Site-Specific Thermodynamics: Understanding Cooperativity in Molecular Recognition

Enrico Di Cera

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, Box 8231, St. Louis, Missouri 63110

Received December 16, 1997 (Revised Manuscript Received April 8, 1998)

Contents

I. Introduction

A prominent feature of biological macromolecules is the ability to accomplish diverse functions using cooperative interactions among structural domains. The best known example of this behavior is offered by hemoglobin, in which binding of oxygen to one heme affects the binding properties of other hemes in the molecule and oxygen release to the tissues is allosterically controlled by the uptake of protons and organic phosphates at other sites.[1] Cooperativity is not limited to ligand binding processes and allosteric proteins. It is an inherent component of protein stability, providing the necessary communication among residues of the protein to maintain the folded structure.[2] It is also a key player in determining secondary structure, as illustrated by the helix−coil transitions of biopolymers.[3] More recently, the developments of recombinant DNA technology have enabled a dissection of ligand recognition at the level of individual residues.[4] Systematic mutagenesis studies of binding epitopes have fostered the notion that cooperativity may be a fundamental ingredient of any recognition event.[5]

Enrico Di Cera was born in Palermo, Italy, in 1960. He received his M.D. degree from the Catholic University School of Medicine in Rome, Italy, in 1985. He was introduced to biological thermodynamics by Stan Gill and Jeffries Wyman during his postdoctoral experience in Boulder, CO. He joined Washington University School of Medicine in 1990, where he is currently Associate Professor. His research interests involve the thermodynamics of molecular recognition, as well as enzyme structure, function, and evolution.

The existence of cooperative interactions in macromolecular systems raises the question of how to decipher the mechanism underlying the communication among structural domains. One can imagine that a cooperative property, F, subject to experimental investigation is the result of the contribution of a number of individual structural domains of the macromolecule, so that

$$F = f_1 + f_2 + ... + f_N \quad (1)$$

where the f's encapsulate the individual contributions. Relevant examples of such properties are the following: protein stability, where the f's represent the contributions of particular folding units to the macroscopic free energy of folding; helix−coil transitions, where the f's represent the helix propensities of individual residues and their contribution to the helix state of the peptide as a whole; binding and linkage phenomena, where the f's denote the probabilities of binding to individual sites and F is the

S0009-2665(96)00135-5 CCC: $30.00
Published on Web 05/09/1998

average number of ligated sites; molecular recognition, where the f's denote contributions arising at each residue of the binding epitope to the binding free energy F.

An important consequence of cooperativity is that the properties of individual components embodied by the functions f's cannot be inferred from the behavior of each structural component separate from the system. Cooperativity changes the behavior of each domain when assembled into the whole macromolecule. Most often, the only quantity amenable to experimental investigation is the *global* quantity F in eq 1. This quantity has obvious limitations because it cannot define uniquely the individual components f's. The close connection between structure and function is embodied by the *site-specific* properties f's that, once specified, uniquely define F. The global quantity F may not reflect the true cooperativity pattern operating at the site-specific level. When individual components are summed, the exact nature of each particular contribution may be obscured by other terms that define the quantity F. Cooperativity can only be understood fully when the contribution of the site-specific components is sorted out.

The need for a description of cooperativity in terms of site-specific properties has been recognized for a long time and first emerged in the pioneering studies of Wegscheider[6] on the ionization reactions of polybasic acids. For many years, however, our understanding of cooperativity has been confined to the global description due to the limitations imposed by experimental techniques. Likewise, previous theoretical treatments of cooperativity have focused on the analysis of global effects.[7,8] Recent advances in various areas, and especially in structural biology and recombinant DNA technology, have made it possible to access information at the site-specific level. Global phenomena can now be dissected in terms of the contribution of individual binding sites, folding units, amino acid residues, or even atoms, thereby revealing the true and extraordinary complexity of cooperative effects in biology. These new advances have fostered the development of a thermodynamic description of site-specific effects.[9] Site-specific thermodynamics expands previous analyses of global effects and provides the conceptual and methodological tools to study cooperativity in a variety of systems. Much of the theory was developed to dissect ligand binding cooperativity.[9] Subsequent developments have encompassed the analysis of mutational effects in proteins[5] and have proved the general applicability of concepts and analytical methods originally introduced for the study of ligand binding processes. Relevant applications of the theory are summarized in this review.

II. Site-Specific Thermodynamics

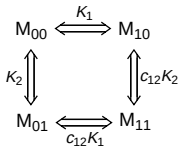
A. Difference between Global and Site-Specific Cooperativity

The simplest and most convincing argument to demonstrate the need for a site-specific description of cooperativity comes from consideration of a macromolecule M composed of two binding sites for a ligand L. We shall assume that temperature and pressure are constant and that the macromolecule and the ligand do not change their aggregation state when free or bound. There are two ways to describe the binding equilibria in the system. The *global* description focuses on the overall behavior of the two sites,[7,8] whereas the *site-specific* description takes into account how binding occurs at each site.[9] In the global description there are two reactions to be considered. These reactions define uniquely the binding isotherm accessible to experimental measurements, from which the average number of ligated sites is obtained as a function of ligand concentration. In the first reaction, $M + L = ML$, the free macromolecule interacts with one molecule of ligand to form the singly ligated intermediate ML. The equilibrium binding constant for this reaction is $2k_1$, where k_1 is the binding affinity for the ligation of a site and the factor accounts for the possible ways of generating the singly ligated intermediate from the unligated form of the macromolecule. In the second reaction, $ML + L = ML_2$, the singly ligated intermediate binds a second molecule of ligand to form the doubly ligated species ML_2. The equilibrium constant for this reaction is $k_2/2$, where k_2 is the binding affinity for the ligation of the second site and the factor at the denominator accounts for the number of ways one can generate the singly ligated species from the doubly ligated intermediate. In general, for the reaction $ML_{j-1} + L = ML_j$, the binding constant is $[(N - j + 1)/j]k_j$, where k_j is the binding affinity for ligation of the jth site, the factor at the numerator accounts for the number of ways of generating ML_j from ML_{j-1} and that at the denominator accounts for the number of ways of generating ML_{j-1} from ML_j. The equilibrium constants measuring the affinity of each ligation step are called *stepwise* binding constants. They do not distinguish between the sites, 1 and 2, although they depend on the properties of both sites. Cooperativity is observed when $k_1 \neq k_2$, in which case binding of the second ligand molecule takes place with an affinity different than that of the first ligand molecule. Positive cooperativity implies $k_1 < k_2$, whereas negative cooperativity demands $k_1 > k_2$. The case $k_1 = k_2$ reflects the absence of cooperativity.

We shall not discuss the graphical manifestations of cooperativity that are dealt with in detail elsewhere.[7-9] These signatures are useful in the analysis of experimental data but bear little on the conceptual framework that we are interested in discussing here. If the system shows cooperative binding of ligand L ($k_1 \neq k_2$), what is the underlying mechanism that produces this effect? On the other hand, if the system shows no presence of cooperativity ($k_1 = k_2$), does it imply that the two sites are independent? These simple and important questions may be difficult to answer in the global description due to the lack of information on the behavior of sites 1 and 2. The stepwise binding constants reflect the average properties of sites 1 and 2, whereas one needs to know the detailed behavior of each site as a function of the ligand concentration. Unraveling this information pertains to the site-specific description. The

binding reactions of ligand L with the macromolecule M in the site-specific description directly identify the two sites in their ligation state. This makes it necessary to introduce two site-specific binding constants, K_1 and K_2, referring respectively to the reaction of ligand L with site 1 when site 2 is unligated and binding to site 2 when site 1 is unligated. The interaction between the sites is expressed by a third independent parameter, c_{12}, that reflects the presence of positive ($c_{12} > 1$), negative ($c_{12} < 1$) or no ($c_{12} = 1$) cooperativity. By virtue of this interaction, binding to site 1 when site 2 is ligated occurs with a binding constant $c_{12}K_1$, and for site 2, when site 1 is ligated, the binding constant is $c_{12}K_2$.

The nature of these parameters and the reactions in the site-specific description are best understood from the thermodynamic cycle



(2)

where M_{00} is the free macromolecule, M_{10} the singly ligated form with site 1 bound and site 2 free, M_{01} the analogous intermediate with site 2 bound and site 1 free, and M_{11} the doubly ligated form. Energy conservation in the cycle gives rise to only three independent parameters to describe the four possible reactions. The reciprocity of site−site interactions is a consequence of energy conservation in the cycle. So, if site 1 affects site 2, site 2 must affect site 1 and to the same extent.

Analysis of binding in terms of the site-specific description raises a seemingly paradoxical issue. In the global description, only two independent parameters (k_1 and k_2) are sufficient to define the properties of the system. In the site-specific description, on the other hand, there are three independent parameters (K_1, K_2, and c_{12}) to be taken into account. What is the origin of this apparent discrepancy? Simple considerations on the equilibria involving the two sites lead to the following relationships between global and site-specific parameters:

$$2k_1 = \frac{[ML]}{[M]x} = \frac{[M_{10}] + [M_{01}]}{[M_{00}]x} = K_1 + K_2 \quad (3)$$

$$\frac{k_2}{2} = \frac{[ML_2]}{[ML]x} = \frac{[M_{11}]}{([M_{10}] + [M_{01}])x} = \frac{c_{12}K_1K_2}{K_1 + K_2} \quad (4)$$

where x is the ligand concentration. The apparent discrepancy arises because it is not possible to uniquely derive the site-specific parameters K_1, K_2, and c_{12} from knowledge of the global parameters k_1 and k_2. On the other hand, if the site-specific parameters are known, then the global parameters can be determined uniquely. Hence, the global description is incapable of deciphering what goes on at the level of individual sites. This fact has been recognized for a long time, and the need for a local description of binding processes finds its origin in the early work on the dissociation of polyvalent substances.[6,10−13]

The limitations of the global description become even more apparent when we examine the nature of cooperativity. The condition for cooperativity in the global description is $k_1 \neq k_2$ and can be formally represented using eqs 3 and 4 as the difference:

$$k_2 - k_1 = \frac{4c_{12}K_1K_2 - (K_1 + K_2)^2}{2(K_1 + K_2)} \quad (5)$$

The sign of the expression

$$\Delta = 4c_{12}K_1K_2 - (K_1 + K_2)^2 =$$

$$K_2^2\left[4c_{12}\frac{K_1}{K_2} - \left(1 + \frac{K_1}{K_2}\right)^2\right] = K_1^2\left[4c_{12}\frac{K_2}{K_1} - \left(1 + \frac{K_2}{K_1}\right)^2\right] \quad (6)$$

defines the nature of cooperativity. $\Delta = 0$ denotes absence of cooperativity and provides the cutoff between positive ($\Delta > 0$) and negative ($\Delta < 0$) cooperativity. The condition for the absence of cooperativity in the global description does not necessarily coincide with the condition $c_{12} = 1$ that reflects the true absence of interactions between sites 1 and 2. Only when $K_1 = K_2$ are the two conditions identical. If the binding sites have different affinities, there is always a value of $c_{12} > 1$ such that $\Delta = 0$. This means that positive interactions between two sites that bind with different affinities may not manifest themselves in the global description as positive cooperativity.

A direct illustration of this fact is given in Figure 1, where the logarithm of c_{12} is plotted versus the logarithm of the ratio K_1/K_2. Plotting versus the logarithm of the ratio K_2/K_1 is completely equivalent because of the symmetry of eq 6. The continuous line represents the relation between c_{12} and the ratio K_1/K_2 such that $\Delta = 0$ in eq 6. On this line, any combination of site-specific binding constants K_1 and K_2 and interaction constant c_{12} yields $k_1 = k_2$ in the global description. The region above this line is characterized by positive cooperativity in the global description ($k_1 < k_2$), whereas the region below the line characterizes negative cooperativity ($k_1 > k_2$). The discontinuous line gives the condition for the absence of true interactions between the sites ($c_{12} = 1$). Above this line the sites are positively linked and below it they are negatively linked. The two lines in the plot define three regions. In region I, defined by $c_{12} < 1$, there is no ambiguity between global and site-specific cooperativity. When binding to one site opposes binding to the other site at the site-specific level, the result is negative cooperativity in the global description. At the boundary $c_{12} = 1$, the system is always negatively cooperative in the global description, unless $K_1 = K_2$ and the two sites bind with the same affinity. Independent sites binding with different affinities therefore mimic negative cooperativity in the global description. In region II negative cooperativity in the global description is observed even though the two sites interact in a positive manner. This is a consequence of the heterogeneity

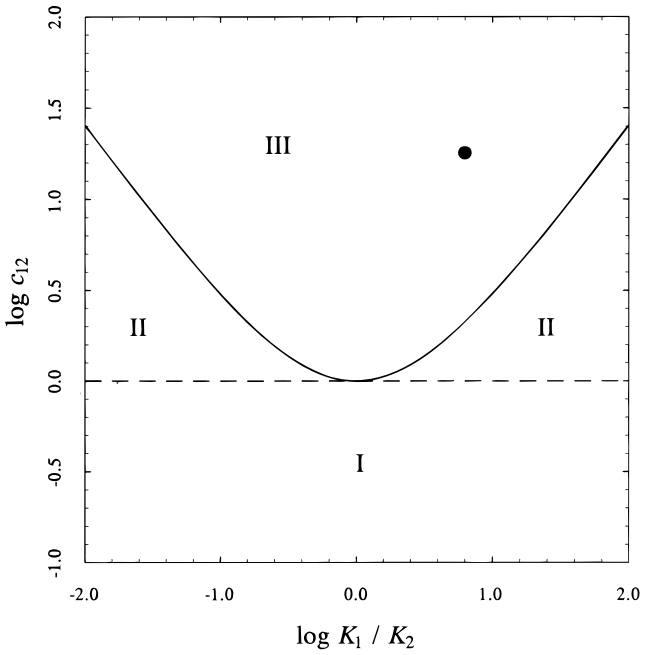


Figure 1. Difference between global and site-specific cooperativity for a two-site system. The logarithm of the interaction constant c_{12} is plotted versus the logarithm of the ratio between the site-specific binding parameters K_1 and K_2. The continuous line depicts the relation between c_{12} and the ratio K_1/K_2 such that $\Delta = 0$ in eq 6. This line defines the boundary between negative (below the line) and positive (above the line) cooperativity in the global description. The discontinuous line, on the other hand, defines the boundary between negative (below the line) and positive (above the line) cooperativity in the site-specific description. The two lines define three regions. In region I, there is no ambiguity between global and site-specific cooperativity. A value of $c_{12} < 1$ always results in negative cooperativity. Likewise, in region III positive interactions between the sites always result in positive global cooperativity. Region II, defined symmetrically between the lines, is ambiguous because positive interactions between the sites ($c_{12} > 1$) result in negative global cooperativity because of the heterogeneity of the sites ($K_1 \neq K_2$). The filled circle depicts the values of site-specific parameters for Ca^{2+} binding to calbindin (see section II.C).

of the binding affinities that opposes the favorable coupling between the sites. Region III again leaves no ambiguity between global and site-specific cooperativity, in this case both positive.

The conclusion to be drawn from Figure 1 is that cooperativity in the global description is at most a crude approximation of the true pattern of interaction between the sites. In the case of positive cooperativity, the true extent of interactions is always underestimated if $K_1 \neq K_2$. In fact, for any given value of $c_{12} > 1$, there is always a value of K_1/K_2 such that $\Delta = 0$, or even $\Delta < 0$. No matter how strongly two sites are positively coupled at the site-specific level, positive cooperativity in the global description is only seen in region III. Positive cooperativity can be made arbitrarily small and even turned into negative cooperativity in the global description if the binding affinities of the two sites differ significantly. For example, in a system where the affinities of the two sites differ by a factor of 100 ($K_1/K_2 = 0.01$), but binding to one site increases the affinity of the other site by a factor of 10 ($c_{12} = 10$), one has $\Delta < 0$ and negative cooperativity appears in the global description even though the sites are strongly coupled in a positive manner. For this system to show positive cooperativity in the global description, the value of c_{12} must exceed 25. In a system where $K_1/K_2 = 0.001$ the value of c_{12} must exceed 250, and so on. The limitations of the global description are not confined to the case of positive cooperativity. When negative cooperativity is observed in the global description, the system can actually be positively or negatively cooperative at the site-specific level. Particularly interesting is the case $c_{12} = 1$, which always leads to negative cooperativity in the global description if $K_1 \neq K_2$. The heterogeneity of the binding affinities of the sites generates per se a misleading pattern of negative interactions.

In summary, cooperativity as assessed by the global parameters k's does not reflect the true pattern of interaction between the sites, unless the sites have the same affinity. If the sites bind with different affinities, then positive cooperativity in the global description always underestimates the coupling between the sites. Negative cooperativity can be totally misleading, since it may be associated at the site-specific level with positive coupling between the sites or absence of interactions. In the case of negative coupling between the sites, on the other hand, the interaction may be overestimated in the global description. Absence of cooperativity in the global description can be misleading as well. Positive coupling between the sites can be exactly countered by heterogeneity of their binding affinity. The true cooperative nature of the interactions between the sites can only be resolved from knowledge of the value of c_{12}, but this requires information on how binding occurs at each site of the system.

B. The Wegscheider Principle and the Analysis of Ionization Reactions

The importance of dissecting cooperativity at the site-specific level was obvious even to early investigators of ionization equilibria of polybasic acids. In 1895, the Austrian chemist Wegscheider introduced an ingenious strategy to structurally perturb a system to mimic the properties of reduced systems containing a fewer number of sites.[6] The comparison of the original system and its reduced versions would then be used to extract information on the behavior of individual ionizable groups. It is quite instructive to comment on Wegscheider's strategy because it helps understand more elaborate, but conceptually similar, strategies currently employed in the study of proteins and nucleic acids.

The question that Wegscheider posed was as follows. If only global properties of a system can be accessed experimentally, is it possible to derive relevant information on site-specific parameters? Wegscheider thought that replacement of the ionizable carboxylate in a polybasic acid with methyl or ethyl esters could mimic the protonated state of the group and reduce the number of sites to be studied by direct titration. By selectively replacing groups at each of the carboxylates of interest he could in turn study how protonation of these groups in the original

Table 1. pK_a's of Glutamic Acid and Its Ethyl Esters[a]

	^{1}pK_a	^{2}pK_a	^{3}pK_a
glutamic acid	2.155	4.324	9.960
α-ethyl glutamate	3.846	7.838	
γ-ethyl glutamate	2.148	9.19	
ethyl glutamate	7.035		

[a] The site-specific parameters for proton binding to glutamic acid, derived from analysis of these pK_a's, are ($\log_{10}$ values are given in parentheses) $K_1 = 5.69 \times 10^4$ M^{-1} (4.755), $K_2 = 9.19 \times 10^9$ M^{-1} (9.960), $K_3 = 1.26 \times 10^5$ M^{-1} (5.101), $c_{12} = 0.00755$ (−2.122), $c_{13} = 0.353$ (−0.452), $c_{23} = 0.17$ (−0.770), and $c_{123} = 0.00042$ (−3.377).

molecule would affect protonation of other ionizable groups and therefore extract important site-specific information from the system.

Neuberger[14] exploited Wegscheider's idea to completely resolve the three ionization reactions of glutamic acid into their site-specific components.[15] For such a system there are three binding sites for the proton: the α-carboxyl (site 1), amino (site 2), and γ-carboxyl (site 3) groups. Titration of glutamic acid yields three pK_a's (Table 1) that reflect the composite behavior of the three sites. The stepwise binding constants and the site-specific parameters are directly defined by these pK_a's as follows:[9,15]

$$10^{9.960} = \frac{[M_{100}] + [M_{010}] + [M_{001}]}{[M_{000}]}\frac{1}{h} = 3k_1 = K_1 + K_2 + K_3 \quad (7)$$

$$10^{4.324} = \frac{[M_{110}] + [M_{101}] + [M_{011}]}{[M_{100}] + [M_{010}] + [M_{001}]}\frac{1}{h} = k_2 = \frac{c_{12}K_1K_2 + c_{13}K_1K_3 + c_{23}K_2K_3}{K_1 + K_2 + K_3} \quad (8)$$

$$10^{2.155} = \frac{[M_{111}]}{[M_{110}] + [M_{101}] + [M_{011}]}\frac{1}{h} = \frac{k_3}{3} = \frac{c_{123}K_1K_2K_3}{c_{12}K_1K_2 + c_{13}K_1K_3 + c_{23}K_2K_3} \quad (9)$$

The M's refers to the various protonated intermediates of glutamic acid, and h is the proton concentration. Knowledge of the three stepwise binding constants derived from direct titration of glutamic acid does not suffice to define uniquely the set of seven independent site-specific parameters necessary to completely solve the ionization reactions at the level of individual sites. In the site-specific description there are three site-specific binding constants, K_1, K_2, and K_3, three second-order coupling constants, c_{12}, c_{13}, and c_{23}, and one third-order coupling constant, c_{123}. Resolution of seven parameters demand at least seven independent constraints from experimental data. To this end, Neuberger synthesized esters of either the α- or γ-carboxyl groups and titrated the other groups (Table 1). The assumption embodied by the Wegscheider principle is that the ethyl ester mimics the protonated state of the carboxyl group. Under this assumption, the following two relations for α-ethyl glutamate apply:

$$10^{7.838} = \frac{[M_{110}] + [M_{101}]}{[M_{100}]}\frac{1}{h} = 2k_1' = c_{12}K_2 + c_{13}K_3 \quad (10)$$

$$10^{3.846} = \frac{[M_{111}]}{[M_{110}] + [M_{101}]}\frac{1}{h} = \frac{k_2'}{2} = \frac{c_{123}K_2K_3}{c_{12}K_2 + c_{13}K_3} \quad (11)$$

They provide two additional constraints for the solution of the problem. Analogous expressions for γ-ethyl glutamate

$$10^{9.190} = \frac{[M_{101}] + [M_{011}]}{[M_{001}]}\frac{1}{h} = 2k_1'' = c_{13}K_1 + c_{23}K_2 \quad (12)$$

$$10^{2.148} = \frac{[M_{111}]}{[M_{101}] + [M_{011}]}\frac{1}{h} = \frac{2k_1''}{2} = \frac{c_{123}K_1K_2}{c_{13}K_1 + c_{23}K_2} \quad (13)$$

complete the set of seven constraints needed to resolve the seven independent site-specific parameters. The results are shown in Table 1. The uniqueness of the results can be tested by calculating the pK_a of the amino group in ethyl glutamate as

$$pK_a = \log_{10}\frac{[M_{111}]}{[M_{101}]h} = \log_{10}\left(\frac{c_{123}}{c_{13}}K_2\right) \quad (14)$$

The predicted value of 7.035 is identical to that found experimentally.

