**Title**
Testing "native-state" assumptions in ligand binding and protein structure analysis



**Author**
Miller, David Warren

**Publication Date**
1997

Peer reviewed|Thesis/dissertation

# Testing "Native-State" Assumptions in Ligand Binding and Protein Structure Analysis

**by**

David Warren Miller

**DISSERTATION**

**Submitted in partial satisfaction of the requirements for the degree of**

**DOCTOR OF PHILOSOPHY**

**in**

Biophysics

**in the**

**GRADUATE DIVISION**

**of the**

**UNIVERSITY OF CALIFORNIA**

**San Francisco**

To my mother, father,

and my brother, Jonathan,

for your love, support, friendship,

humor, passion, and inspiration.

You are my role models.

# Acknowledgments

Thank you first and foremost to my advisor, Ken Dill. His infinite enthusiasm for my research, even during the low times, steered me again and again from what might have been debilitating frustration (and, perhaps more dramatically, an escape to medical school). Thank you to Peter Kollman, David Agard, and Tack Kuntz for much-needed scientific advice and criticism over the years, to Dennis Deen for his counseling and excellent career advice, and to Julie Ransom, an administrative miracle-worker whose support, encouragement, and wise purveyance of both sympathy and swift kicks in the booty, as appropriate, made her nothing less than "mom away from home."

Financial support for the work described in this thesis was provided by a Eugene Cota-Robles Fellowship and a National Science Foundation Fellowship, for which I am very grateful.

# Testing "Native-State" Assumptions in Ligand Binding and Protein Structure Analysis

David Warren Miller

August 1997

## Abstract

We use variations of the two-dimensional HP lattice model to test several common paradigms – used in both protein structure analysis and the characterization of various protein processes – that tend to overemphasize the influence of *native*, as opposed to *non-native*, conformations. This thesis has three parts. In the first, we present a model of amide hydrogen exchange (HX), and test the assumption that exchange in the stable limit occurs by only small perturbations of the native protein structure. We find that while this is sometimes true, exchange can also occur from conformations close *energetically*, but dissimilar *structurally*, to the native state. We show that this results from "bumpiness" of the HP model energy landscape, and that the "funnel" landscape assumption of current HX analyses – namely,

that native-like energies guarantee native-like structures – may lead to an incorrect interpretation of HX data that is unfairly biased towards the native structure. In the second part we model protein-ligand interactions. The model shows many familiar binding behaviors, but also several which violate standard premises of ligand binding. These include the *continuity* premise, that small changes to the ligand structure result in small changes to the bound complex; the *native* premise, that ligand binds tightly only to native-like conformations of proteins; and the *stability* premise, that specific, Michaelis-Menten binding always increases protein global stability. We present experimental evidence that supports our model results. In the third part we present a method for correcting free energy perturbation (FEP) calculations of relative binding energies so that they include contributions from "non-dominant" modes of binding, i.e., modes not observed by NMR or x-ray crystallography. We show an HP model example in which including only the dominant modes results in an incorrect rank ordering of the binding energies of two ligands, and demonstrate how a "multi-mode" approach using several FEP calculations recovers the correct rank ordering and gives a more accurate estimation of the relative binding energy.

# Table of Contents

# List of Tables

## Chapter 3

# List of Figures

## Chapter 1



## Chapter 2

# Introduction

Protein molecules can adopt a large number of conformations in solution, some of which must necessarily be very different from the native structure observed by x-ray crystallography or NMR spectroscopy. Although there are in principal many experimental techniques capable of detecting these non-native conformations, very few can do so under native conditions, because the *ensemble-averaged* structure, which most techniques measure, will be dominated by the single native conformation. Hydrogen exchange (HX), one of few techniques capable of assaying non-native structure even under native conditions, is nevertheless limited in the structural detail it provides, since the measured rates are influenced by local environment in a complex manner, and have contributions from many potentially diverse structures. Other techniques, such as CD spectroscopy, can probe non-native structures only when they represent a significant fraction of the overall population. And even under such conditions, these techniques are often limited by low resolution, providing only a measure of secondary structure, for example, or the local environment of a single fluorescent probe. A more detailed structural analysis of protein non-native states is therefore beyond the capacity of current biochemical and biophysical techniques.

A central goal of this work has been to study through simple models the effects of protein non-native states on experimental measurements, and to show that many common assumptions used to interpret these experiments are fundamentally biased by knowledge of the native, or ensemble-averaged, structure. We focus on two sets of assumptions. The first – used in structure-analysis studies using hydrogen exchange – are that HX rates can be used as a *direct probe* of native structure. Experiments are of two types: (1) The first is HX under *native* conditions, and involves the well-known debate over how solvent molecules are able to reach buried amide hydrogens in the protein interior. Several different mechanisms have been proposed, but an assumption inherent in each of them is that the pertinent fluctuations of

the protein must represent only *small perturbations* of the native structure. This assumption – evident in the names of the mechanisms themselves ("*local* unfolding," for example) – is used to infer protein structure and stability information from HX data. (2) The second type of HX experiment probes structure under *non-native* conditions, and includes kinetic studies of folding pathways, the characterization of stable intermediates, and the identification of folding subunits using curves of HX rate versus denaturant. The assumption made in each of these is that for any particular non-native conformation, the similarity or dissimilarity of HX rates to their native-state equivalents is a direct measure of the conformation's *structural* similarity or dissimilarity to the native state. This assumption is evident in the model "ribbon diagrams" drawn to represent the HX data from stable or folding-pathway intermediates: typically these have essentially native folds, with perhaps somewhat relaxed structure in the faster-exchanging regions.

The second set of assumptions pertain to protein-ligand interactions. We address four premises in which the current paradigms of ligand binding are heavily rooted: *(1) steric fit:* that binding is determined mainly by *shape complementarity*, *(2) native binding:* that ligands mainly bind to *native states*, with an accompanying increase in *global stability*, *(3) locality:* that ligands perturb protein structures mainly at the binding site, and *(4) continuity:* that small changes in ligand or protein structure lead to small changes in binding affinity. Underlying each of these assumptions is the idea that only the native structures of proteins determine where and how tightly a particular ligand will bind.

We examine, and ultimately question, the validity of these assumptions using a simplified model of proteins. Though not structurally realistic, the model is advantageous in that it allows model protein molecules to explore *all* of conformational space, including highly non-native structures, both folded and unfolded. Hence what

the model lacks in atomic detail, it gains in conformational diversity, an ingredient missing both from all-atom models and from the familiar two-state (native versus denatured) analyses of many protein processes. The model is described in detail below. Though not the primary theme, an important goal of this work is to demonstrate how simple models based on a few, basic, physical concepts can be extremely helpful in explaining unusual or perplexing behavior in complex systems.

This thesis is divided into three chapters. Chapter 1 describes a statistical mechanical model for amide hydrogen exchange (HX) in proteins. Unlike the two-state assumptions of other HX models, the HP model allows full enumeration of all non-native conformations, so we are able to determine exactly the types and magnitudes of protein fluctuations that give rise to exchange under varying conditions of protein stability. The model reproduces the familiar biphasic curves of HX rate versus external conditions (temperature or denaturant concentration), and we find that the two "regimes" of HX arise not from two separate mechanisms, but from a natural shifting of the non-native ensemble: from a compact state under stabilizing conditions, to an unfolded state under denaturing conditions. We find that HX can occur from protein conformations close *energetically* to the native state, but very distant *structurally*. This result has two implications. (1) It suggests that the HP model energy landscape, and perhaps that of proteins as well, is very "bumpy," since conformations with highly non-native structures can have very native-like energies. (2) It implies that current interpretations of HX data may be biased towards the known native structure, that is, by a "funnel" landscape assumption. For example, two prevalent assumptions in HX are (a) that fast HX rates measured under *native* conditions imply that only small perturbations of the protein take place during exchange, and (b) that under *denaturing* conditions, or during *folding*, the more that HX rates resemble the native-state rates, the more native-like the local structure

must be. Our model demonstrates that such assumptions may sometimes be incorrect, and that in general, inferring structural information from HX experiments requires additional knowledge about the protein energy landscape. We propose a hydrogen exchange experiment based on our model results which may help to reveal certain properties of these landscapes.

Chapter 2 describes a statistical mechanical model of protein-ligand interactions. Using the HP model we can go beyond the empirical methods of *binding polynomials*, and instead can predict binding behavior a priori, from the ligand structure and amino acid sequence of the protein. First we show how the model reproduces many familiar binding behaviors, such as "lock and key" and "induced fit" types of specific binding, characterized by Michaelis-Menten isotherms; ANS-like binding to compact-denatured states; binding and stabilizing a disordered structure, as heme stabilizes apomyoglobin; weak binding to the protein surface, as observed for various non-perturbing dyes and organic solvents; cooperative binding between two identical ligands; and ligand-induced denaturation. Next, we explore some interesting and unconventional behaviors of the model that contradict standard premises of ligand binding: (1) two ligands with identical shapes but slightly different energetics can bind in two entirely different binding modes. This result, supported by experiment, suggests both (a) that shape complementarity may not be as large a determinant of the binding mode as is often assumed, and (b) that small changes in ligand structure will not necessarily lead to only small changes in either the binding constant or the lowest-energy configuration of the complex. (2) Ligands can bind with native-like affinity to highly *non-native*, even *unfolded*, conformations of a protein, even when binding follows a normal Michaelis-Menten isotherm. This is important not only from a conceptual standpoint (i.e., the idea that tight binding doesn't always mean high population), but also from the standpoint of theoretical drug-design efforts,

since it suggests that modes of ligand binding other than the dominant x-ray mode may sometimes contribute significantly to the binding energetics. We show that these non-native binding interactions, as well as properties of the ligand-protein energy landscape in general, can be detected and quantified using hydrogen exchange, and we show experimental examples that support our model results. (3) Ligand binding can induce global *destabilization* in proteins, even when the interaction is specific and follows a Michaelis-Menten isotherm.

In chapter 3 we use the HP model of binding to examine the validity of an assumption commonly used in theoretical calculations of binding free energies, namely, that only a single dominant binding mode contributes to the energetics. Free energy perturbation (FEP) and other theoretical methods commonly base calculations on only the single mode of binding observed by x-ray crystallography or NMR. We explore this assumption with the HP model, using as a test case two structurally similar ligands, both of which have the identical dominant mode of binding, but one of which also binds in non-dominant modes. We find that assuming only the single dominant mode leads to large errors, and even an incorrect *rank ordering* of the predicted free energies of binding. We propose a method for incorporating non-dominant binding modes in FEP calculations, and for combining terms into a single, rigorous expression for the true relative binding free energy. We demonstrate this method for the HP model example, using Monte Carlo to simulate molecular dynamics (MD) FEP calculations. Using this method, we are able to predict accurate binding free energies for the two ligands, and to recover the correct rank ordering.

# Chapter 1

# A Statistical Mechanical Model for Hydrogen Exchange in Globular Proteins

This chapter is taken from the following published article: Miller DW, Dill KA. 1995. A statistical mechanical model for hydrogen exchange in globular proteins. *Prot Sci 4*: 1860–1873.

## ABSTRACT

We develop a statistical mechanical theory for the mechanism of hydrogen exchange in globular proteins. Using the HP lattice model, we explore how the solvent accessibilities of chain monomers vary as proteins fluctuate from their stable native conformations. The model explains why hydrogen exchange appears to involve two mechanisms under different conditions of protein stability: (i) a "global unfolding" mechanism by which all protons exchange at a similar rate, approaching that of the denatured protein, and (ii) a "stable-state" mechanism by which protons exchange at rates that can differ by many orders of magnitude. There has been some controversy about the stable-state mechanism: does exchange take place inside the protein by solvent penetration, or outside the protein by the local unfolding of a subregion? The present model indicates that the stable-state mechanism of exchange occurs through an ensemble of conformations, some of which may bear very little resemblance to the native structure. While most fluctuations are small-amplitude motions involving solvent penetration or local unfolding, other fluctuations can involve much larger transient excursions to completely different chain folds.

## BACKGROUND

Hydrogen exchange (HX) is a powerful tool in protein research, most notably in studies of structure, folding kinetics, folding intermediates, and ligand binding (Roder et al., 1988; Roder, 1989; Udgaonkar & Baldwin, 1988; Hughson et al., 1990; Englander et al., 1992). The chemical mechanism of exchange is well understood from experiments with extended, unstructured peptides highly solvated in solution (Berger & Linderstrom-Lang, 1957; Eigen, 1964). The reaction involves the acid-

or base-catalyzed exchange of backbone amide-group hydrogens with deuterium or tritium from bulk solvent. Studies of poly-DL-alanine and other polypeptides have yielded a quantitative description of the effects of temperature, pH, and solvent character on the exchange rates of solvent- exposed protons (Englander & Poulsen, 1969; Englander et al., 1972; Bai et al., 1993), and rules have been developed to calculate the sequence-specific inductive effects of neighboring amino acid side chains (Molday et al., 1972; Bai et al., 1993). These studies now make it possible to accurately predict the exchange rates of amide protons in extended oligopeptides (Molday et al., 1972; Bai et al., 1993) and in solvent-exposed, unstructured regions of native proteins (Englander & Staley, 1969; Molday et al., 1972; Yee et al., 1974).

In contrast, the mechanism of exchange in *folded* proteins is complicated by the inhomogeneous structural environments of core and surface protons. Factors such as steric inaccessibility to solvent, local charge distributions, and internal hydrogen bonding, which inhibits the chemical catalysis step, can slow the rate of amide proton exchange considerably. Nevertheless, experiments show that most protons, even those buried deep within the protein, exchange at measurable rates under physiological conditions (Hvidt & Linderstrom-Lang, 1954). How do catalysts from the bulk solvent gain access to the hydrophobic core? The mechanism has been studied by measuring amide proton HX rates for whole proteins and, using NMR, for individual protons, under varying degrees of protein stability (Woodward & Rosenberg, 1971; Woodward et al., 1975; Hilton & Woodward, 1979; Woodward & Hilton, 1980). Two such experiments are shown in Figure 1. Figure 1A shows an Arrhenius plot of the HX rate for several slowly-exchanging protons in the beta-sheet subregion of BPTI. Figure 1B shows the exchange rate as a function of denaturant concentration for several slowly-exchanging protons of cytochrome C. Both proteins are folded over the ranges of temperature and denaturant concentration shown. The results

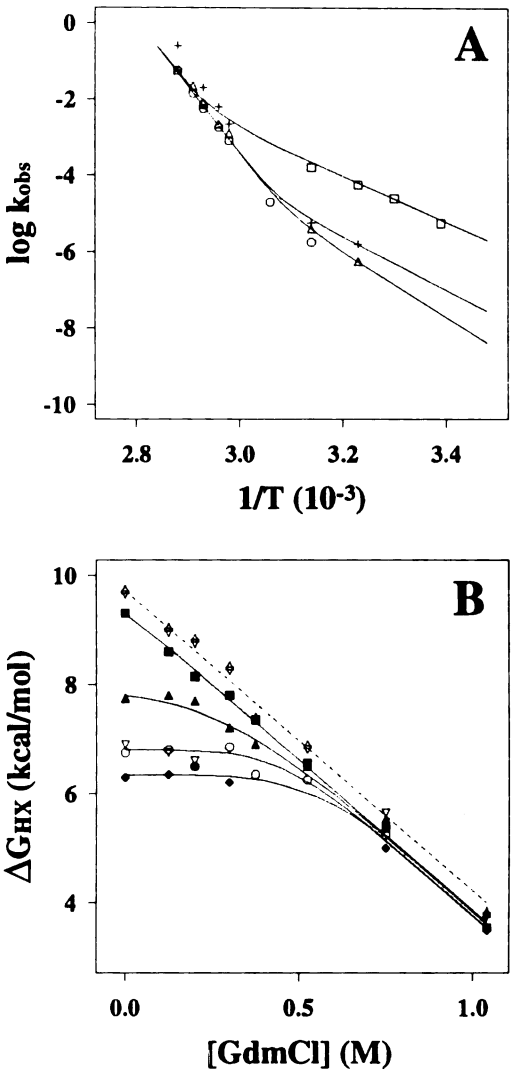


Figure 1: Experiments imply two exchange mechanisms. A: Temperature dependence of the normalized exchange-rate constants for beta-sheet protons of BPTI (adapted from Kim & Woodward, 1993). B: Denaturant dependence of exchange-rate constants for protons of Cyt C (adapted from Bai et al., 1994). In B, $\Delta G = -RT \log(k_{obs}/k_0)$. Under destabilizing conditions, all rates become equal and resemble those of the denatured state. Under stabilizing conditions, curve slopes are smaller and rates differ by orders of magnitude.

in Figure 1, which are typical of the slowly-exchanging protons of several proteins, have led to the following conclusions regarding the mechanism of hydrogen exchange:

(1) The negative slopes indicate that exchange slows exponentially as the stability of the protein is increased via the removal of denaturants or the lowering of temperature. This suggests that there are protons whose exchange from the native conformation would be immeasurably slow in the absence of fluctuations, implying that exchange takes place from conformations involving an opening of the native structure. The native state is said to be "exchange incompetent" for these slow-exchanging protons.

(2) The different limiting behaviors in the low- and high-stability regions suggest that two different types of fluctuations are responsible for exchange. In the high temperature (or high denaturant) limit, all of the buried protons exchange at roughly the same rate, suggesting that a common type of fluctuation governs exchange when protein stability is marginal. Conversely, at low temperature (or denaturant), the curves have uniformly smaller slopes, and exchange rates become strongly dependent on spatial position: within a given protein they can differ by more than eight orders of magnitude (Englander et al., 1972). This suggests that a different type of fluctuation may be responsible for exchange under highly stabilizing conditions. Clare Woodward and her colleagues were first to interpret the biphasic curves as suggesting two distinct exchange mechanisms (the "two-process model"; see Woodward & Rosenberg, 1971).

## Kinetic model of exchange and the two-process model

A simple kinetic model (Hvidt, 1964; Hvidt & Nielsen, 1966) has been helpful

in analyzing HX data from folded proteins under the various conditions of stability. The model involves two components: (i) a structural factor accounting for the rate at which the protein becomes "exchange competent" by losing any structural barriers to exchange, and (ii) a chemical factor accounting for the exchange rate of protons that are fully exposed and unstructured; these are known from reference data in well-characterized systems (see references above). According to this model, the exchange rate of a proton that is intrinsically incompetent to exchange from the native state is described by:

$$\text{HX incompetent} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{HX competent} \xrightarrow{k_0} \text{exchanged} \qquad (1)$$

where $k_{op}$ and $k_{cl}$ are the rate constants for opening and closing of the native conformation respectively, and $k_0$ is the rate constant for exchange from a solvent-exposed, unstructured peptide. Under steady-state conditions, the overall rate constant for (1) is:

$$k_{obs} = \frac{k_{op}\, k_0}{k_{cl} + k_0} \qquad (2)$$

Exchange generally follows one of two types of kinetics, referred to as EX1 and EX2 (Hvidt & Nielsen, 1966). Under EX2 conditions, $k_{cl} \gg k_0$, meaning that exchange-competent and incompetent structures reach a fast equilibrium prior to the slower exchange step. EX2 exchange is characterized by a linear dependence on the pH. Under EX1 conditions, $k_{cl} \ll k_0$. Here exchange depends only on $k_{op}$ and is pH independent. Most experimentally-measured rates show EX2 behavior (Segal & Harrington, 1967; Roder et al., 1985), for which the rate constant in Equation (2) reduces to:

$$k_{obs} = \frac{k_{op}}{k_{cl}} k_0 = K_{op} k_0 \quad (3)$$

where $K_{op} = k_{op}/k_{cl}$ describes the equilibrium between the native state and the exchange-competent state. Thus the exchange rate is the product of the extended-peptide rate multiplied by the fraction of time the protein is transiently unfolded.

Measurement of the ratio of exchange rates, $k_{obs}/k_0$, which is the reciprocal of the protection factor, has been used to determine $K_{op}$, the stability and relative population of the exchange-competent conformations, through Equation (3). Additional information is gained from the Arrhenius plot slopes, which are obtained by expressing Equation (3) in terms of energies:

$$k_{obs} = C_1 e^{-\Delta G_{op}/kT} e^{-E^{\ddagger}/kT} \quad (4)$$

where $\Delta G_{op}$ is the free energy of opening the native-state, $E^{\ddagger}$ is the activation energy for exchange from a solvent- exposed, unstructured peptide, and $C_1$ is a constant. The activation energy, $E^{\ddagger}$, of $k_0$ is observed to be temperature independent (Englander et al., 1979), and the free energy, $\Delta G_{op} = \Delta H_{op} - T\Delta S_{op}$, is often assumed to have temperature-independent values of $\Delta H_{op}$ and $\Delta S_{op}$. Thus the slope, $m$, of the Arrhenius plot for a given proton is obtained by differentiating log $k_{obs}$ with respect to $1/T$:

$$m = -C_2 (\Delta H_{op} + E^{\ddagger}) \quad (5)$$

where $C_2$ is assumed to be temperature independent. Hence the slope is proportional to the enthalpy difference between the native state and the most highly populated exchange-competent structure. Comparisons of these rate-based enthalpies with experimentally-measured unfolding enthalpies have been used to infer the types and

magnitudes of the fluctuations responsible for exchange.

On this basis, it has been possible to provide a structural interpretation for the two different exchange mechanisms proposed by Woodward et al. to occur in the limits of high and low stability (e.g. Ellis et al., 1975; Woodward et al., 1982; Englander & Kallenbach, 1984). (1) Under destabilizing conditions, it has been proposed that exchange occurs by "global unfolding". That is, since the rate-based enthalpies coincide with equilibrium unfolding enthalpies (Woodward et al, 1975), it has been concluded that the exchange-competent states are the denatured conformations. This is consistent with the observation that all protons exchange at roughly the same rate under these conditions. (2) Under stabilizing conditions (low temperature or low denaturant), exchange rates are slower, and rate-based enthalpies are smaller than for global unfolding. This has been taken as evidence that the conformational fluctuations from the native state are correspondingly smaller, and HX sites more protected, than in denatured states. In this limit, protection is highly site-specific, and HX rates vary with the amide proton environment. This second mechanism of exchange has alternately been called "exchange from the folded state" (Woodward & Rosenberg, 1971), the "low activation energy" process (Woodward et al., 1975), and "process b" (Hilton & Woodward, 1979). We refer to it as the "stable-state" mechanism.

### Models for the stable-state mechanism

Two types of structural models have been developed to describe the stable- state mechanism: "solvent penetration" models (e.g. Woodward & Rosenberg, 1971; Lumry & Rosenberg, 1975; Richards, 1979) and the "local unfolding" model (Hvidt & Nielsen, 1966; Englander, 1975). The models are reviewed fully in Englander & Englander, 1978; Gurd & Rothgeb, 1979; Hilton et al., 1981; Woodward et al., 1982;

and Englander & Kallenbach, 1984, and are discussed briefly below.

### A: Solvent penetration models

In penetration models, protons exchange within the protein interior: catalysts enter the protein core through transiently-formed channels and cavities. One penetration model proposes that such channels arise from small, rapid fluctuations of tenths to several angstroms of interior atoms. Other penetration models have proposed that the channels arise from the redistribution of interior hydrogen bonds (Lumry & Rosenberg, 1975; Nakanishi et al., 1973), or from the random association of pre-existing interior cavities (Richards, 1979). In penetration models, the exchange rate of a proton depends both on its average accessibility to solvent, which is a function of its depth of burial and the local mobility of the protein in its vicinity, and on its reactivity with the catalyst, which is affected by local structural features such as packing density and hydrogen bonding. Some penetration models postulate that exchange is highly localized; i.e., protons on adjacent amino acids exchange independently rather than in a correlated way.

### B: Local unfolding models

In the local unfolding model, exchange occurs *outside* the protein when a subregion, such as an alpha-helix, transiently unfolds into the bulk solvent. Protein segments of no more that about ten residues are expected to unfold at one time (the unfolding region is "local"), consistent with the small enthalpies observed for exchange under stable conditions. In this model, the main barrier to exchange is considered to be the interior hydrogen-bonding of amide protons, rather than their depth of burial. Adjacent protons within subregions are predicted to exchange at roughly the same rate, since protons belonging to the same structural unit are un-

folded simultaneously. Such correlated exchange behavior has sometimes been called "cooperativity."

### Experimental evidence

Evidence that has been taken as support for solvent penetration models includes the following:

(a) The magnitudes of fluctuations inferred from X-ray crystallography and NMR spectroscopy under native conditions are thought to be large enough to form the solvent channels envisioned by the penetration model (Wuthrich & Wagner, 1978; Frauenfelder et al., 1979).

(b) Molecular dynamics (MD) calculations suggest that small, picosecond atomic fluctuations provide penetration pathways for $O_2$ in myoglobin (Karplus & McCammon, 1981).

(c) Proton NMR studies in BPTI show that exchange of water molecules between the protein core and bulk solvent can occur on a sub-second time scale under physiological conditions (Tuchsen et al., 1987; Otting et al., 1991).

(d) The pKa's of two histidine residues in ribonuclease S during C-2 proton exchange are more characteristic of a buried than exposed environment (Bradbury et al., 1977, 1980).

(e) Denaturants at low concentrations do not accelerate stable-mechanism HX rates, implying that regional unfolding is not required for exchange (Woodward et al., 1975;

Woodward & Hilton, 1979; Hilton et al., 1981; Mayo & Baldwin, 1993).

(f) Roughly 95% of amide protons in myoglobin crystals exchange with deuterium after perfusion with deuterium solvent (Schoenborn et al., 1978), suggesting that exchange does not require unfolding.

(g) The exchange kinetics in lysozyme is identical in solution and in the crystal, where it is presumed that local unfolding should not occur (Tuchsen & Ottesen, 1979; Tuchsen et al., 1980).

(h) Protons at the interface between trypsin and trypsin inhibitor exchange without dimer dissociation, implying an absence of local unfolding, and the pH and temperature dependence of exchange is characteristic of buried protons in globular proteins (Pershina & Hvidt, 1974; Woodward, 1977).

(i) For the slowest exchanging amides in BPTI, next-neighbor residues can have very different exchange rates, implying a lack of correlation (Hilton & Woodward, 1979).

(j) For the fifteen slow-exchanging amides in BPTI, there is no obvious correlation between H-bond length and the rates of exchange by the stable- state mechanism (Levitt, 1981a; Woodward et al., 1982).

(k) In lysozyme (Radford et al., 1992; Pedersen et al., 1993) and leucine zipper peptides (Goodman & Kim, 1991) there are helices in which buried protons exchange more slowly than solvent-exposed protons.

(l) Studies of the effects of local structure on tyrosine ring flips in folded proteins suggest that the formation of solvent channels, as envisioned by the penetration model, could give rise to the large range of enthalpies observed for the stable-state mechanism of exchange (Karplus & McCammon, 1981).

Evidence taken as support for the local unfolding model includes the following:

(a) In support of neighbor correlations, like-exchanging sets of protons in hemoglobin undergo uniform rate shifts upon the allosteric structural change from deoxy-Hb to oxy-Hb (Englander & Mauel, 1972; Englander, 1975) and these sets are localized within regions of secondary structure (Englander et al., 1983).

(b) As an indication that HX rates depend not only on the degree of burial, but also on hydrogen bonding, small molecules that form intramolecular hydrogen bonds, such as salicylate, exchange up to six orders of magnitude slower than their diffusion-limited rates, even though they have otherwise good solvent accessibility (Haslam & Eyring, 1967; Rose & Stuehr, 1968).

(c) For intermediate and fast (non-beta-sheet) protons in BPTI, trypsin, and ribonuclease, there is no correlation between crystallographic B factors and exchange rates, suggesting that small internal fluctuations are not the primary mechanism of exchange (Levitt, 1981b; Kossiakoff, 1982; Wlodawer & Sjolin, 1982).

