

6.11 The energy required to break interatomic interactions in folded macromolecules gives rise to the peak in heat capacity

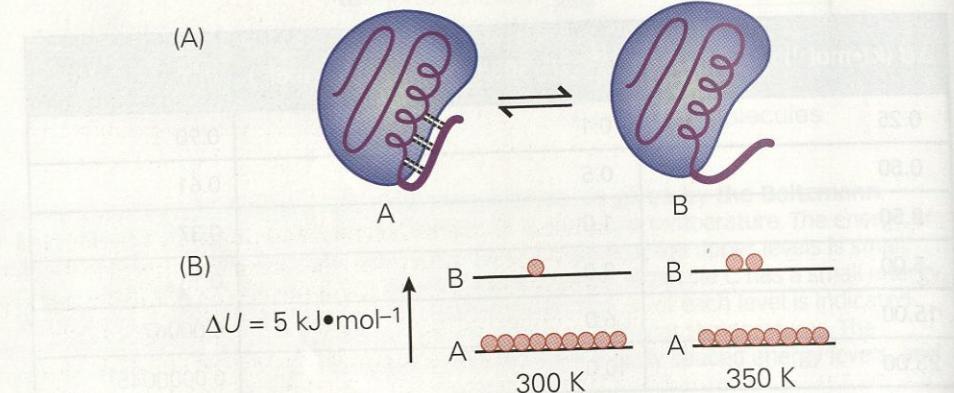
The energy available from collisions at room temperature can excite molecular vibrations and conformational changes, which refer to all the different kinds of fluctuations in the internal structure of the molecules. We can estimate the energy required to excite the vibrations of individual covalent bonds because we know that these are excited by infrared radiation, with wavelengths around 2000 nm. This corresponds to $\sim 60 \text{ kJ}\cdot\text{mol}^{-1}$ ($\sim 24 \text{ k}_B T$ at 300 K). The energy difference between the lowest energy level and the first excited state for bond vibrations must be comparable. Thus, covalent bonds are very stiff, and their vibrations are unlikely to be excited at room temperature.

The key to understanding the heat capacity changes in complex molecules like proteins and DNA is to recognize that they are held together by numerous weak interactions, such as hydrogen bonds and van der Waals interactions. These interactions stabilize the folded structure collectively, but each individual interaction is easily broken. The energy required to break each interaction contributes to the increase in heat capacity as the protein unfolds.

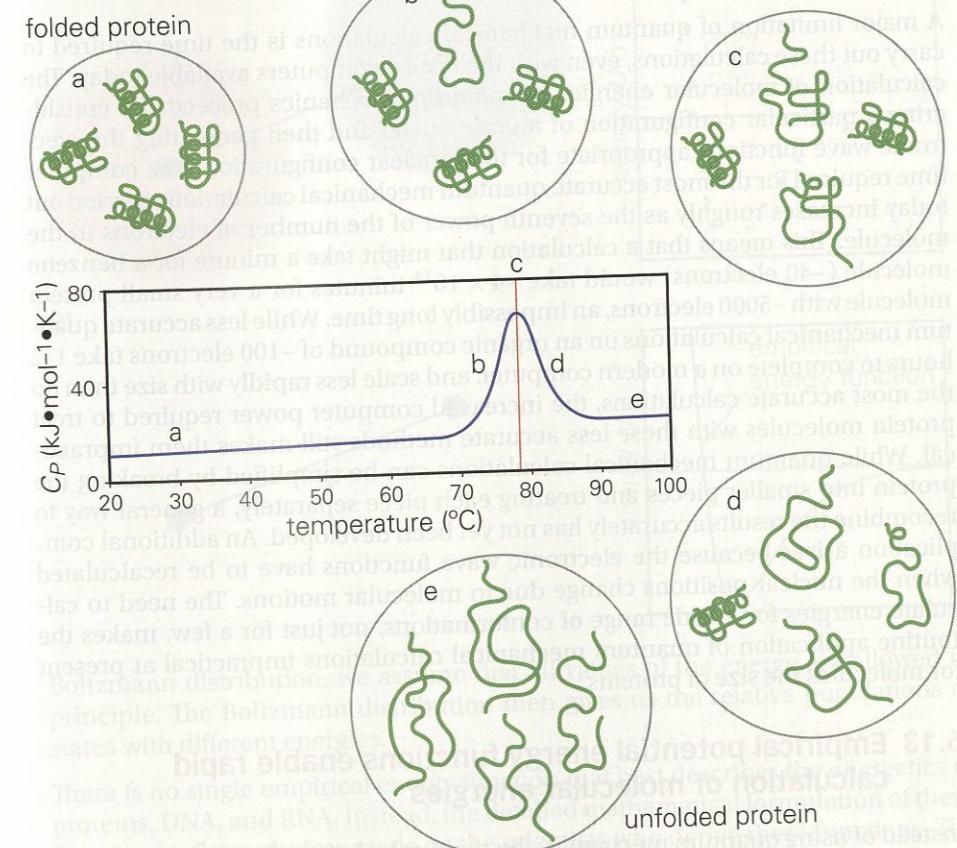
We can now understand the origin of the characteristic heat capacity changes of proteins and DNA as follows. At low temperature, the molecules are folded into stable three-dimensional structures. As the temperature is increased, the structures undergo cooperative unfolding transitions, when the many weak interactions that hold the structure together begin to break. The effect of temperature on the unfolding of a small part of a protein is shown in Figure 6.17, in which states A and B denote the fully folded conformation and a partially unfolded conformation of the protein. We assume that the energy of the protein is $5 \text{ kJ}\cdot\text{mol}^{-1}$ higher when the interactions made by this small region are disrupted in state B. The population of state B increases from $\sim 10\%$ to 20% as the temperature increases from 300 K to 350 K. The increase in population of state B requires energy, which manifests itself as an increase in the heat capacity of the protein solution.

Although the increase in population of the partially unfolded state shown in Figure 6.17 may seem small, there are many such interactions that hold the protein together. Once a small number start to break, other interactions rapidly become unstable, and the protein unfolds cooperatively. Once all of the interactions are broken (that is, the protein is completely unfolded), energy no longer goes into the unfolding process, and the heat capacity is reduced. The maximum value of the heat capacity occurs at the melting temperature, when the folding process is halfway towards completion (Figure 6.18).

Figure 6.17 Protein molecules take up energy as they unfold.
 (A) A schematic drawing of a protein molecule in two states, denoted A and B. A corresponds to the fully folded conformation and B corresponds to a partially unfolded one. Interactions that are made in state A but broken in state B are indicated by dashed lines.
 (B) An energy diagram showing the relative populations of the two states at 300 K and at 350 K. State B is higher in energy by $5 \text{ kJ}\cdot\text{mol}^{-1}$ than state A. The circles indicate the relative populations of the two states at the two temperatures, as given by the Boltzmann distribution.



heat is taken up by the molecules as they unfold: higher heat capacity



C. ENERGETICS OF INTERMOLECULAR INTERACTIONS

6.12 Simplified energy functions are used to calculate molecular potential energies

In order to understand what happens to molecules as they take up and release energy, we need to be able to relate the structure of a molecule to its potential energy. Molecular potential energies can be calculated from first principles if we use quantum mechanics, but these calculations require an enormous amount of computer time. A typical protein or RNA molecule contains thousands of atoms and, even with the fastest computers available today, it is difficult to use quantum mechanical methods for the accurate calculation of potential energies for molecules of this size. In practice what is done instead is to use mathematical expressions known as **empirical potential energy functions**, which are highly simplified functions that allow approximate molecular energies to be calculated from knowledge of the three-dimensional coordinates of all the atoms in a molecule. The energy functions are also used to calculate forces, which are obtained by computing the first derivative of the potential energy with respect to atomic positions.

In quantum mechanical calculations, all of the electrons and nuclei in a molecule are considered explicitly, and probability functions that describe the distribution of electrons around nuclear positions are obtained by solving Schrödinger's equation on a computer. All the properties of the molecule, including its covalent bonded structure and the potential energy, can be derived accurately from the electronic wave functions that correspond to different configurations of the

Figure 6.18 Heat capacity and protein unfolding. The heat capacity curve shown in Figure 6.11 is redrawn here, along with an interpretation of the molecular behavior underlying the changes in the heat capacity. At lower temperatures ($< 40^\circ\text{C}$), the protein molecule is mainly folded. As the temperature is increased the protein begins to unfold and energy goes into breaking the numerous interactions that stabilize the folded structure. This manifests itself as an increased heat capacity. Once most of the protein molecules have unfolded the heat capacity decreases.

Empirical potential energy functions

These are relatively simple mathematical expressions that allow us to calculate the potential energy of a molecule or a collection of molecules, given the conformation (that is, the internal structure) of each molecule and their relative positions. Empirical potential energy functions are approximations to the true quantum mechanical energy of the system.

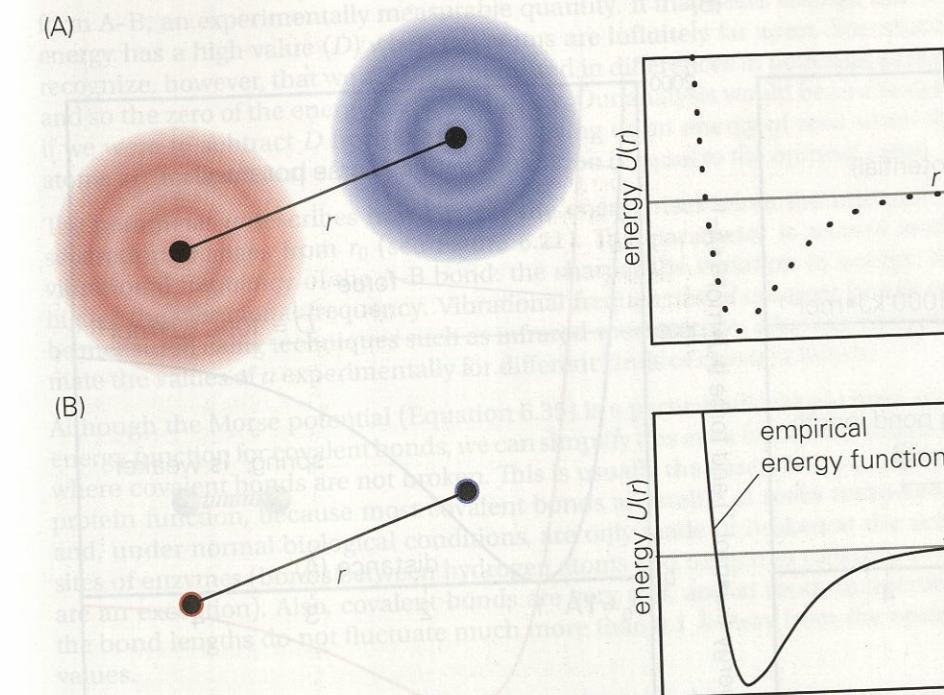
nuclei. This provides the most detailed and accurate information about the structure and electronic properties of molecules and forms the basis for all descriptions of molecular energies.

A major limitation of quantum mechanical calculations is the time required to carry out these calculations, even with the fastest computers available today. The calculation of molecular energies by quantum mechanics proceeds by considering a particular configuration of atomic nuclei and then generating the electronic wave functions appropriate for that nuclear configuration. The computer time required for the most accurate quantum mechanical calculations carried out today increases roughly as the seventh power of the number of electrons in the molecule. This means that a calculation that might take a minute for a benzene molecule (~40 electrons) would take $\sim 4 \times 10^{14}$ minutes for a very small protein molecule with ~5000 electrons, an impossibly long time. While less accurate quantum mechanical calculations on an organic compound of ~100 electrons take 1–2 hours to complete on a modern computer, and scale less rapidly with size than do the most accurate calculations, the increased computer power required to treat protein molecules with these less accurate methods still makes them impractical. While quantum mechanical calculations can be simplified by breaking the protein into smaller pieces and treating each piece separately, a general way to recombine the results accurately has not yet been developed. An additional complication arises because the electronic wave functions have to be recalculated when the nuclear positions change due to molecular motions. The need to calculate energies for a wide range of conformations, not just for a few, makes the routine application of quantum mechanical calculations impractical at present for molecules the size of proteins.

