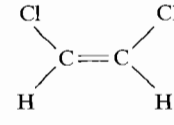
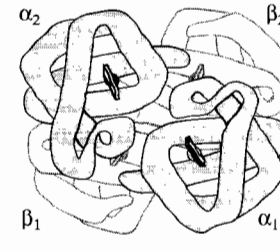
Molecule	Stoichiometry	Geometry (Structure)
cis-1,2-Dichloroethylene	$C_2H_2Cl_2$	
Hemoglobin	$\alpha_2\beta_2$	

Figure 1.1 Examples of molecules in chemistry and macromolecules in biochemistry. The simple compound *cis*-1,2-dichloroethylene is uniquely defined by the stoichiometry of its atomic components and the geometry of the atoms. Similarly, the structure of a biological macromolecule such as hemoglobin is defined by the proportions of the two subunits (the α and β-polypeptide chains) and the geometry by the relative positions of the subunits in the functional complex.

composition of *monomer* units. The stoichiometry of a protein therefore is its amino acid composition. The geometry of a biological molecule is again the unique linear and three-dimensional (3D) arrangements of these components. This is the *structure* of a biochemical molecule.

A *macro*molecule is literally a large molecule. A biological macromolecule or *biopolymer* is typically defined as a large and complex molecule with biological function. We will take a chemical perspective when dealing with macromolecules, so, for this discussion, size will be judged in terms of the number of components (atoms, functional groups, monomers, and so on) incorporated into the macromolecule. Complexity generally refers to the organization of the three-dimensional structure of the molecule. We will treat size and structural complexity separately.

What is considered large? It is very easy to distinguish between molecules at the two extremes of size. Small molecules are the diatomic to multiple-atom molecules we encounter in organic chemistry. At the upper end of large molecules is the DNA of a human chromosome, which contains tens of billions of atoms in a single molecule. At what point do we decide to call something a macromolecule? Since these are biopolymers, their size can be defined by the terms used in polymer chemistry, that is, according to the number of sugar or amino acid or nucleic acid residues that polymerize to form a single molecule. Molecules composed of up to 25 residues are called *oligomers,* while polymers typically contain more than 25 residues. This is an arbitrary distinction, since some fully functional molecules, such as the DNA-condensing J-protein of the virus G4, contain 24 residues.

The structure of biological macromolecules is hierarchical, with distinct levels of structure (Figure 1.2). These represent increasing levels of complexity, and are defined below.

Monomers are the simple *building blocks* that, when polymerized, yield a macromolecule. These include sugars, amino acids, and nucleic acid residues of the polymers described above.

Primary structure (abbreviated as 1°) is the linear arrangement (or sequence) of residues in the covalently linked polymer.

Secondary structure (abbreviated as 2°) is the local regular structure of a macromolecule or specific regions of the molecule. These are the *helical* structures.

Tertiary structure (abbreviated as 3°) describes the global 3D fold or *topology* of the molecule, relating the positions of each atom and residue in 3D space. For macromolecules with a single subunit, the functional tertiary structure is its *native structure.*

Quaternary structure (abbreviated as 4°) is the spatial arrangement of multiple distinct polymers (or subunits) that form a functional complex.

Not all levels of structure are required or represented in all biological macromolecules. Quaternary structure would obviously not be relevent to a protein such as myoglobin that consists of a single polypeptide. In general, however, all biological macromolecules require a level of structure up to and including 2°, and typically 3°

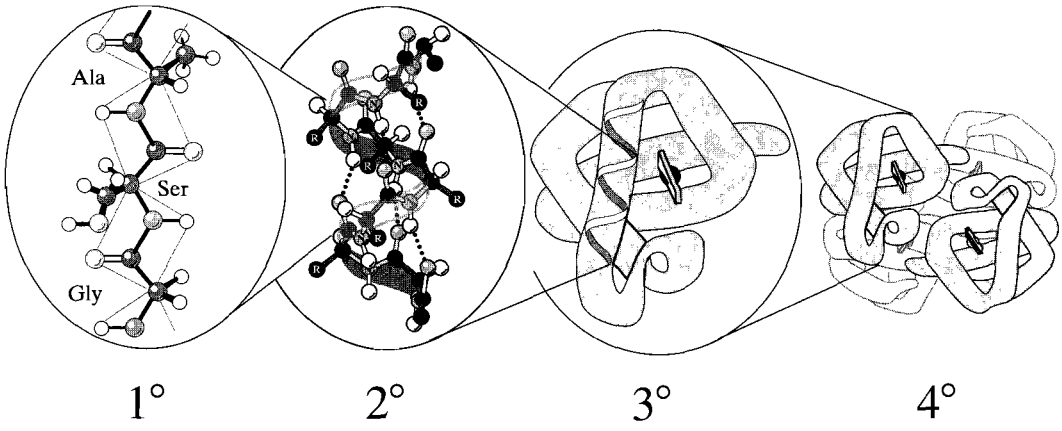


Figure 1.2 Hierarchical organization of macromolecular structure. The structures of macromolecules are organized starting with the simple monomers to form the sequence in the primary structure, which folds into the local regular helices of secondary structure, the global tertiary structure, and the association of folded chains to form complexes in the quaternary structure.

for biological function. The relationship between these levels of structure is often presented in sequential order as 1°, followed by 2°, which is followed by 3°, and finally 4° (if present). This sequential relationship is a convenient means of presenting the increasing complexity of macromolecular structure; however, it is not clear that this is how a molecule folds into its functional form. The most recent models for protein folding suggest that a less compact form of 3° (often called a *molten globule* state, see Section 4.4.3) must occur first in order to form the environment to stabilize helices (2°). One of the goals in physical biochemistry is to understand the rules that relate these levels of structural complexity. This is often presented as the problem of predicting 3D structure (2° to 3°) from the sequence (1°) of the building blocks. The problem of predicting the complete 3D structure of a protein from its polypeptide sequence is the *protein-folding problem.* We can define a similar folding problem for all classes of macromolecules.

We will see how this hierarchical organization of structure applies to the structures of proteins and nucleic acids, but first we need to discuss some general principles that will be used throughout this chapter for describing molecular structure. It should be emphasized that we cannot directly see the structure of a molecule, but can only measure its properties. Thus, a picture of a molecule, such as that in Figure 1.2, is really only a model described by the types of atoms and the positions of the atoms in 3D space. This model is correct only when it conforms to the properties measured. Thus, methods for determining the structure of a molecule in physical biochemistry measure its interactions with light, or with a magnetic or electric field, or against a gradient. In all cases, we must remember that these are models of the structure, and the figures of molecules presented in this book are nothing more than representations of atoms in 3D space. It is just as accurate (and often more useful) to represent the structure as a list of these atoms and their atomic coordinates (x, y, z) in a standard Cartesian axis system.

