The Protein Journal (2019) 38:95–119
https://doi.org/10.1007/s10930-019-09822-x

PERSPECTIVE

Why Proteins are Big: Length Scale Effects on Equilibria and Kinetics

Kenneth A. Rubinson[1,2]

Published online: 23 April 2019

Abstract
Proteins are polymers, and yet the language used in describing their thermodynamics and kinetics is most often that of small molecules. Using the terminology and mathematical descriptions of small molecules impedes understanding why proteins have evolved to be big in comparison. Many properties of the proteins should be interpreted as polymer behavior, and these arise because of the longer length scale of polymer dimensions. For example, entropic rubber elasticity arises only because of polymer properties, and understanding the separation of entropic and enthalpic contributions shows that the entropic contributions mostly reside within the polymer and enthalpy originates mostly at the site of small-molecule binding. Recognizing the physical chemistry of polymers in descriptions of proteins' structure and function can add clarity to what might otherwise appear to be confusing or even paradoxical behavior. Two of these paradoxes include, first, highly selective binding that is, nevertheless, weak, and, second, small perturbations of an enzyme that cause large changes in reaction rates. Further, for larger structures such as proteins every thermodynamic measurement depends on the length scale of the structure. One reason is that the larger molecule can control up to thousands of waters resulting in collective movements with kcal sums of single-calorie-per-molecule solvent energy changes. In addition, the nature of covalent polypeptides commonly leads to multiple binding—i.e., multivalency—and the benefits of multivalent binding can be assessed semiquantitatively drawing from understanding the chelate effect in coordination chemistry. Such approaches clarify the origins, inter alia, of many low energies of protein denaturation, which lie in the range of only a few kcal mol^{-1}, and the difficulties in finding the structures of proteins in the multiple substates postulated within complex kinetic schemes. These models involving longer length scales can be used to elucidate why such observed behavior occurs, and can provide insight and clarity where the phenomena modeled employing experimentally inseparable translational, vibrational, and rotational entropy along with charge, dipole moment, hydration, hydrogen bonding, and van der Waals energies together obscure such origins. The short-distance, long distance separation does not include explaining any enzymatic lowering of activation energies due to stabilization of the intermediate(s) along the reaction path. However, well known small-molecule methods that treat electrostatics and bonding can be used to explain the local chemistry that contributes most of the changes in enthalpy and activation enthalpy for the process.

Keywords Proteins · Flexibility · Multivalency · Viscoelastic · Preordering · Hydration · Lengthscales · Thermodynamics · Kinetics

Contents

✉ Kenneth A. Rubinson
Rubinson@wright.edu; Rubinson@nist.gov

[1] National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

[2] Department of Biochemistry and Molecular Biology, Wright State University, Dayton, OH 45435, USA

1 Preliminaries

1.1 A Division into Two Length Scales

A great deal of effort over many decades has been expended attempting to explain and predict how protein structures arise from amino acid sequences as well as the energetics and control of protein-affiliated chemistries. This work refers only briefly to the energetics of protein structure formation, and deals with the reactions in an embryonic way. That is, the details of the reactions' energetics at the atomic scale are not attempted for reasons that are described below. But the desire to explain the reaction energetics leads to depending on two separate length scales—those of < 0.3 Å (see Sect. 1.4) and those that are longer, i.e., > 0.3 Å. The experimental results compel us to consider these different length scales within the chemistry. As will be described, *every thermodynamic measurement has an associated length scale*, and recognizing this property clarifies why differences of a few calories in molar free energy for, say, each individual water of hydration can, nevertheless, control the biochemistry.

Descriptions of the longer-scale properties may explain some of the difficulties that have been encountered with models that include parameters describing only short range interactions—less than approximately 3 Å, or perhaps double that with coarse graining—while trying to understand proteins that function on the 10 Å to 100 Å length scale. In this work, no account is made for the factors that enter into the chemistry of proteins and their binding such as van der Waals interactions, hydrogen bonds, polarity, charge, hydrophobicity, hydration, and primary through quaternary structures. Instead, I illustrate how at larger distances, models based on simple, general physical principles clarify the origins of how proteins interact with smaller molecules or with other proteins both in their structured and unstructured forms.

Note that lengths are described here in Angstroms rather than nanometers (10 Å = 1 nm) because atomic bond lengths are on the order of unity in Angstroms: most are between 1 and 3 Å. Also, energies are stated in kilocalories or calories (rather than kilojoules or joules, 1 cal = 4.184 J) because RT (the gas constant multiplied by temperature) for a *tenfold change* in equilibrium constant at 298 K is also near unity in kcal, i.e., 1.36 kcal mol^{-1} (= 1.987 calories K^{-1} mol^{-1} · 298 K · 2.303).

1.2 Four Characteristics of Big Proteins

With their specific primary amino acid sequences, proteins form the secondary, tertiary, and quarternary structures that are required for their function. From the properties of these structured polymers come the four fundamental properties that are considered here: *flexibility*, *preorganization*, *multivalency*, and *deep hydration* as outlined next.

First, the polymers are much more *flexible* than the monomers composing them, which has long been recognized [1–11]. At ambient temperature, multi-Angstrom changes in structure are available to the proteins, where structural changes in small molecules are on the order of an Angstrom.

Since the monomers are strung together to make up the polymer chains and then continue to rearrange to form the different levels of structure, it follows that these monomers are found set in specific positions within the much larger molecule; the atomic sites' relative locations are *preorganized* [9, 10, 12–18].

Preorganization of the monomers aligns the various atoms of the monomers into positions where binding other molecules may occur simultaneously at numerous atomic positions of the same protein; there are multiple binding sites to associate with another molecule. So the third property is *multiple binding*, which has also been called *multivalency* [19–21] (see Sect. 3).

One other characteristic of having polymers much larger than the monomers is that the larger structure can cause changes in larger numbers of the surrounding water molecules, which means the influence reaches greater distances into the solvent. The large molecule may be thought of as herding the components of the surrounding solvent causing collective changes. The numbers of water molecules involved can be in the thousands per protein, so even a single-calorie change per water can make large

contributions to the total energy. The large number of waters means that the associated hydration that differs from the bulk is multilayers deep ($\sim$ 10 Å). We can call this characteristic *deep hydration* [22–28].

As will be discussed in detail below, these four characteristics of proteins as polymerized monomers and their function as structured macromolecules means that their thermochemical characteristics cannot be treated in the usual manner relating to their standard states [29]. After all, only by comparing thermochemical quantities to common standard states can quantitative comparisons of stabilities, relative binding energies, and rate constants be made. As a result, some extrathermodynamic tool is needed to help in comparing these polymeric proteins with amino acids or with peptide sequences that constitute parts of the protein structure. A semiquantitative, useful tool to integrate three of the four characteristics—the flexibility, preorganization, and multivalency—is to calculate an effective local concentration for the bond-forming sites [30, 31]. These sites may be designated as reactive groups—i.e., the parts involved directly in the reaction (see Sect. 3.2).

These four characteristics of native proteins that have dimensions of, say, 20 Å to 100 Å can be combined and then compared to the short-distance characteristics of small-molecule reactions. There are two orders of magnitude difference between the proteins' lengthscales and the bond formation and breaking distance of 0.3 Å required at the reacting groups. This large difference requires a change in viewpoint toward the biochemistry. The next section begins the discussion to clarify why the kinetics and thermochemistries of these prestructured, polymeric molecules should not be modeled on the same length scales or in a similar manner as small molecules.

1.3 Thermochemical Cycles and the Energetics of Protein Catalysis

Enzymes catalyze biochemical reactions, and the resulting equilibria arise from the balance between the forward and reverse rates. But catalysts are known to change the rates of reactions by lowering the activation energies "without themselves being changed" [32]. The rates of enzymatic reactions range from as fast as diffusion controlled—such as carbonic anhydrase, where the protein holds a zinc ion exposed to passing carbonic acid substrate—to a rate seven or more orders of magnitude slower. Proceeding from binding a substrate to release of product to binding the next substrate comprises a thermodynamic cycle. The sums of all such cycles' free energy, enthalpy, and entropy are zero. Writing this point mathematically, where any point in the cycle can be chosen as an initial position,

$$\oint dG = \oint dH = \oint dS = 0 \tag{1}$$

(The same cyclic thermochemical properties also include any adjacent solvent and ions that accompany the polymer structural changes.) The thermochemistry described by Eq. 1 is fulfilled by all catalysts; all the applicable thermodynamic changes in the reaction arise from the different free energies of the reactants and products and their changes in concentrations.

But that brings up an important question: If the thermodynamic cycle has all these thermochemical measures zero, how do allosteric processes change the interactions with substrate molecules? How are the rates of reaction controlled through either binding with small molecules (peptides, messengers) or by the influence of supramolecular structures formed with other proteins? [33]

The answer must be that the control of the chemistry resides in the *fraction of time* the structure allows binding to occur compared to the fraction of time the conformations of the polymer are such that the substrate *cannot* bind effectively. As described in Sect. 1.4 below, this time requirement arises since binding can occur only over a relatively small range of the structures available to the proteins, and this chemistry may be described by incorporating an equilibrium between the reactable form and the nonreactable structures. This view has been called conformational selection [34–36] (more in Sect. 2.4).

At the longer length scale of proteins, these nonreactable structures are defined here as a continuum in order to reduce the number of variables needed for the large number of macromolecular structures that differ by sub-Angstrom distances. These structures lie within k_BT energies, so they cannot be delineated; we do not see a set of separate protein structures.

Let us make a brief excursion into the quantity k_BT, because a clear understanding of values in various units is helpful. This is the energy equivalent to a Kelvin temperature T. For 25 °C (298 K), it has the following values: an infrared energy of 207 cm^{-1}; a spectroscopic measure of 25.7 meV; an electrochemical measure for k_BT/e^- of 25.7 mV; and the thermochemical equivalent to 592 cal/mol. Also, the Boltzmann distribution shows that a state with energy higher by k_BT has a population 37% as large as the lower energy state, and one at 2 k_BT is 14% as populated. These all will be mentioned below and are useful in relating measurements on the atomic scale and those made on molar quantities.

1.4 Why is the Short Distance 0.3 Å or Less?

The border between short and longer length scales was stated to be 0.3 Å. This section expands on why that is the dividing boundary chosen.

The static properties of bonds suggests the 0.3 Å distance. We recognize that a typical triple carbon–carbon bond length is 1.20 Å, a double bond at 1.34 Å, and a single bond, 1.54 Å. A linear extrapolation yields 1.75 Å for a zero bond order. This progression suggests moving the nuclei apart by another 0.2 Å means the single bond is broken. Similar linear extrapolations for carbon–nitrogen and nitrogen–nitrogen pairs both yield the same result: lengthening a singly bonded pair by 0.2 Å results in a nonbonded pair. This does not mean that van der Waals interactions are not present, but the single bond is broken.

This 0.3 Å approximate distance has been shown experimentally in coordination chemistry as well [37]. A colorful name for the reactable structure as a "near-attack conformation" was given by Lightstone and Bruice [38]. It is also about the same distance an ion or small molecule can diffuse in the time of a low-frequency vibration—33 cm^{-1} vibrations occur over 10^{-12} s.

Taken from the opposite point of view with carbon as an example, the van der Waals radius of carbon at 1.7 Å means that the carbon–carbon van der Waals distance between two is 3.4 Å. This is 0.6 Å longer than the dissociation distance of 2.8 Å defined by Rosker et al. [39] for the dissociation distance of the I–CN bond in the gas-phase, i.e., without adjacent molecules. If the atoms at k_BT cannot get closer than the van der Waals distance 3.4 Å, no bond will form, and so an alternative viewpoint to the van der Waals diameter of contact is needed to discuss the bond making/breaking in water solution.

The best characterized moiety in water-based bond making/breaking is the hydrogen bond between two oxygens. The expectation values for the positions of the hydrogen between acetic acid pairs has been measured [40]. The bond is not quite linear (164.8°), and the two distances within the hydrogen bond are 1.011 Å and 1.642 Å from an end; these distances differ by 0.631 Å. In a double-well potential, the hydrogen reaches its midpoint between the two oxygens when it moves approximately 0.3 Å from one of its average positions towards the other one, i.e., our delineating 0.3 Å [41–44]. In addition, oxygens with shorter hydrogen bonds (hydrogen bridged) are separated from 2.4 to 2.6 Å [45], while longer ones approach the van der Waals distance [46], e.g., oxygen–oxygen of 3.04 Å [47, 48]. These lengths show the H-bonded oxygens can be about ½ Å closer than the oxygen–oxygen van der Waals distance. But the distance for detachment from one to the other remains about 0.3 Å. The remarkable lack of variation in this 0.3 Å distance results from the strongly distance-dependent activation energy of the hydrogen bond [49–51].

In addition, the proton transfer is fast; a proton transfer time has been measured at ~ 50 fs [52]. As for the activation energy, the transfer probability depends strongly on the distance between the two heavy atoms involved. (This is called the proximity effect or propinquity effect relating to the critical distances [21, 50].) Incidentally, the angle of acceptance for this transfer is relatively wide [49, 50, 53]. During the 50 fs duration, for a typical ion or small-molecule with a diffusion coefficient of $D \approx 2 \times 10^{-5}$ $cm^2\ s^{-1}$ [54], the characteristic distance moved over the 50 fs time is $\Delta x = (2Dt)^{½} \approx (2 \cdot 2 \times 10^{-5} \cdot 50 \times 10^{-15})^{½} = 0.14$ Å. The molecules themselves and the surrounding solvent are nearly translationally immobile during the transfer of the proton [55]. In other words, the step of bond transfer for the proton is fast compared to the time to get to the required, specific structure that allows the old bond to break and the new bond to form.