6.13 Empirical potential energy functions enable rapid calculation of molecular energies

Instead of using quantum mechanics, the rapid calculation of molecular energies for large molecules, such as proteins and nucleic acids, relies on highly simplified descriptions of the relationship between molecular geometry and potential energy. These calculations do not treat electrons explicitly (the effects of the electrons are embedded in the form of the energy functions), and consider molecular structure only in terms of nuclear positions. These relationships are called empirical potential energy functions because they are not derived from first principles, but are instead approximate representations of the empirically observed behavior of molecules in terms of simple mathematical expressions that can be very quickly calculated using computers. The approximations introduced have a price, which is that the gain in speed is offset by a loss in accuracy. The highly accurate calculation of the energies of protein molecules remains a problem that has no really satisfactory solution at the present time. Despite these limitations, empirical energy functions are the only practical way to describe the energies of macromolecules and the forces within them.

While quantum mechanics describes the properties of matter in terms of wave functions, classical or Newtonian mechanics considers the movement of particles or larger objects in terms of potential energies and forces, which depend on the distances between the particles (Figure 6.19). In principle, quantum mechanical calculations on small pieces of protein molecules tell us how their energy changes as a function of the distances between the various atomic nuclei. Such calculations can then be used to derive the form of simple energy functions that approximate the energy as a function of nuclear positions. These energy functions are then used within the framework of classical mechanics to calculate structural and dynamical properties of the molecule. We also ignore the quantization of energy that is a fundamental aspect of quantum mechanics. Because we are treating low-energy vibrations and interactions, this turns out not to be a bad approximation. Thus, instead of considering discrete energy levels as we did in discussing the



Boltzmann distribution, we assume that all values of the energy are allowed in principle. The Boltzmann distribution then gives us the relative populations of states with different energies.

There is no single empirical energy function that best describes the energetics of proteins, DNA, and RNA. Instead, the detailed mathematical formulation of these functions reflects choices made by the scientists who derive these functions. The simplest empirical energy functions typically include mathematical terms that describe the change in the energy of a protein molecule as its covalent bonds and angles are deformed, as the molecule changes its conformation by rotating about its covalent bonds, and as clusters of atoms arranged in particular geometries, such as planar groups, are distorted from their equilibrium configurations. In addition to these terms, which depend on interatomic distances between covalently bonded clusters of atoms, the energy functions also include **noncovalent energy terms** that describe changes in the energy as atoms that are not covalently bonded approach each other (Figure 6.20). These include the van der Waals interactions between atoms and electrostatic interactions, such as ion pairs and hydrogen bonds.

6.14 The energies of covalent bonds are approximated by functions such as the Morse potential

To illustrate how empirical energy functions are formulated, consider two atoms, labeled A and B, that are covalently bonded to each other (Figure 6.21). We can calculate the energy of such a pair of A and B atoms using quantum mechanics. These calculations show us that the atoms experience a repulsive force when the atoms approach each other closely, due to overlap between their electronic orbitals. The repulsive force decreases rapidly when the distance between the atoms is increased. The interatom force becomes zero at a distance corresponding to the optimal bond length, r_0 . If the distance between the atoms is increased beyond r_0 , the atoms experience an attractive force until the bond between them is broken and the force between them again becomes zero. In summary, the potential energy of this two-atom system is zero for large interatomic separations, becomes negative for closer distances that correspond to covalent bond formation, and then rises sharply as the atoms move closer together.

Figure 6.19 Calculation of energy in quantum and classical mechanics.
(A) Two interacting atoms are shown, and their wave functions are illustrated schematically. Values of the ground-state energy, calculated using quantum mechanics for specific values of the interatomic distance, r , are shown as dots in the graph.
(B) The empirical energy function for the interaction between the two atoms is graphed. The empirical energy function is a continuous mathematical function that approximates the quantum mechanical energy.

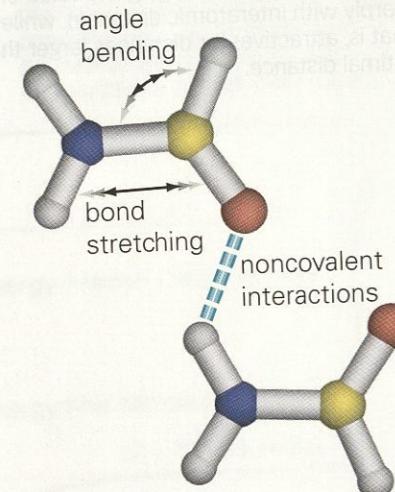


Figure 6.20 Covalent and noncovalent interactions. The deformations of covalently bonded atoms (such as bond stretching) have a high energetic penalty and, as a consequence, the geometry stays close to the optimal values at room temperature. Noncovalent interactions between groups that are not covalently bonded are readily broken.

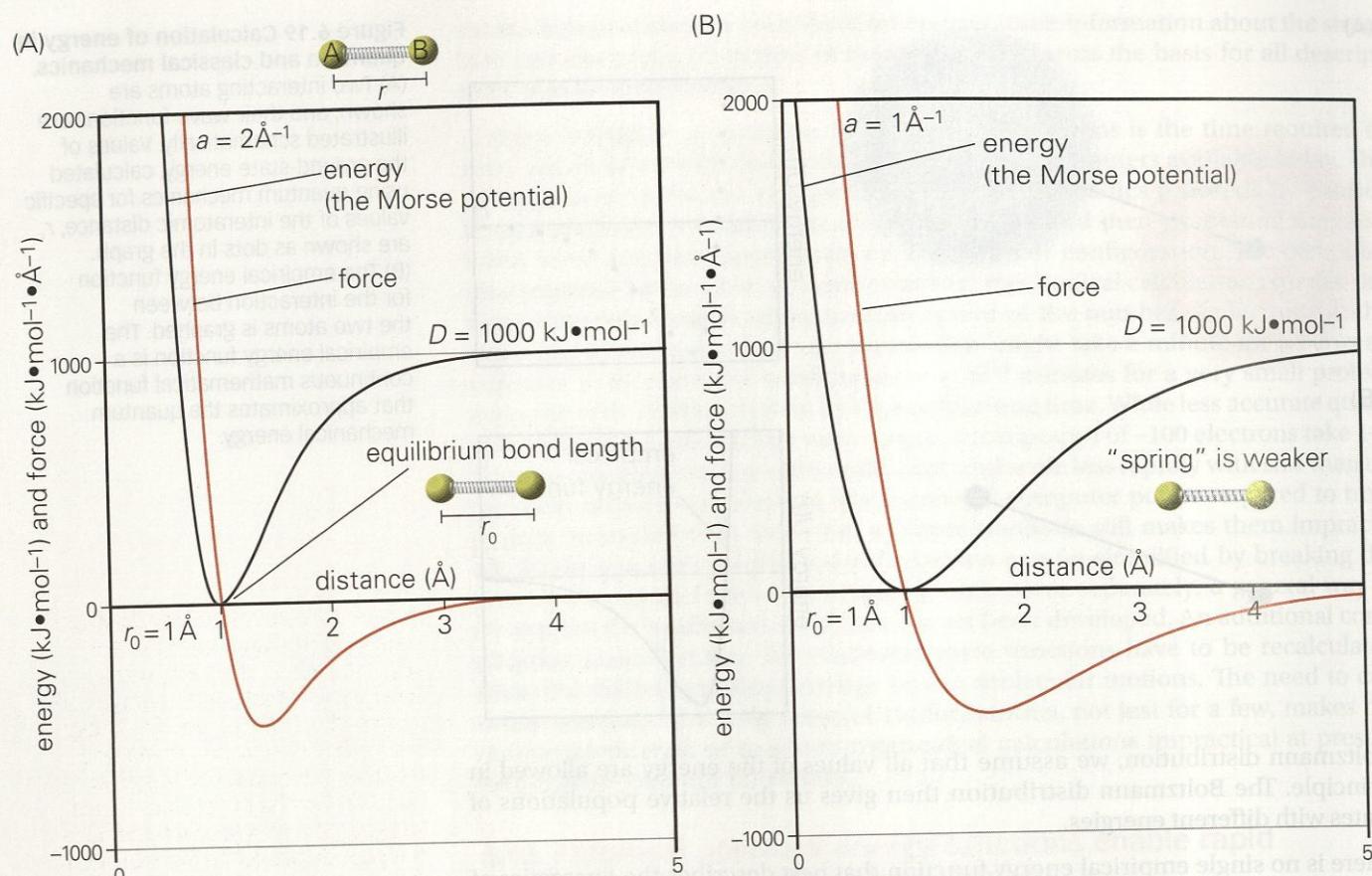


Figure 6.21 The energy and the force on each atom in an A-B covalently bonded pair as a function of the interatomic distance. The energy and force are calculated using the Morse potential (Equation 6.35). For this example, the equilibrium (optimal) distance is $r_0 = 1.0 \text{ \AA}$. (A) The values of the parameters a and D are 2 \AA^{-1} and $1000 \text{ kJ} \cdot \text{mol}^{-1}$, respectively. The parameter D is bond length, r_0 , is 1.0 \AA . (B) The effect of the parameter a on the energy and force is illustrated by keeping all other parameters the same as in (A), but changing the value of the parameter a to 1.0 \AA^{-1} . This results in a "softening" of the spring (the energy rises less sharply with interatomic distance), while the net stabilization of the covalent bond does not change. Note that the force is negative (attractive) for distances larger than the optimal distance (1.0 \AA) and positive (that is, repulsive) for distances shorter than the optimal distance.

The behavior of the interatomic force is modeled by the following simple expression for the potential energy that is based on an exponential function of the interatomic distance, r :

$$U_{\text{bond}}(r) = D [1 - e^{-a(r-r_0)}]^2 \quad (6.35)$$

In Equation 6.35, U_{bond} is the energy of the covalent bond when the interatomic separation is r .

This expression, known as the **Morse potential** (see Figure 6.21), has three adjustable parameters— D , a , and r_0 —that are characteristic of the nature of the covalent bond between A and B. These three parameters are related to the chemical properties of the molecule. For example, when the interatomic separation, r , is equal to r_0 , then:

$$U_{\text{bond}}(r_0) = D [1 - e^{-a(r-r_0)}]^2 = D [1 - e^0]^2 = D \quad (6.36)$$

Thus, the energy is zero when $r = r_0$, and so r_0 is called the **optimal bond length**.

When the atoms are infinitely far apart ($r = \infty$),

$$U_{\text{bond}}(r = \infty) = D [1 - e^{-\infty}]^2 = D \quad (6.37)$$

The parameter D is the amount by which the potential energy of the A-B pair is reduced when the covalent bond is formed. D is therefore related to the enthalpy of making the A-B bond, which is the heat released when A and B atoms fuse to

form A-B, an experimentally measurable quantity. It may seem strange that the energy has a high value (D) when the atoms are infinitely far apart. You should recognize, however, that we are only interested in differences in potential energy, and so the zero of the energy scale is arbitrary. Our analysis would be unchanged if we were to subtract D from the energy, giving us an energy of zero when the atoms are far apart and $-D$ when their separation is equal to the optimal value.

The parameter a describes how sharply the energy rises when the interatomic separation changes from r_0 (see Figure 6.21). This parameter is related to the vibrational frequency of the A-B bond: the sharper the variation in energy, the higher the vibrational frequency. Vibrational frequencies of covalent bonds can be measured using techniques such as infrared spectroscopy, allowing us to estimate the values of a experimentally for different kinds of covalent bonds.

Although the Morse potential (Equation 6.35) is a particularly simple form of an energy function for covalent bonds, we can simplify this even further for situations where covalent bonds are not broken. This is usually the case when we consider protein function, because most covalent bonds are stable at room temperature and, under normal biological conditions, are only made or broken at the active sites of enzymes (bonds between hydrogen atoms and oxygen or nitrogen atoms are an exception). Also, covalent bonds are very stiff, and at room temperature the bond lengths do not fluctuate much more than 0.1 \AA away from the optimal values.