(d) There is no obvious correlation between exchange rate and depth of burial in trypsin and ribonuclease (Kossiakoff, 1982; Wlodawer & Sjolin, 1982).

(e) Englander and Kallenbach (1984) have argued that the free energies would be too high for the catalytic $OH^-$ ions to penetrate a protein core.

(f) The observed shift in $pH_{min}$ for exchange by the stable-state mechanism is argued to be caused by the separation of interior charged groups, which requires a structural change greater in magnitude than the small fluctuations envisioned by the penetration model (Englander & Kallenbach, 1984).

(g) In alpha-helical segments of ribonuclease S (Kuwajima & Baldwin, 1983), BPTI (Wagner & Wuthrich, 1982), and apamin (Wemmer & Kallenbach, 1983), amides exchange at similar rates despite very different solvent accessibilities in the native protein.

(h) Amides in alpha-helical and beta-sheet regions of both synthetic polypeptides and native proteins exchange slowly despite being solvent accessible (Leichtling & Klotz, 1966; Ikegami & Kono, 1967; Nakanishi et al., 1972; Welch & Fasman, 1974; Rohl et al., 1992; Bai et al., 1994).

Both penetration and local unfolding models view the exchange competent conformations as being small excursions from the native state since the rate-based enthalpies are smaller than those for unfolding the protein. The main distinctions between the two mechanisms are: (i) whether the exchange reaction is *external* in bulk solvent (local unfolding model) or *internal* within the protein interior (solvent penetration model), and (ii) whether exchange involves neighbor correlations or whether neighboring residues can exchange independently. Neither model attempts

to make quantitative predictions for individual proton exchange rates.

Our aim here is to avoid making assumptions about the various mechanisms of exchange and to see instead if a unified viewpoint can be derived from an underlying statistical mechanical theory of how amino acids develop access to solvent through conformational fluctuations from native states. Using a simple lattice representation of proteins, we model the non-chemical part of the hydrogen exchange process and explore the mechanism by which buried monomers gain access to solvent under varying conditions of chain stability. Our model involves two components: a general statistical mechanical treatment of monomer accessibility, and a simple lattice model to relate structure to accessibility. While the lattice representation of structure is very simple, our statistical mechanical treatment is more complete and has fewer approximations than previous models. We make no assumptions about mechanisms of fluctuations, about whether the equilibrium is 2-state or 3- state, about the rates of exchange from compact or unfolded conformations, or about how different conformations are affected by external conditions. Rather, these properties are derived from the theory.

We find that the two limiting mechanisms of exchange, under stable and unstable conditions respectively, arise as a natural consequence of different limiting behaviors of native fluctuations. Exchange by the stable- state mechanism is found to resemble both solvent penetration and local unfolding models in some respects, but there are also novel aspects, namely that: (i) fluctuations are usually ensembles of many different conformations, and (ii) the exchange competent conformations under stable conditions can occasionally involve large conformational deviations from the native structure. We explain these results in terms of the shape of the folding energy landscape, and discuss possible implications for biological proteins.

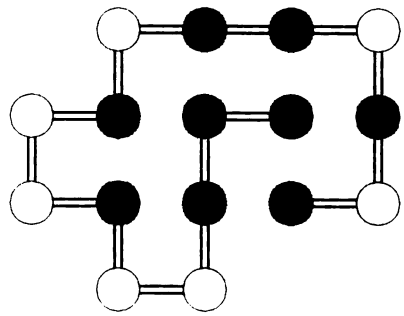

Figure 2: Sample conformation for the two-dimensional HP lattice model. Black and white beads represent H (hydrophobic) and P (polar) monomers, respectively. Non-covalent HH contacts are favored by a free energy, $\epsilon$ ($\epsilon < 0$). This conformation has 8 HH contacts, so the free energy is $8\epsilon$.

## THE LATTICE MODEL

We model proteins using the 2-dimensional HP lattice model (Lau & Dill, 1989, 1990; Chan & Dill, 1991; Dill et al., 1995). A protein is represented by a sequence of H (hydrophobic) and P (other) monomers configured on a two-dimensional lattice (Figure 2). Amino acids are represented by beads, and the protein backbone by lines connecting the beads. The background lattice serves to divide space into amino-acid- sized units. Lattice sites may be either empty or filled by a single bead (the excluded-volume constraint), and empty lattice sites are assumed to contain a solvent molecule. In the HP model, each HH contact, formed when two non-sequential H monomers occupy adjacent lattice sites, is favored by a contact free energy $\epsilon$ ($\epsilon < 0$). This energy is a simple way of capturing the importance of hydrophobic interactions, whatever their underlying atomic origins in hydrogen bonding, van der Waals interactions, and solvent ordering (Dill, 1990; Chan & Dill, 1991). HP and PP contacts are neither favored nor disfavored energetically. Hence the free energy of a conformation is $h\epsilon$, where $h$ is the number of HH contacts. The conformation in

Figure 2 has 8 HH contacts, and therefore has a free energy of $8\epsilon$. The magnitude of $\epsilon$ determines the degree of stability imparted by external conditions; large and negative $\epsilon$ is more stabilizing. Conformational entropy, the driving force for unfolding, enters the model through the exhaustive enumeration of all the possible chain configurations (see discussion below).

The disadvantages of the model are clear: atomic resolution is lost; conformations are restricted to a lattice; it is in 2 dimensions; the energy function is simplified; and chains are unrealistically short. Yet despite these disadvantages, the model has been found useful for modeling protein properties (Lau & Dill, 1989, 1990; Chan & Dill, 1989, 1990, 1994; Shortle et al., 1991; Dill et al., 1995) because it shows several protein-like features, including cooperative collapse, a nonpolar core, secondary and tertiary structures, and multi-stage folding kinetics. And because it characterizes the full conformational space in an exact and unbiased way, we are able to consider all fluctuations of all sizes, without being restricted to small- amplitude motions, short time scales, or mathematical approximations. Most importantly, we believe the model captures the main physics of protein folding—the hydrophobic interactions, conformational freedom of the chain, and the steric restrictions imposed by excluded volume.

This model does not include explicit hydrogen bonding, but we believe this is not a serious limitation. There are three possible roles of hydrogen bonding in hydrogen exchange. First, hydrogen bonding is required to catalyze the exchange reaction, but as in past literature on exchange mechanisms, we calculate ratios (observed rates divided by rates for solvent-exposed, unstructured sites), and thus avoid the need to consider the chemical details of the exchange process. Second, to the extent that hydrogen bonding contributes to protein stability, it is an implicit component of the HH, HP, and PP interactions of the HP model. Because the model produces

hydrophobic cores, and monomer burial, we consider it a simplest model of how the accessibilities of amino acids in proteins depend on stability. Third, inter-residue hydrogen bonds present a structural barrier to exchange because they must be broken before the exchange process can be initiated. The breaking of these bonds can also give rise to correlated exchange, as in the local unfolding model. In the present model the proximity of an amino acid to solvent determines its ability to exchange, as described below. In this regard, all microscopic barriers to exchange, including hydrogen bonding, are treated as a single factor implicitly accounted for by monomer burial. However, neighbor correlations can occur naturally in a sequence-specific way without introducing the additional energy parameter a hydrogen bond would require.

We study HP sequences of 16 monomers. For any 16-mer chain there are exactly 802,075 possible conformations that can be configured on a two- dimensional lattice. These conformations are generated by computer, and each is weighted by a Boltzmann factor according to the number of HH contacts made. Figure 3 shows an energy diagram for a sample sequence. The native structure (ground state) is the conformation with the largest possible number of HH contacts, and thus the lowest contact free energy. We study only those sequences ("non-degenerate")that have a single native conformation, since we believe that they best represent biological proteins, which fold to unique structures. All higher-energy conformations comprise the "denatured" states, and are grouped by energy into "first-excited" states, "second-excited" states, etc., corresponding to successively fewer HH contacts (see Figure 3). Figure 3C shows the density of states, $g(h)$, the distribution of conformations having $h$ HH contacts. For all HP sequences, $g(h)$ diminishes rapidly as $h$ increases. That is, there are far more open, high-energy conformations than compact, low-energy conformations.

Our treatment of hydrogen exchange is divided into two parts. First we develop

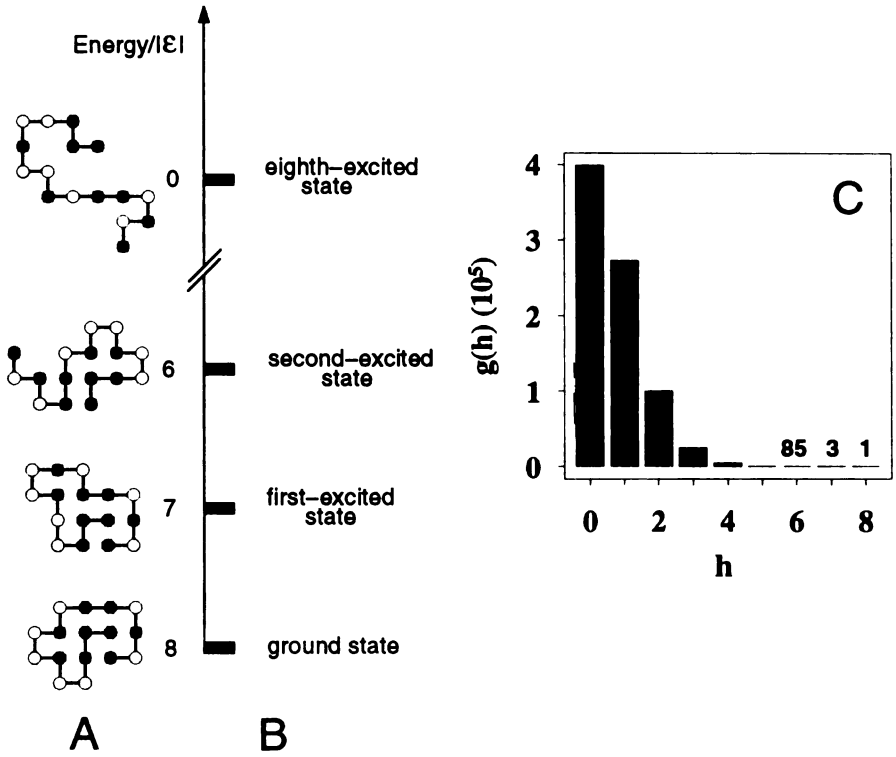


Figure 3: The "energy ladder" diagram for the 16-mer HP sequence in Figure 2. The native conformation is at the bottom of the ladder; this single lowest-energy conformation has 8 HH contacts. Increasing steps up the energy ladder represent successively fewer HH contacts; these levels are the first-excited states, second-excited, etc. Sample conformations of each energy level are shown in A. C: Density of states distribution, $g(h)$, the number of conformations of this HP sequence that have the given number, $h$, of HH contacts.

a statistical mechanical theory to describe how monomer sites that are exchange incompetent in the native state become competent to exchange through Boltzmann populations of non-native conformations. This theory is completely general and does not depend on lattices or on any other particular model. Second, we introduce the HP model to give a concrete basis for relating protein structure and stability to exchange competence.

### Statistical mechanical model of exchange competence

At a temperature $T$, any conformation, $c$, of a protein will be populated according to its Boltzmann probability, $P_c$:

$$P_c = \frac{e^{-E_c/kT}}{\sum_{c=1}^{N} e^{-E_c/kT}} \tag{6}$$

where $E_c$ is the free energy of conformation $c$, $k$ is Boltzmann's constant, and $N$ is the number of conformations available to the protein. Equation (6) is a general, model-independent equation for any equilibrium system. At low temperatures, only the native conformation will be populated since its Boltzmann factor will be dominant. At high temperatures, the denatured conformations will be populated because there are so many more non-native than native conformations, and the entropically-favored states will dominate.

The average value of any observable conformational property $A_c$ can be computed as a sum over Boltzmann probabilities:

$$\langle A \rangle = \sum_{c=1}^{N} A_c P_c = \frac{\sum_{c=1}^{N} A_c \, e^{-E_c/kT}}{\sum_{c=1}^{N} e^{-E_c/kT}} \tag{7}$$

In particular, the observable property of interest here is the exchange competence of

an individual proton site. If $A_{c,i}$ characterizes the exchange competence of proton $i$ in conformation $c$, Equation (7) gives the average exchange-competence of that proton as a function of temperature.

Equation (7) is completely general and model independent. Now we introduce the HP model. The conformational energy depends only on the number of HH contacts: $E_c = h\epsilon$, where $h$ is the number of HH contacts in conformation $c$, and $\epsilon$ is the HH contact free energy. Written in terms of $h$, Equation (7) becomes (see appendix):

$$\langle A \rangle_i = \frac{\sum_{h=0}^{h_N} A_{h,i}\, g(h)\, e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} g(h)\, e^{-h\epsilon/kT}} \tag{8}$$

where $h_N$ is the number of HH contacts in the native conformation, $g(h)$ is the density of states, and $A_{h,i}$ is the average value of $A_{c,i}$ for the $g(h)$ conformations having $h$ HH contacts.

In the present model, the exchange competence of a monomer is given by its accessibility to solvent: we define $A_{c,i} = 0$ if monomer $i$ is completely surrounded by four other monomers in conformation $c$, and $A_{c,i} = 1$ if monomer $i$ is adjacent to a lattice site containing a solvent molecule. The model is illustrated in Figure 4, which shows the native and three first- excited states of the sequence in Figure 3. The five monomers buried in the native state are exchange incompetent, and four of them become competent to exchange in at least one of the first-excited states. In the present model, buried monomers do not exchange, while those adjacent to a solvent- occupied lattice site exchange at a generalized rate, $k_0$, corresponding to a solvated, unstructured peptide. Therefore the exchange rate of any monomer, $i$, is:

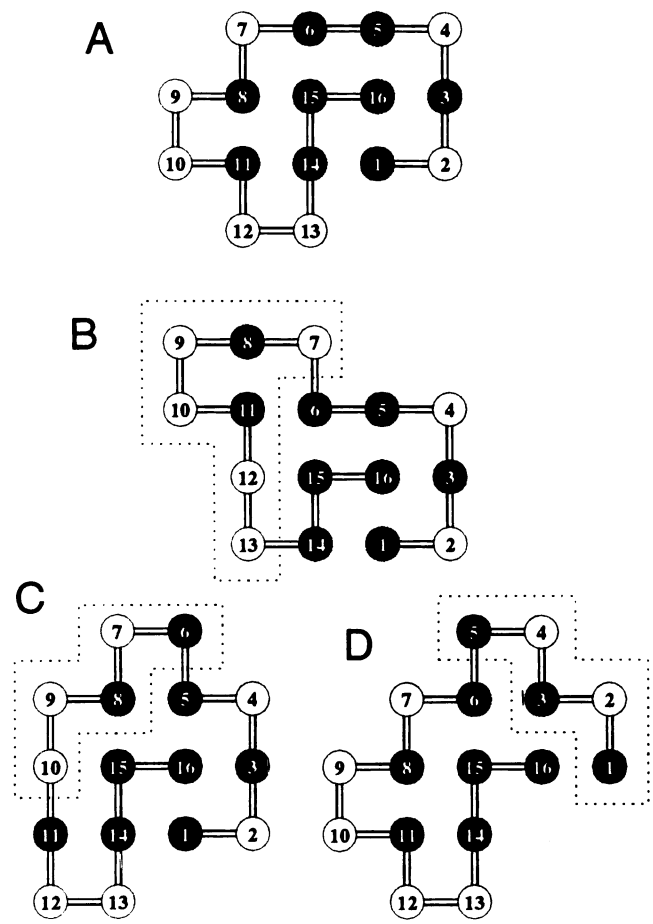


Figure 4: Native and fluctuational conformations. A: Native state of the sequence in Figure 3. Monomers 8, 11, 14, 15, and 16 are each surrounded by four other monomers and so are "buried," and hence exchange incompetent, in the native state. B–D: The three first-excited states. Monomers 8, 11, 14, and 16 become exchange competent in the first-excited states because each of those monomers makes a contact with a solvent lattice site in at least one of the three conformations. Monomer 15 cannot exchange in any of the three first excited states, but can exchange in the second-excited states (not shown). Boxed regions represent the smallest section of chain that must reconfigure to reach the native state. Dissimilarities (see text) between the first-excited states in B–D and the native state are 0.275, 0.274, and 0.219, respectively.

$$k_i = \langle A \rangle_i \, k_{0,i} \tag{9}$$

where $k_{0,i}$ is the extended-peptide exchange rate for monomer $i$. Equation (9), which can be determined by exact enumeration for monomers of any short HP sequence, is analogous to Equation (3) of the Hvidt model for EX2-type exchange, and gives the exchange rate in terms of average monomer solvent accessibility.

Since $\langle A \rangle_i$ of Equation (8) depends on $\epsilon$, monomer exchange rates can be determined under various conditions of chain stability. We use the single parameter $|\epsilon/kT|$ as a measure of the degree of stabilization due to external conditions. When $|\epsilon/kT|$ is large (corresponding to strongly stabilizing conditions such as low temperatures or denaturants), only the native state is populated to any appreciable degree, so the accessibilities of monomers, and their corresponding exchange rates, can be determined directly from the native structure. In the other limit, when $|\epsilon/kT|$ is small (high temperature or denaturants), most molecules are denatured, and the average solvent accessibility and exchange rate of a monomer will reflect the distribution of accessibilities in the ensemble of denatured conformations. Thus for a monomer buried in the native state, average solvent accessibility and the exchange rate increase as $|\epsilon/kT|$ becomes smaller, because the Boltzmann populations of non-native exchange-competent states increase.

Below we show results of calculations of the ratio $k_i/k_{0,i}$, the reciprocal of the protection factor, as a function of $|\epsilon/kT|$, for monomers of the sample 16-mer sequence shown in Figures 3 and 4.

## RESULTS

### Model exchange curves

Figure 5 shows the predicted exchange rate, $k_i/k_{0,i}$ versus $|\epsilon/kT|$ for the five monomers buried in the native conformation of Figure 4A. These five monomers are exchange incompetent in the native state, so this calculation corresponds to the experiments shown in Figure 1. Since the parameter $|\epsilon/kT|$ is equivalent to $1/T$ for fixed $\epsilon$, the plot in Figure 5A is analogous to Arrhenius plots of the type shown in Figure 1A. Figure 5B shows the same curves with both axes reversed, so that it corresponds to the denaturant plot of Figure 1B. The theoretical results in Figure 5 show the same general features as the experimental results in Figure 1, with exchange rates that converge to a single value under destabilizing conditions and diverge under stabilizing conditions, and a transition region suggestive of the two limiting exchange mechanisms proposed by Woodward.

The basis for these results is illustrated in Figure 6, which shows how the populations of different conformations of the 16-mer change with temperature. Under destabilizing conditions (6A and B), the chain populates an ensemble of conformations having large energies and correspondingly unfolded structures. Exchange under these conditions occurs from the denatured states, i.e., by global unfolding. Since solvent exposure is high in these unfolded conformations, exchange rates are similar and converge toward the maximum rate as native stability is decreased. Conversely, under stabilizing conditions (6C and D), only the lowest-energy excited states contribute to the fluctuations. Exchange of buried monomers therefore occurs from a restricted set of conformations with native-like energy and compactness. For monomers 8, 11, 14, and 16, stable-state exchange is from the first-excited states, since exchange competence is achieved in these conformations (see Figure 4). For

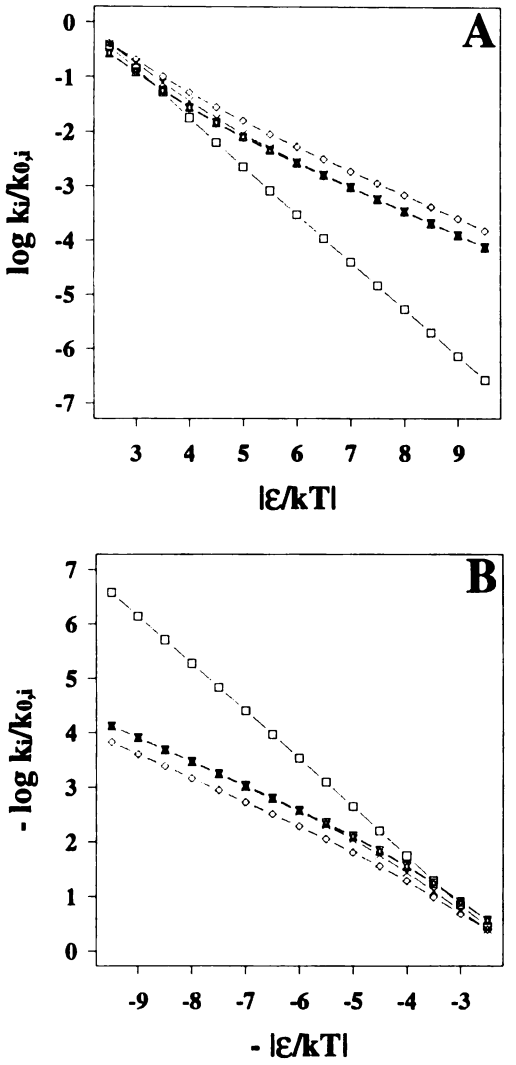


Figure 5: HX curves computed from the model, vs. HH-contact energy, $\epsilon$, for the five buried monomers of Figure 4A. Curves correspond to monomers 8($\bigtriangledown$), 11($\bigtriangleup$), 14($\diamond$), 15($\square$), and 16($\times$). A: Analog of the Arrhenius plots of Figure 1A. B: Analog of the denaturant curves of Figure 1B. Both figures show converging exchange rates under destabilizing conditions, diverging rates under stable conditions, and a transition region suggesting two exchange mechanisms.

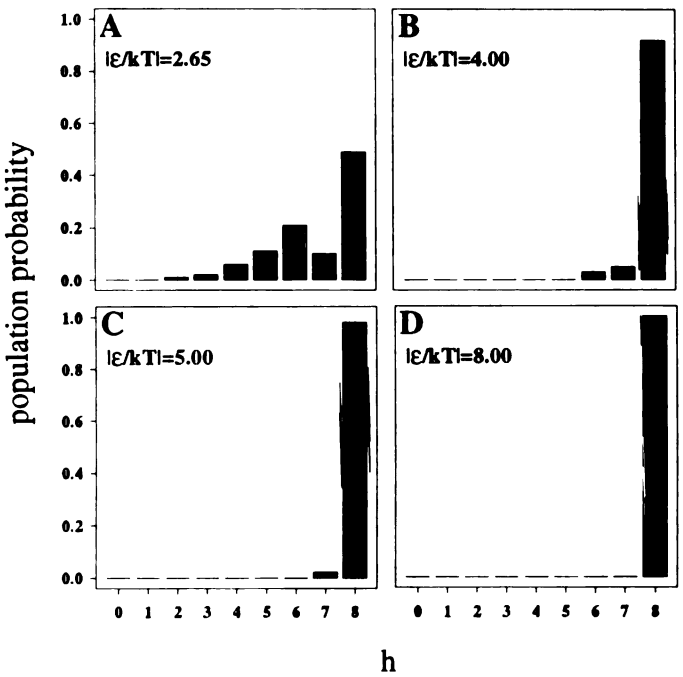


Figure 6: Populations of conformations. A–D: populations at $|\epsilon/kT| = 2.65$, 4.00, 5.00, and 8.00, respectively. The melting temperature, where the native state is 50% populated, is $|\epsilon/kT| = 2.65$.

monomer 15, which is buried in all three first-excited states, exchange first occurs only when the chain fluctuates to the second-excited states.

The two apparent exchange mechanisms in Figures 1 and 5 are explained as a shift in the class of conformations from which exchange occurs: from the broad set of unfolded conformations to a much smaller set of highly compact conformations. This result is consistent with the "global unfolding" and "stable state" mechanisms of Woodward's two-process model, and supports the belief that the shift in mechanisms results from changes in protein structure rather than from complex exchange chemistry or a shift from EX2- to EX1-type exchange. The slopes of the curves are obtained from Equation (8) (see appendix), and are proportional to the energy required for a fluctuation from the native state to the most highly-populated exchange-competent conformation: $m(\epsilon)_i = \langle h \rangle_{acc,i} - h_N$, where $\langle h \rangle_{acc,i}$ is the weighted average of $h$ for all conformations in which monomer $i$ is solvent accessible. Under destabilizing conditions, the most highly populated exchange-competent conformations are unfolded, and are the $h = 6$ conformations (Figure 6A) for the sample 16-mer of Figure 4. Hence $\langle h \rangle_{acc,i} = 6$ in the limit of small values of $|\epsilon/kT|$, and the limiting slope is -2. Under these conditions monomer exchange rates are similar and approach the maximum value, $k_{0,i}$, as in the global unfolding model. Conversely, under stable conditions only the exchange conformations of lowest energy contribute to exchange, so $\langle h \rangle_{acc,i}$ is equal to $h_i^*$, the number of HH contacts in the lowest-energy exchange conformations of monomer $i$ (see appendix). For monomers 8, 10, 14, and 16, which can exchange from the first-excited states (Figure 4), $h_i^* = 7$, and the limiting slope for large $|\epsilon/kT|$ is -1. For monomer 15, $h_i^* = 6$, so the limiting slope is -2. The diversity of monomer exchange rates in this stable limit arises from differences in the energy required for the minimum fluctuation to an exchange-competent structure. This result is evident from Equation (8) in the limit of large $|\epsilon/kT|$, where the

average solvent accessibility becomes:

$$\langle A \rangle_i \longrightarrow A_{h_i^*,i} \frac{g(h_i^*)}{g(h_N)} e^{-(h_i^* - h_N)\epsilon/kT} \tag{10}$$

Equation (10) is a single exponential, so the exchange curve of a given monomer is a straight line with slope $m(\epsilon)_i = h_i^* - h_N$. Monomer 15 requires greater fluctuational energy to exchange ($2\epsilon$ vs. $\epsilon$), so as stability is increased its exchange rate decreases more rapidly than the other buried monomers.

The approximate linearity of the curves in the two regions arises from the temperature independence of $\langle h \rangle_{acc,i}$ in the limits of large and small $|\epsilon/kT|$. For large $|\epsilon/kT|$, this result follows directly from equation (10). For small $|\epsilon/kT|$, conformations from many energy levels contribute to exchange, so a constant slope is not necessarily expected. Whether or not conformations of a single energy level will dominate exchange, as is the case with the sequence shown here, depends on the density of states, $g(h)$, which is highly sequence specific. The density of states also affects the curve shapes.

Although the model denaturant curves in Figure 5B do not become as flat as the experimental curves of Figure 1B, this is undoubtedly due to the relatively large size of the energy spacing in our short-chain model compared to real proteins, which have many more monomer contacts. Short chains also restrict the possible range of stable-limit HX rates among monomers. The extent of rate diversity observed in experiments (Figures 1A and B) would be obtained in our model if more energy levels were available, as is the case for real proteins.

### Structures of exchange-competent states

Figures 4B–D show the three first-excited states of the given 16-mer sequence. Under stable conditions, these are the predominant exchange-competent conforma-

tions for four of the five buried monomers. The structural fluctuations from the native state represented by these conformations are consistent with both solvent penetration and local unfolding models. In Figure 4B for example, monomer 8 has become solvent accessible by a small shift in its local subregion (monomers 7–13) that has forced monomer 8 out into the solvent, much like a local unfolding event. And in Figure 4D, monomers 14 and 16 have both become accessible by the movement of a sequentially-distant chain region (monomers 1–5), resulting in the exposure of both monomers to bulk solvent. This latter example could be considered consistent with solvent penetration, since monomers 14 and 16 are part of an intact core of residues that become exposed by removal of some chain surface. In our model examples, the chains are too short to have enough "core" for us to interpret either penetration or local unfolding unambiguously. Nevertheless, from this and other examples, we believe this model supports the general types of protein motions characteristic of both solvent penetration and local unfolding mechanisms. As for neighbor correlations, the sequence in Figure 3 shows independent, rather than correlated exchange; monomers 14, 15, and 16 form a buried subregion in the native conformations, but monomer 15 exchanges at much slower rate. In general, both correlated and independent exchange are possible depending on the monomer sequence.

