

3. Solutions

Many biological processes occur between molecules in aqueous solution. In addition, many protein and nucleic acid molecules adopt three-dimensional structure (“fold”) in aqueous solution. It is therefore important to understand solutions and their thermodynamic properties.

Consider mixing of two pure solvents A and B with molecules of comparable size. There is an entropy increase associated with mixing because of the greater number of ways of arranging the molecules in the mixed solution.

Consider the separate A and B solvents. Solvent A contains N_A molecules on N_A “sites” in the liquid and solvent B contains N_B molecules on N_B sites in the liquid. Because the A molecules diffuse in the liquid, they are considered indistinguishable and $W = 1$ ($S_A = 0$) in the neat liquid. Similarly, $S_B = 0$.

After mixing A and B, consider the mixed solution as N_A A molecules on N_A sites and N_B B molecules on N_B sites with $N = N_A + N_B$ total sites. The total number of ways of arranging the molecules in the solution (see Eq. 1.25):

$$W = N!/(N_A!N_B!) \quad (3.1)$$

Using Stirling’s approximation $\ln(N!) = N\ln N - N$:

$$\Delta S_{\text{mix}} = k \times (N_A \ln N + N_B \ln N - N_A \ln N_A - N_B \ln N_B) \quad (3.2)$$

If we define the A mol fraction as $x = N_A/N$ and the B mol fraction as $1 - x$:

$$\Delta S_{\text{mix}} = Nk \times \{-x \ln x - [(1 - x) \times \ln(1 - x)]\} \quad (3.3)$$

ΔS_{mix} will have maximum value at $x = 0.5$, $0.693 \times Nk$ or 5.76 J/mol-K .

Free energy of mixing will also include an energetic term. Each molecule can be considered as having energetic contact with a number (z) of other molecules. The energies of the A-A, B-B, and A-B contacts, w_{AA} , w_{BB} , and w_{AB} , will likely be different from one another.

The total number of A-A, B-B, and A-B contacts in the solution are m_{AA} , m_{BB} , and m_{AB} , respectively. Each A-A contact is shared between two A molecules, each B-B contact is shared between two B molecules, and each A-B contact is shared between one A and one B molecule. The equations follow:

$$z \times N_A = 2m_{AA} + m_{AB}$$

$$z \times N_B = 2m_{BB} + m_{AB}$$

$$z \times N = (m_{AA} + m_{BB} + m_{AB}) \times 2 \quad (3.4)$$

The upper two lines of Eq. 3.4 can be rewritten:

$$\begin{aligned} m_{AA} &= [(z \times N_A) - m_{AB}]/2 \\ m_{BB} &= [(z \times N_B) - m_{AB}]/2 \end{aligned} \quad (3.5)$$

Eqs. 3.5 can be incorporated into the total intermolecular energy of the solution:

$$\begin{aligned} U &= (m_{AA} \times w_{AA}) + (m_{BB} \times w_{BB}) + (m_{AB} \times w_{AB}) \\ &= [(z \times w_{AA}) \times N_A/2] + [(z \times w_{BB}) \times N_B/2] + \{[w_{AB} - (w_{AA} + w_{BB})/2] \times m_{AB}\} \end{aligned} \quad (3.6)$$

The point of this derivation is that the energy depends on m_{AB} but not on m_{AA} and m_{BB} . The quantity m_{AB} can be estimated by considering that a fraction N_B/N of the contacts of A molecules are A-B contacts. This is a “mean-field” approximation:

$$m_{AB} \approx [(z \times N_A) \times N_B/N] = zN_A x (1 - x) \quad (3.7)$$

The left equality of Eq. 3.7 can be substituted into Eq. 3.6:

$$U = [(z \times w_{AA}) \times N_A/2] + [(z \times w_{BB}) \times N_B/2] + (kT \times \chi_{AB} \times N_A N_B/N) \quad (3.8)$$

with the dimensionless “exchange parameter”:

$$\chi_{AB} = \{z \times [w_{AB} - (w_{AA} + w_{BB})/2]\}/kT \quad (3.9)$$

This parameter represents the difference between the energy of A-B interactions and the average energy of A-A and B-B interactions (relative to kT).