Outside of proton transfer, few experimental examples can be found where structural rearrangement is not needed before a bonding event occurs so that the bond forming can be separated from the transport. One example is described in Appendix 2.

A corollary of needing a specific distance to allow a new bond to form is that if the distance between the incipient-bond-forming atoms is outside the 0.3 Å range, the bond will *not form* [56]. The reactants may diffuse toward the correct position or they may diffuse away and eventually return to the correct position for a subsequent reaction. In either case, the transport is slower than the bond-rearrangement step. And in this, the polymeric protein has a crucial difference from most small molecules. The protein is viscoelastic [57], so the reacting site may reside in a position to react but also can be pulled back by its connected chain to be blocked behind other parts of the molecule and inaccessible even when other binding sites are correctly positioned or already bound. Smaller molecules in water also diffuse but are not subject to retraction. As will be elucidated, this ability to move within some greater or lesser limits of structures is a major reason why proteins are big. Also, the effective length of the chain to which the binding/reacting group is attached determines the range of structures allowed within the k_BT of the viscoelastic environment. In Sect. 2, these limits of structure are described as being analogous to heat induced motions of the potentially reactable atomic sites while restrained by springs. But first, let us look at some possible confusion in finding an applicable ΔG°.

1.5 Difficulties with the Standard State: What Part of a Protein's Structural Free Energy Matters for Equilibria and Kinetics?

Preordering of proteins, which has been mentioned already, includes primary, secondary, tertiary, and quarternary structures, which, however, may overall be only marginally stable [58–60]. Meanwhile, the shorter-longer lengthscale

division suggests that it is worth asking the question, What part of the structural free energy needs to be considered when describing proteins' structures and functions and especially any changes in those functions? In answering the question, consider this slightly absurd thought experiment.

In the thermodynamics of small molecules, the standard states are pure solids, pure liquids, STP gases, and solutions of molar concentrations (activities). An enzyme's standard state should, then, be a solution with 1 M of the amino acids in the protein. We then ask, What would be the probability of an enzyme-catalyzed reaction occurring in a solution with the enzyme chopped up into the individual amino acids in, say, an initially 10 μM solution of the enzyme? Of course the rate of reaction would be close to that without the enzyme present. Neither the molar standard state nor any diluted amino acid solution standard state appear to be applicable.

Then let us link the amino acids together in their correct order to form the protein's primary sequence. Each of the peptide bonds' formation would be some value of free energy in the range + 0.5 kcal mol^{-1} to + 4 kcal mol^{-1}; the expected values [61–63] depend both on the identities and locations of neighboring amino acids as well as solution pH [61, 63]. In other words, free energy is *needed* to assemble the amino acids into the primary sequence: so $\Delta G^{\circ} > 0$.

Should we take the covalently linked, primary structure at 1 M as the standard state? Let's put the unfolded test protein into the solution with the substrate. The reaction still would occur at a rate close to that in water alone. So the 1 M polypeptide solution also is not a standard state for a functioning enzyme.

Next consider the contributions to the free energy of formation of α-helices and β-sheets. For the α-helices, studies of stability with various amino acid sequences are measured relative to the value of a polyalanine α-helix, where for an alanine 21-mer containing three nonconsecutive arginines, the helix forms with a ΔG° smaller than − 0.5 kcal mol^{-1} at 25 °C and mammalian physiological ionic strength [64]. On average, each amino acid contributes $\Delta G^{\circ}_{\text{formation}} \approx -24\text{cal}$ to the helix formation. However, the average $\Delta\Delta G^{\circ}$ deviation from alanine found for the other 19 amino acids was approximately + 0.5 kcal mol^{-1} [65, 66]. The small free energies involved indicate at best marginal stability for α-helices, with the stability dependent on phenomena such as side chain interactions and hydrogen bonding [67–69]. Because the α-helices are barely stable, their presence is not expected to change the overall positive free energy required for forming the polypeptide in solution from the individual amino acids.

The β-sheets also contribute little extra stability. These structures are also only marginally stable and are unfolded in the absence of bound metal ions. Their net free energy for stabilization arises from the metal-ion binding [70]. The formation of β-sheet also does not change the free energy of peptide bond formation being overall positive. In addition, efficient enzyme catalysis does not occur with only primary and secondary structures present. So none of the fully or partially denatured forms—even with their secondary structure intact—will be useful as a standard state.

Finally, only when the last few kcal mol^{-1} of structure formation occurs (the negative of the denaturation free energies such as listed in Table 2) does the enzyme become active—where it is able to engage the short–lengthscale interactions. In other words, none of the energy of connecting the amino acids, folding its secondary, tertiary, and quaternary structures, nor the fluctuations in the structure of the functional protein [9, 71, 72] enters into the thermodynamics of the enzymatic function. Standard states must then be those involving only the products and reactants. For proteins (and for other types of biopolymers as well) binding to form supramolecular structures, the bookkeeping for the thermodynamics begins with the average free energy of the unbonded reactant proteins in their native forms.

2 Molecular Flexibility, Enthalpy Versus Entropy, and a Spring Model

2.1 Molecular Springs for the Two Molecular Length Scales

Let us compare the force constants for structural displacements in small molecules compared to polymers; the language of mechanical springs is unavoidable. We begin with a single spring (Sect. 2.2) followed by connecting a small-molecule spring to a polymer spring (Sect. 2.3). With the familiar physics of attaching two springs in series, the energy distribution between the two dissimilar springs can be specified and applied to a general substrate-protein interaction.

2.2 Stretching a Single Spring: A Model for Entropic and Enthalpic Changes with Structure

Consider two hydrocarbons: propane and its much longer cousin polyethylene. Assume that we can pull on one of the methyl protons on each end of the propane (as shown by the arrows in Fig. 1a) and try to stretch them apart by a few hundredths of an Angstrom in the directions of the arrows. Most of the energy will go into bending the bonds in the

path between the two pulled hydrogens. Meanwhile, a small amount of the energy of the pulling will stretch the two terminal H–C bonds, as well as the two C–C bonds. The energy versus pulled distance is illustrated in Fig. 1b. (In the limit of infinite stiffness, the attempted stretching would result in no structural work/energy change at all.) Upon releasing the molecule, it springs back into its original, unperturbed structure by giving up the potential energy that was imparted.

The force constant for this stretch is 0.72×10^5 dyne cm^{-1}, typical for a single-bond *bending* motion as can be seen from the 2nd row of Table 1. For this motion, a stretch distance of only about 0.075 Å reaches an energy equaling k_BT (298 K).

Pulling the ends of an unbranched polyethylene chain is illustrated in Fig. 2a [73, 74]. The polymer is assumed to be immersed in its theta solvent—a solvent that neither swells nor collapses the polymer. When the ends are pulled apart the same distance as before, bond stretching and bond-angle bending are negligible; energy is not stored in the unstretched and unbent bonds. Instead, the pulling causes a large number of slight changes in the average bond rotational angles. But the energy changes for these slight rotations are within distortions produced by the thermal motions of the surroundings. The pulling does, however, slightly reduce the thermal disorder of the chain, and, as a result, the polyethylene is called an entropic spring, such as is found in rubber. The force constant is about *four to five orders of magnitude* smaller than bond stretching and bond bending, as listed in Table 1. A k_BT-equivalent displacement is expected to be two orders of magnitude larger—on the order of a few Å to 10 Å. Let us simply call this the "polymer potential" compared to the small-molecule potential as shown in Fig. 1b. For this relatively broad, low polymer potential, the thermal motions populate the end-to-end probability density shown by the Gaussian curve that is the well known distribution of end-to-end distances (the chain vector) in ideal polymers [73, 75]. Further, the end-to-end distance movements change essentially only the entropy and not the enthalpy of the chain [57].

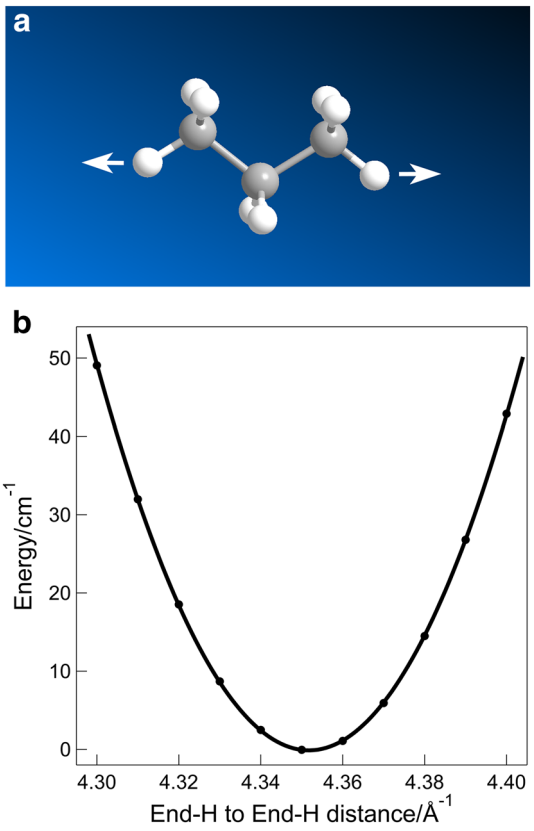

Fig. 1 **a** Stretching propane. **b**. The energy to distort propane in the manner illustrated in **a** was calculated with the DFT method B3LYP/6-31G(d). The resulting potential-energy curve for small distortions is plotted here. The curve is well fit by a parabola with a return force constant of 0.72×10^5 dyne cm^{-1}. This is typical for a single-bond bending motion as can be seen from the second row of Table 1

Finally, when the ends of the polyethylene are released after pulling, they are unlikely to return to the same spatial positions since there are so many different conformations that are able to be populated through the thermal motions within k_BT. The presence of multiple conformations represents the other well known interpretation of entropy as a measure of disorder.

From this simple illustration, you can see that when changes in chemical structure occur over the longer length scales, entropy changes occur, and the structure can only be defined by statistical averages. However, on the short length scale, with a changing structural range of only a small fraction of an Angstrom, the enthalpy (or internal energy) changes. This dichotomy illustrates the important idea about entropic and enthalpic contributions to the free energy. They are divided by the partition of energy into structures that can be changed by energies above k_BT (enthalpy) and below k_BT (entropy). This division between enthalpy and entropy is discussed in greater detail in Appendix 3. It is interesting to note that between propane and the polymer sizes, some chain length must exist that will have the enthalpy and entropy contributing equally to the free energy of the work of stretching.

The contrast of the close structural fits for the short, chemical bond forming distance and the relatively broad polymer potential extent was demonstrated experimentally by Rajasekaran et al. [76]. With NMR, changes in chemical shifts within proteins were measured at distances up to tens of Å away from point mutations. However, within the measurement precision, while the local structure changes were able to be registered with NMR, they found "no apparent change in the 3D structures."

Table 1 Representative force constants

Identification	Force constant[a] $\times 10^{-5}$ dyne cm^{-1} = $\times 10^{-5}$ pN nm^{-1}	Length equivalent to k_BT (Å)[b]	References
Single-bond stretch	5–8	0.028–0.022	[208, 209]
Single-bond bend	1	0.064	[208–210]
HCl bond stretch	0.5	0.09	[211]
Hydrogen-bond stretch	0.3	0.1	[212, 213]
α-Helix stretch	0.3	0.1	[214]
Hydrogen-bond bend	0.03	0.32	[212, 213]
Single-bond torsion	0.02	0.45	[209]
Sodium channel of nerve[c]	0.0002	4.5	[74]
dsDNA 10 μm (entropic range)	10^{-5}	14	[215]

[a]Units are: for stretching, value $\times 10^{+5}$ dyne cm^{-1} or value $\times$ mdyne Å^{-1}; for bending and torsion, value $\times 10^{+11}$ dyne cm rad^{-2} or value$\times$ mdyne Å rad^{-2}

[b]Assume a harmonic potential. For bending, the k_BT-equivalent distance is along an arc at a lever arm length of 1.5 Å. (k_BT = 4.10 pN nm at 25 °C)

[c]The force constant shown in the table for the sodium channel of nerve likely arises from the shortest of the flexible, hydrophilic polymer chains that connect the transmembrane helices

The work done separating the ends of the polymer also perturbs the solvent structure that surrounds the chain. For proteins in water, these positional displacements subtly change the low-energy vibrational modes of a large numbers of local solvent waters as well. Because of the strong coupling between the protein and solvent molecules (also see Sect. 2.5), it would be surprising if the subtleties of the energetics of these interactions could be simulated well as if the protein chain lies within an assumed homogeneous environment.

It is common to approach calculating the entropy affiliated with the proteins and the chemical binding of various types within and on the surface of the protein by summing over tens to hundreds of modes that are found through knowing the structure and applying appropriate force constants for local stretching, bending, and twisting of groups of a few atoms through the entire molecule [77–79]. Such calculations are, in effect, determining the polymer potential from local force constants with the added difficulty of separating the entropy of binding and rates measured experimentally to changes in the protein and solvent that do not contribute to the reaction [78, 80–82] This is in contrast to using the polymer potential where the entropy can be calculated from measured changes in the available volumes occupied by the reacting atoms. (See Sects. 2.4, 3.2, and Appendix 1.)