1.1.2 Configuration and Conformation

The arrangement of atoms or groups of atoms in a molecule is described by the terms *configuration* and *conformation*. These terms are not identical. The configuration of a molecule defines the position of groups around one or more nonrotating bonds or around *chiral* centers, defined as an atom having no plane or center of symmetry. For example, the configuration of *cis*-1,2-dichloroethylene has the two chlorine atoms on the same side of the nonrotating double bond (Figure 1.3). To change the configuration of a molecule, chemical bonds must be broken and remade. A conversion from the *cis*- to *trans*-configuration of 1,2-dichloroethylene requires that we first break the carbon-carbon double bond, rotate the resulting single bond, then remake the double bond. In biological macromolecules, configuration is most important in describing the stereochemistry of a chiral molecule. A simple chiral molecule

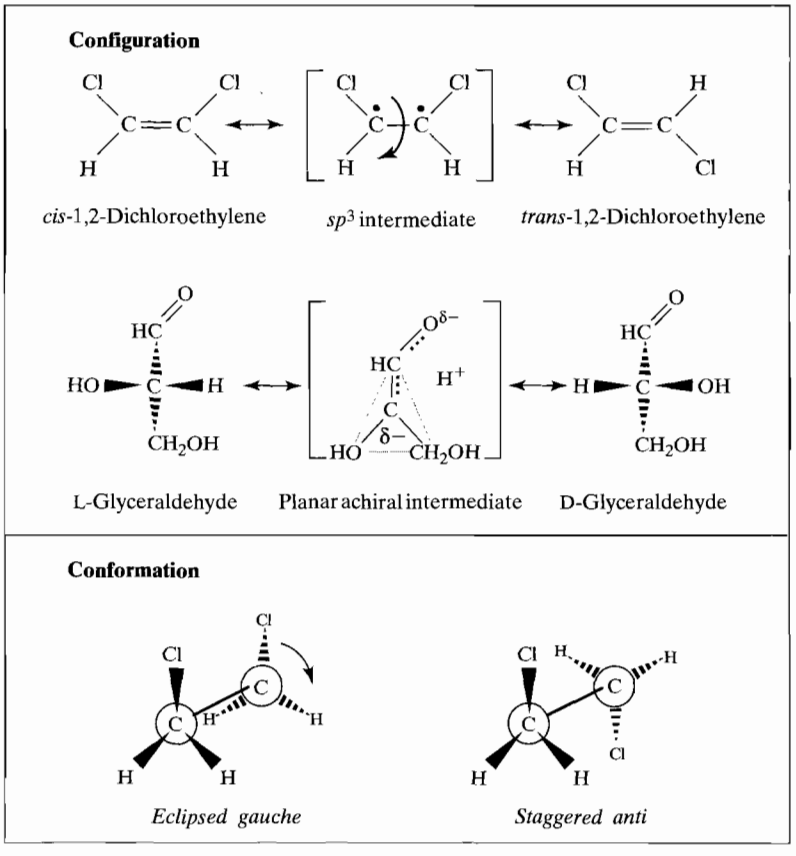


Figure 1.3 Configuration and conformation both describe the geometry of a molecule. The configuration of a molecule can be changed only by breaking and remaking chemical bonds, as in the conversion of a *cis*-double bond to one that is in the *trans*-configuration, or in converting from the L- to the D-stereoisomer of a chiral molecule. Conformations can be changed by simple rotations about a single bond.

has four unique chemical groups arranged around a tetrahedral atom (usually a carbon atom with sp^3 hybridization). To change the configuration or chirality of this molecule, we must break one bond to form a planar *achiral* intermediate, and reform the bond on the opposite side of the plane. The resulting molecule is the *stereoisomer* or *enantiomer* of the starting structure. The stereoisomers of a molecule, even though they are identical in chemical composition, are completely different molecules with distinct properties, particularly their biological properties. Sugars that have more than one chiral center have more complex stereochemistry.

The conformation of a molecule, on the other hand, describes the spatial arrangement of groups about one or more freely rotating bonds. For example, 1,2-dichloroethane, the saturated version of dichloroethylene, has no restrictions to rotation about the chemical bonds to prevent the chlorine atoms from sitting on the same or opposite sides of the central carbon-carbon bond. These positions define the *gauche* and *anti* structural isomers, respectively. In addition, the conformation can be *eclipsed* or *staggered,* depending on whether the groups are aligned or misaligned relative to each other on either side of the carbon-carbon bond. The conformation of a molecule thus describes the structural isomers generated by rotations about single bonds (Figure 1.3). A molecule does not require any changes in chemical bonding to adopt a new conformation, but may acquire a new set of properties that are specific for that conformation.

The stereochemistry of monomers. The monomer building blocks of biological macromolecules are *chiral* molecules, with only a few exceptions. There are many conventions for describing the stereochemistry of chiral molecules. The stereochemistry of the building blocks in biochemistry has traditionally been assigned according to their absolute configurations. This provides a consistent definition for the configuration of all monomers in a particular class of biopolymer. For example, the configurations of sugar, amino acid, and nucleic acid residues are assigned relative to the structures of L- and D-glyceraldehyde (Figure 1.4). In a standard projection formula, the functional groups of D-glyceraldehyde rotate in a clockwise direction around the chiral carbon, starting at the aldehyde, and going to the hydroxyl, then the hydroxymethyl, and finally the hydrogen groups. The configuration of the building blocks are therefore assigned according to the arrangement of the analogous functional groups around their chiral centers. Since glyceraldehyde is a sugar, it is easy to see how the configurations of the carbohydrate building blocks in polysaccharides are assigned directly from comparison to this structure. Similarly, the configuration of the ribose and deoxyribose sugars of the nucleic acids can be assigned directly from glyceraldehyde. Biopolymers are typically constructed from only one enantiomeric form of the monomer building blocks. These are the L-amino acids in polypeptides and the D-sugars in polysaccharides and polynucleotides.

For an amino acid such as alanine, the chiral center is the C_α carbon directly adjacent to the carboxylic acid. The functional groups around the C_α carbon are analogous but not identical to those around the chiral center of glyceraldehyde. The L-configuration of an amino acid has the carboxylic acid, the amino group, the α-hydrogen and the methyl side chain arranged around the C_α carbon in a

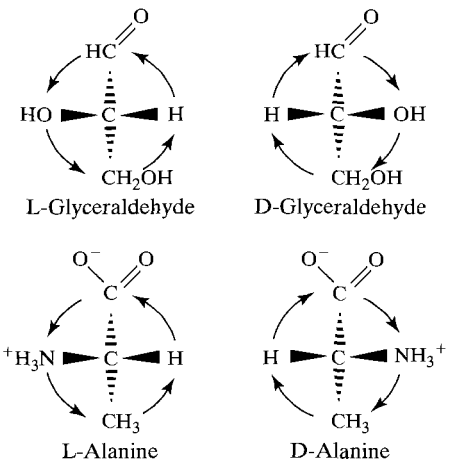


Figure 1.4 Absolute configuration of monomer building blocks. The stereochemistry of the monomers in biopolymers are assigned relative to L- and D-glyceraldehyde. Carbohydrates and the sugars of nucleic acids are assigned directly according to the rotation starting at the carbonyl group. For amino acids, the stereochemistry is defined according to the rotation starting at the analogous carboxyl group.

manner analogous to the aldehyde, hydroxyl, hydrogen, and hydroxymethyl groups in L-glyceraldehyde.

Conformation of molecules. Unlike the configuration of a macromolecule, the number of possible conformations of a macromolecule can be enormous because of the large number of freely rotating bonds. It is thus extremely cumbersome to describe the conformation of a macromolecule in terms of the alignment of each group using the *gauche/anti* and *eclipsed/staggered* distinctions. It is much more convenient and accurate to describe the *torsion angle* θ about each freely rotating bond. The torsion angle is the angle between two groups on either side of a freely rotating chemical bond. The convention for defining the torsion angle is to start with two nonhydrogen groups (A and D) in the *staggered anti* conformation with $\theta = -180°$. Looking down the bond to be rotated (as in Figure 1.5) with atom A closest to you, rotation of D about the B—C bond in a clockwise direction gives a positive rotation of the bond. Thus, the values for θ are defined as 0° for the

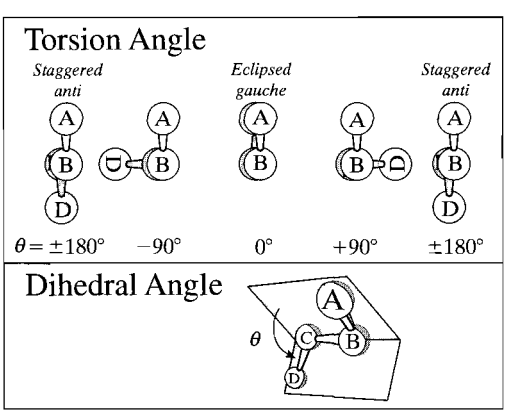


Figure 1.5 Torsion angles and dihedral angles (θ). The rotation around a single bond is described by the torsion angle of the four atoms around the bond (A—B—C—D) and the dihedral angle θ relating the planes defined by atoms A—B—C and by B—C—D.

eclipsed gauche conformation to +180° for the *staggered anti* conformation. Notice that the start and end points ($\theta = \pm 180°$) are identical.

