21. Binding and Association

Molecular associations are at the heart of biological processes. Specific functional interactions are present at every level of cellular activity. Some of the most important:

1) Proteins Interacting with Small Molecules and Ions
    - Enzyme/substrate interactions and catalysis
    - Ligand/receptor binding
    - Chemical energy transduction (for instance ATP)
    - Signaling (for instance neurotransmitters, cAMP)
    - Drug or inhibitor binding
    - Antibody binding antigen
    - Small molecule and ion transport
        - $Mb + O_2 \rightarrow MbO_2$
        - Ion channels and transporters
2) Protein–Protein Interactions
    - Signaling and regulation networks
    - Receptors binding to ligands activate receptors
        - GPCRs bind agonist/hormone for transmembrane signal transduction
    - Assembly and function of multi-protein complexes
        - Replication machinery in replisome consists of multiple proteins including DNA polymerase, DNA ligase, topoisomerase, helicase
        - Kinetochore orchestrate interactions of chromatin and the motor proteins that separate sister chromatids during cell division
3) Protein–Nucleic Acid Interactions
    - All steps in the central dogma
    - Transcription factor binding
    - DNA repair machinery
    - Ribozymes

In all of these examples, the common thread is a macromolecule, which typically executes a conformational change during the interaction process. Conformational flexibility and entropy changes during binding play an important role in describing these processes.

Thermodynamics of Bimolecular Reactions

To begin, we recognize that binding and association processes are bimolecular reactions. Let's describe the basics of this process. The simplest kinetic scheme for bimolecular association is

$$A + B \rightleftarrows C \tag{1}$$

A and *B* could be any two molecules that interact chemically or physically to result in a final bound state; for instance, an enzyme and its substrate, a ligand and receptor, or two specifically interacting proteins. From a mechanistic point of view, it is helpful to add an intermediate step:

$$A + B \rightleftarrows AB \rightleftarrows C$$

Here *AB* refers to transient encounter complex, which may be a metastable kinetic intermediate or a transition state. Then the initial step in this scheme reflects the rates of two molecules diffusing into proximity of their mutual target sites (including proper alignments). The second step is recognition and binding. It reflects the detailed chemical process needed to form specific contacts, execute conformational rearrangements, or perform activated chemical reactions. We separate these steps here to build a conceptual perspective, but in practice these processes may be intimately intertwined.

Equilibrium Constant

Let's start by reviewing the basic thermodynamics of bimolecular reactions, such as reaction scheme (1). The thermodynamics is described in terms of the chemical potential for the molecular species in the system ($i = A,B,C$)

$$\mu_i = \left(\frac{\partial G}{\partial N_i} \right)_{p,T,\{N_j, j \neq i\}}$$

where N_i is the number of molecules of species *i*. The dependence of the chemical potential on the concentration can be expressed as

$$\mu_i = \mu_i^0 + RT \ln \frac{c_i}{c^0} \tag{2}$$

c_i is the concentration of reactant *i* in mol L^{-1}, and the standard state concentration is $c^0 = 1$ mol L^{-1}. So the molar reaction free energy for scheme (1) is

$$\begin{aligned} \Delta \bar{G} &= \sum\nolimits_i \nu_i \mu_i \\ &= \mu_C - \mu_A - \mu_B, \\ &= \Delta \bar{G}^0 + RT \ln K \end{aligned} \tag{3}$$

ν_i is the stoichiometric coefficient for component *i*. *K* is the reaction quotient

$$K = \frac{(c_C / c^0)}{(c_A / c^0)(c_B / c^0)}$$

At equilibrium, $\Delta\bar{G} = 0$, so

$$\Delta\bar{G}^0 = -RT \ln K_a \tag{4}$$

where the association constant K_a is the value of the reaction quotient under equilibrium conditions. Dropping c^0, with the understanding that we must express concentration in M units:

$$K_a = \frac{c_C}{c_A c_B} \tag{5}$$

Since it is defined as a standard state quantity, K_a is a fundamental constant independent of concentration and pressure or volume, and is only dependent on temperature. The inverse of K_a is K_d the equilibrium constant for the C dissociation reaction $C \rightleftarrows A + B$.

Concentration and Fraction Bound

Experimentally one controls the total mass $m_{TOT} = m_A + m_B + m_C$, or concentration

$$c_{TOT} = c_C + c_A + c_B \tag{6}$$

The composition of system can be described by the fraction of concentration due to species i as

$$\theta_i = \frac{c_i}{c_{TOT}}$$
$$\theta_A + \theta_B + \theta_C = 1 \tag{7}$$

We can readily relate K_a to θ_i, but it is practical to set some specific constraint on the composition here. If we constrain the $A{:}B$ composition to be 1:1, which is enforced either by initially mixing equal mole fractions of A and B, or by preparing the system initially with pure C, then

$$K_a = \frac{4\theta_C}{(1-\theta_C)^2 c_{TOT}} \qquad (\theta_A{=}\theta_B)$$
$$= \frac{(1-2\theta_A)}{\theta_A^2 c_{TOT}} \tag{8}$$

This expression might be used for mixing equimolar solutions of binding partners, such as complementary DNA oligonucleotides. Using eq. (6) (with $c_A{=}c_B$) and (7) here, we can obtain the composition as a function of total concentration fraction as a function of the total concentration

$$\theta_C = \left(1+\frac{2}{K_a c_{TOT}}\right) - \sqrt{\left(1+\frac{2}{K_a c_{TOT}}\right)^2 - 1}$$

$$\theta_A = \frac{1}{2}\left(1-\theta_C\right)$$

In the case where A=B, applicable to homodimerization or hybridization of self-complementary oligonucleotides, we rewrite scheme (1) as the association of monomers to form a dimer