The first two terms in brackets in Eq. 3.8 correspond to U for the neat A and B solvents.

For the free energy of mixing, DB, Chapter 15 uses the Helmholtz free energy ($F = U - TS$) rather than the Gibbs free energy ($G = U - TS + pV$). Because pressure is typically constant in biochemical processes, and volume changes little, Helmholtz and Gibbs free energy changes are very close to one another. Using Eqs. 3.3, 3.7, and 3.8:

$$\begin{aligned} F &= NkT \times \{(x \times \ln x) + [(1 - x) \times \ln(1 - x)] + [zw_{AA}x/2kT] + [zw_{BB}(1 - x)/2kT] \\ &\quad + [\chi_{AB} \times x \times (1 - x)]\} \end{aligned} \quad (3.10)$$

$$\Delta F_{\text{mix}} = NkT \times \{x \ln x + [(1 - x) \times \ln(1 - x)] + [\chi_{AB} \times x \times (1 - x)]\} \quad (3.11)$$

In many cases, $\chi_{AB} > 0$ and mixing is entropy-driven. For example, for oil and water, $\chi_{AB} \sim 5$, and at 300 K, $\Delta F_{\text{mix}} \approx 0$ for $x = 10^{-2}$. In this model, oil and water would only mix to this mol fraction.

The chemical potentials with $x_A = x$ and $x_B = 1 - x$:

$$\begin{aligned}\mu_A/kT &= (\partial F/\partial N_A)_{NB,T} = \ln(x_A) + (zw_{AA}/2kT) + [\chi_{AB} \times (1 - x_A)^2] \\ \mu_B/kT &= (\partial F/\partial N_B)_{NA,T} = \ln(x_B) + (zw_{BB}/2kT) + [\chi_{AB} \times (1 - x_B)^2]\end{aligned}\quad (3.12)$$

See DB, page 272 for further details.

For proteins in aqueous solution, formation of 3D structure likely reflects lower energy of the 3D structure relative the unfolded ensemble of coil structures. This energy difference must more than compensate for the higher entropy of the coil structures (high multiplicity) relative to the folded 3D structure (low multiplicity). Any real calculation/simulation should also consider the energy and entropy of the surrounding water for the two different structures. The energies of membrane proteins are likely more complex because of the mixture of the apolar interior of the membrane and polar headgroup and aqueous environments.

For proteins in both aqueous and membrane environments, it is believed that “hydrophobicity” makes a significant contribution to the stabilization of 3D structure. Hydrophobicity will be discussed in more detail later in the course. At this time, I want to discuss some of the experiments which underlie our understanding of hydrophobicity or the “hydrophobic effect”.

These experiments measure the partitioning of an organic solute between an organic solvent and aqueous solution; i.e. the concentrations in the immiscible organic and aqueous layers at equilibrium.

Denote “s” as the organic solute, “A” as aqueous solvent, “B” as organic solvent, μ_{sA} as the chemical potential of s in A, μ_{sB} as the chemical potential of s in B, x_{sA} as the mol fraction s in A, and x_{sB} as the mol fraction s in B.

Using Eq. 3.12:

$$\begin{aligned}\mu_{sA}/kT &= \ln(x_{sA}) + (zw_{ss}/2kT) + [\chi_{sA} \times (1 - x_{sA})^2] \\ \mu_{sB}/kT &= \ln(x_{sB}) + (zw_{ss}/2kT) + [\chi_{sB} \times (1 - x_{sB})^2]\end{aligned}\quad (3.13)$$