The angle between the two groups of atoms can also be defined by the *dihedral angle*. Mathematically, the dihedral angle is defined as the angle between two planes. Any three atoms about a freely rotating bond (two atoms in the bond, plus one extending from that bond, as in A—B—C and B—C—D in Figure 1.5) defines a plane. Thus, we can see from this definition that the torsion and dihedral angles are identical.

Changing the conformation of a molecule does not make a new molecule, but can change its properties. The properly folded conformation of a protein, referred to as the *native* conformation, is its functional form, while the unfolded or *denatured* conformation is nonfunctional and often targeted for proteolysis by the cell. Thus, both the configuration and conformation of a molecule are important for its shape and function, but these represent distinct characteristics of the molecule and are not interchangeable terms. The conformations of polypeptides and polynucleic acids will be treated in greater detail in later sections.

1.2 MOLECULAR INTERACTIONS IN MACROMOLECULAR STRUCTURES

The configurations of macromolecules in a cell are fixed by covalent bonding. The conformations, however, are highly variable and dependent on a number of factors. The sequence-dependent folding of macromolecules into secondary, tertiary, and quaternary structures depends on a number of specific interactions. This includes the interactions between atoms in the molecule and between the molecule and its environment. How these interactions affect the overall stability of a molecule and how they can be used to construct models of macromolecules are discussed in greater detail in Chapter 3. In this introductory chapter, we define some of the characteristics of these interactions, so that we can have some understanding for how the various conformations of proteins and polynucleic acids are held together.

1.2.1 Weak Interactions

The covalent bonds that hold the atoms of a molecule together are difficult to break, releasing large amounts of energy during their formation and concomitantly requiring large amounts of energy to break (Figure 1.6). For a stable macromolecule, they can be treated as invariant. The conformation of a macromolecule, however, is stabilized by weak interactions, with energies of formation that are at least one order of magnitude less than that of a covalent bond. The weak interactions describe how atoms or groups of atoms are attracted or repelled to minimize the energy of a conformation.

These are, in general, distance-dependent interactions, with the energies being inversely proportional to the distance r or some power of the distance (r^2, r^3, etc.) separating the two interacting groups (Table 1.1). As the power of the inverse

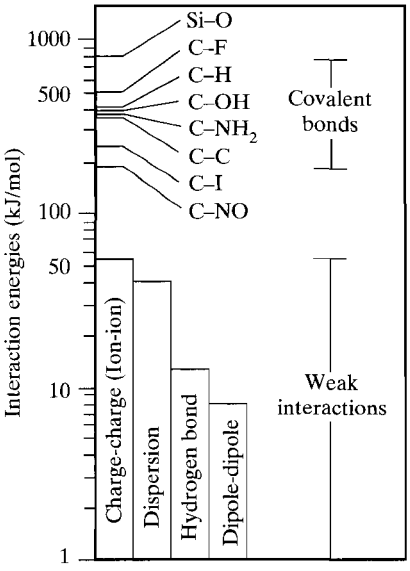


Figure 1.6 Energies of molecular interactions. The interactions that define the structure of a molecule range from the strong interactions of covalent bonds (200 to 800 kJ/mol) to the weak charge-charge (or ion-ion), dipole-dipole, dispersion, and hydrogen-bonding interactions (0 to 60 kJ/mol).

distance dependency increases, the interaction approaches zero more rapidly as r increases, and thus becomes a shorter range interaction. The interaction energy between two charges varies as $1/r$; this is a long-range interaction. At the other extreme are the induced dipole-induced dipole (or *dispersion*) interactions. These interactions describe the natural tendency of atoms to attract, regardless of charge and polarity, because of the polarizability of the electron clouds. Its dependence on $1/r^6$ defines this as a very short-range interaction, having a negligible interaction energy at about 1 nm or greater. Directly opposing this attraction, however, is steric repulsion, which does not allow two atoms to occupy the same space at the same time. This repulsion occurs at even shorter distances and is dependent on $1/r^{12}$. Together, the attractive dispersion and repulsive exclusion interactions define an optimal distance separating any two neutral atoms at which the energy of interaction is a minimum. This optimal distance thus defines an effective radius (the *van der Waals radius,* or r_{vdw}) for each type of atom. The potential energy functions for

Table 1.1 Relationship of Noncovalent Interactions to the Distance Separating the Interacting Molecules, r

Type of Interaction	Distance Relationship
Charge-charge	$1/r$
Charge-dipole	$1/r^2$
Dipole-dipole	$1/r^3$
Charge-induced dipole	$1/r^4$
Dispersion	$1/r^6$
Repulsion	$1/r^{12}$

each interaction and their application to simulating the thermodynamic properties of macromolecules are treated in detail in Chapter 3.

The energies associated with long-range interactions (charge-charge, charge-dipole, and dipole-dipole) are dependent on the intervening medium. The interaction between two charged atoms, for example, becomes shielded in a polar medium and is therefore weakened. The least polarizable medium is a vacuum, with a dielectric constant of $\kappa\epsilon_O = 4\pi 8.85 \times 10^{-12}\ C^2\ J \cdot m$, where $\epsilon_O = 8.85 \times 10^{-12}\ C^2\ J \cdot m$ and $\kappa = 4\pi$ for a point charge. The polarizability of a medium is defined as its dielectric constant D relative to that of a vacuum. The expressions for the energy of long-range interactions are all inversely related to the dielectric of the medium and are therefore weakened in a highly polarizable medium such as water.

With the dielectric constant, we introduce the environment as a factor in stabilizing the conformation of a macromolecule. How the environment affects the weak interactions is discussed in the next section. In the process, two additional interactions (hydrogen bonds and hydrophobicity) are introduced that are important for the structure and properties of molecules.

1.3 THE ENVIRONMENT IN THE CELL

The structures of macromolecules are strongly influenced by their surrounding environment. For biopolymers, the relevant environment is basically the solvent within the cell. Because the mass of a cell is typically more than 70% water, there is a tendency to think of biological systems primarily as aqueous solutions. Indeed, a large majority of studies on the properties of biological macromolecules are measured with the molecule dissolved in dilute aqueous solutions. This, however, does not present a complete picture of the conditions for molecules in a cell. First, a solution that is 70% water is in fact highly concentrated. In addition, the cell contains a very large surface of membranes, which presents a very different environment for macromolecules, particularly for proteins that are integral parts of the bilayer of the membranes. The interface between interacting molecules also represents an important nonaqueous environment. For example, the recognition site of the TATA-binding protein involves an important aromatic interaction between a phenylalanyl residue of the protein and the nucleotide bases of the bound DNA.

In cases where solvent molecules are observed at the molecular interfaces (for example, between the protein and its bound DNA), the water often helps to mediate interactions, but is often treated as part of the macromolecule rather than as part of the bulk solvent. In support of this, a well-defined network of water molecules has been observed to reside in the minor groove of all single-crystal structures of DNA duplex. Results from studies using nuclear magnetic resonance (NMR) spectroscopy indicate that the waters in this spine do not readily exchange with the bulk solvent and thus can be considered to be an integral part of the molecule. We start by briefly discussing the nature of the aqueous environment because it is the dominant solvent system, but we must also discuss in some detail the nonaqueous environments that are also relevant in the cell.