It is from knowledge of the site-specific parameters for the ionization reactions of glutamic acid that a more close connection with the structure of the amino acid can be drawn. When all groups are deprotonated, binding of the proton occurs with high affinity to the amino group and with low affinity to the carboxyl groups. The α-carboxyl group binds the proton with slightly lower affinity compared to the γ-carboxyl group, due to the proximity of the amino group and the unfavorable electrostatic coupling experienced when this group is protonated. The interaction constants are all less than 1, indicating the presence of site-specific negative cooperativity in the protonation reactions. This effect acts in concert with the extreme heterogeneity of the sites to produce a very pronounced negatively cooperative proton binding curve for glutamic acid, as shown in Figure 2. Negative coupling among the sites is expected from electrostatic considerations and decreases with the distance between neighbor groups. The two carboxyl groups are located far enough away that $c_{13} \approx 1$. On the other hand, protonation of the amino group has an effect almost 3 orders of magnitude larger on the α- than the γ-carboxyl group, due to the proximity of the former group.

An alternative solution to eqs 7−9 can be found by assuming $c_{12} = c_{13} = c_{23} = c_{123} = 1$ and solving for the three independent site-specific binding constants.

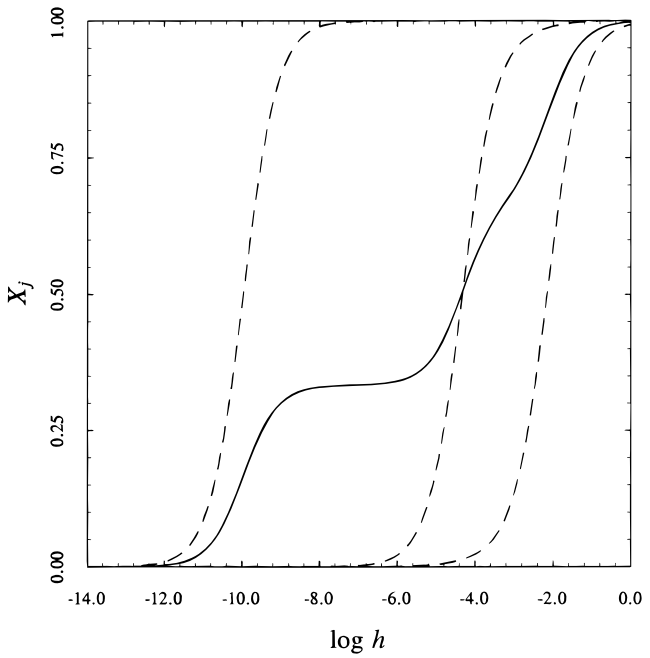


Figure 2. Proton binding curves of the three ionizable groups of glutamic acid: α-carboxyl group (discontinuous curve at right), amino group (discontinuous curve at left), and γ-carboxyl group (discontinuous curve in the middle). The sum of these three curves, divided by the number of sites, gives the global proton binding curve measured experimentally (continuous line). Curves were drawn using the parameter values listed in Table 1.

The result is $K_1 = 10^{2.158}$, $K_2 = 10^{9.960}$, and $K_3 = 10^{4.321}$, which suggests that the pK_a's measured from titration of glutamic acid correspond quite closely to the three site-specific pK_a's. This erroneous conclusion would make the affinity of the α-carboxyl group 2 orders of magnitude lower than that of the γ-carboxyl group and nearly 3 orders of magnitude lower than the value correctly derived from the analysis of Neuberger's results. The large discrepancy found between the simplifying assumption and the correct result in a simple molecule like glutamic acid represents a serious warning for more complex cases involving ionization reactions in proteins. It has become common practice to assume that ionization of protein residues occurs almost independently of other proton binding events in the protein in an attempt to simplify the calculations. This assumption is likely to be wrong, and the pK_a's estimated from independent ionization reactions may have little bearing on the actual binding affinities of the ionizable groups.

Can the strategy embodied by the Wegscheider principle be extended to macromolecular systems? Consider the problem of calculating the pK_a of ionizable groups in a protein.[16,17] In this case, experimental measurements of the site-specific proton binding curve of each ionizable group may be unfeasible. For a protein containing 20 such groups, there are a total of $2^{20} \approx 10^6$ total configurations and as many site-specific parameters to be resolved from experimental data. No currently available technique can provide such information. However, many residues such as Asp, Glu, Arg, Lys, and Tyr do not ionize in the pH range of physiological interest and can be treated as fully protonated or deprotonated. The remaining groups of the protein form a reduced system in the Wegscheider sense whose description requires a significantly reduced number of parameters. The idea of fixing the ionization state of selected residues in a protein to calculate more efficiently the ionization properties of other groups has been implemented by Bashford and Karplus and is known as the "reduced-site" model.[18] Reduced systems can also be generated empirically by replacing the residue of interest, say a His, with groups that mimic the protonated or unprotonated state of that residue. However, if these substitutions affect other properties of the macromolecule, such as the ionization of other groups, or the coupling of His with other residues, the assumption central to the entire approach is invalidated. In general, the assumption that a given substitution actually mimics a particular ligation state for the His may be questioned. More importantly, even though this strategy may be successful in the case of ionization reactions, it will certainly fail in the case of other ligands such as metal ions, peptides, or nucleic acids, whose recognition by proteins entails extended structural domains. For example, in the specific case of Ca^{2+} binding to an EF-hand,[19] it is difficult to envision simple substitutions that can exactly mimic the Ca^{2+}-bound or the Ca^{2+}-free form of the site. When binding of a ligand involves several protein residues, any perturbation is likely to be extensive and the expectation that it mimics a specific ligation state becomes unrealistic. Nonetheless, a refined version of the Wegscheider principle is still applicable under certain circumstances and provides a powerful approach to the study of site-specific energetics.

C. Using Mutants To Resolve Site-Specific Parameters: Ca^{2+} Binding to Calbindin

Under suitable conditions, site-specific parameters can be obtained from analysis of global properties. Although these conditions may be difficult to reproduce in general for any system of interest, it is instructive to consider the potential advantages of the approach when feasible. Consider a system containing N sites, each existing in two possible states, free or bound. There are a total of 2^N possible configurations in this system, $2^N - 1$ of which are independent if one is chosen as reference. The sum of the concentrations of all possible configurations relative to the concentration of the reference species, the unligated state, defines the partition function of the system Ψ.[9] The partition function is a polynomial expansion in the ligand concentration x of degree N.[7-9] From the partition function, all of the relevant global properties of the system can be derived. For example, the average number of ligated sites, X, accessible to experimental measurements through titration is $X = \mathrm{d}\ln\Psi/\mathrm{d}\ln x$. The configurations defining the partition function can be split in two sets: one containing all configurations with a given site, say site j, unligated and the other containing all configurations with site j ligated. These sets define partition functions of reduced or contracted systems. Let ${}^0\Psi_j$ and ${}^1\Psi_j$ be the partition functions

of the two sets constrained by the particular ligation, free (0) or bound (1), of site j. The partition function of the system can then be written as[9]

$$\Psi = {}^0\Psi_j + {}^1\Psi_j K_j x \tag{15}$$

The factor $K_j x$ arises because all intermediates with site j bound contain this term in the partition function. The binding probability to site j is evidently

$$X_j = K_j x \frac{{}^1\Psi_j}{\Psi} = 1 - \frac{{}^0\Psi_j}{\Psi} \tag{16}$$

with the conservation relation analogous to eq 1

$$X = X_1 + X_2 + ... + X_N \tag{17}$$

The quantity X_j cannot be accessed experimentally if only global properties such as X are measured, as already pointed out in the Introduction. However, the function X_j can be reconstructed indirectly using ad hoc substitutions at site j. This assists in the resolution of some of the site-specific parameters in the system.

Consider a perturbation of site j, say a chemical modification induced by a site-directed mutation. The perturbation either can affect the properties of the site where it applies or can carry over to other sites. Assume that the perturbation remains localized at site j, with a negligible secondary effect on other sites. We speak in this case of a first-order perturbation that changes the value of K_j to K_j', while it leaves the value of all other parameters unchanged. The smaller the perturbation, the more likely it will cause a first-order effect. Then, for the perturbed system, we have

$$\Psi' = {}^0\Psi_j + {}^1\Psi_j K_j' x \tag{18}$$

The contracted partition functions are the same as those for the wild-type, or unperturbed system, since they do not depend on K_j. Both Ψ and Ψ' are accessible experimentally from integration of measurements of the average number of ligated sites as a function of ligand concentration. Hence, the function

$$f_j = 1 - \frac{\Psi'}{\Psi} = \left(1 - \frac{K_j'}{K_j}\right) X_j = \eta_j X_j \tag{19}$$

can be constructed from measurements on the wild-type and mutant systems. Except for a constant factor η_j, which is easily obtained from f_j in the limit $x \rightarrow \infty$, this function is the same as the quantity of interest X_j in the unperturbed, wild-type system. In the limiting case where the mutation abolishes binding to site j, eq 19 yields eq 16. In general, eq 19 only requires η_j to be finite and therefore generalizes the Wegscheider approach that strictly demands the perturbation to mimic specifically a ligated state of the site.

An approach analogous to that embodied by eqs 15−18 has been used by Qian[20] in the analysis of the effects of single-residue substitutions on the stability of α-helices in homopolypeptides[21] and by Wrabl and Shortle[22] in the analysis of the effects of site-directed mutations on the unfolded state of a protein. Particularly important for these approaches is to verify the uniqueness of the solution obtained. In the case of ligand binding, this can be done by examining a number of possible perturbations at site j to guarantee a robust reconstruction of the site-specific binding probability X_j. Expressions equivalent to eq 19 can be derived when the perturbation at site j carries over to other sites,[9] but the parameters defining these expressions are difficult to resolve experimentally. Therefore, a successful use of this strategy should be expected only in the case of first-order perturbations.

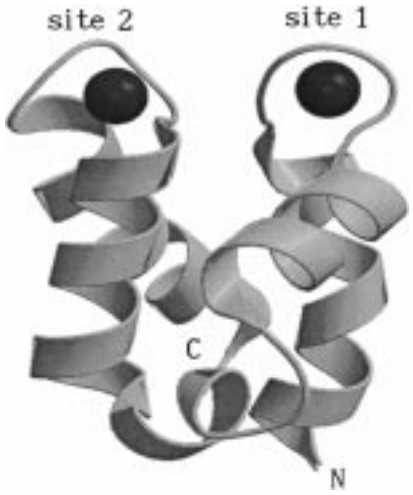


Figure 3. Ribbon representation of calbindin. The two bound Ca^{2+} are depicted by circles.

Table 2. Stepwise Binding Constants (M^{-1}) for Ca^{2+} Binding to Calbindin and Its Mutants

	k_1	k_2	k_2/k_1
wild-type	1.0×10^8	7.9×10^8	7.9
P20G	3.3×10^7	1.6×10^7	0.48
P20G, ΔN21	8.5×10^7	1.0×10^6	0.012
ΔP20	4.0×10^7	1.0×10^6	0.25
Y13F	1.9×10^8	4.9×10^8	2.6
E17Q	1.3×10^7	2.5×10^8	19
D19N	2.0×10^7	2.0×10^8	10
E26Q	3.2×10^7	5.0×10^8	16
E60Q	5.0×10^7	6.4×10^8	13
E17Q, D19N	1.3×10^7	2.5×10^7	1.9
E17Q, E26N	3.2×10^6	1.3×10^8	41
D19N, E26Q	4.0×10^6	8.0×10^7	20
E17Q, D19N, E26Q	3.2×10^6	1.0×10^7	3.1

An application of the approach embodied by eq 19 has been reported for the analysis of cooperative Ca^{2+} binding to calbindin,[9] one of the smallest members of the calmodulin superfamily.[23] Calbindin is composed of two helix-loop-helix motifs responsible for Ca^{2+} binding (Figure 3). The C-terminal site (site 2) has the amino acid sequence and fold of an archetypal EF-hand, while the N-terminal site (site 1) differs from the usual EF-hand in that it contains two additional residues.[24] The protein binds Ca^{2+} with positive cooperativity and a difference between k_1 and k_2 of a factor of 8[25−27] (Table 2). Due to their structural differences, the extent of coupling between the sites as revealed by k_1 and k_2 may be underestimated. Assessment of the exact extent of coupling between the sites is important for the mechanism of

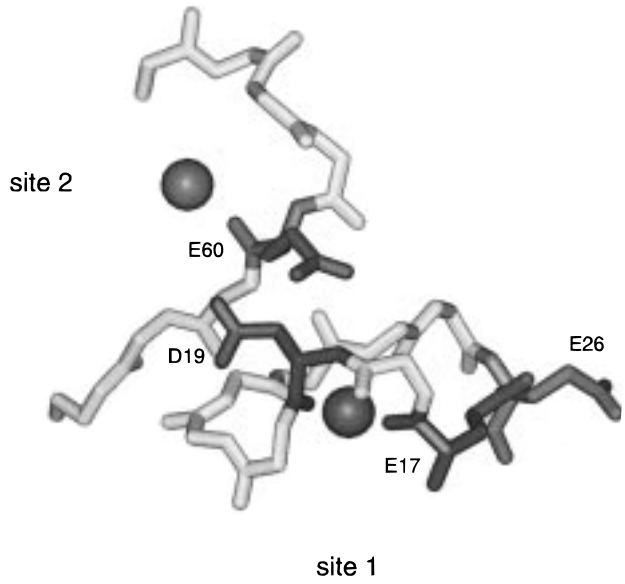


Figure 4. Molecular environment of Ca^{2+} binding sites 1 and 2 of calbindin, showing the side chains of residues whose mutation produces a first-order perturbation of site 1.

transduction of signals from one site to the other and for correctly reproducing the energetics of Ca^{2+} binding to calbindin using computational approaches. The arguments discussed in section II.A assume particular relevance in the case of this protein.

Experimental evidence from NMR data suggests that Cd^{2+} binds with higher affinity to site 2 and that the binding pathways involving site 1 or site 2 as possible singly ligated intermediates elicit distinct structural transitions in the molecule.[28−30] Spectroscopic and kinetic studies on the Ca^{2+} binding properties of calbindin suggest that site heterogeneity may be a factor of 4−6.[25−27,31,32] No direct evidence has been provided for Ca^{2+} binding with higher affinity to site 1 or site 2 and direct determination of Ca^{2+} binding to either site in the wild-type protein has been lacking. Electrostatic calculations suggest that site 2 may have only a slightly higher affinity.[33] On the other hand, valence maps indicate that site 1, rather than site 2, may bind Ca^{2+} with higher affinity.[34] A number of mutants of residues in and around site 1, and partially site 2, have been made to assess the role of electrostatic contributions to the binding of Ca^{2+}, and the global binding parameters have been resolved for all of them[25−27,32] (Table 2). Some of the residues mutated in the wild-type are shown in Figure 4. Particularly interesting is the observation that mutations around site 1 remain localized at this site and do not propagate to site 2.[25,32] This set of mutants was used to resolve the site-specific parameters for calbindin according to eq 19.

Mutations around site 1 can be assumed to produce a first-order perturbation. The assumption is supported by NMR data showing that the mutants E17Q, E26Q, and E60Q have practically the same structure as wild-type protein.[28,29] The mutant P20G and the deletion mutants P20G, ΔN21, and ΔP20 show drastic perturbation of the global cooperativity pattern (Table 2). Nonetheless, there is evidence that

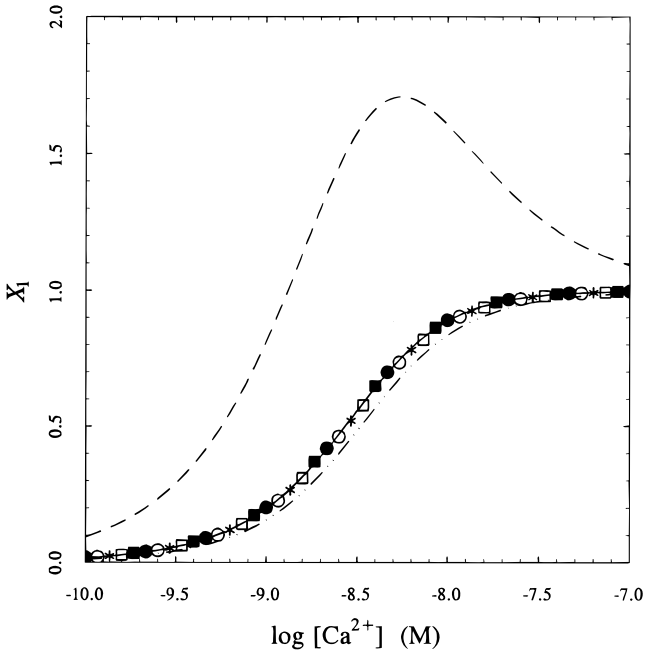


Figure 5. Site-specific binding curve of site 1 of calbindin derived from the approach based on the first-order perturbation hypothesis and eq 22 in the text, using the parameters listed in Table 2 and the partition functions $\Psi = 1 + k_1x + k_1k_2x^2$ and $\Psi = 1 + k_1'x + k_1'k_2'x^2$. The five mutants E17Q (○), D19N (●), E26Q (*), E60Q (□), and E17Q/D19N (■) predict a consensus binding curve (continuous line) that yields site-specific parameter values: $K_1 = 1.7 \times 10^8\ M^{-1}$, $K_2 = 2.7 \times 10^7\ M^{-1}$, and $c_{12} = 18$. The mutant ΔP20 yields the discontinuous dotted line significantly different from the consensus curve. The mutant Y13F predicts a physically implausible curve (discontinuous line) for which $X_1 > 1$.

the structural perturbation is confined to site 1.[25] The perturbation of the global binding constants in these mutants may indeed reflect a perturbation of the site-specific binding constant K_1 only. Construction of the predicted binding curve for site 1 according to eq 19 is shown in Figure 5. The mutant Y13F predicts a binding curve that is physically implausible ($X_1 > 1$). The Y13F replacement violates the assumption of first-order perturbation and eq 19 based on it. The mutation induces changes that must propagate to site 2 and may also involve the communication between the sites. Other mutations, like ΔP20, produce physically plausible results, but five mutants in particular produce a consensus binding curve for site 1. These mutations are all isosteric substitutions of negatively charged residues around site 1 (Figure 4). The predicted site-specific binding curves are practically identical for all of these mutants, although the perturbation of the overall binding constants is quite different in each case (Table 2). Site 1 binds Ca^{2+} with an affinity of about $1.7 \times 10^8\ M^{-1}$, which is nearly 6-fold the affinity of site 2. As a result of cooperative coupling between the sites, the affinity of either site increases by a factor of 18 when the other site is ligated. These parameters map on the point in region III in Figure 1, corresponding to positive cooperativity in the global description. The site heterogeneity is not large enough to overcome the positive interaction between the sites.

These results are consistent with the crystal structure of calbindin.[24] The structure of site 1 (Figure 4) suggests that it is unlikely that mutation of E26 or E17 would be carried over to site 2. These residues are about 10 Å closer to the Ca^{2+} in site 1 than that in site 2. Residues D19 and E60 are close to both sites and, in principle, mutations of these residues should perturb sites 1 and 2. In practice, however, mutation of these residues yields effects similar to mutation of E26 and E17. A molecular dynamics simulation of calbindin shows that E60 may be part of the coordination sphere of Ca^{2+} in site 1.[33] This residue, and possibly D19, may be closer to site 1 in solution, contrary to the conclusions drawn from the crystal structure.[24] It is likely that these residues are more strongly coupled to site 1 than site 2, so that perturbation of site 2 can be neglected for all practical purposes. Finally, comparison of the structure of apo-calbindin with the full Ca^{2+} form shows minor changes at the level of site 1 and more significant rearrangements of the side chains around site 2.[23] Given the similar coordination geometry at the two sites, it is expected that the enthalpy of binding will be similar for Ca^{2+} binding to site 1 or site 2. However, the preformed structure of site 1 should reduce the entropy loss of binding to this site compared to site 2, thereby making site 1 the high-affinity site.