At equilibrium, $\mu_{sA} = \mu_{sB}$, analogous to thermal or pressure equilibrium:

$$\begin{aligned}\ln(x_{sA}) + [\chi_{sA} \times (1 - x_{sA})^2] &= \ln(x_{sB}) + [\chi_{sB} \times (1 - x_{sB})^2] \\ \ln(K_{eq}) = \ln(x_{sA}/x_{sB}) &= [\chi_{sB} \times (1 - x_{sB})^2] - [\chi_{sA} \times (1 - x_{sA})^2]\end{aligned}\quad (3.14)$$

A classic experiment for understanding hydrophobicity is organic solute (e.g. benzene) partitioning into aqueous solvent from pure solvent (e.g. benzene). In this case, $x_{sB} = 1$ and $x_{sA} \ll 1$.

$$\ln(K_{eq}) = \ln(x_{sA}) \approx -\chi_{sA}$$

$$\Delta G_{\text{transfer}} = RT \times \chi_{\text{sA}} \quad (3.15)$$

For example, benzene partitioning into aqueous solution from pure benzene at $T = 310 \text{ K}$ yields $x_{\text{sA}} \approx 4 \times 10^{-4}$, $\chi_{\text{sA}} \approx 8$, and $\Delta G_{\text{transfer}} \approx 20 \text{ kJ/mol}$.

4. Intermolecular Interactions

There are several interactions which underlie macromolecular structure. The first interaction is chemical bonds. These are well-understood from small molecules.

In proteins, the CO–N bond is of special importance because it has partial double bond character which leads to formation of peptide planes.

In addition, there are different dihedral angles associated with the planes. There are cis- and trans- isomers of the C=O and N–H bonds in the peptide plane. In most cases, the trans-isomer has energy ~8 kJ/mol lower than the cis-isomer because of steric effects. Only the trans-isomer is observed. The exception is the residue preceding a proline and ~10% of these residues are cis-isomers.

The other important set of dihedral angles are the ϕ and ψ dihedral angles associated with N–C α and C α –CO bonds, respectively. These angles determine the relative orientation of adjacent peptide planes. For residues other than Gly or Pro, steric constraints generally limit the dihedral angles to two regions: strand ($-180^\circ < \phi < -60^\circ$, $90^\circ < \psi < 180^\circ$) and right-handed helix ($-75^\circ < \phi < -45^\circ$, $-70^\circ < \psi < -30^\circ$). Residues have a smaller probability of dihedral angles outside these regions. A small fraction of residues are observed in the left-handed helix region around $\phi = 50^\circ$, $\psi = 40^\circ$.

For Pro, the ring structure restricts ϕ to -60° . Because the “R” group of Gly is the proton, there is much less steric restriction. Most ϕ , ψ values are observed for Gly with less probability in $-40^\circ < \phi < 40^\circ$, $-15^\circ < \psi < 15^\circ$.

There are of course sidechain chemical bonds and torsion angles which are sterically constrained (similar to organic molecules or synthetic polymers).

Electrical interactions also play a role in protein and nucleic acid structure. The strongest interactions are between charges. The charge-charge energy:

$$u_{c-c} = C \times (q_1 q_2 / Dr) = (Ce^2) \times (z_1 z_2 / Dr) \quad (4.1)$$

where q_1 and q_2 are the electrical charges, e is the electron charge with value 1.602×10^{-19} Coulomb, z_1 and z_2 are the numerical charges on the two particles, D is the dielectric constant, r is the inter-charge distance, and C is 8.968×10^9 J-m/Coulomb².

The dielectric constant represents shielding of the charges by the intervening medium and plays an important role in biomolecular systems. The values of D in vacuum, water, and heptane are 1, ~78, and ~2, respectively. In vacuum, there is no shielding; i.e. the charges interact to their maximum extent. In water, the polar ends of the water molecules will orient around the charges and will reduce the interaction. In heptane, there will be polarization of the heptane molecules close to the charges which also qualitatively reduces their values.

The value of D in the membrane interior will be close to that of heptane. The value of D in the protein interior is ~ 4 , greater than that of heptane because there are polar backbone as well as apolar sidechain parts of protein residues (see VJH, pp. 116-117).