$$2M \rightleftarrows D$$

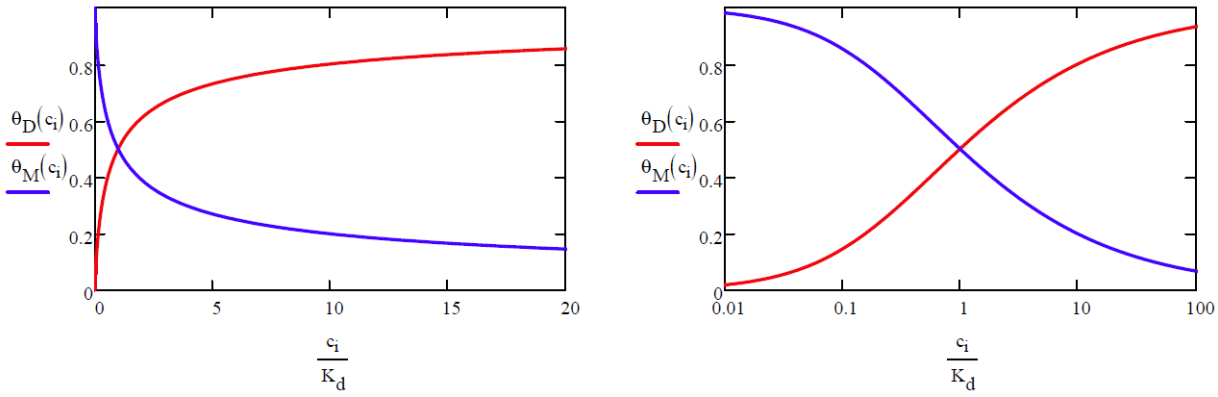
and find:

$$K_a = \theta_D / 2(1-\theta_D)^2 c_{TOT} \qquad (9)$$
$$K_a = (1-\theta_M)/2\theta_M{}^2 c_{TOT}$$

$$\theta_D = 1+\frac{1}{4c_{TOT}K_a}\left(1-\sqrt{1+8c_{TOT}K_a}\right) \qquad (10)$$

$$\theta_M = 1-\theta_D$$

These expressions for the fraction of monomer and dimer, and the corresponding concentrations of monomer and dimer are shown below. An increase in the total concentration results in a shift of the equilibrium toward the dimer state. Note that $c_{TOT} = (9K_a)^{-1} = K_d/9$ at $\theta_M = \theta_D = 0.5$,



For ligand receptor binding, ligand concentration will typically be much greater than that of the receptor, and we are commonly interested in fraction of receptors that have a ligand bound, θ_{bound}. Re-writing our association reaction as

$$L + R \rightleftarrows LR \qquad K_a = \frac{c_{LR}}{c_L c_R}$$

we write the fraction bound as

$$\theta_{\text{bound}} = \frac{c_{LR}}{c_R + c_{LR}} = \frac{c_L K_a}{1 + c_L K_a}$$

This is equivalent to a Langmuir absorption isotherm.

Temperature Dependence

The temperature dependence of K_a is governed by eq. (4) and the fundamental relation

$$\Delta G^0(T) = \Delta H^0(T) - T\Delta S^0(T) \tag{11}$$

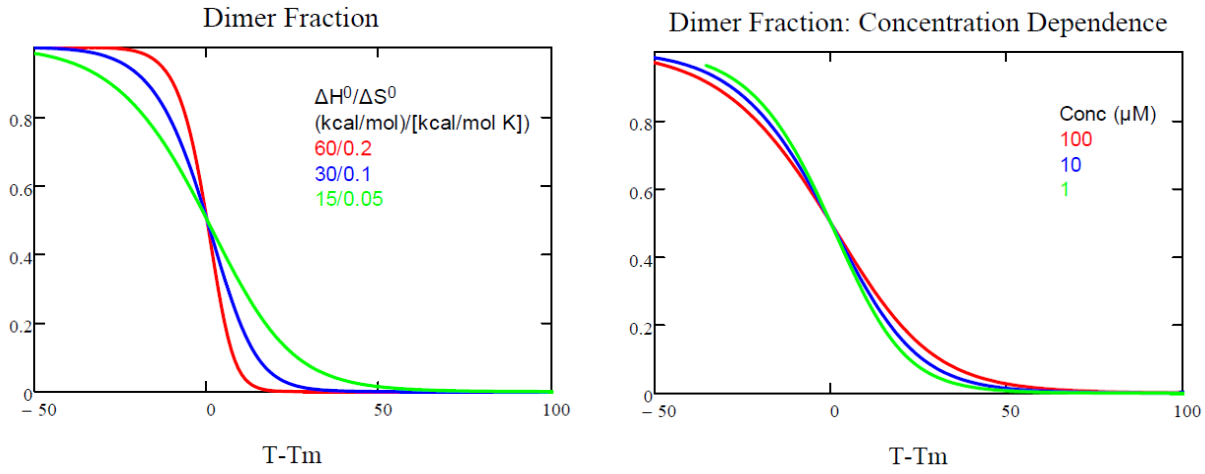
Under the assumption that ΔH^0 and ΔS^0 are temperature independent, we find

$$K_a(T) = \exp\left[-\frac{\Delta H_a^0}{RT} + \frac{\Delta S_a^0}{R}\right] \tag{12}$$

This allows us to describe the temperature-dependent composition of a system using the expressions above for θ_i. While eq. (12) allows you to predict a melting curve for a given set of thermodynamic parameters, it is more difficult to use it to extract those parameters from experiments because it only relates the value of K_d at one temperature to another.

Temperature is often used to thermally dissociate or melt dsDNA or proteins, and the analysis of these experiments requires that we define a reference temperature. In the case of DNA melting, the most common and readily accessible reference temperature is the melting temperature T_m defined as the point where the mole fractions of ssDNA (monomer) and dsDNA (dimer) are equal, $\theta_M = \theta_D = 0.5$. This definition is practically motivated, since DNA melting curves typically have high and low temperature limits that correspond to pure dimer or pure monomer. Then T_m is commonly associated with the inflection point of the melting curve or the peak of the first derivative of the melting curve. From eq. (9), we see that the equilibrium constants for the association and dissociation reaction are given by the total concentration of DNA: $K_a(T_m) = K_d(T_m)^{-1} = c_{\text{tot}}^{-1}$ and $\Delta G_d^0(T_m) = -RT_m \ln c_{tot}$. Furthermore, eq. (12) implies $T_m = \Delta H^0/\Delta S^0$.

The examples below show the dependence of melting curves on thermodynamic parameters, T_m, and concentration. These examples set a constant value of T_m ($\Delta H^0/\Delta S^0$). The concentration dependence is plotted for $\Delta H^0 = 15$ kcal mol^{-1} and $\Delta S^0 = 50$ cal mol^{-1} K^{-1}.