There is a caveat about the effect of flexibility at Ångstrom distances but longer than the 0.3 Å bond formation distance. With the idea of docking substrates or other molecules as putting together parts of a precisely fitting (0.3 Å) puzzle, the flexibility can either increase or decrease the various rates, and so change the equilibrium. A decrease may occur as derived quantitatively here, namely having a greater range of structures available and so spending a smaller fraction of time in the fittable position. However, an increase might result from the flexibility allowing the puzzle pieces to fit at all. This latter effect is, in essence, the reverse of the idea that the viscoelastic protein not allowing the site to be accessible.

2.3 Energy Partition Between Two Different, Connected Springs

With the order(s) of magnitude difference in extent of the polymer potentials compared to those of bonded atomic pairs', an interesting question is, "How is the energy distributed when a bond is formed between an atom attached to a protein and one on a small molecule?" This interaction of small-molecule substrate site to a polymer-connected protein binding site can be modeled by a weaker spring attached to a much stronger spring as shown in Fig. 3 [74, 83]. The dot represents the bond formed (a separate step), and let us assume both springs are assumed to be stretched from their unstretched lengths.

At equilibrium, the steady state forces at the connection are equal and opposite for the weak (w) and strong (s) springs. That is, $-k_w \Delta x_w = F_w = -F_s = k_s \Delta x_s$. As these equalities indicate, the displacements from the positions of unstretched springs are inversely proportional to each force constant. However the energy is quadratic in the

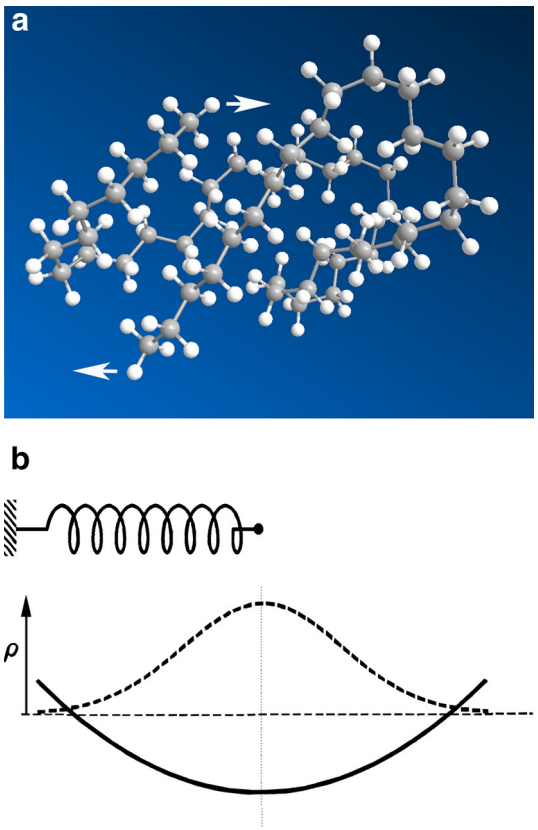

Fig. 2 **a** Stretching polyethylene from the ends. **b** The polymer potential (bottom, solid line) can be considered as arising from a spring (top) holding the binding atom (dot) and anchored on the other end by a relatively immobile part of the protein (hatched block). The spring analog does not necessarily correspond to a single polypeptide chain segment, but can represent a network of chains. The length range is on the 1- to 10-Ångstrom scale. The dashed line indicates the gaussian probability density of the positioning in space of the atom that can form a bond with, e.g., a substrate

Fig. 3 A widely stretched, weak spring (a polymer) at the left in equilibrium with a slightly stretched, strong spring (a small molecule). Forces at the connecting dot are equal and opposite. The hatched blocks are considered at fixed positions. This represents the interaction of a small-molecule substrate with a binding site anchored on a peptide chain of a protein

displacement; that is, $U_w = ½\ k_w\ (\Delta x_w)^2$, and $U_s = ½\ k_s\ (\Delta x_s)^2$. As derived in detail in Appendix 1, the work required to bring each of the springs from their unstretched states to their positions determined by their equality of force are related by

$$U_w = \frac{\Delta x_w^2}{\Delta x_s^2}\ U_s = \frac{k_s}{k_w}\ U_s \qquad (2)$$

So the fraction of the energy stored in each of the springs is inversely proportional to its force constant. The cause of this partitioning is simply that when a bond forms (the dot) between the binding atoms—one restrained weakly by the protein and the other held on the much stiffer substrate—the forces for further conformational change are equilibrated, but the work of reaching their respective, stretched positions were not equal.

Equilibrated at k_BT, the partitioning of the energy is quite uneven. For example, if the distance accessible to the polymer binding site is three times that of the substrate site to which it is bound ($\Delta x_w = 3\ \Delta x_s$), then 90% of the distortion energy resides in the polymer. If the accessible distance for the protein binding site is five times that of the bound substrate ($\Delta x_w = 5\ \Delta x_s$), then more than 96% of the conformational energy resides in the polymer. In other words, for reactions of proteins binding small substrates, the reaction entropy is mostly within the protein, and the enthalpy is due to the binding itself and some level of distortion of the small molecule.

This generality of the model allows us to state the following about, e.g., enzyme catalysis, polysaccharide properties [84], and intrinsically disordered proteins:

- Entropy of *structural* changes reside primarily in the polymer chain network and associated solvent and not in the small molecule(s) bonded to it. Any reduced fraction of time the protein resides in its reactable form is a reflection of the entropy within the polymer potential.
- The enthalpy of the reaction comes primarily from the bond formation/breaking and concomitant desolvation/solvation at the binding location.

The balance between enthalpy and entropy tends to be more equal when the interactive force constant is larger such as occurs when the effective chain length is short; perhaps the protein spring is stiffened by hydrogen bonds such as in an α-helix or β-sheet. These two rules also hold for protein–protein binding, i.e., when no separate, small molecules are involved.

2.4 A Binding Site's Spatial Position, the Entropy, and Its k_BT Surroundings

As noted above, the atoms that react at the binding sites of a protein must be quite precisely in the correct positions for the incipient bonds to form. If the binding sites' positions are not correct when its restraining chain sits at the minimum-energy position, each binding group must have migrated to its reactable position through a random walk, i.e., diffusion, comprised of a large number of small (at most tenths of Angstroms), random displacements due to

k_BT-forced Brownian motions. As a result, the chain-attached atoms can move away from the chain's minimum-energy position (bottom of polymer potential) at the expense of lower probability at the higher energy structure where the reaction can occur. (The probability distribution is shown by the dashed gaussian in Fig. 2b.) By this mechanism, the intramolecular Brownian motion to reach a higher energy structure can be utilized when the enzyme binds to the substrate to provide some conformational forces that can perturb the stiffer substrate toward an activated-state structure. However, as pointed out by Levitt [85], most of the free energy is then contributed by the chemical bonding with the substrate at each correctly positioned site. Once the binding atoms are released and the product leaves, the acquired polymer strain energy may be coupled back to the surroundings at k_BT with diffusion in the reverse direction.

The polymer potential, illustrated in Figs. 2b and 4, reduces the description of a multidimensional system to a 1-D coordinate. It can be used to understand the ongoing chemistry because the other motions either average since they are much faster than the structural change within the potential or these other motions (modes) are isolated from the process of interest.

The polymer potential has one especially interesting property; it does not change its vertical position on the internal energy axis easily. One way to cause a vertical shift is by changing the solvent's dielectric constant, which really means substituting the solvent. In this case, the energy change results from the response of the charges and dipoles within the protein's volume. Another way to induce a vertical energy shift is to apply a linear electric field extending across the entire polymer potential. But a long-range field is difficult to sustain since it is dissipated by mobile countercharges both in the protein and the surrounding solution. So we can consider the polymer potential(s) of a nonbonded, functional (i.e., not denatured) protein to be fixed on the internal energy scale.

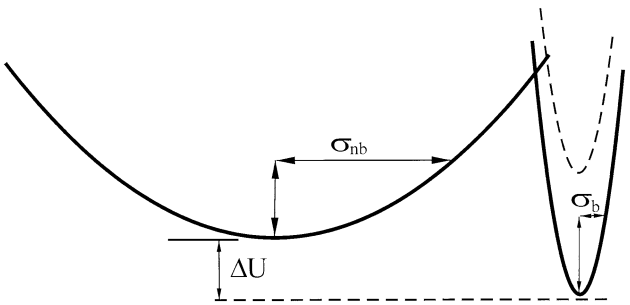

Fig. 4 Definitions of the width parameters for the polymer potentials for the energy *vs*. position of a binding site attached to a polypeptide. The potential representing the nonbonded (nb) form is the broad potential, and the narrow potential applies when bonded (b) at the position shown. The energy difference between the potentials is ΔU. The height at which the σ values are defined is k_BT above the minima. The higher, dashed, narrow parabola indicates binding with a different binding energy

However, that vertical immobility is lost when a bond forms with the reactive group of the protein. The new, lower vertical position is localized at the binding site and measures the enthalpic contribution of the binding. In addition, the bond formation reduces the potential's 1-D spatial extent, and the difference in the parabola width represents the protein's entropic change [83].

This entropic contribution to the free energy from the changes in the available spatial volume of the binding group(s) is able to be calculated quantitatively from experimental measurements in the following way. On the time scale of nanoseconds and longer in a solution, the motions of the atoms attached to the protein chains move by diffusion [86–88]. (An extreme example of such motion is seen in adenylate kinases that have a mobile "lid," where the front can move 32 Å from its extreme position to enclose the active site [89, 90]. Another occurs where a side chain group moves 9.2 Å between the free and bound forms of a T cell receptor [91].) When such translational motions are present, the isothermal entropy change can be characterized by changes in accessible volume [92, 93].

Mathematically,

$$\Delta S = R \ln(V_1/V_2) \tag{3}$$

Each atom has some accessible volume when alone (say, V_1) and generally a smaller volume V_2 when bound to a substrate. (The 1-D equivalent are the σ_{nb} and σ_b of Fig. 4.) The protein's entropic contributions can be measured experimentally using the proportionality to the thermal (Debye–Waller) factors in X-ray crystallography—specifically, those of the bond-forming atoms in X-ray crystal structures of the proteins before and after binding. (These Debye–Waller factors are usually shown by spheres proportional to their magnitudes centered at each atom's position.) The Debye–Waller factor $B \propto \sigma^2$, where σ is the 1-D spatial gaussian distribution's standard deviation [83]. When the entropies calculated from the changed σ values for all the individual binding sites are summed, it accounts for the protein's contribution to the free energy. This was true for a β-Lactamase [83].

The reason this analysis for ΔS works is that the contribution of the polymer chain network attached to the binding sets cancel out for the two different volumes (bound and unbound) [83], so the volume accessible to the binding site alone accounts for the entropy measured in this way. This suggests that calculating distortional modes for the entire protein to find their associated entropies is unnecessary.

2.5 The Inability to Separate Protein and Water Energy Changes

An extensive literature exists describing the intimate coupling of protein structures and motions with the solvating water surrounding them. As is well understood, the complexity of this coupling is significant [56, 94–111] with many efforts to classify and calculate the enthalpies and entropies associated with each chemical subclass such as buried/surface, side-chain/main-chain, hydrophillic/hydrophobic, charged/uncharged, and H-bonded/van der Waals association.

However, for the thickness of the hydration layer around proteins, a consensus seems to have developed in locating a division between hydration and bulk water. The consensus is 3 to 4 layers of water with a thickness around 7 Å to 10 Å is found employing numerous techniques: time-dependent fluorescence of natural and substituted tryptophans [112, 113]; THz spectrometry over the range 0.25 THz to 400 THz [114] and an investigation over a narrower THz range comparing the signals over a range of concentration [115–117]; and similar results were found with NMR [25, 27]. In addition, as found with 2D ultrafast infrared spectrometry, the surrounding hydration waters' dynamics were influenced by the presence of lysozyme to distances of 15 to 20 Å [118]. Also, THz measurements on *monomeric* deoxyribonucleotides indicate four hydration layers [119].

Similar deep hydration has been found for polysaccharides [120], lipid bilayers (easily seen from spacing of multilayer lipids and multilamellar liposomes) [22, 23, 121], and between layers of the water-soluble polymer polyethylene glycol (PEG) [26]. The last-mentioned work estimated the water binding energies within the individual layers by relating the spacing between the PEG sheets (found by neutron scattering at various PEG concentrations) and the measured water activity of 2 kDa and 4 kDa PEG solutions [122]. Starting with the *second* layer from the polymer surface, with bulk water as the standard state, these estimates were:

$$\Delta G^{\circ}_{2\text{ndlayer}} = -2.1\,\text{cal mol}^{-1},\ \Delta G^{\circ}_{3\text{rdlayer}} = -0.8\,\text{cal mol}^{-1},$$
$$\Delta G^{\circ}_{4\text{thlayer}} = -0.4\,\text{cal mol}^{-1}.$$

These are small energy differences.

The formation of the PEG layers separated through large numbers of weakly bound waters indicates a significant general attribute about structure formation at different length scales. For small, covalent molecules, their structures arise from bonds with energies greater than 100 kcal mol^{-1} $\approx$ 170 k_BT. However no permanent structure would appear if the energies of the bonds connecting the atoms fell to near k_BT.