In biomolecular systems, it is often useful to describe electrical energy in terms of a single charge q and a surrounding “electrical potential” or “voltage” Ψ :

$$u_{e\Psi} = q\Psi = ze\Psi \quad (4.2)$$

The potential is based on the other nearby charges and could be written:

$$\Psi = C \times \sum_j q_j / (Dr_j) \quad (4.3)$$

where the sum is over these other charges and r_j is the distance between q and q_j . The typical units for Ψ are Volts (V) $\equiv$ J/Coulomb.

In molecules, charges are usually delocalized over a volume which is approximated as charge delocalized over a sphere with radius a . Consider the free energy (or electrical work) which must be done to take the sphere from no charge to charge q . Energy and work are required because at charges intermediate between 0 and q , there is electrical repulsion between each aliquot of added charge and the existing charge on the sphere. A straightforward derivation leads to the Born “self-energy”:

$$G_{\text{Born}} = (Cq^2)/(2Da) \quad (4.4)$$

See DB, pp. 424-425 for the derivation. This is an important equation in biochemistry because it tells us that there is positive energy associated with individual charges and this energy is proportional to $1/D$. For a charge of magnitude e and $a = 3 \text{ \AA}$, $G_{\text{Born}, D=3.5} - G_{\text{Born}, D=78} = 63 \text{ kJ/mol}$; i.e. there is a very large free energy difference between having a charge in water and a charge in the interior of a protein. The difference is even larger for the charge in the membrane interior with $D = 2$. It is therefore likely that charges in protein interior or in membranes will be close to opposite charges so that the attractive electrical energy compensates for the large Born self-energies. These opposite charges are called salt bridges.

A dipole refers to a separation of equal and opposite charges and its magnitude can be written:

$$\mu = qa = zea \quad (4.5)$$

where $q = ze$ is the charge magnitude and a is the separation distance of the charges. You will sometimes see μ in units of Debye. If $q = e$ and $a = 1 \text{ \AA}$, $\mu = 4.8 \text{ Debye}$.

For proteins, the C=O and N-H bonds in the backbone are dipoles and there are also dipoles associated with these groups and with O-H groups in sidechains. In addition, water molecules have large dipoles.

There is preference for aligning opposite charges on two dipoles. The approximate magnitude of the dipolar energy:

$$u_{dd} \sim C \times (\mu_1 \mu_2 / Dr^3) = (Ce^2) \times (a_1 a_2 / r^2) \times (z_1 z_2 / Dr) \quad (4.6)$$

As one example, consider a C=O group and N-H group in a hydrogen bond in the protein interior. The values of D , z_1 , z_2 , a_1 , a_2 , r are approximately 4, 0.4, 0.3, 1.2 Å, 1.0 Å, and 3.0 Å, respectively, so that $u_{dd} \sim 2$ kJ/mol. There is clearly some energetic preference for formation of backbone hydrogen bonds in the protein interior or in membranes.

Van der Waals interaction is an attractive electrical interaction between two chemical groups in which at least one of the groups does not have a permanent dipole, i.e. a dipole associated with a chemical bond. In this group (or both groups), there is an induced dipole; i.e. there is charge separation induced by charges in the other group. The magnitude of the induced dipole is proportional to the group's polarizability (α). The charges in the other group could be due to a permanent charge and/or a dipole or could be a transient dipole.

Van der Waals interactions are significant between apolar amino acid sidechains in the protein interior. For these apolar sidechains, $u_{VdW} \propto r^{-6}$.

In addition, at close interatomic distance for non-bonded atoms, there will be repulsion between the electrons and nuclei. The energy increases quickly in the close-approach domain and leads to "Van der Waals" radii for individual atoms. To a good approximation, the distance of closest approach of two atoms is the sum of their Van der Waals radii.

A common semi-empirical potential for two atoms is the Lennard-Jones potential:

$$u_{LJ} = A/r^{12} - B/r^6$$

where A and B are constants specific for the two atoms and represent repulsion and attraction, respectively. VJH Table 3.4 has Lennard-Jones parameters for nuclear pairs common in biomolecules.