CHAPTER 3

Molecular Thermodynamics

3.1 COMPLEXITIES IN MODELING MACROMOLECULAR STRUCTURE

One of the basic principles in biochemistry is that the information needed to fold a macromolecule into its native three-dimensional (3D) structure is contained within its sequence. This principle was first demonstrated by Anfinsen (1961, 1973), who showed that unfolded (*denatured*) ribonuclease spontaneously refolded (*renatured*) to an enzymatically active form. A long-sought goal in physical biochemistry is to accurately predict the 3D structure of a macromolecule starting with its sequence—this is the *folding problem* for macromolecules. In this chapter, we describe methods to model the conformations of macromolecules by using the basic principles of thermodynamics at the level of individual molecules. This includes a description of the interactions that facilitate the proper folding of macromolecules and how these interactions are formulated into energy functions that are useful for modeling macromolecular structures and behavior through *molecular simulation*. However, we stress that the folding problem has not been solved and that the principles described here represent only steps toward a general solution to the problem.

In theory, all the chemical properties of a macromolecule, including its 3D structure, can be predicted from an accurate description of the total thermodynamic state of the system at the atomic level. Although atoms are most accurately described by *quantum mechanics* (Chapter 8), most descriptions to date are only approximate. In this chapter, we review the methods of classical Newtonian physics to describe the thermodynamic properties of macromolecules.

Even with this classical approach, modeling macromolecular structures is complicated by the large number of atoms in the system. Insulin, a small protein of 51 amino acid residues, is composed of over 760 atoms, nearly 400 of which are non-hydrogen atoms. If we include ions and water molecules that are associated with each

molecule, the system becomes even more unwieldy. For example, an insulin molecule in a 1 mM solution would, on average, interact with over 10^8 water molecules.

Finally, the conformations of the molecules are highly variable, with both the macromolecule and its solvent environment assuming any of a large number of different structures at any time. Each macromolecule has many degrees of freedom, not only in rotation and translation in solution, but also internal degrees of freedom at each freely rotating bond. Thus, the problem in trying to model macromolecular structures de novo stems from the large size and complexity of the system. The strategy, therefore, is to simplify the system. We will start by discussing some of these simplifications and the assumptions required to make them work.

3.1.1 Simplifying Assumptions

One of the first simplifications in trying to model a system of biological molecules in solution is to assume that the average behavior of the system can be represented by a single molecule (Figure 3.1). In this model, a single macromolecule and its associated solvent is isolated in a box. If the single molecule in this box is truly dynamic, it will eventually sample all the possible conformations accessible to the system. The properties of a population of molecules is thus represented by the time-averaged behavior of a single macromolecule in an isolated box. By making this a periodic (repeating) box, the contents of one box are identical to that of all the other boxes, and anything that leaves the box from one direction must simultaneously enter it from the opposite direction so that the concentration of material remains constant. Molecules are not allowed to move freely in or out of the box except in this periodic manner. This limitation may appear trivial, but in fact it is important if we consider that many biological molecules are dramatically affected by self-association, association with salts, and so on. One of the logistical problems in this type of model is to define a box that is large enough to accurately simulate the system, including the behavior

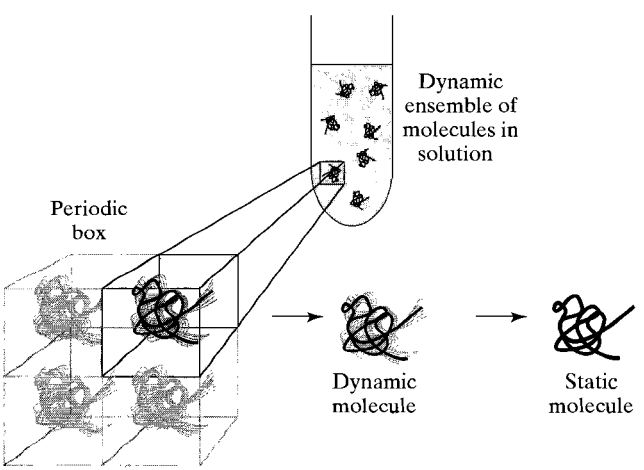


Figure 3.1 Simplifying macromolecular systems. An ensemble of dynamic molecules in solution is first simplified to a single molecule in a periodic box surrounded by solvent, then to an isolated molecule in the absence of solvent, and finally to a static molecule at a low-energy state.

of the bulk solvent. From an experimental point of view, there are now a number of very exciting methods to detect and study the structure and behavior of single molecules (Chapter 15). However, one needs to be cautioned that the characteristics of each individual molecule is unique and, therefore, many single-molecule measurements must be made in order to understand the average properties of a population.

Many macromolecular systems are still too unwieldy to simulate even when isolated in a periodic box. The next step in reducing the size and complexity of the system is to remove much or all the individual solvent and ions in the box. Obviously, if a molecule is treated in vacuo, in the absence of a solvent environment, there is no need for a physical box to contain the system. The problems with this simplification are obvious. Nearly all macromolecules make some contact with water, with the degree of solvent interaction defining the properties of the molecule. Thus, both hydrophilic and hydrophobic effects are largely ignored. However, a simulation in vacuo does reduce the number of atoms in the system by at least an order of magnitude. A number of approximations have been incorporated into the various molecular simulation methods as attempts to include the effects of the solvent without explicitly including solvent molecules as part of the system. We will discuss some of these approximations for treating the solvent and how they contribute to our understanding of macromolecular folding.

Finally, we can make the assumption that the native conformation of a molecule is the one with the lowest overall potential energy. The dynamic properties of the system are ignored in this case. Nonetheless, this general principle lays a foundation for methods that try to study and predict macromolecular structure, and does help to simplify the overall system.

We will start at the lowest level of molecular structure (the atom), and work through methods that attempt to rebuild the original complex system in a series of manageable steps.

3.2 MOLECULAR MECHANICS

3.2.1 Basic Principles

The best predictions of the structure and physical properties for a molecule come from an exact quantum mechanical treatment of every atom within a molecular system (Chapter 8). However, this is only analytically possible for the hydrogen atom. Using approximations to the wavefunctions for larger atoms introduces errors that are compounded as the molecule increases in size and complexity. The alternative is to apply a classical rather than quantum mechanical treatment to describe the interactions between atoms.

According to classical mechanics, the total energy E within a system includes both the kinetic K and the potential energies V, as discussed in Chapter 2 and as summarized in Eq. 3.1.

$$E = K + V \tag{3.1}$$

The kinetic energy used here includes all the motions of the atoms in the system, and thus is the sum of their kinetic energies. The potential energy of a macromolecule in its various conformations can be represented by a multidimensional surface with hills and valleys (Figure 3.2); any particular conformation corresponds to one point on this surface.

The description of molecular interactions is based on the principles of Newtonian physics. This is *molecular mechanics*, where molecular motions are determined by the masses of and the forces acting on atoms. The nuclei contribute the mass while the electrons provide the force of interaction between atoms. Thus, in classical molecular mechanics, the electrons and nuclei are treated together.

The basic relationship in molecular mechanics is Newton's second law of motion, which relates the force F along a *molecular trajectory* (the distance vector $\mathbf{r}$) to the acceleration a of a mass m along that trajectory.

$$F = ma \tag{3.2}$$

To simplify this discussion, we will only consider one component (for example x) of $\mathbf{r}$. The potential energy surface, therefore, becomes an *energy profile* (Figure 3.2), but the relationships that we derive in one dimension apply to all three directions for the molecular trajectory.