**"Conformational Distant Relatives"**

How different are the hydrogen-exchanging conformations compared to the native state in the stable limit? To answer this question, we must explore the relatedness of native and first-excited conformations. Figures 4B–D show examples of relatively small fluctuations in which the native chain fold and hydrophobic core are largely preserved in the first-excited conformations. At the other extreme, Figure 7 shows an example of a native state (7A) and one of its first excited-states (7B) that are

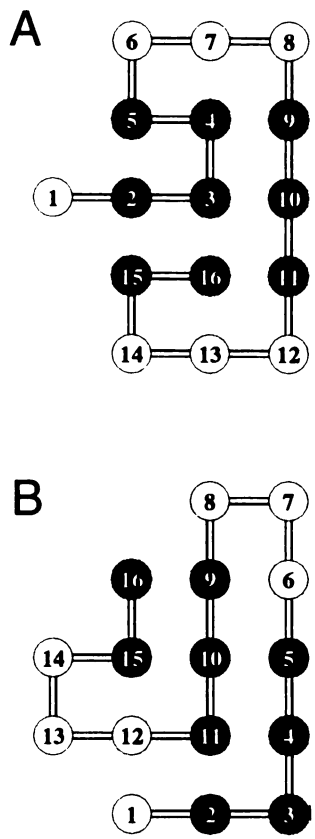


Figure 7: Example of conformational distant relatives. A: Sample 16-mer native state ($h = 6$). B: One of its first-excited states. No native contacts are conserved. The dissimilarity is 0.809.

completely different, sharing no common residue contacts. This is an example of a fluctuation in which a very *small* change in *energy* corresponds to a very *large* change in *structure.*

Are such large structural fluctuations rare or common? Figure 8A shows the average magnitude of structural change under stable-state conditions, obtained by comparing all possible 16-mer native states to their first- excited states and measuring structural differences between them. We use the dissimilarity measure of Yee & Dill (1993), which scores the conformational relatedness of two structures: 0 means two structures are identical. Figure 8A shows that fluctuations from native to first-excited states display a wide range of magnitudes. For comparison, Figure 8B shows

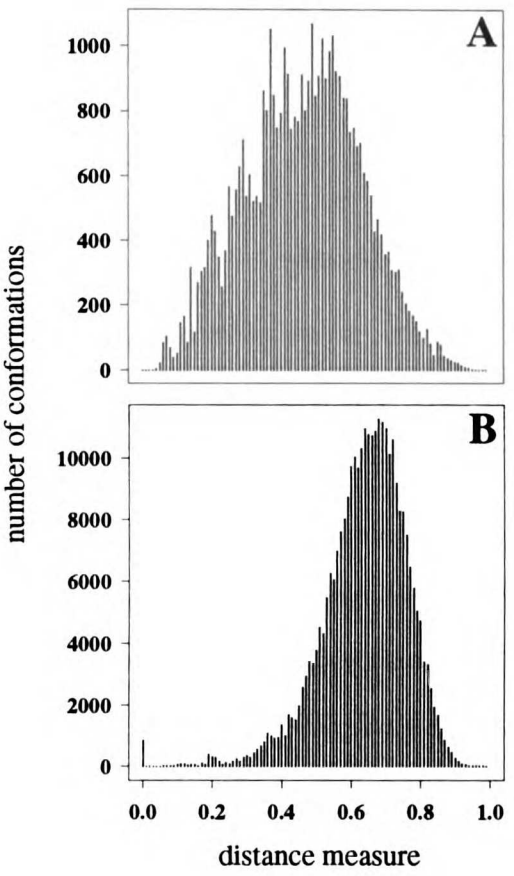


Figure 8: Dissimilarity distributions. A: Comparison of each non- degenerate 16-mer native state with each of its first-excited states. B: Pairwise comparison of all non-degenerate 16-mer native states. Horizontal axis shows the dissimilarity measure of Yee & Dill (1993). The average dissimilarities for A and B are 0.50 and 0.65 respectively. This figure shows that first-excited states can have a broad distribution of structures, including very non-native ones.

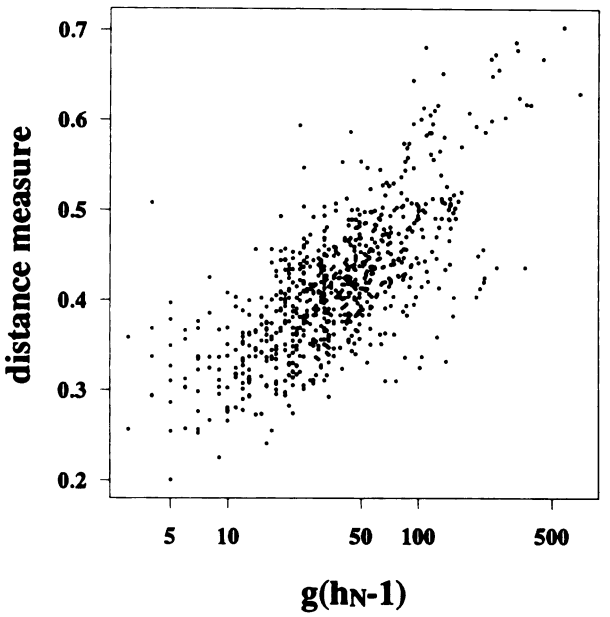


Figure 9: Average dissimilarities between native and their first-excited states for all non-degenerate 16-mers, vs. the number of first-excited states, $g(h_N - 1)$. Fluctuations are more distant from the native structure for HP sequences that have higher numbers of first-excited conformations.

a similar distribution in which each of the possible 16-mer native states is compared to each other native state rather than to its first-excited states. This second distribution involves unrelated conformations, and so should represent approximately the complete range of dissimilarity that is possible for compact 16-mer structures. The significant overlap of the two distributions shows that although most first-excited states represent small structural excursions from the native state, occasional low-energy fluctuations result in conformations that show near-maximal dissimilarity from the native. We call such states "conformational distant relatives" (CDR) of the native structure, being only slightly higher in free energy, but very different in conformation. Although the percentage of CDRs among first-excited states is small and highly sequence dependent (data not shown), we find that their relative population increases with the number of first-excited states, $g(h_N - 1)$ (Figure 9). That is, proteins that have a greater number of first-excited states also have more CDRs.

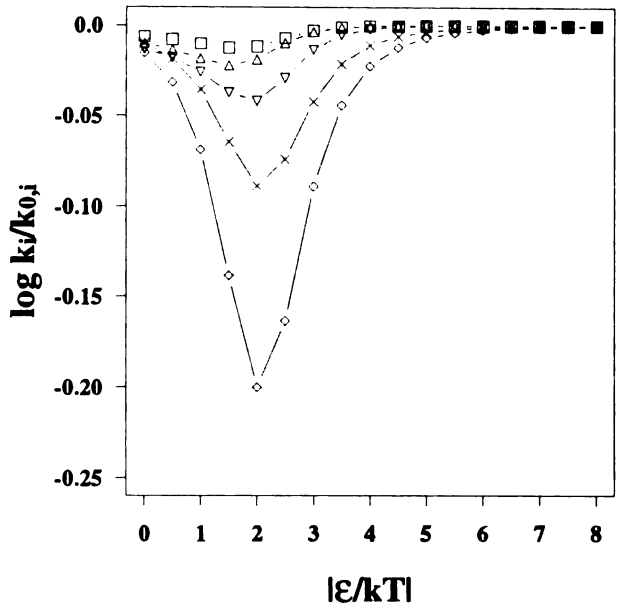


Figure 10: HX rates for monomers that are solvent accessible in the native state of Figure 4A. Curves are for monomers 2(□), 5(×), 6(◇), 7(▽), and 12(△). As $|\epsilon/kT|$ decreases towards the denaturation midpoint, HX rates first decrease due to burial in compact denatured conformations, then increase as the chain becomes unfolded.

This result is to be expected, since typically only a few first-excited states can be configured by small-amplitude fluctuations of the native state.

Figure 10 shows an unexpected prediction from the model. In the studies described above, we have considered protons that are exchange incompetent in the native state: for example, they may be interior protons buried within the core or surface protons that are involved in amide hydrogen bonds. As conditions become increasingly denaturing, such protons become increasingly competent to exchange, and their HX rates increase. Now what happens to protons that are exchange *competent* in the native state, such as those in exposed loop regions where hydrogen bonding is absent? In our lattice model of HX, such protons are represented by the monomers that are solvent accessible in the native state. Figure 10 indicates the interesting possibility that while such protons should exchange at the maximum rate under both stabilizing and destabilizing conditions, under *intermediate* conditions

of stability, they may exchange at a slightly *slower* rate. This is because surface protons that are locked into an exposed and unstructured position in the native state can become buried (or H-bonded) in the compact non-native states that are increasingly populated as native stability is slightly decreased. As stability is decreased further, all proton sites eventually become fully exposed and unstructured, and exchange rates increase towards the maximum. Hence just as small amounts of denaturing agents tend to expose hydrophobic residues, they may also tend to bury polar monomers. Figure 10 shows that such effects correspond only to small rate differences, perhaps accounting for why they have not been previously reported, as far as we know.

## DISCUSSION

One of the main issues addressed in the present study concerns the magnitude of protein fluctuations. A common assumption, implicit in both solvent penetration and local unfolding models is: (1) that such fluctuations must, under stable conditions, involve only small changes in energy relative to the native state, and (2) that conformations similar in energy to the native state must also be similar in structure. We agree with (1), but the present study suggests that (2), while true most of the time, may not be required. The fluctuations observed in our model can involve small changes in energy relative to the native state but large changes in conformation.

Are conformational distant relatives important for biological proteins? Two fundamental issues are: (1) whether low-energy states can be structurally dissimilar to the native state, and (2) whether such non-native structures, if they exist, can be visited on the time scales relevant to hydrogen exchange. We contend that (1)

is not an artifact of the lattice model, but follows from any model based on the idea that proteins have "rugged" energy landscapes (Bryngelson & Wolynes, 1989; Frauenfelder et al., 1991; Bryngelson et al., 1995; Dill et al., 1995; Wolynes et al., 1995), i.e., that very non-native structures can have very low energies (Figure 11A). The HP lattice model has such ruggedness. At the other extreme, protein folding might be guided by a smooth "funnel- like" energy landscape (Fig. 11B), whereby conformational dissimilarity with the native state implies a large difference in free energy. In that case, fluctuations involving small differences in energy relative to the native state must involve small conformational changes. This is the implicit assumption in existing local unfolding and solvent penetration models.

There is considerable experimental evidence that some proteins have energy landscapes that involve at least some degree of ruggedness. Some proteins fold on time scales of seconds and longer, rather than the much shorter time scales (perhaps nanoseconds to microseconds) that might be expected for smooth funnels. Both prion protein PrPC (Pan et al., 1993; Huang et al., 1994; Cohen et al., 1994) and alpha-lytic protease (Baker & Agard, 1994) have been found to exist in both a native and highly non-native conformation under physiological conditions, separated by a large kinetic barrier. Computer threading experiments (Novotny et al., 1984) have found that a single amino-acid sequence can be constrained to form two entirely different structures (one alpha-helix and the other beta-sheet) with negligible differences in the *in vacuo* CHARMM free energy.

While protein landscapes are not fully understood, they can be studied to some degree by kinetic refolding experiments, since it is the shape of the landscape that determines folding rates under native conditions. On this basis, evidence suggests that small, single-domain proteins, which typically show relatively simple kinetics, may have landscapes that look more like funnels. Proteins with more complex

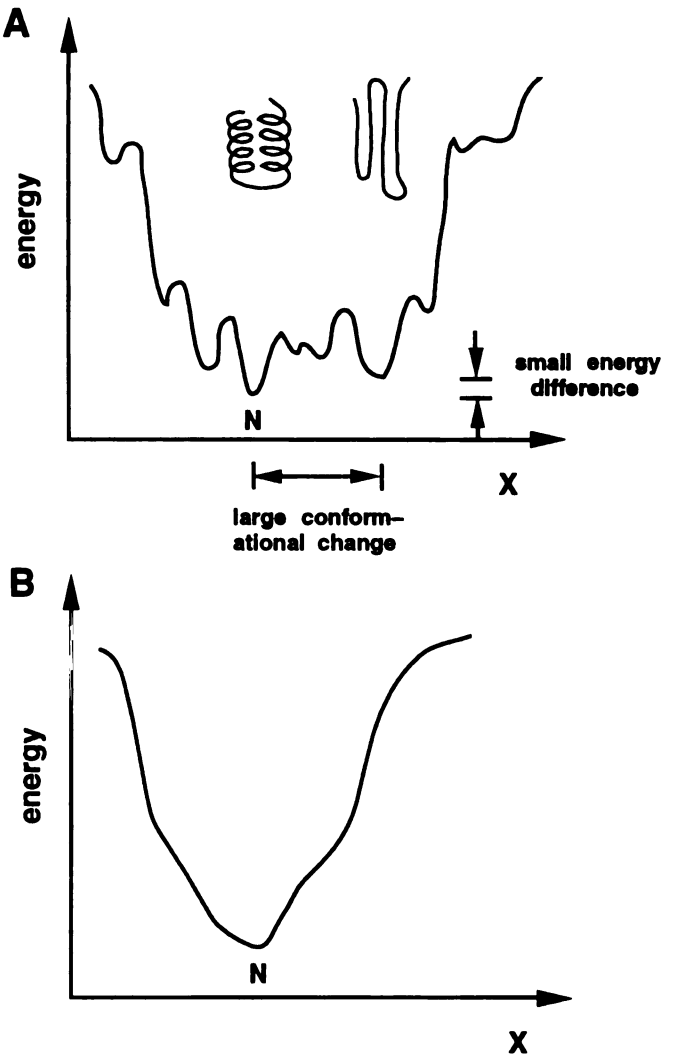


Figure 11: Two possible landscapes for protein folding. The y-axis is conformational free energy, and $x$ is one "reaction coordinate" for folding. $N$ denotes the native state. A: On rugged landscapes, there are multiple minima in free energy, so a small energy difference can correspond to a large conformational change. This is just a schematic to emphasize this point, and is not meant to imply that helical and sheet structures interconvert. B: On funnel-like landscapes, large distances from the native state imply a large increase in free energy.

folding kinetics, such as large, multi-domain proteins, probably have more rugged landscapes. It is not clear whether any biological protein has as much ruggedness as the HP lattice model. However, if the amount of ruggedness is significant (i.e., if there are multiple deep minima in free energy), then it follows, irrespective of the model used to describe it, that some fluctuations may bear little resemblance to the native structure.

The second issue is kinetic accessibility. Can the native structure of a protein fluctuate to very non-native low-energy conformations on the time scale of HX? The time scales of experiments dictate the heights of kinetic barriers that can be crossed. The large range of time scales observed for EX2 exchange—picoseconds to several months at low pH—suggests that some non-native conformations could be kinetically accessible, although the evidence is currently limited.

How can we determine if native protein fluctuations involve distant relatives? Such fluctuations are probably very difficult to observe experimentally because the native structure is so dominant under stabilizing conditions, and the denatured conformations are so dominant under unfolding conditions. Moreover these conformations need not be on the kinetic folding pathway, and so may be undetectable in refolding experiments. HX in the high-stability limit can detect native fluctuations, but as we noted before, slopes of Arrhenius plots tell us only that the *energy* changes (from the native state) are small; they cannot tell us how big the *structural* changes are. Native fluctuations can also be studied by NMR or by X-ray crystallographic B factors, but in both cases the measured fluctuations involve averages over multiple non-native states, so no one such conformation can be studied in isolation.

The present model appears consistent with data cited as support for local unfolding and solvent penetration models, because each of these models describes a subset of the fluctuations indicated by the present model. A main point of our study is

that local unfolding and solvent penetration may not be mutually exclusive: the native structures in our model show fluctuations of both types. At the same time, we find that structural fluctuations may sometimes be to conformations that differ considerably from the native state. Our model also predicts that knowledge of the native structure is not sufficient in itself to predict monomer exchange rates; these rates are strongly affected by the structures of the first-excited states, which cannot be predicted in a simple way from knowledge of the native structure. This result seems consistent with previously- reported difficulties in correlating HX rates with features of native structures.

An interesting question raised by these results is whether molecular dynamics calculations, which sample only small excursions from native structures, could detect CDR conformations.

Acknowledgments: The idea to do these calculations arose from conversations with Clare Woodward and Sarina Bromberg, to whom we are most grateful. We thank Robert L. Baldwin, Walter Englander, Tobin Sosnick, and Clare Woodward for teaching us about HX and for many very helpful discussions, Karen Tang for comments on the manuscript, and NIH for financial support. D.W. Miller has received support from the National Science Foundation Fellowship GER-9255688 (94-95).

## APPENDIX

### Accessibilities and Arrhenius slopes in the HP Model

In this appendix, we first convert Equation (7) into a form more convenient for the HP model. Equation (7) gives the Boltzmann-averaged solvent accessibility,

$\langle A \rangle_i$, of monomer $i$, for a set of $N$ conformations, $c$, of free energy $E_c$. It is more convenient to group the $N$ conformations by their numbers of HH contacts, $h$. Since the number of conformations with $h$ HH contacts is the density of states, $g(h)$, we can express Equation (7) as:

$$\langle A \rangle_i = \frac{\sum_{h=0}^{h_N} \sum_{c=1}^{g(h)} A_{c,i}\, e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} \sum_{c=1}^{g(h)} e^{-h\epsilon/kT}} \tag{11}$$

where $h_N$ is the number of HH contacts in the native conformation. Since the Boltzmann exponential factors depend only on $h$, and not on particular conformations, $c$, Equation (11) becomes:

$$\langle A \rangle_i = \frac{\sum_{h=0}^{h_N} e^{-h\epsilon/kT} \sum_{c=1}^{g(h)} A_{c,i}}{\sum_{h=0}^{h_N} g(h)\, e^{-h\epsilon/kT}} \tag{12}$$

Equation (12) can be further simplified by defining a quantity, $A_{h,i}$, which is the average solvent accessibility of monomer $i$ for all $g(h)$ conformations having $h$ HH contacts:

$$A_{h,i} \equiv \frac{1}{g(h)} \sum_{c=1}^{g(h)} A_{c,i} \tag{13}$$

For example, Figures 4B–D show that monomer 14 has solvent accessibilities of 1, 0, and 1 in the three conformations with $h = 7$, so its average solvent accessibility at this energy level is $A_{7,14} = 2/3$. Substituting Equation (13) into Equation (12) the average solvent accessibility becomes:

$$\langle A\rangle_i \;=\; \frac{\sum_{h=0}^{h_N} A_{h,i}\; g(h)\; e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} g(h)\; e^{-h\epsilon/kT}} \tag{14}$$

which is the result given in Equation (8). Since some monomers are solvent- inaccessible in the native ($A_{h_N,i} = 0$) and other compact states, we define $h_i^*$ as the largest value of $h$ for which $A_{h,i} \neq 0$. Thus for any monomer, $i$, buried in the native state, $h_i^* = h_{N-1}$ if it becomes solvent accessible in any of the first-excited states, $h_i^* = h_{N-2}$ if it is remains buried until the second-excited states, and so on. Figure 4 shows that $h_i^* = 7$ for monomers 8, 11, 14, and 16, all of which are buried in the native state but solvent accessible in the first-excited states. For monomer 15, which is buried in the native state and all three first- excited states, $h_i^* = 6$. Since $A_{h,i} = 0$ for $h > h_i^*$, the upper summation limit in the numerator becomes $h_i^*$, and Equation (14) becomes:

$$\langle A\rangle_i \;=\; \frac{\sum_{h=0}^{h_i^*} A_{h,i}\; g(h)\; e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} g(h)\; e^{-h\epsilon/kT}} \tag{15}$$

The slope of the temperature plot in Figure 5A is obtained by differentiating $\log \langle A\rangle_i$ with respect to $|\epsilon/kT|$. From Equation (15):

$$m(\epsilon)_i \;=\; \frac{\sum_{h=0}^{h_i^*} h\; A_{h,i}\; g(h)\; e^{-h\epsilon/kT}}{\sum_{h=0}^{h_i^*} A_{h,i}\; g(h)\; e^{-h\epsilon/kT}} \;-\; \frac{\sum_{h=0}^{h_N} h\; g(h)\; e^{-h\epsilon/kT}}{\sum_{h=0}^{h_N} g(h)\; e^{-h\epsilon/kT}} \tag{16}$$

The second term in Equation (16) is simply the average number of HH contacts for the chain, $\langle h\rangle$. To first approximation, $\langle h\rangle = h_N$ under native conditions, since

Figure 6 shows that the native population dominates over the experimental range in $|\epsilon/kT|$. The first term in Equation (16) can be simplified by assuming that solvent accessibility is two-state for each free energy level, i.e., that $A_{h,i}$ must equal 0 or 1. In other words, once a buried monomer becomes solvent accessible (in one or more of the $h = h_i^*$ conformations), it is assumed to be completely solvent accessible in every conformation of equal or more positive contact energy. This approximation does not significantly affect the slope of the exchange curve, although it does have a small effect on the two-slope transition temperature (data not shown). With the approximations given above, Equation (16) becomes:

$$m(\epsilon)_i = \frac{\sum_{h=0}^{h_i^*} h\, g(h)\, e^{-h\epsilon/kT}}{\sum_{h=0}^{h_i^*} g(h)\, e^{-h\epsilon/kT}} - h_N = \langle h \rangle_{acc,i} - h_N \qquad (17)$$

where $\langle h \rangle_{acc,i}$ is the average value of $h$ for only those conformations that are solvent accessible for monomer $i$, i.e., those for which $h \leq h_i^*$. Many exchange-competent conformations can contribute to exchange, so $\langle h \rangle_{acc,i}$ is determined by those that are most highly populated for a given $|\epsilon/kT|$. Since the contact free energy is $h\epsilon$, Equation (17) shows that to first approximation, the slope of the exchange curve is proportional to the free energy difference between the native state and predominant exchange- competent state.

## REFERENCES


Bai Y, Milne JS, Mayne L, Englander SW. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins Struct Funct Genet 17*: 75-86.

Bai Y, Milne JS, Mayne L, Englander SW. 1994. Protein stability parameters measured by hydrogen exchange. *Proteins Struct Funct Genet 20*: 4-14.

Baker D, Agard DA. 1994. Kinetics versus thermodynamics in protein folding. *Biochemistry 33*: 7505-7509.

Berger A, Linderstrom-Lang K. 1957. Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch Biochem Biophys 69*: 106-118.

Bradbury JH, Chapman B, Crompton MW, Norton R, Teh J. 1980. Hydrogen-deuterium exchange of the C-2 protons of histidine and histidine peptides and proteins. *J Chem Soc Perkin 2*: 693-699.

Bradbury JH, Crompton MW, Teh JS. 1977. Nuclear-magnetic-resonance study of the histidine residues of S-peptide and S-protein and kinetics of 1H-2H exchange of ribonuclease A. *Eur J Biochem 81*: 411-422.

Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins Struct Funct Genet 21*: 167-195.

Bryngelson JD, Wolynes PG. 1989. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J Phys Chem 93*: 6902-6915.

Chan HS, Dill KA. 1989. Compact polymers. *Macromolecules 22*: 4559-4573.

Chan HS, Dill KA. 1990. Origins of structure in globular proteins. *Proc Natl Acad Sci USA 87*: 6388-6392.

Chan HS, Dill KA. 1991. "Sequence space soup" of proteins and copolymers. *J Chem Phys 95*: 3775-3787.

Chan HS, Dill KA. 1994. Transition states and folding dynamics of proteins and heteropolymers.

*J Chem Phys 100*: 9238-9257.

Cohen FE, Pan KM, Huang Z, Baldwin M, Fletterick RJ, Prusiner SB. 1994. Structural clues to prion replication. *Science 264*: 530-531.

Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*: 7133-7155.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding. A perspective from simple exact models. *Protein Sci 4*: 561-602.

Eigen M. 1964. Proton transfer, acid-base catalysis, and enzymatic hydrolysis. *Angew Chem Int Ed Engl 3*: 1-19.

Ellis L M, Bloomfield VA, Woodward CK. 1975. Hydrogen-tritium exchange kinetics of soybean trypsin inhibitor. Solvent accessibility in the folded conformation. *Biochemistry 14*: 3413-3419.

Englander JJ, Calhoun DB, Englander SW. 1979. Measurement and calibration of peptide group hydrogen-deuterium exchange by ultraviolet spectrophotometry. *Anal Biochem 92*: 517-524.

Englander JJ, Rogero JR, Englander SW. 1983. Identification of an allosterically sensitive unfolding unit in hemoglobin. *J Mol Biol 169*: 325-344.

Englander SW. 1975. Measurement of structural and free energy changes in hemoglobin by hydrogen exchange methods. *Ann NY Acad Sci 244*: 10-27.

Englander SW, Downer NW, Teitelbaum H. 1972. Hydrogen exchange. *Annu Rev Biochem 41*: 903-924.

Englander SW, Englander JJ. 1978. Hydrogen-tritium exchange. *Methods Enzymol 49G*: 24-39.

Englander SW, Englander JJ, McKinnie RE, Ackers GK, Turner GJ, Westrick JA, Gill SJ. 1992. Hydrogen exchange measurement of the free energy of structural and allosteric change in hemoglobin. *Science 256*: 1684-1687.

Englander SW, Kallenbach NR. 1984. Structural dynamics of proteins and nucleic acids. *Q Rev Biophys 16*: 521-655.

Englander SW, Mauel C. 1972. Hydrogen exchange detection of discrete ligand-induced changes

in hemoglobin. *J Biol Chem 247*: 2387-2394.

Englander SW, Poulsen A. 1969. Hydrogen-tritium exchange of the random chain polypeptide. *Biopolymers 7*: 329-393.

Englander SW, Staley R. 1969. Measurement of the free and the H-bonded amides of myoglobin. *J Mol Biol 45*: 277-295.

Frauenfelder H, Petsko GA, Tsernoglu D. 1979. Temperature dependent X-ray diffraction as a probe of protein structural dynamics. *Nature 280*: 558-563.

Frauenfelder H, Sligar SG, Wolynes PG. 1991. The energy landscapes and motions of proteins. *Science 254*: 1598-1603.

Goodman E, Kim P. 1991. Periodicity of amide proton exchange rates in a coiled-coil leucine zipper peptide. *Biochemistry 30*: 11615-11620.

Gurd FRN, Rothgeb M. 1979. Motions in proteins. *Adv Protein Chem 33*: 73-165.

Haslam JL, Eyring EM. 1967. Deuterium oxide solvent isotope effects on N-H...O, O-H...N, and N-H...N intramolecular hydrogen bonds. *J Phys Chem 71*: 4470-4475.

Hilton BD, Trudeau K, Woodward CK. 1981. Hydrogen exchange rates in pancreatic trypsin inhibitor are not correlated to thermal stability in urea. *Biochemistry 20*: 4697-4703.

Hilton BD, Woodward CK. 1979. On the mechanism of isotope exchange kinetics of single protons in bovine pancreatic trypsin inhibitor. *Biochemistry 18*: 5834-5841.