Hydrogen bonds play an important role in protein and nucleic acid structure. They are generally between Hs in O-H or N-H groups and Os or Ns. Hydrogen bonds between backbone protein nuclei stabilize secondary structure (e.g. helix or strand). In addition, there are hydrogen bonds between water molecules and nuclei on the surface of the protein.

In one sense, the hydrogen bond is a simple favorable dipole-dipole interaction (e.g. $\delta^+C=O^{\delta-} \cdots \delta^+H-N^{\delta-}$). However, the $O \cdots H$ distance is ~ 1.9 Å and is less than the 2.7 Å sum of Van der Waals radii. In addition, hydrogen bond energies (4 – 50 kJ/mol) are somewhat greater than the energy expected from dipole-dipole interaction (see above). There have also been measurements of NMR "J-couplings" through hydrogen bonds. J-coupling is only observed with a "chemical" bond (i.e. shared electrons).

Full simulations of a protein in aqueous or membrane environment are in principle possible if one has a realistic model for the sum of the bonding and nonbonding energies among all of the atoms. A “force field” can then be created as the derivative of the potential with respect to displacement of the atoms. If each atom j moves for a short time τ under its force $\mathbf{F}_j$, the incremental displacement of the atom will be $\Delta \mathbf{r}_j = (\mathbf{F}_j/2m_j)(\tau^2)$. The potential and forces are then calculated at the new atomic coordinates and new displacements done for an additional time τ . The process is repeated until the coordinates corresponding to an energy minimum are found. At a minimum, the forces will be close to zero.

Simulations often add some random motion in addition to the force field-applied motion. This motion allows the atoms to move over potential barriers and reach the global energy minimum structure. The average magnitude of the random motion is set by a “temperature” parameter. From Eq. 1.106, the thermal $\langle v_{xj} \rangle = \langle p_{xj}/m_j \rangle = (kT/m_j)^{1/2}$. So, the average magnitude of thermal displacement during time τ will be $\langle \Delta r_{xj} \rangle = (kT\tau^2/m_j)^{1/2}$. In a real simulation, there will be a distribution of thermal displacements of the atoms with average magnitude $\langle \Delta r_{xj} \rangle$ and with no net average direction. During a single step, the magnitude and direction of the thermal displacement for a single atom is randomly generated.

In calculating aspects of the potential relating to permanent dipoles, the dipoles may not be considered explicitly and instead partial charges are assigned to each atom (see VJH Table 3.2). This allows all of the non-Van der Waals electrical interactions to be calculated using Eq. 4.1.

An accurate description of the non-Van der Waals electrical interactions requires knowledge of the dielectric constant which will be different in the protein interior (primarily apolar – low value of D) and the protein exterior (primarily polar – high value of D). For proteins in membranes, the situation is even more complex because there is the low dielectric membrane interior and higher dielectric headgroup and aqueous environments.

5. Water and the Hydrophobic Effect

Biomolecules are typically dissolved in aqueous solution or embedded in membranes. Water is a hydrogen-bonding solvent and there are often hydrogen bonds between water and biomolecules.

For proteins in aqueous solution, it is generally believed that the “hydrophobic effect” is the most important interaction for creation of 3D structure.

On the most basic level, the hydrophobic effect refers to the non-mixing of oil and water. Because of this effect, the hydrophobic sidechains of proteins “phase-separate” to form the protein interior and the hydrophilic sidechains form the protein exterior. This effect would generally explain protein folding in aqueous solution although it would not explain why specific 3D structure is attained.

The hydrophobic effect also explains the formation of membranes from lipid molecules. The lipid bilayer is one way in which the hydrophobic acyl chains of lipids can phase-separate from the aqueous phase.