For conformational changes in macromolecules, it is expected that the enthalpy and entropy will be temperature dependent. Drawing from the definition of the heat capacity,

$$C_p = \left(\frac{\partial H}{\partial T}\right)_{N,P} = T\left(\frac{\partial S}{\partial T}\right)_{N,P}$$

we can describe the temperature dependence of ΔH^0 and ΔS^0 by integrating from a reference temperature T_0 to T. If ΔC_p is independent of temperature over a small enough temperature range, then we obtain a linear temperature dependence to the enthalpy and entropy of the form

$$\Delta H^0(T) = \Delta H^0(T_0) + \Delta C_p\left[T - T_0\right] \tag{13}$$

$$\Delta S^0(T) = \Delta S^0(T_0) + \Delta C_p \ln\left(\frac{T}{T_0}\right) \tag{14}$$

These expressions allow us to relate values of ΔH^0, ΔS^0, and ΔG^0 at temperature T to its value at the reference temperature T_0. From these expressions, we obtain a more accurate description of the temperature dependence of the equilibrium constant is

$$K_d(T) = \exp\left[-\frac{\Delta H_m^0}{RT} + \frac{\Delta S_m^0}{R} - \frac{\Delta C_p}{R}\left[1 - \frac{T_m}{T} - \ln\left(\frac{T}{T_m}\right)\right]\right] \tag{15}$$

where $\Delta H_m^0 = \Delta H^0(T_m)$ and $\Delta S_m^0 = \Delta S^0(T_m)$ are the enthalpy and entropy for the dissociation reaction evaluated at T_m.

Statistical Thermodynamics of Bimolecular Reactions

Statistical mechanics can be used to calculate K_a on the basis of the partition function. The canonical partition function Q is related to the Helmholtz free energy through

$$F = -k_B T \ln Q \tag{16}$$

$$Q = \sum_{\alpha} e^{-E_\alpha / k_B T} \tag{17}$$

where the sum is over all microstates (a particular configuration of the molecular constituents to a macroscopic system), Boltzmann weighted by the energy of that microstate E_α. The chemical potential of molecular species i is given by

$$\mu_i = -k_B T \left(\frac{\partial \ln Q}{\partial N_i} \right)_{V,T,\{N_{j \neq i}\}} \tag{18}$$

We will assume that we can partition Q into contributions from different molecular components of a reacting system such that

$$Q = \prod_i Q_i \tag{19}$$

The ability to separate the partition function stems from the assumption that certain degrees of freedom are separable from each other. When two sub-systems are independent of one another, their free energies should add ($F_{TOT} = F_1 + F_2$) and therefore their partition functions are separable into products: $Q_{TOT} = Q_1 Q_2$. Generally this separability is a result of being able to write the Hamiltonian as $H_{TOT} = H_1 + H_2$, which results in the microstate energy being expressed as a sum of two independent parts: $E_\alpha = E_{\alpha,1} + E_{\alpha,2}$. In addition to separating the different molecular species, it is also very helpful to separate the translational and internal degrees of freedom for each species, $Q_i = Q_{i,trans} Q_{i,int}$. The entropy of mixing originates from the translational partition function, and therefore will be used to describe concentration dependence.

For N_i non-interacting, indistinguishable molecules, we can relate the canonical and molecular partition function q_i for component i as

$$Q_i = \frac{q_i^{N_i}}{N_i!} \tag{20}$$

and using Sterling's approximation we obtain the chemical potential,

$$\mu_i = -RT \ln \frac{q_i}{N_i} \tag{21}$$

Following the reasoning in eqs. (2)–(5), we can write the equilibrium constant as

$$K_a = \frac{N_C}{N_A N_B} = \frac{q_C}{q_A q_B} V \tag{22}$$

This expression reflects that the equilibrium constant is related to the stoichiometrically scaled ratio of molecular partition functions per unit volume $K_a = \prod_i \left(q_i / V\right)^{\nu_i}$. Then the standards binding free energy is determined by eq. (4).

DNA Hybridization[2]

To illustrate the use of statistical thermodynamics to describe binding, we discuss simple models for the hybridization or melting of DNA. These models are similar to our description of the helix–coil transition in their approach. These do not distinguish the different nucleobases, only considering nucleotides along a chain that are paired (*bp*) or free (*f*).

Consider the case of the pairing between self-complementary oligonucleotides.

$$S + S \rightleftarrows D$$

S refers to any fully dissociated ssDNA and *D* to any dimer forms that involve two strands which have at least one base pair formed. We can then follow expressions for monomer–dimer equilibria above. The equilibrium constant for the association of single strands is

$$K_a = \frac{c_D}{c_S^2} \tag{23}$$

This equilibrium constant is determined by the concentration-dependent free-energy barrier for two strands to diffuse into contact and create the first base pair. If the total concentration of molecules present is either monomer or dimer, the form is

$$C_{tot} = c_S + 2c_D \tag{24}$$

then the fraction of the DNA strands in the dimer form is

$$\theta_D = \frac{2c_D}{C_{tot}} \tag{25}$$

and eq. (10) leads to

$$\theta_D = 1 + (4K_a C_{tot})^{-1} - \sqrt{(1 + (4K_a C_{tot})^{-1})^2 - 1} \tag{26}$$

We see that at the total concentration, which results in a dimer fraction $\theta_D = 0.5$, the association constant is obtained from $K_a = (9C_{tot})^{-1}$. This is a traditional description of the thermodynamics of a monomer–dimer equilibrium.

We can calculate K_a from the molecular partition functions for the *S* and *D* states:

$$K_a = \frac{q_D}{q_S^2}$$

2. C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*. (W. H. Freeman, San Francisco, 1980), Ch. 20; D. Poland and H. A. Scheraga, *Theory of Helix–Coil Transitions in Biopolymers*. (Academic Press, New York, 1970).