On the other hand, if these small-molecule bond energies were found between atoms on surfaces of separate protein molecules, a large aggregation of connected proteins would soon form a precipitate. The structure of the precipitate would be essentially amorphous at length scales longer than that of an individual protein in the aggregation. For specific, monodisperse structures of proteins to form or to form monodisperse supramolecular protein structures, the process involves many sites binding weakly and dissociating, and rebinding repetitively over the large, interacting area. This continues until the most stable bound structure is reached and could involve hundreds of atom pairs. In effect, with weak bonding between surface atoms of the proteins (or between the waters bound to those surfaces), the molecules are able to rearrange until they form the final equilibrium structure. Only then would each protein or each supramolecular structure find the same energy minimum and form the same structure. These qualitative descriptions lead to an important general rule about chemical structures at different length scales:

> For small molecules, regular structure depends on high bond-formation energies ($\gg k_BT$), while large molecules form regular structures by depending on many small binding energies (on the order of a calorie) spread over a larger area such that the *total* for the binding surfaces becomes $\gtrsim$ k_BT.

As another example of this summation of small free energies to control structures consider a spherical protein with a diameter of 65 Å (volume $\approx$ 1.4 $\times$ 10^5 Å^3); this possesses the approximate volume of human serum albumin. The protein's four layers of hydration water fill a shell about 10 Å thick with a volume of 1.8 $\times$ 10^5 Å^3—a volume greater than the protein's. At a density of one g/cm^3, a molecule of water occupies about 30 Å^3, so the hydration shell of the 65 Å-diameter protein contains about 6000 molecules of water. How does this deep hydration affect the protein energetics?

Let us assume 1500 waters in each of the 2nd, 3rd, and 4th layers with their respective energies estimated from the PEG hydration [26] of − 2.1, − 0.8, and − 0.4 cal mol^{-1}. Then the outer three layers would contribute an additional − 5 kcal mol^{-1} to the energetic contributions of the first hydration layer. Compare this − 5 kcal mol^{-1} additional binding free energy with the representative denaturation free energies shown in the second column of Table 2. The hydration free energies are the same magnitude as the measured denaturations. These properties of the deep hydration leads us to ask the question, Does the stabilization energy of a folded protein lie within its structure, or within the surrounding water? An additional question is, Can we expect that such a subtle but important contribution to the free energy as the intimate interactions of the waters

Table 2 Free energy of unfolding-denaturation of selected proteins (low to high)

Protein	ΔG° (kcal mol^{-1})[f]	References
Bovine serum albumin[c]	1.2 ± 0.9	[216]
Staphylococcal nuclease[a]	3.15 ± 0.05	[217]
Bacterial histidine-containing protein[b]	4.21	[218]
α-Lactalbumin (pH 7.0)[b]	4.3 ± 0.1	[219]
Phosphoglycerate kinase yeast[d]	5.3 ± 0.1	[220]
Ribonuclease A (pH 7.0)[b]	7.3 ± 0.2	[219]
Myoglobin (pH 6.6)[b]	7.9 ± 0.2	[219]
Lysozyme (pH 7.0)[b]	8.9 ± 0.1	[219]
Lysozyme (pH 7.0)[c]	9.1 ± 0.7	[216]
Cytochrome *c*[c]	9.1	[221]
Ribonuclease[c]	10.8	[221]
Phosphoglycerate kinase	11.8 ± 0.2	[220]
T. Thermophilus[d]		
Chymotrypsin[c]	12.3	[219]
Myoglobin[c]	12.3	[221]
AmpC β-lactamasee	14.0	[222]
E. coli		

[a]Pressure monitor with NMR

[b]Extrapolate fully unfolding to zero denaturant

[c]Thermal methods, 25 °C

[d]CD and Fluorescence, $\Delta H^{\circ} = 0$

[e]Thermal methods, 54.6 °C

[f]Uncertainties listed where reported. None of the references define a confidence interval for these uncertainties

of the ionic solution be modeled successfully as a continuum environment as is commonly done? Even for smaller organic molecules, the answer is no [123].

Numerous other biological structures such as membranes may have rigid lengths larger than proteins, and, apparently the depth of the hydration depends on that length scale; longer lengthscales involve greater depth with, e.g. charge density remaining the same [22–24, 124, 125]. (An extreme example of this effect is seen where the water mobility is affected to greater than μm depths from a polyacrylic acid gel surface [126].) Presently, however, not enough experimental data is available to formulate a quantitative dependence between the hydration depth and the lengthscale of the structural correlations (i.e., stiff is highly correlated in its structure, and floppy has shorter structurally correlated distances). The hydration depth also depends on other surface characteristics including charge density, hydrophobicity, as well as properties of the solution such as ion content and distances between the various other molecules present [127]. Nevertheless, it is expected that the surrounding water has a large effect on the protein stability where the water contributing to the stability is not limited to the first hydration layer, which is commonly the only one modeled explicitly. However, for protein folding and unfolding, given the changing hydrated surface area and the expected but not quantitated local changes in depth and strength of hydration, it is unclear how to sum all the local changes in the solvation energies for the processes.

From the discussion above, it is clear that collective, long distance (up to 20 Å) interactions are likely to contribute significantly to the structural stabilities of proteins and their supramolecular complexes. Nevertheless, to the author's knowledge, such distance-dependent free energies over that lengthscale have not been incorporated into modeling. Also, from the discussion in this section, within that 20 Å lengthscale, including the contribution from the surrounding solvent is essential.

2.6 When Does Entropy Compensate Enthalpy?

Regardless of the origin of an equilibrium, experiments measuring changes in the equilibrium are most precise when the largest changes in the equilibrium accompany changes in a reagent concentration or conditions. This optimum point occurs where the equilibrium quotient, Q in the equation for free energy

$$\Delta G = -RT \ln Q = \Delta H - T\Delta S \quad (4)$$

is near unity; $\Delta G = 0$. Measurements outside of the range $10 > Q > 0.1$ are far less precise. But a tenfold change in measured Q changes the free energy by only $\pm$ 1.4 kcal M^{-1}. A reasonable interpretation would be to claim the presence of enthalpy-entropy compensation [128–130]. However, in this case, the compensation is due to obtaining the data where the equilibrium changes can be measured most precisely, i.e., close to $Q = 1$.

3 Multivalency and Preorganization

Carbonic anhydrase utilizes a single zinc atom in catalysis, but most enzymes involve more than one reactive atom, and their mechanisms are multivalent [131]. Since the binding sites are connected to a flexible protein's structure, they are anchored near their reactive positions. As mentioned earlier, these sites incorporated in the active protein can be characterized as preorganized [17, 21, 132, 133]. A definition of preorganization: The concept of multiple parts of a molecule being held close to its chemically binding form prior to reaction. This concept appeared in the chemical literature almost four decades ago [12], where it applied to cryptands and chelates [134–136]. Preorganization for enzyme catalysis has been considered nearly as long [13].

Examples of localizing the amino acids of the peptide chain in their preorganized positions include formation of hydrophobic clusters, hydrogen bonding, and sterically limited chain arrangements extending over all the intramolecular length scales from adjacent atoms to the longest distance within the structure [137–140]. In addition, since the chains to which the reactive sites are attached can move around in some localized volume, the proteins themselves act as viscoelastic solvents surrounding the reaction site. Similarly, multivalent viscoelastic materials with preorganized reactive sites also can be used to describe proteins that are part of reversibly formed structures such as microtubules. Preorganization also is implicit to the iterative process creating better aptamer specificity. Similarly, the idea permeates the quantitative statistical modeling of the chemistry underlying epistatic structural alterations [141–143]. (Epistatic effects describe the structural changes measured upon making one or more single-site mutations in a protein.)

Forming a bond at each of the binding sites is a short–lengthscale process, while the motions involved both in the catalytic cycle and in the formation of supramolecular structures occur with long-lengthscale motions. Parallel to these—but with simpler structures that have the same properties—are the chelates in inorganic chemistry, such as the one shown in Fig. 5. Both of these structures are

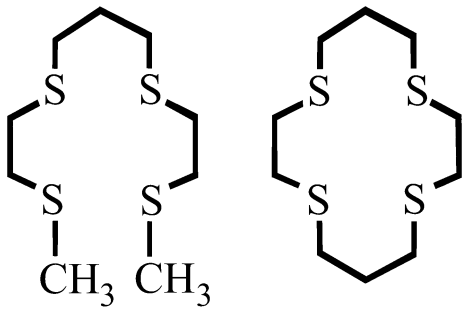

Fig. 5 For the chelate (left) $K_{CuL} = 94$, and for the macrocycle (right) $K_{CuL} = 21{,}800$ in water at 25 °C

characterized by multivalency [19, 21] and preorganization [14, 21, 144, 145]. Their structural, noncovalent, and solvation contributions to binding stabilities have been thoroughly discussed [146]. Even for these apparently simple complexes, all these contributions are difficult to disentangle. Next, in Sect. 3.1, I show a way around this perplexing interplay of factors.

3.1 Multivalency with Different Amounts of Preordering: The Chelate and Macrocycle Effects

These parallels between proteins and the smaller chelate molecules can be applied semiquantitatively to account for the multivalency and preorganization contributions to protein equilibria and kinetics. Understanding multivalency requires comparing a multivalent molecule with the same binding groups and the same number of each but binding individually. For example, compare one ethylenediamine with two methylamines. The standard state of 1 M ethylenediamine has the same number of binding sites as a 2 M solution of methylamine, which is not the latter's standard state. As a result, to compare equilibria involving a given number of *individual* ligands and the same number and identities of covalently connected binding necessarily requires an extrathermodynamic accounting because of the lack of a joint standard state [29]. The covalently-connected set is generally more stable: this is the chelate effect [19, 21, 133, 147].

Numerous attempts over the past 60 years to understand the underlying causes of the chelate effect have not been successful due to the complexity of describing theoretically all the details of the factors involved. Numerous studies have recognized the contribution of preorganization [14, 21, 105, 144, 145, 148] but with different explanations. For example, one framework involves tradeoffs between topology and conformational entropy to provide a "valency-corrected" description [19]. Another divides the free energy into bond strain—named the "connection free

energy"—and the "intrinsic binding energy" that changes with translational and rotational entropy [149]. A third attributes the changes to different vibrational modes [81]. Nonlinear or nonadditive contributions further confound such explanations.

The difficulties found when attempting such detailed descriptions become especially clear when comparing an open-chain chelate to its related, cyclized macrocyle such as the pair shown in Fig. 5. The macrocyle binds ionic copper 230 times more strongly than the chelate, where the chelate and macrocycle have the same enthalpy of binding. The added stability of the macrocycle compared to the related chelate has been called the macrocycle effect [150]. None of the three models noted above can explain the difference. In this specific case, the divergence was attributed to direct methyl–methyl interaction and some level of structural strain [151]. For more complicated multivalent reactants such as proteins, a more general method to explain such changes is presented in the next section.

3.2 Extrathermodynamic Approach to Multivalency Stability: The Local Effective Concentrations

I suggest that one can bypass the failed models for the preordering effects on the equilibria and kinetics by recognizing that the concentration of a reactant only can be locally effective; reactions occur at the short–length scale. This idea has been understood implicitly for redox reactions where electron transfer occurs at electrode surfaces and also for a diffusion controlled reactions in a homogeneous solution. Calculating a semiquantitative effective local concentration is a straightforward and natural extension of the short-longer lengthscale division [30, 31].

The formalism of effective local concentrations uses the concepts of kinetics and equilibria and not those of statistical mechanics mainly because of the difficulty of ascertaining the identities and changes in vibrational modes, hydration depths, charge, polarization, degrees of freedom, and degeneracies for translations, rotations, and vibrations. These are all incorporated into an effective concentration when contrasting reactions, e.g., where ethylenediamine is compared to two methylamines or for proteins where a denatured form differs from its native structure. The viewpoint presented here *differs* from the concept of effective molarity [21, 147, 152], the value of which multiplies a reference equilibrium constant to match an equilibrium that shows greater stability compared to that reference. Effective molarities are suggested to have values up to 10^8 M [147]. As pointed out more than 40 years ago by Illuminati, et al. [153], "effective molarities clearly do not represent real concentrations and may be accounted for by well-known chemical principles." Clearly, effective molarities are proxies that includes other factors as well.

A convenient structure-based, semiquantitative way to make the extrathermodynamic adjustment is to find the effective local concentration for each binding group of a multivalently binding unit. Smith and Margerum alluded to effective concentrations to explain the difference in the strengths of binding of a cryptand (a multidentate molecule that binds ions in a 3-D cavity—a crypt) compared to the associated open chelate [154]. They explained that the difference is due to a reduction in conformational entropy. However the structural basis introduced here can be more generally applied.

The short-length long-length division supports this view since once a multidentate ligand binds one of a protein's sites, any other, subsequent binding groups attached covalently to this first binder are held more closely in proximity to the next binding sites instead of still migrating freely in solution. This proximity has the same effect as having a higher concentration at their required short-distance for reaction [155]. Here, binding at a multivalent site results from sequential steps of bond formation until all the sites are bound; some initial site is found by migration through the solution, which is followed by bonds that subsequently form at the now higher effective local concentrations. (To be discussed later is that a site may be held away from a bindable position by the attached viscoelastic chain or blocked by some local groups, i.e., steric clash.)