Huang Z, Gabriel JM, Baldwin MA, Fletterick RJ, Prusiner SB, Cohen FE. 1994. Proposed three-dimensional structure for the cellular prion protein. *Proc Natl Acad Sci USA 91*: 7139-7143.

Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science 249*: 1544-1548.

Hvidt A. 1964. Discussion of the pH dependence of the H-D exchange of proteins. *C R Trav Lab Carlsberg 34*: 299-317.

Hvidt A, Linderstrom-Lang K. 1954. Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochim Biophys Acta 14*: 574-575.

Hvidt A, Nielsen SO. 1966. Hydrogen exchange in proteins. *Adv Protein Chem 21*: 287-386.

Ikegami A, Kono N. 1967. Tritium-hydrogen exchange of polypeptides in aqueous solutions. *J Mol Biol 29*: 251-274.

Karplus M, McCammon JA. 1981. The internal dynamics of globular proteins. *CRC Crit Rev Biochem 9*: 293-349.

Kim KS, Woodward C. 1993. Protein internal flexibility and global stability: effect of urea on hydrogen exchange rates of bovine pancreatic trypsin inhibitor. *Biochemistry 32*: 9609-9613.

Kossiakoff AA. 1982. Protein dynamics investigated by the neutron diffraction-hydrogen exchange technique. *Nature 296*: 713-721.

Kuwajima K, Baldwin RL. 1983. Exchange behavior of the H-bonded amide protons in the 3-13 helix of ribonuclease S. *J Mol Biol 169*: 299-324.

Lau KF, Dill KA. 1989. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules 22*: 3986-3997.

Lau KF, Dill KA. 1990. Theory for protein mutability and biogenesis. *Proc Natl Acad Sci USA 87*: 638-642.

Leichtling BH, Klotz IM. 1966. Catalysis of H-D exchange in polypeptides. *Biochemistry 5*: 4026-4036.

Levitt M. 1981a. Hydrogen bond and internal solvent dynamics of BPTI protein. *Ann NY Acad Sci 367*: 162-181.

Levitt M. 1981b. Molecular dynamics of hydrogen bonds in bovine pancreatic trypsin inhibitor protein. *Nature 294*: 379-380.

Lumry R, Rosenberg A. 1975. The mobile defect hypothesis of protein function. *Colloq Int CNRS 246*: 55-63.

Mayo SL, Baldwin, RL. 1993. Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science 262*: 873-876.

Molday RS, Englander SW, Kallen RG. 1972. Primary structure effects on peptide group hydrogen

exchange. *Biochemistry 11*: 150-158.

Nakanishi M, Tsuboi M, Ikegami A. 1972. Fluctuations of the lysozyme structure. *J Mol Biol 70*: 351-361.

Nakanishi M, Tsuboi M, Ikegami A. 1973. Fluctuations of the lysozyme structure. II. Effects of temperature and binding of inhibitors. *J Mol Biol 75*: 673-682.

Novotny J, Bruccoleri R, Karplus M. 1984. An analysis of incorrectly folded protein models. *J Mol Biol 177*: 787-818.

Otting G, Liepinsh E, Wuthrich K. 1991. Protein hydration in aqueous solution. *Science 254*: 974-980.

Pan KM, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang Z, Fletterick RJ, Cohen FE. 1993. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci USA 90*: 10962-10966.

Pedersen T, Thomsen N, Andersen K, Madsen J, Poulsen F. 1993. Determination of the rate constants k1 and k2 of the Linderstrom-Lang model for protein amide hydrogen exchange. *J Mol Biol 230*: 651-660.

Pershina L, Hvidt A. 1974. A study by the hydrogen-exchange method of the complex formed between the basic pancreatic trypsin inhibitor and trypsin. *Eur J Biochem 48*: 339-344.

Radford S, Dobson C, Evans P. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature 358*: 302-307.

Richards FM. 1979. Packing defects, cavities, volume fluctuations, and access to the interior of proteins. *Carlsberg Res Commun 44*: 47-63.

Roder H. 1989. Structural characterization of protein folding intermediates by proton magnetic resonance and hydrogen exchange. *Methods Enzymol 176*: 446-473.

Roder H, Elove GA, Englander SW. 1988. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature 335*: 700-704.

Roder H, Wagner G, Wuthrich K. 1985. Amide proton exchange in proteins by EX1 kinetics: stud-

ies of the basic pancreatic trypsin inhibitor at variable p2H and temperature. *Biochemistry 24*: 7396-7407.

Rohl CA, Scholtz JM, York EJ, Stewart JM, Baldwin RL. 1992. Kinetics of amide proton exchange in helical peptides of varying chain lengths. Interpretation by the Lifson-Roig equation. *Biochemistry 31*: 1263-1269.

Rose MC, Stuehr J. 1968. Kinetics of proton transfer reactions in aqueous solution: rates of internally hydrogen-bonded systems. *J Am Chem Soc 90*: 7205-7209.

Schoenborn BP, Hanson JC, Darling GD, Norvell JC. 1978. Real space refinement of neutron diffraction data from carbonmonoxide sperm whale myoglobin. *Acta Crystallogr A34*: 65.

Segal DM, Harrington WF. 1967. The tritium-hydrogen exchange of myosin and its proteolytic fragments. *Biochemistry 6*: 768-787.

Shortle D, Chan HS, Dill KA. 1991. Modeling the effects of mutations on the denatured states of proteins. *Protein Sci 1*: 201-215.

Tuchsen E, Hayes J, Ramaprasad S, Copie V, Woodward C. 1987. Solvent exchange of buried water and hydrogen exchange of peptide NH groups hydrogen bonded to buried waters in bovine pancreatic trypsin inhibitor. *Biochemistry 26*: 5163-5172.

Tuchsen E, Hvidt A, Ottesen M. 1980. Enzymes immobilized as crystals. Hydrogen isotope exchange of crystalline lysozyme. *Biochimie 62*: 563-566.

Tuchsen E, Ottesen M. 1979. A simple hydrogen exchange method for cross-linked protein crystals. *Carlsberg Res Commun 44*: 1-10.

Udgaonkar JB, Baldwin RL. 1988. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature 335*: 694-699.

Wagner G, Wuthrich K. 1982. Amide proton exchange and surface conformation of the basic pancreatic trypsin inhibitor (BPTI) in solution: studies with two dimensional NMR. *J Mol Biol 160*: 343-361.

Welch WH, Fasman GD. 1974. Hydrogen-tritium exchange in polypeptides. Models of alpha-helical

and beta conformations. *Biochemistry 13*: 2455-2466.

Wemmer DE, Kallenbach NR. 1983. The structure of apamin in solution: 2D NMR study. *Biochemistry 22*: 1901-1906.

Wlodawer A, Sjolin L. 1982. Hydrogen exchange in ribonuclease A: neutron diffraction study. *Proc Natl Acad Sci USA 79*: 1418-1422.

Wolynes PG, Onuchic JN, Thirumalai D. 1995. Navigating the folding routes. *Science 267*: 1619-1620.

Woodward CK. 1977. Dynamic solvent accessibility in the soybean trypsin inhibitor-trypsin complex. *J Mol Biol 11*: 509-515.

Woodward CK, Ellis LM, Rosenberg A. 1975. Solvent accessibility in folded proteins: studies of hydrogen exchange in trypsin. *J Biol Chem 250*: 432-439.

Woodward CK, Hilton BD. 1979. Hydrogen exchange kinetics and internal motions in proteins and nucleic acids. *Annu Rev Biophys Bioeng 8*: 99-127.

Woodward CK, Hilton BD. 1980. Hydrogen isotope exchange kinetics of single protons in bovine pancreatic trypsin inhibitor. *Biophys J 32*: 561-576.

Woodward CK, Rosenberg A. 1971. Studies of hydrogen exchange in proteins. VI. Urea effects on RNase hydrogen exchange kinetics leading to a general model for hydrogen exchange from folded proteins. *J Biol Chem 246*: 4114-4121.

Woodward CK, Simon I, Tuchsen E. 1982. Hydrogen exchange and the dynamic structure of proteins. *Mol Cell Biochem 48*: 135-160.

Wuthrich K, Wagner G. 1978. Internal motion in globular proteins. *Trends Biochem Sci 3*: 227-230.

Yee DP, Dill KA. 1993. Families and the structural relatedness among globular proteins. *Protein Sci 2*: 884-899.

Yee RY, Englander SW, von Hippel PH. 1974. Native collagen has a two-bonded structure. *J Mol Biol 83*: 1-16.

# Chapter 2

## Ligand Binding To Proteins: The Binding Landscape Model

This chapter is taken from the following published article: Miller DW, Dill KA. 1997. Ligand binding to proteins: the binding landscape model. *Prot Sci* (in press).

## ABSTRACT

Models of ligand binding are often based on four assumptions: **(1) steric fit:** that binding is determined mainly by *shape complementarity*, **(2) native binding:** that ligands mainly bind to *native states*, **(3) locality:** that ligands perturb protein structures mainly at the binding site, and **(4) continuity:** that small changes in ligand or protein structure lead to small changes in binding affinity. Using a generalization of the 2D HP lattice model, we study ligand binding and explore these assumptions. We first validate the model by showing that it reproduces typical binding behaviors. We observe ligand-induced denaturation, ANS and heme-like binding, and "lock-and-key" and "induced-fit" specific binding behaviors characterized by Michaelis-Menten or more cooperative types of binding isotherms. We then explore cases where the model predicts violations of the standard assumptions. For example, very different binding modes can result from two ligands of identical shape. Ligands can bind highly denatured states more tightly than native states and yet have Michaelis-Menten isotherms. Even low-population binding to denatured states can cause changes in global stability, hydrogen-exchange rates, and thermal B-factors, contrary to expectations, but in agreement with experiment. We conclude that ligand binding, like protein folding, may be better described in terms of energy landscapes than in terms of simpler mass-action models.

## STANDARD ASSUMPTIONS IN BINDING MODELS

Ligand binding is important for protein function. A quantitative understanding of many biological binding processes has been gained through *binding polynomial* models (Wyman & Gill, 1990; Di Cera, 1995; Ackers et al., 1992), empirical

equations which relate the fractional occupation of binding sites to free ligand concentration. This approach underlies Michaelis-Menten kinetics, Hill and Scatchard plots, and cooperativity and allostery as embodied in the MWC (Monod et al., 1965) and KNF (Koshland et al., 1966) models, for example. Despite the great power and widespread usage of binding polynomial models in biochemistry, they are incomplete and phenomenological. For a given protein and ligand, binding polynomial models do not tell us where or how tightly the ligand will bind, whether or not binding will induce a conformational change, or whether the ligand will denature the protein, for example. Binding polynomial models begin by *assuming* some mass-action scheme for the binding process, and the binding and cooperativity constants are then determined by curve-fitting to experimental data. Finding the right binding model for a given ligand/protein system is a matter of trial and error.

A more complete binding model would *predict* the binding sites, the binding constants and cooperativity, and the perturbations of the protein, based on knowledge of the ligand structure and the amino acid sequence of the protein. Such models are far beyond the current scope of computational biochemistry. Nevertheless we take a step in that direction by using a simplified statistical mechanical model of protein/ligand interactions, for which we can exactly enumerate all possible protein conformations and binding modes. Our aim here is to describe such a model and its predictions for binding.

First we show that the model leads to many of the familiar types of protein/ligand binding, including "lock and key" and "induced fit" specific binding, ANS binding to molten globules, and ligand-induced denaturation, among others. We then explore some interesting and unconventional behaviors predicted by the model, many of which are not readily interpreted using simpler mass-action models. In particular, we address four premises in which the current paradigm of ligand binding is heavily

rooted: (1) Binding is largely *shape- determined*, as embodied in the terms "lock-and-key" and "induced *fit*". (2) Ligands bind principally to the *native states* of proteins with little or no interaction with the *unfolded* states. Here is a typical description: "A general consequence of ligand binding is that the protein is stabilized against unfolding and is less flexible .... [This is] a consequence of the ligand binding more tightly to the fully folded conformation (N) than to the fully unfolded state (U) and any distorted or partially unfolded forms that result from flexibility of the structure (Creighton, 1993)." (3) Binding is highly *localized.* The main perturbations of the protein structure are assumed to be near the binding site. (4) Small changes in the structure of a ligand or protein lead to only small changes in the bound complex.

These premises are usually supported by x-ray and NMR structures of unbound and complexed proteins, and by the successes of structure-based drug design methods (Kuntz, 1992; Shoichet et al., 1993; Bohacek & McMartin, 1994; Strynadka et al., 1996). Nevertheless a few recent experimental results, particularly from hydrogen exchange, are puzzling when interpreted using these premises. Here we develop a model that, unlike binding polynomial models, aims to connect structure to thermodynamics. Although the physical model is simple, the statistical mechanics is rigorous so we can test such premises, rather than assume them. For reasons that will become clear below, we call ours the Binding Landscape Model, to contrast it to those based on the premises above, such as the Lock-and-Key and Induced Fit models. The groundwork for connecting ligand binding to energy landscapes has been described in theoretical and experimental work on small molecule binding to globins by Frauenfelder, Wolynes and others (Frauenfelder et al., 1991).

### Modeling proteins using the HP lattice model.

We model proteins using the 2-dimensional HP lattice model (Lau & Dill, 1989,

1990; Chan & Dill, 1991; Dill et al., 1995). A protein is represented as a sequence of H (hydrophobic) and P (other) monomers on a two-dimensional lattice. Lattice sites may be either empty or filled by a single monomer, and empty lattice sites are assumed to contain a solvent molecule. Each HH contact, formed when two non-sequential H monomers occupy adjacent lattice sites, is favored by a free energy $\epsilon$ ($\epsilon < 0$), which is meant to capture the importance of hydrophobic interactions in protein collapse and global stability (Dill, 1990). Hence the free energy of a conformation is $h\epsilon$, where $h$ is the number of HH contacts. The magnitude of $\epsilon$ determines the stability imparted by external conditions: large and negative $\epsilon$ reflects conditions that are more stabilizing, such as lower temperature or lower denaturant concentrations. Conformational entropy, the driving force for unfolding, enters the model through the exhaustive enumeration of all the possible chain configurations (see below).

The disadvantages of the model are clear: atomic resolution is lost; conformations are restricted to a lattice; it is in 2 dimensions; the energy function is simplified; and chains are unrealistically short. Yet despite these disadvantages, the model has been found useful for modeling protein properties (Lau & Dill, 1989, 1990; Chan & Dill, 1989, 1990, 1994; Shortle et al., 1991; Dill et al., 1995; Miller & Dill, 1995) because it shows several protein-like features, including cooperative collapse, native structures having a nonpolar core and definable secondary structures, multi-stage folding kinetics, and molten globule states. Most importantly, we believe the model captures the main physics of protein folding—the hydrophobic interactions, conformational freedom of the chain, and the steric restrictions imposed by excluded volume.

We study HP sequences having 16 monomers. For any 16-mer chain there are exactly 802,075 possible conformations that can be configured on a two-dimensional

square lattice. These conformations are generated by computer, and each is weighted by a Boltzmann factor according to the number of HH contacts made. Figure 1A shows an energy diagram for a sample HP sequence (called "sequence A"). The native structure (ground state) is the conformation with the largest possible number of HH contacts, and thus the lowest free energy at low temperatures. We study only non-degenerate sequences, i.e., those having a single native conformation, since we believe they best represent biological proteins, which fold to unique structures. All higher-energy conformations comprise the non-native, or "denatured" states, and are grouped by energy into "first-excited" states, "second-excited" states, etc., corresponding to successively fewer HH contacts. For any HP sequence, there are far more open, high- energy conformations than compact, low-energy conformations (see Figure 1B).

**Modeling the ligand and its interactions with the protein.**

We model ligands as single, monomer-sized beads (Figure 1C). A protein-ligand contact occurs when the ligand occupies a lattice site adjacent to a chain monomer. Here we consider only nonpolar ligands: a contact between a ligand and an H monomer (LH contact) is favored by a free energy $b\epsilon$, where $\epsilon$ is the HH contact energy, and $b$ is a positive constant ($0 < b < 1$). In order to have the simplest possible model of binding, we assume the interaction energy is zero between a ligand and a P monomer, and zero between ligands. The total contact energy, $E_s$, for any protein-ligand configuration ("ligation state"), $s$, is therefore

$$E_s = h\epsilon + mb\epsilon \tag{1}$$

where $m$ is the total number of LH contacts. The ligation state in Figure 1C has 3 HH contacts and 6 LH contacts, so the total energy is $3\epsilon + 6b\epsilon$.

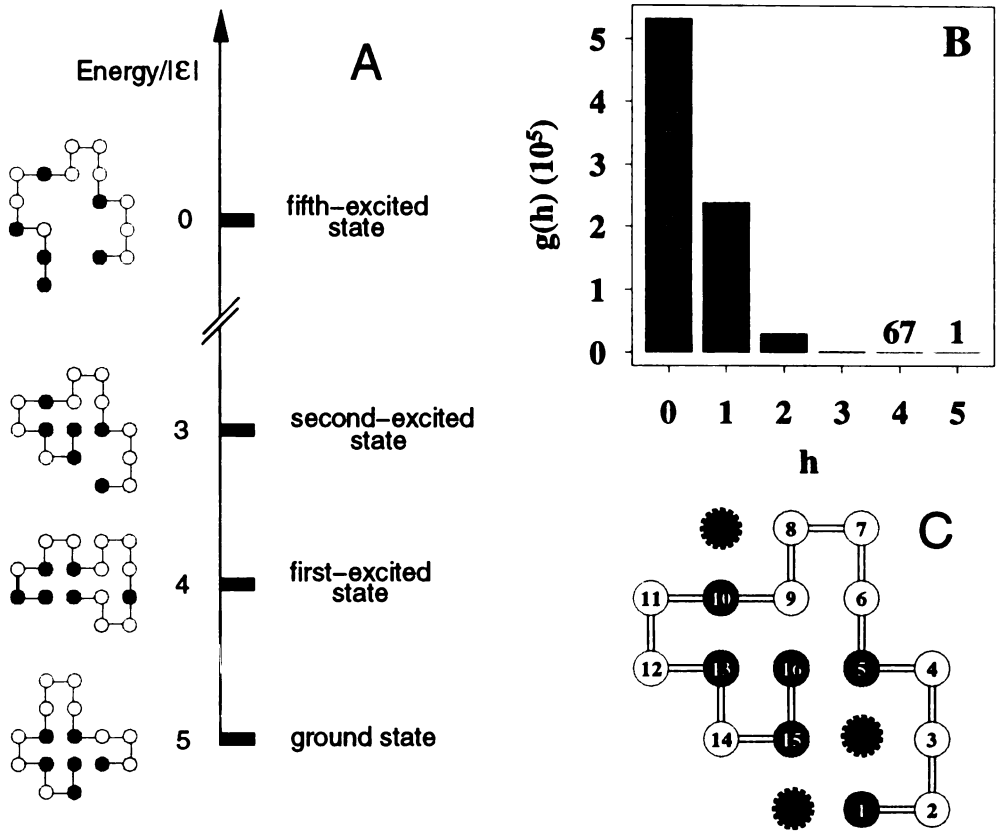


Figure 1: **Energy ladder diagram** for 16-mer HP sequence A. Each hydrophobic (HH) contacts is favored by a free energy $\epsilon$. The native conformation, with 5 HH contacts, is at the bottom, and each step up the ladder represents the loss of one HH contact. **B**: Density of states, $g(h)$: the numbers of conformations of sequence A having $h$ HH contacts. **C**: A particular ligation state with 3 ligands bound (protein monomers are numbered). Ligands are highlighted "beads." Contacts between a ligand and H monomer (LH contacts) are favored by a free energy $b\epsilon$, where $b$ is a constant. This ligation state makes 6 LH contacts and 3 HH contacts, so the energy is $3\epsilon + 6b\epsilon$ (Equation 1).

Below we show how average properties of protein-ligand complexes are derived through exact enumeration of all the possible protein-ligand configurations. First, we present the general statistical mechanical theory, which is independent of the lattice or any other specific model. Then we introduce the HP lattice model to relate the binding thermodynamics to the corresponding structures. A more detailed derivation of the theory is given in the appendix.

**Statistical mechanics of ligand binding.**

The probability, $P_s$, of any protein-ligand ligation state, $s$, is given by the grand canonical distribution function:

$$P_s = \frac{e^{-E_s/kT} e^{\mu N_s/kT}}{\sum_{s=1}^{\Gamma} e^{-E_s/kT} e^{\mu N_s/kT}} \tag{2}$$

where $T$ is absolute temperature, $\mu$ is the ligand chemical potential, $E_s$ is the total energy of the ligation state (i.e. due to intrachain contacts plus ligand contacts), $N_s$ is the number of ligands bound, $k$ is Boltzmann's constant, and $\Gamma$ is the total number of ligation states available to the protein-ligand complex.

Equation 2 gives the probability of a specific ligation state as a function of temperature and ligand concentration (related to the ligand's chemical potential; see the appendix). In broad terms, this equation predicts the following behaviors. At low temperature and low ligand concentration (large negative $\mu$), proteins are folded and have few ligands bound. Increasing the ligand concentration lowers the unfavorable translational entropy of binding, and more ligands bind. Increasing the temperature denatures the protein. It is the interplay of these behaviors, for different ligands and different proteins, that we explore more fully below.

Equation 2 is general and model-independent. It permits the calculation of var-

ious average protein properties (see appendix), provided that all possible ligation states can be enumerated. The HP model permits this, and thus provides a way to relate the binding thermodynamics to the structure and stability of the protein. Using Equation 1 for the ligation-state energy, $E_s$, of the HP model protein-ligand complex, Equation 2 becomes

$$P_s = \frac{e^{-h\epsilon/kT} e^{-mb\epsilon/kT} e^{\mu N_s/kT}}{\sum_{s=1}^{\Gamma} e^{-h\epsilon/kT} e^{-mb\epsilon/kT} e^{\mu N_s/kT}} \tag{3}$$

Equation 3 can be computed exactly by exhaustive enumeration for any short HP sequence, thus providing average properties of model proteins as a function of temperature, ligand concentration, binding constants, and monomer sequence.

Despite its simplicity, the HP model offers several advantages for studying principles of protein-ligand interactions. (1) We can consider all model protein conformations, so there is no approximation or partial sampling of the protein conformational space, allowing a complete study of the effects of binding on protein structure. (2) We can compute all possible ligation states, for every possible chain conformation, so we are not limited to one or a few ligands bound at a time. We can explore a full range of ligand concentrations, from zero to denaturing. (3) The model has only two energy parameters, so we can explore the physics in a complete way. We make no assumptions about the locations or numbers of binding sites, about the mechanisms of ligand-induced conformational changes, or about how binding is affected by external conditions. Rather, these properties are derived from the theory.

Our aim is not to describe the biology of the various interactions, which may be quite complex, but rather to show how a diverse collection of binding phenomena can be understood through a simple unified picture that relies on only a few basic, physical concepts.

## (1) VALIDATION OF THE MODEL: PROTEIN-LIKE BINDING

The model shows a range of behaviors that mimic real protein-ligand interactions. These can be divided into two classes, which we call *specific* and *non-specific* binding. We refer to binding as specific when a ligand binds to high-affinity (2 or more LH contact) sites on the protein, and when binding follows a site-specific (Michaelis-Menten) isotherm as described below. We call binding non-specific when many ligands bind, either to the native or denatured states of the protein, and when the binding isotherms do not show site-specific thermodynamics.

### Non-specific binding behaviors of the model

#### Denaturation.

High-affinity ligands (roughly $0.5 < b < 1.0$) at high concentration can induce unfolding in model proteins (see Figure 2). As ligand concentration increases, the translational entropy of binding becomes more favorable, so LH contacts are favored at the expense of HH contacts, driving the protein to unfold. This model result is similar to protein denaturation by urea and guanidinium chloride. Other simple models of denaturation have been explored previously (Alonso & Dill, 1991; Thomas & Dill, 1993).

#### Dyes and weak solvents.

Low-affinity ligands ($b < 0.2$) have little effect on model protein structure, even at high ligand concentrations. For these ligands, binding is sufficiently weak that

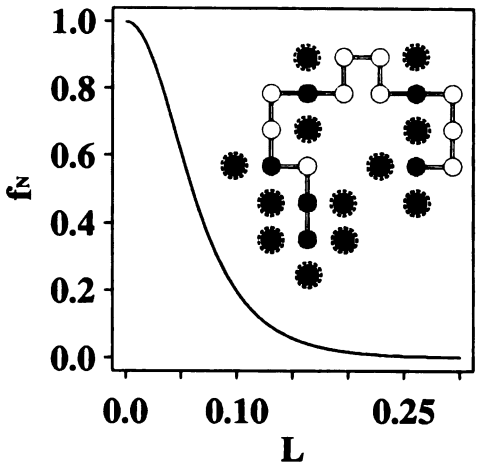


Figure 2: **Denaturation by ligand.** The fractional population of native proteins, $f_N$, vs. ligand concentration. High-affinity ligands at high concentrations denature model proteins. ($b = 0.50$, $\epsilon = -10$). **Inset:** A sample ligation state from the ensemble of denatured conformations.

breaking HH contacts to make LH contacts is always unfavorable. Thus at high ligand concentrations all hydrophobic sites on the protein surface become saturated, but the native structure remains intact. This model mimics the behavior of certain dyes and organic solvents (Allen et al., 1996; Mattos & Ringe, 1996) which bind, but do not perturb the native structure.