If we ignore chemical reactions and chemical groups that can release or accept protons, we can replace the Morse potential with a simpler form that describes the energy and force associated with stretching the bond in terms of a Hooke's law spring:

$$U_{\text{bond}}(r) = \frac{1}{2} K_b (r - r_0)^2 \quad (6.38)$$

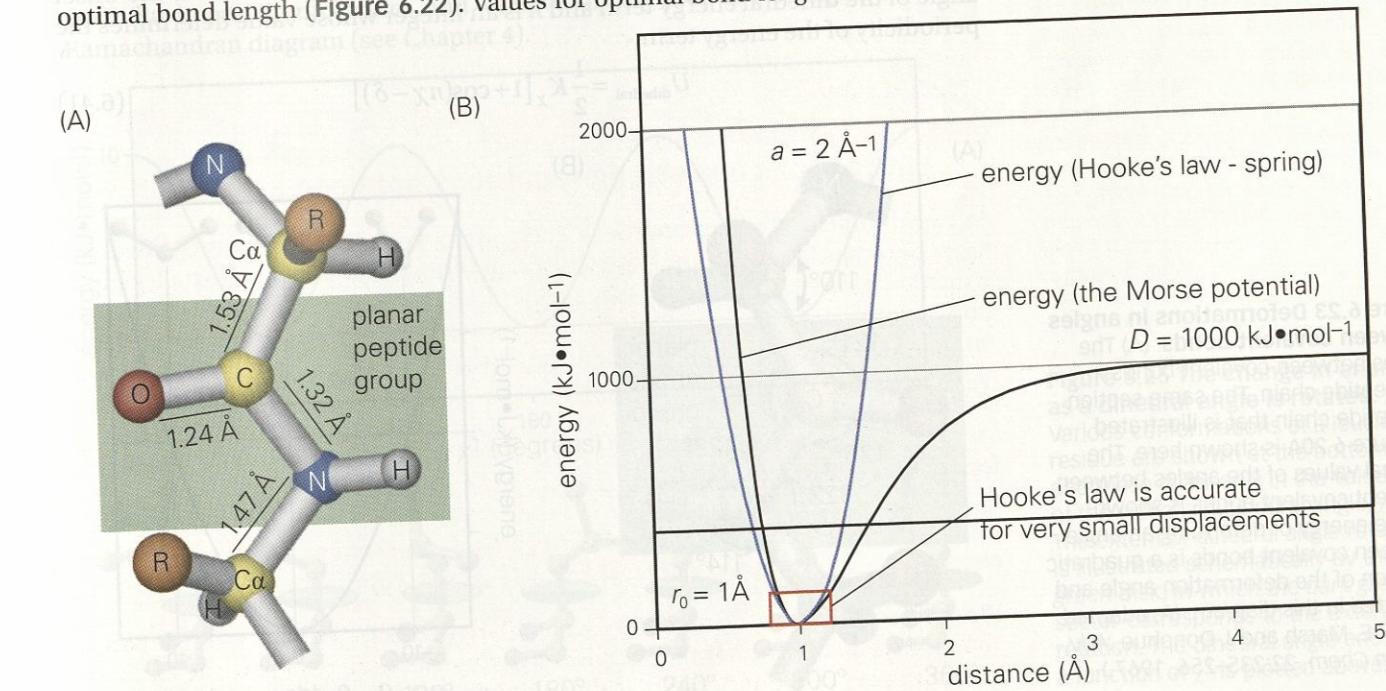
and

$$F(r) = -K_b(r - r_0) \quad (6.39)$$

In these equations, r_0 is the optimal bond length and K_b is known as the **force constant**. We can relate these Hooke's law equations to the Morse potential by choosing a value for the force constant, K_b , such that the Hooke's law expression matches the Morse potential as well as possible for distances that are close to the optimal bond length (Figure 6.22). Values for optimal bond lengths are derived

Figure 6.24 Dihedral angles. (A) Dihedral angles describe rotations about bonds in the structure and can be thought of as the angles between two planes, which are indicated in red. As discussed in Chapter 4, the peptide backbone has two dihedral angles that rotate easily, ϕ and ψ . The dihedral angle, ϕ , is not free to rotate very much. (B) A dihedral angle, referred to as γ^2 , in the sidechain of leucine.

Figure 6.22 Covalent bonds in the peptide backbone. (A) A section of protein chain is shown, with amino acid sidechains in orange and indicated by "R." The peptide linkage between two adjacent amino acid residues is indicated by the green box. The lengths of covalent bonds between the nonhydrogen atoms in the peptide backbone are indicated. (B) A Hooke's law function that matches the Morse potential for small displacements about the optimal bond length is shown. Note that the energy rises sharply without limit as the interatomic separation increases, so this function does not model covalent bond breakage. (A, adapted from R.E. Marsh and J. Donohue, *Adv. Protein Chem.* 22: 235–256, 1967.)



from crystal structures of amino acids and peptides, as shown in Figure 6.22A for the peptide group. The values of K_b for covalent bonds are typically in the range of $\sim 400\text{--}2000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$.

6.15 Other terms in the energy function describe torsion angles and the deformations in the angles between covalent bonds

There are at least two other energetic terms associated with covalent bonding that need to be considered when calculating the potential energy. One term describes changes in the energy of the molecule that arise from angular distortions in the covalently bonded structure. Two adjacent covalent bonds in a molecule are constrained to an angular separation, θ_0 , known as the optimal bond angle (Figure 6.23). If the bond angle is distorted, a spring-like force tends to restore the angular separation between the bonds back to the optimal value. The energy for bond deformations is given by:

$$U_{\text{angle}} = \frac{1}{2} K_\theta (\theta - \theta_0)^2 \quad (6.40)$$

where θ is the value of the bond angle and K_θ is the force constant for deformation of the angle. Like the covalent bonds, the bond angles in a protein are also stiff. The values of K_θ range from 50 to $200 \text{ J}\cdot\text{mol}^{-1}\cdot\text{degree}^{-2}$. Again, as for bond vibrations, the Boltzmann distribution tells us that only very small displacements of bond angles are likely to occur at room temperature. Like the lengths of covalent bonds, the angles between covalent bonds do not change very much when a protein molecule alters its conformation or binds to other molecules.

Another term in the energy function concerns the relative rotation or torsion of two segments of the molecule about a covalent bond. Such a rotation is known as a **dihedral angle rotation**, because its magnitude is given by the angle between two planes formed by adjacent pairs of covalent bonds (a “dihedral” is the angle between two planes; Figure 6.24). Dihedral angle rotations about single bonds are “soft,” in that the barrier to rotation about these bonds is usually not very high relative to the value of $k_B T$. Rotations about the ϕ and ψ angles in the backbone of the polypeptide chain correspond to such “soft” dihedral rotations (see Figure 6.24). For soft dihedral angles, the dependence of the energy on the dihedral angle, χ , is approximated by a term involving $\cos(n\chi - \delta)$, where δ is the offset angle, χ , is the angle of the dihedral energy term and n is an integer whose value determines the periodicity of the energy term:

$$U_{\text{dihedral}} = \frac{1}{2} K_\chi [1 + \cos(n\chi - \delta)] \quad (6.41)$$

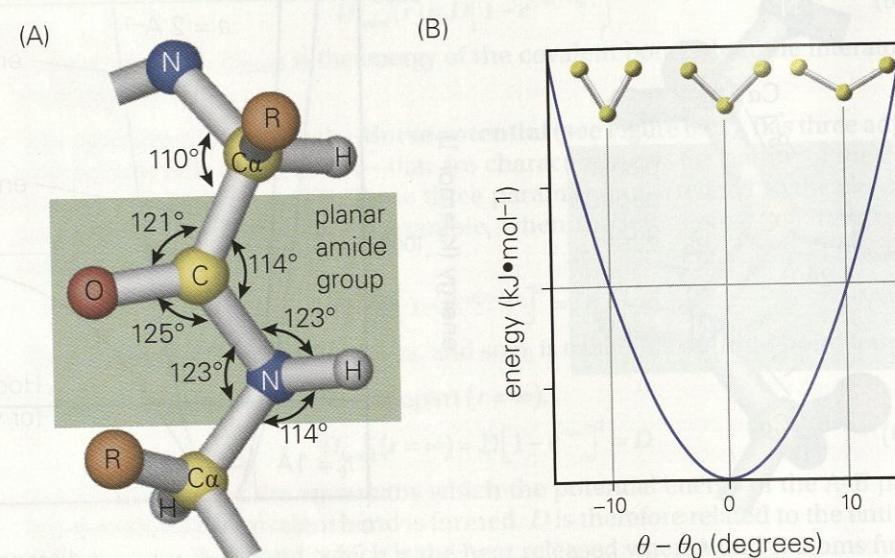
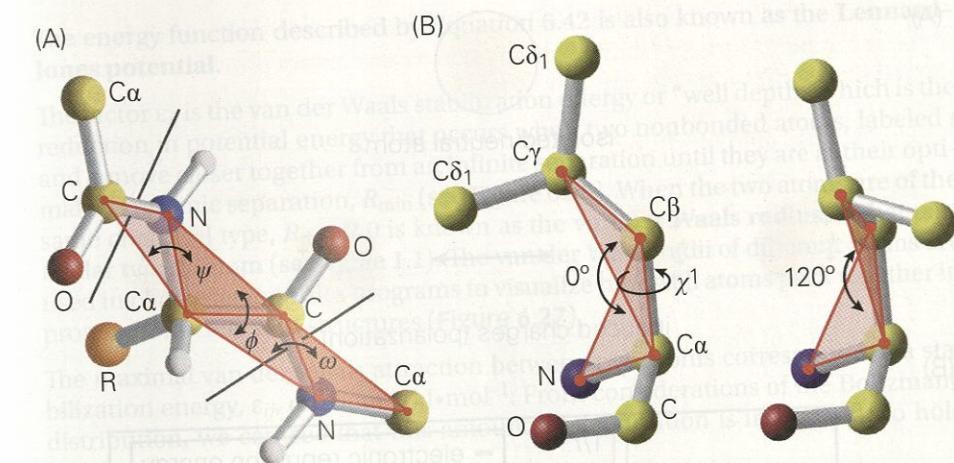


Figure 6.23 Deformations in angles between covalent bonds. (A) The angles between covalent bonds in the peptide chain. The same section of peptide chain that is illustrated in Figure 6.20A is shown here. The optimal values of the angles between adjacent covalent bonds is shown. (B) The energy of deforming an angle between covalent bonds is a quadratic function of the deformation angle and is plotted in this diagram. (A, adapted from R.E. Marsh and J. Donohue, *Adv. Protein Chem.* 22:235–256, 1967.)



The application of Equation 6.41 is illustrated in Figure 6.25, which shows how the value of U_{dihedral} changes as a sidechain dihedral angle in leucine, denoted χ^1 , is varied from 0° to 360° . For this dihedral angle, which corresponds to rotation about the $C_\alpha-C_\beta$ bond, the value of δ is zero, and so the energy has a maximum value when the value of χ^1 is zero. The value of n is 3, and so the energy is maximal for three values of χ^1 ($0^\circ, 120^\circ, 240^\circ$). These correspond to conformations in which the substituents of the C_α and C_β atoms are in eclipsed configurations, and so these values of χ^1 are hindered sterically. There are also three values of χ^1 for which the energy has minimal values ($60^\circ, 180^\circ$ and 300°). The substituents of the C_α and C_β atoms are in staggered configurations for these values of χ^1 and so there is minimal steric overlap. You can work out that the value of the parameter K_χ corresponds to the difference in energy between the eclipsed and staggered conformations. The value of K_χ is in the range of $1\text{--}20 \text{ kJ}\cdot\text{mol}^{-1}$ for dihedral angles in protein sidechains.

As you can see from this example, when $n > 1$, the energy function given by Equation 6.41 has more than one value of the dihedral angle, χ , with the same value for the potential energy. According to the Boltzmann distribution, each of the alternative values of the dihedral angles is equally likely, and the molecule will hop between these different allowed conformations. This is the primary source of flexibility in protein molecules, and we have already seen its consequences in the Ramachandran diagram (see Chapter 4).