Much of the fundamental understanding of the hydrophobic effect is based on measurements of the partitioning of organic solutes from their pure liquids to water (see Eq. 3.15). DB Chapter 30 uses $\Delta\mu$, Δh , Δs , and Δc_p to refer to the changes in free energy, enthalpy, entropy, and heat capacity with this process and this convention will be followed.

The partition coefficient will yield $\Delta\mu$. In addition, calorimetric measurements have been made to determine Δc_p for partitioning.

There were several results from these small organic molecule experiments. First, there is temperature dependence to solubility with a solubility minimum near room temperature. In addition, Δc_p is large and positive with value ~ 350 J/mol-K; i.e. the heat capacity of the organic molecule in water is much higher than in organic liquid.

The equations for Δh , Δs , and $\Delta\mu$ are adapted from Eqs. 1.93-1.94. In the adaptation, different reference temperatures T_1 and T_2 are used for Δh and Δs , respectively, instead of a single temperature T_m . The temperatures in parentheses refer to the temperature of the state function:

$$\Delta h(T) = \Delta h(T_1) + \{\Delta c_p \times [T - T_1]\}$$

$$\Delta s(T) = \Delta s(T_2) + \{\Delta c_p \times \ln[T/T_2]\}$$

$$\Delta\mu(T) = \Delta h(T_1) - [\Delta s(T_2) \times T] + \{\Delta c_p \times [T - T_1]\} - \{\Delta c_p T \times \ln[T/T_2]\} \quad (5.1)$$

As described in DB page 579, T_1 and T_2 are chosen to correspond to $\Delta h(T_1) = 0$ and $\Delta s(T_2) = 0$. The typical T_1 and T_2 values are ~ 295 K and 390 K, respectively. The Δh values have an approximate linear dependence on T and are negative below T_1 and positive above T_1 . The Δs values also have an approximate linear dependence on T and are negative below T_2 and positive above T_2 .

The values of Δh and Δs are somewhat unusual for partitioning of organic molecules into polar solvents. In particular, $\Delta h \approx 0$ near room temperature for water while partitioning into other polar solvents often has positive Δh . This latter more typical “regular solution” behavior is understood as preferential solvation of an organic molecule by other organic molecules and preferential solvation of polar molecules by polar molecules.

At lower temperature, organic solutes may order the neighboring water molecules and lead to better hydrogen bonding than exists for water molecules neighboring only water molecules. This improved hydrogen bonding near an organic solute may mitigate the loss in number of hydrogen bonds for these water molecules. Evidence which supports this water ordering is molecular “clathrate” solids which contain an organic molecule surrounded by a finite number of ordered water molecules.

At higher temperature, it is hypothesized that the more positive Δh is due to a decrease in the ordering of the neighboring water molecules.

The ordering of the water molecules can also explain the temperature dependence of Δs . At lower temperature, the ordering of the water molecules would be consistent with more negative Δs and at higher temperature, the loss in ordering of the water molecules would be consistent with more positive Δs .

The water molecule ordering contribution to Δs can be estimated by considering that an ordered water molecule neighboring an organic solute has about half of the orientational configurations relative to a water molecule surrounded by water molecules. If there are 15 water molecules neighboring an organic solute, then $\Delta s = R \times \ln(W_{\text{solute}}/W_{\text{no solute}}) = R \times \ln(0.5)^{15} \approx -85 \text{ J/mol-K}$. For benzene transfer to aqueous solution at 280 K, the experimental $\Delta s \approx -85 \text{ J/mol-K}$. There is semi-quantitative agreement between the model and experiment.

It is believed that the positive Δc_p is also related to the water molecules neighboring the organic solute. Comparison of Δc_p for different solutes shows a linear relationship between Δc_p and the expected number of water molecules neighboring the solute.

In aqueous solution, the interior of a folded protein contains many apolar sidechains typically in Van der Waals contact with exclusion of water. This type of close-packing yields minimal protein volume and surface area and therefore the minimum number of water molecules associated with the protein surface. There would therefore be the minimum Δc_p and minimal $\Delta \mu$ from Eqs. 5.1.