A quantitative calculation can be illuminating. Molecules at 10 μM are, on average, 590 Å apart [156], but when a reaction site on a molecule is being held within an average distance of 10 Å, its equivalent concentration is then 2 M. Figure 6 illustrates the process.

Here, two interacting molecules (ellipses) have four separate binding pairs that are located on the framework such that eventually all four simultaneously bond in the final complex. Assume that each of the four sets are chemically identical. As represented in the upper right

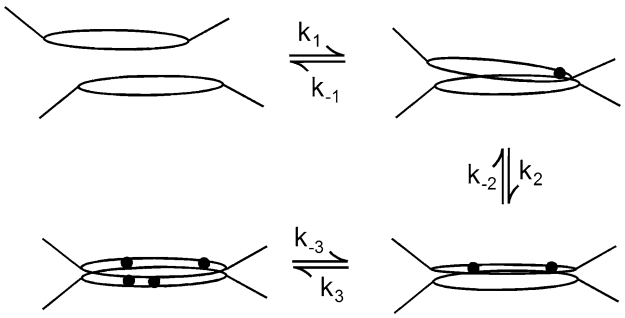

Fig. 6 The sequential process of a multivalent equilibrium binding by four sites. The last two binding sites are close together and act as a single, third site for formation but together are harder to break. This fourth site is not included in the analysis presented in the main text

structure, the complex initially binds a single site through diffusion of the multisite molecules and subsequent approach by one of the binding pairs closely enough to form a bond. The dot indicates the first bond formed. Many events of binding and breaking might occur until the first bond forms at its ultimate, "correct" position. The forward process for this formation occurs at the rate k_1, and the reverse rate is k_{-1}.

The rate of the first bond's formation and its equilibrium constant are

$$R_1 = k_1[C_A][C_B]; \quad K_1 = k_1[C_A][C_B]/(k_{-1}[\text{intermediate}]) \tag{5}$$

where $[C_A]$ and $[C_B]$ are the concentrations of the two multidentate reactants. But Eq. 5 is the last time that the solution concentrations describe the chemistry.

After attaching at one site, the two parts now are linked together. In this way, the supramolecular structure held together by the now-bonded pair will be inhibited from moving apart, but can move by rotating about the first bond's position while the bonded form holds.

Subsequently, a second bond forms with the same rate constant since all four are identical. However, the local concentration for the second site is much higher than that of the molecules in solution since the second site is preordered. In a quantitative calculation, we must ask, "What is a reasonable value to use for the effective local concentration vis a vis its companion site?" In other words, "What value should be used for the local effective concentration between this second pair?" A simplifying approximation can be made here that does *not* attempt to account for all those factors mentioned before: conformational entropy changes, limitations to translation, rotation, or nutations of various moieties within the molecule, any changes in structure of the surrounding solutes and solvent [98, 157, 158], nor statistical factors. This approximation is based on concentrations of pure, organic liquids such as shown in Table 3. For these representative, pure liquids, the average value of ~20 M can be taken as the local concentration of a pair of binding sites within the short length distance, i.e., immediately adjacent to each other.

Given that rotation around the first bond is allowed, the second site occupies a volume defined by that motion, and the volume is defined by the distance between the first and second site given the steric restrictions. Let us simply propose a single effective local concentration of 5 M for the second, incipient binding site "diluted" from 20 M. The exact value of such an effective local concentration is not critical since the value eventually appears only as its logarithm in a free energy calculation. Because of this logarithmic functionality, the uncertainty in this effective concentration will still offer at least semiquantitative insight despite an absence of most molecular and atomic details [72].

Table 3 Molar concentrations of pure, organic liquids

Compound	MW (Da)	Density (g/cm^3)	Molarity (mol L^{-1})
Methanol	32.04	0.791	24.7
Dimethyl ether	46.07	0.735	16.0
Acetic acid	60.05	1.05	17.5
Methylamine	31.06	0.656	21.1
Dimethylamine	45.08	0.65	14.4

With the approximation for the local concentration of the second bond-forming pair set at 5 M and with the same rate constants for all the sites, we can write

$$R_2 = k_1[5\text{ M}]; \quad K_2 = k_1[5\text{ M}]/(k_{-1}[\text{intermediate}]) \tag{6}$$

If one of the two now-bonded sites' bond breaks, the other stays intact, and because the local 5 M effective concentration applies, the broken bond can be remade rapidly. This whole interactive process is equivalent to a long-time cage effect [147, 159–161].

Perhaps after the first or second site has formed a bond and dissociated several times, a third and possibly yet other pairs of sites will all finally arrive at their bindable positions as illustrated at the lower right in Fig. 6. Each of the nearby, bindable (but as yet unbound) pairs will then have an equivalent concentration locally of about 20 M. As a result,

$$R_3 = k_1[20\text{ M}]; \quad K_3 = k_1[20\text{ M}]/(k_{-1}[\text{intermediate}]) \tag{7}$$

For a three-site binding process where C_A and C_B are the free solution concentrations of the two reactants, the three-site equilibrium constant with its effective concentrations can be expressed as

$$\begin{aligned} K &= \frac{k_1[C_A][C_B]\cdot[5\ M]k_1\cdot[20\ M]k_1}{k_{-1}k_{-1}k_{-1}[\text{intermediate}]^3} \\ &= [100\ M^2]\frac{k_1^3[C_A][C_B]}{k_{-1}^3[\text{intermediate}]^3} \end{aligned} \tag{8}$$

The term [100 M^2] on the right expresses the approximate contribution from preordered, multiple binding compared to an equal number of individual sites at the initial protein solution concentrations. This is the semiquantitative extrathermodynamic multiplier for the multiple-site binding. The *estimated apparent* increase in free energy for the binding is

$$\begin{aligned} \Delta G^\circ &\approx -RT\ln[100M^2] = -2.303\cdot 600\text{ cal}\cdot\log[100M^2] \\ &= -2.8\text{ kcal/mol} \end{aligned} \tag{9}$$

This is the free energy that is expected to be measured experimentally due to the preordering in a multidentate complex. In addition, the free energy equivalents for a tetradentate, pentadentate, and hexadentate within this formalism are approximately − 4.6, − 6.3, and − 8.1 kcal/mol. Also, as indicated by the rightmost term of Eq. 9, this free energy is mostly entropic in character; it enters the scheme as a concentration change.

Note in the cartoon that a fourth site is so close to the third one that it binds in the same step. This illustrates some ambiguity that can arise in this extrathermodynamic analysis. For such a pair, the off rate will also be characterized by requiring simultaneous breaking of both bonds.

These ΔG-equivalents for the preordering also provide *minimum values* for the threshold of protein denaturation. Since the derivation assumed that $\Delta H^{\circ} = 0$, then for any experimental denaturation/stability measurement, the free energy characterizing the stability must include the sum of all pairs' binding enthalpies (and local entropies) added to the preorder $T\Delta S$ term. As can be seen in Table 2, the least stable of these structured proteins agrees with the minimum preorder value. Here, the label "denatured" perhaps should be in quotes since it only reflects the extent of loss of the active form into possibly multiple structures that are not active [17, 162–170]. In other words, the experimental free energies of denaturation follow from a simplified equilibrium involving only the active/inactive fraction [171]. The short vs. longer lengthscale division clearly underlies these findings that there are more different structures in the collection of the denatured forms.

The cartoon model of Fig. 6 lies at one stiffness extreme; all the binding pairs are anchored on rigid supporting structures and are correctly localized to bind simultaneously eventually. However, to the extent the anchoring structure allows motion so that the pairs can move within some larger volume both before and after binding, these less rigid structures result in lower effective local concentrations. This change in accessible (at k_BT) volume also corresponds to a change in entropy, as was introduced qualitatively in Sect. 2.4. That is,

$$\Delta S = \Sigma \Delta S_i = \Sigma R \ln (V_{i1}/V_{i2}) \quad (10a)$$

When reduced to a one-dimensional geometry for k_BT driven displacement, the standard deviation of the gaussian position distribution is $\sigma = (k_BT/k)^{1/2}$, where k is the force constant. Then,

$$\Delta S = \Sigma R \ln (\sigma_{i1}/\sigma_{i2}) \quad (10b)$$

(The site entropy related to this thermal motion through the temperature factor and more formally as the experimental Debye–Waller factor, B, through $B = 8\pi^2\langle u^2\rangle$, and since $u = \sigma$, then $\sigma = B^{1/2}/8.88$.) [83]

This treatment appears to be quantitative when describing a macroscopically measured change in kinetics for β-lactamase upon an amino acid substitution distant from the active site [83]. This quantitative agreement was surprising since the analysis required normalizing the B-values due to different qualities of the crystals and different temperatures of the data collection for the various substituted β-lactamases.

As imprecise as using local effective concentrations may appear, when compared to atomic-level or coarse-grained modeling, it uses a minimal number of parameters to model the causes and consequences of having different measurable accessible volumes. At the same time, this straightforward viewpoint, together with the characteristics of deep hydration, explains some of the difficulties found matching atomic or coarse-grained modeling with the experimental results for protein structures and energies [172, 173].

The idea of effective concentration also has been applied at the long lengthscale. Differences in a tether length between two subunits were recognized as changing local probabilities of reaction building up from the ends of growing microtubules [174]. The tethers were long with a minimum of 40 amino acids. The presence of the tether produces a higher effective concentration of the second-binding subunit, increasing its rate of addition over that of the first-binding part. The modeling did not include the short distance range over which binding occurs.

3.3 Reactable or Non-reactable?: Population Shifts and Rates

The effectiveness of enzymes arises from the general properties discussed above: preordering, flexibility, multivalency, deep hydration, and bond forming only on the short lengthscale. (Warshal et al. [71] argues definitively that the control of enzyme-enhanced rates are due to structure and not due to vibrational modes interacting temporally with the active region.) In addition, the existence of allosteric control and the cyclic thermodynamics of catalysts (Eq. 1) requires that the fraction of time various possible structures exist is central. An alternative, equivalent description is that the proteins have some probability to have a structure that fits the optimal binding of substrates (and/or regulators) as well as simultaneously a fraction far less strongly binding. This differentiation has been given a number of labels. For example, over the last half century numerous groups differentiated productive versus nonproductive binding [10, 175–177]. More specifically, Kokkinidis et al. [7] and many others [5, 79, 178–181] view proteins as a dynamic ensemble of conformations, where each substrate is bound to different conformations with different affinities. All are examples consistent with the required structural specificity for

binding small molecules as substrates or effectors. Since the less strongly binding structures tend to have orders-of-magnitude lower binding, there is no loss of applicability to separate them into two forms, bindable and nonbindable, for the rate equations.

For a general second order reaction, we write the rate equation for the species E for the enzyme and S for the substrate (not to be confused with ΔS for the entropy).

$$R_i = k_i[E][S] \tag{11}$$

However, consider the reaction of calcium with the chelate ethylenebis(oxyethylenenitrilo)]tetraacetic acid (egta), which can be used to form a calcium-concentration buffer: ignoring protons, $Ca^{2+} + egta^{-4} \rightleftharpoons Ca\ egta^{2-}$. The chelated calcium is present in the solution but is prevented from reacting as hydrated, free calcium; it is sequestered. We know that the total calcium concentration present, $[Ca^{2+}]_{Tot}$ is not the buffer level, $[Ca^{2+}]_{Free}$. The treatment of enzymes with their reactable and nonreactable populations can be recognized as parallel to this chelate system.

The reactable protein has a time-average concentration lower by some fraction than its nominal concentration.

$$[Enzyme]_{Total} = [E]_{reactable} + [E]_{nonreactable} \tag{12}$$

Equation 12 may be recast to

$$\begin{aligned} Rate = k_i\,[E]_{reactable}\,[S] \\ \uparrow\downarrow \\ [E]_{nonreactable} \end{aligned} \tag{13}$$

Equation 13 expresses the conditions that the changes of reactable $\rightleftharpoons$ nonreactable occur more slowly than both the mutual diffusion of the reactants and the picosecond bond make/break step. To include the presence of some fraction of nonreactable protein is to recognize the scheme of Eq. 13 as being a general statement of allosterism [182, 183].

This equilibrium between reactive and nonreactive forms has often been combined with the reaction step through the use of an equilibrium constant equivalent to both together. Separating the contributions of the reactive and nonreactive forms should be edifying as to the contributions of, for example, the proteins entropy versus the contribution from water.

Equation 13 indicates that control of the observed rate changes do not arise from changes in the rate constant [184] but from changes in reactable concentration. In other words, the reactable form is so narrow in its structural range that the rate constant should be considered to remain constant because it applies to a single, congruent population among all those achievable by the flexible proteins. Such behavior is observed directly in the fluctuations between open and closed ion channels [185, 186] (where the equilibrium populations are controlled by an electric field [187]) as well as more recently in single-molecule measurements made with fluorescently labeled reactants [188–190]. In addition, significant local information about these population shifts has been measured by NMR [191–193].

However, depending on the lifetimes of various steps of a reaction, the reactable-nonreactable equilibrium of Eq. 13 may not be needed. For example, the Henri-Michaelis–Menten equation holds for

$$E + S \rightleftharpoons ES \rightleftharpoons E + P \tag{14}$$

when the species ES dissociates to E + P faster than to E + S, and when the substrate is in large excess. If the formation of ES is fast enough, and the change in free energy of the bound form great enough (right hand parabola of Fig. 4), the substrate binding drives the reaction towards [ES] up to the point where $[E]_{tot} = [E]_{reactable}$. Then the kinetics will be described satisfactorily by Eq. 11.