**ANS-like ligands.**

For an intermediate range of affinities in our model ($0.2 < b < 0.5$), ligands are too weak to induce full denaturation, but are strong enough to shift the native-denatured equilibrium. This class of ligands binds preferentially to the *compact denatured* states of the model proteins. As a result, the average number of bound ligands follows a bell-shaped curve (Figure 3A) as external conditions are changed from native to denaturing. The maximum in the number of bound ligands occurs under the intermediate solvent conditions at which the compact denatured states

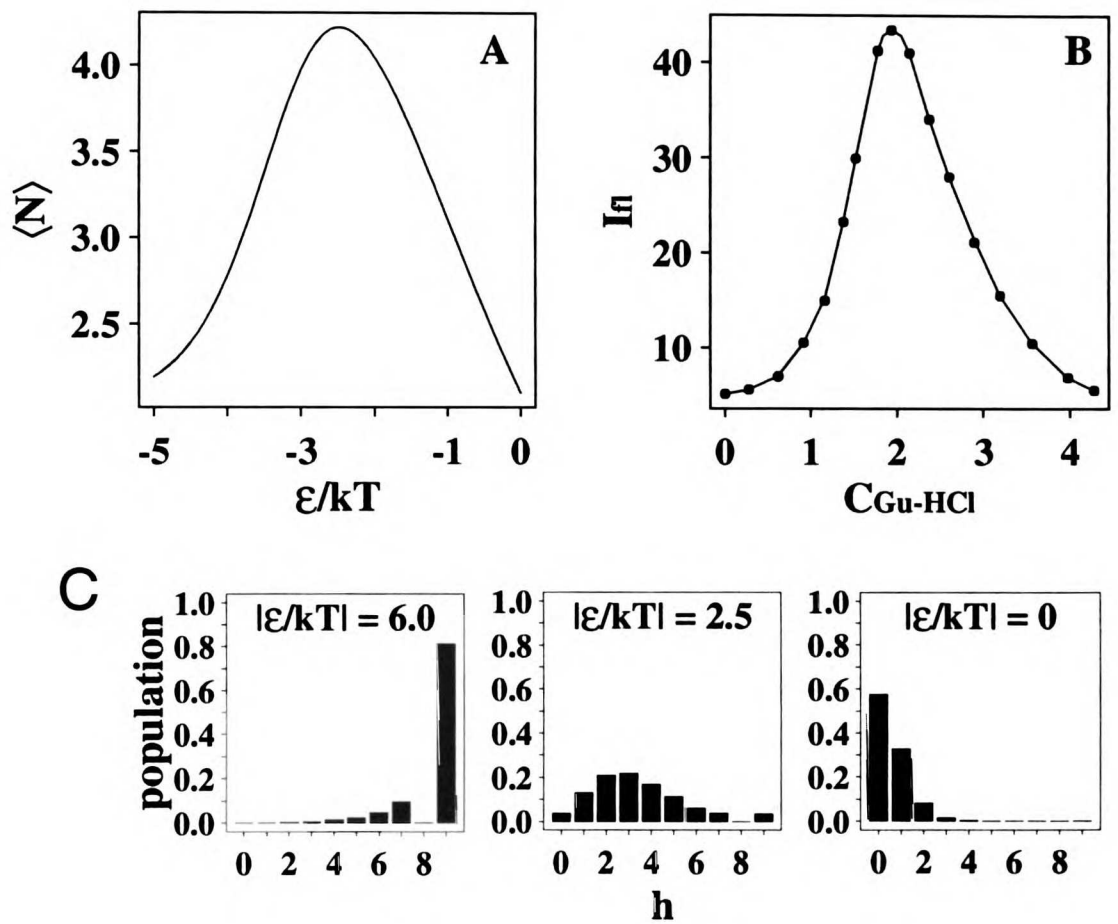


Figure 3: **Model of ANS binding.** Ligands with *intermediate* affinities bind preferentially to compact-denatured states, much like ANS binds to molten globules. **A**: Average number of bound ligands vs. external conditions $\epsilon/kT$ (native stabilizing conditions to the left; denaturing conditions to the right) with $L = 0.15$ and $b = 0.50$. **B**: Corresponding experiment: fluorescence-intensity changes from ANS binding to bovine carbonic anhydrase B, vs. Gu-HCl concentration (adapted from Semisotnov et al., 1991). **C**: Energy-ladder histograms for the protein conformations, showing that under conditions where binding is greatest ($\epsilon/kT = 2.5$ in A), the molecules have intermediate compactness ($h \approx 3 - -4$, middle of C).

are most stable. This behavior resembles that of ANS, a hydrophobic dye that binds preferentially to the molten globule states of proteins (Semisotnov et al., 1987; Semisotnov et al., 1991; Shi et al., 1994) (see Figure 3B).

Figure 3C explains this behavior. The number of ligands bound depends on a product of two factors: the number of accessible hydrophobic sites on the protein, and the strength of the binding interaction under the given conditions. Under native conditions (i.e., large $|\epsilon/KT|$), binding is strong (see Equation 1), but there are few available hydrophobic sites on the predominantly native protein, so there is little binding. Under denaturing conditions, the protein is unfolded and there are many exposed H sites, but binding is limited by the weak LH attraction. The number of bound ligands per protein is highest between these extremes, where the compact-denatured states are highly populated.

Semisotnov et al. (1987; 1991) have interpreted ANS binding as requiring "hydrophobic clustering" in the protein. In our model, hydrophobic clustering happens too, but clustering is a *consequence* of the balance between binding strength and number of binding sites, not a special mechanism of binding. In our model, there is nothing different about binding to a cluster than to any other arrangement of the same number of hydrophobic monomers.

## Specific binding behaviors of the model

To define site-specific binding, we begin with the traditional mass-action description. Equation 4 illustrates the equilibrium between a native conformation, $N$, and its bound state, $NL$.

$$N + L \overset{K_{bN}}{\rightarrow} NL \tag{4}$$

where $K_{bN}$ is the binding equilibrium constant and $L$ represents free ligand.

The fraction, $f_b$, of bound $N$ molecules is (see appendix)

$$f_b = \frac{K_{bN}L}{1 + K_{bN}L} \tag{5}$$

which approaches unity as ligand concentration increases. We refer to behavior described by Equation 5 as site-specific, or "Michaelis-Menten" binding.

**Lock-and-key and induced-fit binding.**

For some HP model proteins, and for high-affinity ligands ($0.5 < b < 1.0$), binding is localized to a single site on the protein and follows a Michaelis-Menten binding isotherm. Figure 4 shows a "lock and key" example of specific binding, in which the protein binds in its native conformation. The computed binding isotherm closely follows the Michaelis-Menten binding of Equation 5.

Site-specific binding might be considered surprising in this model, for two reasons. First, the model allows large numbers of alternate sites among the non-native and native states, since each H monomer is a potential contact. Second, the model protein/ligand interactions are *orientationally non-specific*, lacking the geometric requirements of hydrogen bonding, for example. The specificity in our model arises instead from the ability of the protein to configure in a specific way, namely with a compatible pocket: the protein and ligand cannot mutually find any lower-energy configuration. While real proteins often take advantage of chemically specific interactions such as hydrogen bonds or salt bridges, our minimal model shows that binding specificity does not require it.

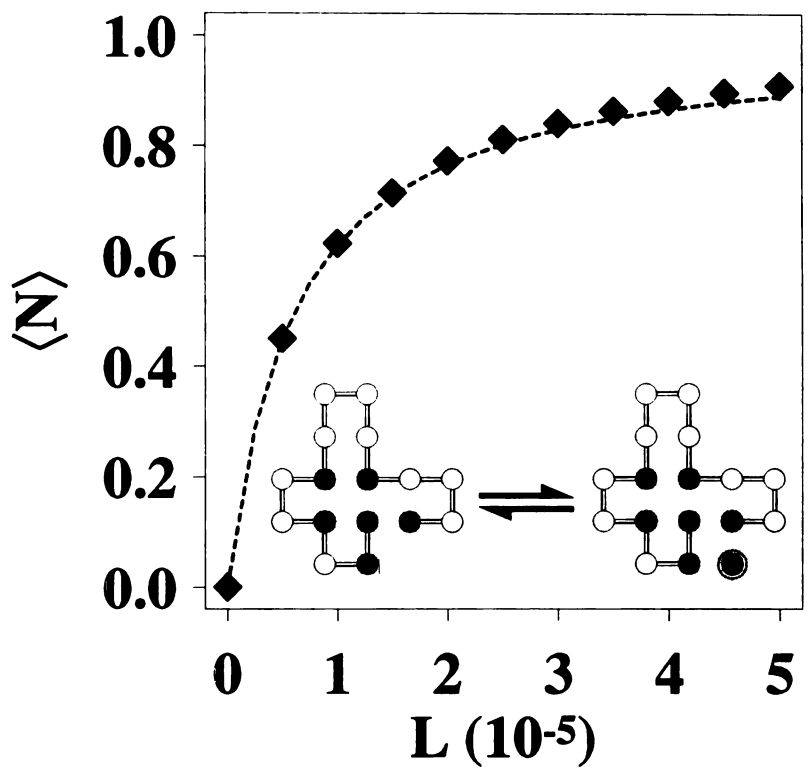


Figure 4: **Lock-and-key binding.** The average number of bound ligands, $\langle N \rangle$ (diamonds), vs. ligand concentration closely follows the theoretical curve (dashed line) expected for site-specific (Michaelis-Menten) binding. ($\epsilon = -10$ and $b = 0.6$). **Inset:** The ligand binds at a single site on the native protein, without perturbing its structure.

What is the basis for the specific binding in the model? There are two reasons the protein has relatively few binding options under native conditions. First, if a ligand is to bind a non-native conformation, the system must pay an energetic price to "excite" the protein from the native state to the non-native state. Even though the numbers of protein conformations and potential binding sites grow dramatically with increasing steps up the energy ladder, Boltzmann's law dictates that the ligand prefers to choose from the relatively few ligation states low on the energy ladder. Second, the binding of more than one ligand is disfavored by the high price in translational entropy at low ligand concentrations. We observe lock-and-key binding for approximately 11% of unique-folding 16-mer HP sequences. These are all sequences in which there are at least 2 H monomers exposed in the native state in the form of a model "binding pocket."

Our model also shows induced-fit binding, in which a ligand specifically binds a low energy, but non-native, conformation of the protein. In these cases, the energy price in inducing the conformational change is more than compensated by the energy gain upon ligand binding. We observe induced fit in roughly 17% of HP sequences.

**Modeling hemes and cofactors.**

For some HP sequences that do not fold to unique structures by themselves, a ligand can induce the "selection" of single conformation (see Figure 5). This is a model for proteins that populate a small conformational ensemble in the absence of cofactor, substrate, or prosthetic group, but which become structured in the bound complex. The heme-induced shift from apomyoglobin to myoglobin is an example.

**Cooperative binding.**

Figure 6 shows an example of binding cooperativity between two identical lig-

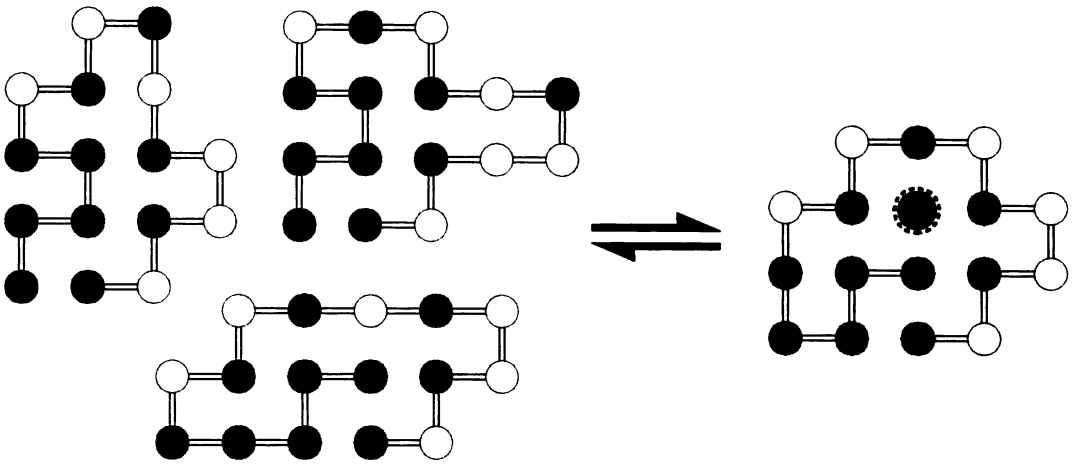

Figure 5: **Model heme binding.** An HP sequence having 14 lowest-energy conformations (only three are shown) is locked into a single lowest-energy structure when the ligand binds.

ands. Binding of the first ligand enhances the binding of the second, resulting in a cooperative binding isotherm. Depending on the HP sequence, binding can have different degrees of cooperativity and may or may not progress through a singly-bound intermediate state.

## (2) UNCONVENTIONAL BINDING BEHAVIOR IN THE MODEL

### Identically-shaped ligands can bind in different modes.

Figure 7 shows one of the most interesting results of the model. It bears on two standard premises: (1) the *steric premise*, that the binding mode is predominantly determined by the shape of the ligand, and (2) the *continuity premise*, that a small change in the structure of a ligand should lead to a small change in the structure of the bound complex. Figure 7 shows a case in which both of these premises are

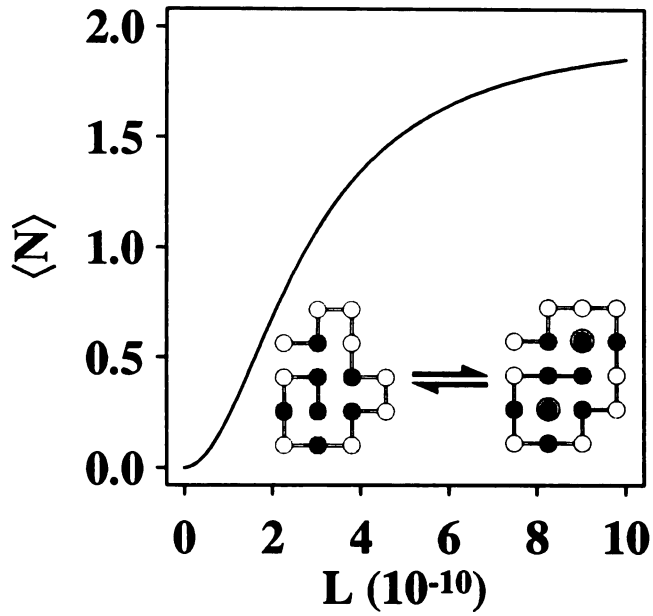


Figure 6: **Binding cooperativity.** Average number of bound ligands, $\langle N \rangle$, vs. ligand concentration. Binding of one ligand facilitates binding of the second. **Inset:** unbound native state and unique excited state with two ligands bound.

violated. Two identically-shaped ligands bind in different locations depending on whether the binding is tight or weak. The tight-binder ($b = 1$) binds to a second-excited state while the lower-affinity ligand ($b = 0.6$) binds to the native state in a lock-and-key fashion.

Since the two ligands have identical shapes, the choice of binding mode in this case is not based on shape complementarity alone, but also on the balance between the energy *lost* in inducing fit, and the energy *gained* in the binding. The tighter binder overcomes an unfavorable distortion of the protein since the resulting binding complex has a lower overall energy. This result resembles a recent experiment by Morton & Matthews (1995), which showed that ligands of very similar shape but different hydrophobicities can change the protein structure in different ways.

We believe there are two implications for drug design. First, even small structural differences between ligands could lead to large differences in the binding mode or

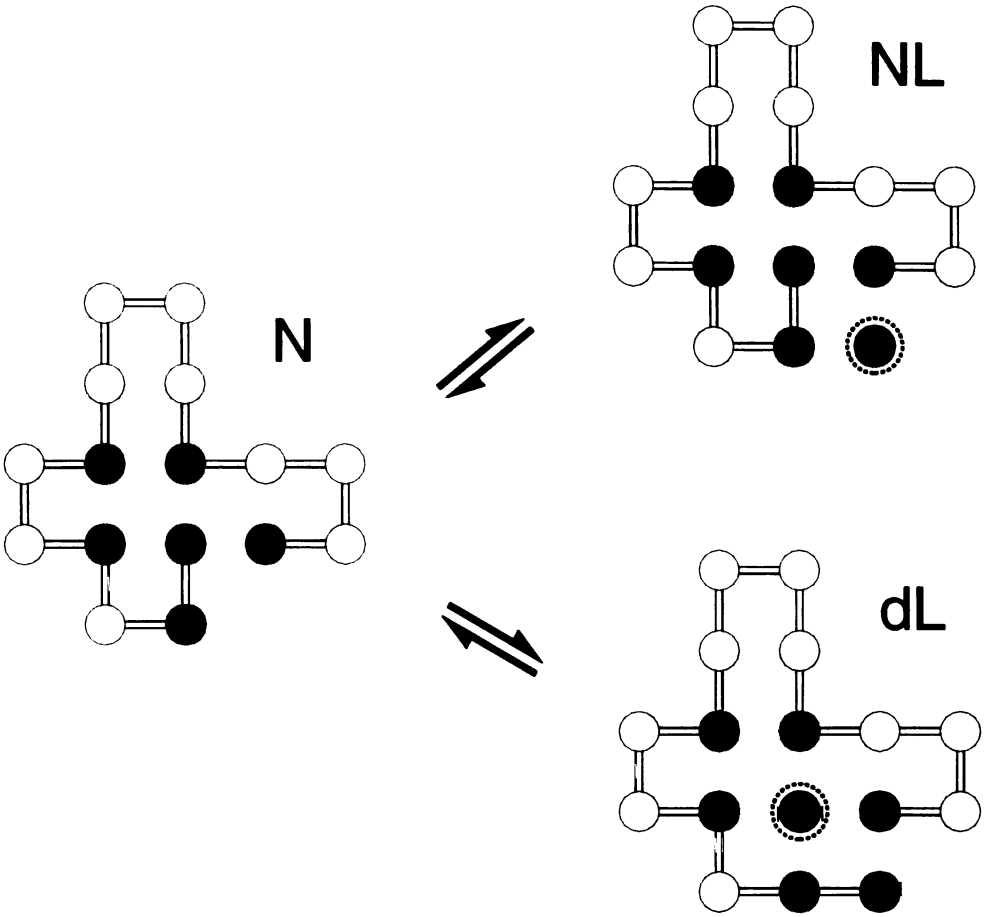


Figure 7: **Binding modes depend on more than ligand shape.** Identically-shaped ligands with different binding constants, $b$, cause different structural changes upon binding. When $b = 0.6$, binding favors the native-state complex, $NL$ (lock and key), while $b = 1$ favors an induced-fit complex, $dL$, where $d$ is a second-excited state. The low-energy complex is determined not by shape complementarity alone, since the two ligands have identical shapes, but by a balance between free energy *lost* in induced-fit structural changes, and free energy *gained* through the protein-ligand interaction.

binding site. This could make the prediction of relative binding affinities sometimes difficult. Second, it implies that one possibly underappreciated determinant of the binding mode of a ligand is not structural. In Figure 7, the change in binding mode can be driven by temperature, or more generally with solvent conditions, both of which regulate the binding strength. A ligand may choose its binding mode based on external conditions rather than on the shapes of the ligand or receptor, and might hop from one site to another when the solvent is changed.

### Ligands can bind tightly to non-native states

Equation 4 describes Michaelis-Menten binding of a ligand to the native conformation, $N$. Now suppose a ligand may interact with *either* the native state or a particular non-native conformation, $d$. The corresponding mass-action equation becomes

$$\begin{array}{ccc} & \overset{K_{bN}}{} & \\ N + L & \rightarrow & NL \\ \downarrow K'_u & & \downarrow \\ & \overset{K_{bd}}{} & \\ d + L & \rightarrow & dL \end{array} \qquad (6)$$

where $K_{bd}$ is the equilibrium binding constant for the non-native conformation, and $K'_u$ is the equilibrium constant for $N$ unfolding to $d$. We use the symbols $d$ and $D$ to represent non-native states: $d$ is one particular non-native conformation (a microscopic state), and $D$ represents the full ensemble of all non- native conformations (the macroscopic denatured state). The native state, too, in reality is an ensemble of microstates, but in the lattice model, we approximate $N$ as a single microscopic conformation.

The fractional population of native conformations bound to ligand, $f_b$, is

$$f_b = \frac{K_{bN}L}{1 + K_{bN}L + K'_u K_{bd} L} \tag{7}$$

When a ligand can bind $d$, the native state can never be fully saturated, since some ligand molecules bind to non-native conformations ($NL$ and $dL$ are assumed to be experimentally distinguishable). However, Equation 7 also shows that even if a ligand binds the non-native state *more tightly* than it binds the native state (i.e., $K_{bd} > K_{bN}$), the effect on the binding curve, $f_b$, can be vanishingly small, provided that the non-native state has a very small population, $K'_u \ll 1$. Thus tight binding doesn't necessarily imply high populations.

Two common objections to the notion of tight binding to non-native states are readily addressed: (1) How could a non-native conformation form a tighter binding site than the active site? The chain could "envelop" the ligand more completely than the native structure does, creating more LH contacts. (2) How could denatured-state binding overcome the unfavorable entropy required to restrict the presumably flexible unfolded conformation to a single, rigid structure in the bound complex? We distinguish between the *macroscopic* denatured state, $D$, and the large number of individual *microscopic* conformations, $d$, which comprise it. Any given denatured conformation $d$ requires no more or less conformational entropy of binding than the native state does, because each is a single conformation. Hence ligand binding to individual non-native conformations need not be intrinsically opposed by conformational entropy.

While Equation 7 gives the mass-action scheme for non-native state binding, it gives no insight into the structural basis for this behavior. Figure 8A illustrates the structural basis using the HP model. The figure shows one example of a non-native state, $d$, of sequence A that binds ligand more tightly than the native state does, forming 3 LH contacts as opposed to only 2 for the native. Yet as shown in Fig-

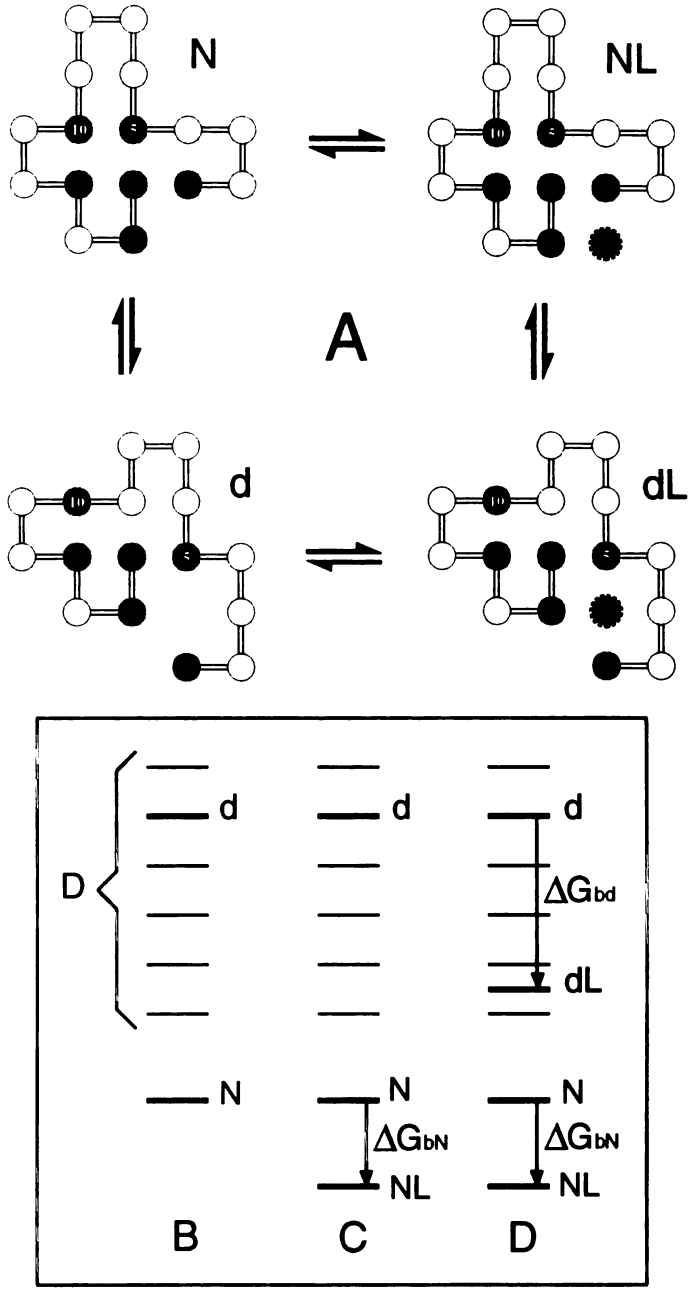


Figure 8: **Denatured-state binding.** **A**: The ligand binds more tightly to a denatured (second-excited) state, $d$, than to $N$, since there are 3 LH contacts in the $dL$ complex vs. only 2 in the $NL$ complex. But $NL$ remains the dominant complex, and binding is Michaelis-Menten, because the cost of unfolding to $d$ is large compared to the binding energy (see Figure 4). **B**: Protein energy ladder, no ligand. **C**: Ligand binds only to the native state; the free energy of binding is $\Delta G_{bN}$. **D**: Ligand binds *d more tightly* than $N$ ($\Delta G_{bd} > \Delta G_{bN}$), as in A. Since the $dL$ complex remains high in energy, it does not affect the binding isotherm.

ure 4, the calculated binding curve indicates binding to only the native conformation. Why?

The reason is illustrated in Figures 8B–D, which show the corresponding free energy diagrams. Figure 8B shows the energy ladder of the protein conformations alone, in the absence of ligand. Figure 8C shows conventional binding, in which the ligand stabilizes the protein by binding the native state. Figure 8D shows the unconventional behavior represented in A: ligand stabilizes the native state, but it stabilizes a particular non-native conformation, $d$, even more. In this case, binding *increases* the population of $d$, but because the $dL$ conformation is so low in population relative to $NL$, the binding isotherm is insensitive to this non-native binding. Under the Michaelis-Menten conditions of Figure 4, all of the possible $dL$ complexes of sequence A have a combined population of only $1 \times 10^{-3}$ at saturating ligand concentration.

### Hydrogen exchange can detect non-native binding

How can tight binding to low-population conformations be detected? Hydrogen exchange (HX) is a technique capable of detecting events of extremely low probabilities in proteins (Woodward & Hilton, 1979; Englander & Kallenbach, 1984; Miller & Dill, 1995; see appendix for brief description). While most experimental measurements of proteins give only ensemble averages, and are therefore overwhelmed by the native signal under native conditions, HX sees no contribution from the native state for protected hydrogens in the cores of proteins. In hydrogen exchange, the detection limit for low-population states is set by the longest exchange time that can be measured, which can be thousands of hours. Hence HX routinely detects structures that are only sparsely populated and cannot be observed by other experimental methods.

When a ligand binds a protein, there are two ways it might affect the exchange rates of amide hydrogens from the hydrophobic core. (1) *Exchange from the unbound state.* If the ligand binds to only the native state (Figure 8C), or at most to only a small number of non-native states, then then there will exist other non-native conformations which *cannot* be populated while the ligand is bound. For the amide hydrogens whose exchange is predominantly from these conformations, the ligand must transiently dissociate from the protein before exchange can occur. The free energy cost for this dissociation is equal to the binding free energy, so the resulting increase in HX rate is $\Delta\Delta G_{hx} = -\Delta G_b$ (see appendix). For a typical protein this represents an increase of several kcal/mol. (2) *Exchange from the bound state.* If the ligand *can* bind non-native conformations (Figure 8D), there will be some amide hydrogens in the protein core that can exchange from *bound* protein structures. Since ligand binding will affect the population of these non-native "exchange" conformations, the corresponding HX rates will also change, by an amount equal to

$$\Delta\Delta G_{hx} = -kT \log \frac{K_{bd}}{K_{bN}} \tag{8}$$

(see appendix) where $K_{bd}$ and $K_{bN}$ are the binding constants for the non-native and native states, respectively (note that $K_{bN}$ approximates the experimentally-measured binding constant, $K_b$, if $D$ is low in population). Equation 8 can therefore be used to calculate the non-native state binding free energy for a given change in the HX rate.

Can HX detect strong interactions between a ligand and a highly non-native protein conformation? Our model predicts three experimental conditions which must be satisfied. (1) Exchange must be slow in the absence of ligand, i.e., $\Delta G_{hx} \approx \Delta G_u$ (see appendix). This suggests that the exchange conformation is significantly non-native in structure. (2) Exchange must occur from the bound state, not the unbound

state. This occurs when the change in HX rate is less than the binding energy, that is, when $\Delta\Delta G_{hx} < -\Delta G_b$. (3) The interaction of ligand with the exchange conformation must be strong. This is evidenced, through Equation 8, by a change in HX rate that is either negative or near zero, the former suggesting even *tighter* binding to the non-native than to the native state.

We use the HP model to simulate the effects of ligand binding on protein HX rates. In this model, the HX rate is proportional to the Boltzmann-averaged solvent accessibility of an amino acid over all possible protein conformations (Miller & Dill, 1995; see appendix). In any one particular protein structure, a monomer is considered to exchange fully with solvent if it is adjacent to a solvent site. It is considered fully protected if it is surrounded on all four sides either by protein monomers or by ligand. Model HX rates are equal to the conformational ensemble average of this solvent accessibility quantity (either 0 or 1). For example, in Figure 8A, monomers 5 and 10 are buried in the native structure, and thus have much slower HX rates than the monomers exposed on the surface. But while monomer 5 is able to exchange from a *first*-excited state (see Figure 1A), monomer 10 is buried in all of the first-excited states (not shown), and so can exchange only from a *second*- or higher-excited state. Hence it exchanges much more slowly than monomer 5.