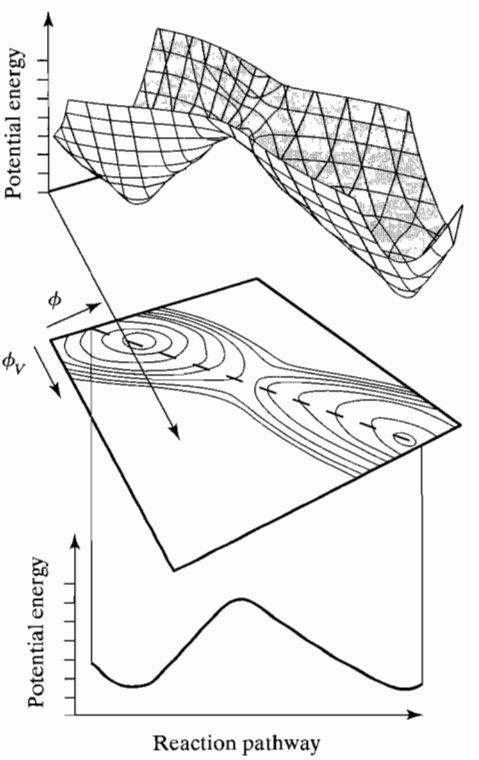


Figure 3.2 The potential energy surface for rotating a tyosine side chain in the protein pancreatic trypsin inhibitor. The potential energy is plotted as a function of the ring dihedral angle ϕ defined by the atoms C_α-C_β-C_γ-$C_{\delta 1}$ of the tyrosyl residue and a virtual dihedral angle ϕ_V, which measures the angle of the ring relative to the plane of the peptide bond (defined by the atoms C_β-C_γ-$C_{\delta 2}$ of the tyrosyl residue and the amino nitrogen of the next amino acid). The three-dimensional surface is projected onto a flat topographical map of the surface, where each contour represents an isoenergy level. The cross-section through the two-dimensional map (dotted line through) shows the profile of the potential energy along the reaction pathway. [Data from J. A. McCammon et al. (1983), *J. Am. Chem. Soc.* **105**, 2232–2237.]

The force exerted in the direction of r is related to the potential energy by

$$F = -\frac{\partial V}{\partial r} \tag{3.3}$$

The force applied on an atom thus depends on how the potential energy changes as the distance between interacting atoms changes. The local gradient in the potential energy defines the *force field* in molecular mechanics. There are a number of different force fields used for modeling the structures of macromolecules, each having its own distinctive advantages and disadvantages. Our discussion will focus on energy functions that are common features of macromolecular force fields.

Newton's first law of motion is the special case for a system at equilibrium, where the net force is defined as $F = 0$. A system at equilibrium thus has $-\partial V/\partial r = 0$, which means that the molecule sits at a potential energy minimum. This is the basic principle behind *energy minimization* methods, which attempt to find the lowest energy conformation of a macromolecule, providing a static picture of the system at equilibrium.

The kinetic energy K of an atom is related to its velocity v or, equivalently, its momentum p.

$$K = \frac{1}{2}mv^2 = \frac{1}{2}\frac{p^2}{m} \tag{3.4}$$

The parameter K describes the dynamic change in the atomic positions at any time t. Thus, the methods of *molecular dynamics* are used to simulate the time-dependent changes in a system.

We discuss the application of molecular force fields to the simulation of molecular properties using energy minimization and molecular dynamics in greater detail later in this chapter. First, we must define the potential energy functions describing atomic and molecular interactions that are common to macromolecular force fields. This is followed by a discussion of how protein and nucleic acid structures are stabilized through these interactions.

3.2.2 Molecular Potentials

A description of the total potential energy in a macromolecular system must include the intermolecular interactions among molecules and the intramolecular interactions among atoms within the molecule. The potential energy of a single, isolated molecule depends only on the intramolecular interactions. The total intramolecular potential energy V_{total} is thus the sum of two types of interactions, the bonding V_{bonding} and the nonbonding interactions $V_{\text{nonbonding}}$. For N number of atoms in the molecule, we can write

$$V_{\text{total}} = \sum_{i=1}^{N} (V_{\text{bonding}} + V_{\text{nonbonding}})_i \tag{3.5}$$

Every conceivable conformation of a system has a corresponding value of V_{total}. If we plot V_{total} as a multidimensional surface where the coordinates are the positions

of all atoms in the system, this is the energy surface for the system. To calculate the total potential, it is often sufficient to add together the interactions between pairs or small groups of adjacent atoms. These may be either *bonded* or *nonbonded* pairs or groups.

The bonding interactions are the covalent bonds that hold the atoms together, while the nonbonding interactions include electrostatic, dipolar, and steric interactions. In molecular mechanics, the potential energy functions are derived empirically, based on how molecules behave, as opposed to *ab initio* derivations from quantum mechanics. In many instances, the empirically derived functions are more accurate because they are based on the macroscopic properties that are actually observed for macromolecules, while quantum mechanical functions are often very approximate. We should note, however, that many parameters for these properties (for example, partial charges of atoms in peptide bonds) are derived from quantum mechanical calculations.

3.2.3 Bonding Potentials

The chemical bond that holds two atoms together is conceptually the easiest interaction in the total potential energy surface to understand. It also dominates the surface because of its magnitude ($\sim$150 to $>$1000 kJ/mol). The bond energy is the energy absorbed in breaking a bond or released in forming a bond (Table 3.1). Bond energies are therefore enthalpic energies.

The bonding energy is modeled in a force field as a distance-dependent function. An equilibrium distance r_0 is defined for the standard length of the chemical bond where, according to Eq. 3.3, no force is exerted on either of the bonded atoms. For any distance $r \neq r_0$ the atoms are forced toward r_0 by the potential energy surface. Therefore, how this force is applied depends on how V is defined.

The potential energy profile for a chemical bond is an anharmonic function with a minimum value at r_0, a steep ascending curve for $r < r_0$, and a more gradual ascending curve that approaches $V = 0$ for $r > r_0$ (Figure 3.3). The simplest form of the potential treats the chemical bond as a spring. Like a bond, a spring has an

Table 3.1 Average Dissociation Energies of Chemical Bonds in Organic Molecules

Bond Type	Bond Energy (kJ/mol)
C—H	408
C—C	342
C=C	602
C=N	606
C=O	732
O—H	458

Source: From A. Streitwieser, Jr., and C. H. Heathcock (1976), *Introduction to Organic Chemistry*, Macmillan, New York.