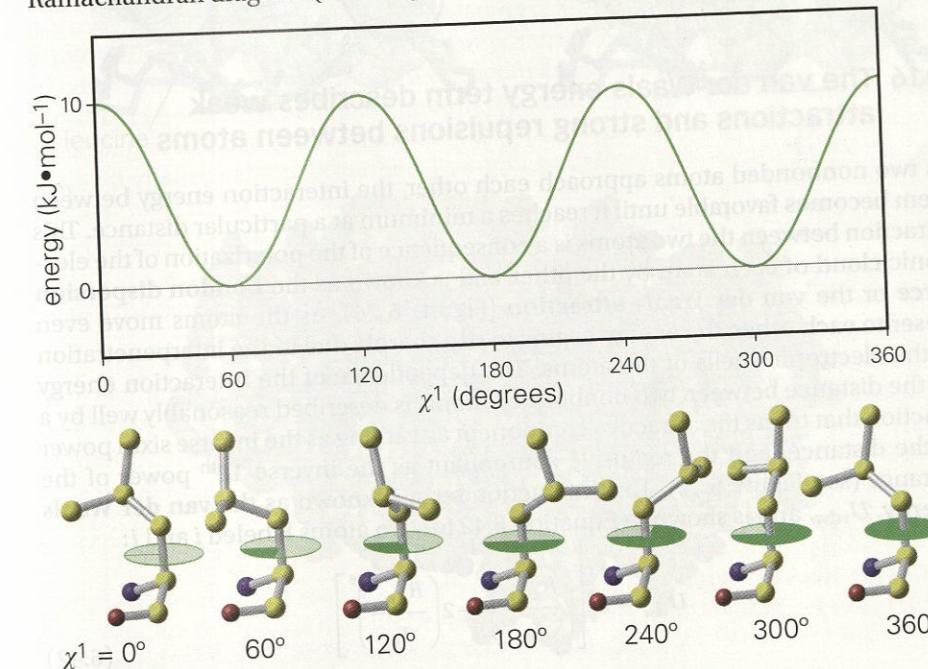
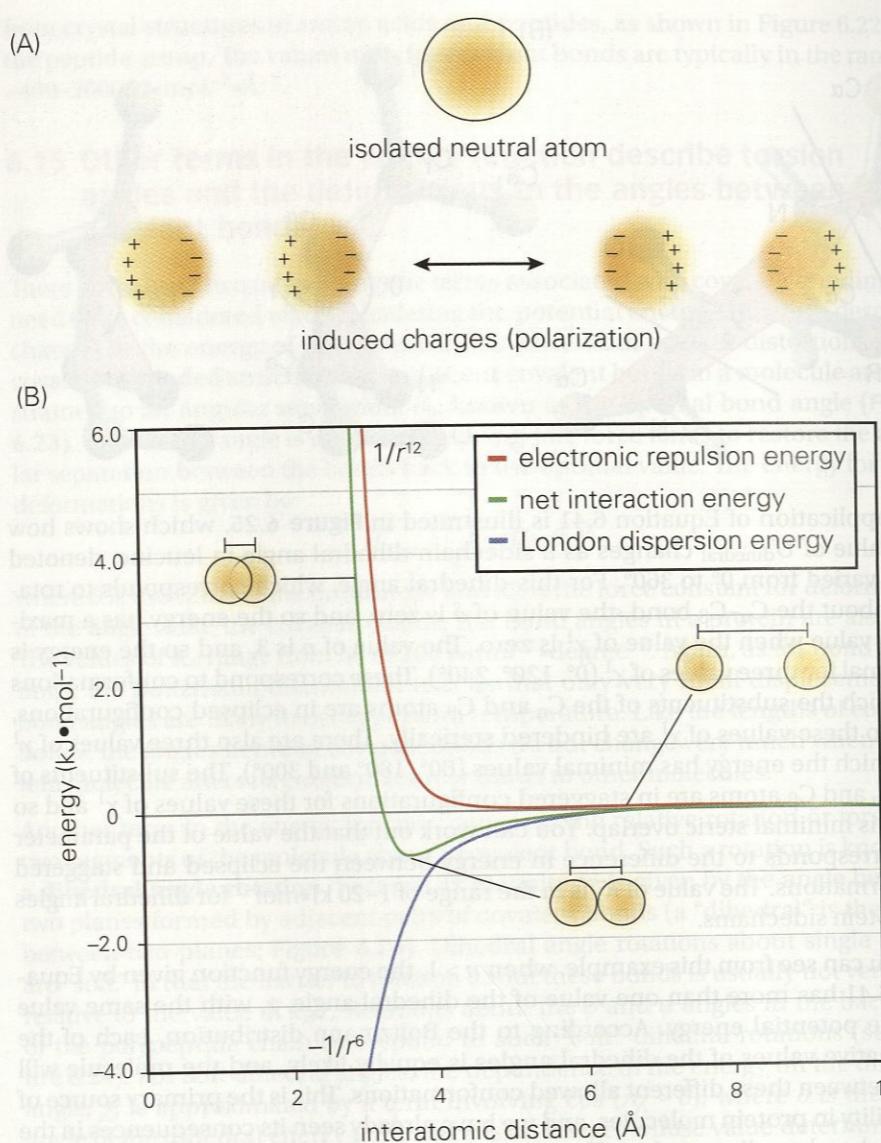


Figure 6.25 The change in energy as a dihedral angle is rotated. Various conformations of a leucine residue are shown at the bottom, and these differ only in the value of the sidechain dihedral angle, χ^1 . The extent of dihedral angle rotation is illustrated schematically by the green disk, in which the dark green sector corresponds to the extent of rotation. The dihedral angle energy as a function of χ^1 is plotted above.

Figure 6.26 van der Waals interactions between atoms.

(A) The attractive part of the van der Waals interaction is a consequence of induced dipoles in atoms. (B) The diagram shows how the van der Waals interaction energy between two atoms (shown as yellow spheres) changes as a function of interatomic distance. The van der Waals energy consists of two terms, a repulsive component that varies as $(1/r^{12})$ and an attractive component that varies at $(-1/r^6)$. These two components are plotted as red and blue lines in the diagram, and the total van der Waals energy is indicated by the green line. Note that the energy scale here is ~1000 times smaller than that shown for the bond energy in Figure 6.22.



6.16 The van der Waals energy term describes weak attractions and strong repulsions between atoms

As two nonbonded atoms approach each other, the interaction energy between them becomes favorable until it reaches a minimum at a particular distance. This attraction between the two atoms is a consequence of the polarization of the electronic cloud of each atom by the other and is known as the **London dispersion force** or the **van der Waals attraction** (Figure 6.26). As the atoms move even closer to each other, the energy begins to rise sharply due to the interpenetration of the electronic shells of the atoms. The dependence of the interaction energy on the distance between two nonbonded atoms is described reasonably well by a function that treats the attractive component as varying as the inverse sixth power of the distance, and the repulsive component as the inverse 12th power of the distance (see Figure 6.26). This interaction term is known as the **van der Waals energy**, U_{vdw} and is shown in Equation 6.42 for two atoms labeled i and j :

$$U_{vdw} = \varepsilon_{ij} \left[\left(\frac{R_{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min}}{r_{ij}} \right)^6 \right] \quad (6.42)$$

The energy function described by Equation 6.42 is also known as the **Lennard-Jones potential**.

The factor ε_{ij} is the van der Waals stabilization energy or “well depth,” which is the reduction in potential energy that occurs when two nonbonded atoms, labeled i and j , move closer together from an infinite separation until they are at their optimal interatomic separation, R_{\min} (see Figure 6.26). When the two atoms are of the same chemical type, $R_{\min}/2.0$ is known as the **van der Waals radius** of that particular type of atom (see Table 1.1). The van der Waals radii of different atoms are used in computer graphics programs to visualize how well atoms pack together in protein or nucleic acid structures (Figure 6.27).

The maximal van der Waals attraction between two atoms corresponds to a stabilization energy, ε_{ij} , of about $1 \text{ kJ}\cdot\text{mol}^{-1}$. From considerations of the Boltzmann distribution, we can see that this amount of stabilization is insufficient to hold

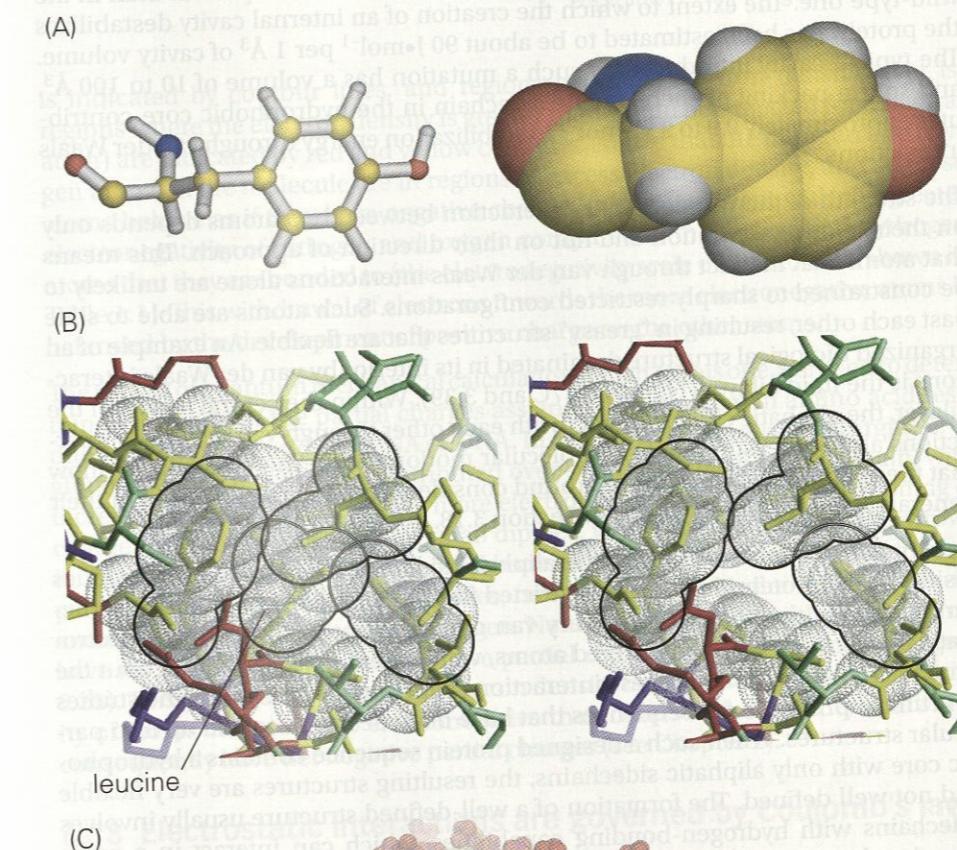


Figure 6.27 van der Waals interactions. (A) A tyrosine residue is shown, with spheres with radii corresponding to the van der Waals radii of the atoms. (B) A number of hydrophobic sidechains in the core of a protein are shown in this diagram. Dotted spheres are shown around each atom, and the radii of the spheres correspond to the van der Waals radii of the atoms. The effect of removing a leucine sidechain is shown in the panel on the right. The leucine sidechain and three of its neighbors are outlined in black. The atoms in these sidechains are in van der Waals contact. The projection view shown here makes it appear that the sidechains overlap, but in reality they merely touch each other. (C) The structure of a typical lipid bilayer. The phosphate and oxygen atoms of the head groups are shown in purple and red, respectively. Carbon and hydrogen atoms are shown in various colors, and these are the only atoms that are present in the interior of the bilayer.

two atoms together—thermal fluctuations can easily move them apart ($\epsilon_{ij} < k_B T$ at room temperature). Any individual van der Waals contact provides only a small degree of stabilization, but van der Waals interactions can add up to provide a significant net stabilization to the structures of proteins, because the interiors of folded proteins contains hundreds of atoms that are very tightly packed.

The importance of van der Waals interactions in stabilizing protein structures can be appreciated by analyzing the effects of replacing residues that have large sidechains, such as leucine, with alanine. If such a mutation is introduced within the hydrophobic core, the protein usually becomes less stable (see Figure 6.27). Most of the reduction in stability is due to a reduction in the hydrophobic effect (as you can see in Figure 4.74, alanine is less hydrophobic than leucine and, consequently, contributes less to the stability of the hydrophobic core). One component of the destabilization does arise, however, from the loss of van der Waals interactions due to the creation of packing defects or cavities around the smaller alanine sidechain in the mutant protein. Sidechains that are adjacent to the cavity participate in fewer van der Waals interactions in the mutant protein than in the wild-type one. The extent to which the creation of an internal cavity destabilizes the protein has been estimated to be about $90 \text{ J}\cdot\text{mol}^{-1}$ per 1 \AA^3 of cavity volume. The typical cavity introduced by such a mutation has a volume of 10 to 100 \AA^3 , and so the packing of each leucine sidechain in the hydrophobic core contributes approximately 0.9 to $9 \text{ kJ}\cdot\text{mol}^{-1}$ in stabilization energy through van der Waals interactions.

The strength of the van der Waals interaction between two atoms depends only on their relative separation and not on their direction of approach. This means that atoms that interact through van der Waals interactions alone are unlikely to be constrained to sharply restricted configurations. Such atoms are able to slide past each other, resulting in “greasy” structures that are flexible. An example of an organized biological structure dominated in its interior by van der Waals interactions is the lipid bilayer (Figures 6.27C and 3.39). Within the central region of the bilayer, the aliphatic chains interact with each other through van der Waals interactions alone. Measurements of molecular motion within the lipid bilayer show that these aliphatic chains move around considerably and are quite fluid in their general properties, as discussed in Section 3.18.

In contrast to membrane lipids, the aliphatic sidechains of amino acid residues inside folded proteins have more restricted flexibility. The key to imposing structural order on the otherwise slippery van der Waals interactions is the electrostatic interactions between charged atoms, which are distributed throughout the protein. The importance of these interactions has been made clear by the studies of artificial proteins with sequences that have been designed to fold up into particular structures. When such a designed protein sequence contains a hydrophobic core with only aliphatic sidechains, the resulting structures are very flexible and not well defined. The formation of a well-defined structure usually involves sidechains with hydrogen-bonding capabilities, which can interact in a more directional manner than is possible with the van der Waals interactions alone.