Different models for hybridization will vary in the form of these partition functions. For either state, we can separate the partition function into contributions from the conformational degrees of freedom relevant to the base-pairing and hybridization, and other degrees of freedom, $q_i = q_{i,conf}q_{i,ext}$. Assuming that the external degrees of freedom will be largely of an entropic nature, we neglect an explicit calculation and factor out the external degrees of freedom by defining the variable γ:

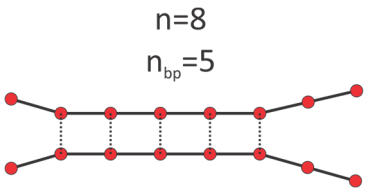
$$\gamma = \frac{q_{D,ext}C_{tot}}{q_{S,ext}^2}$$

then

$$\theta_D = 1 + \frac{q_{S,\text{int}}^2}{4\gamma q_{D,\text{int}}} - \sqrt{\left(1 + \frac{q_{S,\text{int}}^2}{4\gamma q_{D,\text{int}}}\right)^2 - 1}$$

Short Oligonucleotides: The Zipper Model

For short oligonucleotide hybridization, a common (and reasonable) approximation is the single stretch model, which assumed that base-pairing will only occur as a single continuous stretch of base pairs. This is reasonable for short oligomers ($n < 20$) where two distinct helical stretches separated by a bubble (loop) are unlikely given the persistence length of dsDNA. The zipper model refers to the single-stretch case with "perfect matching", in which only pairing between the bases in precisely sequence-aligned DNA strands is counted. As a result of these two approximations, the only dissociated base pairs observed in this model appear at the end of a chain (fraying).



The number of bases in a single strand is n and the number of bases that are paired is n_{bp}. For the dimer, we consider all configurations that have at least one base pair formed. The dimer partition function can be written as

$$\begin{aligned} q_{\text{D,int}}(n) &= \sigma \sum_{n_{bp}=1}^{n} g(n, n_{bp}) s^{n_{bp}} \\ &= \sigma \sum_{n_{bp}=1}^{n} (n - n_{bp} + 1) s^{n_{bp}} \end{aligned} \tag{27}$$

Here g is the number of ways of arranging n_{bp} continuous base pairs on a strand with length n; σ is the statistical weight for nucleating the first base pair; and s is the statistical weight for forming a base pair next to an already-paired segment: $s = e^{-\Delta\varepsilon_{bp}/k_BT}$. Therefore, in the zipper model, the equilibrium constant in eq. (23) between ssDNA and dimers involving at least one

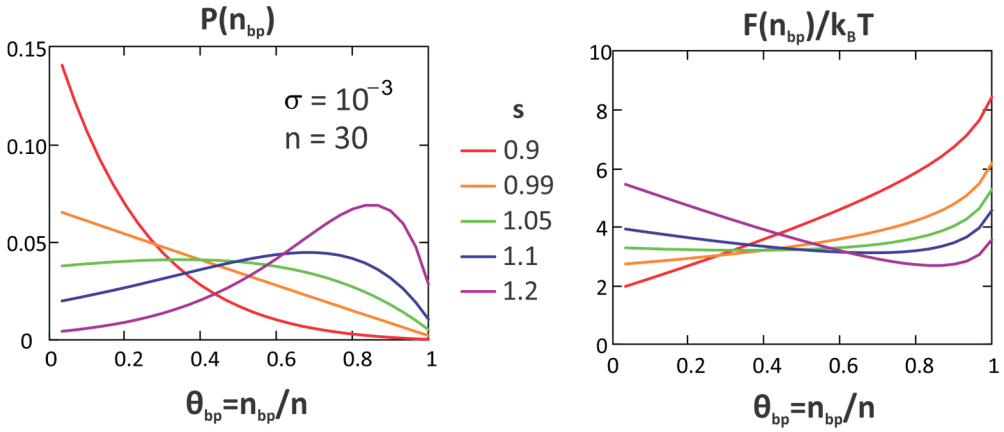
intact base pair is: $K_{\text{zip}} = \sigma s$. In the case of homogeneous polynucleotide chains, in which sliding of registry between chains is allowed: $q_{\text{D,int}}(n) = \sigma \sum_{n_{bp}=1}^{n} (n - n_{bp} + 1)^2 s^{n_{bp}}$. The sum in eq. (27) can be evaluated exactly, giving

$$q_{D,\text{int}}(n) = \frac{\sigma s}{(s-1)^2}\left[s^{n+1} - (n+1)s + n \right] \tag{28}$$

In the case that $s > 1$ ($\Delta\varepsilon_{bp} < 0$) and $n \gg 1$, $q_{D,\text{int}} \rightarrow \sigma s^n$. Also, the probability distribution of helical segments is

$$P_{bp}(n, n_{bp}) = \frac{(n - n_{bp} + 1)\sigma s^{n_{bp}}}{q_{D,\text{int}}} \qquad 1 \le n_{bp} \le n$$

The plot below shows illustrations of the probability density and associated energy landscape for a narrow range of s across the helix–coil transition. These figures illustrate a duplex state that always has a single free-energy minimum characterized by frayed configurations.



In addition to the fraction of molecules that associate to form a dimer, we must also consider the fraction of contacts that successfully form a base pair in the dimer state

$$\theta_{bp} = \frac{\langle n_{bp} \rangle}{n}$$

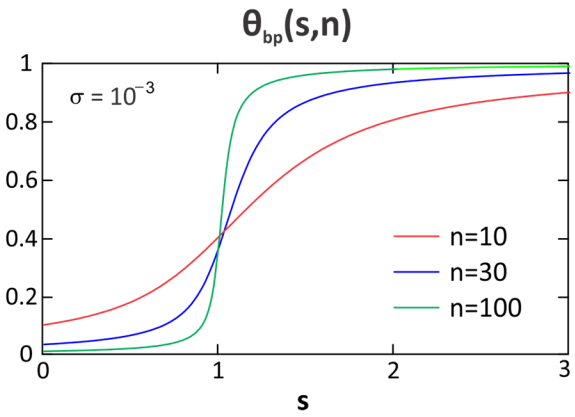
We can evaluate this using the identity

$$\langle n_H \rangle = \frac{s}{q}\frac{\partial q}{\partial s}$$

Using eq. (28) we have

$$\theta_{bp} = \frac{n s^{n+2} - (n+2)s^{n+1} + (n+2)s - n}{n(s-1)(s^{n+1} - s(n+1) + n)}$$

Similar to the helix–coil transition in polypeptides, θ_{bp} shows cooperative behavior with a transition centered at $s = 1$, which gets steeper with increasing n and decreasing σ.



Finally, we can write the total fraction of nucleobases that participate in a base pair as the product of the fraction of the DNA strands that are associated in a dimer form, and the average fraction of bases of the dimer that are paired.

$$\theta_{tot} = \theta_D \theta_{bp}$$

Bimolecular Kinetics

Returning to our basic two-state scheme, we define the rate constants k_a and k_d for the association and dissociation reactions:

$$A + B \underset{k_d}{\overset{k_a}{\rightleftharpoons}} C$$

From detailed balance, which requires that the total flux for the forward and back reactions be equal under equilibrium conditions:

$$K_a = \frac{k_a}{k_d}$$

The units for K_a are M^{-1}, $M^{-1}s^{-1}$ for k_a, and s^{-1} for k_d.