D. Practical Limitations

Given the importance of site-specific parameters in deciphering cooperativity, it is desirable to have a general strategy of approach that works for any system of interest. Measuring site-specific binding curves is one way to obtain information on these parameters. In the case of cytochromes, each redox center has distinct spectral properties and enables direct site-specific measurements.[35,36] In the case of λI repressor binding to its operator, footprint titrations yield binding isotherms for the three individual sites of the operator.[37] Although site-specific probes can be exploited in many systems to obtain information on the behavior of individual sites, it should be noted that knowledge of site-specific binding curves may be insufficient to resolve all independent parameters in the system. Hence, the possibility of experimentally measuring binding events at individual sites by no means guarantees that the site-specific energetics of the system will be fully dissected. For a system of N sites, the N independent site-specific binding constants K_j's can be resolved from the N site-specific binding curves.[9] However, the partition function of the system also contains $N(N-1)/2$ second-order coupling constants, $N(N-1)/(N-2)/6$ third-order coupling constants, and in general $\binom{N}{m}$ mth-order coupling constants that need to be resolved from experimental data. As soon as $\binom{N}{m} > N$ for a particular value of m, measurements of the N site-specific binding curves become insufficient to resolve all site-specific parameters. This limitation arises already for $N = 4$. Even if one could measure all the site-specific binding curves in a macromolecule containing four binding sites, unique resolution of the six second-order coupling constants would be impossible. The only way to overcome this problem is to determine directly the ligated intermediates in the partition function, but this cannot be done in the vast majority of cases. The cryogenic quenching technique developed by Perrella[38] is unique in its ability to resolve all ligated intermediates of hemoglobin[39] ($N = 4$), but this experimental strategy exploits peculiar properties of this protein and has no applicability to other cooperative systems.

The foregoing considerations may lead one to conclude that site-specific thermodynamics is a theory of limited applicability to small systems ($N < 4$) or to particular cases such as hemoglobin where all intermediates of the partition function can be accessed directly. This is indeed the case when the theory is applied to ligand binding cooperativity. The limitations vanish when dealing with another class of processes where the theory finds its most ideal and general applicability. These processes include site-directed mutagenesis of residues aimed at understanding the molecular signatures of stability and ligand recognition. The various intermediates of the system, whose characterization is so problematic in ligand binding studies, are generated directly from the perturbations introduced in the system in the form of site-directed mutations. Analytical and conceptual tools developed for ligand binding processes can be exploited in the analysis of mutational effects using a basic analogy to be described in the next section. This brings site-specific thermodynamics into the main stream of current studies of structure−function relations, protein stability, and ligand recognition.

III. Structural Mapping of Energetics

A. A Basic Analogy

Site-directed mutagenesis[4] has made it possible to perturb the structure of a protein at the level of individual residues and study the origin of protein stability and ligand recognition with unprecedented detail. Residues in a protein can be replaced by any of the 20 natural amino acids. Extension of this strategy to include unnatural amino acids has further expanded the ability to manipulate protein structure and function.[40,41] Site-directed mutagenesis is a modern incarnation of the Wegscheider principle and uses the structural perturbation created by the site-directed substitution as a source of information on the properties of the residue being substituted. A significant advantage of this technique is that it can directly generate all intermediates of interest in a site-specific analysis, overcoming the intrinsic limitations imposed by the number of binding sites seen in ligand binding cooperativity (see section II.D).

Although mutational effects are intrinsically different from ligand binding processes, their thermodynamic treatment is extraordinarily similar to that of cooperative ligand binding once a basic analogy is considered. The free → bound transition at a given binding site is analogous in energetic terms to the wild-type → mutant transition of a given residue. This enables use of the same formalism developed for the analysis of ligand binding cooperativity in the

analysis of mutational effects. Each residue subject to mutational perturbation represents a site in the system. The energetic balance of the wild-type → mutant transition at site *j*, when all other sites are wild-type, specifies the site-specific free energy ΔG_j analogous to the site-specific binding free energy $\Delta G_j = -RT \ln K_j$ (R is the gas constant and T the absolute temperature) for binding to site *j* when all other sites are free. Cooperativity in mutational effects can be expected when substitutions are made at multiple sites and is treated in a manner analogous to that of ligand binding processes. The coupling free energy ΔG_{ij} between mutations at site *i* and *j* is defined from the thermodynamic cycle analogous to eq 2 that involves the intermediates with both sites mutated or wild-type and the two singly mutated species (see section III.C). This quantity is analogous to the coupling free energy $\Delta G_{ij} = -RT \ln c_{ij}$ for the binding of two ligand molecules to sites *i* and *j*.[5,9] The number of intermediates to be characterized is set by the number of residues subject to site-directed mutagenesis. Unlike ligand binding, this number is not imposed by intrinsic properties of the system but is determined entirely by the experimentalist. When N sites are targeted with a single substitution, 2^N possible intermediates are to be considered. Of these, only $2^N - 1$ are independent if one is chosen as reference. This gives rise to N independent site-specific free energies ΔG_j's and a total of $2^N - N - 1$ coupling free energies from second up to Nth order. Resolution of all these independent parameters provides information relevant to the behavior of each residue in the process under investigation and the nature of cooperative interactions.

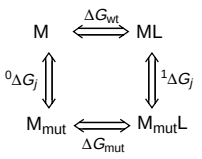
B. Ala Scans

Targets for site-directed mutagenesis are often identified from available structural information. In the analysis of protein stability, particular attention is devoted to residues buried in the interior of the protein and defining hydrophobic cores.[42−44] Other targets are found in residues involved in ionic interactions,[45] especially if screened from the solvent.[46,47] In the analysis of ligand recognition, targets are identified from residues involved in polar and hydrophobic interactions in the bound complex.[48,49] In the absence of structural information on the bound protein, solvent accessibility can successfully guide a mutagenesis screen.[50−52]

There are a number of questions to be addressed when identifying epitopes for protein stability or ligand recognition. First, one would like to know what are the residues important for the energetics of binding or stability. Identification of these residues then raises the question of whether they act independently or in a cooperative manner. Finally, if cooperativity is involved, one would like to know what are the factors responsible for it. Answers to all these questions can be found by application of the principles of site-specific thermodynamics to mutational effects.[5,9]

The first question is addressed by replacing residues that are thought to be involved in stability or recognition. Definition of a structural epitope specifies the degrees of freedom of the system and the number of residues that are energetically relevant to the phenomenon under study. There are 20 possible choices for any given residue in a protein, and therefore, if one specific residue is to be replaced, there are 19 possibilities. In practice, the residue of choice for the replacement is Ala. The rationale behind Ala-scanning mutagenesis is that all interactions of a side chain, except for the C_β atom, are eliminated.[53,54] The contribution of the deleted groups relative to the methyl moiety of Ala is assessed from the difference between the properties of the wild-type relative to the Ala mutant. For this strategy to be effective, it is necessary that the Ala substitution eliminates interactions without introducing new properties. In principle, this should be the case for almost all amino acids except Gly, for which the Ala substitution can introduce new nonpolar interactions, and Cys, for which the Ala substitutions can disrupt an important disulfide bond, generating global destabilizing effects on the protein. In addition, for Gly and Pro, the Ala substitution can introduce perturbations of the protein backbone that becomes less flexible (Gly → Ala substitution) or less rigid (Pro → Ala substitution). Ala scanning mutagenesis has found myriad applications in the identification and energetic characterization of structural epitopes recognizing specific ligands[50−52] or the structural determinants of protein stability,[42−44,55−57] enzyme mechanism,[58] and specificity.[59]

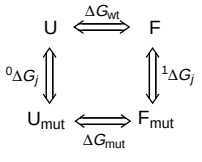
Free energies of binding in the ground or transition state, or free energies of unfolding are used to quantify the effect of the Ala substitution at any given site. In the case of ligand recognition, the effect of Ala replacements is quantified from the properties of the following thermodynamic cycle:[5]



(20)

ΔG_{wt} measures the free energy of binding L to the wild-type macromolecule. The same process in the mutant gives ΔG_{mut}, and the difference $\Delta\Delta G = \Delta G_{mut} - \Delta G_{wt} = \Delta G_c$ is a measure of the effect of the site-directed mutation on the binding process.[5] This difference is the coupling free energy of the cycle and measures the linkage between the binding of L and the mutation. The same cycle applies to binding in the transition state, where the free energy is directly related to the specificity constant $s = k_{cat}/K_m$ and $\Delta\Delta G = RT \ln(s_{wt}/s_{mut}) = \Delta G_c$. When $\Delta G_c > 0$, the mutation reduces specificity, whereas enhanced specificity is reflected by $\Delta G_c < 0$ and no effect is seen for $\Delta G_c = 0$.

In the case of protein stability, a thermodynamic cycle analogous to eq 20 can be constructed as follows:



(21)

The unfolded state of the protein, U, replaces M and the folded state, F, replaces ML.[5] The value of ΔG_c measures the linkage between the mutation and the folding of the protein. When ΔG_c is positive, the mutation reduces stability. Enhanced stability is reflected by a negative value of ΔG_c and no effect is seen for $\Delta G_c = 0$.

Definition of ΔG_c in the cycles in eqs 20 and 21 implies that the effect of the mutation cannot be attributed entirely to the bound or folded state of the protein, as far too often assumed. In fact, ΔG_c measures the difference in free energy between the folded mutant and wild-type relative to the same difference in the unfolded state. In the case of ligand recognition, ΔG_c measures the difference ${}^1\Delta G_j - {}^0\Delta G_j$ and reflects the perturbation introduced by the mutation on the ML complex relative to the free macromolecule M. To assign ΔG_c entirely as a perturbation of the folded state, one must assume that the free energy of the unfolded state is not affected by the mutation. However, this is in contrast with a large body of experimental data obtained in a variety of systems.[5] Likewise, identification of a structural epitope strictly demands that the mutation perturbs specifically the bound state of the macromolecule. However, if a mutation has $\Delta G_c > 0$ and destabilizes the binding of L, the effect is not necessarily due to destabilization of the complex ML. A mutation that stabilizes the free form of the macromolecule (${}^0\Delta G_j < 0$) and has no effect on the bound form (${}^1\Delta G_j = 0$) also gives $\Delta G_c > 0$ and can be confused with a mutation that directly affects recognition of the ligand. In this case, the residue mutated is mistakenly associated with the epitope recognizing the ligand L, although it plays no role in the binding event. A value of $\Delta G_c > 0$ only means that the effect of the mutation has reduced the stability of the complex more than that of the free form. Assignment of the perturbation to the bound complex requires experimental demonstration that the free form of the macromolecule is not affected by the mutation (${}^0\Delta G_j = 0$). In the absence of this information, interpretation of the results may be problematic and must rely on other criteria like the spatial proximity of residues affecting ligand binding or the involvement of these residues in ligand recognition based on structural information. Only when the Ala substitution does not alter the properties of the unfolded state, or removes contacts important for interaction with the ligand, can maps of the regions involved in stability and ligand recognition be constructed from the effect of the mutation on ΔG_c.

In addition to the potential problems outlined above, single-site Ala replacements neglect a priori the contribution of possible site−site interactions to protein stability and ligand recognition. The analogy with ligand binding reveals the limits of such an approach. Single-site Ala scans only provide information on the equivalent of the site-specific binding constants K_j's, and there is no way one can assess the binding properties of a cooperative system from knowledge of these parameters alone. In the absence of interactions, these constants are indeed sufficient to characterize the properties of the system. In a cooperative system where interactions are predominant, these parameters represent only a small fraction of the total number of independent parameters needed to characterize the energetics. Results from the limited number of studies where the importance of site−site interactions in mutational effects has been addressed experimentally have fostered the somewhat misleading notion that residues tend to participate independently in stability and recognition[60−62] and that interactions only occur among residues close in space.[56,58,62−64] It has now been recognized that interactions may involve residues as far as 30 Å away from each other.[65−73] Hence, there is good reason to believe that interactions are present in nearly every system and provide the most important ingredient to protein stability and ligand recognition.

A compelling argument in favor of the existence of cooperativity in ligand recognition and protein stability is as follows. The key assumption of Ala-scanning mutagenesis is that the Ala replacement has the only effect of eliminating the interactions of the side chain beyond the C_β.[53,54] If this assumption is at all valid and the Ala replacement is an unbiased probe of the energetic contribution of a given residue to binding, then the Ala mutation at any position of the epitope should convert the free energy contribution to zero. If this is not the case, then the Ala replacement has introduced new properties at the site, thereby invalidating the assumption. If a functional epitope for binding or stability were composed exclusively of independent residues, then these residues would contribute to the energetics in an additive manner and their contribution would be unraveled by single-site Ala scans. Furthermore, the sum of the free energy changes due to Ala replacement over all sites in the epitope, with changed sign, would be close to the actual free energy of binding or stability measured experimentally for the wild-type. Inspection of the results in Table 3 for a number of systems shows that this is not the case.

A large discrepancy exists between the calculated and experimentally determined values. In the case of human growth hormone[48] or granulocyte colony stimulating factor[76] binding to their receptors, the binding affinity calculated from the results of the Ala scan is greatly overestimated, and so is the stability of Arc repressor[46] and staphylococcal nuclease.[74] In the cases of BPTI binding to trypsin,[49] tissue factor binding to coagulation factor VIIa,[51] or linolenate binding to intestinal fatty acid binding protein,[75] the binding affinity is grossly underestimated. When the affinity is underestimated, it may be argued that the functional epitope might have been incompletely characterized thereby missing important interactions. This can hardly be the case for the interaction of tissue factor with VIIa, where 112 residues were targeted by mutagenesis, or intestinal fatty acid binding protein, where 23 important residues in the binding cavity were replaced. On the other hand, when the affinity is overestimated, it may be argued that the functional epitope might have included sites of marginal importance. Again, this can hardly be the case in the interaction of human growth hormone

Table 3. Comparison of Free Energy Values (kcal/mol) for Stability and Ligand Recognition Measured Experimentally and Calculated from Single-Site Ala Scans

system	process	Ala replacements	ΔG_{calc}[i]	ΔG_{exp}	$\delta\Delta G_{coop}$	ref
Arc repressor	unfolding	51[a]	58.2	13.8	−44.4	46
Staphylococcal nuclease	unfolding	14[b]	39.1	5.5	−33.6	74
hGH-hGHbp[c]	binding	30	−25.9	−12.3	13.6	48
BPTI-chymotrypsin	binding	15	−6.4	−10.7	−4.3	49
VIIa-TF[d]	binding	112	−9.7	−15.4	−5.7	51
I-FABP[e] (palmitate)	binding	23	−6.8	−10.9	−4.1	75
I-FABP[e] (stearate)	binding	23	−13.1	−11.7	1.4	75
I-FABP[e] (oleate)	binding	23	−8.5	−10.7	−2.2	75
I-FABP[e] (linoleate)	binding	23	−5.4	−10.0	−4.6	75
I-FABP[e] (linolenate)	binding	23	2.4	−9.1	−11.5	75
I-FABP[e] (arachidonate)	binding	23	−3.6	−9.5	−5.9	75
GCSF-GCSF receptor[f]	binding	27	−14.5	−11.3	3.2	76
RANTES-CCR1[g]	binding	16	−5.2	−12.3	−7.1	52
RANTES-CCR3[g]	binding	16	−10.5	−13.0	−2.5	52
insulin receptor	binding	26	−8.9	−11.2	−2.3	77
insulin receptor	binding	38	−23.4	−11.2	12.1	77−83
VIII-mAb413[h]	binding	10	−16.6	−13.6	3.0	84

[a] Only the Ala replacements of residues W14, N29, R31, S32, E36, R40, S44, K47, E48, and R50 forming hydrogen bonds and ion pairs protected from the solvent were included in the calculations. [b] Only the Ala replacements of large hydrophobic residues were included in the calculations. [c] Human growth hormone (hGH) binding to the extracellular domain of its first bound receptor (hGHbp). [d] Tissue factor (TF) binding to coagulation factor VIIa. [e] Intestinal fatty acid binding protein. [f] Granulocyte colony stimulating factor (GCSF). [g] CC-chemokine regulated upon activation normal T-cell expressed and secreted interacting with its receptors. [h] Monoclonal antibody 413 binding to coagulation factor VIII. [i] ΔG_{calc}, ΔG_{exp}, and $\delta\Delta G_{coop}$ are defined in eq 22.

with its receptor where the functional epitope is a small hot spot, or for Arc repressor, where the calculated value of stability was taken from the sum of only 11 out of 52 mutated residues, or else for staphylococcal nuclease where only the effect of Ala replacements of 14 large hydrophobic side chains was considered. The results of intestinal fatty acid binding protein are particularly instructive insofar as they show that the discrepancy between calculated and experimentally determined values depends on the particular ligand examined. The difference changes from −11.5 kcal/mol for linolenate to 1.4 kcal/mol for stearate. Given the comparable size of the fatty acids listed in Table 3 and their comparable binding affinity, this large difference cannot be due to intrinsic properties of the ligand. Rather, it suggests the presence of communication among the protein residues that is sensitive to the particular ligand bound. In the case of insulin binding to its receptor, the affinity is underestimated when 26 residues are mutated to Ala.[77] However, when the results are combined with other Ala scans under identical conditions[78−83] to cover a total of 38 residues, the affinity is grossly overestimated. A similar situation is encountered in the binding of a monoclonal antibody to coagulation factor VIII.[84] Again, when the Ala scan involves most of the residues responsible for binding, a large discrepancy is seen between calculated and experimentally determined values for the binding of the ligand to the wild-type, underscoring the important role that interactions among residues play in the recognition process.

It may seem paradoxical that an epitope containing all residues replaced by Ala should bind a ligand with a $\Delta G = 0$, regardless of the system studied, if the residues are truly independent. A binding free energy of zero means that the ligand experiences no net energetic change in going from the free to the bound state and that the all-Ala binding epitope is energetically neutral. Similar arguments apply to protein stability. Although this scenario is hypothetical, its validity within reasonable energetic terms is key to the approach based on Ala scans. If the large discrepancy in Table 3 is the result of specific favorable or unfavorable contributions to stability and recognition introduced by the presence of Ala at any given site, the assignment of epitopes with Ala-scanning mutagenesis becomes context dependent and highly questionable. It is possible that Ala replacements may introduce additional properties at the site of mutation and that these properties may bias the energetic balance of the substitution. However, this bias is likely to be small. We propose that the large discrepancy documented in Table 3 is indicative of a more general problem, i.e., the neglect of energetic contributions arising from possible site−site interactions that cannot be quantified by single-site Ala scans.

In the case of ligand binding, the presence of cooperativity in the recognition event may be the result of some general rules through which biological specificity is encoded into the structure of a protein. A similar scenario may apply to protein stability, where recognition involves domains of the same protein. The stability of a protein is thought to result from the balance of two large and opposite forces, one favorable due to the hydrophobic and electrostatic effects and the other unfavorable due to conformational entropy loss.[2] The balance is usually comparable in magnitude to the free energy involved in only a few polar or charged interactions. In view of this well-established fact, it may be argued that the nonadditivity documented in Table 3 for the perturbation of protein stability may be due to the disruption of favorable interactions, without compromising the unfavorable contributions. Hence, the disruption of a few contacts independent of one another may result in a loss of stability comparable to that of the entire protein, and the balance of similar perturbations over a large number of residues will necessarily

exceed the stability of the protein by a large factor. If this argument is correct, it should be possible to find a significant number of mutations in a protein that only affect the unfavorable contributions and produce large increases in stability. However, stabilizing mutations are rather exceptional, whereas destabilizing mutations are very common. We propose that some of the unfavorable contributions to protein stability result from the negative interactions among residues that contribute to protein stability. In the absence of these important interactions, underlying the complex cooperative nature of the folding process, proteins would be orders of magnitude more stable. This hypothesis explains the results in Table 3 and accounts for the large prevalence of destabilizing effects observed in single-site Ala scans of proteins.

An approximate measure of the extent of interactions among residues is given by the difference, $\delta\Delta G_{coop}$, between the experimentally determined, ΔG_{exp}, and calculated, ΔG_{calc}, values of the free energy of binding or stability, i.e.

$$\delta\Delta G_{coop} = \Delta G_{exp} - \Delta G_{calc} = \Delta G_{wt} + \sum_{j=1}^{N}({}^{j}\Delta G_{mut} - \Delta G_{wt}) = \Delta G_{wt} + \sum_{j=1}^{N}{}^{j}\Delta\Delta G_{mut} \quad (22)$$

The value of ΔG_{exp} is the same as the free energy of the wild-type, ΔG_{wt}. The calculated ΔG_{calc} is the sum of the differences between the free energy of the mutant and wild-type for all N mutants in the epitope, with changed sign. In the absence of interactions among the residues being mutated to Ala, and under the assumption that the Ala substitution is energetically neutral, $\delta\Delta G_{coop}$ should be as close as possible to zero. Hence,

$$\Delta G_{wt} = -\sum_{j=1}^{N}{}^{j}\Delta\Delta G_{mut} \quad (23)$$

is the expected result for an epitope composed of independent residues.