6.17 Atoms in proteins and nucleic acids are partially charged

Most of the atoms in a protein or nucleic acid molecule have an associated electric charge. In chemistry and biochemistry, we usually express charge in terms of the magnitude of the charge on the electron, e , which is known as the **elementary charge**. That is, when we say that an atom has a +1 charge, we mean that the positive charge is equivalent in magnitude to that of an electron—namely, 1.602×10^{-19} coulombs (C). An atom may be partially charged, which means that it bears a charge that is a fraction of that of an electron (for example, $+0.5 e$ or $-0.25 e$).

The charges on the atoms in a molecule are determined by quantum mechanical calculations of the electron density. The results of such a calculation are shown in Figure 6.28 for a small molecule that mimics an amino acid. The electron density

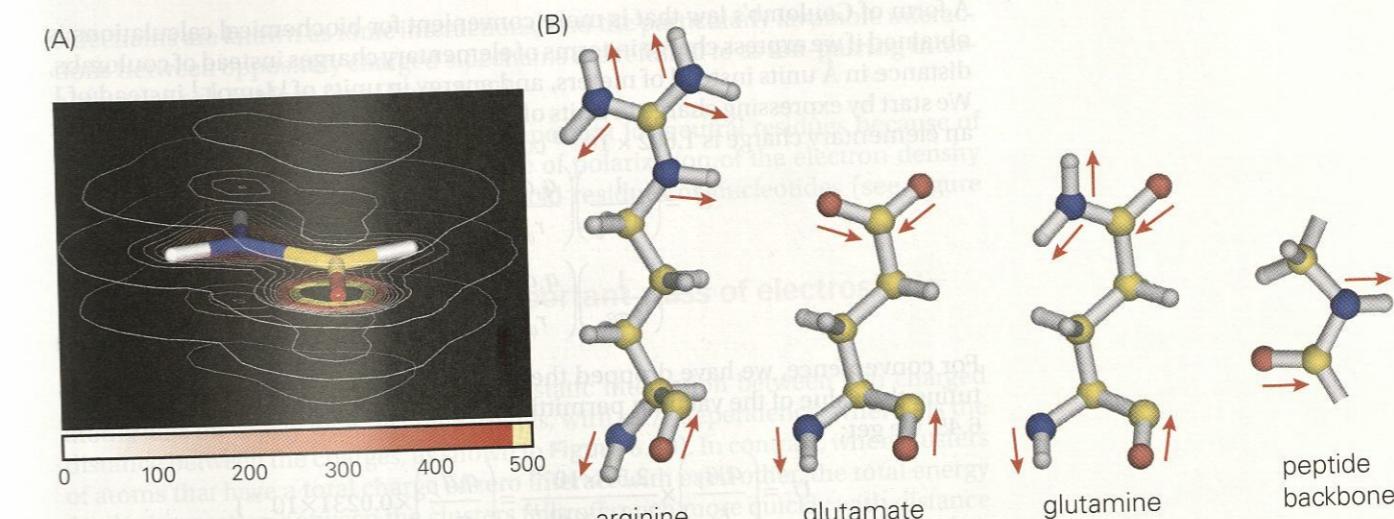


Figure 6.28 Partial charges on atoms. (A) The results of quantum mechanical calculations of electron density are shown for a small molecule that mimics an amino acid. The electron density is indicated by contour lines, with regions of excess electron density colored red and yellow. (B) Partial charges in amino acid residues. The partitioning of charge between adjacent atoms leads to the formation of electrostatic dipoles, indicated by arrows that point from the atom with partial negative charge to the one with partial positive charge. All of the atoms in the molecules have some degree of partial charge on them, but only the largest dipoles are shown. Neutral amino acids, such as glutamine, have no net charge, but they too have partial charges on their atoms. (A, adapted from A.R. Leach, Molecular Modelling: Principles and Applications. Upper Saddle River, NJ: Prentice Hall, 2001.)

The results of quantum mechanical calculations, such as this one, are used to determine the values of the partial charges assigned to each atom in amino acid residues and the nucleotides in RNA or DNA. In a charged residue, such as arginine or glutamate, the net charge is distributed over several atoms in the sidechain, with increased negative charge on the more electronegative atoms. This partitioning of charge leads to the formation of small dipoles, which are indicated by the arrows in Figure 6.28. The atoms in the backbone of the amino acid residue also have partial charges, leading to two dipoles, one associated with the carbonyl group, and one with the nitrogen and hydrogen. Neutral amino acids, such as glutamine, have no net charge, but they too have partial charges on their atoms (see Figure 6.28). This polarization results in electron density being concentrated in the vicinity of atoms that are more electronegative than others in the residue, and which consequently have net negative partial charges (see Chapter 2).

6.18 Electrostatic interactions are governed by Coulomb's law

The electrostatic energy of interaction, $U_{\text{electrostatic}}$, between two atoms, i and j , is described by Coulomb's law:

$$U_{\text{electrostatic}} = \left(\frac{1}{4\pi\epsilon_0} \right) \left(\frac{q_i q_j}{r_{ij}} \right) \quad (6.43)$$

where q_i and q_j are the charges on the two atoms, and r_{ij} is the distance separating them. The parameter ϵ_0 is known as the permittivity of vacuum and is given by:

$$\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2 \cdot \text{N}^{-1} \cdot \text{m}^{-2} \quad (6.44)$$

Comparing Equations 6.43 and 6.44 we can see that if the charges are expressed in coulombs (C) and the interatomic distance in meters (m), then Equation 6.43 yields the energy in joules (1 J = 1 N·m). The permittivity of water is ~80 times higher than that of vacuum, and the effects of this on the interactions between atoms will be discussed in Section 6.23.

A form of Coulomb's law that is more convenient for biochemical calculations is obtained if we express charge in terms of elementary charges instead of coulombs, distance in Å units instead of meters, and energy in units of $\text{kJ}\cdot\text{mol}^{-1}$ instead of J. We start by expressing charge in units of the elementary charge, using the fact that an elementary charge is 1.602×10^{-19} coulombs:

$$\begin{aligned} U &= \left(\frac{1}{4\pi\epsilon_0} \right) \left(\frac{q_i q_j}{r_{ij}} \right) (1.602 \times 10^{-19})^2 \text{ J} \\ &= \left(\frac{1}{4\pi\epsilon_0} \right) \left(\frac{q_i q_j}{r_{ij}} \right) \times 2.57 \times 10^{-38} \text{ J} \end{aligned} \quad (6.45)$$

For convenience, we have dropped the subscript for U in Equation 6.45. Substituting the value of the vacuum permittivity, ϵ_0 , from Equation 6.44 into Equation 6.45, we get:

$$U = \left(\frac{q_i q_j}{r_{ij}} \right) \times \frac{2.57 \times 10^{-38}}{111.3 \times 10^{-12}} = \left(\frac{q_i q_j}{r_{ij}} \right) \times 0.0231 \times 10^{-26} \text{ J} \quad (6.46)$$

Expressing distance in Å units rather than meters, and using the fact that $1 \text{ \AA} = 10^{-10}$ meters, we get:

$$U = \left(\frac{q_i q_j}{r_{ij}} \right) \times 0.0231 \times 10^{-26} \times 10^{10} = \left(\frac{q_i q_j}{r_{ij}} \right) \times 0.0231 \times 10^{-16} \text{ J} \quad (6.47)$$

The energy in Equation 6.47 is in units of joules (J). If we wish to express the energy in units of $\text{kJ}\cdot\text{mol}^{-1}$, instead, Equation 6.47 becomes:

$$\begin{aligned} U &= \left(\frac{q_i q_j}{r_{ij}} \right) \times 0.0231 \times 10^{-16} \times 6.022 \times 10^{23} \times 10^{-3} \\ &= \left(\frac{q_i q_j}{r_{ij}} \right) \times 1391 \text{ kJ}\cdot\text{mol}^{-1} \end{aligned} \quad (6.48)$$

In Equation 6.48, we have multiplied the right-hand side by Avogadro's number (6.022×10^{23}) in order to convert to units of $\text{J}\cdot\text{mol}^{-1}$, and then by 10^{-3} to convert $\text{J}\cdot\text{mol}^{-1}$ to $\text{kJ}\cdot\text{mol}^{-1}$. In Equation 6.48, the two charges are expressed in terms of elementary charges, and the interatomic distance is in Å units.

To see how to use this equation, consider an ion with a unit positive charge that is 4.0 \AA from an ion with a unit negative charge. We use Equation 6.48 to calculate the energy as follows:

$$U = \frac{-1 \times 1}{4} \times 1391 = -347.8 \text{ kJ}\cdot\text{mol}^{-1} \quad (6.49)$$

This is a very strong energy of interaction, much greater than the value of $k_B T$ at room temperature ($\sim 2.5 \text{ kJ}\cdot\text{mol}^{-1}$). This energy is calculated for the two charges interacting in a vacuum. As discussed in Section 6.23, water weakens the strength of electrostatic interactions by about 80-fold. Thus, the same two ions interacting in water will have an interaction energy of only $\sim -4.34 \text{ kJ}\cdot\text{mol}^{-1}$, which is less than $2k_B T$. Thus, interactions between charges are easily broken in water.

The strongest electrostatic interactions arise between atoms in residues that have a net charge, because atoms in these residues bear the largest individual charges (Figure 6.29). At neutral pH, lysine and arginine residues each have a net positive charge of +1, and aspartate and glutamate are negatively charged (-1). Histidine, which has a pK_a near 7, may be positively charged or neutral, depending on its environment and the pH of the solution (the concept of the pK_a of an amino acid sidechain is explained in Section 10.17). Interactions between oppositely charged

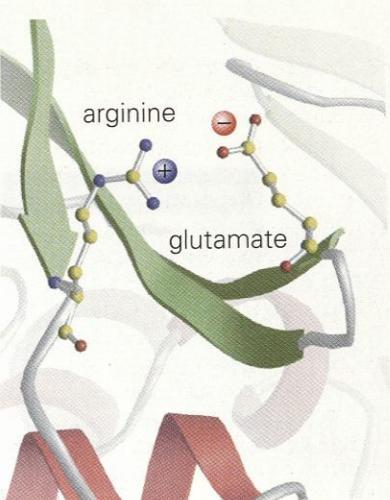


Figure 6.29 Ion-pairing interactions. When oppositely charged residues interact closely, as shown here for sidechains on the surface of a protein, they are said to form an ion pair, or a salt bridge.

sidechains are known as ionic interactions, and the particularly favorable interactions between oppositely charged sidechains are referred to as **ion-pairing interactions** or **salt bridges** (see Figure 6.29).

Electrostatic interactions may also be important for neutral residues because of partial charges on the atoms, or because of polarization of the electron density that can occur for any atom in amino acid residues or nucleotides (see Figure 6.28).

6.19 Hydrogen bonds are an important class of electrostatic interactions

According to Coulomb's law, the electrostatic interaction between two charged atoms falls off slowly with distance (that is, with a $1/r$ dependence, where r is the distance between the charges, as shown in Figure 6.30). In contrast, when clusters of atoms that have a total charge of zero interact with each other, the total energy for the interaction between the clusters falls off much more quickly with distance than $1/r$. Since the net charge on each cluster is zero, at large distances the interactions between individual charged atoms in the two clusters cancel out, leaving no net attraction. When the clusters are closer together, the distances between various pairs of atoms are not the same, and there is a net electrostatic interaction. For example, the energy of interaction between favorably aligned dipoles falls off as $-1/r^3$, where r is the distance between the interacting dipoles. This function decays to zero much more quickly than the $-1/r$ fall-off in the interaction energy between two opposite charges (see Figure 6.30).

The interaction energy for two dipoles depends strongly on their relative orientation. The most favorable interaction occurs when the dipoles approach each other co-linearly, with the positive pole of one dipole pointing towards the negative pole of the other one. If the orientation of one of the dipoles is flipped, this attractive interaction is converted to a repulsive one. The strongly directional nature of the short-range interactions between the dipoles or larger clusters of partially charged atoms provides the structural glue that holds specific protein conformations in place.