For the case where we explicitly consider the AB encounter complex:

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} (AB) \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} C$$

Schemes of this sort are referred to as reaction–diffusion problems. Note, this corresponds to the scheme used in Michaelis–Menten kinetics for enzyme catalysis, where AB is an enzyme–substrate complex prior to the catalytic step.

The kinetic equations corresponding to this scheme are often solved with the help of a steady-state approximation ($\partial[AB]/\partial t \approx 0$), leading to

$$\frac{d[C]}{dt} = k_a[A][B] - k_d[C]$$

$$k_a = \frac{k_1 k_2}{(k_{-1} + k_2)} \qquad k_d = \frac{k_{-1} k_{-2}}{k_{-1} + k_2}$$

Let's look at the limiting scenarios:

1) Diffusion controlled reactions refer to the case when reaction or final association is immediate once A and B diffusively encounter one another, i.e., $k_2 \gg k_{-1}$. Then the observed rate of product formation $k_a \approx k_1$, and we can then equate k_1 with the diffusion-limited association rate we have already discussed.
2) Pre-Equilibrium. When the reaction is limited by the chemical step, an equilibrium is established by which A and B can associate and dissociate many times prior to reaction, and the AB complex establishes a pre-equilibrium with the unbound partners defined by a nonspecific association constant $K'_a = k_1/k_{-1}$. Then the observed association rate is $k_a = k_2 K'_a$.

What if both diffusion and reaction within encounter complex matter? That is the two rates $k_1 \approx k_2$.

$$A + B \underset{}{\overset{k_a}{\rightleftharpoons}} AB \underset{}{\overset{k_{rxn}}{\rightleftharpoons}} C$$

Now all the rates matter. This can be solved in the same manner that we did for diffusion to capture by a sphere, but with boundary conditions that have finite concentration of reactive species at the critical radius. The steady-state solution gives:

$$k_{eff} = \frac{k_a k_{rxn}}{k_a + k_{rxn}}$$
$$k_{eff}^{-1} = k_a^{-1} + k_{rxn}^{-1}$$

k_{eff} is the effective rate of forming the product C. It depends on the association rate k_a (or k_1) and k_{rxn} is an effective forward reaction rate that depends on k_2 and k_{-1}.

Competing Factors in Diffusion–Reaction Processes

In diffusion–reaction processes, there are two competing factors that govern the outcome of the binding process. These are another manifestation of the familiar enthalpy–entropy compensation effects we have seen before. There is a competition between enthalpically favorable contacts in the bound state and the favorable entropy for the configurational space available to the unbound partners. Overall, there must be some favorable driving force for the interaction, which can be expressed in terms of a binding potential $U_{AB}(R)$ that favors the bound state. On the other hand, for any one molecule A, the translational configuration space available to the partner B will grow as R^2.

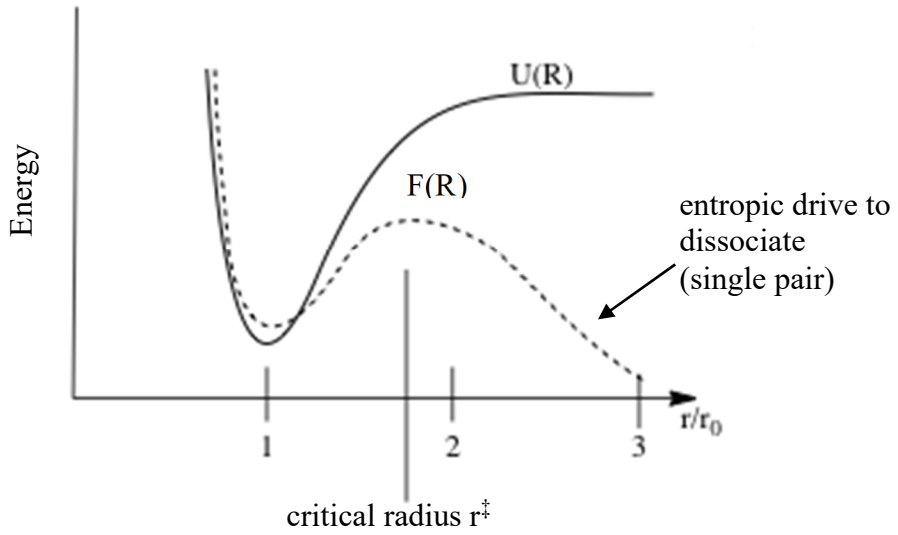
We can put these concepts together in a simple model.[1] The probability of finding B at a distance R from A is

$$P(R)dR = Q^{-1} e^{-U(R)/kT} 4\pi R^2 dR$$

where Q is a normalization constant. Then we can define a free energy along the radial coordinate

$$\begin{aligned} F(R) &= -k_B T \ln P(R)dR \\ &= U(R) - k_B T \ln R^2 - \ln Q \end{aligned}$$

1. D. A. Beard and H. Qian, *Chemical Biophysics; Quantitative Analysis of Cellular Systems*. (Cambridge University Press, Cambridge, UK, 2008).



Here $F(R)$ applies to a single *A*-*B* pair, and therefore the free energy drops continuously as R increases. This corresponds to the infinitely dilute limit, under which circumstance the partners will never bind. However, in practice there is a finite volume and concentration for the two partners. We only need to know the distance to the nearest possible binding partner $\langle R_{AB} \rangle$. We can then put an upper bound on the radii sampled on this free energy surface. In the simplest approximation, we can determine a cut off radius in terms of the volume available to each *B*, which is the inverse of the B concentration: $\frac{4}{3}\pi r_c^3 = [B]^{-1}$. Then, the probability of finding the partners in the bound state is