The presence of interactions invalidates the energetic assignments derived from single-site Ala scans because the contribution of a given residue to stability or ligand binding will depend on the state (wild-type or mutated) of other residues. The extent to which interactions affect the assignments based on single-site Ala scans must be evaluated in each case and complicates the identification of epitopes. In cooperative processes such as protein stability or ligand recognition, the contribution of a given residue involves effects of multiple order. A first-order contribution comes from contacts made directly with the ligand or with another residue in the protein. Higher-order contributions may come from the coupling between the residue and other structural components. The residue recognizing the ligand may be involved in a number of interactions with other residues via short-range van der Waals coupling, long-range electrostatic coupling, or large-scale conformational transitions. For interactions of second-order, the construction of double mutations becomes necessary to assess the energetic contribution to stability and ligand recognition, and so forth for higher-order interactions. If an epitope contains N residues, a complete single-site Ala scan requires N mutations and a double-site Ala scan requires $N(N-1)/2$ mutations. The problem of correctly assessing the energetic contribution of residues in a functional epitope using site-directed mutagenesis is combinatorially challenging and demands elucidation of site−site coupling patterns. This calls for a new method of analysis of mutational effects in proteins where the role of interactions is explicitly taken into account.

C. Double-Mutant Cycles

Cooperativity between single-site mutations is typically assessed from the properties of double-mutant cycles. Consider the general case of a system composed of N sites that can exist in two states, 0 (wild-type) and 1 (mutant). ΔG_j is the free energy change associated with the $0 \rightarrow 1$ transition at site j when all other sites are in state 0. This term is the difference in free energy between the configuration with site j perturbed and the wild-type resulting in the loss ($\Delta G_j > 0$) or gain ($\Delta G_j < 0$) of specificity or stability due to perturbation of that site. There are N such terms to be taken into account, one for each site. Consider, then, the double perturbation at sites i and j. The free energy change for such perturbation can be written as the sum $\Delta G_i + \Delta G_j + \Delta G_{ij}$, where ΔG_{ij} is the interaction free energy between sites i and j when the perturbation is applied at both sites. ΔG_{ij} is the same as the coupling free energy in the thermodynamic cycle analogous to eq 2:

$$\begin{array}{ccc} M_{00} & \overset{\Delta G_i}{\Longleftrightarrow} & M_{10} \\ {\scriptstyle \Delta G_j}\Big\updownarrow & & \Big\updownarrow{\scriptstyle \Delta G_j + \Delta G_{ij}} \\ M_{01} & \underset{\Delta G_i + \Delta G_{ij}}{\Longleftrightarrow} & M_{11} \end{array} \quad (24)$$

where the suffix denotes the state, wild-type or mutant, of sites i and j. A negative value of ΔG_{ij} indicates positive coupling between the perturbations at sites i and j in enhancing specificity or stability, or negative coupling in reducing it, and *vice versa* for a positive value. A value of $\Delta G_{ij} = 0$ indicates the absence of coupling between the perturbations.

Some properties of double-mutant cycles have been discussed previously.[62−65,73,85,86] Horovitz and Fersht[86] pointed out that these cycles can also be used to dissect more complex interactions involving multiple sites. This becomes necessary if one wants to understand the origin of the coupling between two sites. The cycle in eq 24 can help establish the presence of coupling between mutations introduced at two sites, but it cannot reveal the origin of the coupling. Once the existence of coupling is established, is this the result of direct interactions between the sites or is it mediated indirectly via other sites? Horovitz and Fersht[86] suggested that comparison of the values of the coupling free energy obtained in the two states, wild-type and mutated, of a third residue

can determine whether the third residue affects the interaction between the two sites. This approach can be extended to an arbitrary number of sites by constructing a hierarchy of perturbed cycles. First, the effect of a third site is examined on the coupling between two sites. Then the effect of a fourth site is studied on the coupling between the third site and the first two sites, and so forth.

A more straightforward and informative method to assess the origin of coupling between mutations at two different sites exploits a key property of the coupling free energy.[5,9] The mechanism of coupling is unraveled by studying how the coupling between any two sites is affected by the configuration of other sites. The coupling free energy between two mutations at sites i and j is defined in the double-mutant cycle in eq 24 by implicitly assuming that all other sites are in state 0. A cycle analogous to that in eq 24 can be constructed for any configuration of the other $N - 2$ sites. There are 2^{N-2} such configurations and $N(N - 1)/2$ distinct pairs of sites, i and j, leading to a total of $N(N - 1)2^{N-3}$ possible thermodynamic cycles and coupling free energies. Not all cycles are independent because the system only contains $2^N - 1$ independent terms, N of which are site-specific free energies of perturbation ΔG_j's and the remainder are coupling free energy terms from second up to Nth order. Construction of double-mutant cycles cannot generate more information than that contained in the independent coupling terms. Hence, of the $N(N - 1)2^{N-3}$ possible cycles, only $2^N - 1 - N$ are necessarily independent. However, once any pair of sites i and j is chosen, the 2^{N-3} coupling free energy values generated by all configurations of the other $N - 2$ sites are all independent. Therefore, there are two alternative and equivalent ways to characterize the interactions of a system. One is based on the second- and higher-order interaction free energies that define the intermediates of the system; the other casts these free energies in terms of the coupling between two sites in any possible configurations of the other sites.

Once coupling free energies are calculated for all possible configurations of the system, it is possible to decipher the code for site−site interactions using the following property of a thermodynamic cycle whose mathematical proof is given elsewhere:[9]

THEOREM: *If the coupling between two sites is direct and involves only second-order interactions, then the coupling free energy is independent of the configuration of other sites. Otherwise, the coupling is indirect and involves interactions higher than second order.*

To understand the significance of this property, it is useful to consider two key examples of direct and indirect coupling. Direct coupling is peculiar of models of nearest-neighbor interactions, like the Koshland−Nemethy−Filmer model of ligand binding cooperativity.[87] In this model, interactions are all pairwise and second order. Coupling of higher order is simply the result of additive contributions from second-order coupling terms. No matter how two sites are linked to each other and to the rest of the system, the coupling between them remains energetically the same regardless of the configuration of other sites. This has the nontrivial consequence that, when the coupling between a pair of sites is not affected by a third site, one cannot conclude that the third site is not coupled to the pair as the Horovitz−Fersht approach would mistakenly imply.[86] In fact, in any nearest-neighbor model where the third site is coupled to each site in the pair, the state of the third site is inconsequential on the coupling free energy of the pair. Though somewhat counterintuitive, this conclusion can be proved mathematically[9] and provides an important reference point for the correct interpretation of coupling free energy profiles.

The case of the ionization reactions of glutamic acid dealt with in section II.B is particularly relevant in this regard. The third-order coupling constant c_{123} is the same as the product $c_{12}c_{13}c_{23}$ (Table 1). Hence, the third-order coupling free energy $\Delta G_{123} = -RT \ln c_{123}$ is the sum of the three second-order coupling free energies $\Delta G_{12} = -RT \ln c_{12}$, $\Delta G_{13} = -RT \ln c_{13}$ and $\Delta G_{23} = -RT \ln c_{23}$. As a result, the coupling between any two ionizable groups in glutamic acid is not influenced by the ionization state of the third group, although all groups are coupled. The amino group influences protonation of the α- and γ-carboxyl groups but has no influence on the negative coupling between these groups, which remains the same whether the amino group is protonated or not.

Indirect coupling manifests itself in a more obvious manner. An example is provided by the Monod−Wyman−Changeux model of concerted allosteric transitions[88] where interactions involve all sites through a linked global conformational change. In this model, sites are always positively coupled and the order of coupling changes according to the state of other sites as the protein switches from one state to another. Combination of the Koshland−Nemethy−Filmer and Monod−Wyman−Changeux models into a more general hybrid model accounts for arbitrarily complex mechanisms of coupling.[5,9]

The mechanism of coupling can be identified from analysis of double-mutant cycles but requires the availability of a high-dimensional manifold of perturbations where the coupling between two sites can be studied as a function of a relatively large number of configurations of other sites. This poses challenging tasks from an experimental standpoint because construction and expression of triple or higher order mutants in a protein may be problematic. The analysis based on the properties of the coupling free energy appears to be ideally suited for the site-specific dissection of ligand recognition when most of the perturbations are introduced in small peptides that bind to the protein. Large libraries of peptides containing all the relevant mutant forms can be constructed with ease and when combined with perturbations in the protein generate the complexity necessary to dissect all interactions in the system. An example of how this new and powerful approach based on the principles of site-specific thermodynamics can be implemented in practice is offered in the next section.

IV. Site-Specific Dissection of Thrombin Specificity

A. Substrate Recognition by Serine Proteases

The principles outlined in the previous sections find an ideal application to the study of enzyme specificity. Understanding the molecular origin of enzyme specificity is important for structure−function and evolutionary studies and also bears on rational drug design. One of the best characterized class of enzymes is that of serine proteases of the chymotrypsin family.[89,90] These enzymes participate in key physiological functions such as digestion, blood coagulation, fibrinolysis, complement, and development. Proteases involved in digestive processes, like trypsin, have wide specificity and are also found in bacteria. In contrast, proteases involved in blood coagulation, fibrinolysis, and complement have narrow specificity and are found almost exclusively in vertebrates.[91−93] Among these more specialized proteases, activity and specificity is controlled allosterically by the binding of Na^+, whereas more primitive proteases and those involved in fibrinolysis are apparently devoid of such important property.[94,95]

Serine proteases of the chymotrypsin family share a common fold composed of two six-stranded β-barrels of similar structure that pack together asymmetrically to host at their interface the residues of the catalytic triad H57, D102, and S195.[96] Although they have a common catalytic mechanism,[97] these enzymes differ widely in specificity. The exact molecular origin of this difference remains in the most part elusive. The preference of trypsin-like enzymes for cleavage at Arg residues is due to the presence of D189 at the bottom of the catalytic pocket. In chymotrypsin, residue 189 is a Ser and the preference is for bulky aromatic side chains. However, the D189S replacement in trypsin does not result in a chymotrypsin-like specificity. This is instead obtained by more substantial replacements involving the surface loops 185−188 and 221−225 with the homologous regions in trypsin,[98] though none of the residues in these loops contacts the bound substrate. These observations suggest a molecular origin of protease specificity that depends on multiple critical sites.

The classical approach to the study of protease specificity takes into account interactions made by the enzyme with the substrate at the level of individual sites.[99] This approach lends itself to application of the principles of site-specific thermodynamics developed for the study of binding cooperativity.[9] Residues of the substrate interacting with the enzyme are labeled with a P and a number from 1 to *N*, starting from the scissile bond and moving to the N-terminus. Residues of the enzyme making contacts with the substrate are called *specificity sites* and are labeled with an S. The amino acid at P1 of the substrate makes contacts with the specificity site S1 of the enzyme, P2 contacts S2, and so forth. Residues on the C-terminal portion of the scissile bond of the substrate are numbered P1′, P2′, and so forth and the corresponding specificity sites on the enzyme are S1′, S2′, and so on. The scissile bond is positioned between P1 and P1′. The existence of multiple recognition sites effectively narrows down specificity by reducing the probability that the required sequence is found in a random sample of potential substrates. The longer the consensus sequence interacting with the enzyme, the smaller the probability that it will occur in another potential substrate.

The recognition model based on binding to multiple specificity sites brings about a number of important questions, including the assessment of the free energy cost of a replacement made at a P or S site and whether the P or S sites contribute to recognition additively or cooperatively. These questions are central to the analysis of mutational effects discussed in section III and are addressed below in the specific case of thrombin−substrate interactions.

B. Thrombin Structure and Function

The serine protease thrombin is capable of two important and opposite roles that are at the basis of the efficiency of blood coagulation. The procoagulant role entails the conversion of fibrinogen into the insoluble fibrin clot, the promotion of platelet aggregation, the stabilization of the ensuing clot by activation of factor XIII and inhibition of fibrinolysis, and the feedback enhancement of its own generation from prothrombin by activation of factors V, VIII, and XI. The anticoagulant role involves the thrombomodulin-assisted conversion of protein C into an active component that cleaves and inactivates factors VIIIa and Va together with protein S, thereby limiting the conversion of prothrombin into thrombin catalyzed by the prothrombinase complex.[100,101] In addition to its primary roles in coagulation, thrombin elicits a variety of important effects on a number of cell lines upon binding to its receptors.[102,103]

Na^+ is required for the optimal conversion of fibrinogen into fibrin monomers, which is catalyzed by the procoagulant fast (Na^+-bound) form with high specificity.[104] The slow (Na^+-free) form of thrombin performs the same task with lower specificity. This form, on the other hand, has higher specificity than the fast form toward protein C[105,106] and plays predominantly an anticoagulant role. As a result of the different affinity of the two allosteric forms, Na^+ is actively exchanged in the transition state upon binding of fibrinogen or protein C. Fibrinogen binds to the fast forms with higher affinity and promotes the slow → fast conversion and Na^+ binding. On the other hand, binding of protein C promotes the fast → slow conversion and Na^+ release. Hence, Na^+ binding and dissociation are important molecular components of substrate recognition by thrombin.

Thrombin is composed of two polypeptide chains of 36 (A chain) and 259 (B chain) residues that are covalently linked through a disulfide bond.[107] The B chain carries the functional epitopes of the enzyme and has an overall architecture similar to that of pancreatic serine proteases (Figure 6). The extraordinary specificity of thrombin toward fibrinogen arises not only from contacts made in the interior of the active site (see below) but also from interactions with exosite I located about 20 Å away from the

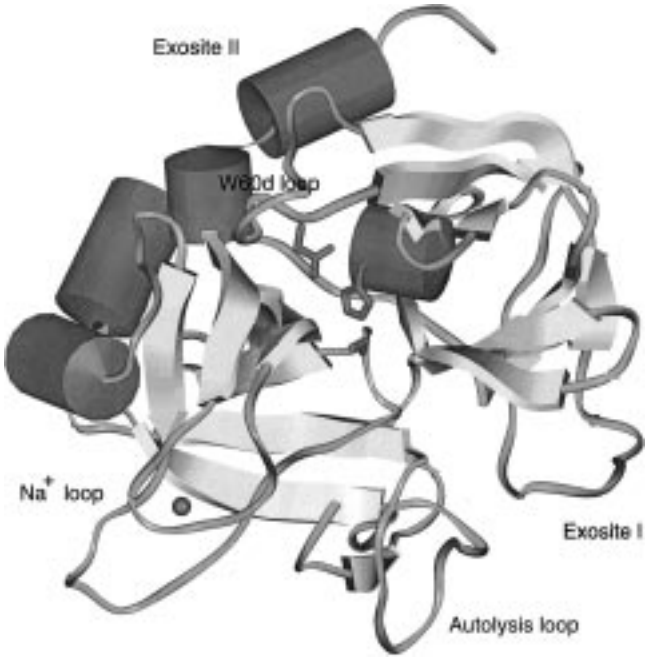


Figure 6. Ribbon representation of thrombin showing the residues of the catalytic triad. Important regions of the enzyme are noted.

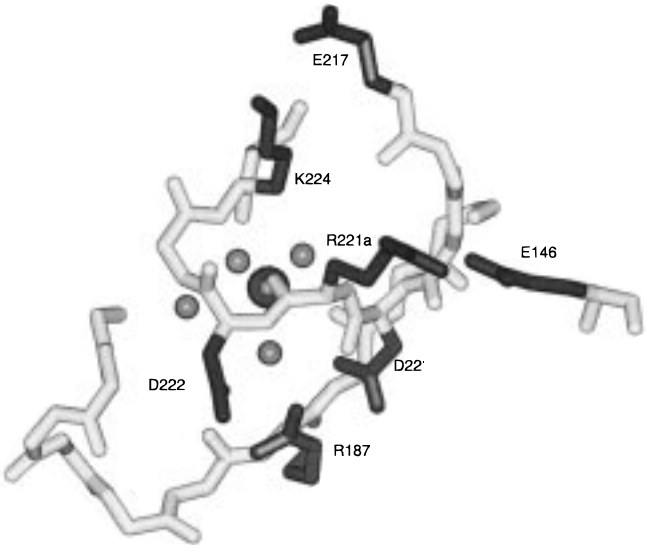


Figure 7. Molecular environment of the Na^+ binding site of thrombin. The bound Na^+ (black circle) is coordinated octahedrally by the carbonyl O atoms of K224 and R221a and four water molecules (gray circles). The site seems to be stabilized by three ion pairs: R221a-E146, D221,D222-R187, and K224-E217.

active site.[108,109] Exosite I serves as an extended primed recognition site. Binding of hirudin derivatives or thrombomodulin to this site also enhances allosterically Na^+ binding and switches the enzyme to the fast form, thereby changing activity and specificity.[110−112] Another factor that influences thrombin specificity is the W60d insertion loop that is unique to thrombin and shapes the apolar specificity site S2. This loop narrows significantly the access to the active site by protruding into the solvent. Replacement of W60d with the less bulky Ala or Ser profoundly affects the interaction of thrombin with the natural inhibitor antithrombin III[113] or fibrinogen.[111,114] A similar function has been hypothesized for the autolysis loop shaping the lower rim of the access to the active site. Deletion of the entire loop results in a selective loss of fibrinogen binding.[115]

The Na^+ binding site (Figure 7) displays octahedral coordination involving the carbonyl O atoms of R221a and K224 and four buried water molecules tetrahedrally coordinated by protein atoms and other water molecules[116,117] that altogether define a complex hydrogen-bonding network within the catalytic pocket.[118] Some of the hydrogen bonds in the network are conserved with trypsin.[119] Others are specific to thrombin and are associated with Na^+ and its coordination shell. The bound Na^+ is located 15−20 Å away from the catalytic triad and lies within 5 Å from D189 in the specificity site S1 with a water molecule mediating a hydrogen-bonding interaction with $O^{\delta 2}$ of D189. The Na^+ site also appears to be stabilized by three ion pairs. R221a is ion-paired to E146 of the autolysis loop, K224 is ion-paired to E217, while D221 and D222 form a bidentate ion pair with R187. Altering the bidentate ion pair with the double substitution D221A/D222K results in reduced activity toward fibrinogen but enhanced activity toward protein C.[116] Perturbation of the ion pair in the R187Q thrombin Greenville produces a reduced clotting activity, consistent with reduced Na^+ binding.[120] Similar effects of reduced clotting activity due to reduced Na^+ binding are seen upon disruption of the R221aA-E146[106,121] or the K224-E217[106,122] ion pairs.

C. Library of Site-Specific Probes

The molecular strategy used by thrombin to achieve specificity toward fibrinogen and protein C is deeply rooted in the mechanism through which Na^+ binding affects the environment of the active site of the enzyme. The main question is how the Na^+-induced slow → fast conversion enhances specificity toward fibrinogen and small chromogenic substrates. A related question is which allosteric form should be targeted with active-site inhibitors to guarantee optimal specificity. In both cases, the answer resides primarily in the properties of the specificity sites of the enzyme and warrants a quantitative assessment of their energetic contribution in the transition state.

Substrate libraries generated from combinatorial chemistry or phage display to identify consensus sequences for binding[123,124] can be used as powerful probes of the molecular environment of the specificity sites of the enzyme to elucidate how they contribute to recognition in the transition state. If perturbations are made in the sequence of a substrate to generate a library containing all species required for a site-specific analysis, much information can be derived on the energetic contributions of the specificity sites that is difficult to obtain from mutagenesis of the enzyme. To understand the molecular origin of the higher specificity of the fast form toward fibrinogen, the chromogenic tripeptide substrate FPR (Table 4) was synthesized[59] to mimic the interaction of the natural substrate with the active site of the enzyme.[108,109] Like fibrinogen, FPR is cleaved by the fast form with a specificity 30-fold higher than that of the slow form[95] (Table 5). The crystal structure

Table 4. Substrate Library

abbrev	substrate	site(s) perturbed
FPR	H-D-Phe-Pro-Arg-*p*-nitroanilide	none
FPK	H-D-Phe-Pro-Lys-*p*-nitroanilide	P1
FGR	H-D-Phe-Gly-Arg-*p*-nitroanilide	P2
VPR	H-D-Val-Pro-Arg-*p*-nitroanilide	P3
FGK	H-D-Phe-Gly-Lys-*p*-nitroanilide	P1 and P2
VPK	H-D-Val-Pro-Lys-*p*-nitroanilide	P1 and P3
VGR	H-D-Val-Gly-Arg-*p*-nitroanilide	P2 and P3
VGK	H-D-Val-Gly-Lys-*p*-nitroanilide	P1, P2, and P3

of thrombin inhibited with H-D-Phe-Pro-Arg-CH_2Cl[107] provides information on the interactions of the P1−P3 groups of FPR with the enzyme. Arg at P1 makes an ion pair with D189 at S1 at the bottom of the catalytic pocket, Pro at P2 interacts with the apolar moiety of S2 defined by P60b, P60c, and W60d, whereas Phe at P3 forms a favorable edge-to-face interaction with the aromatic ring of W215 at S3 (Figure 8). The D enantiomer at P3 mimics the interaction of F8 at P9 of fibrinogen with W215 of thrombin.[125] The chromogenic group *p*-nitroanilide attached to the C-terminus enables quantitative spectroscopic measurements of the released *p*-nitroaniline upon cleavage by thrombin at the P1−*p*-nitroanilide scissile bond.