A particularly important electrostatic interaction in biology occurs between oxygen or nitrogen atoms bearing hydrogen (known as the donor atoms), and oxygen or nitrogen atoms bearing lone pairs of electrons (the acceptor atoms, Figure 6.31). The hydrogen atom "bridges" the donor and acceptor atoms, and such an interaction is known as a **hydrogen bond** (see Section 1.4). The stabilization that results from the formation of a hydrogen bond can be understood in terms of electrostatics. The hydrogen attached to each of the donor atoms has a small net positive charge and the more electronegative donor atom has a small net negative charge, forming a dipole (see Figure 6.31A). Likewise, the acceptor atom bearing

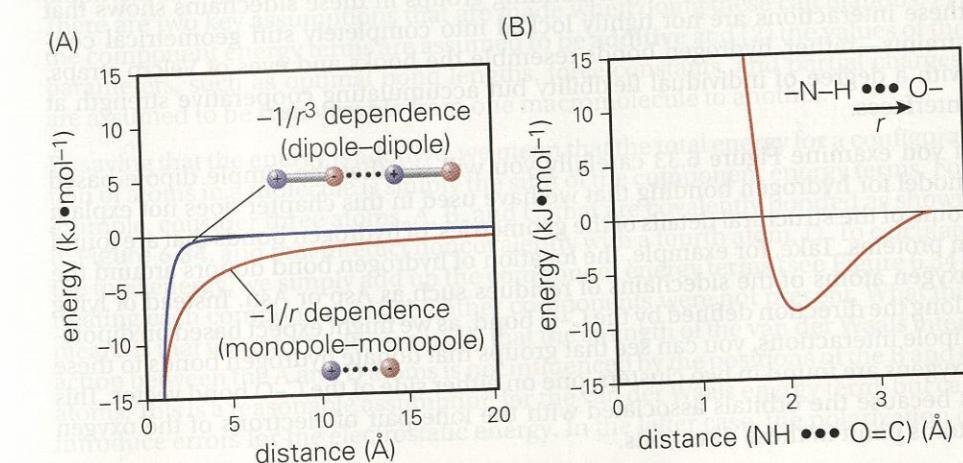


Figure 6.30 Electrostatic energy and hydrogen bonds. (A) The dependence of electrostatic energy on distance. The interactions between two opposite charges falls off with distance as $(-1/r)$, where r is the distance between the charges. The interaction between two pairs of charges (dipoles) falls off much more quickly with distance and has a $(-1/r^3)$ dependence. (B) The change in energy as the distance between the hydrogen and oxygen atoms in a $\text{NH} \cdots \text{O}=\text{C}$ hydrogen bond is increased. The energy shown here is the sum of the van der Waals and electrostatic energy terms.

Figure 6.32 α helices and β sheets allow the formation of repetitive and stable amide-carbonyl hydrogen bonds. This diagram shows the hydrogen-bonding networks that are present in an α helix (A), a parallel β sheet (B), and an antiparallel β sheet (C).

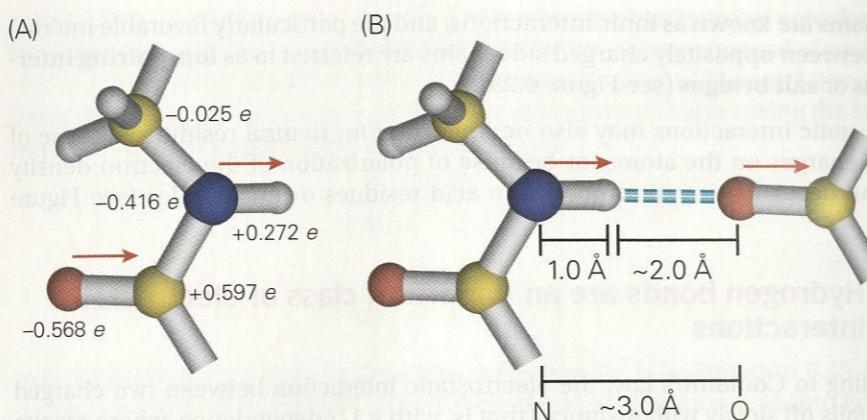


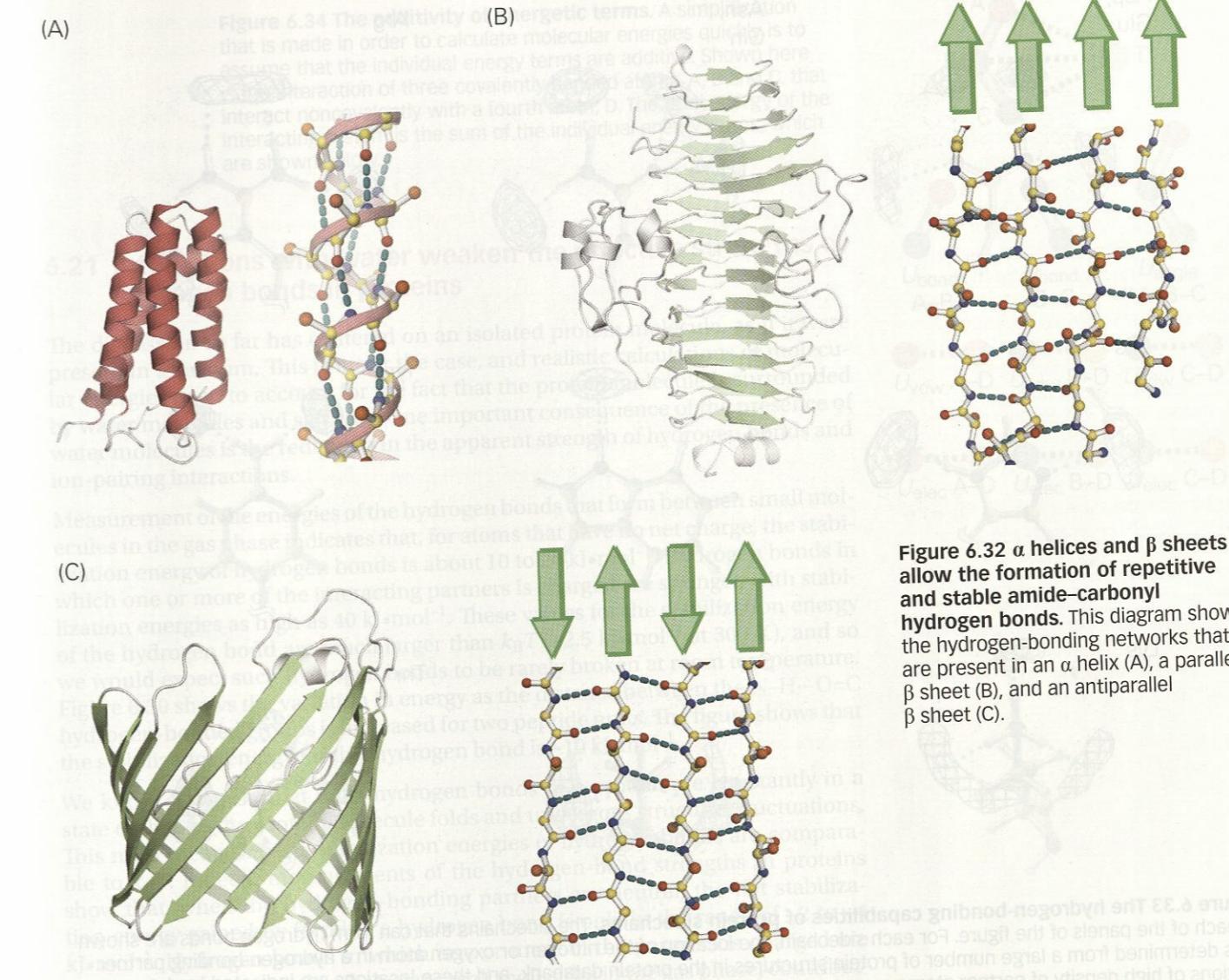
Figure 6.31 A hydrogen bond between two peptide groups. (A) Partial charges on the atoms in the backbone of a peptide, shown as multiples of an elementary charge, e. (B) The amide (N-H) and carbonyl (C=O) groups in the backbone interact by forming hydrogen bonds that align the dipoles (indicated by arrows that point towards the positively charged atom in the pair). The formation of this kind of backbone hydrogen bond underlies the stability of α helices and β sheets in proteins.

the lone pair of electrons has a small net negative charge and the atom that it is covalently bonded to has a small net positive charge, forming another dipole (see Figure 6.31B). In this simplest kind of hydrogen bond, the electrostatic interaction between two groups of atoms can be described as a dipole-dipole interaction, embedded within the van der Waals interactions that determine the distance of closest approach between the various atoms. Particularly tight hydrogen bonding interactions have the hydrogen and oxygen atoms separated by about 2 Å.

As we have seen already, hydrogen bonds are very important in determining the conformations of proteins and nucleic acids. In proteins, the amide NH and carbonyl C=O groups of the polypeptide backbone interact with each other to form arrays of hydrogen bonds in α helices and β sheets (Figure 6.32; see also Figures 1.36–1.38). The minimum energy configuration of interacting dipoles is one in which the dipoles are co-linear. α helices and β sheets are arrays of dipoles and, in particularly stable and well-formed elements of secondary structure, the N-H...O-C interaction tends to be linear. However, hydrogen-bonding geometry is not sharply restricted, as can be seen in the variation of hydrogen-bonding angles in the parallel β sheet shown in Figure 6.32C.

Figure 6.33 shows the results of a survey of high-resolution protein structures, indicating the directional tendencies of the hydrogen bonds formed by protein sidechains. As for the backbone hydrogen bonds, the spread of orientations in the alignments of the hydrogen-bonding groups in these sidechains shows that these interactions are not tightly locked into completely stiff geometrical constraints—rather, hydrogen bonds resemble the hooks and eyes of Velcro straps, with a degree of individual flexibility but accumulating cooperative strength at interfaces.

If you examine Figure 6.33 carefully you will see that the simple dipole-based model for hydrogen bonding that we have used in this chapter does not explain some of the structural details of the geometries of hydrogen bonds that are found in proteins. Take, for example, the location of hydrogen bond donors around the oxygen atoms of the sidechains of residues such as Asp or Asn. Instead of lying along the direction defined by the C=O bond, as we might expect based on dipole-dipole interactions, you can see that groups that donate hydrogen bonds to these oxygens are found in two clusters, one on either side of the C=O bond vector. This is because the orbitals associated with the lone pair of electrons of the oxygen atoms point in these directions.



6.20 Empirical energy functions are used in computer programs to calculate molecular energies

Given a three-dimensional structure for a protein or nucleic acid molecule, empirical energy functions can be used to calculate the energy of the molecule. There are two key assumptions that are made when doing these calculations: (1) the component energy terms are assumed to be **additive** and (2) the values of the parameters, such as optimal bond lengths, force constants, and partial charges, are assumed to be **transferable** from one macromolecule to another.

By saying that the energy is additive, we mean that the total energy for a configuration of atoms in a molecule is simply the sum of the component energy terms. For example, consider three atoms, A, B, and C, that are covalently bonded as shown in Figure 6.34, and that interact noncovalently with a fourth atom, D. To calculate the total energy, we simply add up the component energy terms (see Figure 6.34), treating each component as if the other components were not present. What this means is that we assume, for example, that the strength of the van der Waals interaction between the A and D atoms is not influenced by the positions of the B and C atoms. This is a reasonable assumption for the van der Waals energy term, but can introduce errors for the electrostatic energy. In the latter case, the polarization of