As previously described, many possible structures are included in the nonreactable concentration, and can be shifted by allosteric regulation toward or away from the catalytically binding structure. For example, binding at a distant site can shift the relative concentrations of the two forms or even effectively lock the protein in one of them. The $[E]_{reactable}/[E]_{nonreactable}$ also can be shifted by blocking the path to the bond-forming position [3]. With the narrow structural range of the multiple binding being less than an atomic radius, it is reasonable to say that the kinetics is a more precise measure of local structure than found even from an X-ray analysis of an excellent crystal or from a cryo-EM structure.

3.4 Induced Fit? Lock and Key?

Depending on the time frame that is chosen, the path to multivalency can be viewed as following a sequence of binding events involving many moieties: protein binding sites, water, ions, and other smaller or polymeric molecules. (Within a longer time frame, all these processes may be viewed as, or are measured as, a single event.) As covered in Sect. 3.2, the time progression of multibinding shows one site's binding increases the rate of a second, and so forth. We may call this a *progression* toward the conclusion of the process. Nevertheless, the short lengthscale of the reactable structure for the multiple binding is contrary to the concept of "induced fit" postulated by Koshland [1] and reviewed a decade later by Koshland and Neet [2] and often discussed subsequently [34–36, 194].

To clarify this last statement, a careful definition of induction also is needed. Again, the time frame, number of

sites, and distances are involved. For times longer than a few picoseconds and distances more than an Å, the reactants reorganize randomly until the positions are reached where a bond can be made. On the other hand, induction requires that a directional force greater than k_BT pulls and rotates the parts to fit together faster than this diffusional process. In a condensed phase outside the single Å and ps time and distance limits, it is highly unlikely a directional force of sufficient magnitude could be discerned; only by showing a sufficient directional force exists could the fit be considered induced.

Another postulated general mechanism is lock and key, also known as conformational selection. If we refer back to Fig. 6, illustrating the multiple binding process extending over some timeframe, it would be highly unlikely that every site were sufficiently precisely set up to bind "simultaneously." In that case, it appears difficult to apply the idea of a single step of a key binding in the well fitting lock.

In considering both induced fit and lock and key and whether a useful assignment of either is warranted, there are three properties to consider. These three properties can be presented as questions. First, what time frame is being considered? Second, how many local sites are involved? And third, within what distances are the various binding steps? I suggest that without a clear answer to these three questions in each case that may be considered, neither of these two points of view is well defined.

Similarly within the viewpoint discussed throughout this work, both the separation of these two classifications as well as their application to process and mechanism do not appear to be terribly helpful. Minimally, if one or the other is desired to be assigned to a reaction, the limits of time, site number, and distances must be designated, otherwise they have little meaning.

3.5 A Limit to the Short-Distance Long-Distance View: Stabilization of Reaction Intermediates

Interpreting structural changes in proteins as being within polymer potentials, and recognizing that reactions occur only when the structures are at "near-attack conformations," provides a method to visualize and understand how many of the phenomena associated with protein structures and functions occur using only a few parameters that are experimentally accessible. What it cannot do in the case of enzymes is to clarify the contribution that the preordering of the active site structures stabilize intermediates as a part of the catalysis.

One of the reasons that the stabilization's contribution is difficult to determine arises from the difficulty of delineating the contribution from longer-range preordering. For example, while it is it attractive to suggest that a 1 μM enzyme that has three sites involved in catalysis such that their effective local concentrations are each 10 M at the active site makes the probability of a reaction occurring to be 10^{21} times more likely, that argument is specious. Two reasons why are that the reaction mechanism in the absence of the enzyme may be different, and the numerical multiplier differs for trivial changes for the enzyme: say for being 5 μM.

So while the changes in accessible volume of the reacting atoms can show the entropic contribution from the protein site, the changes in electronic structure at distances of the size of the active site is not part of the formalism. However, fortunately, the local electrostatics and bonding are changing within a length scale of approximately 4 Å, which is well treated by the small-molecule methods that are well known in chemistry [195–200]. As mentioned earlier and illustrated in Fig. 4, this local chemistry contributes mostly to the enthalpy of the binding. Also, it is this local binding, and not the preordering of the site that contributes to the lowering of activation energies due to stabilization of the intermediate(s) along the reaction path [85].

4 Summary

Chemical bonds are formed and broken at distances on the order of a quarter of an atomic radius; proteins' diameters are more than 100 times larger. In addition, proteins are intimately coupled with about four layers of hydrating waters that are not the same as the bulk; these waters rearrange along with motions of the protein chains. All the energy differences between the hydrating waters and the bulk are on the order of 1 cal, but they can act collectively with enough free energy to have significant effects on the protein structures' intramolecular and intermolecular equilibria and function. This deep hydration suggests that proteins' influence each other at distances up to 20 Å between surfaces (2 × 4 hydration layers).

A simple physical model of two connected springs with different force constants suggests that the entropy and changes of the entropy of reactions resides in the flexible polymer, and that the enthalpy of enzymatic reactions results primarily from the chemistry at the active site, and not from protein structural changes. Because of proteins' flexibility, the range of ineffective structures is much larger than the narrow range of structures that can bind and function. As a result, allosteric control resides in modifying the fraction of time that proteins can bind and function, which can be modeled as an equilibrium between bindable and non-bindable structures of the viscoelastic protein. Inherent in this formalism is that the rate constant itself does not change. This additional equilibrium is not needed

under conditions where the majority of the protein can be found in the bound form.

With the recognition that polymeric proteins are flexible, preorganized, multivalent, and deeply hydrated, a useful, extrathermodynamic model is presented that can estimate semiquantitatively the free energy measured that is due to the flexibility, preorganization, and multivalency. The hydrating waters are intimately coupled to these changing structures. These free energies are involved in protein denaturation, multivalent active site binding, and formation of supramolecular structures.

Acknowledgements I am pleased to thank the following colleagues for reading and commenting on the typescript: Lawrence Prochaska, Gerald Alter, David Hoogerheide, Joseph Hubbard, Lawrence Berliner, Susana Teixeira, and Curt Meuse, and to Karl Irikura for calculating the propane's potential. This work utilized neutron scattering facilities supported in part by the National Science Foundation under Agreement No. DMR-1508249.

Appendix 1

Relations of the Energies, Force Constants, Gaussian Widths, and Entropies

For the stretched, connected springs of Fig. 3,

$$-k_w\,\Delta x_w = F_w = -F_s = k_s\,\Delta x_s \tag{15}$$

where the subscript w represents properties of the weak spring on the left, and subscript s represents the properties of the strong spring on the right. F is the force, k the force constant, and Δx the distance the point of connection resides from its position at the rest point of the unconnected spring. The signs of Δx_s and Δx_w are opposite.

The internal energies involved in stretching each spring from its rest point so the bond can be formed are, respectively,

$$U_w = \tfrac{1}{2}k_w\,\Delta x_w^2; \quad U_s = \tfrac{1}{2}k_s\,\Delta x_s^2 \tag{16}$$

From Eq. 15, we know that $\Delta x_s = -(k_w/k_s)\Delta x_w$ and, substituting for Δx_s in Eq. 16, then $U_s = \tfrac{1}{2}\,(k_w^2/k_s)\,\Delta x_w^2$. Putting $\Delta x_w^2 = 2U_s\,(k_s/k_w^2)$ into Eq. 16 and by rearranging get

$$U_w = \frac{k_s}{k_w}\,U_s \tag{17}$$

which is Eq. 2 in the main text.

The quadratic potentials of Fig. 4 show the energies versus stretch position for a reactive group attached to a polypeptide and elastically restrained by it. This group has a greater probability of being at the center, unstretched position than at the ends. The quantitative probability distribution of the possible locations of the binding groups are derived next.

With the energy U of the binding group at a location Δx from the minimum of the polymer potential U is described by

$$U = U_{0P} + \frac{1}{2}k(\Delta x)^2 \tag{18}$$

where k is the Hookean force constant, Δx is the distance moved from the minimum of the potential, and U_{0P} is the energy minimum of the polymer potential. For the remainder of this derivation, we assign the value of $U_{0P} = 0$. To simplify the appearance of the equations, from now on, x is used in place of Δx.

The normalized probability distribution of the population over the parabola is

$$\rho(x) = \left(\sqrt{2\pi k_B T/k}\right)^{-1} \exp\left(-\frac{kx^2}{2k_B T}\right) \tag{19}$$

where k_B is the Boltzmann constant, and k the force constant. This is a Gaussian centered at the energy minimum. This Gaussian distribution can be characterized by its standard deviation. By comparing Eq. 19 to the general equation for a normalized Gaussian distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{20}$$

shows that the Gaussian's characterizing width

$$\sigma = \sqrt{\frac{k_B T}{k}} \tag{21}$$

Applying Eq. 21 for both springs of Fig. 4 at the same $k_B T$, we can find the three equalities of Eq. 22

$$\sigma_w^2 k_w = k_B T = \sigma_s^2 k_s; \quad \sqrt{\frac{k_s}{k_w}} = \frac{\sigma_w}{\sigma_s}; \quad \frac{k_s}{k_w} = \frac{\sigma_w^2}{\sigma_s^2} \tag{22}$$

The σ values are a measure of 1-D displacement, so $\Delta S = -R\ \ln\ (\sigma_w/\sigma_s)$ expresses the entropy change between the site before the bond is formed and when bonded. In addition, Eqs. 21 and 22 show that if the force constant changes, so does the width. So the protein's structural entropy can be changed by stiffening the tether. Possible molecular mechanisms to do so include structuring the chain by intrachain hydrogen bonding or forming intramolecular bonds or bonding with adjacent molecules to shorten or strengthen the polymer spring.

In three dimensions, this change in σ transforms to a difference in accessible volume. For a reaction of bond formation, let the smaller, bonded volume accessible be V_1, and the initial, larger unbonded accessible volume be V_2. Then,

$$\Delta S = R \ \ln \frac{V_2}{V_1}$$

which appears as Eqs. 3 and 10a and 10b. This result only holds for the quadratic approximation for a continuum system for each binding group [83] and the approximation of homogeneous distributions within the volumes. As the attachment becomes looser, such as on a "long" unstructured polypeptide, the gaussian distribution may need to be explicitly included.

Appendix 2

Another Experimental Example of Fast Bond Formation

Another experimental example for bond formation times (compared to relatively slow mass migration) is for bonds between gold atoms, where the preorganization is fulfilled by the proximity of weak clusters of $Au(CN)_2^-$ in water. Kim et al. [201] used femtosecond x-ray scattering to follow photoactivated bond formation between the gold atoms that reside at van der Wall distances within these $Au(CN)_2^-$ clusters. The bond formation occurred within about a picosecond over which time three adjacent golds having distances of 3.9 Å and 3.3 Å from the central gold and forming a 101° angle transformed to a linear structure with both bond distances 2.8 Å. (This bond-forming process is surprisingly fast when we realize the characteristic diffusion distance over that time at ambient temperature is around 0.6 Å.) This measurement of the bond-forming supports the fraction-of-an-Angstrom motion of bond formation/breaking in solution.

Appendix 3

More About Separating Entropy and Enthalpy and the Importance of k_BT

At the end of Sect. 2.3, it was noted that molecular structures can be altered by energies less than k_BT—an entropic mechanism—or altered by energies greater than k_BT that contribute enthalpy. For the quantitative support of that description, I will be switching back and forth between molecular lengthscale descriptions and descriptions using terms that ordinarily only apply to macroscopic systems. When descriptions involve molecular characteristics such as vibrational modes, the description should be interpreted as being that of a single molecule, which resides in a form that is an average of the ensemble from which the molecule is taken. Similarly, since a single molecule is assumed to represent the ensemble average, entropy will be used for single-molecule differences that are due to entropy on the macro scale. In addition, when thermodynamic terms are used on the molecular scale, enthalpy has no clear meaning since pressure-volume (PV) work is part of the enthalpy. Nevertheless since free energy involves both the enthalpy and entropy, let us use enthalpy as the term and ignore the difference between internal energy and enthalpy.

To clarify the nature of entropy, consider a protein residing in its equilibrium form that exhibits many low energy vibrations [202] (i.e., $< 200\ cm^{-1}$ at 25 °C). Now add some energy within a few ps [42, 101, 203]. The added energy first partitions among these same vibrational modes and eventually also migrates and adds to the solvent's vibrations, rotations, and translations that define the temperature. The added energy is eventually distributed in a volume so large relative to the protein that the temperature can be considered unchanged. This added energy cannot be recovered because no temperature difference remains which would be able to drive some energy back to, e.g., change the molecule's structure. The energy distributed to these low-energy modes of the protein and of the water is, then, the entropic part; the entropy is the energy "dispersed" or "dissipated," [204] which are useful terms to describe the progression of events described in this paragraph.

On the other hand, if the work going into the molecule is distributed into modes that are not occupied at k_BT (say, arbitrarily, 10 k_BT) or changes the Boltzmann occupancy of some modes at lower energies but still greater than k_BT, then the energy put in can be recovered by removal from the same modes, which drops the molecular energy to lower states. This energy can be recovered since the temperatures of those modes—as characterized by the Boltzmann distribution—are greater than the surroundings, which allows the energy to flow back. This recoverable energy is the enthalpic part.