Starting from FPR, seven substitutions were made to generate the library in Table 4.[59] The rationale behind these substitutions was to introduce enough perturbation at P1, P2, and P3 while retaining sufficient specificity for accurate experimental measurements. The perturbation would then act as the source of information on the environment of the specificity sites of the enzyme S1, S2, and S3. H-D-Phe was replaced with H-D-Val in VPR, VPK, VGR, and VGK, to replace the aromatic moiety with a hydrophobe. Pro was replaced with Gly in FGR, FGK, VGR, and VGK, to avoid steric hindrance with S2 and relieve the rigidity of the P2−P3 bond. Arg was replaced with Lys in FPK, FGK, VPK, and VGK, to preserve the positive charge at P1 needed to contact D189 at S1. The substitutions were combined to generate all possible intermediates from the parent substrate FPR: the three singly substituted substrates FPK, FGR, and VPR, the three doubly substituted substrates FGK, VPK, and VGR, and the triply substituted substrate VGK.

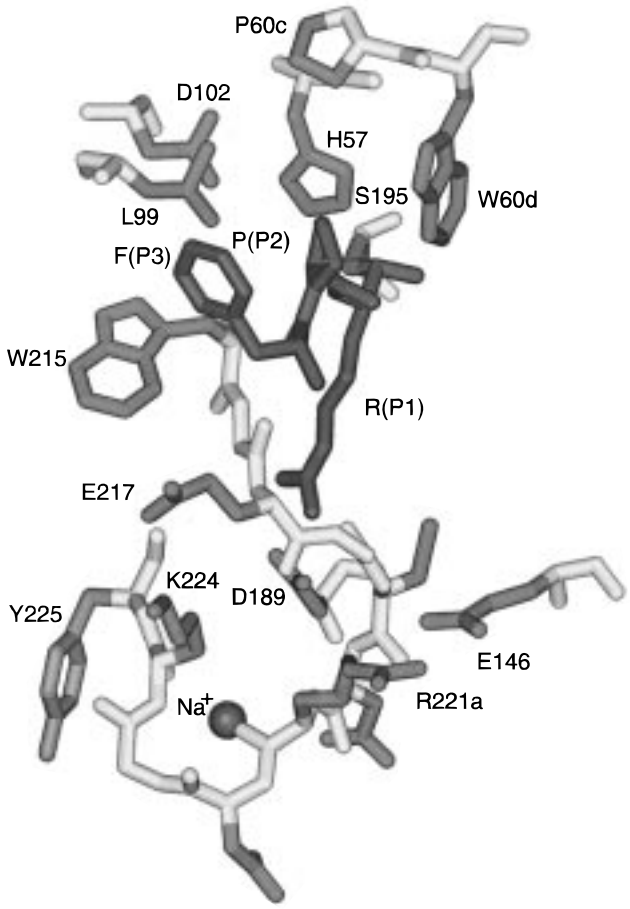


Figure 8. Contacts between the irreversible inhibitor H-D-Phe-Pro-Arg-CH_2Cl and the active site of thrombin. Shown are thrombin residues D189, P60c, W60d, L99, and W215 that interact with the inhibitor. The guanidyl group of the Arg at P1 makes an ion pair with the carboxyl group of D189 at S1 at the bottom of the active site. Pro at P2 packs in the S2 apolar cavity provided by the W60d loop. H-D-Phe at P3 makes favorable hydrophobic contacts in the cleft with L99 and especially a perpendicular aryl−aryl edge-on interaction with W215 at S3.

To obtain the relevant free energy changes associated with the perturbations, the specificity constant $s = k_{cat}/K_m$ for substrate hydrolysis was measured in all cases (Table 5) to estimate the free energy of stability of the transition state. The value for FPR was used to scale energetically all others to obtain

Table 5. Specificity Constants k_{cat}/K_m (μM^{-1} s^{-1}) for the Hydrolysis of Synthetic Substrates by Thrombin, Trypsin, and Plasmin

	FPR	FPK	FGR	VPR	FGK	VPK	VGR	VGK
				Thrombin Fast Form				
wild type	90	7.9	2.0	100	0.021	2.1	0.34	0.0047
R221aA	80	4.6	0.75	36	0.011	0.96	0.14	0.0024
K224A	44	7.7	0.93	24	0.027	1.4	0.17	0.0044
R221aA/K224A	26	3.2	0.33	13	0.011	0.70	0.049	0.0017
				Thrombin Slow Form				
wild type	3.0	0.35	0.86	6.7	0.0026	0.11	0.17	0.00079
R221aA	1.6	0.040	0.042	1.0	0.00038	0.0097	0.0086	0.00013
K224A	0.47	0.034	0.012	0.28	0.00039	0.0063	0.0020	0.00013
R221aA/K224A	0.34	0.010	0.0025	0.077	0.00021	0.0018	0.00063	0.000063
trypsin	8.9	0.95	2.2	6.9	0.22	0.75	0.67	0.069
plasmin	0.031	0.047	0.0018	0.028	0.0048	0.058	0.0016	0.0037

[a] Experimental conditions: 5 mM Tris, I = 200 mM, 0.1% PEG, pH 8.0, at 25 °C. The slow form was studied in the presence of 200 mM choline chloride. The properties of the fast form refer to the limit $[Na^+] \rightarrow \infty$, at constant I = 200 mM. Errors are typically ±2%.

Table 6. Free Energy Values (kcal/mol) Due to Perturbation of the P1−P3 Sites of FPR[a]

	ΔG_1	ΔG_2	ΔG_3	ΔG_{12}	ΔG_{13}	ΔG_{23}	ΔG_{123}
Thrombin Fast Form							
wt	1.4	2.3	−0.1	1.3	0.8	1.1	2.2
R221aA	1.7	2.8	0.5	0.8	0.5	0.5	1.2
K224A	1.0	2.3	0.4	1.1	0.7	0.6	1.8
R221aA/K224A	1.2	2.6	0.4	0.8	0.5	0.7	1.5
Thrombin Slow Form							
wt	1.3	0.7	−0.5	2.2	1.2	1.4	3.3
R221aA	2.2	2.2	0.3	0.6	0.6	0.7	1.0
K224A	1.6	2.2	0.3	0.5	0.7	0.8	0.8
R221aA/K224A	2.1	2.9	0.9	−0.6	0.1	−0.1	−0.8
trypsin	1.3	0.8	0.2	0.0	−0.0	0.6	0.6
plasmin	−0.3	1.7	0.1	−0.3	−0.2	0.0	−0.2

[a] Values were obtained from the specificity constants in Table 5 using eqs 25−31 in the text. Errors are ±0.1 kcal/mol or less.

the relevant free energy changes in the transition state (Table 6) as follows:[59]

$$\Delta G_1 = -RT \ln(s_{FPK}/s_{FPR}) \quad (25)$$

$$\Delta G_2 = -RT \ln(s_{FGR}/s_{FPR}) \quad (26)$$

$$\Delta G_3 = -RT \ln(s_{VPR}/s_{FPR}) \quad (27)$$

$$\Delta G_{12} = -RT \ln(s_{FGK}s_{FPR}/s_{FPK}s_{FGR}) \quad (28)$$

$$\Delta G_{13} = -RT \ln(s_{VPK}s_{FPR}/s_{FPK}s_{VPR}) \quad (29)$$

$$\Delta G_{23} = -RT \ln(s_{VGR}s_{FPR}/s_{FGR}s_{VPR}) \quad (30)$$

$$\Delta G_{123} = -RT \ln(s_{VGK}s_{FPR}^2/s_{FPK}s_{FGR}s_{VPR}) \quad (31)$$

ΔG_1, ΔG_2, and ΔG_3 are the changes in specificity due to the single-site substitutions at P1, P2, and P3. ΔG_{12}, ΔG_{13}, and ΔG_{23} are the second-order coupling free energies for substitutions made at the three possible pairs of sites, and ΔG_{123} is the third-order coupling free energy for the triple substitution. These terms reflect interactions between substitutions made at different sites that may reduce ($\Delta G > 0$) or enhance ($\Delta G < 0$) specificity beyond simple additivity. The terms in eqs 25−31 define the free energy level of any substrate in the library relative to FPR. For example, the relative free energy level of FGK is $\Delta G_1 + \Delta G_2 + \Delta G_{12}$ and that of VGK is $\Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_{123}$. Similar measurements were carried out with the three thrombin mutants R221aA, K224A, and R221a/K224A to assess the role of the ion pairs that seem to stabilize the Na^+ binding environment (Figure 7). This resulted in the complete dissection of a five-dimensional manifold of species in both the slow and fast forms of the enzyme from which detailed information can be derived on how perturbations of the substrate are coupled to each other and to perturbations in the enzyme. The five sites perturbed are P1, P2, and P3 in the substrate and R221a and K224 in the enzyme. The site-specific parameters relative to the 32 possible intermediates in the manifold are listed in Table 6 for each thrombin form.

D. Cooperativity in Substrate Recognition

Inspection of Table 6 reveals the presence of large and significant cooperativity in the effects induced by perturbations of the P1−P3 sites in the case of wild-type thrombin. The extent of cooperativity changes for each pair of substitutions and is also affected by the allosteric state of the enzyme and mutations made around the Na^+ binding environment. In contrast, no interactions are seen for trypsin and plasmin, two cognate proteases. The different response elicited by the substrate library in different enzymes lends validity to the strategy of probing the environment of the specificity sites.

The free energy change due to replacing Arg with Lys at P1 in all possible combinations of the state of P2 and P3 is summarized in Table 7. The values are all positive in both the slow and fast forms, for wild-type and mutant thrombins, indicating that the Arg → Lys replacement at P1 always causes a loss of specificity. The cost of this replacement is about 1 kcal/mol in both the slow and fast forms when no replacement is made at P2 and P3, which suggests that the same mechanism may cause the loss of specificity in both allosteric forms. The changes in catalytic parameters observed in the fast → slow conversion of thrombin for both synthetic substrates and fibrinogen involve a decrease in k_{cat} and an increase in K_m.[104,126] This would suggest that binding of Na^+ orients the side chain of D189 for optimal coordination of the guanidinium group of Arg at P1, perhaps using water 447 that bridges the bound Na^+ and the $O^{\delta 2}$ atom of D189.[118] In this case, however, the loss of specificity with the Arg → Lys substitution at P1 would be more pronounced in the fast form. The similarity of effects seen for the two forms argues against a direct influence of the allosteric switch on the position of the side chain of D189. This conclusion is consistent with the observation that water 447 is also present in trypsin,[118,119] which does not bind Na^+,[94] where it bridges the $O^{\delta 2}$ atom of D189 to the carbonyl O atom of K224. The origin of the increased specificity of the fast form must therefore reside at other specificity sites.

Due to the strong interactions among the P1−P3 sites, the cost of replacing Arg with Lys at P1 depends on the residue at P2 and P3 (Table 7) and reveals the importance of cooperativity in substrate recognition. With Gly at P2, the cost of the Arg → Lys replacement at P1 increases by 1.3 kcal/mol in the fast form and 2.1 kcal/mol in the slow form, introducing a significant difference of −0.7 kcal/mol between the two forms. This difference measures the coupling between the replacement at P1 and the slow → fast transition. A negative value indicates that the replacement promotes the slow → fast conversion in the transition state or that the replaced residues binds preferentially to the slow form. A positive value signals a stabilization of the slow form or that the replaced residues binds preferentially to the fast form. The presence of a small, but significant coupling when Gly is present at P2 suggests that the environment around D189 in the transition state may be different in the slow and fast form. When P3 is substituted, the energetic penalty for the P1 substitution increases by nearly 1 kcal/mol in both thrombin forms. The extent of interaction of P2 and P3 with P1 is significant. When Gly is present at P2,

Table 7. Free Energy Change (kcal/mol) in Specificity Due to Perturbation of the P1−P3 Sites of FPR[a]

	fast form			R221aA/	slow form			R221aA/	coupling			R221aA/
	wt	R221aA	K224A	K224A	wt	R221aA	K224A	K224A	wt	R221aA	K224A	K224A
Replacement at P1 (Arg→Lys)												
FPX	1.4	1.7	1.0	1.2	1.3	2.2	1.6	2.1	0.2	−0.5	−0.5	**−0.8**
FGX	2.7	2.5	2.1	2.0	3.4	2.8	2.0	1.5	**−0.7**	−0.3	0.1	0.5
VPX	2.3	2.1	1.7	1.7	2.4	2.7	2.2	2.2	−0.1	−0.6	−0.6	−0.5
VGX	2.5	2.4	2.2	2.0	3.2	2.5	1.6	1.4	−0.6	−0.1	0.5	0.6
Replacement at P2 (Pro → Gly)												
FXR	2.3	2.8	2.3	2.6	0.7	2.2	2.2	2.9	**1.5**	0.6	0.1	−0.3
FXK	3.5	3.6	3.3	3.4	2.9	2.8	2.6	2.3	0.6	**0.8**	**0.7**	**1.1**
VXR	3.4	3.3	2.9	3.3	2.2	2.8	2.9	2.8	**1.2**	0.5	0.0	0.5
VXK	3.6	3.5	3.4	3.6	2.9	2.6	2.3	2.0	**0.7**	**1.0**	**1.1**	**1.6**
Replacement at P3 (Phe → Val)												
XPR	−0.1	0.5	0.4	0.4	−0.5	0.3	0.3	0.9	0.4	0.2	0.1	−0.5
XPK	0.8	0.9	1.0	0.9	0.7	0.8	1.0	1.0	0.1	0.1	0.0	−0.1
XGR	1.0	1.0	1.0	1.1	1.0	0.9	1.1	0.8	0.1	0.1	−0.1	0.3
XGK	0.9	0.9	1.1	1.1	0.7	0.6	0.7	0.7	0.2	0.3	0.4	0.4

[a] wt = wild type. Errors are ±0.1 kcal/mol or less. Values were obtained from the data in Table 6. The difference between the values for the fast and slow forms gives the coupling between the substitution and the slow → fast transition. Positive values are indicative of stabilization of the slow form in the transition state, whereas negative values signal stabilization of the fast form. Values of the coupling in excess of $\pm RT$ (0.6 kcal/mol) are in bold type.

the interaction with P1 actually exceeds the cost of the replacement at P1 itself in the slow form.

The free energy change due to replacing Pro with Gly at P2 in all possible combinations of the state of P1 and P3 is summarized in Table 7. As for the substitution at P1, the values are significantly positive. In this case, the effects tend to be more pronounced in the fast form, underscoring an obvious change in the environment of the S2 site in the slow → fast transition. The significant difference is conducive to stabilization of the slow form in the transition state when Pro is replaced by Gly. The apolar site S2 of thrombin is formed by residues in the W60d loop, which has no counterpart in other serine proteases. Residues in the apolar site must be oriented differently in the slow and fast forms, causing a better discrimination of the residue at P2 in the fast form. W60d may play a key role in this respect because replacement of the bulky side chain with Ser in W60dS abolishes the differences between the slow and fast forms in recognizing substrates with Pro or Gly at P2.[111] The indole ring of W60d likely produces steric hindrance in the slow form, but not in the fast form. The perturbation at P2 depends strongly on the residue present at P1 and P3. The cost of the Pro → Gly replacement increases by 1.2 kcal/mol in the fast form and 2.2 kcal/mol in the slow form as a result of the substitution at P1. This effect is exactly (taking into account roundoff error) the same as that seen for the perturbation at P1 when P2 is perturbed, as a consequence of the reciprocity of the linkage between the perturbations at P1 and P2.

The free energy change due to replacing Phe with Val at P3 in all possible combinations of the state of P1 and P2 is summarized in Table 7. The unexpected finding is that Val at P3 does not cause a loss of specificity. Rather it increases specificity slightly in the slow form. The hydrophobic group at P3 may interact favorably with the hydrophobic moiety of L99 (Figure 8), which is close to the apolar site S2. Interestingly, residue Y3 of hirudin contacts W215 of thrombin in a manner similar to Phe at P9 of the fibrinogen Aα chain,[127] but replacement of Y3 with more hydrophobic residues significantly enhances the binding affinity,[128,129] consistent with the enhanced specificity of VPR compared to FPR. The energetic effect linked to replacement of the residue at P3 is of the same magnitude in both forms and excludes a direct involvement of the S3 site in the slow ↔ fast equilibrium. The perturbation at P3 depends strongly on the state of P1 and P2. The cost of the Phe → Val replacement increases by 0.9 kcal/mol in the fast form and 1.2 kcal/mol in the slow form as a result of the substitution at P1 and is the reciprocal of the effect seen for the perturbation at P1 when P3 is perturbed.

The data in Tables 6 and 7 reveal the presence of coupling among perturbations at P1, P2, and P3. The coupling is the result of constraints imposed by the enzyme on the bound substrate in the transition state and is therefore revealing of the molecular environment underlying the recognition process. The coupling free energies for the three possible pairs of P sites in the two possible states of the third site are listed in Table 8. The values are constructed from the specificity constants pertaining to the four species in the double-mutant cycle in eq 24, where the mutations are replaced by substitutions at the P sites. For example, the coupling between P1 and P2 is ${}^0\Delta G_{12} = -RT \ln(s_{FGK}s_{FPR}/s_{FPK}s_{FGR})$ in the absence of perturbation at P3 and ${}^1\Delta G_{12} = -RT \ln(s_{VGK}s_{VPR}/s_{VPK}s_{VGR})$ when P3 is perturbed. The value of ${}^0\Delta G_{12}$ is the same as ΔG_{12} in Table 6. The coupling free energies in the case of wild-type thrombin are mostly positive and quite significant, demonstrating that perturbations at the P1, P2, and P3 sites are negatively coupled in enhancing specificity and that the residues at P1−P3 are negatively coupled in the binding to the S1−S3 sites. When a site is perturbed, perturbation at a second site reduces specificity beyond simple additivity. Furthermore, the coupling between any two sites is enhanced by more than 1 kcal/mol when the third site is perturbed underlying an even stronger cooperative effect in reducing specificity that progresses with the extent of pertur-

Table 8. Coupling Free Energies (kcal/mol) for Perturbation of the P1−P3 Sites of FPR[a]

	fast form				slow form			
	wt	R221aA	K224A	R221aA/K224A	wt	R221aA	K224A	R221aA/K224A
$^0\Delta G_{12}$	1.3	0.8	1.1	0.8	2.2	0.6	0.5	−0.6
$^1\Delta G_{12}$	0.2	0.3	0.6	0.3	0.7	−0.3	−0.6	−0.9
$^0\Delta G_{13}$	0.8	0.5	0.6	0.5	1.2	0.6	0.7	0.1
$^1\Delta G_{13}$	−0.2	−0.1	0.1	−0.0	−0.2	−0.3	−0.4	−0.1
$^0\Delta G_{23}$	1.1	0.5	0.6	0.7	1.4	0.7	0.7	−0.1
$^1\Delta G_{23}$	0.1	−0.0	0.1	0.2	0.0	−0.2	−0.3	−0.3

[a] Listed are the two possible configurations of the third P site (0 = wild-type (wt), 1 = mutant). Errors are ±0.1 kcal/mol or less.

bation in the substrate. There are six possible coupling free energy values for the three pairs, but only four are independent. Hence, the difference between any two values for each pair is exactly the same for all pairs. From the property of the coupling free energy (section III.C), we conclude that the sites are coupled indirectly through interactions higher than second order.

E. Origin of the Higher Specificity of the Fast Form

The Arg → Lys replacement at P1 slightly promotes the slow → fast transition when Gly is present at P2. On the other hand, the Pro → Gly replacement at P2 strongly stabilizes the slow form. The replacement at P3 is inconsequential on the allosteric equilibrium. Hence, the slow → fast transition affects mostly the environment of the S2 site, with modest effects on the S1 site and no effect on the S3 site. Constraints at the S2 site accounts for the lower specificity of the slow form compared to the fast form and become inconsequential if the substrate acquires flexibility with a Gly at P2 and can readjust in the active site to compensate for the increased steric hindrance of the S2 site in the slow form. These findings explain why the thrombin mutant W60dS cleaves FPR with the same specificity in the slow and fast forms[111] and suggest the bulky side chain of W60d as the likely origin of the constraints at S2.

The dominant factors that control specificity are the rigidity of the P2−P3 bond and the strength of the P1−S1 interaction. When the P2−P3 bond is rigid, the substrate finds a more favorable S2 environment in the fast form. Flexibility of the P2−P3 bond relaxes the optimal interaction of Arg at P1 with D189 at S1, this effect being favored by a more accessible active site in the fast form.[104,110] The coupling between substitutions at P1 and P2 comes partially from an intrinsic effect on the substrate, the loss of rigidity of the P2−P3 bond, and partially from the different environment of the enzyme in the slow and fast forms. The less constrained environment of the specificity sites in the fast form also act to reduce the extent of negative coupling among the various perturbations in the substrate, causing the interactions to essentially disappear as more substitutions are introduced at the P sites.