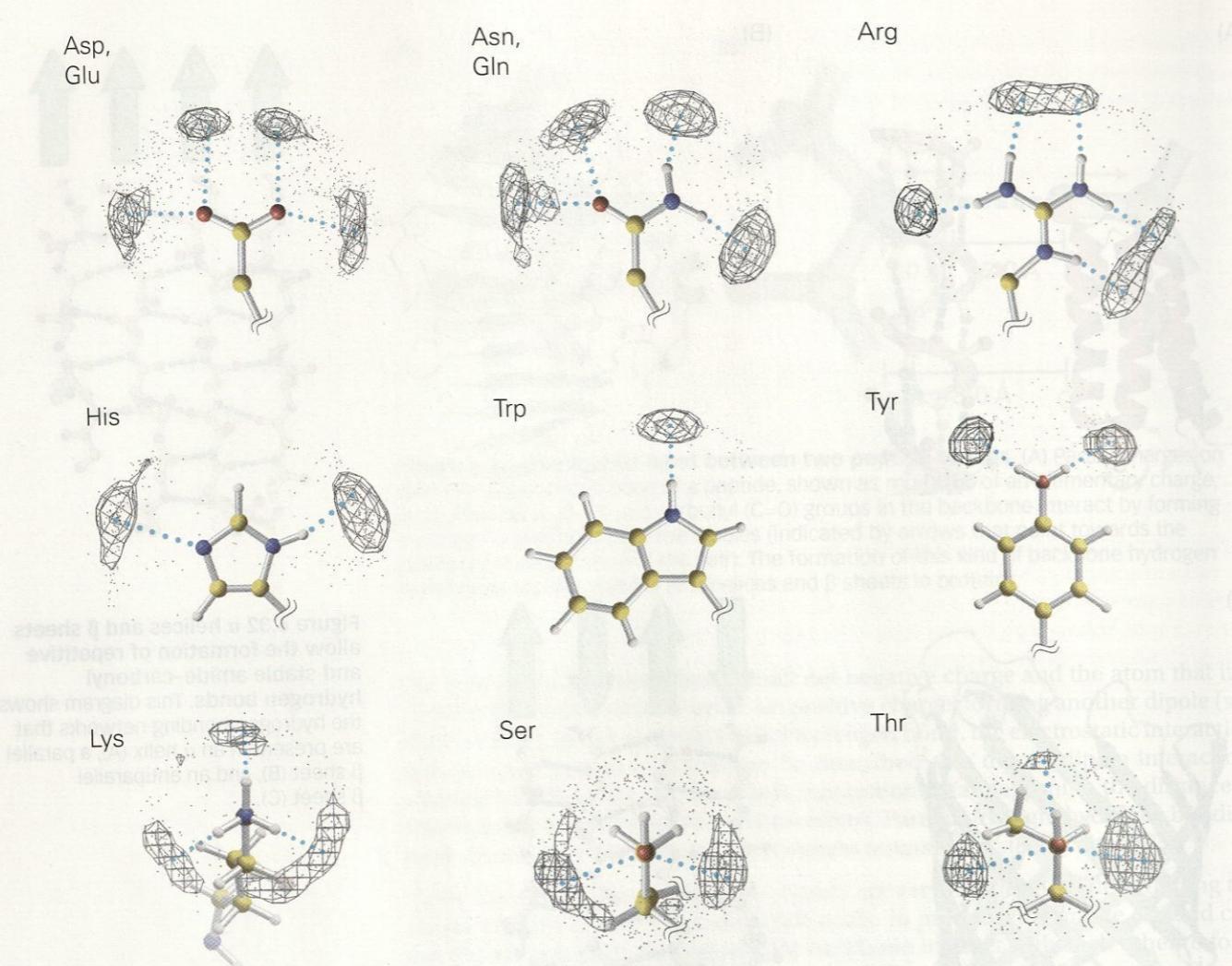


Figure 6.33 The hydrogen-bonding capabilities of protein sidechains. The sidechains that can form hydrogen bonds are shown in each of the panels of the figure. For each sidechain, the location of the nitrogen or oxygen atom in a hydrogen-bonding partner was determined from a large number of protein structures in the protein databank, and these locations are indicated by dots. Regions of high density of partner atoms are encompassed by meshes, and these represent regions of most favorable hydrogen-bond formation. For Tyr there are two equivalent positions for the hydroxyl proton, both of which are indicated. For Ser and Thr the C_β -O_y bond is pointed directly out of the plane of the page, and the hydroxyl proton is pointed up (the other protons are from C_β). (Adapted from J.A. Ippolito et al., and D. Christianson, *J. Mol. Biol.* 215(3): 457–471, 1990.)

atoms can be affected by the presence of other atoms, thereby changing the partial charges of the atoms and the electrostatic energy. Despite this complication, the assumption of additivity is still necessitated by limitations imposed by the speed of present-day computers. The calculation of energy would simply take too long if we had to recalculate the atomic polarization for each configuration of atoms.

The assumption of parameter transferability is another simplification that is introduced in order to make the energy calculation computationally tractable. We assume that parameters in the energy function can be defined once for all the component amino acids and nucleic acids, and that these parameters can be transferred without modification for use in energy calculations for any particular protein or oligonucleotide. For example, the force constant (K) and the optimal bond length (r_0) for the C_α - C_β bond in a leucine residue is assumed to be the same for all leucine residues in all proteins (see Equation 6.36). This is likely to be a good assumption for all energy terms, except for electrostatics once again, because polarization effects can make the transferability of electrostatic parameters problematic.

Figure 6.34 The additivity of energetic terms. A simplification that is made in order to calculate molecular energies quickly is to assume that the individual energy terms are additive. Shown here is the interaction of three covalently bonded atoms, A, B and C, that interact noncovalently with a fourth atom, D. The total energy of the interacting system is the sum of the individual energy terms, which are shown below.

6.21 Interactions with water weaken the effective strengths of hydrogen bonds in proteins

The discussion so far has centered on an isolated protein molecule, as if it were present in a vacuum. This is never the case, and realistic calculations of molecular energies need to account for the fact that the protein molecule is surrounded by water molecules and salt ions. One important consequence of the presence of water molecules is the reduction in the apparent strength of hydrogen bonds and ion-pairing interactions.

Measurement of the energies of the hydrogen bonds that form between small molecules in the gas phase indicates that, for atoms that have no net charge, the stabilization energy of hydrogen bonds is about 10 to 20 kJ•mol⁻¹. Hydrogen bonds in which one or more of the interacting partners is charged are stronger, with stabilization energies as high as 40 kJ•mol⁻¹. These values for the stabilization energy of the hydrogen bond are much larger than $k_B T$ (~2.5 kJ•mol⁻¹ at 300 K), and so we would expect such hydrogen bonds to be rarely broken at room temperature. Figure 6.30 shows the variation in energy as the distance between the N-H...O=C hydrogen-bonded groups is increased for two peptide units. The figure shows that the stabilization energy for this hydrogen bond is ~10 kJ•mol⁻¹.

We know, however, that most hydrogen bonds in a protein are constantly in a state of flux, as the protein molecule folds and undergoes structural fluctuations. This must mean that the stabilization energies of hydrogen bonds are comparable to $k_B T$. Indeed, measurements of the hydrogen-bond strengths in proteins show that, when the hydrogen-bonding partners are neutral, the net stabilization energy gained by forming a hydrogen bond is only in the range of ~2 to ~4 kJ•mol⁻¹. Hydrogen bonds in which one or more of the partners is charged are also weaker in proteins than between small molecules in the gas phase, contributing ~4 to ~8 kJ•mol⁻¹ of stabilization energy.

The crucial difference between the strengths of hydrogen bonds in proteins and those formed by small molecules in the gas phase is that the energy of hydrogen-bond formation in proteins is strongly influenced by water (Figure 6.35). Water is particularly good at forming hydrogen bonds with the backbone and sidechains of the amino acid residues in a protein. The change in energy when a hydrogen bond within a protein is broken is, consequently, a result of the difference in energy between the internal hydrogen bond and the hydrogen bonds that the protein residue can make with water. This results in the relatively small net energies associated with the formation of hydrogen bonds in proteins.

We can gain some appreciation for the effect of water on hydrogen-bond strengths by studying the behavior of a small organic compound, acetamide (see Figure 6.35C). Acetamide contains hydrogen-bonding groups that are similar to those found in proteins. When acetamide is dissolved in chloroform, a nonpolar solvent that cannot form hydrogen bonds, it can only form hydrogen bonds with itself. In this case, the net enthalpic stabilization of these hydrogen bonds has been measured experimentally to be -17 kJ•mol⁻¹. When acetamide is dissolved in water, with which it can form hydrogen bonds, the net enthalpic stabilization due to hydrogen bonds between molecules of acetamide becomes essentially zero. This is explained by the fact that dimer formation by acetamide now requires breaking hydrogen bonds to water (see Figure 6.35C).

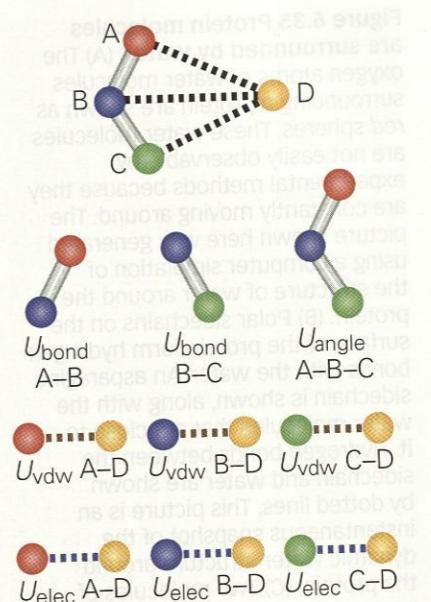
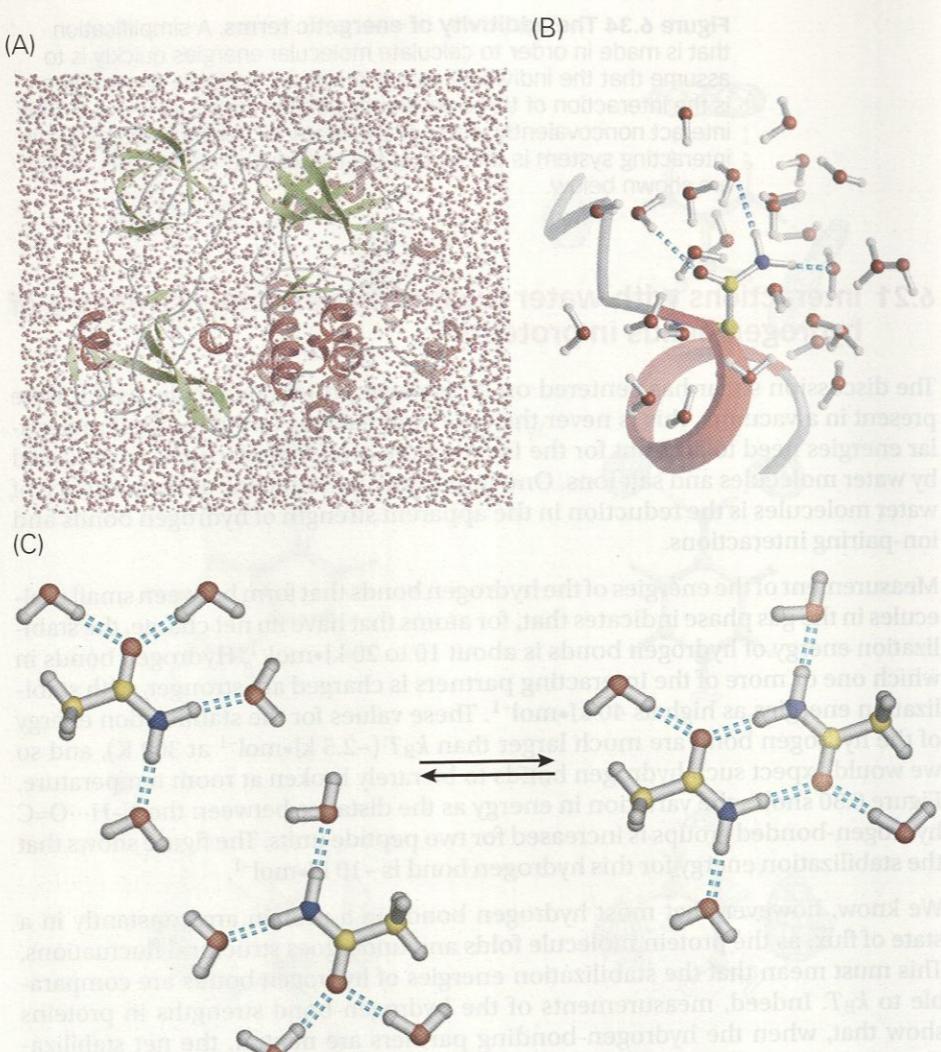


Figure 6.35 Protein molecules are surrounded by water. (A) The oxygen atoms of water molecules surrounding a protein are shown as red spheres. These water molecules are not easily observable by experimental methods because they are constantly moving around. The picture shown here was generated using a computer simulation of the structure of water around the protein. (B) Polar sidechains on the surface of the protein form hydrogen bonds with the water. An asparagine sidechain is shown, along with the water molecules that are close to it. Hydrogen bonds between the sidechain and water are shown by dotted lines. This picture is an instantaneous snapshot of the dynamic water structure around the protein. (C) Two molecules of acetamide are shown interacting with water, but not with each other. The picture on the right shows the two molecules forming hydrogen bonds with each other, which results in the loss of hydrogen bonds to water. This weakens the effective strength of the hydrogen bonds between molecules of acetamide.



6.22 The presence of hydrogen-bonding groups in a protein is important for solubility and specificity

If the competition with water makes hydrogen bonds only marginally stabilizing, why does the protein have so many hydrogen-bonding groups placed throughout its structure? One reason is that the polar groups of the protein provide for solubility and specificity. Were a protein to lack polar groups, it would be completely insoluble in water and would aggregate rather than fold. In addition, a polymer chain composed only of hydrophobic atoms would be unlikely to fold into a specific structure because the nondirectional van der Waals interactions would make it difficult to discriminate one condensed conformation of the chain from another. The constraints on conformation that are imposed by the need to maintain hydrogen-bonding complementarity is probably the single most important factor that determines the ability of protein chains to fold up into specific and well-defined three-dimensional structures.