Figure 9A shows how ligand binding affects HX rates in the model. For monomer 5, the HX rate is decreased when the ligand binds, and the curve of HX rate versus ligand concentration indicates that exchange occurs predominantly from *unbound* conformations. This is because the first-excited states, the most important exchange conformations for monomer 5, do not bind the ligand as tightly as the native state does, making only 1 LH contact, on average, versus 2 for the native state. Hence for this monomer the fastest route to exchange is by dissociation of the ligand, followed by unfolding of the protein. In contrast, the HX rate of monomer 10 *increases* upon

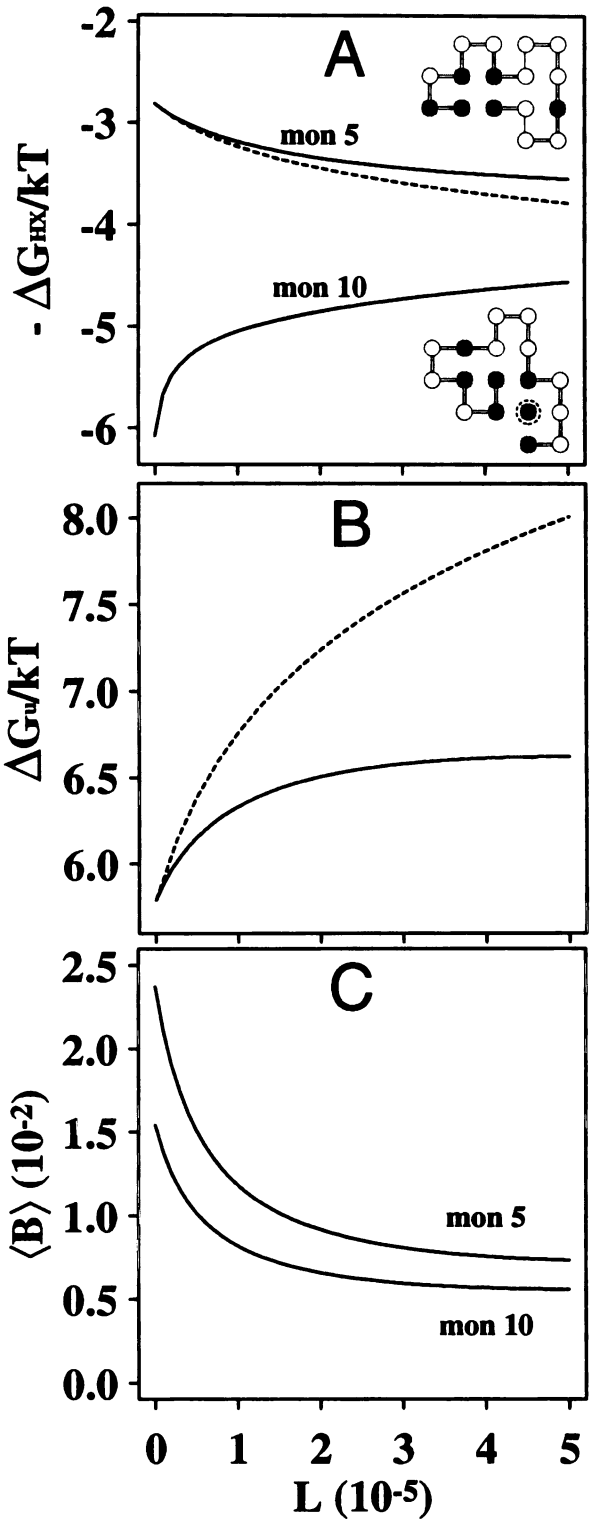


Figure 9: (caption on the following page)

ligand binding. As indicated in Figure 8A, this is due to tight binding to non-native states, in particular a small group of second-excited states. Using Equation 8, we find that the average binding constant, $K_{bd}$, for these second-excited state conformations is roughly 40 times that of the native state. Hence we can predict the binding constants for extremely low-population $d$ states using the measured change in HX rate. (See further discussion in the appendix).

Figure 9: **Effects of binding on HX rates, stability, and flexibility. A**: Binding *increases* the HX rate of monomer 10 by stabilizing *second*-excited states (one shown), which are the fastest-exchanging states for this monomer. Binding *decreases* the HX rate of monomer 5 by stabilizing the *native* state relative to the weak-binding *first*-excited states (one shown), which are the fastest-exchanging states for monomer 5. The HX rate of monomer 5 is close to the theoretical rate (dashed line) corresponding to HX from only the *unbound* states. **B**: Binding stabilizes the native state (increases $\Delta G_u$) because binding is weaker to the full denatured ensemble than to the native state. The dashed line is the expected stability change when *only* the native state is bound. The actual stabilization is smaller because of binding to denatured states. **C**: Average thermal factors, $\langle B \rangle$, decrease with ligand binding for both monomers 5 and 10. The thermal motions depend mainly on first-excited states, which bind ligand weakly, while the HX of monomer 10 depends on second-excited states. For A–C, the model parameters are identical to Figure 4.

## Experimental results from hydrogen exchange

There is experimental evidence from hydrogen exchange that ligands can bind tightly to highly non-native conformations. The HX technique has been used to study the effects of ligand binding on the structure and dynamics of cytochrome c (Paterson et al., 1990), lysozyme (Benjamin et al., 1992), serine-protease inhibitor (Werner & Wemmer, 1992), staph nuclease (Loh et al., 1993), barnase (Meiering et al., 1993), protein G (Orban et al.,1994), and acyl coenzyme A binding protein (ACBP) (Kragelund et al., 1995). In all of these proteins, the general effect of ligand binding is to decrease the HX rates of the majority of amide hydrogens. For many of these hydrogens, binding decreases the HX rate by the maximum amount, indicating

that exchange occurs from the unbound states. Monomer 10 in Figure 9A shows this behavior.

However, the HX rates of many hydrogens either *increase* in the bound conformation, with a typical range in $\Delta\Delta G_{hx}$ of roughly -0.5 to -2.0 kcal/mol, or else they decrease by significantly *less* than the maximum amount. Either of these observations indicates that exchange takes place from *bound* conformations. Among such hydrogens, many are among the *slowest* exchanging in the protein ($\Delta G_{hx} \approx \Delta G_u$), suggesting that exchange may take place from very non-native, highly unfolded conformations. We believe these hydrogens are analogous to monomer 10 in Figures 8 and 9, and may be examples in which the ligand binds a largely unfolded protein conformation with *at least* as high an affinity as the native state.

Are there other explanations for this data? Several possible explanations have been ruled out. First, it is generally found that the increases in HX rate are not due to large structural changes upon binding. In all the examples cited above, there are only very small differences between the bound and unbound structures (from NMR or X-ray crystallography) near the relevant hydrogens. Second, there is no evidence that exchange is increased by bound water molecules trapped at the protein-ligand interface. Third, Benjamin et al. (1992) have argued on the basis of electrostatic calculations that ligand binding does not significantly affect the intrinsic chemical rate, $k_x$, of the exchange process (see appendix). Finally, while it has been argued that HX might occur by "solvent penetration", rather than by structural unfolding (e.g. Woodward & Rosenberg, 1971; Lumry & Rosenberg, 1975; Richards, 1979), and thus that slow exchange might *not* imply highly-unfolded exchange structures, HX experiments in urea (e.g., Hilton et al., 1981; Bai et al., 1994) suggest that for the *slowest-exchanging* hydrogens in the protein core, exchange does involve considerable unfolding of the protein.

Other explanations for increased HX rates upon ligand binding can be found in the hydrogen exchange literature. These involve either: (1) changes in protein structure too subtle to be detected with the measurement techniques used (NMR and X-ray crystallography); (2) a ligand-induced decrease in the "regional stability" of the protein (Kragelund et al., 1995); or (3) an increase in protein dynamics (Benjamin et al., 1992; Kragelund et al., 1995). But there are problems with each of these explanations. For (1), rate increases may be due to subtle changes in *structure* (Equation 7), but they cannot be due to subtle changes in the binding *energetics.* Our model predicts that in order for hydrogens to undergo even a *small* increase in HX rate, the ligand must bind the non-native exchange conformation *at least* as tightly as it binds the native state. The problem with explanation (2) is only that "regional stability" is not very clearly defined. Moreover, *global* stability has been observed to *increase* even when the HX rates of some hydrogens also increase (Loh et al., 1993; Kragelund et al., 1995). Similar behavior arises from our model, as shown below. Explanation (3) appears inconsistent with the poor correlation between changes in HX rates and changes in protein flexibility (Kragelund et al., 1995, using $T_1$ times). As shown below, our model too, predicts a poor correlation.

Our model results are fundamentally different from the previous proposals listed above. Our explanation puts less emphasis on the native state alone, and more emphasis on ligand binding to non-native conformations. Non-native state binding can account for affects on HX rates even when there is little or no detectable change in the protein's native structure.

### How can ligand binding stabilize the native state while increasing the HX rate?

Figure 9B shows that ligand binding stabilizes the native state of sequence A,

while at the same time increasing the HX rate of monomer 10 (Figure 9A). This surprising result has a simple explanation: the ligand affects some few non-native conformations, $d_1, d_2, d_3, \ldots$, differently than it affects the overall denatured ensemble, $D$. In particular, the non-native conformations contributing most to the $D$ ensemble, and thus to protein stability, are the first-excited states, which for this HP sequence bind the ligand less favorably than the native state does. On the other hand, the most important conformations for the HX of monomer 10 are second-excited states, many of which bind ligand *more tightly* than the native state does (Figure 8A). These results are in agreement with experiments from ACBP (Kragelund et al., 1995), which show that both global stability and HX rates can be simultaneously increased by ligand binding, and from early experiments by Woodward (Hilton et al., 1981) showing that HX rates and global stability can be poorly correlated.

**Michaelis-Menten binding can sometimes be destabilizing**

Figure 10 illustrates another unexpected result: that favorable ligand binding can be globally *destabilizing.* Figure 10A shows the binding curve of a particular HP sequence that appears to bind in a lock-and-key fashion to only the native conformation. Figure 10B shows that the same binding event *decreases* the global stability. The reason is that ligand interacts strongly with conformations of the denatured ensemble, thereby increasing the denatured-state population relative to the native. The difference between this and the previous example, where binding *increased* stability, is that here the individual $dL$ conformations *are* highly representative of the full $D$ ensemble, because they are among the first-excited states. Nevertheless, their population *relative to the native state* remains very small. Thus while the fraction of molecules that are denatured changes several fold upon binding, causing a significant decrease in global stability, the fraction itself remains too small to affect the

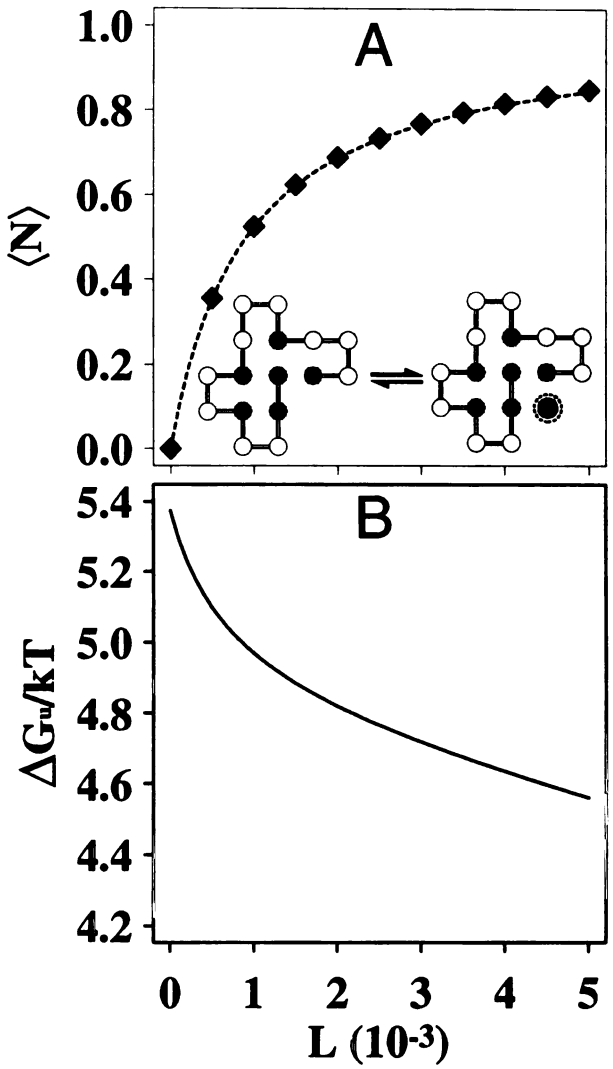


Figure 10: **Michaelis-Menten binding can be globally destabilizing. A**: A sample HP sequence shows lock-and-key, Michaelis-Menten binding (same as Figure 4). **B** Ligand binding *decreases* the the free energy of unfolding, $\Delta G_u$, by stabilizing first-excited states (not shown), and thus increasing the $D$ population relative to $N$. However, since the overall $D$ population remains small (less than 1% in this example), binding to $d$ states has only a small effect on the binding curve. ($b = 0.35$ and $\epsilon = -10$).

binding isotherm (i.e., $K'_u \ll 1$ in Equation 7). This result suggests that observing Michaelis-Menten binding does not necessarily rule out the possibility of highly non-native protein configurations in the complex. A possible example of this behavior is the protein ACBP (Kragelund et al., 1995), which shows an increase in stability much smaller than expected given the strength of the interaction with its inhibitor (see appendix).

### HX rates do not always correlate with protein flexibility

Figure 9C shows that ligand binding can restrict the thermal motions of two residues, even though one residue has an increased HX rate and the other has a decreased HX rate. We calculate the thermal motion in the model as the Boltzmann-average of the fluctuations over all the intra-monomer distances, resembling a B factor in x-ray crystallography (see appendix). B factors are often taken as a measure of the flexibility of the protein, and our model, like experiments, shows a general correlation between thermal motions and HX rates. However, this is not true of the *changes* that occur in these measurements as result of ligand binding. Figure 9C shows that while binding *increases* the HX rate of monomer 10, it *decreases* the B factor. This result has the same explanation as above: B factors reflect ensemble averages over the first-excited states, while the HX of monomer 10 reflects binding to the small subpopulation of second-excited states. While there is often at least some correlation observed between absolute HX rates and protein dynamics, such as in BPTI (Levitt, 1981), ribonuclease A (Wlodawer & Sjolin, 1982), and trypsin (Kossiakoff, 1982), there is at least one observation in which ligand binding *increases* the HX rate while simultaneously *decreasing* flexibility (Kragelund et al., 1995, with ACBP using $T_1$ times), consistent with the model result shown in Figure 9C.

We conclude that HX rates and thermal B factors may not always reflect the

same property of a protein. For the slowest-exchanging hydrogens in particular, the structures that determine HX rates are higher up the energy ladder, while the structures that determine the flexibility are on the first rung. If the ligand can interact differently with the two classes of conformations, then our model suggests that changes in HX rates will not correlate well with changes in flexibility.

## DISCUSSION

Proteins are flexible and only marginally stable. But current models of ligand/protein interactions often assume that proteins are rigid and unperturbable, and that there is a single dominant mode of binding to the native structure. To test the validity of these ideas requires models that go beyond binding polynomials, so that structural consequences of binding are *derived* from principles and structures, rather than *assumed.* Here we present a first step in this direction. Using the 2-dimensional HP lattice model, we search exhaustively through all of protein conformational space, and exhaustively through all possible ligation states of each conformation, to find the global free-energy minimum. In this way, we *compute* the structures and thermodynamics of binding from the protein sequence and the ligand structure, rather than postulating them.

Despite the simplicity of this model, it shows a wide range of protein/ligand binding behaviors. Very hydrophobic ligands at high concentrations denature model proteins, much like urea and guanidine hydrochloride. Weaker ligands bind to compact denatured states, modeling the behavior of ANS. Even weaker ligands bind native states, as dyes and probes do, without perturbing the protein structure. Model ligands can also bind specifically, that is, at single specific sites, and with

Michaelis-Menten (Langmuir) isotherms. Such binding specificity does not require chemical specificity of the underlying interactions, such as hydrogen bonding or ion pairing. The model shows that such binding can be lock-and-key, to a preformed site on the native structure, or induced fit, to an excited-state conformation having a higher energy than the native state. We also find examples of a ligand stabilizing a disordered ensemble, much like heme stabilizes apomyoglobin, as well as examples of binding cooperativity between two ligands.

The model is useful for explaining non-traditional binding behaviors, many of which have been observed experimentally. Two ligands of identical shape can bind in very different binding modes. Ligands can bind denatured states more tightly than the native state and still show Michaelis-Menten binding isotherms. Binding that appears to be "lock and key" may actually destabilize the native structure. Binding may cause model HX rates to increase, while at the same time increasing global stability and decreasing thermal B factors. The model suggests that many of these non-traditional behaviors may have the same physical origin, namely that a ligand can interact differently with a few non-native states than it does with the vast sea of other denatured conformations in general. Experimental evidence of denatured-state binding is found in HX data from barnase, lysozyme, and ACBP.

There are two broader implications for understanding protein/ligand interactions. First, this model suggests that the premise of structure-based drug design—that knowledge of the native structure of a protein is sufficient to rationalize a binding interaction—may not always be true. Knowledge of non-native structures may sometimes be necessary. It suggests that certain information about ligand binding may be hidden and unavailable for structural interpretation by x-ray crystallography and NMR spectroscopy. Second, in the same way that the modeling of protein folding has moved from simple macroscopic mass-action models to more microscopic models

that are rooted in the language of ensembles, energy ladders and energy landscapes (Dill & Chan, 1997), ligand binding too can benefit from more microscopic models and from the language of energy landscapes. Most importantly, proteins are not single conformations. Even under native conditions, proteins populate a broad ensemble of structures. Some of these fluctuations may be important for binding, and may not be well represented by the average properties of the denatured ensemble.

Acknowledgments: We would like to thank NIH for financial support. D.W. Miller has received support from the National Science Foundation Fellowship GER-9255688 (96-97).

## APPENDIX

### Derivation of the partition function

Here we calculate the total partition function for our model of ligand binding, given by the denominator of Equation 3. This requires summing the individual Boltzmann factors for all possible combinations of ligand and model protein. We use "ligation state" to denote a particular combination of ligand and model protein, and "conformation" to denote the HP chain structure alone. Each chain *conformation* has many *ligation states.* A "binding site" is any vacant site adjacent to an H monomer. For a given HP sequence in a given conformation, the total number of ligation states is found by counting the number of ways that indistinguishable ligands can be distributed among the binding sites, beginning with zero ligands and ending with the number that fills all of the sites of the particular conformation.

First we calculate the number of ways that ligands can bind to one particular

conformation, $c$. We define four numbers, $M_i$, $i = 1,2,3,4$, equal to the numbers of binding sites that contact $i$ monomers of type H. The chain conformation in Figure 1C, for example, has $M_1 = 2$, $M_2 = 1$, $M_3 = 1$, and $M_4 = 0$. Now for a particular *ligation state*, $s$, the number of ligands bound to each of the four types of sites is denoted by $n_i$, and can vary from 0 to $M_i$. The particular ligation state in Figure 1C has $n_1 = 1$, $n_2 = 1$, $n_3 = 1$, and $n_4 = 0$. The total number of bound ligands, $N_s$, in the ligation state is obtained by adding the four $n_i$ values:

$$N_s = \sum_{i=1}^{4} n_i \tag{9}$$

There is more than one way that $n_i$ ligands can be distributed among the $M_i$ sites. The total number of combinations, $g_i$, is given by the expression:

$$g_i = \binom{M_i}{n_i} = \frac{M_i!}{n_i!\,(M_i - n_i)!} \tag{10}$$

Similarly, the number of ways that all $N_s$ ligands can be arranged such that $n_1$ of them bind among the $M_1$ single-contact sites, $n_2$ of them bind among the $M_2$ two-contact sites, and so on, is given by the product of four terms:

$$g_{total} = \binom{M_1}{n_1}\binom{M_2}{n_2}\binom{M_3}{n_3}\binom{M_4}{n_4} = \prod_{i=1}^{4}\binom{M_i}{n_i} \tag{11}$$

The partial partition function, $\Xi_c$, for the given HP conformation, $c$, is obtained by summing the Boltzmann factors for each possible ligation state, over all values of $n_i$:

$$\Xi_c = \sum_{n_1=0}^{M_1}\sum_{n_2=0}^{M_2}\sum_{n_3=0}^{M_3}\sum_{n_4=0}^{M_4}\left(\left(\prod_{i=1}^{4}\binom{M_i}{n_i}\right)e^{-h\epsilon/kT}e^{-mb\epsilon/kT}e^{\mu N_s/kT}\right) \tag{12}$$

where $h$ is the number of HH contacts for the conformation; $N_s$ is the total number of ligands bound in ligation state $s$ (Equation 9); and $m$ is the total number of ligand

contacts in the ligation state:

$$m \;=\; n_1 + 2n_2 + 3n_3 + 4n_4 \;=\; \sum_{i=1}^{4} i\, n_i \tag{13}$$

The total partition function, $\Xi$, is found by adding up the partial partition functions for all $\Omega$ possible HP chain conformations:

$$\Xi \;=\; \sum_{c=1}^{\Omega} \Xi_c \tag{14}$$

where $\Omega = 802,075$ in the case of a 16-mer.

**Comparison with the binding polynomial**

Our microscopic binding partition function (Equation 14) can be related directly to the thermodynamic binding polynomial, $Q$. The binding polynomial is related to the free energy of binding, $\Delta G_b$, by (Schellman, 1975):

$$\Delta G_b \;=\; -kT \log Q \tag{15}$$

To compare the partition function with the binding polynomial, it is useful to express Equation 14 in a different form. Written as a product of four sums, Equation 14 becomes

$$\Xi \;=\; \sum_{c=1}^{\Omega} e^{-h\epsilon/kT} \prod_{i=1}^{4} \left\{ \sum_{n_i=0}^{M_i} \binom{M_i}{n_i} \left( e^{-i\,b\epsilon/kT} e^{\mu/kT} \right)^{n_i} \right\} \tag{16}$$

The analogous binding polynomial, $Q$, for macromolecular binding has been derived by Schellman (1975) and is given by

$$Q \;=\; \sum_{c=1}^{\Omega} K_c \prod_{i=1}^{4} \left\{ \sum_{n_i=0}^{M_i} \binom{M_i}{n_i} (K_{bi} L)^{n_i} \right\} \tag{17}$$

where $\Omega$ is the number of molecular conformations, $c$; $K_c$ are the equilibrium constants with respect to the $c = 1$ conformation ($K_1 = 1$); $K_{bi}$ are the binding constants describing the four types, $i$, of binding site; $M_i$ are the number of each type of site; $n_i$ are the number of ligands bound to each type, and $L$ is free ligand concentration.

Comparison of Equations 16 and 17 show the relationships between the microscopic Boltzmann factors and the corresponding thermodynamic quantities:

$$K_c = e^{-\Delta h\epsilon/kT} \tag{18}$$

$$K_{bi} = e^{-ib\epsilon/kT} \tag{19}$$

$$L = e^{\mu/kT} \tag{20}$$

## Hydrogen Exchange

We assume that HX of amide hydrogens from the protein interior is governed by the equilibrium unfolding of the native structure (Englander & Kallenbach, 1984):

$$N(H) \underset{k_f}{\overset{k_u}{\rightleftharpoons}} d(H) \overset{k_x}{\rightarrow} d(D) \tag{21}$$

where $N(H)$ is the native structure; $d(H)$ and $d(D)$ describe an individual denatured conformation prior to and after the exchange of a hydrogen for a deuterium; $k_u$ and $k_f$ are the rate constants for partial unfolding and folding of the native structure; and $k_x$ is the exchange rate for a non-bonded, solvent-exposed hydrogen. The observed rate constant, $k_{obs}$, for this process is (Segal & Harrington, 1967; Roder et al., 1985; Hvidt & Nielsen, 1966)

$$k_{obs} = K\, k_x \tag{22}$$

where $K = k_u/k_f$ is the equilibrium constant for the unfolding process. The observed rate is the product of the "maximum" rate, $k_x$, with $K$, the fraction of time the protein is in the exchanging conformation, $d$. Equation 22 can also be expressed as $\Delta G_{hx}$, the free energy required for the protein to undergo the transition from the native to the exchanging conformation:

$$\Delta G_{hx} = -kT \log K = -kT \log \frac{k_{obs}}{k_x} \tag{23}$$

This free energy is close to zero for surface hydrogens exchanging near the theoretical limit, $k_x$, and reaches a maximum roughly equal to the free energy of global unfolding, $\Delta G_u$, for those hydrogens in the hydrophobic core which require complete unfolding for exchange.

In the HP lattice model, HX rates are calculated from Boltzmann-averaged solvent accessibilities, $A$. $A = 0$ if a monomer is fully protected, i.e., completely surrounded by other monomers or ligands, and $A = 1$ if the monomer is solvent-accessible, i.e., is adjacent to an empty lattice site. The HX rate from the monomer is

$$k = \langle A \rangle \, k_x \tag{24}$$

which is identical to Equation 22 except that the two-state equilibrium constant is replaced by the ensemble-averaged accessibility, $\langle A \rangle$.

### Binding to non-native states

The effects of non-native state binding on global stability, hydrogen exchange, and the binding isotherm can be considered using the simple thermodynamic binding equilibrium:

$$\begin{array}{ccc} & \overset{K_{bN}}{} & \\ N + L & \rightarrow & NL \\ & & \\ \downarrow K'_u & & \downarrow \\ & \underset{K_{bd}}{} & \\ d + L & \rightarrow & dL \end{array} \tag{25}$$

where $N$ and $d$ represent the native state and a particular denatured conformation, respectively; $K_{bN}$ and $K_{bd}$ are the binding constants for $N$ and $d$; and $K'_u$ is the equilibrium constant for $N$ unfolding to $d$.

If the ligand can bind to the native state, $N$, but not to $d$ (i.e., if $K_{bd} = 0$), it follows that:

(1) The fractional population of bound native-state molecules, $f_b$, as a function of ligand concentration is

$$f_b = \frac{NL}{N + d + NL} = \frac{K_{bN}L}{1 + K_{bN}L} \tag{26}$$

where we have assumed $K'_u \ll 1$ in the last step. This is the equation for Michaelis-Menten binding, and the denominator is equal to the Michaelis-Menten binding polynomial, $Q$.

(2) The corresponding change in global stability upon ligand binding, i.e., the change in the free energy, $\Delta\Delta G_u$, is

$$\Delta\Delta G_u = -kT \log \frac{d}{N + NL} + kT \log \frac{d}{N} = -kT \log \frac{1}{1 + K_{bN}L} \tag{27}$$

which is simply the free energy of binding, $-\Delta G_b$, from Equation 15, since $Q = 1 + K_{bN}L$ is the binding polynomial for Michaelis-Menten binding. Thus if ligand binds to only the native state, all of the binding energy contributes toward stabilizing the native relative to the unfolded states.

(3) The change in HX rate, $\Delta\Delta G_{hx}$, of a hydrogen whose fastest route to HX is through conformation $d$ is given by an expression identical to Equation 27:

$$\Delta\Delta G_{hx} = -kT \log \frac{1}{1 + K_{bN}L} = -\Delta G_b \tag{28}$$

Hence if ligand binds only the native state, the HX rate is expected to decrease by an amount equal to the binding energy, $-\Delta G_b$, which is typically several kcal/mol. This is true for *all* hydrogens, regardless of the particular fluctuation $d$ required for exchange.