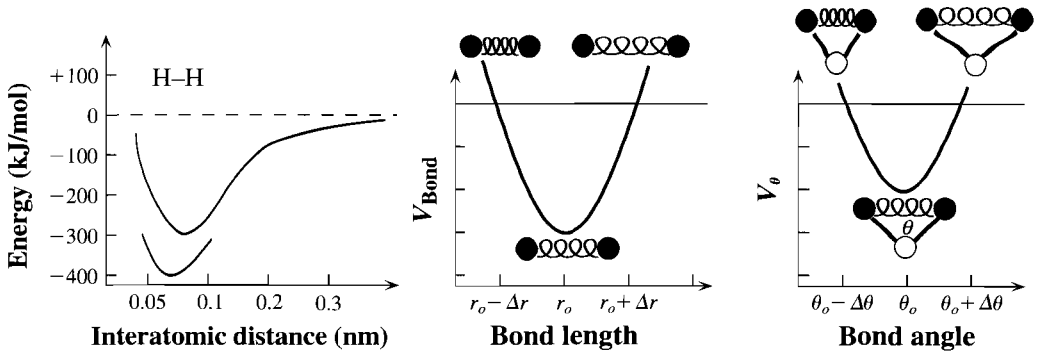


Figure 3.3 The potential energies for a hydrogen-hydrogen bond, and for deforming a covalent bond V_{bond} and the bond angle between three bonded atoms V_θ as treated by simple spring models. The potential energies calculated from quantum mechanics (upper curve) is compared to the experimental values (lower curve) for a hydrogen-hydrogen covalent bond (H—H). The deformations associated with any bond length are modeled, in the simplest case, as harmonic springs, with a spring constant for stretching $+\Delta r$ or compressing $-\Delta r$ a bond from the equilibrium bond length r_o. The spring model greatly overestimates the potential energy for stretching to large Δr. Deformations to the bond angle θ_o can be similarly treated as a spring between the 1-3 atoms of the three bonded atoms.

equilibrium length, and stretching or compressing the spring from r_0 requires an applied force. The potential energy of a chemical bond V_{bond} is thus dependent on the equilibrium potential (V^0_{bond}) and a function that describes the deformation of the spring from r_0, and the spring constant k_{bond}.

$$V_{bond} = V^0_{bond} + k_{bond}(r - r_0)^2 \tag{3.6}$$

The treatment of the chemical bond as a spring is approximate. V_{bond} as described by Eq. 3.6 is symmetric about r_0 and is thus a simple *harmonic* function. Consequently, the spring model matches the steep ascent of the potential energy profile for compressing a chemical bond, as described by the quantum mechanical model. However, it also defines a steep potential for bond stretching and therefore does not allow the extension to long distances that ultimately leads to dissociation. Thus, this approximation depicts molecules more tightly bonded than they really are. The harmonic spring model obviously could not be applied in simulating a true chemical reaction, such as those catalyzed by enzymes. However, for macromolecules at or near equilibrium, where the atomic fluctuations are small, this is a good approximation.

In addition to being stretched and compressed, a bond can be bent and twisted. The lateral bending of a bond is not explicitly treated in most molecular mechanics force fields. For a *three-atom center* held by two bonds, bending falls into the category of deformations to the bond angle. The bond angle θ is defined as the angle between three bonded atoms A—B—C (Figure 3.3). We can think about a deformation to the bond angle as a compression or extension of a spring that connects atom A to atom C. Thus, the potential energy function V_θ for the bond angle can be

treated in a manner similar to V_{bond} with θ_0 defining the equilibrium bond angle and k_θ the spring constant for deforming this angle.

$$V_\theta = V_\theta^0 + k_\theta(\theta - \theta_0)^2 \tag{3.7}$$

Alternatively, since A and C are the number 1 and 3 atoms of the three-atom center, the distance between the two atoms is called the 1–3 distance ($r_{1\text{-}3}$), and the potential for the bond angle is often referred to as the 1–3 potential ($V_{1\text{-}3}$). $V_{1\text{-}3}$ is defined identically as V_θ, where r_θ is the equilibrium distance for the standard bond angle, and $k_{1\text{-}3}$ is the spring constant for bending the bond angle.

$$V_{1\text{-}3} = V_{1\text{-}3}^0 + k_{1\text{-}3}(r - r_0)^2 \tag{3.8}$$

The twisting of a bond defines the dihedral angle ϕ around the central bond between atoms B and C of the four-atom center A—B—C—D (Chapter 1). The potential energy function for ϕ, V_ϕ, depends on the type of the bond connecting B to C. Single bonds, for example, are relatively free to rotate, while double bonds have very distinct energy minima at $\phi = 0°$ and 180°. V_ϕ is not treated as a simple harmonic spring function, but takes the form of the periodic function.

$$V_\phi = V_\phi^0 + V_n \cos\left(\gamma + \frac{\phi}{n}\right) \tag{3.9}$$

In Eq. 3.9, V_n is the torsion force constant (equivalent to a spring constant), n is the period of the function, and γ is the phase angle that defines the position of the minima. For a single bond, the function defines three minima, at $\phi = 60°$, 180°, and 300° (Figure 3.4), associated with the *staggered* conformations around the bond. The height of each potential barrier is V_n. For a double bond, there are two minima, at $\phi = 0°$ and 180°. Thus, $n = 2$ and $\gamma = -180°$ in Eq. 3.9. In analogy to the bond

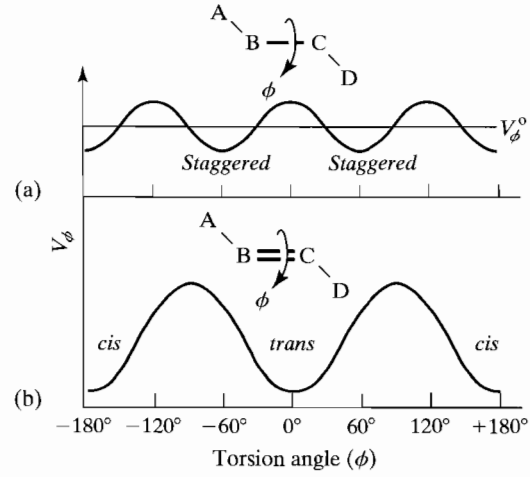


Figure 3.4 Potential energies for rotations about the dihedral angle ϕ for a single bond (a) and a double bond (b). Curves were calculated from Eq. 3.9, with $n = 3$ and $\gamma = 0°$ for a single bond and $n = 2$, and $\gamma = \pm 180°$ for a double bond.

angle, the potential for the dihedral angle is often referred to as the 1-4 potential. However, it is more difficult to think of the 1-4 potential in terms of a spring connecting atoms A to D.

These are the minimum definitions of the bonding potentials. Certain force fields include explicit functions to define the planarity of aromatic groups, such as the bases of nucleic acids and the aromatic amino acid side chains. However, much of this can be handled by stringent definitions of the bond angles and dihedral angles. In general, most force fields reduce the bonding interactions between atoms in a molecule to that of simple harmonic springs. The potential energies for these interactions are very large, but they do not drive the folding of macromolecules because they are approximately the same for all conformations of a molecule. The conformations of macromolecules are defined by the weaker interactions between nonbonded atoms.

3.2.4 Nonbonding Potentials

The nonbonding potentials define all of the interactions that are not directly involved in covalent bonds. We describe these briefly in Chapter 1. Here, we provide a more detailed discussion of the potential energy functions for each interaction. The two broad categories of noncovalent interactions are the intermolecular interactions (those between molecules, and between a molecule and the solvent) and the intramolecular interactions (those between the atoms or groups of atoms within a single molecule). Both types of interactions include charge-charge, dipolar, dispersion, and steric interactions. The potential energy functions for nonbonding interactions (Table 2.1) have two common features. First, they are distance dependent and, second, the long and medium range interactions (electrostatic and dipolar) are strongly dependent on the polarizability of the intervening medium, as measured by the dielectric constant ($D = 4\pi\varepsilon_0$).

The potential energy functions for the nonbonding interactions are inversely related to some power n of the distance (r) between atoms (as in $1/r^n$). The range at which a particular interaction becomes dominant depends on n. For large r, $1/r^n$ approaches zero more rapidly for higher values of n. Conversely, for small r, $1/r^n$ approaches ∞ more rapidly for higher values of n (Figure 3.5). Thus, functions that depend on high powers of r (where n is large) are short-range interactions, while those with low powers of r (n is small) are long-range interactions.