The two ion pairs R221a-E146 and K224-E217 stabilizing the Na^+ binding environment (Figure 7) provide other constraints in the slow form. The R221aA mutant has a reduced Na^+ affinity,[106] suggesting that disruption of the R221aA-E146 ion pair may destabilize the fast form. However, disruption of the R221aA-E146 ion pair affects specificity more in the slow than the fast form. The parameters pertaining to the fast form are practically unchanged relative to wild-type, while those in the slow form show enhanced sensitivity to perturbation at P1 and P2. This perturbation is also less dependent on the state of other groups, indicating a reduction in the coupling among substitutions at the P1−P3 sites (Tables 6 and 7).

Disruption of the R221a-E146 ion pair has a direct influence on the specificity sites S1 and S2 of the enzyme in the slow form and affects the way these sites discriminate between Arg and Lys at P1 or Pro and Gly at P2. This ion pair maintains the correct architecture of the S1 site, especially in the slow form, but also influences the S2 site located some 17 Å away. The molecular basis of this effect may be due to enhanced mobility of the autolysis loop on the Glu side of the ion pair upon disruption of the contact. The enhanced mobility may interfere with substrate recognition in the slow form. The R221a-E146 ion pair contributes to the integrity of the S1 environment in the slow form, but not in the fast form because the perturbation is practically abolished by Na^+ binding.

As for the R221aA mutant, mutation of K224 to Ala reduces the Na^+ affinity,[106] suggesting that disruption of the K224-E217 ion pair may destabilize the fast form, but again, this proposal is contradicted by the experimental data that document a larger perturbation of the slow form (Tables 6 and 7). Disruption of the K224-E217 ion pair produces effects very similar to those seen for the R221aA mutant, with a reduction of the coupling among the P1−P3 sites especially in the slow form. The ion pair between K224 and E217 bridges two residues on the last two β-strands of the B chain contributes to the integrity of the S1 and S2 environments in the slow form. The region in immediate proximity to K224 and E217 plays a key role in substrate selectivity and is absolutely conserved in thrombin from different species.[130] The state of this ion pair can therefore control the access of substrates into the bottom of the catalytic pocket where the specificity site S1 is located.

The two ion pairs interact slightly in the slow form, but not in the fast form, as demonstrated by the results on the double mutant R221aA/K224A (Tables 6 and 7). The perturbation induced by the double mutation is more drastic and almost abolishes Na^+ binding.[106] The mutation affects the response to

Table 9. Coupling Free Energies (kcal/mol) for Perturbation of the P1−P3 Sites of FPR and Residues R221A and K224 of Thrombin[a]

	000	100	010	001	110	101	011	111	coupling	mediated by
					Fast Form					
P1-P2	1.3	0.2	0.8	1.1	0.3	0.5	0.8	0.3	indirect	P3
P1-P3	0.8	−0.2	0.5	0.6	−0.1	0.1	0.5	−0.0	indirect	P2
P1-R221a	0.2	−0.2	−0.1	0.2	−0.1	−0.1	0.0	−0.2	none	
P1-K224	−0.4	−0.6	−0.6	−0.4	−0.4	−0.5	−0.4	−0.4	none	
P2-P3	1.1	0.1	0.5	0.6	−0.0	0.1	0.7	0.2	indirect	P1
P2-R221a	0.5	0.1	−0.1	0.3	−0.1	0.0	0.4	0.1	none	
P2-K224	0.0	−0.2	−0.4	−0.2	−0.2	−0.2	0.0	0.0	none	
P3-R221a	0.2	0.1	−0.1	0.0	0.0	−0.1	0.1	0.0	none	
P3-K224	0.4	0.2	−0.0	−0.1	0.2	−0.0	0.1	0.2	none	
R221a-K224	0.2	0.2	0.0	−0.2	0.1	−0.0	0.2	0.2	none	
					Slow Form					
P1-P2	2.2	0.7	0.6	0.5	−0.3	−0.6	−0.6	−0.9	indirect	P3, R221a, K224
P1-P3	1.2	−0.2	0.6	0.7	−0.3	−0.4	0.1	−0.1	indirect	P2
P1-R221a	0.9	−0.6	0.3	0.5	−0.7	−0.6	−0.0	−0.3	indirect	P2
P1-K224	0.3	−1.4	−0.2	−0.1	−1.6	−1.3	−0.5	−1.1	indirect	P2
P2-P3	1.4	0.0	0.7	0.7	−0.2	−0.3	−0.1	−0.3	indirect	P1, R221a, K224
P2-R221a	1.4	−0.1	0.6	0.7	−0.4	−0.4	−0.1	−0.3	indirect	P1, P3, K224
P2-K224	1.4	−0.3	0.7	0.7	−0.6	−0.5	0.0	−0.6	indirect	P1, P3, R221a
P3-R221a	0.7	0.1	−0.0	0.6	−0.1	0.0	−0.2	0.1	indirect	P2
P3-K224	0.8	0.3	0.1	0.6	−0.0	0.2	−0.1	0.1	indirect	P2
R221a-K224	−0.2	−0.6	−0.9	−0.4	−0.8	−0.7	−1.1	−0.6	indirect	P2

[a] Listed are all possible configurations of the other sites (0 = wild-type, 1 = mutant) in the order P1, P2, P3, R221a, and K224. Errors are ±0.1 kcal/mol or less. Indirect coupling requires values that differ by at least $\pm RT$ (0.6 kcal/mol). Direct coupling of less than $\pm RT$ on the average is considered zero.

perturbations at the P1−P3 site, with an effect more pronounced in the slow form. The site-specific parameters are profoundly altered in the slow form and, interestingly, the pairwise coupling pattern shows the disappearance of indirect coupling in both the slow and fast forms, with the onset of positive second-order direct coupling between P1 and P2 (Table 8). This effect is peculiar to the double substitution, though it is somewhat anticipated by the single substitutions. The molecular basis for the synergism between the R221a-E146 and K224-E217 ion pairs in the slow form is in the participation of residues R221a and K224 in Na^+ and water coordination. In the fast form, the carbonyl O atoms of R221a and K224 directly ligate the Na^+. Mutation of these residues reduces the Na^+ affinity, but high concentrations of Na^+ oppose the structural perturbation induced by the mutation restoring a molecular environment for the specificity sites that is essentially that of the fast form of wild-type. When Na^+ is released, the carbonyl O atom of K224 may reorient as seen in the structure of trypsin and may hydrogen bond to water 447 in concert with the carbonyl O atom of R221a. Water 447 hydrogen bonds to the side chain of D189 in the specificity pocket S1 and through the switching mechanism any perturbation of R221a and K224 changing the orientation of the carbonyl O atoms will not be compensated as in the case of the fast form and therefore may lead to more drastic structural changes.[118]

We conclude that the more constrained environment in the slow form of thrombin is partially due to stronger ion pairs formed by R221a and K224 in the Na^+ binding loop with E146 in the autolysis loop and E217 in the penultimate β-strand of the B chain. The integrity of these ion pairs is essential for maintaining the correct architecture of the specificity sites through the effect on the water molecules in the channel that embeds the specificity site S1. The role of the ion pairs in the fast form appears to be less critical and their disruption can be compensated by the binding of Na^+. The origin of the reduced Na^+ affinity in these mutants should be seen in a perturbation of the slow form leading to an impaired ability to switch to the fast form.[131] The foregoing analysis is invaluable to structure−function studies and to practical issues revolving around the design of better active-site inhibitors. Improvement in the potency of these molecules can be obtained by reducing the negative coupling among the P1−P3 sites. This effect is obtained by keeping a rigid backbone around the P2−P3 position that facilitates the coordination with D189 at S1 and by breaking the ion pairs R221a-E146 and K224-E217.

F. Molecular Origin of the Cooperativity among the P1–P3 Sites

The coupling pattern emerged from the analysis of the substrate library (Table 8) is conducive to negatively cooperative interactions higher than second order. To elucidate the origin of this coupling, derived from the property of the coupling free energy, the entire five-dimensional manifold of species should be considered. This manifold is composed of the sites P1, P2, P3, R221a, and K224, and the relevant free energies are calculated by operating on the values listed in Table 6. Analysis of the coupling pattern involving all possible pairs (Table 9) shows how interactions change with the state of other sites. Considering only differences of at least $\pm RT$ (0.6 kcal/mol) in the coupling free energy, the patterns can be analyzed to identify the nature of the interaction.

In the fast form, only the P1−P3 sites are significantly coupled and in an indirect way. Perturbation of any P site influences the coupling at other sites.

In the slow form all sites are strongly coupled. Each coupling can be dissected to identify the element perturbing the interaction. A direct way to illustrate the effect of a third site on the coupling between two sites is to calculate the difference in coupling free energy of a pair due to the $0 \rightarrow 1$ transition of a third site, in all possible configurations of the remaining sites. The P2 site emerges as a major node of interaction. In the slow form, the state of P2 influences all interactions (Table 9). The state of R221a and K224 influences the P1−P2 and P2−P3 interactions but has no effect on the P1−P3 coupling that is influenced by P2. Finally, the coupling between R221a and K224 is influenced by P2 only. As a result, the Ala replacements at these thrombin residues produce additive effects on specificity when Pro is at P2 but are positively linked when Pro is replaced by Gly.

It is of interest to note that the molecular determinants of cooperativity among the specificity sites in thrombin, like the region around W60d in the S2 site and the R221a-E146 and K224-E217 ion pairs, are not present in trypsin and plasmin. These proteases, unlike thrombin, show simple additivity of the effects of perturbing individual sites in the substrate (Table 6). Disruption of the R221a-E146 and K224-E217 ion pairs in thrombin produces a trypsin-like energetic profile. The coupling free energies reflect the strain imposed by the enzyme on the substrate in the transition state. More constrained environments, like thrombin in the slow form, tend to couple more the substitutions made at different P sites. In more relaxed environments, like thrombin in the fast form or trypsin, the coupling is greatly reduced or absent. The energetic signatures of substrate recognition in these proteases correlate well with the known structural features of the enzymes. Trypsin has a more accessible environment in the specificity sites than thrombin.[119] The information is also valuable when the structure is not known, as in the case of plasmin. The results in Table 6 suggest that the environment of the specificity sites of plasmin is more similar to that of trypsin than thrombin.

The approach based on site-specific thermodynamics is capable of effectively probing the environment of the specificity sites of the enzyme in the transition state. Extension to other proteases, or to other mutant forms of thrombin, may further elucidate the structural determinants of enzyme specificity and the role of cooperativity in substrate recognition. The approach can also be extended to the analysis of ligand binding coupled to mutational effects. The substrate library provides an exceptionally sensitive probe of the molecular environment of the specificity sites of the enzyme and can be used to assess the effect on these sites caused by the binding of allosteric ligands. In the case of thrombin, the library can unravel the effect of thrombomodulin on the specificity sites of the enzyme and help understand the mechanism of action of this important cofactor.

F. How Thrombomodulin Really Works

Thrombomodulin is a cofactor present on the surface of endothelial cells that markedly (~1000-fold) increases the ability of thrombin to activate protein C while it inhibits in a competitive manner fibrinogen binding.[132] It has been proposed that such an effect is borne out by a thrombomodulin-induced change in thrombin conformation,[132,133] but convincing experimental support to this hypothesis has been lacking. The substrate library (Table 4) was therefore used to dissect the effect of thrombomodulin on the specificity sites of thrombin. The site-specific approach reveals important new information on the molecular mechanism of thrombomodulin function.

Table 10. Specificity Constants k_{cat}/K_m (μM^{-1} s^{-1}) for the Hydrolysis of Synthetic Substrates by Thrombin in the Presence of Thrombomodulin or Hir$^{55-65}$ [a]

	FPR	FPK	FGR	VPR	FGK	VPK	VGR	VGK
Fast Form								
thrombomodulin	94	10	2.1	96	0.035	3.7	0.44	0.010
hir$^{55-65}$	117	6.3	1.7	98	0.023	2.0	0.32	0.0064
r_{TM}a	1.0	1.3	1.0	1.0	1.7	1.8	1.3	2.1
r_{hir}b	1.3	0.8	0.8	1.0	1.1	1.0	0.9	1.4
Slow Form								
thrombomodulin	20	3.4	4.4	27	0.020	1.6	0.73	0.0061
hir$^{55-65}$	21	1.6	3.1	24	0.0084	0.57	0.66	0.0027
r_{TM}a	6.7	9.7	5.1	4.0	7.7	15	4.3	7.7
r_{hir}b	7.0	4.6	3.6	3.6	3.2	5.2	3.9	3.4

[a] Experimental conditions: 5 mM Tris, $I = 200$ mM, 0.1% PEG, pH 8.0, at 25 °C, 100 nM thrombomodulin or 100 μM hir$^{55-65}$ The slow form was studied in the presence of 200 mM choline chloride. The properties of the fast form refer to the limit $[Na^+] \rightarrow \infty$, at constant $I = 200$ mM. Errors are typically ±2%. [b] Ratio of specificity relative to the absence of thrombomodulin (see Table 5). [c] Ratio of specificity relative to the absence of hir$^{55-65}$ (see Table 5).

When thrombomodulin binds to the fast form, there is at most a 2-fold enhancement of specificity for all substrates in the library that differ up to 5 orders of magnitude in specificity (Table 10). Binding of thrombomodulin to the slow form produces a consistently higher increase in specificity by as much as 15-fold (Table 10). As a result, thrombomodulin binding tends to abolish the differences between the slow and fast forms, consistent with the observation that the cofactor binds to the fast form with higher affinity.[105,112] This effect is also seen with the natural substrate protein C, which is cleaved by the slow form with significantly higher specificity in the absence but not in the presence of thrombomodulin.[105,106] Interestingly, the slow form becomes more specific in the case of substrates such as FGR and VGR when thrombomodulin binds, suggesting that the cofactor may elicit other effects in addition to the slow $\rightarrow$ fast transition of thrombin. All of these effects, however, are not peculiar to thrombomodulin because the hirudin C-terminal fragments 55−65 (hir),[55-65] reproduce them almost identically, whereas it has no effect on protein C activation. Thrombomodulin and hir$^{55-65}$ share common epitopes on exosite I of thrombin.[134,135] These epitopes may provide the structural basis for the allosteric effects observed on the hydrolysis of chromogenic substrates.

These results have a bearing on the mechanism that leads to the drastic (~1000-fold) enhancement of thrombin specificity toward protein C upon thrombomodulin binding, which is seen in both the slow

and fast forms.[105] The effect of thrombomodulin on the specificity sites S1, S2, and S3 of the enzyme produces a change in specificity that is either small (fast form) or at most 15-fold (slow form). Hence, thrombin must enhance its specificity toward protein C using sites other than those probed by the library of chromogenic substrates, but this has little experimental support.[95] A more reasonable hypothesis is that thrombomodulin exerts its physiologically important function by influencing the conformation of the bound protein C in the thrombin−thrombomodulin−protein C ternary complex, thereby enhancing the specificity of the enzyme by turning protein C into a better substrate. It is unlikely that the structural domains responsible for the enhancement in specificity are entirely located in regions of the enzyme other than the critical sites within the catalytic pocket that can be probed with the substrate library. It is also unlikely that thrombomodulin would induce a large conformational transition in thrombin not linked to a large change in heat capacity.[112] Thrombomodulin makes extensive contacts with thrombin through its EGF domains 5 and 6, whereas its EGF domain 4 may contact protein C.[136] The thrombin−thrombomodulin complex would also have the W60d loop and especially the Na^+ binding loop available for contacting protein C to form the ternary complex. It is conceivable that the bound protein C would make contacts with the bound thrombomodulin, perhaps at the level of the external portion of W60d loop of thrombin. If this were the case, a chromogenic substrate contacting only the interior of the catalytic pocket would not experience the large change in specificity observed for protein C because it would lack the critical direct interaction with the cofactor. This model explains the similarity of effects seen on the chromogenic substrates with thrombomodulin and hir^{55-65} but the lack of effect of hir^{55-65} on protein C hydrolysis. It also predicts that it should be possible to find mutations of thrombomodulin that do not affect binding to thrombin but reduce the ability of thrombin to cleave protein C or mutations of protein C that affect cleavage by thrombin to different extent in the presence and absence of thrombomodulin. A number of such mutations have been reported recently for protein $C^{137-139}$ and provide much support to our proposed mechanism for thrombomodulin function. Thrombomodulin is therefore a competitive inhibitor of fibrinogen binding to thrombin and a cofactor of protein C.

V. New Formalism for the Analysis of Mutational Effects

The analogy between binding and mutational effects introduced in section III can be extended further to develop a new formalism for the analysis of mutational effects. In the case of ligand binding, the quantities accessible to experimental measurements, like the binding isotherm, are continuous. Much can be learned from the shape and properties of the binding isotherm,[7−9] and its analysis yields discrete, site-specific parameters as shown in section II. In the case of mutational effects, on the other hand, the discrete site-specific parameters are determined directly without the need for analyzing continuous quantities that are functions of these parameters. Although this is certainly advantageous, a continuous representation of the energetics may come in quite handy when some general properties of the system are to be illustrated. For example, cooperativity is known to profoundly affect the shape of the binding isotherm.[7−9] Therefore, the analysis of mutational effects in terms of quantities equivalent to the binding isotherm may help elucidate features of the cooperative nature of the process that are difficult to grasp from inspection of the site-specific parameters alone.

A pivotal quantity in the analysis of ligand binding cooperativity is the partition function of the system, Ψ, that lists all intermediates involved in the binding equilibria as defined by the law of mass action.[7−9] The partition function is the sum of the concentrations of all intermediates relative to the concentration of the unligated species used as reference. Once the partition function is correctly defined, the important quantities of the system can be derived by differentiation.

A partition function can also be defined for mutational effects. To this end, we define a general equilibrium reaction $M + jP = M_j$, where M is the wild-type macromolecule, P is a generic site-directed mutation applied to it, and M_j is the macromolecule bearing j such site-specific mutations. The reaction so defined is analogous to the binding equilibrium involving M and j ligand molecules. The free energy for the equilibrium reaction is defined from the difference in chemical potential between the product M_j and the parent species M and P. The difference between the chemical potentials of M_j and M is given by the difference in stability or specificity calculated experimentally. For example, in the case of substrate binding, the ratio of the concentrations of macromolecular species $[M_j]/[M]$, is given by s_j/s, where s refers to the specificity. Consideration of the site-specific intermediates in the system gives the partition function

$$\Psi = \sum_{\alpha=0}^{1}\sum_{\beta=0}^{1}\cdots\sum_{\omega=0}^{1}\frac{[M_{\alpha\beta\ldots\omega}]}{[M_{00\ldots0}]}x^{\alpha+\beta+\ldots+\omega} \qquad (32)$$

$M_{00\ldots0}$ is the reference wild-type macromolecule to which mutations at sites 1, 2, ... N can be introduced. The variable $0 \le x \le \infty$ is a dummy quantity analogous to the ligand concentration and can be thought of as the driving force responsible for $0 \rightarrow 1$ or wild-type $\rightarrow$ mutant transition taking place at each site of the macromolecule. The discrete nature of the process describing the perturbations at each site is given a continuous description through the variable x, just like ligand binding to discrete sites of a macromolecule is given a continuous description in terms of the polynomial expansion analogous to eq 32. In general, any coefficient of the partition function can be written in terms of a coupling free energy of order $\alpha + \beta + \ldots + \omega$ plus the sum of $\alpha + \beta + \ldots + \omega$ site-specific perturbation free energies. For example, the partition function for the case of mutations at two sites characterized in terms of enzyme

specificity is

$$\Psi = \frac{[M_{00}]}{[M_{00}]} + \left(\frac{[M_{10}]}{[M_{00}]} + \frac{[M_{01}]}{[M_{00}]}\right)x + \frac{[M_{11}]}{[M_{00}]}x^2 = 1 + \left(\frac{s_{10}}{s_{00}} + \frac{s_{01}}{s_{00}}\right)x + \frac{s_{11}}{s_{00}}x^2 = 1 + \left[\exp\left(-\frac{\Delta G_1}{RT}\right) + \exp\left(-\frac{\Delta G_2}{RT}\right)\right]x + \exp\left(-\frac{\Delta G_1 + \Delta G_2 + \Delta G_{12}}{RT}\right)x^2 \quad (33)$$

where the *s*'s refer to specificity constants of the various intermediates. Alternatively, eq 33 can describe the effect of mutating two sites on the stability of a protein (see below). Identification of x with the Ca^{2+} concentration turns eq 33 into the partition function for ligand binding to calbindin (see section II.C).

Differentiation of the logarithm of the partition function relative to the logarithm of x gives a quantity, X, analogous to the average number of ligated sites in ligand binding processes. More informative for mutational effects is however the derivative of X analogous to the binding capacity.[9] The quantity

$$\chi = \frac{dX}{d \ln x} = \frac{d^2 \ln \Psi}{d \ln^2 x} \quad (34)$$

defines the global *susceptibility* of the system, or the probability density that a given perturbation in the system will cause a certain free energy change in specificity or stability. Specifically, the product χN^{-1} d ln x is the probability density that a mutation produces a free energy perturbation of specificity or stability comprised between RT ln x and RT(ln x + d ln x). This information is obtained directly from a plot of χ versus $\Delta G = RT$ ln x and is of immediate practical relevance. In the susceptibility plot, the dummy variable x assumes physical meaning through the free energy change ΔG caused by a given perturbation introduced in the system. The response of the system to the perturbation is proportional to the value of χ for a given value of ΔG.