While energy added to the protein can flow into modes that can be characterized as either entropic or enthalpic, the delineation between the two is not sharp [189]. The reason the boundary remains indistinct is that while the ambient temperature is the dividing line, the Boltzmann distribution indicates that energy states greater than k_BT are also occupied at the system temperature. (As mentioned earlier, a state with energy k_BT above ground is 37% as populated as the ground state, and one at $2k_BT$ is 14% as populated.) As a result, some normally unrecoverable energy (when $\Delta T = 0$) can be recoverable by removing energy from the partially occupied higher-energy states. This blurring in energy may be reasonably assumed to be about 3 k_BT wide. If no states exist in that range, the enthalpy-entropy separation is sharp. For a protein in solution, this sharp separation is not possible.

One reason for this impossibility is that the longer the length scale of a vibration, the lower the vibrational energy levels that are available in the (molecule + solvent). This relationship can be seen in column 3 of Table 1. The lower energy vibrations are associated with larger groups of bonded atoms: examples are librations of groups of four atoms and vibrations that occur over so many atoms they are classified as acoustic modes [205–207]. This length-scale-frequency relationship brings us full circle to the differences between distorting alkane chains being enthalpic (propane) and entropic (polyethylene) as discussed in Sect. 2.2. The long chain has low energy vibrational modes that can be excited by the random thermal motions of the surroundings. The short chain can have only the methyl rotations excited this way. The indistinct boundary between enthalpy and entropy is confirmed since chains with lengths greater than propane decrease the enthalpic contribution, and shortening the polyethylene eventually increases the enthalpic contribution until reaching some length of chain that has both mechanisms contributing equally to the restoring force. The boundary of enthalpic and entropic contributions is indistinct.

References

1. Koshland DE (1958) Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci USA 44:98–104
2. Koshland JDE, Neet KE (1968) The catalytic and regulatory properties of enzymes. Ann Rev Biochem 37:359–411
3. Careri G, Fasella P, Gratton E (1979) Enzyme dynamics: the statistical physics approach. Ann Rev Biophys Bioeng 8:69–97
4. Britt BM (1997) For enzymes, bigger is better. Biophys Chem 69:63–70
5. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D (2005) Intrinsic dynamics of an enzyme underlies catalysis. Nature 438(3 November):117–121
6. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Dorothee Kern D (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature 450(6 December):913–918
7. Kokkinidis M, Glykos NM, Fadouloglou VE (2012) Protein flexibility and enzymatic catalysis. Adv Prot Chem Struct Biol 87:181–218
8. Hong L, Glass DC, Nickels JD, Perticaroli S, Yi Z, Madhusudan T, O'Neill H, Zhang Q, Sokolov AP, Smith JC (2013) Elastic and conformational softness of a globular protein. Phys Rev Lett 110:028104
9. Warshel A, Bora RP (2016) Perspective: defining and quantifying the role of dynamics in enzyme catalysis. J Chem Phys 144:180901
10. Bar-Even A, Milo R, Noor E, Tawfik DS (2015) The moderately efficient enzyme: futile encounters and enzyme floppiness. Biochemistry 54:4969–4977
11. Glantz-Gashai Y, Meirson T, Samson AO (2016) Normal modes expose active sites in enzymes. PLoS Comput Biol 12:e1005293
12. Cram DJ, Lein GM, Kaneda T, Helgeson RC, Knobler CB, Maverick E, Trueblood KN (1981) Augmented and diminished spherands and scales of binding. J Am Chem Soc 103(20):6228–6232
13. Krishtalik LI, Topolev VV (1983) The intraglobular electrostatic field of an enzyme. 1. The primary field created by the polypeptide core, functional groups and ions of the alpha-chymotrypsin molecule. Molekuliarnaia Biologiia 17(5):1034–1041
14. Reinhoudt DN, Dijkstra PJ (1988) Role of preorganization in host-guest-chemistry. Pure Appl Chem 60(4):477–482
15. Bruice TC, Benkovic SJ (2000) Chemical basis for enzyme catalysis. Biochemistry 39(21):6267–6274
16. Rajamani D, Thiel S, Vajda S, Camacho CJ (2004) Anchor residues in protein–protein interactions. Proc Natl Acad Sci USA 101(31):11287–11292
17. Nienhaus GU (2006) Exploring protein structure and dynamics under denaturing conditions by single-molecule FRET analysis. Macromol Biosci 6:907–922
18. Wittenberg JB, Isaacs L (2012) Complementarity and preorganization. In: Steed JW, Gale PA (eds) Supramolecular chemistry: from molecules to nanomaterials. https://doi.org/10.1002/9780470661345.smc004/full
19. Kitov PI, Bundle DR (2003) On the nature of the multivalency effect: a thermodynamic model. J Am Chem Soc 125:16271–16284
20. Badjić JD, Nelson A, Cantrill SJ, Turnbull WB, Stoddart JF (2005) Multivalency and cooperativity in supramolecular chemistry. Acc Chem Res 38:723–732
21. Sun H, Hunter CA, Navarro C, Turega S (2013) Relationship between chemical structure and supramolecular effective molarity for formation of intramolecular H-bonds. J Am Chem Soc 135:13129–13141
22. Rand RP (1981) Interacting phospholipid bilayers: measured forces and induced structural changes. Ann Rev Biophys Bioeng 10:277–314
23. Pabst G, Rappolt M, Amenitsch H, Laggner P (2000) Structural information from multilamellar liposomes at full hydration: full *q*-range fitting with high quality x-ray data. Phys Rev E 62(3):4000–4009
24. Filfil R, Chalikian TV (2003) The thermodynamics of protein–protein recognition as characterized by a combination of volumetric and calorimetric techniques: the binding of turkey ovomucoid third domain to α-chymotrypsin-chymotrypsin. J Mol Biol 326:1271–12288
25. Fisette O, Päslack C, Barnes R, Isas JM, Langen R, Heyden M, Han S, Schäfer LV (2016) Hydration dynamics of a peripheral membrane protein. J Am Chem Soc 138:11526–11535
26. Rubinson KA, Meuse CW (2013) Deep hydration: poly(ethylene glycol) M_w 2000-8000 Da probed by vibrational spectrometry and small-angle neutron scattering and assignment of ΔG° to individual water layers. Polymer 54:709–723
27. Cheng C-Y, Varkey J, Ambroso MR, Langen R, Han S (2013) Hydration dynamics as an intrinsic ruler for refining protein structure at lipid membrane interfaces. Proc Natl Acad Sci USA 110:16838–16843
28. Xu Y, Havenith M (2015) Perspective: watching low-frequency vibrations of water in biomolecular recognition by THz spectroscopy. J Chem Phys 143:170901
29. Munro D (1977) Misunderstandings over the chelate effect. Chem Britain 13(3):100–105
30. Cotton FA, Harris FE (1956) The thermodynamics of chelate formation. II. A Monte Carlo study of the distributions of configurations in short chains. J Phys Chem 60(10):1451–1454
31. Carter MJ, Beattie JK (1970) The kinetic chelate effect. Chelation of ethylenediamine on platinum(II). Inorg Chem 9(5):1233–1238

32. Wolfenden R, Snider MJ (2001) The depth of chemical time and the power of enzymes as catalysts. Acc Chem Res 34(12):938–945
33. Naganathan AN (2019) Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. Curr Opin Struct Biol 54:1–9
34. Vogt AD, Di Cera E (2012) Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. Biochemistry 51:5894–5902
35. Vogt AD, Di Cera E (2013) Conformational selection is a dominant mechanism of ligand binding. Biochemistry 52:5723–5729
36. Gianni S, Dogan J, Jemth P (2014) Distinguishing induced fit from conformational selection. Biophys Chem 189:33–39
37. Cabbiness DK, Margerum DW (1970) Effect of macrocyclic structures on the rate of formation and dissociation of copper(II) complexes. J Am Chem Soc 92(7):2131–2133
38. Lightstone FC, Bruice TC (1996) Ground state conformations and entropic and enthalpic factors in the efficiency of intramolecular and enzymatic reactions. 1. Cyclic anhydride formation by substituted glutarates, succinate, and 3,6-endoxo-D^4-tetrahydrophthalate monophenyl esters. J Am Chem Soc 118:2595–2605
39. Rosker MJ, Dantus M, Zewail AH (1988) Femtosecond clocking of the chemical bond. Science 241:1200–1202
40. Jönsson P-G (1971) Hydrogen bond studies. XLIV. neutron diffraetion study of acetie acid. Acta Cryst B27:893–898
41. Robl C, Hentschel S, McIntyer GJ (1992) Hydrogen bonding in $Be[C_2(COO)_2]\cdot 4H_2O$–a neutron diffraction study at 15 K. J Solid State Chem 96:318–323
42. Elles CG, Crim FF (2006) Connecting chemical dynamics in gases and liquids. Annu Rev Phys Chem 57:273–302
43. Goryainov SV (2012) A model of phase transitions in double-well Morse potential: application to hydrogen bond. Phys B 407:4233–4237
44. Ishikita H, Saito K (2016) Proton transfer reactions and hydrogen-bond networks in protein environments. J R Soc Interface 11:20130518
45. Blinc R, Hadži D, Novak A (1960) The relation between the bridge length of short hydrogen bonds, the potential curve, and the hydroxyl stretching frequency. Beri Bunsenge Phys Chem 64(5):567–571
46. Emsley J (1980) Very strong hydrogen bonding. Chem Soc Rev 1:91–124
47. Batsanov SS (2001) Van der Waals Radii of Elements. Inorg Mater 37(9):871–885
48. Mantina M, Chamberlin AC, Valero R, Cramer CJ, Truhlar DG (2009) Consistent van der Waals Radii for the Whole Main Group. J Phys Chem A 113:5806–5812
49. Menger FM (1983) Directionality of organic reactions in solution. Tetrahedron 39(7):1013–1040
50. Menger FM (1985) On the source of intramolecular and enzymatic reactivity. Acc Chem Resh 18:128–132
51. Karaman R (2010) A general equation correlating intramolecular rates with 'attack' parameters: distance and angle. Tetrahedron Lett 51:5185–5190
52. Rini M, Kummrow A, Dreyer J, Nibbering ETJ, Elsaesser T (2002) Femtosecond mid-infrared spectroscopy of condensed phase hydrogen-bonded systems as a probe of structural dynamics. Faraday Discuss 122:27–40
53. Menger FM, Chow JF, Kaiserman H, Vasquez PC (1983) Directionality of proton transfer in solution. three systems of known angularity. J Am Chem Soc 105:4996–5002
54. Robinson RA, Stokes RH (1959) Electrolyte solutions, 2nd edn. Butterworths, Appendices 11.1 & 11.2, London
55. Rubinson KA (1984) Regularity in protonation and rate constants and the structures in solution of reactants containing a benzene ring. J Phys Chem 88:148–156
56. Rühmann EH, Rupp M, Betz M, Heine A, Klebe G (2016) Boosting affinity by correct ligand preorganization for the S2 pocket of thrombin: a study by isothermal titration calorimetry, molecular dynamics, and high-resolution crystal structures. ChemMedChem 11:309–319
57. Urry DW, Hugel T, Seitz M, Gaub HE, Sheiba L, Dea J, Xu J, Parker T (2002) Elastin: a representative ideal protein elastomer. Philo Trans R Soc Lond B 357:169–184
58. Bryan PN, Orban J (2010) Proteins that swich folds. Curr Opin Struct Biol 20:482–488
59. Proter LL, He Y, Chen Y, Orban J, Bryan PN (2015) Subdomain interactions foster the design of two protein pairs with $\sim$ 80% sequence identity but different folds. Biophys J 108:154–162
60. Wong K-B, Yu H-A, Chan C-H (2012) Energetics of protein folding. In: Egelman EH (ed) Comprehensive biophysics, vol 3. Academic Press, New York, pp 19–33
61. Dobry A, Fruton JS, Sturtevant JM (1952) Thermodynamics of hydrolysis of peptide bonds. J Biol Chem 195:149–154
62. Borsook H (1953) Peptide bond formation. Adv Prot Chem 8:127–174
63. Martin RB (1998) Free energies and equilibria of peptide bond hydrolysis and formation. Biopolymers 45:351–353
64. Xiong K, Asciutto EK, Madura JD, Asher SA (2009) Salt dependence of an α-helical peptide folding energy landscapes. Biochem 48:10818–10826
65. Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HX, O'Neil KT, DeGrado WF (1995) Protein design: a hierarchic approach. Science 270:935–941
66. Pace CN, Scholtz JM (1998) A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 75:422–427
67. Scholtz JM, Qian H, Robbins VH, Baldwin RL (1993) The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. Biochemistry 32:9668–9676
68. Muñoz V, Serrano L (1995) Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. J Mol Biol 245:275–296
69. Yang J, Spek EJ, Gong Y, Zhou H, Kallenbach NR (1997) The role of context on α-helix stabilization: host-guest analysis in a mixed background peptide model. Prot Sci 6:1264–1272
70. Kim C, Berg JM (1993) Thermodynamic β-sheet propensities measured using a zinc-finger host peptide. Nature 362:267–270
71. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM (2006) Electrostatic basis for enzyme catalysis. Chem Rev 106(8):3210–3235
72. Kamerlin SCL, Warshel A (2009) At the dawn of the 21st century: is dynamics the missing link for understanding enzyme catalysis? Proteins 78:1339–1375
73. Flory PJ (1953, Chap. XI) Principles of polymer chemistry. Cornell University Press, Ithaca
74. Rubinson KA (1986) Closed channel-open channel equilibrium of the sodium channel of nerve: simple models of macromolecular equilibria. Biophys Chem 25:57–72
75. Flory PJ (1988, Chap. VIII) Statistical mechanics of chain molecules. Hanser, Munich
76. Rajasekaran N, Sekhar A, Naganathan AN (2017) A Universal pattern in the percolation and dissipation of protein structural perturbations. J Phys Chem Lett 8:4779–4784
77. Fischer S, Smith JC, Verma CS (2001) Dissecting the vibrational entropy change on protein/ligand binding: burial of a water molecule in bovine pancreatic trypsin inhibitor. J Phys Chem B 105:8050–8055