Although the tendency of water to compete for hydrogen bonds with the protein ends up reducing the stability of the folded protein structure, this competition with water is crucial for protein function. As we shall see throughout this book, the ability of protein molecules to change their conformations in response to external cues is a necessary aspect of their function. Water molecules allow hydrogen bonds to be broken readily and remade without major expenditures of energy. Had the hydrogen bonds retained the full strength that is latent in them, then

protein molecules would be rigidly locked into particular conformations and would be functionally inert.

Perhaps the most important conclusion to be drawn from the analysis of hydrogen bonding in proteins is that it is not so much the energy associated with a particular hydrogen bond that is important, as much as the fact that the hydrogen bond is made. If a sidechain with hydrogen-bonding capability, such as a tyrosine sidechain, were to be folded into the protein structure in such a way that the sidechain could not form a hydrogen bond, then the folded protein would be destabilized by the withdrawal of such a sidechain away from its interactions with water. However, by providing compensating hydrogen bonds within the folded structure, the protein can maintain the balance of interactions. The secondary structural elements of the protein are particularly good at satisfying the hydrogen-bonding requirements of the backbone. Whenever polar sidechains are found inside the protein structure, they are almost invariably associated with one or more suitable hydrogen-bonding partners.

Amino acid sidechains that bear a full charge interact particularly strongly with water, even when they form an ion-pairing interaction with another sidechain (see Figure 6.29 for an illustration of an ion pair). Removal of a charged sidechain from water consequently has a large energetic penalty, known as the **desolvation energy**. The desolvation energy of charged sidechains is so high that they are rarely found in the interiors of proteins.

6.23 The water surrounding protein molecules strongly influences electrostatic interactions

Water molecules are polarizable, and they respond to the presence of charged atoms near them. The presence of adjacent charges increases the dipole moment of water so as to balance the charge. This tends to reduce the electrostatic interaction energy between two charges in water. The attenuation of the interaction energy between two charges in a polarizable medium, such as water, is accounted for by modifying Coulomb's law. If the polarizability of the medium is uniform throughout space, then the energy of interaction between two charges is attenuated by a factor, ϵ , known as the **dielectric constant** of the medium. This leads to the following modified form of Coulomb's law:

$$U = \frac{1}{\epsilon} \left(\frac{q_i q_j}{r_{ij} \text{\AA}} \right) \times 1390 \text{ kJ} \cdot \text{mol}^{-1} \quad (6.50)$$

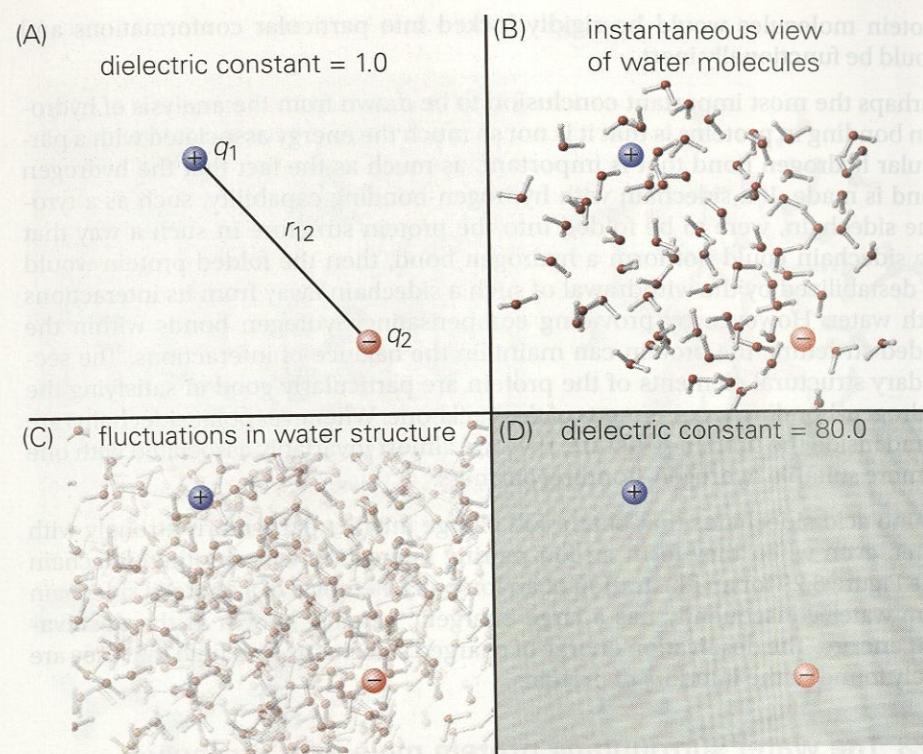
Equation 6.50 is derived from Equation 6.48 (Coulomb's law for charges interacting in a vacuum, with q_i and q_j being elementary charges, and the interatomic distance being expressed in Å units). This modified form of Coulomb's law can be used to calculate the interaction between two charges in water if we ignore the molecular details of the structure of water and simply assume that the strength of the interaction between the charges is reduced by an appropriate amount. The effective dielectric constant of water is 80, and this relatively large value reflects the ability of water to attenuate electrostatic interactions strongly. The effective interaction energy between the two charges in water is then given by Coulomb's law, with the dielectric constant set to 80 (Figure 6.36).

Two charged atoms in water do not actually see a medium of uniform dielectric around them, but are instead surrounded by discrete molecules of water (see Figure 6.36B). Each charged atom in water is surrounded and stabilized by the dipoles of water, and consequently interacts much less strongly with other charged atoms in the solution. The structure of the surrounding water fluctuates constantly due to thermal collisions (see Figure 6.36C) and, if we average over a long time and over many configurations of the water molecules, we find that the effective or averaged interaction energy between the two charges is reduced by a factor of 80 (the effective dielectric constant of water) relative to their interaction in a vacuum.

Dielectric constant

The dielectric constant, denoted ϵ , is a scale factor that reduces the magnitude of the electrostatic energy as calculated using Coulomb's law. The dielectric constant accounts for the effect of the environment in weakening the interaction between charges. The dielectric constant of bulk water is 80. The interior of a protein, which is slightly polar, has a much smaller dielectric constant (~2).

Figure 6.36 The electrostatic interaction between charged atoms that are surrounded by water molecules is attenuated. In (A), two ions, one positively charged (blue) and one negatively charged (red) are shown in a vacuum. The electrostatic energy is given by Coulomb's law. (B) The two ions are shown surrounded by water molecules. For this configuration of waters and ions, we can still calculate the energy using Coulomb's law, but we need to sum over all the atoms in the system, including the oxygen and hydrogen atoms of the water molecules. (C) All the atoms in the system are continually moving, and the net electrostatic energy of the system is the average over all instantaneous configurations. Over time, the effect of the water molecules on the ions is blurred out, and the detailed structure of water becomes less important. (D) We can approximate the effect of water molecules on the interactions between the ions by reducing the strength of the electrostatic energy by a factor known as the dielectric constant. This affords us a great simplification, because we are back to considering just the positions of the two ions.



When the molecular environment surrounding two interacting charges is less polarizable than water, the attenuation of the electrostatic interaction is correspondingly smaller. For example, a charge inside a protein molecule is surrounded by many chemical groups that are not very polarizable (aliphatic groups in sidechains, for example), and by some groups that are fairly polarizable (such as the amide and carbonyl groups of the backbone). In contrast to water, which has a very dynamic structure and can readily reorient to interact with charges, atoms in the interior of a protein are relatively rigid and are therefore limited in their ability to attenuate electrostatic interactions. The effective dielectric constant for the interior of a protein molecule turns out to be very low and is between 2 and 4, depending on the particular environment.

An electrostatic calculation that approximates the effect of polarizable groups by a choice of suitable dielectric constants for the medium is referred to as a **continuum dielectric model** for electrostatics. The application of continuum dielectric models to the consideration of electrostatic effects in proteins is a very powerful tool in understanding protein function, because it allows us to visualize effects that would be otherwise impractical to calculate if we had to account for each individual water molecule. Without the use of the continuum dielectric approximation, we would have to average the electrostatic energies over a large number of water configurations, which would make such calculations extremely time consuming.

In the continuum dielectric approach to electrostatics, individual solvent molecules are replaced by a region of uniform dielectric constant outside the protein, which represents the averaged effects of the solvent. The calculation of electrostatic effects in a protein involves the consideration of at least two regions with different dielectric constants (Figure 6.37). The region outside the protein, corresponding to the bulk solvent, has an effective dielectric constant of 80, that of water. Inside the protein, the structural environment is quite inhomogeneous, but in the simplest approximation this is treated as a region with a uniformly low dielectric constant (typically this is set to a value between 2 and 4). Because of the nonuniformity of the dielectric medium, Coulomb's law is no longer applicable.

Figure 6.37 A protein surrounded by water creates dielectric boundaries. The difficulty in calculating electrostatic energies for a protein molecule immersed in water is illustrated. The region within the protein (colored white) has a low dielectric constant ($\epsilon = 2$). The region outside the protein, filled with water molecules (not shown), has a high dielectric constant ($\epsilon = 80$). Consider two charged atoms, one inside the protein (blue) and one outside (red). The interaction energy between them cannot be calculated using Coulomb's formula, because the space between them has different dielectric constants in different regions. Instead, the Poisson equation has to be solved in order to calculate the spatial distribution of the electrostatic potential.

Instead, the fundamental equation describing the electrostatic potential generated by the set of charges in the protein is a differential equation known as the **Poisson equation**, which we shall not discuss in detail.

The solution to the Poisson equation relates the spatial distribution of charges within the protein to the electrostatic potential, $\phi(\vec{r})$, at any point in space, \vec{r} . The value of the electrostatic potential at any point is the energy required to move a unit positive charge from infinity to that point (this concept is discussed in more detail in Section 11.10). Solution of the Poisson equation yields a three-dimensional map of the electrostatic potential as a function of position in space (Figure 6.38). This map allows us to calculate the interaction energy between charged atoms and the protein.

Protein molecules normally function in environments with appreciable ionic strength, corresponding typically to a concentration of 100–150 mM NaCl. The effect of ionic strength on the electrostatic potential generated by the protein can be calculated by using a modified form of the Poisson equation, known as the **Poisson-Boltzmann equation**. The Poisson-Boltzmann equation is derived by assuming that the ions in the solution are distributed throughout the solvent region as given by a Boltzmann distribution based on the electrostatic potential. The Poisson-Boltzmann equation takes account of the fact that the ions will be distributed more towards regions of electrostatic potential that are energetically favorable, and this results in a screening of electrostatic effects by the ions.

6.24 The shapes of proteins change the electrostatic fields generated by charges within the protein

The results of solving the Poisson-Boltzmann equation depend on the shape of the protein under consideration. The shape of the protein determines the shape

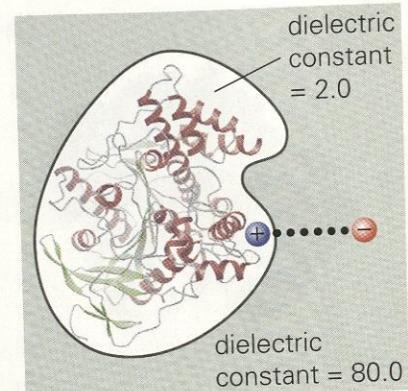


Figure 6.38 The electrostatic potential generated by a protein.

(A) The structure of trypsin, a protease enzyme, is shown in yellow. The electrostatic potential generated by the charges in trypsin is illustrated in color for a plane passing through the active site of trypsin (white circle). Blue and red indicate regions of positive and negative electrostatic potential, respectively. The dielectric constant is assumed to be 80 throughout space. An inhibitor protein that binds to trypsin is also shown, slightly separated from it. The inhibitor protein was not included in the electrostatic calculation. A lysine residue in the inhibitor binds to trypsin (white arrow). The positively charged lysine residue is in a region of positive electrostatic potential, which would not favor binding to trypsin. (B) The electrostatic calculation was repeated, but this time the dielectric constant was set to 2 inside trypsin and 80 outside. Even though the charges that generate the electrostatic potential are the same as in (A), the map of the potential looks very different. Note that the active site of trypsin is now in a region of negative electrostatic potential, which favors binding of the inhibitor. (From B. Honig and A. Nicholls, *Science* 268: 1144–1149, 1995. With permission from the AAAS.)