The information generated at the site-specific level with mutational perturbations, as in the case of thrombin dealt with in section IV, is sufficient to define susceptibilities for each perturbed site. As for ligand binding cooperativity, the quantities defined in the global description can be decomposed into their site-specific contributions as implied by the basic eq 1. To define X_j, we make use of contracted forms of the partition function containing all configurations with site j perturbed, ${}^1\Psi_j$, or wild-type, ${}^0\Psi_j$, as shown in section II for ligand binding. From the definition of X_j it also follows that

$$\chi_j = \frac{dX_j}{d \ln x} \quad (35)$$

with the conservation relationship

$$\chi = \sum_{j=1}^{N} \chi_j \quad (36)$$

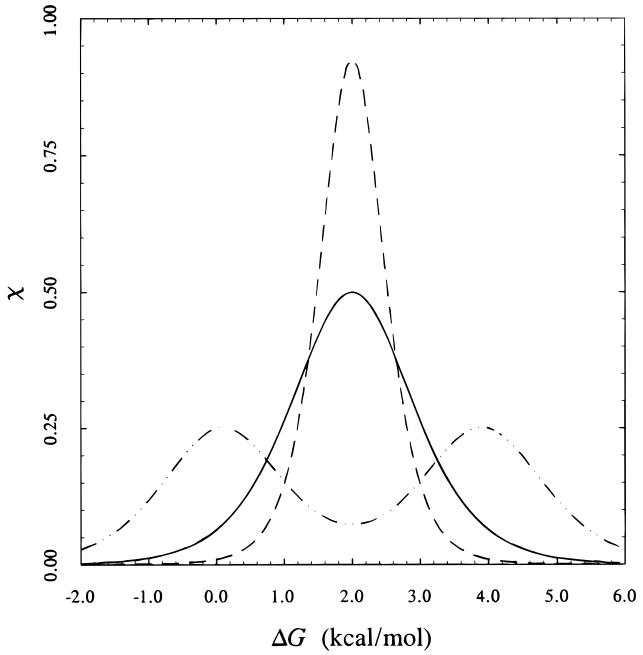


Figure 9. Global susceptibility profiles for the case of two identically perturbed and independent sites (continuous line; $\Delta G_1 = 2$ kcal/mol, $\Delta G_2 = 2$ kcal/mol, and $\Delta G_{12} = 0$ kcal/mol). The cases of two positively (discontinuous line; $\Delta G_1 = 3.5$ kcal/mol, $\Delta G_2 = 3.5$ kcal/mol, and $\Delta G_{12} = -3$ kcal/mol) or negatively (discontinuous-dotted line; $\Delta G_1 = 0.5$ kcal/mol, $\Delta G_2 = 0.5$ kcal/mol, and $\Delta G_{12} = 3$ kcal/mol) linked sites are also shown for comparison. The three cases are constructed so to have the same value of $\Delta G_m = 2$ kcal/mol, using the partition function in eq 33 in the text.

The site-specific susceptibility defines the probability density χ_j d ln x that a given perturbation or mutation in the system will cause a free energy perturbation at site j comprised between RT ln x and RT(ln x + d ln x).

Integration of the susceptibility profile yields important information on the energetics that is difficult to obtain from the site-specific parameters. The information is particularly relevant in the site-specific case when cooperativity is present. The first moment of the global susceptibility defines the *mean free energy of perturbation*, i.e.

$$\Delta G_m = RT \ln x_m = \frac{1}{N}\left[\int_{-\infty}^{+\infty} \chi \Delta G \, d\Delta G / \int_{-\infty}^{+\infty} \chi \, d\Delta G\right] \quad (37)$$

ΔG_m measures the average perturbation per site defined as the free energy change in going from the wild-type to the fully perturbed configuration, divided by the number of sites. In fact, it can be shown that[9]

$$\Delta G_m = (\Delta G_1 + \Delta G_2 + ... + \Delta G_N + \Delta G_{12...N})/N \quad (38)$$

The quantity x_m is the analogue of Wyman's median ligand activity in ligand binding processes[8,140] that defines the value of x where the unligated and fully ligated configurations are equally populated.

The global susceptibility for a system composed of two identical and independent sites is illustrated in Figure 9. A steeper distribution is indicative of positive interactions among the sites, whereas a more

broader distribution is conducive to negative interactions or site heterogeneity in response to the mutational perturbations. The area under the curve is constant and gives the number of sites N.

The site-specific susceptibilities define quantities analogous to ΔG_m, i.e.

$$\Delta G_{m,j} = \int_{-\infty}^{+\infty} \chi_j \Delta G \, d\,\Delta G / \int_{-\infty}^{+\infty} \chi_j \, d\,\Delta G \qquad (39)$$

The value of $\Delta G_{m,j}$ is the mean free energy of perturbation at site j when a given mutation is introduced in the system. This value is particularly informative in the presence of interactions among the sites because of the potential ambiguity in defining the cost of a mutation at site j (Table 7). Unlike ΔG_m, $\Delta G_{m,j}$ cannot be expressed as a simple function of the site-specific parameters, except in the trivial case of independent sites. The value of $\Delta G_{m,j}$ must be obtained by integration of the susceptibility profile. We also note the conservation relationship

$$\Delta G_m = \frac{1}{N}\sum_{j=1}^{N} \Delta G_{m,j} \qquad (40)$$

as a direct consequence of eq 36. The sum of the mean free energies of perturbation for the N individual sites gives the mean free energy of perturbation in the global description times the number of sites. Alternatively, the mean free energy of perturbation in the global description is the average of the site-specific mean free energies of perturbation.

As an illustrative example, we calculate the global and site-specific susceptibility profiles for the slow and fast forms of thrombin interacting with the substrate library discussed in section IV. The results are shown in Figure 10. The partition function for the system carrying perturbations at the P1−P3 sites of the substrate FPR is

$$\begin{aligned}\Psi = \frac{[M_{000}]}{[M_{000}]} + \left(\frac{[M_{100}]}{[M_{000}]} + \frac{[M_{010}]}{[M_{000}]} + \frac{[M_{001}]}{[M_{000}]}\right)x + \\ \left(\frac{[M_{110}]}{[M_{000}]} + \frac{[M_{101}]}{[M_{000}]} + \frac{[M_{011}]}{[M_{000}]}\right)x^2 + \frac{[M_{111}]}{[M_{000}]}x^3 = \\ 1 + \left(\frac{s_{100}}{s_{000}} + \frac{s_{010}}{s_{000}} + \frac{s_{001}}{s_{000}}\right)x + \left(\frac{s_{110}}{s_{000}} + \frac{s_{101}}{s_{000}} + \frac{s_{001}}{s_{000}}\right)x^2 + \\ \frac{s_{111}}{s_{000}}x^3 = 1 + \left[\exp\left(-\frac{\Delta G_1}{RT}\right) + \exp\left(-\frac{\Delta G_2}{RT}\right) + \right. \\ \left.\exp\left(-\frac{\Delta G_3}{RT}\right)\right]x + \left[\exp\left(-\frac{\Delta G_1 + \Delta G_2 + \Delta G_{12}}{RT}\right) + \right. \\ \exp\left(-\frac{\Delta G_1 + \Delta G_3 + \Delta G_{13}}{RT}\right) + \\ \left.\exp\left(-\frac{\Delta G_2 + \Delta G_3 + \Delta G_{23}}{RT}\right)\right]x^2 + \\ \exp\left(-\frac{\Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_{123}}{RT}\right)x^3 \qquad (41)\end{aligned}$$

where the suffix denotes perturbation at site 1 (P1), 2 (P2), and 3 (P3) in order. For example, s_{FPR} is s_{000} and s_{VGR} is s_{110} (see eqs 25−31). The various ΔG's

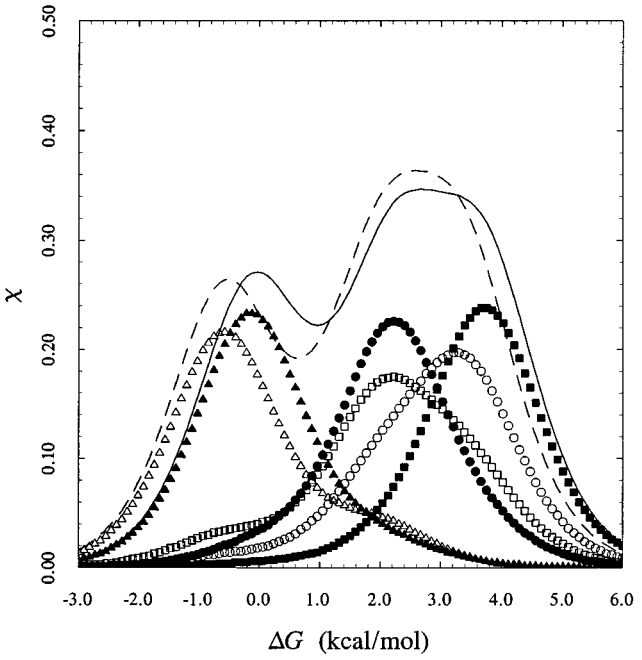


Figure 10. Global susceptibility profiles of the slow (discontinuous line) and fast (continuous line) forms of thrombin, constructed from eqs 41−44 in the text using the parameters listed in Table 6. The profiles are decomposed into the site-specific components showing the susceptibility of S1 to perturbation in P1 (○,●), S2 to perturbation in P2 (□,■), and S3 to perturbation in P3 (△,▲) in the slow (○,□,△) and fast (●,■,▲) forms. The values of mean free energy of perturbation are (slow form) $\Delta G_m = 1.6$ kcal/mol, $\Delta G_{m,1} = 2.9$ kcal/mol, $\Delta G_{m,2} = 2.2$ kcal/mol, and $\Delta G_{m,3} = -0.2$ kcal/mol and (fast form) $\Delta G_m = 1.9$ kcal/mol, $\Delta G_{m,1} = 2.2$ kcal/mol, $\Delta G_{m,2} = 3.6$ kcal/mol, and $\Delta G_{m,3} = 0.0$ kcal/mol.

are the same as those listed in Table 6. The expressions for the site-specific quantities X_1, X_2 and X_3 are

$$\begin{aligned}X_1 = 1 - \left[1 + \left(\frac{s_{010}}{s_{000}} + \frac{s_{001}}{s_{000}}\right)x + \frac{s_{011}}{s_{000}}x^2\right] \Big/ \\ \left[1 + \left(\frac{s_{100}}{s_{000}} + \frac{s_{010}}{s_{000}} + \frac{s_{001}}{s_{000}}\right)x + \left(\frac{s_{110}}{s_{000}} + \frac{s_{101}}{s_{000}} + \frac{s_{011}}{s_{000}}\right)x^2 + \right. \\ \left. \frac{s_{111}}{s_{000}}x^3\right] \qquad (42)\end{aligned}$$

$$\begin{aligned}X_2 = 1 - \left[1 + \left(\frac{s_{100}}{s_{000}} + \frac{s_{001}}{s_{000}}\right)x + \frac{s_{101}}{s_{000}}x^2\right] \Big/ \\ \left[1 + \left(\frac{s_{100}}{s_{000}} + \frac{s_{010}}{s_{000}} + \frac{s_{001}}{s_{000}}\right)x + \left(\frac{s_{110}}{s_{000}} + \frac{s_{101}}{s_{000}} + \frac{s_{011}}{s_{000}}\right)x^2 + \right. \\ \left. \frac{s_{111}}{s_{000}}x^3\right] \qquad (43)\end{aligned}$$

$$\begin{aligned}X_3 = 1 - \left[1 + \left(\frac{s_{100}}{s_{000}} + \frac{s_{010}}{s_{000}}\right)x + \frac{s_{110}}{s_{000}}x^2\right] \Big/ \\ \left[1 + \left(\frac{s_{100}}{s_{000}} + \frac{s_{010}}{s_{000}} + \frac{s_{001}}{s_{000}}\right)x + \left(\frac{s_{110}}{s_{000}} + \frac{s_{101}}{s_{000}} + \frac{s_{011}}{s_{000}}\right)x^2 + \right. \\ \left. \frac{s_{111}}{s_{000}}x^3\right] \qquad (44)\end{aligned}$$

and the site-specific susceptibilities are derived from differentiation of eqs 42−44 according to eq 35.

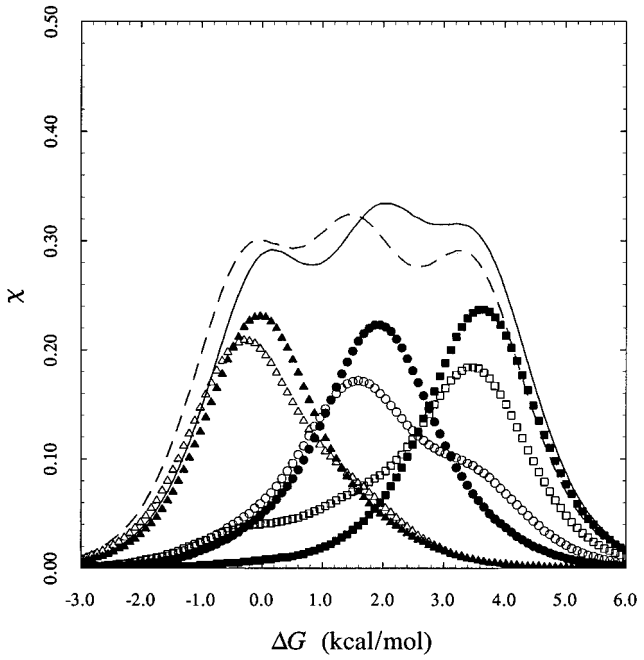


Figure 11. Effect of thrombomodulin on the global susceptibility profiles of the slow (discontinuous line) and fast (continuous line) forms of thrombin, constructed from eqs 41−44 in the text using the specificity values listed in Table 10. The profiles are decomposed into the site-specific components showing the susceptibility of S1 to perturbation in P1 (○,●), S2 to perturbation in P2 (□,■), and S3 to perturbation in P3 (△,▲) in the slow (○,□,△) and fast (●,■,▲) forms. Comparison with the data in Figure 10 show how the profiles tend to become similar in the two forms at both the global and site-specific level. The values of mean free energy of perturbation are (slow form) $\Delta G_m = 1.6$ kcal/mol, $\Delta G_{m,1} = 1.9$ kcal/mol, $\Delta G_{m,2} = 2.7$ kcal/mol, and $\Delta G_{m,3} = 0.0$ kcal/mol and (fast form) $\Delta G_m = 1.8$ kcal/mol, $\Delta G_{m,1} = 1.9$ kcal/mol, $\Delta G_{m,2} = 3.5$ kcal/mol, and $\Delta G_{m,3} = 0.1$ kcal/mol.

The global susceptibility (Figure 10) spans 9 kcal/mol, consistent with the presence of negative coupling among the sites and heterogeneous response of the P1−P3 sites to structural perturbation. There is very little difference in the profiles of the two forms, as also indicated by the similar values of ΔG_m (1.9 kcal/mol for the fast form and 1.6 kcal/mol for the slow form). However, the similarity in the global susceptibility is contrasted by significant differences in the site-specific susceptibilities. There is a profound difference in the response to perturbation at P1, with the fast form being less susceptible by 0.7 kcal/mol. The slow → fast transition affects the environment of D189 at S1 by making it less susceptible to the Arg → Lys replacement at P1. This difference is not borne out by a difference in the site-specific perturbation free energy ΔG_1 (Table 6), but rather by a different coupling between S1 and S2 in the two forms. The susceptibility profile is in this case very informative because it reveals an important property of the system that is not easily anticipated upon inspection of the site-specific parameters. Likewise, there is a profound difference in the response to perturbation at P2, with the fast form being more susceptible by 1.4 kcal/mol. The shape of the susceptibility is in this case very different in the two forms. The slow → fast transition affects the environment of the apolar site S2 by making it more susceptible to the Pro → Gly replacement at P2. The cooperativity at this site is also higher in the fast form, indicating a reduction of the negative linkage with S1 and S3. These effects are primarily accounted for by the differences in the perturbation free energy ΔG_2, which is smaller in the slow form, and by the stronger negative coupling between P1 and P2 in this form. The response to perturbation at P3 shows no significant differences in the two forms.

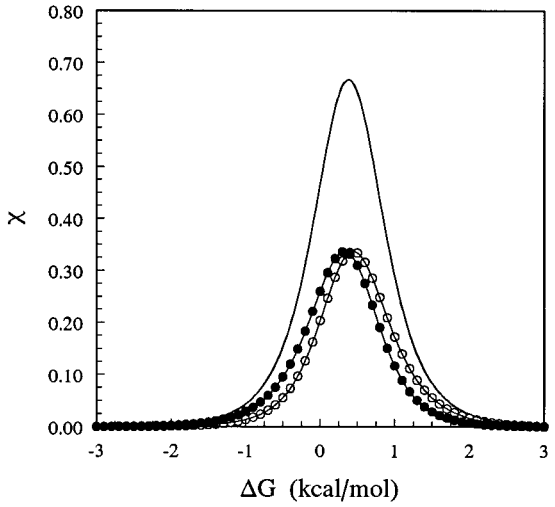


Figure 12. Susceptibility profiles for the Ala substitution of the Glu-Arg pair at position 44 and 53 of a variant of the immunoglobulin G binding domain of streptococcal protein G. The global susceptibility (continuous line) is decomposed into the site-specific components showing the susceptibility of E44 (●) and R53 (○) to the Ala substitution. Comparison with the data in Figure 13 shows the difference in cooperativity between the pair. Curves were drawn using eqs 45−47 in the text with parameter values: $\Delta G_1 = 0.84$ kcal/mol, $\Delta G_2 = 1.47$ kcal/mol, and $\Delta G_{12} = -1.09$ kcal/mol.

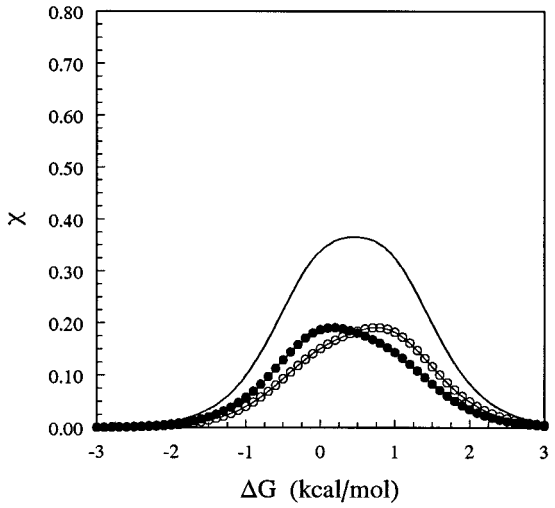


Figure 13. Susceptibility profiles for the Ala substitution of the Thr-Ile pair at position 44 and 53 of a variant of the immunoglobulin G binding domain of streptococcal protein G. The global susceptibility (continuous line) is decomposed into the site-specific components showing the susceptibility of T44 (●) and I53 (○) to the Ala substitution. Comparison with the data in Figure 12 shows the difference in cooperativity between the pair. Curves were drawn using eqs 45−47 in the text with parameter values $\Delta G_1 = 0.24$ kcal/mol, $\Delta G_2 = 0.49$ kcal/mol, and $\Delta G_{12} = 0.74$ kcal/mol.

The susceptibility profiles also help understand the effect of thrombomodulin on the two forms of thrombin (Figure 11). The most notable effect is that the profiles for perturbation at P1 and P2 become more similar in the slow and fast forms, in contrast to what is seen in the absence of cofactor (Figure 10). The overall effect of thrombomodulin binding is to change the environment of the slow form around the S1 and S2 sites into a fast-like conformation, so that the differences between the two allosteric states are reduced. The susceptibility profiles illustrate directly the mechanism of action of this important cofactor.

As another example we analyze the elegant work of Smith and Regan on the energetics of β-sheet side chain interactions.[141] They determined the effect on protein stability resulting from the substitution of cross-strand pairs of side chains on an antiparallel β-sheet of a variant of the immunoglobulin G binding domain. Several residues were substituted at position 44 and 53 facing each other on the opposite strands and the results were expressed relative to the stability of the Ala residues at either position. These data can be analyzed using the same formalism as developed for the analysis of enzyme catalysis and ligand binding. The partition function for the system carrying Ala perturbations at positions 44 (site 1) and 53 (site 2) is

$$\Psi = \frac{[M_{00}]}{[M_{00}]} + \left(\frac{[M_{10}]}{[M_{00}]} + \frac{[M_{01}]}{[M_{00}]}\right)x + \frac{[M_{11}]}{[M_{00}]}x^2 = 1 + \left[\exp\left(-\frac{\Delta G_1}{RT}\right) + \exp\left(-\frac{\Delta G_2}{RT}\right)\right]x + \exp\left(-\frac{\Delta G_1 + \Delta G_2 + \Delta G_{12}}{RT}\right)x^2 \quad (45)$$

ΔG_1 is the free energy change in stability due to replacement of the residue at position 44 with Ala, ΔG_2 is analogous free energy change in stability due to replacement of the residue at position 53 with Ala, and ΔG_{12} is the coupling free energy between the substitutions. The expressions for the site-specific quantities X_1 and X_2 are

$$X_1 = 1 - \left[1 + \exp\left(-\frac{\Delta G_2}{RT}\right)x\right] \Big/ \left\{1 + \left[\exp\left(-\frac{\Delta G_1}{RT}\right) + \exp\left(-\frac{\Delta G_2}{RT}\right)\right]x + \exp\left(-\frac{\Delta G_1 + \Delta G_2 + \Delta G_{12}}{RT}\right)x^2\right\} \quad (46)$$

$$X_2 = 1 - \left[1 + \exp\left(-\frac{\Delta G_1}{RT}\right)x\right] \Big/ \left\{1 + \left[\exp\left(-\frac{\Delta G_1}{RT}\right) + \exp\left(-\frac{\Delta G_2}{RT}\right)\right]x + \exp\left(-\frac{\Delta G_1 + \Delta G_2 + \Delta G_{12}}{RT}\right)x^2\right\} \quad (47)$$

and the site-specific susceptibilities are derived from differentiation according to eq 35.