Conversely, if the ligand *can* bind to the unfolded state, $d$, with a binding constant $K_{bd}$, it follows that:

(1) The fractional population of bound native molecules, $f_b$, becomes

$$f_b = \frac{NL}{N + d + NL + dL} = \frac{K_{bN}L}{1 + K'_u + K_{bN}L + K'_u K_{bd}L} \tag{29}$$

Although $f_b$ in Equation 29 is less than for the Michaelis-Menten case, it reduces to normal Michaelis-Menten binding when the unfolded-state population is very small ($K'_u \ll 1$), even when binding is *stronger* to $d$ than to $N$ ($K_{bd} > K_{bN}$).

(2) The change in global stability will deviate from the Michaelis-Menten case above only if the *single* bound conformation $d$ represents a significant population of the *full* denatured ensemble, $D$. In this case the change in stability is

$$\Delta\Delta G_u = -kT \log \frac{d + dL}{N + NL} + kT \log \frac{d}{N} = -kT \log \frac{1 + K_{bd}L}{1 + K_{bN}L} \tag{30}$$

Equation 30 implies the well-known result that ligands that prefer to bind the native state will stabilize the protein, while ligands that prefer to bind $D$ will destabilize the protein.

(3) If hydrogens can exchange from the bound state, $dL$, then the change in HX rate is given by

$$\Delta\Delta G_{hx} = -kT \log \frac{d + dL}{N + NL} + kT \log \frac{d}{N} = -kT \log \frac{1 + K_{bd}L}{1 + K_{bN}L} \quad (31)$$

This is equal to the stability change of Equation 30, but Equation 31 is valid for *any* denatured conformation, $d$, no matter how low in population. At saturating ligand concentration, Equation 31 becomes

$$\Delta\Delta G_{hx} = -kT \log \frac{K_{bd}}{K_{bN}} \quad (32)$$

which is the result given in Equation 8. If $\Delta\Delta G_{hx}$ is known, Equation 32 can be used to estimate $K_{bd}$, the binding affinity of the ligand for the particular denatured state, $d$. This estimate will be less than the *maximum* $dL$ affinity because Equation 32 assumes that *every* exchanging $d$ state binds ligand equally. For monomer 10 of sequence A, only 3% of the exchanging second-excited states binds ligand with the 3-LH contact affinity shown in Figure 8A, so the estimated affinity represents an average that is less than this maximum 3-contact affinity.

### HP model of protein flexibility

Here we describe our measure of thermal motion, $B$ in the model proteins. For any monomer, $i$, of an HP sequence we can derive a flexibility parameter, $B_i$, by taking all possible conformations and calculating the average deviation of the monomer from its position in the native state, $N$. To illustrate, consider a single non-native conformation, $X$. The "position" of monomer $i$ in this conformation is given by its distance matrix, $d_i(X)$, i.e., the sum of its intra-monomer distances:

$$d_i(X) = \sum_{j \neq i} r_{ij}^2(X) \tag{33}$$

where $j$ is the monomer index and $r_{ij}(X)$ is the distance between monomers $i$ and $j$, in arbitrary lattice units.

The difference in position of monomer $i$ in conformation $X$ and in the native state, $N$, is calculated from a relative distance matrix, $B_i(X)$:

$$B_i(X) = \sum_{j \neq i} \frac{|r_{ij}^2(X) - r_{ij}^2(N)|}{\left(r_{ij}^2(X) + r_{ij}^2(N)\right)/2} \tag{34}$$

where the denominator of Equation 34 is a normalization factor. The flexibility, $B_i$, of monomer $i$ is then calculated from the Boltzmann average of $B_i(X)$ over all possible conformations, $X$.

The quantity $B_i$ represents an *equilibrium* measure of flexibility rather than a *dynamic* one, i.e., flexibility is determined by the populations of low-energy non-native states, not by the heights of kinetic barriers that separate them, which we do not treat here. However, since HX is shown to be an equilibrium process for many proteins, $B_i$ captures the component of flexibility most relevant to HX measurements.

# REFERENCES


Ackers GK, Doyle ML, Myers D, Daugherty MA. 1992. Molecular code for cooperativity in hemoglobin. *Science 255*: 54–63.

Allen KN, Bellamacina CR, Ding X, Jeffery CJ, Mattos C, Petsko GA, Ringe D. 1996. An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem 100*: 2605–2611.

Alonso DOV, Dill KA. 1991. Solvent denaturation and stabilization of globular proteins. *Biochemistry 30*: 5974–5985.

Bai Y, Milne JS, Mayne L, Englander SW. 1994. Protein stability parameters measured by hydrogen exchange. *Proteins Struct Funct Genet 20*: 4–14.

Benjamin DC, Williams DC, Smith-Gill SJ, Rule GS. 1992. Long-range changes in a protein antigen due to antigen-antibody interaction. *Biochemistry 31*: 9539–9545.

Bohacek RS, McMartin C. 1994. Multiple highly diverse structures complementary to enzyme binding sites: results of extensive application of a de novo design method incorporating combinatorial growth. *J Am Chem Soc 116*: 5560–5571.

Chan HS, Dill KA. 1989. Compact polymers. *Macromolecules 22*: 4559–4573.

Chan HS, Dill KA. 1990. Origins of structure in globular proteins. *Proc Natl Acad Sci USA 87*: 6388–6392.

Chan HS, Dill KA. 1991. "Sequence space soup" of proteins and copolymers. *J Chem Phys 95*: 3775–3787.

Chan HS, Dill KA 1994. Transition states and folding dynamics of proteins and heteropolymers. *J Chem Phys 100*: 9238–9257.

Creighton TJ. 1993. *Proteins: structures and molecular properties. 2nd ed.*, pp. 339. New York, New York: W.H. Freeman and Company.

Di Cera, Enrico. 1995. *Thermodynamic theory of site-specific binding processes in biological macro-*

*molecules.* Cambridge, England: Cambridge University Press.

Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*: 7133–7155.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding—A perspective from simple exact models. *Protein Sci 4*: 561–602.

Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nature Struct Biol 4*: 10–19.

Englander SW, Kallenbach NR. 1984. Structural dynamics of proteins and nucleic acids. *Q Rev Biophys 16*: 521–655.

Frauenfelder H, Sligar SG, Wolynes PG. 1991. The energy landscapes and motions of proteins. *Science 254*: 1598–1603.

Hilton BD, Trudeau K, Woodward CK. 1981. Hydrogen exchange rates in pancreatic trypsin inhibitor are not correlated to thermal stability in urea. *Biochemistry 20*: 4697–4703.

Hvidt A, Nielsen SO. 1966. Hydrogen exchange in proteins. *Adv Protein Chem 21*: 287–386.

Koshland DE, Nemethy G, Filmer D. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry 5*: 365–385.

Kossiakoff AA. 1982. Protein dynamics investigated by the neutron diffraction-hydrogen exchange technique. *Nature 296*: 713–721.

Kragelund BB, Knudsen J, Poulsen FM. 1995. Local perturbations by ligand binding of hydrogen deuterium exchange kinetics in a four-helix bundle protein, acyl coenzyme A binding protein (ACBP). *J Mol Biol 250*: 695–706.

Kuntz, ID. 1992. Structure-based strategies for drug design and discovery. *Science 257*: 1078–1082.

Lau KF, Dill KA. 1989. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules 22*: 3986–3997.

Lau KF, Dill KA. 1990. Theory for protein mutability and biogenesis. *Proc Natl Acad Sci USA 87*: 638–642.

Levitt M. 1981. Molecular dynamics of hydrogen bonds in bovine pancreatic trypsin inhibitor protein. *Nature 294*: 379–380.

Loh SN, Prehoda KE, Wang J, Markley JL. 1993. Hydrogen exchange in unligated and ligated staphylococcal nuclease. *Biochemistry 32*: 11022–11028.

Lumry R, Rosenberg A. 1975. The mobile defect hypothesis of protein function. *Colloq Int CNRS 246*: 55–63.

Mattos C, Ringe D. 1996. Locating and characterizing binding sites on proteins. *Nature Biotech 14*: 595–599.

Meiering EM, Bycroft M, Lubienski MJ, Fersht AR. 1993. Structure and dynamics of barnase complexed with 3'-GMP studied by NMR spectroscopy. *Biochemistry 32*: 10975–10987.

Miller DW, Dill KA. 1995. A statistical mechanical model for hydrogen exchange in globular proteins. *Protein Sci 4*: 1860–1873.

Monod J, Wyman J, Changeux JP. 1965. On the nature of allosteric transitions: A plausible model. *J Mol Biol 12*: 88–118.

Morton A, Matthews BW. 1995. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. *Biochemistry 34*: 8576–8588.

Orban J, Alexander P, Bryan P. 1994. Hydrogen-deuterium exchange in the free and immunoglobulin G-bound protein G B domain. *Biochemistry 33*: 5702–5710.

Paterson Y, Englander SW, Roder H. 1990. An antibody binding site on cytochrome c defined by hydrogen exchange and two-dimensional NMR. *Science 249*: 755–759.

Richards FM. 1979. Packing defects, cavities, volume fluctuations, and access to the interior of proteins. *Carlsberg Res Commun 44*: 47–63.

Roder H, Wagner G, Wuthrich K. 1985. Amide proton exchange in proteins by EX1 kinetics: Studies of the basic pancreatic trypsin inhibitor at variable p2H and temperature. *Biochemistry 24*: 7396–7407.

Schellman JA. 1975. Macromolecular binding. *Biopolymers 14*: 999–1018.

Segal DM, Harrington WF. 1967. The tritium-hydrogen exchange of myosin and its proteolytic fragments. *Biochemistry 6*: 768–787.

Semisotnov GV, Rodionova NA, Kutyshenko VP, Elbert B, Blank J, Ptitsyn OB. 1987. Sequential mechanism of refolding of carbonic anhydrase B. *FEBS lett 224*: 9–13.

Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI. 1991. Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers 31*: 119–128.

Shi L, Palleros DR, Fink AL. 1994. Protein conformational change induced by 1, 1'-Bis (4-anilino-5-naphthalenesulfonic acid): Preferential binding to the molten globule of DnaK. *Biochemistry 33*: 7536–7546.

Shoichet BK, Stroud RM, Santi DV, Kuntz ID, Perry KM. 1993. Structure-based discovery of inhibitors of thymidylate synthase. *Science 259*: 1445–1450.

Shortle D, Chan HS, Dill KA. 1991. Modeling the effects of mutations on the denatured states of proteins. *Protein Science 1*: 201–215.

Strynadka NC, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F, Olson A, Duncan B, Rao M, Jackson R, Sternberg M, James MNG. 1996. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. *Nature Struct Biol 3*: 233–239.

Thomas PD, Dill KA. 1993. Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. *Protein Sci 2*: 2050–2065.

Werner MH, Wemmer DE. 1992. Identification of a protein-binding surface by differential amide hydrogen-exchange measurements. Application to bowman-birk serine-protease inhibitor. *J Mol Biol 225*: 873–889.

Wlodawer A, Sjolin L. 1982. Hydrogen exchange in ribonuclease A: neutron diffraction study. *Proc Natl Acad Sci USA 79*: 1418–1422.

Woodward CK, Hilton BD. 1979. Hydrogen exchange kinetics and internal motions in proteins and nucleic acids. *Ann Rev Biophys Bioeng 8*: 99–127.

Woodward CK, Rosenberg A. 1971. Studies of hydrogen exchange in proteins. VI. Urea effects on

RNase hydrogen exchange kinetics leading to a general model for hydrogen exchange from folded proteins. *J Biol Chem 246*: 4114–4121.

Wyman J, Gill SJ. 1990. *Binding and linkage. Functional chemistry of biological macromolecules.* Mill Valley, California: University Science Books.

# Chapter 3

## Computer Simulations of Ligand Binding: Are Non-Dominant Binding Modes Important?

This chapter is taken from the following article: Miller DW, Dill KA. 1997. Computer simulations of ligand binding: are non-dominant binding modes important? *Prot Sci* (submitted).

## ABSTRACT

In computer simulations of protein-ligand interactions, it is common to assume that ligands bind in only a single dominant binding mode. We call this the *single-mode assumption*, and it is the basis for molecular dynamics (MD) simulations of protein-ligand complexes, and for free energy perturbation (FEP) calculations of ligand binding free energies. All-atom models cannot rigorously test this assumption, because they sample conformational space too sparsely. We test the single-mode assumption using the HP lattice model, for which we can explore the full space of protein conformations and ligand binding modes by exact enumeration. We find that even when a wild-type ligand $L_w$ binds in a single mode, a mutant ligand $L_m$ with very similar structure may sometimes require the addition of several *non-dominant* modes to properly account for its binding energetics, even when its *dominant* mode of binding is the same as the wild type's. In such cases FEP theory can sometimes fail to compute the correct difference in binding free energy for the two ligands, and sometimes even the correct rank ordering. We develop a multi-mode FEP theory that may be useful for improving the accuracies of computer simulations of binding energetics.

## THE SINGLE-MODE ASSUMPTION IN LIGAND BINDING

How should computer simulations be used to determine free energies of ligand binding? Following theories of binding and linkage (Cantor & Schimmel, 1980; Wyman & Gill, 1990; Ackers et al., 1992), the rigorous approach would be to determine all possible configurations, free and bound, of the protein and ligand, determine the energy of each configuration, and weight them properly according to Boltzmann's

law. Such an approach should accurately predict equilibrium binding constants.

But given current limitations in computational resources, it is impossible to sample the configurational space of protein-ligand complexes exhaustively. Instead, it is common to make what we call the "single-mode" binding assumption, namely that ligands bind to proteins in only a single dominant mode, such as those observed in x-ray crystallography or NMR experiments. (Here the term "mode" denotes a collection of protein-ligand configurations of highly similar structure. This is discussed in more detail below and in the appendix.) According to the single-mode assumption, any "non-dominant" binding modes can be assumed to make negligible contributions to the binding energetics. Hence in order to achieve adequate conformational sampling of a protein-ligand complex, it is assumed necessary only to perform a sufficient search *within* the dominant mode, while all other weak binding modes can be neglected.

The single-mode assumption is the basis for both structure-based drug design methods and free energy perturbation (FEP) calculations. To illustrate the former, we consider a ligand $L_w$ that binds with high affinity to a protein $P$. Suppose that in the atomic structure of the complex, $L_wP$, the ligand is found to bind in a single conformation at a single site. Often such information is then used to design a second ligand, $L_m$, that is similar in structure to $L_w$. The underlying assumption is that $L_m$ (the "mutant" ligand) will bind in a single mode that is identical to the mode of the "wild-type" ligand $L_w$. The binding energetics of the two ligands are thus presumed to be very similar.

The free energy perturbation (FEP) method is then often used to calculate the relative binding free energies of the two structurally similar ligands (see Postma et al., 1982; Tembe & McCammon, 1984; Bash et al., 1987; Singh et al., 1987; Kollman, 1993). Calculations are based on the following thermodynamic binding cycle:

$$\begin{array}{ccc} & \Delta G_b(w) & \\ P + L_w & \rightarrow & PL_w \\ \Delta G_L \downarrow & & \downarrow \Delta G_{PL} \\ & \Delta G_b(m) & \\ P + L_m & \rightarrow & PL_m \end{array} \qquad (1)$$

where $P$ is protein, and $L_w$ and $L_m$ are two structurally similar ligands. The free energies of binding for the two ligands are $\Delta G_b(w)$ and $\Delta G_b(m)$, while $\Delta G_L$ and $\Delta G_{PL}$ are the free energies of mutating the ligand, free and bound to the protein, respectively, from the wild-type to the mutant structures. The difference in binding free energy for the two ligands, $\Delta\Delta G_b$, can be calculated using Equation 1:

$$\Delta\Delta G_b \;=\; \Delta G_b(m) - \Delta G_b(w) \;=\; \Delta G_{PL} - \Delta G_L \qquad (2)$$

The "mutation" free energies $\Delta G_{PL}$ and $\Delta G_L$, though corresponding to non-physical phenomena, can be calculated using FEP. This requires two molecular dynamics (MD) simulations on the wild-type ligand: one on the ligand bound to the protein and one on the ligand free in solution. For each time frame of the simulation, the coordinates and energy parameters are changed to those of the mutant ligand, and the difference in energy, $\Delta U$, is calculated. The mutation free energy, either $\Delta G_{PL}$ or $\Delta G_L$, is then readily calculated using:

$$\Delta G \;=\; -kT \log\langle e^{-\Delta U/kT}\rangle \qquad (3)$$

where the brackets indicate a Boltzmann average over the full simulation. Equations 2–3 are then used to calculate the relative binding free energy.

The FEP method, like structure-based design, assumes that the mutant ligand binds in a single mode, and that the wild-type and mutant modes are identical. On

this basis, it is assumed that an MD simulation of the *wild-type* complex will sample all configurations important to the *mutant* complex, and will therefore give an accurate prediction of the relative binding free energies for the two ligands. Simulations are therefore restricted to the crystallographic binding mode, that is, to conformations representing only small perturbations from the known crystal structure of the wild-type complex. And although considerable effort is made to sample sufficiently *within* the dominant mode — such as by making the ligand mutation in several small steps (see Pearlman & Kollman, 1989; Kollman, 1993) — this is done only to improve the convergence of the simulation, and not to explore the possibility of other important modes of binding.

Is the single-mode assumption valid? To test the assumption requires a model in which the full space of ligand and protein conformations can be explored exhaustively. Otherwise, errors due to the single-mode assumption cannot be unambiguously distinguished from those due to incomplete conformational sampling. For this reason, we use the HP lattice model of proteins, rather than all-atom simulations or other more detailed models. The virtue of the simplifications is that we can explore all conformations of the model protein and all possible binding modes with rigor and completeness, so we can determine the validity of the single-mode assumption unambiguously. Moreover, the simplifications are not limitations. The lattice model is simply an illustration of principle, and the issues it raises pertain to real proteins, irrespective of the geometric simplifications the model entails. A benefit of the present study is that it points the way to practical fixes for problems in more realistic model simulations.

From the lattice model simulations presented here, we observe the following principle. Even when a mutant ligand has the same dominant binding mode as a wild-type ligand (so that the bound complexes may be very similar), the weaker,

*non-dominant* modes of the mutant ligand can still be a major determinant of the differences in binding energetics between the two ligands. Thus calculations that neglect non-dominant binding modes can predict incorrect free energy differences for the two ligands, and in some cases even an incorrect *rank ordering* of the binding constants. Using the HP model to simulate FEP experiments, we show a way to repair these problems in all-atom simulations.

## HP MODEL EXAMPLE

We model ligand binding using a variation of the two-dimensional HP lattice model of proteins. The binding model is described in Miller & Dill (1997). The HP model is described in (Lau & Dill, 1989, 1990; Chan & Dill, 1991; Dill et al., 1995). A protein is represented as a sequence of H (hydrophobic) and P (other) monomers on a two-dimensional lattice. Lattice sites may be either empty or filled by a single monomer, and empty lattice sites are assumed to contain a solvent molecule. Each HH contact, formed when two non-sequential H monomers occupy adjacent lattice sites, is favored by a contact free energy $\epsilon$ ($\epsilon < 0$), meant to capture the importance of hydrophobic interactions in protein collapse and global stability (Dill, 1990). Hence the free energy of a conformation is $h\epsilon$, where $h$ is the number of HH contacts. The magnitude of $\epsilon$ determines the degree of stability imparted by external conditions: large and negative $\epsilon$ reflects conditions that are more stabilizing, such as lower temperature or lower denaturant concentrations. Conformational entropy, the driving force for unfolding, enters the model through the exhaustive enumeration of all the possible chain configurations (see below).

The disadvantages of the model are clear: atomic resolution is lost; conformations

are restricted to a lattice; it is in 2 dimensions; the energy function is simplified; and chains are unrealistically short. Yet despite these disadvantages, the model has been found useful for modeling protein properties (Lau & Dill, 1989, 1990; Chan & Dill, 1989, 1990, 1994; Shortle et al., 1991; Dill et al., 1995; Miller & Dill, 1995, 1997) because it shows several protein-like features, including cooperative collapse, native structures having a nonpolar core and definable secondary structures, multi-stage folding kinetics, and molten globule states. Most importantly, we believe the model captures the main physics of protein folding—the hydrophobic interactions, conformational freedom of the chain, and the steric restrictions imposed by excluded volume.

We study HP sequences having 16 monomers. For any 16-mer chain there are exactly 802,075 possible conformations that can be configured on a two-dimensional lattice. These conformations are generated by computer, and each is weighted by a Boltzmann factor according to the number of HH contacts made. Figure 1A shows an energy diagram for a sample HP sequence (called "sequence A"). The native structure (ground state) is the conformation with the largest possible number of HH contacts, and thus the lowest contact free energy. We study only non-degenerate sequences, i.e., those having a single native conformation, since we believe they best represent biological proteins, which fold to unique structures. All higher-energy conformations comprise the non-native, or denatured states, and are grouped by energy into "first-excited" states, "second-excited" states, etc., corresponding to successively fewer HH contacts. For any HP sequence, there are far more open, high-energy conformations than compact, low-energy conformations (see Figure 1B).

### Modeling protein-ligand interactions

We model ligands as single, monomer-sized beads (Figure 1C) (Miller & Dill,

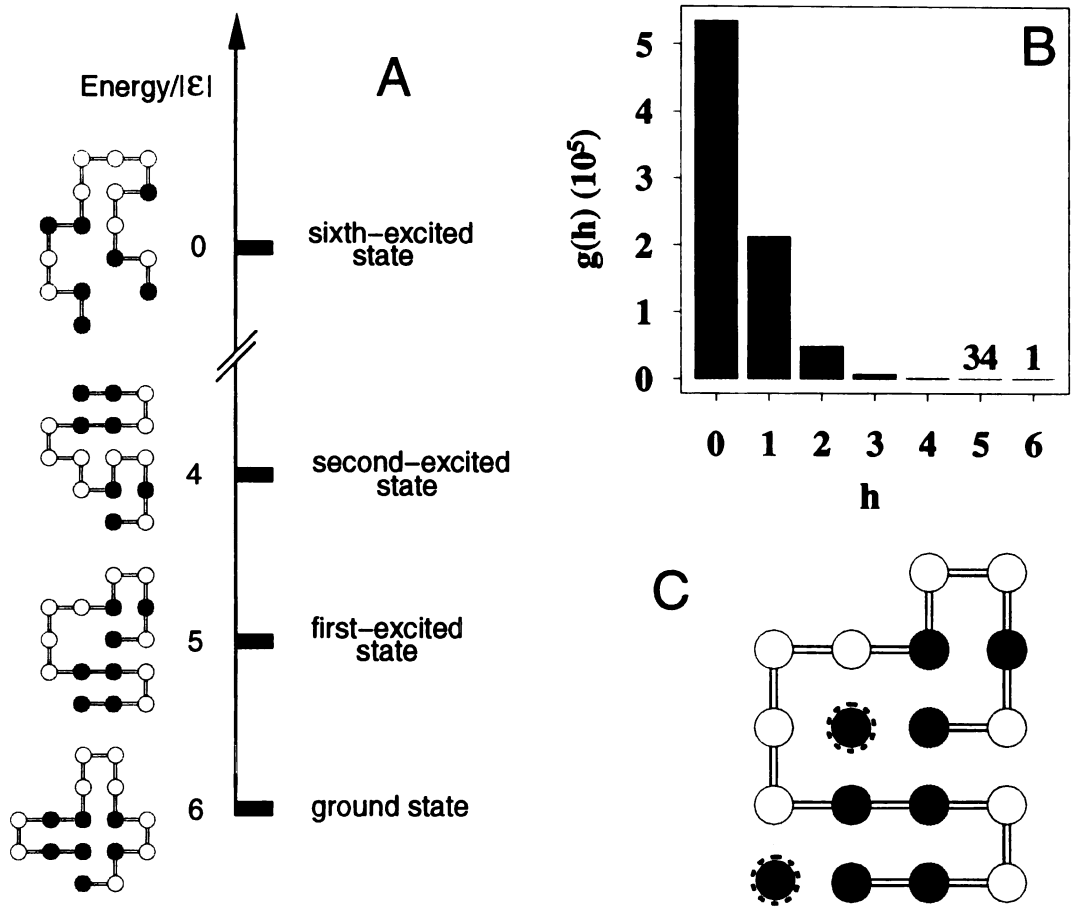


Figure 1: **Energy ladder representation of protein conformations.** Energy ladder diagram for the 16-mer HP sequence A, shown in its single native conformation at the bottom. The native conformation has 6 HH contacts (H monomers are colored black), while each step up the ladder represents one fewer HH contact. Example conformations are shown. **B**: Density of states, $g(h)$: the numbers of conformations of this HP sequence that have $h$ HH contacts. **C**: Sequence A in a given ligation state, having 2 ligands bound. Each contact between a ligand and H monomer (LH contact) is favored by a free energy $b\epsilon$, where $b$ is a positive constant, and $\epsilon$ is the HH contact energy. Each LP contact is favored by a free energy $E_p$. In this ligation state, 2 ligands make 3 LH contacts and 3 LP contacts. Since $h = 5$, the energy is $5\epsilon + 3b\epsilon + 3E_p$ (Equation 4).

1997). A protein-ligand contact occurs when the ligand occupies a lattice site adjacent to a chain monomer. Ligands may interact with both hydrophobic and polar groups: a contact between a ligand and H monomer (LH contact) is favored by a free energy $b\epsilon$, where $b$ is a positive constant ($0 < b < 1$) and $\epsilon$ is the HH contact energy. Polar (LP) contacts are favored by a free energy $E_p$ ($E_p < 0$). In order to have the simplest possible model of binding, we assume the interaction energy is zero between ligands. The total contact energy, $E_s$, for any protein-ligand configuration ("ligation state"), $s$, is therefore

$$E_s = h\epsilon + m_h b\epsilon + m_p E_p \tag{4}$$

where $m_h$ and $m_p$ are the total numbers of LH and LP contacts, respectively. The ligation state shown in Figure 1C represents 5 HH contacts, 3 LH contacts, and 3 LP contacts, so the total energy is $5\epsilon + 3b\epsilon + 3E_p$.

The fractional population, $P_s$, of any protein-ligand ligation state is given by the grand canonical distribution function:

$$P_s = \frac{e^{-E_s/kT} e^{\mu N_s/kT}}{\sum_{s=1}^{\Gamma} e^{-E_s/kT} e^{\mu N_s/kT}} \tag{5}$$

where $T$ is absolute temperature, $\mu$ is the ligand chemical potential (related to ligand concentration by $L = e^{\mu/kT}$), $E_s$ is the total energy of the ligation state (structural energy plus binding energy), $N_s$ is the number of ligands bound to the protein, $k$ is Boltzmann's constant, and $\Gamma$ is the total number of ligation states available to the protein-ligand complex. By exact enumeration of all possible ligation states (also referred to as "binding modes"), Equation 5 can be used to calculate average properties of protein-ligand complexes, as well as absolute free energies of binding. This is described in the appendix.

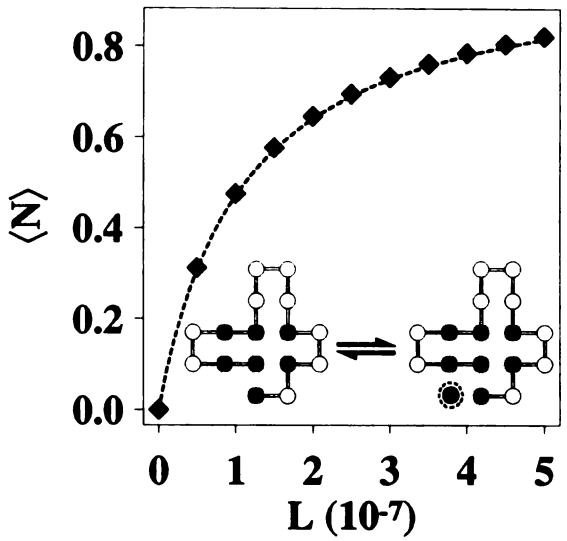


Figure 2: **Lock-and-key binding.** For the wild-type ligand ($b = 0.8$, $E_p = 0$), the average number of bound ligands, $\langle N \rangle$ (diamonds), vs. ligand concentration, $L = e^{\mu/kT}$, closely follows the theoretical curve (dashed line) expected for Michaelis-Menten (lock-and-key) binding to the single site shown. In this example $\epsilon = -10$.