3.2.5 Electrostatic Interactions

The treatment of the electrostatic potential V_e between two unit charges Z_1 and Z_2 is given by Coulomb's law

$$V_e = \frac{Z_1 Z_2 e^2}{Dr} \tag{3.10}$$

The interaction is directly proportional to the product of the two charges (the charge of a proton, $e = 1.602 \times 10^{-19}$ C), and inversely proportional to the dielectric

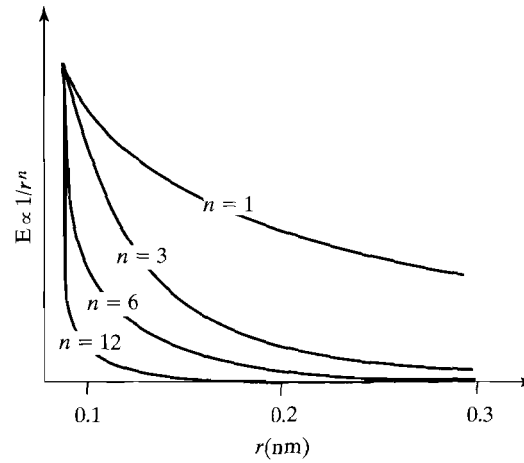


Figure 3.5 Potential energy functions inversely related to r^n. As the distance r increases, functions that are dependent on higher powers (n) of r approach E = 0 more rapidly and are therefore shorter-range interactions. Electrostatic interactions have an $n = 1$ relationship, while dipole-dipole interactions have $n = 3$, London dispersion forces have $n = 6$, and steric repulsion forces have $n = 12$.

constant of the medium D and the distance separating the two charged species r. V_e is thus defined for charges that are paired—it is a *pairwise* interaction. The potential for single isolated charges in a dielectric medium is modeled by the self-energy (see Chapter 1 for a discussion of the self-energy). It should be noted that Coulomb's law as expressed in Eq. 3.10 does not explicitly account for shielding of the charges from counterions in solution. Methods to treat the more complex electrostatic potentials of macromolecules in solution are discussed later in this chapter. However, the simple potential function in Eq. 3.10 is the form typically incorporated into molecular mechanics force fields.

The effect of the dielectric constant D on electrostatic interactions is discussed in some detail in Chapter 1. An accurate treatment of the dielectric constant allows molecular mechanics force fields to account for the effects of a solvent on molecular structure without explicitly incorporating solvent atoms into the model. There are many strategies for assigning the dielectric constant to a macromolecule such as a globular protein. One approach is to define a boundary that distinguishes the interior from the exterior of the protein. The dielectric constant for exposed atoms can then be set to a value similar to that of the solvent. If the solvent is water, $D = 78.5\kappa\varepsilon_0$ at the exterior of the protein. The interior of the protein is then treated as a low-dielectric cavity. Typical values for the dielectric constant for the interior of a protein range from $1\kappa\varepsilon_0$ to $20\kappa\varepsilon_0$, with a good approximation being $3.5\kappa\varepsilon_0$. This is, of course, a rough estimate, since the true dielectric character must vary continuously throughout the molecule. Therefore, other, more sophisticated models have been sought.

One approach is to treat the dielectric constant as a distance-dependent variable. This strategy is based on the assumption that two interacting atoms are likely to be separated by a polarizable medium in the intervening space at long distances, while two closely spaced atoms will have fewer intervening polarizable atoms. A simple function to describe a distance-dependent dielectric is

$$D = f(r)\kappa\varepsilon_0 \tag{3.11}$$

From this definition, the dielectric constant approaches that for a vacuum at close distances (or zero, if the distances between atoms are allowed to approach $r = 0$), while at long distances it approaches that of water. At intermediate distances, the dielectric constant would be estimated to be somewhere between the two extremes.

Alternatively, the dielectric constant can be described as a local function of the protein density ρ at each point and the dielectric constants of water D_W and the protein D_P.

$$D = (1 - \rho)D_W + \rho D_P \tag{3.12}$$

The assumption here is that the highly compact core of a folded molecule excludes water and is thus more dense, while less dense regions will be mixtures of protein and solvent. The problem is that such a function is difficult to incorporate into a standard force field for atomic interactions, since density is a gross measure of molecular structure. However, this relationship has been used to simulate the gross topology of proteins from simple models (see Chapter 4).

3.2.6 Dipole-Dipole Interactions

A separation of the centers of positive and negative charges (δ_+ and δ_-, indicating full and partial charges) in a group give rise to a dipole. This is characterized by a dipole moment, $\boldsymbol{\mu}$, a vector quantity whose magnitude is given by the product δr, where r is the charge-charge distance. The direction of the vector is conventionally taken from δ_- to δ_+. The Coulombic interaction between two dipoles can be approximated by considering only the distance between the dipoles and the dielectric constant of the medium separating two dipoles. The significance of direction is illustrated by considering the interaction between two dipoles oriented in different directions (Figure 3.6). In this analysis, we consider two dipole moments separated by a distance vector $\mathbf{r}$. A simple potential energy function for dipole-dipole interaction is given in Table 2.1.

If the two dipoles lie side by side, their moments can be oriented either parallel or antiparallel to each other. In the antiparallel orientation, where the positive ends interact with the negative ends, we would expect an attractive force (or a negative potential). In contrast, we would expect a repulsive force (or a positive potential) for the parallel orientation of the dipoles. The dipole-dipole interaction is

$$V_{\text{dd}} = \frac{\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2}{D|\mathbf{r}|^3} - \frac{3(\boldsymbol{\mu}_1 \cdot \mathbf{r})(\boldsymbol{\mu}_2 \cdot \mathbf{r})}{D|\mathbf{r}|^5} \tag{3.13}$$

which depends on the orientation of the two dipole moments relative to the distance vector. If the two dipoles are oriented either parallel or antiparallel to each other, but are arranged side by side, then both $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ will be perpendicular to $\mathbf{r}$. Thus, $\boldsymbol{\mu}_1 \cdot \mathbf{r} = 0$ and $\boldsymbol{\mu}_2 \cdot \mathbf{r} = 0$, and Eq. 3.13 reduces to

$$V_{\text{dd}} = \frac{\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2}{D|\mathbf{r}|^3} \tag{3.14}$$

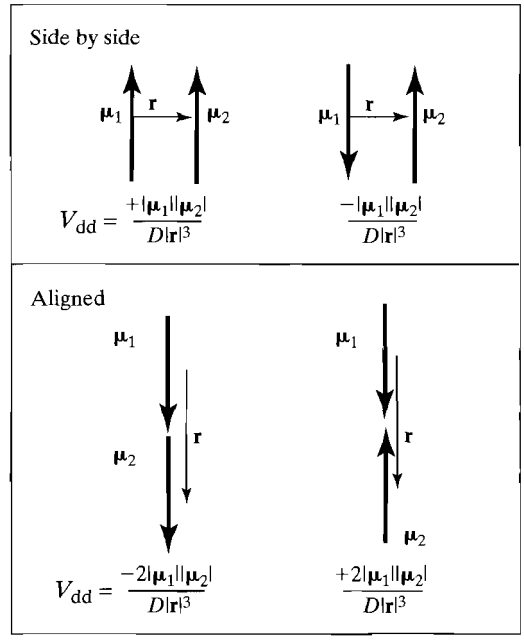


Figure 3.6 Potential energy functions for dipole-dipole interactions. Dipoles that are arranged side by side in parallel and antiparallel directions have dipole moments that are perpendicular to the distance vector **r**. The potential energy is calculated by Eq. 3.14. The potential energies of the head-to-tail and head-to-head alignments of dipoles, however, must be evaluated using the more general relationship in Eq. 3.13.

V_{dd} for dipoles that are aligned parallel and antiparallel are $-2|\boldsymbol{\mu}_1||\boldsymbol{\mu}_2|/D|\mathbf{r}|^3$ and $+2|\boldsymbol{\mu}_1||\boldsymbol{\mu}_2|/D|\mathbf{r}|^3$, respectively, from Eq. 3.13.