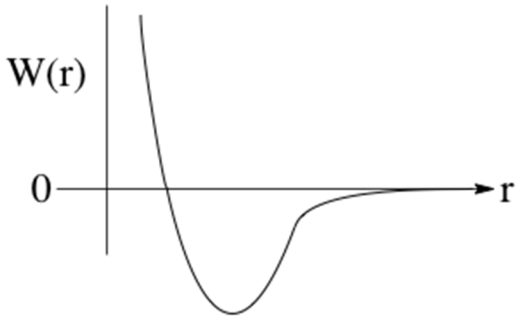
$$P_a = \frac{\int_0^{r^*} e^{-F(r)/k_B T} 4\pi r^2 dr}{\int_0^{r_c} e^{-F(r)/k_B T} 4\pi r^2 dr}$$

At a more molecular scale, the rates of molecular association can be related to diffusion on a potential of mean force. $g(r)$ is the radial distribution function that describes the radial variation of *B* density about *A*, and is related to the potential of mean force $W(r)$ through $g(r) = \exp[-W(r)/k_B T]$. Then the association rate obtained from the flux at a radius defined by the association barrier ($r = r^\dagger$) is

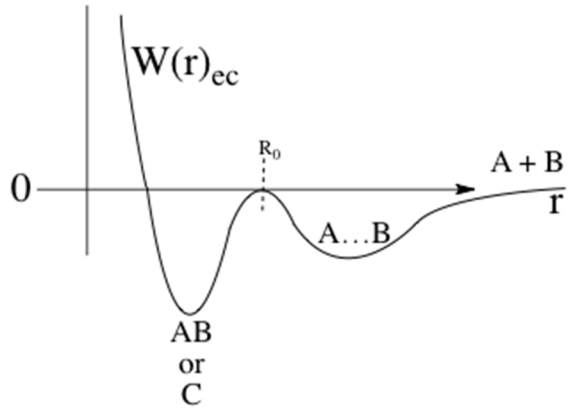
$$k_a^{-1} = \int_{r^\dagger}^{\infty} dr \left[4\pi r^2 D(r) e^{-W(r)/k_B T} \right]^{-1}$$

Here $D(r)$ is the radial diffusion coefficient that describes the relative diffusion of *A* and *B*. The spatial dependence reflects the fact that at small r the molecules do not really diffuse independently of one another.

For weakly attractive:



or structural/large solutes
or encounter complex



Diffusion-Limited Reactions[2]

Association Rate

The diffusion-limited association rate is typically approximated from the expression for the relative diffusion of A and B with an effective diffusion constant $D = D_A + D_B$ to within a critical encounter radius $R_0 = R_A + R_B$, as described earlier.

$$k_a = 4\pi R_0 f\left(D_A + D_B\right)$$

One can approximate association rates between two diffusing partners using the Stokes–Einstein expression: $D_A = k_B T / 6\pi\eta R_A$. For two identical spheres (i.e., dimerization) in water at T = 300 K, where η ~ 1 cP = 100 kg m^{-1} s^{-1},

$$k_a = \frac{8k_B T}{3\eta} = 6.6\times 10^9 M^{-1}s^{-1}$$

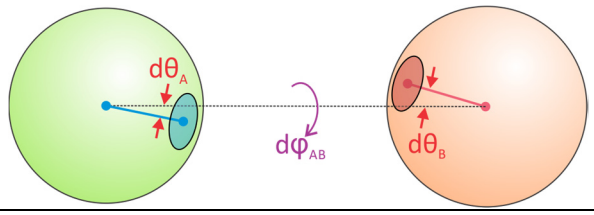
Note that this model predicts that the association rate is not dependent on the size or mass of the object.

For bimolecular reactions, the diffusion may also include those orientational factors that bring two binding sites into proximity. Several studies have investigated these geometric effects.

Example: Spheres with small binding patches

The combined probability that two binding patches are correctly oriented in both reference frames and that both are rotated into the correct azimuthal angle is:

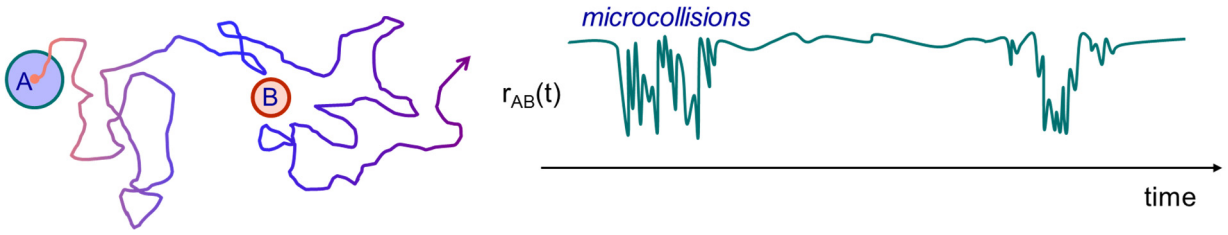
$$P_r = \frac{1}{2}(1-\cos\delta\theta_A)\frac{1}{2}(1-\cos\delta\theta_B)\frac{\delta\varphi_{AB}}{\pi}$$
$$\approx \frac{1}{16\pi}\delta\theta_A^2\delta\theta_B^2\delta\varphi_{AB}$$



During diffusive encounter in dilute solution, once two partners collide but do not react, there is a high probability of re-colliding with the same partner before diffusing over a longer range to a

2. D. Shoup, G. Lipari and A. Szabo, Diffusion-controlled bimolecular reaction rates. The effect of rotational diffusion and orientation constraints, Biophys. J. **36** (3), 697-714 (1981); D. Shoup and A. Szabo, Role of diffusion in ligand binding to macromolecules and cell-bound receptors, Biophys. J. **40** (1), 33-39 (1982).

new partner. Depending on concentration and the presence of interaction potentials, there may be 5–50 microcollisions with the same partner before encountering a new partner.



Diffusion-Limited Dissociation Rate

For the limit where associations are weak, k_1 and k_{-1} are fast and in equilibrium, and the dissociation is diffusion limited. Then we can calculate k_{-1}

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} AB$$

Now we consider boundary conditions for flux moving away from a sphere such that

$$C_B(\infty) = 0$$

$$C_B\left(R_0\right) = \left(\frac{4}{3}\pi R_0^3\right)^{-1}$$

The boundary condition for concentration at the surface of the sphere is written so that the number density is one molecule per sphere.