The susceptibility profiles are shown in Figures 12 and 13 for the Glu-Arg and Thr-Ile pairs of residues 44−53. In the case of the Glu-Arg pair, the coupling is positive and the susceptibility profile is peaked around the mean free energy of perturbation. The existence of positive coupling makes the site-specific susceptibilities look very similar, although the site-specific free energies of perturbation differ by nearly 0.7 kcal/mol. In the case of the Thr-Ile pair, on the other hand, the coupling is negative and the susceptibilities spread out over a range of 4 kcal/mol. As a result, the site-specific susceptibilities look different and peak at free energy values about 1 kcal/mol apart, although the intrinsic free energies of perturbation are within 0.2 kcal/mol.

VI. Conclusions

Site-specific thermodynamics[5,9] provides a general theory of cooperativity and extends previous treatments based exclusively on global effects.[7,8] The contribution of individual sites or residues to protein stability and ligand recognition can be dissected by introducing site-specific perturbations by means of recombinant DNA technologies. The combination of this experimental tool with the principles of site-specific thermodynamics results in a powerful new strategy that can detect the extent, nature, and origin of cooperativity in the system. Structural perturbations should be introduced in the system in a rational manner to generate all intermediates for a site-specific dissection of the energetics. We have illustrated how to implement this strategy in practice using substrate recognition by thrombin. The amount of information and the degree of detail on the recognition process to be gained from such novel approach is unprecedented. Extension of the same strategy to other systems is possible and highly desirable. The conceptual framework discussed in this review article will certainly appeal to biochemists and biophysicists involved in studies of structure−function relationships and molecular recognition in proteins and nucleic acids.

VII. Acknowledgments

I am grateful to Prof. Peter Lollar for providing unpublished data on the interaction of factor VIII with the monoclonal antibody 413 (Table 3). This work was supported by NIH Research Grants HL49413 and HL58141 and was carried out under the tenure of an Established Investigator Award in Thrombosis from the American Heart Association and Genentech.

VIII. References

(1) Perutz, M. F. *Q. Rev. Biophys.* **1989**, *22*, 139−236.
(2) Creighton, T. E. *Protein Folding*; Freeman: New York, 1992.
(3) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526−535.
(4) Smith, M. *Annu. Rev. Genet.* **1985**, *19*, 423−462.
(5) Di Cera, E. *Adv. Protein Chem.* **1998**, *51*, 59−119.
(6) Wegscheider, R. *Monatsch. Chem.* **1895**, *16*, 153−158.
(7) Hill, T. L. *Cooperativity Theory in Biochemistry*; Springer-Verlag: Berlin, 1984.
(8) Wyman, J.; Gill, S. J. *Binding and Linkage*; University Science Books: Mill Valley, CA, 1990.
(9) Di Cera, E. *Thermodynamic Theory of Site-Specific Binding Processes in Biological Macromolecules*; Cambridge University: Cambridge, U.K., 1995.
(10) Adams, E. Q. *J. Am. Chem. Soc.* **1916**, *38*, 1503−1510.
(11) Simms, H. S. *J. Am. Chem. Soc.* **1926**, *48*, 1239−1261.

(12) Edsall, J. T.; Blanchard, M. H. *J. Am. Chem. Soc.* **1933**, *55*, 2337−2353.
(13) Edsall, J. T.; Wyman, J. *Biophysical Chemistry*; Academic: New York, 1958.
(14) Neuberger, A. *Biochem. J.* **1936**, *30*, 2085−2094.
(15) Hill, T. L. *J. Chem. Phys.* **1944**, *12*, 56−61.
(16) Linderstrøm-Lang, K. *Compt. Rend. Lab. Carlsberg* **1924**, *15*, 1−29.
(17) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 5333−5339.
(18) Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 9556−9561.
(19) Kretsinger, R. H. *CRC Crit. Rev. Biochem.* **1980**, *8*, 119−174.
(20) Qian, H. *Biopolymers* **1993**, *33*, 1605−1616.
(21) Chakrabarty, A.; Schellman, J. A.; Baldwin, R. L. *Nature* **1991**, *351*, 586−588.
(22) Wrabl, J. O.; Shortle, D. *Protein Sci.* **1996**, *5*, 2343−2352.
(23) Skelton, N. J.; Kördel, J.; Akke, M.; Forsén, S.; Chazin, W. J. *Nat. Struct. Biol.* **1994**, *1*, 239−245.
(24) Szebenyi, D. M. E.; Moffat, J. *J. Biol. Chem.* **1986**, *261*, 8761−8777.
(25) Linse, S.; Brodin, P.; Drakenberg, T.; Thulin, E.; Sellers, P.; Elmdén, K.; Grundström, T.; Forsén, S. *Biochemistry* **1987**, *26*, 6723−6735.
(26) Linse, S.; Brodin, P.; Johansson, C.; Thulin, E.; Grundström, T.; Forsén, S. *Nature* **1988**, *335*, 651−652.
(27) Linse, S.; Johansson, C.; Brodin, P.; Grundström, T.; Drakenberg, T.; Forsén, S. *Biochemistry* **1991**, *30*, 154−162.
(28) Akke, M.; Forsén, S.; Chazin, W. J. *J. Mol. Biol.* **1991**, *220*, 173−189.
(29) Carlström, G.; Chazin, W. J. *J. Mol. Biol.* **1993**, *231*, 415−430.
(30) Linse, S.; Chazin, W. J. *Protein Sci.* **1995**, *4*, 1038−1044.
(31) Brodin, P.; Johansson, C.; Forsén, S.; Drakenberg, T.; Grundström, T. *J. Biol. Chem.* **1990**, *265*, 11125−11130.
(32) Martin, S. R.; Linse, S.; Johansson, C.; Bayley, P. M.; Forsén, S. *Biochemistry* **1990**, *29*, 4188−4193.
(33) Ahlström, P.; Teleman, O.; Kördel, J.; Forsén, S.; Jönsson, B. *Biochemistry* **1989**, *28*, 3205−3211.
(34) Nayal, M.; Di Cera, E. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 817−821.
(35) Kojima, N.; Palmer, G. *J. Biol. Chem.* **1983**, *258*, 14908−14913.
(36) Hendler, R. W.; Subba Reddy, K. V.; Shrager, R. I.; Caughey, W. S. *Biophys. J.* **1986**, *49*, 717−729.
(37) Senear, D. F.; Ackers, G. K. *Biochemistry* **1990**, *29*, 6568−6577.
(38) Perrella, M.; Rossi-Bernardi, L. *Methods Enzymol.* **1981**, *76*, 133−143.
(39) Ackers, G. K.; Doyle, M. L.; Myers, D.; Daugherty, M. A. *Science* **1992**, *255*, 54−63.
(40) Judice, J. K.; Gamble, T. R.; Murphy, E. C.; de Vos, A. M.; Schultz, P. G. *Science* **1993**, *261*, 1578−1581.
(41) Cornish, V. W.; Schultz, P. G. *Curr. Opin. Struct. Biol.* **1994**, *4*, 601−607.
(42) Matthews, B. W. *Annu. Rev. Biochem.* **1993**, *62*, 139−160.
(43) Yu, M.-H.; Weissman, J. S.; Kim, P. S. *J. Mol. Biol.* **1995**, *249*, 388−397.
(44) Shortle, D. *FASEB J.* **1996**, *10*, 27−34.
(45) Meeker, A. L.; Garcia-Moreno, B.; Shortle, D. *Biochemistry* **1996**, *35*, 6443−6449.
(46) Milla, M. E.; Brown, B. M.; Sauer, R. T. *Nat. Struct. Biol.* **1994**, *1*, 518−523.
(47) Garcia-Moreno, E., B.; Dwyer, J. J.; Gittis, A. G.; Lattman, E. E.; Spencer, D. S.; Stites, W. E. *Biophys. Chem.* **1997**, *64*, 211−224.
(48) Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383−386.
(49) Castro, M. J. M.; Anderson, S. *Biochemistry* **1996**, *35*, 11435−11446.
(50) Tsiang, M.; Jain, A. K.; Dunn, K. E.; Rojas, M. E.; Leung, L. L. K.; Gibbs, C. S. *J. Biol. Chem.* **1995**, *270*, 16854−16863.
(51) Dickinson, C. D.; Kelly, C. R.; Ruf, W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14379−14384.
(52) Pakianathan, D. R.; Kuta, E. G.; Artis, D. R.; Skelton, N. J.; Hebert, C. A. *Biochemistry* **1997**, *36*, 9642−9648.
(53) Lau, F. T.-K.; Fersht, A. R. *Nature* **1987**, *326*, 811−812.
(54) Cunningham, B. C.; Wells, J. A. *Science* **1989**, *244*, 1081−1085.
(55) Green, S. M.; Meeker, A. K.; Shortle, D. *Biochemistry* **1992**, *31*, 5717−5728.
(56) Horovitz, A.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 733−740.
(57) Fersht, A. R.; Serrano, L. *Curr. Opin. Struct. Biol.* **1993**, *3*, 75−83.
(58) Carter, P.; Wells, J. A. *Nature* **1988**, *332*, 564−568.
(59) Vindigni, A.; Dang, Q. D.; Di Cera, E. *Nat. Biotechnol.* **1997**, *15*, 891−895.
(60) Sandberg, W. S.; Terwilliger, T. C. *Science* **1989**, *245*, 54−57.
(61) Shirley, B. A.; Stanssen, P.; Steyaert, J.; Pace, C. N. *J. Biol. Chem.* **1989**, *264*, 11621−11625.
(62) Wells, J. A. *Biochemistry* **1990**, *29*, 8509−8517.
(63) Carter, P. J.; Winter, G.; Wilkinson, A. J.; Fersht, A. R. *Cell* **1984**, *38*, 835−840.
(64) Mildvan, A. S.; Weber, D. J.; Kuliopulos, A. *Arch. Biochem. Biophys.* **1992**, *294*, 327−340.
(65) Shortle, D.; Meeker, A. L. *Proteins: Struct., Funct., Genet.* **1986**, *1*, 81−89.
(66) Perry, K. M.; Onuffer, J. J.; Gittelman, M. S.; Barmat, L.; Matthews, C. R. *Biochemistry* **1989**, *28*, 7961−7970.
(67) Howell, E. E.; Booth, C.; Farnum, M.; Kraut, J.; Warren, M. S. *Biochemistry* **1990**, *29*, 8561−8568.
(68) LiCata, V. J.; Speros, P. C.; Rovida, E.; Ackers, G. K. *Biochemistry* **1990**, *29*, 9771−9783.
(69) Scrutton, N. S.; Berry, A.; Perham, R. N. *Nature* **1990**, *343*, 38−43.
(70) Green, S. M.; Shortle, D. *Biochemistry* **1993**, *32*, 10131−10139.
(71) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1993**, *32*, 13909−13918.
(72) Robinson, C. R.; Sligar, S. G. *Protein Sci.* **1993**, *2*, 826−832.
(73) LiCata, V. J.; Ackers, G. K. *Biochemistry* **1995**, *34*, 3133−3159.
(74) Shortle, D.; Stites, W. E.; Meeker, A. L. *Biochemistry* **1990**, *29*, 8033−8041.
(75) Richieri, G. V.; Low, P. J.; Ogata, R. T.; Kleinfeld, A. M. *J. Biol. Chem.* **1997**, *272*, 16737−16740.
(76) Young, D. C.; Zhan, H.; Cheng, Q.-L.; Hou, J.; Matthews, D. J. *Protein Sci.* **1997**, *6*, 1228−1236.
(77) Kristensen, C.; Kjeldsen, T.; Wiberg, F. C.; Schaffer, L.; Hach, M.; Havelund, S.; Bass, J.; Steiner, D. F.; Andersen, A. S. *J. Biol. Chem.* **1997**, *272*, 12978−12983.
(78) Cosmatos, A.; Cheng, K.; Okada, Y.; Katsoyannis, P. G. *J. Biol. Chem.* **1978**, *253*, 6586−6590.
(79) Nakagawa, S. H.; Tager, H. S. *Biochemistry* **1992**, *31*, 3204−3214.
(80) Marki, F.; Gasparo, M. D.; Eisler, K.; Kambler, B.; Riniker, B.; Rittel, W.; Sieber, P. *Hoppe-Seyler's Z. Physiol. Chem.* **1979**, *360*, 1619−1632.
(81) Nakagawa, S. H.; Tager, H. S. *J. Biol. Chem.* **1991**, *266*, 11502−11509.
(82) Kobayashi, M.; Ohgaku, S.; Iwasaki, M.; Maegawa, H.; Watanabe, N.; Takada, Y.; Shigeta, Y.; Inouye, K. *Biomed. Res.* **1984**, *5*, 267−272.
(83) Mirmira, R. G.; Tager, H. S. *Biochemistry* **1991**, *30*, 8222−8229.
(84) Lubin, I. M.; Healey, J. F.; Barrow, R. T.; Scandella, D.; Lollar, P. *J. Biol. Chem.* **1997**, *272*, 30191−30195.
(85) Ackers, G. K.; Smith, F. R. *Annu. Rev. Biochem.* **1985**, *54*, 597−629.
(86) Horovitz, A.; Fersht, A. R. *J. Mol. Biol.* **1990**, *214*, 613−617.
(87) Koshland, D. E.; Nemethy, G.; Filmer, D. *Biochemistry* **1966**, *5*, 365−385.
(88) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88−118.
(89) Rawlings, R. D.; Barrett, A. J. *Biochem. J.* **1993**, *290*, 205−218.
(90) Rawlings, R. D.; Barrett, A. J. *Methods Enzymol.* **1994**, *244*, 19−61.
(91) Neurath, H. *Science* **1984**, *224*, 350−357.
(92) Doolittle, R. F.; Feng, D. F. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 869−874.
(93) Patthy, L. *Blood Coagulation Fibrinolysis* **1990**, *1*, 153−166.
(94) Dang, Q. D.; Di Cera, E. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10253−10256.
(95) Di Cera, E.; Dang, Q. D.; Ayala, Y. M. *Cell. Mol. Life Sci.* **1997**, *53*, 701−730.
(96) Lesk, A. M.; Fordham, W. D. *J. Mol. Biol.* **1996**, *258*, 501−537.
(97) Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J. K. *Biochemistry* **1989**, *28*, 3629−3637.
(98) Hedstrom, L.; Szilagyi, L.; Rutter, W. J. *Science* **1992**, *255*, 1249−1253.
(99) Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157−162.
(100) Mann, K. G.; Nesheim, M. E.; Church, W. R.; Haley, P.; Krishnaswamy, S. *Blood* **1990**, *76*, 1−16.
(101) Davie, E. W.; Fujikawa, K.; Kisiel, W. *Biochemistry* **1991**, *30*, 10363−10370.
(102) Grand, R. J. A.; Turnell, A. S.; Grabham, P. W. *Biochem. J.* **1996**, *313*, 353−368.
(103) Ishihara, H.; Connolly, A. J.; Zeng, D.; Kahn, M. L.; Zheng, Y. W.; Timmons, C.; Tram, T.; Coughlin, S. R. *Nature* **1997**, *386*, 502−506.
(104) Wells, C. M.; Di Cera, E. *Biochemistry* **1992**, *31*, 11721−11730.
(105) Dang, Q. D.; Vindigni, A.; Di Cera, E. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5977−5981.
(106) Dang, Q. D.; Guinto, E. R.; Di Cera, E. *Nat. Biotechnol.* **1997**, *15*, 146−149.
(107) Bode, W.; Turk, D.; Karshikov, A. *Protein Sci.* **1992**, *1*, 426−471.
(108) Martin, P. D.; Robertson, W.; Turk, D.; Huber, R.; Bode, W.; Edwards, B. F. P. *J. Biol. Chem.* **1992**, *267*, 7911−7920.
(109) Stubbs, M.; Oschkinat, H.; Mayr, I.; Huber, R.; Angliker, H.; Stone, S. R.; Bode, W. *Eur. J. Biochem.* **1992**, *206*, 187−195.
(110) Ayala, Y. M.; Di Cera, E. *J. Mol. Biol.* **1994**, *235*, 733−746.
(111) Guinto, E. R.; Di Cera, E. *Biophys. Chem.* **1997**, *64*, 103−109.
(112) Vindigni, A.; White, C. E.; Komives, E. A.; Di Cera, E. *Biochemistry* **1997**, *36*, 6674−6681.
(113) Rezaie, A. R. *Biochemistry* **1996**, *35*, 1918−1924.

(114) Guinto, E. R.; Vindigni, A.; Ayala, Y.; Dang, Q. D.; Di Cera, E. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11185−11189.
(115) Dang, Q. D.; Sabetta, M.; Di Cera, E. *J. Biol. Chem.* **1997**, *272*, 19649−19651.
(116) Di Cera, E.; Guinto, E. R.; Vindigni, A.; Dang, Q. D.; Ayala, Y. M.; Wuyi, M.; Tulinsky, A. *J. Biol. Chem.* **1995**, *270*, 22089−22092.
(117) Zhang, E.; Tulinsky, A. *Biophys. Chem.* **1997**, *63*, 185−200.
(118) Krem, M. M.; Di Cera, E. *Proteins: Struct., Funct., Genet.* **1998**, *30*, 34−42.
(119) Bartunik, H. D.; Summers, L. J.; Bartsch, H. H. *J. Mol. Biol.* **1989**, *210*, 813−828.
(120) Henriksen, R. A.; Dunham, C. K.; Miller, L. D.; Casey, J. T.; Menke, J. B.; Knupp, C. L.; Usala, S. J. *Blood* **1998**, *91*, 2026−2031.
(121) Miyata, T.; Aruga, R.; Umeyama, H.; Bezeaud, A.; Guillin, M. C.; Iwanaga, S. *Biochemistry* **1992**, *31*, 7457−7462.
(122) Gibbs, C. S.; Coutre, S. E.; Tsiang, M.; Li, W.-X.; Jain, A. K.; Dunn, K. E.; Law, V. S.; Mao, C. T.; Matsumura, S. Y.; Mejza, S. J.; Paborsky, L. R.; Leung, L. L. K. *Nature* **1995**, *378*, 413−416.
(123) Smith, G. P.; Petrenko, V. A. *Chem. Rev.* **1997**, *97*, 391−410.
(124) Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359−1472.
(125) Ni, F.; Ripoll, D. R.; Martin, P. D.; Edwards, B. F. P. *Biochemistry* **1992**, *31*, 11551−11557.
(126) Vindigni, A.; Di Cera, E. *Biochemistry* **1996**, *35*, 4417−4426.
(127) Rydel, T. J.; Tulinsky, A.; Bode, W.; Huber, R. *J. Mol. Biol.* **1991**, *221*, 583−601.
(128) De Filippis, V.; Vindigni, A.; Altichieri, L.; Fontana, A. *Biochemistry* **1995**, *34*, 9552−9564.
(129) De Filippis, V.; Quarzago, D.; Vindigni, A.; Di Cera, E.; Fontana, A. Submitted for publication.
(130) Banfield, D. K.; MacGillivray, R. T. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2779−2783.
(131) Lai, M.-T.; Di Cera, E.; Shafer, J. A. *J. Biol. Chem.* **1997**, *272*, 30275−30282.
(132) Esmon, C. T. *J. Biol. Chem.* **1989**, *264*, 4743−4746.
(133) Ye, J.; Esmon, N. L.; Esmon, C. T.; Johnson, A. E. *J. Biol. Chem.* **1991**, *266*, 23016−23021.
(134) Mathews, I. I.; Padmanabhan, K. P.; Tulinsky, A.; Sadler, J. E. *Biochemistry* **1994**, *33*, 13547−13552.
(135) Vijayalakshmi, J.; Padmanabhan, K. P.; Mann, K. G.; Tulinsky, A. *Protein Sci.* **1994**, *3*, 2254−2271.
(136) Sadler, J. E. *Thromb. Haemostasis* **1997**, *78*, 392−395.
(137) Richardson, M. A.; Gerlitz, B.; Grinnell B. W. *Nature* **1992**, *360*, 261−264.
(138) Grinnell, B. W.; Gerlitz, B.; Berg, D. T. *Biochem. J.* **1994**, *303*, 929−933.
(139) Gerlitz, B.; Grinnell, B. W. *J. Biol. Chem.* **1996**, *271*, 22285−22288.
(140) Wyman, J. *Adv. Protein Chem.* **1964**, *19*, 223−286.
(141) Smith, C. K.; Regan, L. *Science* **1995**, *270*, 980−982.

CR960135G