For varying binding strengths, ligand concentrations, and temperatures, the model shows a range of behaviors that mimic real protein-ligand interactions, including lock-and-key and induced-fit mechanisms of specific binding, ligand-induced denaturation, ANS and heme-like binding, and cooperative binding between several ligands. Details are given in Miller & Dill (1997).

### Examples of non-dominant binding modes

Figure 2 shows the interaction between an HP sequence, which we call sequence A, and a wild-type ligand, $L_w$. In this example the wild-type ligand is hydrophobic: LH interactions are favorable ($b = 0.8$), while LP interactions have zero energy ($E_p = 0$). The protein-ligand complex shown in the figure is at the global minimum in free energy relative to all possible conformations of the protein and all possible binding configurations. We call this two-state lock-and-key binding because: (1) it involves a specific geometric site on the unperturbed native structure

of the protein, (2) the binding isotherm closely follows the theoretical curve describing pure Michaelis-Menten binding to a single site, and (3) the population of all other binding modes is negligible throughout the binding isotherm (see Figure 3B).

Figure 3 shows how changing the chemistry of the ligand can change its binding energetics via the population of non-dominant binding modes. Figure 3A shows the dominant (native) binding mode, $P_0$, and two of the many non-dominant modes, $P_1$ and $P_2$, for HP sequence A complexed with ligand. For the *wild-type* ligand, the fractional populations of these non-dominant modes are negligible (Figure 3B): each contributes only $4.45 \times 10^{-5}$ of the total population at saturating ligand concentrations. But Figure 3C shows that for a *mutant* ligand that differs only slightly from the wild type, the non-dominant modes can become very important. In this example the mutant ligand is defined as having a slightly weaker hydrophobic interaction ($b = 0.79$ compared to $b = 0.8$ for the wild type) but a strong polar interaction ($E_p = -4.5$ compared to $E_p = 0$ for the wild type). For the mutant ligand, the non-dominant modes comprise roughly 40% of the total population, while the native contributes 56%. In fact, summed over all possible binding configurations, the mutant ligand binds the protein *more tightly* than the wild-type ligand does, with a free energy difference $\Delta\Delta G_b = -0.37kT$ (see appendix). Thus even though the mutant ligand binds the native conformation $P_0$ *less* favorably (since $b$ has decreased), it binds more tightly overall, via interactions with the non-native conformations $P_1$ and $P_2$.

### Single-mode FEP can yield errors in predicted binding constants

If an FEP calculation begins in the dominant wild-type configuration $P_0L_w$, it may sample configurations $P_1$ and $P_2$ too sparsely to correctly predict the relative binding constants of the two ligands. To test this assertion, we mimic MD FEP

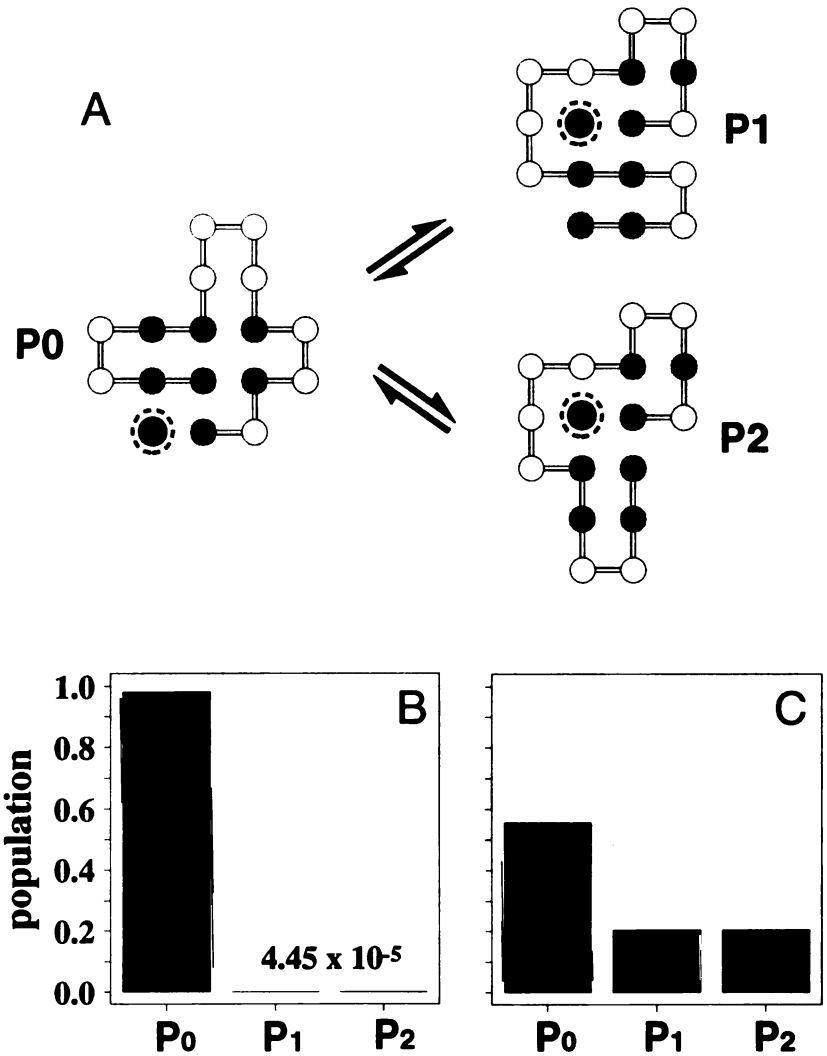


Figure 3: **The mutant ligand binds in non-dominant modes. A**: The dominant binding mode, $P_0$, of sequence A with ligand, and two of the many possible "non-dominant" modes ($P_1$ and $P_2$). **B**: Fractional populations of $P_0$, $P_1$, and $P_2$ for the wild-type ligand ($b = 0.8$, $E_p = 0$). **C**: Populations for the mutant ligand ($b = 0.79$, $E_p = -4.5$). The non-dominant modes $P_1$ and $P_2$ are negligibly populated for the wild-type ligand, but have high populations for the mutant ligand, due to its polar interactions.

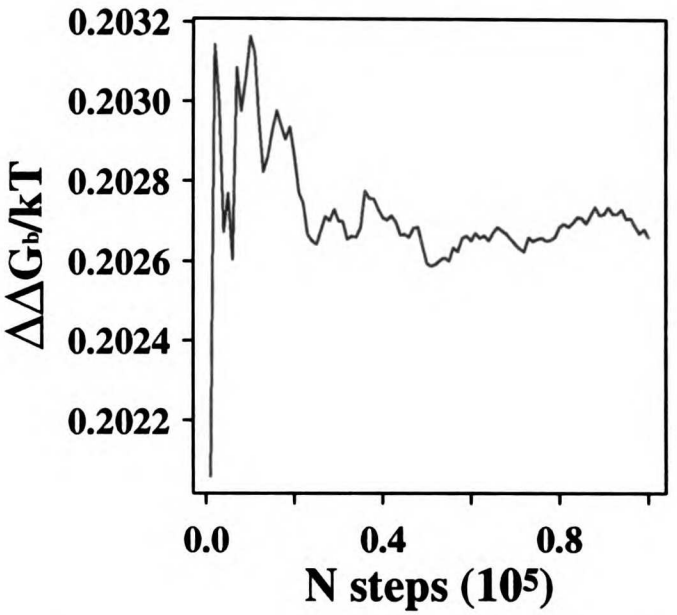


Figure 4: **Relative free energy calculation by traditional FEP.** Free energy difference $\Delta\Delta G_{b0}$ of binding to HP sequence A (wild-type versus the mutant ligand), versus number of Monte Carlo time steps. The curve shows an average of 10 simulations, each calculated using the HP model of free energy perturbation (FEP) according to Equation 3. The actual free energy difference is $-0.37kT$. Despite the obvious convergence of the simulation, sampling within only the native mode leads to an incorrect rank ordering of the binding free energies.

simulations using Monte Carlo, as described in the appendix. Simulations are begun with the HP protein and wild-type ligand bound in the native configuration ($P_0$), and small, random movements of the chain are either accepted or rejected based on the resulting changes in energy. For each structure along the simulation trajectory, we mutate the ligand, calculate the resulting change in energy, and compute a free energy using Equation 3.

Figure 4 shows the time course (in Monte Carlo steps) for an HP model FEP calculation. We chose the number of steps to be large enough that the simulation would converge, but small enough that it would mimic the incomplete sampling of conformational space by MD methods. For the wild-type/mutant ligand pair $L_w$ and $L_m$, the converged free energy difference is $\Delta\Delta G_b = 0.203kT$, an excellent

approximation to $0.2kT$, the value that would be expected if *only* the dominant mode $P_0$ were important (since from Equation 4, $\Delta E_s = m_h \Delta b \epsilon$, with $m_h = 2$, $\Delta b = -0.01$, and $\epsilon = -10$). In fact, the non-native structures $P_1$ and $P_2$ are never sampled in the simulation, and only very small perturbations around the native configuration are explored. As a result, the predicted free energy difference is wrong, as is the *rank ordering* of the binding constants: the wild-type ligand is incorrectly predicted to bind more tightly than the mutant.

Hence this example illustrates a failure of the single-mode assumption, and shows that traditional FEP simulations may give significant errors by insufficiently sampling non-dominant modes that are crucial for an accurate representation of the mutant complex. Moreover, this example shows that the *convergence* of an FEP simulation — ordinarily taken as a sign that equilibrium has been reached — does not guarantee that the single-mode assumption is valid. In this example the simulation clearly converges, but to an incorrect value of the free energy.

## THE MULTI-MODE THEORY FOR LIGAND BINDING

Since the HP lattice model can unambiguously identify failures of the single-mode binding assumption, it is also a testing ground for proposed fixes. Clearly, the problem with the single-mode assumption is its neglect of binding modes other than the dominant one. We propose a way to compute more accurate differences in binding free energies, in two steps. First, a rapid simulation is performed to identify potentially important binding modes, for example by taking several possible conformations from a DOCK algorithm search (Kuntz et al., 1982). Second, binding free energies are computed using these few sampled binding modes, in conjunction

with the multi-mode FEP theory described below. If the sampling is very poor or incomplete, this will lead, at worst, to the usual single-mode model of binding, but if any important modes are sampled, then the binding free energies should be more accurate than those resulting from traditional FEP methods.

One such fast search strategy (step 1) has been described by Wilson et al. (1991) for ligands of alpha-lytic protease. To identify non-dominant binding modes, Wilson et al. systematically rotated each side chain in the binding site through its lowest-energy rotomers, and calculated a binding free energy for each conformation using an empirical free energy function. In this case the terms were combined to give an overall binding energy, but they could also be evaluated separately to identify which binding modes make the largest contributions.

Here we derive the *multi-mode* binding theory (step 2). For a protein-ligand interaction involving several independent binding modes, 0, 1, ..., where 0 is the dominant binding mode and 1 is the first non-dominant mode, etc., the free energy of binding is given by

$$\Delta G_b \;=\; -kT \log \left( e^{-\Delta G_{b0}/kT} + e^{-\Delta G_{b1}/kT} + \cdots \right) \tag{6}$$

where $\Delta G_{b0}$, $\Delta G_{b1}$, ..., represent free energies of binding in the individual modes (see appendix).

The difference in binding energies, $\Delta\Delta G_b$, for two different ligands is then

$$\Delta\Delta G_b \;=\; \Delta\Delta G_{b0} - kT \log \frac{1 + e^{-\Delta\Delta_1 G_b(m)/kT} + \cdots}{1 + e^{-\Delta\Delta_1 G_b(w)/kT} + \cdots} \tag{7}$$

where $\Delta\Delta G_{b0}$ is the difference in the free energy of binding in the dominant mode, 0, and

$$\Delta\Delta_1 G_b(w) \;=\; \Delta G_{b1}(w) - \Delta G_{b0}(w) \tag{8}$$

$$\Delta\Delta_1 G_b(m) \;=\; \Delta G_{b1}(m) - \Delta G_{b0}(m) \tag{9}$$

are the free energy differences for the wild-type and mutant ligands, respectively, for binding in mode 1 versus mode 0. The first term in Equation 7 can be calculated by FEP using the traditional binding cycle in Equation 1. Likewise, the two terms in Equations 8–9 can be calculated by FEP using the following cycle (shown for wild-type only):

$$\begin{array}{ccc} & \Delta G_{b0}(w) & \\ P_0 + L_w & \rightarrow & P_0 L_w \\ \Delta G_P \downarrow & & \downarrow \Delta G_{PL} \\ & \Delta G^*_{b1}(w) & \\ P_1 + L_w & \rightarrow & P_1 L_w \end{array} \tag{10}$$

where there are now two binding modes represented in Equation 10 by the two protein structures, $P_0$ and $P_1$. (The different modes could also be different orientations of the *ligand* within a *fixed* binding site). Since $P_0$ represents the dominant, lowest-energy mode, the total free energy of binding, $\Delta G_{b1}$, in the non-dominant mode $P_1$ is the free energy for the transition from $P_0$ to $P_1$ in the absence of ligand, plus the free energy of binding to $P_1$ after the transition has occurred. In Equation 10 these free energies are $\Delta G_P$ and $\Delta G^*_{b1}$, respectively. Thus the free energy difference in binding the wild-type ligand to $P_1$ versus $P_0$ is, according to Equation 10

$$\Delta\Delta_1 G_b(w) \;=\; \Delta G_{PL} \tag{11}$$

The free energy $\Delta G_{PL}$ can be calculated using FEP, where the mutation now involves not a *chemical* change, but rather a change in *protein structure* — such as the rotation of a side chain — from one binding mode structure to another, while the ligand remains bound in the site. Hence one additional MD simulation is required

| Number of non-dominant modes | $\Delta\Delta G_b$ (est) (kT) | $\Delta\Delta G_b$ (true) (kT) | error (%) |
|---|---|---|---|
| 0 | 0.203 | -0.367 | 155.3 |
| 1 | -0.110 | -0.367 | 70.0 |
| 2 | -0.348 | -0.367 | 5.1 |

Table 1: **Correction using multi-mode FEP theory.** Results of HP model FEP calculations when: only the dominant mode ($P_0$), the dominant and one non-dominant mode ($P_1$), and the dominant and both non-dominant modes ($P_1$ and $P_2$), are used in the calculation. Contributions from non-dominant modes are calculated using Equation 10. Including one non-dominant mode recovers the correct rank ordering of the wild-type and mutant binding constants; including both gives the correct prediction of the relative binding energies to within 5% error.

for each non-dominant binding mode. Below we illustrate this method using the HP model example.

**Illustration of the multi-mode FEP approach in the HP model.**

We now repeat the HP model FEP calculation already described, but we include contributions from non-dominant modes. Now at least two simulations are required, each starting with the native configuration, but each now involving a different type of mutation. In the first simulation, used to calculate the dominant-mode contribution $\Delta\Delta G_{b0}$, the mutation is from the wild-type ligand to the mutant ligand, as shown in Equation 1 and Figure 4. To compute the $\Delta\Delta_i G_b$ term for each non-dominant mode (Equations 8–11), a separate simulation is run in which the mutation is from the native protein structure to the structure that represents non-dominant mode $i$. The total free energy difference is then obtained using Equation 7.

Table 1 shows the results of including the non-dominant modes $P_1$ and $P_2$ (from Figure 3) in the FEP calculation. Table 1 shows that when both modes are included, the predicted free energy difference improves dramatically, to within roughly 5% of

the correct value, and that the proper rank ordering of the two binding constants is recovered. Though not shown in the table, the inclusion of additional non-dominant modes makes a negligible difference in the free energy calculations, as expected, since only the three modes shown in Figure 3 make a detectable contribution to the Boltzmann ensembles of the wild-type and mutant ligand complexes.

## DISCUSSION

In modeling protein-ligand interactions by computer, it is common to assume that ligands bind in only a single most important mode. We call this the *single-mode binding assumption*, and it is the basis for rational drug-design efforts and free energy perturbation (FEP) calculations of relative free energies of binding. According to the single-mode assumption, if a wild-type ligand binds in a single dominant mode (as evidenced by x-ray crystallography or NMR), then a mutant ligand of similar structure will also bind in a single mode, and the mutant and wild-type modes should be identical.

The single-mode assumption neglects other "non-dominant" binding modes that may contribute substantially to the binding energetics of the mutant ligand, even if the *dominant* binding mode of the two ligands is the same. Examples of non-dominant modes may include different positions or conformations of the ligand within the same rigid binding site, or different configurations or locations of the site itself. Traditional FEP methods may be unable to predict accurate free energies in cases where non-dominant modes are important. Although MD simulations do sample a range of configurations, it is a very narrow range, typically involving deviations from the native complex of only fractions of an angstrom. Even small

perturbations, such as the rotations of protein side chains within the binding site, may not be well sampled during the course of an FEP calculation. And the artificial constraints that are often imposed in computer simulations even further reduce the sampling of non-dominant binding modes.

We describe a lattice model test that indicates how non-dominant binding modes can contribute significantly to the binding energetics. Even when a mutant ligand has the same dominant binding mode as a wild-type ligand, the proper signs and magnitudes of the relative binding energies may require accounting for multiple non-dominant modes that would not be sampled using traditional FEP methods. This is not an issue of convergence or sampling hysteresis. A perfectly converged simulation can converge to the wrong free energy if all of the relevant binding modes are not properly sampled. To correct these errors, we present a method to search for putative non-dominant modes, and to account for them in a multi-mode FEP approach. We show that including these modes may lead to improved estimates of binding free energies by computer simulation.

Acknowledgments: We thank the NIH for financial support. D.W. Miller has received support from the National Science Foundation Fellowship GER-9255688 (96-97).

## APPENDIX

### Derivation of Binding Free Energies

Schellman (1975) has derived the free energy of ligand binding to a protein having independent, non-interacting sites labelled $0, 1, \ldots$ :

$$\Delta G_b = -kT \log (1 + K_0 L + K_1 L + \cdots) \tag{12}$$

where the individual equilibrium binding constants, $K$, capture both the energetics of the protein-ligand interaction and any change in protein conformation caused by ligand binding. When the ligand is at saturating concentrations (this is normally the case in experiments, and is implicit in MD simulations), Equation 12 becomes

$$\Delta G_b = -kT \log (K_0 L + K_1 L + \cdots) \tag{13}$$

or, at $L = 1M$ standard state,

$$\Delta G_b = -kT \log (K_0 + K_1 + \cdots) \tag{14}$$

Equation 14 can also be written

$$\Delta G_b = -kT \log \left( e^{-(kT \log K_0)/kT} + e^{-(kT \log K_1)/kT} + \cdots \right) \tag{15}$$

or, in terms of free energies,

$$\Delta G_b = -kT \log \left( e^{-\Delta G_{b0}/kT} + e^{-\Delta G_{b1}/kT} + \cdots \right) \tag{16}$$

where $\Delta G_{b0}$ and $\Delta G_{b1}$ represent the free energies of binding in the individual modes. Although Equation 15 is rigorous, the $\log K$ terms can be written as free energies only if the set of microscopic states that defines each binding mode is independent of every other mode on the experimental time scale, i.e., if the rate of transition between modes is slow. In this case the individual binding modes are distinct "binding sites," insofar as they are independent and non-interacting.

However, whether or not the $\Delta G_b$ terms in Equation 16 represent true free energies or are fictitious is irrelevant here. The method we propose requires only that the

$-kT \log K$ terms of Equation 15 be fully independent, that is, that the dominant-mode MD simulation does not sample non-dominant modes, however the latter are defined. In other words, if binding modes are defined as distinct rotomer conformations for side chains within the binding site, then such rotations must not occur during the simulation unless it is by deliberate structural mutation, as required by our multi-mode method (Equations 10–11). Structural changes of this magnitude are not normally observed in MD simulations, so this definition of binding modes may serve for practical purposes.

The difference in binding free energy, $\Delta\Delta G_b$, for a wild-type and mutant ligand is found by taking the difference of the individual free energies, $\Delta G_b(w)$ and $\Delta G_b(m)$. Assuming the same modes $0, 1, \ldots$ are considered for each ligand (e.g., 0 is the dominant mode and 1 is a non-dominant mode), the free energy difference is

$$\Delta\Delta G_b = -kT \log \left( \frac{e^{-\Delta G_{b0}(m)/kT} + e^{-\Delta G_{b1}(m)/kT} + \ldots}{e^{-\Delta G_{b0}(w)/kT} + e^{-\Delta G_{b1}(w)/kT} + \ldots} \right) \tag{17}$$

Factoring out the native-state mode (0), Equation 17 becomes

$$\Delta\Delta G_b = \Delta\Delta G_{b0} - kT \log \left( \frac{1 + e^{\Delta\Delta_1 G_b(m)/kT} + \ldots}{1 + e^{\Delta\Delta_1 G_b(w)/kT} + \ldots} \right) \tag{18}$$

where $\Delta\Delta G_{b0}$ is the relative free energy of binding of the two ligands to the dominant mode, and where the two $\Delta\Delta_1 G_b$ terms, defined in Equations 8–9, represent for each ligand the difference in binding free energy between the dominant mode and the non-dominant mode 1. An analogous term is added for each new non-dominant mode. Each of these terms can be calculated from a single MD simulation, as shown in Equations 8–11.

### Free energies in the HP model

In the HP model, free energies of binding are calculated from the analog of Equation 12. For saturating ligand concentration the standard state free energy is

$$\Delta G_b = -kT \log \left( e^{-E_{b0}/kT} + e^{-E_{b1}/kT} + \cdots \right) \tag{19}$$

where $E_b$ is the interaction free energy for a single binding mode (see Equation 4):

$$E_b = \Delta h \epsilon + m_h b \epsilon + m_p E_p \tag{20}$$

Here $\Delta h$, the change in the number of HH contacts, accounts for the free energy of any structural change from the native state caused by ligand binding.

### FEP in the HP model

We use Monte Carlo simulations for our FEP calculations. Allowed moves for the HP chain are the standard "end flips" and "corner flips," as described elsewhere (Miller et al., 1992; Camacho & Thirumalai, 1993). For simplicity, ligand movements are not included in the move set. Rather, for each protein conformation, the ligand is forced to bind in the most energetically favorable site. This is assumed when evaluating the change in energy.

All moves resulting in a lower-energy complex are accepted automatically. Those resulting in higher energy are accepted conditionally according to the Boltzmann distribution. The number of steps is kept small (between $1 \times 10^4$ and $1 \times 10^6$) so that only a fraction of the full configurational space is explored, as is true in MD simulations.

In all calculations, we make the approximation that no more than one ligand can bind at any time. This is done to make our simulations resemble typical MD trajectories, which involve single ligand interactions. For the particular HP sequence

and conditions used in this paper, this approximation introduces only small errors (less than 5%).

### Other approaches for correcting the single-mode assumption

Our multi-mode theory requires several MD simulations to calculate the binding free energy. It would be advantageous to use a *single* simulation able to sample many non-dominant modes. Monte Carlo may be one such approach, since it can sample conformations farther away from the native structure (Horiuchi & Go, 1991; Hao & Scheraga, 1994; Kidera, 1995). MD may be another approach, provided that activation barriers between modes could be reduced, for example by increased temperature. These and other approaches may merit further study.

# REFERENCES


Ackers GK, Doyle ML, Myers D, Daugherty MA. 1992. Molecular code for cooperativity in hemoglobin. *Science 255*: 54–63.

Bash PA, Singh UC, Langridge R, Kollman PA. 1987. Free energy calculations by computer simulation. *Science 236*: 564–568.

Camacho CJ, Thirumalai D. 1993. Kinetics and thermodynamics of folding in model proteins. *Proc Natl Acad Sci USA 90*: 6369–6372.

Cantor CR, Schimmel PR. 1980. *Biophysical chemistry. Part III. The behavior of biological macromolecules.* New York. W. H. Freeman and Company.

Chan HS, Dill KA. 1989. Compact polymers. *Macromolecules 22*: 4559–4573.

Chan HS, Dill KA. 1990. Origins of structure in globular proteins. *Proc Natl Acad Sci USA 87*: 6388–6392.

Chan HS, Dill KA. 1991. "Sequence space soup" of proteins and copolymers. *J Chem Phys 95*: 3775–3787.

Chan HS, Dill KA. 1994. Transition states and folding dynamics of proteins and heteropolymers. *J Chem Phys 100*: 9238–9257.

Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*: 7133–7155.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding—A perspective from simple exact models. *Protein Sci 4*: 561–602.

Hao MH, Scheraga HA. 1994. Monte Carlo simulation of a first-order transition for protein folding. *J Phys Chem 98*: 4940–4948.

Horiuchi T, Go N. 1991. Projection of Monte Carlo and molecular dynamics trajectories onto the normal mode axes: human lysozyme. *Proteins Struct Funct Genet 10*: 106–116.

Kidera A. 1995. Enhanced conformational sampling in Monte Carlo simulations of proteins: application to a constrained peptide. *Proc Natl Acad Sci USA 92*: 9886–9889.

Kollman PA. 1993. Free energy calculations: applications to chemical and biochemical phenomena. *Chem Rev 93*: 2395–2417.

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. 1982. A geometric approach to macromolecule-ligand interactions. *J Mol Biol 161*: 269–288.

Lau KF, Dill KA. 1989. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules 22*: 3986–3997.

Lau KF, Dill KA. 1990. Theory for protein mutability and biogenesis. *Proc Natl Acad Sci USA 87*: 638–642.

Miller DW, Dill KA. 1995. A statistical mechanical model for hydrogen exchange in globular proteins. *Protein Sci 4*: 1860–1873.

Miller DW, Dill KA. 1997. Ligand binding to proteins: the binding landscape model. *Protein Sci.* (in press).

Miller R, Danko CA, Fasolka J, Balazs AC, Chan HS, Dill KA. 1992. Folding kinetics of proteins and copolymers. *J Chem Phys 96*: 768–780.

Pearlman DA, Kollman PA. 1989. Free energy perturbation calculations: problems and pitfalls along the gilded road. In: van Gunsteren WF, ed. ***Computer simulation of biomolecular systems: theoretical and experimental applications.*** Princeton, New Jersey, USA. pp. 101–119.

Postma J, Berendsen H, Haak J. 1982. Thermodynamics of cavity formation in water. A molecular dynamics study. *Faraday Symp Chem Soc 17*: 55–67.

Schellman JA. 1975. Macromolecular binding. *Biopolymers 14*: 999–1018.

Shortle D, Chan HS, Dill KA. 1991. Modeling the effects of mutations on the denatured states of proteins. *Protein Science 1*: 201–215.

Singh JC, Brown FK, Bash PA, Kollman PA. 1987. An approach to the application of free energy perturbation methods using molecular dynamics. *J Am Chem Soc 109*: 1607–1614.

Tembe BL, McCammon JA. 1984. Ligand receptor interactions. *Comput Chem 8*: 281–283.

Wilson C, Mace JE, Agard DA. 1991. Computational method for the design of enzymes with altered substrate specificity. *J Mol Biol 220*: 495–506.

Wyman J, Gill SJ. 1990. *Binding and linkage. Functional chemistry of biological macromolecules.* Mill Valley, California: University Science Books.


UC
San Francisco
LIBRARY
UNIVERSITY OF CALIFORNIA · UNIVERSITY OF CALIFORNIA