The steady state distribution of B is found to be

$$C_B(r) = \frac{3}{4\pi R_0^2 r}$$

The dissociation flux at the surface is

$$J = -D_B\left(\frac{\partial C_B}{\partial r}\right)_{r=R_0} = \frac{3D_B}{4\pi R_0^4}$$

and the dissociation frequency is

$$\frac{J}{4\pi R_0^2} = \frac{3D_B}{R_0^2}$$

When we also consider the dissociative flux for the other partner in the association reaction,

$$k_{-1} = k_d = 3\left(D_A + D_B\right)R_0^{-2}$$

Written in a more general way for a system that may have an interaction potential

$$k_d = \frac{4\pi D e^{U(R_0)/kT}}{\frac{4}{3}\pi R_0^3 \int_{R_0}^{\infty} e^{U(r)/kT} r^{-2} dr} = 3DR^* R_0^{-3}$$

Note that equilibrium constants do not depend on D for diffusion-limited association/dissociation

$$K_D = \frac{k_D}{k_A} = \frac{3DR_0^{-2}}{4\pi R_0 D} = \frac{3}{4\pi R_0^3}$$

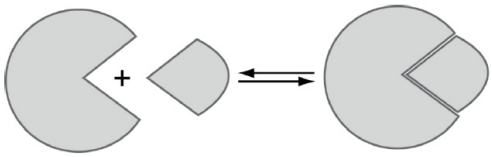
Note this is the inverse of the volume of a sphere.

Protein Recognition and Binding

The description of how a protein recognizes and binds a target is commonly discussed in terms of conceptual models.

Enzyme/Substrate Binding

Lock-and-Key (Emil Fisher)

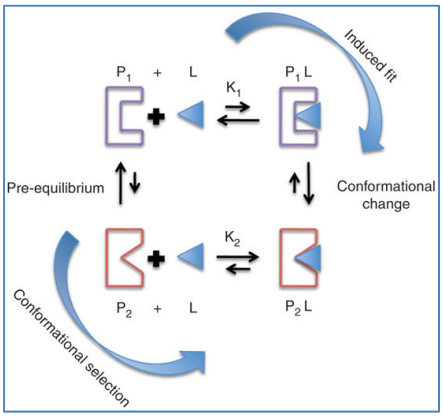


- Emphasizes shape complementarity
- Substrate typically rigid
- Concepts rooted in initial and final structure
- Does not directly address recognition

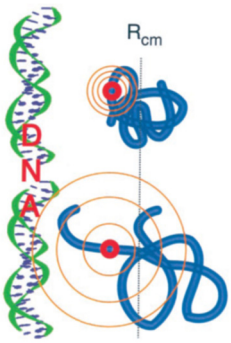
But protein-binding reactions typically involve conformational changes. Domain flexibility can give rise to dramatic increase in binding affinity. A significant conformational change/fluctuation may be needed to allow access to the binding pocket.

For binding a substrate, two models vary in the order of events for conformational change vs. binding event:

1) Induced fit (Daniel Koshland)
2) Conformational selection:Pre-existing equilibrium established during which enzyme explores a variety of conformations.



Protein–Protein Interactions

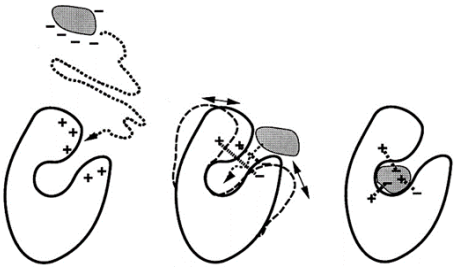


- Appreciation that structure is not the only variable
- Coupled folding and binding
  - Fold on contact
  - Fly-casting
- Both partners may be flexible

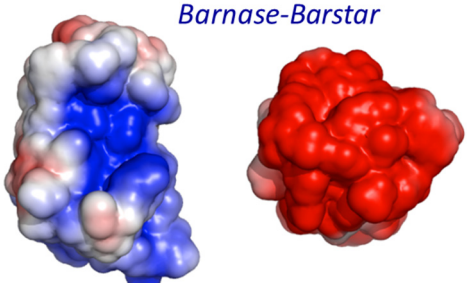
Forces Guiding Binding

Electrostatics

- Electrostatics play a role at long and short range
    - Long-range nonspecific interactions accelerate diffusive encounter
    - Short range guides specific contacts
- Electrostatic complementarity
- Electrostatic steering
- van der Waals, π-π stacking

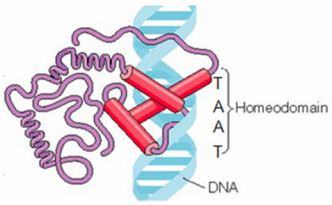


Barnase-Barstar



Shape and Geometry

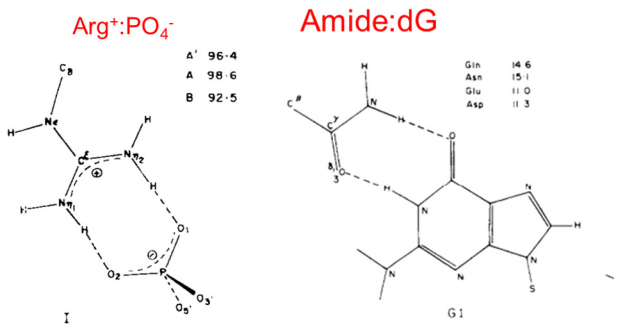
- Shape complementarity
- Orientational registry
- Folding
- Anchoring residues



Repressors: Helix-turn-helix
TAAT homeodomain

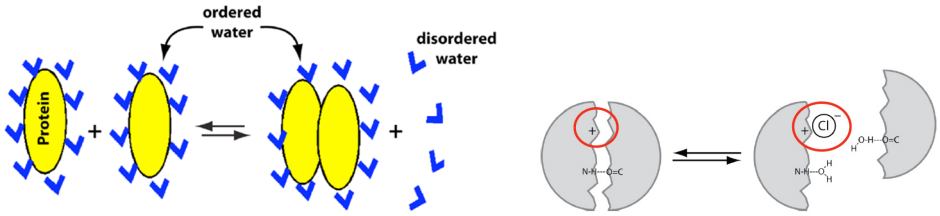
Hydrogen Bonding

- Short range
- Cross over from electrostatic to more charge transfer with strong HBs (like DNA, protein–DNA binding)
- Important in specificity