Equation 3.13 gives us an intuitive understanding of dipole-dipole interactions. However, in many force fields, V_{dd} is incorporated into the potential function for electrostatic interactions by treating each atom as a monopole having a defined partial valence (Table 3.2). Then Coloumb's law in Eq. 3.10 can be used directly to evaluate the interaction between the individual atoms that constitute the dipole. This approach also treats charge-dipole interactions without the need for a separate function in the force field. The interaction between charges or permanent dipoles with induced dipoles are not normally treated separately in molecular mechanics force fields. Induced dipole-induced dipole interactions are included in the van der Waals interactions, which are discussed in Section 3.2.7.

3.2.7 van der Waals Interactions

Two noble gas atoms will attract each other, although neither has a permanent charge or dipole moment. The attractive force derives from an instantaneous and short-lived imbalance in the electron distribution of an atom that generates a temporary dipole. This temporary dipole induces the electron distribution of a neighboring atom to polarize in order to minimize electron-electron repulsion between the atoms. The resulting synchronous interaction is thus an induced dipole-induced

Table 3.2 Examples of Partial Charges of Atoms in Proteins Calculated from Quantum Mechanics

Amino Acid	Atom Type	Charge
Backbone	N	−0.36
	H_N	+0.18
	C_α	+0.06
	H_α	+0.02
	C	+0.45
	O	−0.38
Ser	C_β	+0.13
	H_β	+0.02
	O_γ	−0.31
	H_γ	+0.17
Tyr	O_η	−0.33
	H_η	+0.17
Cys	S_γ	+0.01
	H_γ	+0.01

Source: From Momany et al. (1975), *J. Phys. Chem.* **79**, 2361–2381.

dipole interaction. London (1937) showed that this attraction, known as *London dispersion forces*, is a natural consequence of quantum mechanics.

The magnitude of the attractive potential is dependent on the volume and the number of polarizable electrons in each interacting group. The London dispersion potential V_L between two uncharged atoms is

$$V_L = -\frac{3I\alpha_1\alpha_2}{4r^6} \tag{3.15}$$

where I is the ionizing energy, and α_1 and α_2 are the polarizability of each atom. The inverse relationship at r^6 makes this a very short-range interaction, with the attraction between atoms dropping off dramatically for even a small increase in distance. All atoms are polarizable to some extent and therefore show this short-range attractive interaction. However, for groups that have a permanent charge or dipole moment, this interatomic attraction is dwarfed by the larger electrostatic interactions at longer distances.

With all atoms attracted to each other, we would expect r to approach zero to minimize the potential energy. We know, however, that this cannot be the case. Counteracting this attraction is a repulsive force, acting at extremely short distances, that keep atoms at respectable distances. At the atomic level, this is associated with the repulsion of electrons clouds and, to a lesser extent, from nucleus-nucleus repulsion. These two repulsions dominate the potential energy function for two atoms at closest approach.

The simpest model to accommodate the repulsive force is the *hard sphere approximation*, which treats each atom as an impenetrable spherical volume. However, this model is too stringent to accurately represent the behavior of atoms. A

more accurate model is to define a repulsive potential V_R that acts at extremely short ranges. A typical function is

$$V_R = \frac{k}{r^m} \tag{3.16}$$

where m is some value between 5 and 12. The function approaches that of a hard sphere for the typical value of $m = 12$.

Together, the attractive London dispersion and the repulsive potentials produce an equilibrium distance at which the two opposing energies become equal. The van der Waals radii r_{vdw} for neutral atoms are defined so they sum to give the equilibrium distance (Table 3.3). The two interactions are treated as a single van der Waals potential V_{vdw}. The function that best describes this balance between attraction and repulsion is thus

$$V_{vdw} = \frac{A}{r^m} - \frac{B}{r^6} \tag{3.17}$$

where A and B are constants that describe the magnitude of the repulsive and attractive terms, respectively, and m is the power of the repulsive term (usually between 5 and 12). In a *6-12 potential* or *Lennard-Jones potential,* $m = 12$. This approaches the hard sphere model for steric interactions (Figure 3.7).

At r_{vdw}, the net force on the two atoms is zero. Thus, A and B can be estimated from Eqs. 3.3 and 3.17 for any pair of atoms (Table 3.4). The steep increase in V_{vdw} at distances that are significantly shorter than the sum of the r_{vdw} for two atoms is overcome only by formation of a chemical bond.

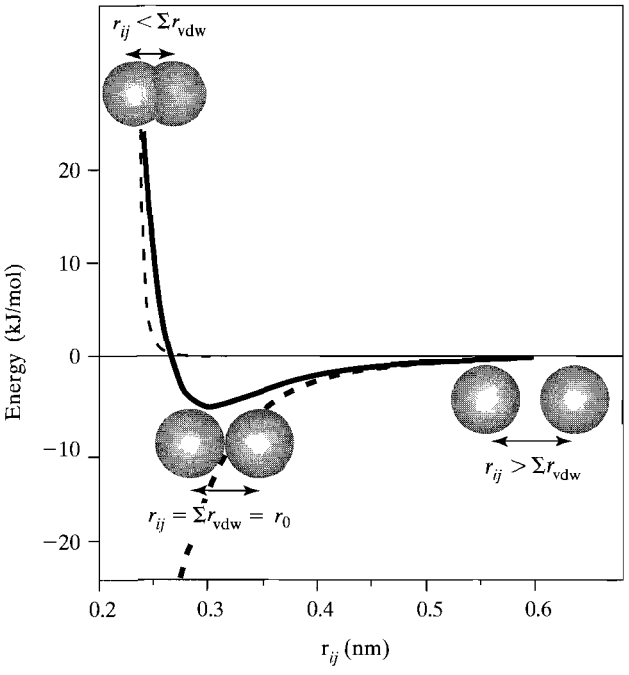


Figure 3.7 The van der Waals potential is a sum of the very short-range attraction between atoms (London dispersion forces) and the extremely short-range steric repulsion between atoms. Together, the two functions define an optimal distance r_0, which is sum of the van der Waals radii r_{vdw} of the two atoms.

Table 3.4 Coefficients for the Repulsive (A) and Attractive (B) Terms of the 6-12 van der Waals Potential in Eq. 3.17

Atomic Interaction	A (kJ-nm^{12}/mol)	B (kJ-nm^6/mol)	r_0 (nm)
H···H	1.84×10^{-8}	1.92×10^{-4}	0.240
H···C	1.57×10^{-7}	5.27×10^{-4}	0.290
H···N	1.11×10^{-7}	5.15×10^{-4}	0.275
H···O	1.03×10^{-7}	5.11×10^{-4}	0.272
H···P	6.35×10^{-7}	1.43×10^{-3}	0.310
C···C	1.18×10^{-6}	1.52×10^{-3}	0.340
C···N	8.90×10^{-7}	1.51×10^{-3}	0.325
C···O	8.49×10^{-7}	1.51×10^{-3}	0.322
C···P	4.49×10^{-6}	4.12×10^{-3}	0.360
N···N	6.63×10^{-7}	1.50×10^{-3}	0.310
N···O	6.30×10^{-7}	1.50×10^{-3}	0.307
N···P	3.44×10^{-6}	4.08×10^{-3}	0.345
O···O	5.97×10^{-7}	1.51×10^{-3}	0.304
O···P	3.28×10^{-6}	4.10×10^{-3}	0.342
P···P	1.68×10^{-5}	1.12×10^{-2}	0.380

The equilibrium distance r_0 is the sum of the van der Waals radii of the two interacting atoms.
Source: From F. Jordan (1973), *J. Theor. Biol.* **30**, 621–630.