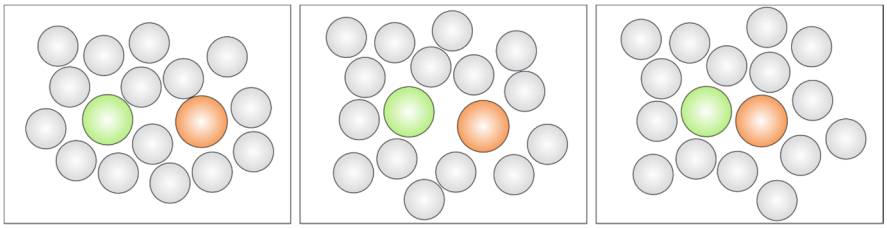
Solvation/Desolvation

- To bind, a ligand needs to desolvate the active site
- Bimolecular contacts will displace water
- Water often intimate binding participant (crystallographic waters)
- Hydrophobic patches
- Charge reconfiguration in electrolyte solutions at binding interface
- Electrostatic forces from water

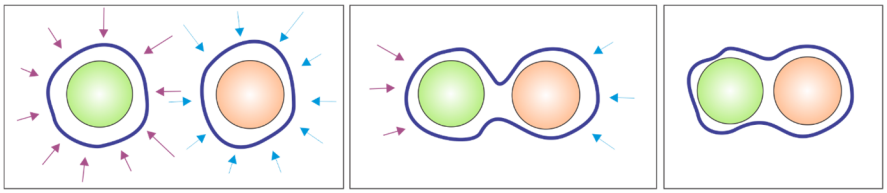


Depletion Forces

- Entropic effect
- Fluctuations that lead to an imbalance of forces that drives particles together
    - Crowding/Caging

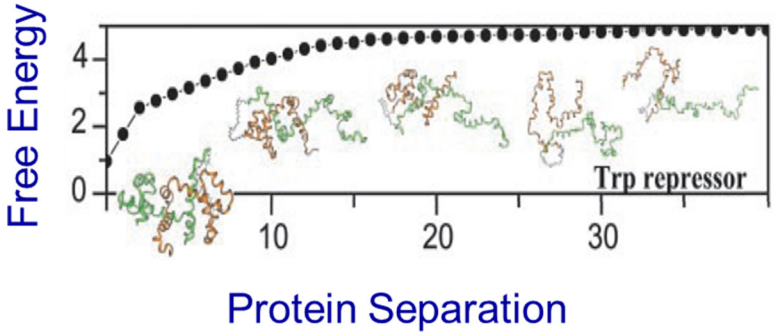


- Hydrophobicity
    - Dewetting and Interfacial Fluctuations



Folding/Conformational Change

- Disorder increases hydrodynamic volume
- Coupled folding and binding
    - Fly-casting mechanism
        - Partially unfolded partners
        - Long-range non-native interaction
        - Gradual decrease in free energy



Specificity in Recognition and Binding

Specificity in Recognition

What determines the ability for a protein to recognize a specific target amongst many partners? To start, let's run a simple calculation. Take the case that a protein (transcription factor) has to recognize a string of n sequential nucleotides among a total of N bases in a dsDNA.

- Assume that each of the four bases (ATGC) is present with equal probability among the N bases, and that there are no enthalpic differences for binding to a particular base.
- Also, the recognition of a particular base is independent of the other bases in the sequence. (In practice this is a poor assumption).
- The probability of finding a particular n nucleotide sequence amongst all n nucleotide strings is

$$\left(\frac{1}{4}\right)^n$$

- For a particular n nucleotide sequence to be unique among a random sequence of N bases, we need

$$\left(\frac{1}{4}\right)^n \geq \frac{1}{N}$$

- Therefore we can say

$$n \geq \frac{\ln N}{\ln 4}$$

Example

For the case that you want to define a unique binding site among N = 65k base pairs:

- A sequence of n = ln (65000)/ln(4) ≈ 8 base pairs should statistically guarantee a unique binding site.
- $n = 9 \rightarrow$ 262 kbp

This example illustrates that simple statistical considerations and the diversity of base combinations can provide a certain level of specificity in binding, but that other considerations are important for high fidelity binding. These considerations include the energetics of binding, the presence of multiple binding motifs for a base, and base-sequence specific binding motifs.

Energetics of Binding

We also need to think about the strength of interaction. Let's assume that the transcription factor has a nonspecific binding interaction with DNA that is weak, but a strong interaction for the target sequence. We quantify these through:

ΔG_1: nonspecific binding

ΔG_2: specific binding

Next, let's consider the degeneracy of possible binding sites:

g_n: number of nonspecific binding sites = $(N - n)$ or since $N \gg n$: $(N - n) \approx N$

g_s: number of sites that define the specific interaction: n

The probability of having a binding partner bound to a *nonspecific* sequence is

$$\begin{aligned} P_{\text{nonsp}} &= \frac{g_n e^{-\Delta G_1/kT}}{g_n e^{-\Delta G_1/kT} + g_s e^{-\Delta G_2/kT}} \\ &= \frac{(N-n)e^{-\Delta G_1/kT}}{(N-n)e^{-\Delta G_1/kT} + ne^{-\Delta G_2/kT}} \\ &= \frac{1}{1+\frac{n}{N}e^{-\Delta G/kT}} \end{aligned}$$

where $\Delta G = \Delta G_2 - \Delta G_1$.

We do not want to have a high probability of nonspecific binding, so let's minimize P_{nonsp}. Solving for ΔG, and recognizing $P_{\text{nonsp}} \quad 1$,

$$\Delta G \le -k_B T \ln\left[\frac{N}{n P_{\text{nonsp}}}\right]$$

Suppose we want to have a probability of nonspecific binding to any region of DNA that is $P_{\text{nonsp}} \le 1\%$. For $N = 10^6$ and $n = 10$, we find

$$\Delta G \approx -16 k_B T \quad \text{or} \quad -1.6 k_B T/\text{nucleotide}$$

for the probability that the partner being specifically bound with $P_{\text{sp}} > 99\%$.

Readings

1. G. Schreiber, G. Haran and H. X. Zhou, Fundamental aspects of protein−protein association kinetics, Chem. Rev. **109** (3), 839-860 (2009).
2. D. Shoup, G. Lipari and A. Szabo, Diffusion-controlled bimolecular reaction rates. The effect of rotational diffusion and orientation constraints, Biophys. J. **36** (3), 697-714 (1981).
3. D. Shoup and A. Szabo, Role of diffusion in ligand binding to macromolecules and cell-bound receptors, Biophys. J. **40** (1), 33-39 (1982).