78. Nicolaï A, Delarue P, Senet P (2015) Intrinsic localized modes in proteins. Sci Rep 5:18128
79. Kalescky R, Zhou H, Liu J, Tao P (2016) Rigid residue scan simulations systematically reveal residue entropic roles in protein allostery. PLoS Comput Biol 12(4):e1004893
80. Smith JC (1991) Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. Q Rev Biophys 24(3):227–291
81. Houk KN, Leach AG, Kim SP, Zhang X (2003) Binding affinities of host–guest, protein–ligand, and protein–transition-state complexes. Angew Chem Int Ed Engl 42:4872–4897
82. Fitter J (2003) A measure of conformational entropy change during thermal protein unfolding using neutron spectroscopy. Biophys J 84:3924–3930
83. Rubinson KA (1998) The polymer basis of kinetics and equilibria of enzymes: the accessible-volume origin of entropy changes in a class Aβ-lactamase. J Protein Chem 17(8):771–787
84. Christensen BE, Smidsrød O, Stokke BT (1996) Metastable, partially depolymerized xanthans and rearrangements toward perfectly matched duplex structures. Macromolecules 29:2939–2944
85. Levitt M (2014) Birth and future of multiscale modeling for macromolecular systems. Angew Chem Int Ed Engl 53:10006–10018
86. Bailey RT, North AM, Pethrick RA (1981, Ch. 13) Molecular motion in high polymers. Clarendon Press, Oxford
87. Richards EG (1980) An introduction to physical properties of large molecules in solution. Cambridge University Press, Cambridge
88. Frey E, Kroy K (2005) Brownian motion: a paradigm of soft matter and biological physics. Ann Phys (Leipzig) 14:20–50
89. Schulz GE (1992) Induced-fit movements in adenylate kinases. Farad Disc 93:85–93
90. Berry MB, Meador B, Bilderback T, Liang P, Glaser M, Phillips GN Jr (1994) The closed conformation of a highly fexible protein: the structure of E. coli adenylate kinase with bound AMP and AMPPNP. Proteins Struct Funct Genet 19:183–198
91. Natarajan K, McShan AC, Jiang J, Kumirov VK, Wang R, Zhao H, Schuck P, Tilahun ME, Boyd LF, Ying J, Bax A, Margulies DH, Sgourakis NG (2017) An allosteric site in the T-cell receptor Cb domain plays a critical signalling role. Nat Commun 8:15260
92. Jacobson H, Stockmayer WH (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems. J Chem Phys 18:1600–1606
93. Shoemaker BA, Wang J, Wolynes PG (1997) Structural correlations in protein folding funnels. Proc Natl Acad Sci USA 94:777–782
94. Rupley JA, Gratton E, Careri G (1983) Water and globular proteins. TIBS 8(1):18–22
95. Thanki N, Thornton JM, Goodfellow JM (1988) Distributions of water around amino acid residues in proteins. J Mol Biol 202:637–657
96. Menger FM (1993) Enzyme reactivity from and organic perspective. Acc Chem Res 26:206–212
97. Krokoszyńska I, Otlewski J (1996) Thermodynamic stability effects of single peptide bond hydrolysis in protein inhibitors of serine proteinases. J Mol Biol 256:793–802
98. Siebert X, Amzel LM (2004) Loss of translational entropy in molecular associations. Proteins 54(1):104–115
99. Fenimore PW, Frauenfelder H, McMahon BH, Yound RD (2004) Bulk-solvent and hydration-shell fluctuations, similar to α- and β-fluctuations in glasses, control protein motions and functions. Proc Natl Acad Sci USA 101(40):14408–14413
100. Ball P (2008) Water as an active constituent in cell biology. Chem Rev 180:74–108
101. Paciaroni A, Cornicchi E, Marconi M, Orecchini A, Petrillo C, Haertlein M, Moulin M, Sacchetti F (2009) Coupled relaxations at the protein–water interface in the picosecond time scale. J R Soc Interface 6:S635–S640
102. Biela A, Betz M, Heine A, Klebe G (2012) Water makes the difference: rearrangement of water solvation layer triggers non-additivity of functional group contributions in protein–ligand binding. ChemMedChem 7:1423–1434
103. Le Caër S, Klein G, Ortiz D, Lima M, Devineau S, Pin S, Brubach J-B, Roy P, Pommeret S, Leibl W, Righini R, Renault JP (2014) The effect of myoglobin crowding on the dynamics of water: an infrared study. Phys Chem Chem Phys 16:22841–22852
104. Nibali VC, Havenith M (2014) New insights into the role of water in biological function: studying solvated biomolecules using terahertz absorption spectroscopy in conjunction with molecular dynamics simulations. J Am Chem Soc 136:12800–12807
105. Pace CN, Scholtz JM, Grimsley GR (2014) Forces stabilizing proteins. FEBS Lett 588:2177–2184
106. Carugo O (2016) When proteins are completely hydrated in crystals. Intl J Biol Macromol 89:137–143
107. Krimmer SG, Cramer J, Betz M, Fridh V, Karlsson R, Heine A, Klebe G (2016) Rational design of thermodynamic and kinetic binding profiles by optimizing surface water networks coating protein-bound ligands. J Med Chem 59:10530–10548
108. Aoki K, Shiraki K, Hattori T (2016) Salt effects on the picosecond dynamics of lysozyme hydration water investigated by terahertz time-domain spectroscopy and an insight into the Hofmeister series for protein stability and solubility. Phys Chem Chem Phys 18:15060–15069
109. Comez L, Paolantoni M, Sassi P, Corezzi S, Morresi A, Fioretto D (2016) Molecular properties of aqueous solutions: a focus on the collective dynamics of hydration water. Soft Matter 12:5501–5514
110. Bellissent-Funel M-C, Hassanali A, Havenith M, Henchman R, Pohl P, Sterpone F, van der Spoel D, Xu Y, Garcia AE (2016) Water determines the structure and dynamics of proteins. Chem Rev 116:7673–7697
111. Jha A, Ishii K, Udgaonkar JB, Tahara T, Krishnamoorthy G (2011) Exploration of the correlation between solvation dynamics and internal dynamics of a protein. Biochemistry 50:397–408
112. Pal SK, Peon J, Zewail AH (2002) Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. Proc Natl Acad Sci USA 99(4):1763–1768
113. Qin Y, Zhang L, Wang L, Zhong D (2017) Observation of the global dynamic collectivity of a hydration shell around apomyoglobin. J Phys Chem Lett 8:1124–1131
114. Shiraga K, Ogawa Y, Kondo N (2016) Hydrogen bond network of water around protein investigated with terahertz and infrared spectroscopy. Biophys J 111:2629–2641
115. Ebbinghaus S, Kim SJ, Heyden M, Yu X, Heugen U, Gruebele M, Leitner DM, Havenith M (2007) An extended dynamical hydration shell around proteins. Proc Natl Acad Sci USA 104:20749–20752
116. Sushko O, Dubrovka R, Donnan RS (2015) Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater deistances than previously detected. J Chem Phys 142:055101
117. Ding T, Li R, Zeitler JA, Huber TL, Gladden LF, Middelberg APJ, Falconer RJ (2010) Terahertz and far infrared Spectroscopy of alanine-rich peptides having variable ellipticity. Opt Express 18(26):27431–27444

118. King JT, Arthur EJ, Brooks CL, Kubarych KJ (2013) Crowding induced collective hydration of biological macromolecules over extended distances. J Am Chem Soc 136:188–194
119. Glancy P, Beyermann WP (2010) Dielectric properties of fully hydrated nucleotides in the terahertz frequency range. J Chem Phys 132:245102
120. Heyden M, Bründermann E, Heugen U, Niehues G, Leitner DM, Havenith M (2008) Long-range influence of carbohydrates on the solvation dynamics of watersanswers from terahertz absorption measurements and molecular modeling simulations. J Am Chem Soc 130:5773–5779
121. Higgins MJ, Polcik M, Fukuma T, Sader JE, Nakayama Y, Jarvis SP (2006) Structured water layers adjacent to biological membranes. Biophys J 91:2532–2542
122. Grossmann C, Tintinger R, Zhu J, Maurer G (1995) Aqueous two-phase systems of poly(ethylene glycol) and dextran–experimental results and modeling of thermodynamic properties. Fluid Phase Equilib 106:111–138
123. Basdogan Y, Keith JA (2018) A paramedic treatment for modeling explicitly solvated chemical reaction mechanisms. Chem Sci 9:5341
124. Pashley RM, Kitchener JA (1979) Surface forces in adsorbed multilayers of water on quartz. J Colloid Interface Sci 71(3):491–500
125. Soper AK (2007) Joint structure refinement of x-ray and neutron diffraction data on disordered materials: application to liquid water. J Phys 19:335206
126. J-m Zheng, Chin W-C, Khijniak E, Khijniak E Jr, Pollack GH (2006) Surfaces and interfacial water: evidence that hydrophilic surfaces have long-range impact. Adv Colloid Interface Sci 127:19–27
127. Urry DW (1997) Physical chemistry of biological free energy transduction as demonstrated by elastic protein-based polymers. J Phys Chem B 101:11007–11028
128. Grunwald E, Comeford LL (1995) Thermodynamic mechanisms for enthalpy-entropy compensation. In: Gregory RB (ed) Protein-solvent interactions. Marcel Dekker, New York, pp 421–443
129. Sharp K (2001) Entropy–enthalpy compensation: fact or artifact? Prot Sci 10(3):661–667
130. Lumry R (2003) Uses of enthalpy–entropy compensation in protein research. Biophys Chem 105:545–557
131. Mammen M, Choi S-K, Whitesides GM (1998) Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors. Angew Chem Int Ed Engl 37:2754–2794
132. Williams DH, Stephens E, O'Brien DP, Zhou M (2004) Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. Angew Chem Int Ed Engl 43:6596–6616
133. Oshovsky GV, Reinhoudt DN, Verboom W (2007) Supramolecular chemistry in water. Angew Chem Int Ed Engl 46:2366–2393
134. Cram DJ (1986) Preorganization—from solvents to spherands. Angew Chem Int Ed Engl 25(12):1039–1057
135. Martell AE, Hancock RD, Motekaitis RJ (1994) Factors affecting stabilities of chelate, macrocyclic, and macrobicyclic complexes in solution. Coord Chem Rev 133:39–65
136. Piguet C (2010) Five thermodynamic describers for addressing serendipity in the self-assembly of polynuclear complexes in solution. Chem Commun 46:6209–6231
137. DiMaio J, Gibbs B, Munn D, Lefebvre J, Konishi Y (1990) Bifunctional thrombin inhibitors based on the sequence of hirudin$^{45\text{-}65}$. J Biol Chem 265(35):21698–21703
138. Fersht AR (1997) Nucleation mechanisms in protein folding. Curr Opin Struct Biol 7:3–9
139. Pappu RV, Srinivasan R, Rose GD (2000) The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. Proc Natl Acad Sci USA 97:12565–12570
140. Fitzkee NC, Rose GD (2004) Reassessing random-coil statistics in unfolded proteins. Proc Natl Acad Sci USA 101:12497–12502
141. Kauffman S, Levin S (1987) Towards a general theory of adaptive walks on rugged landscapes. J Theor Biol 128(1):11–45
142. Figliuzzi M, Jacquier H, Schug A, Tenaillon O, Weigt M (2016) Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1. Mol Biol Evol 33(1):268–280
143. Hopf TA, Ingraham JB, Poelwijk FJ, Schärfe CPI, Springer M, Sander C, Marks DS (2017) Mutation effects predicted from sequence co-variation. Nat Biotech 35(2):128–135
144. Izatt RM, Pawlak K, Bradshaw JS, Bruening RL (1991) Thermodynamic and kinetic data for macrocycle interactions with cations and anions. Chem Rev 91(8):1721–2085
145. Lashley MA, Ivanov AS, Bryantsev VS, Dai S, Hancock RD (2016) Highly preorganized ligand 1,10-phenanthroline-2,9-dicarboxylic acid for the selective recovery of uranium from seawater in the presence of competing vanadium species. Inorg Chem 55:10818–10829
146. Schneider H-J (2009) Binding mechanisms in supramolecular complexes. Angew Chem Int Ed Engl 48:3924–3977
147. Page MI, Jencks WP (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. Proc Natl Acad Sci USA 68:1678–1683
148. Newberry RW, Raines RT (2016) A prevalent intraresidue hydrogen bond stabilizes proteins. Nat Chem Biol 12(12):1084–1088
149. Jencks WP (1981) On the attribution and additivity of binding energies. Proc Natl Acad Sci USA 78(7):4046–4050
150. Cabbiness DK, Margerum DW (1969) Macrocyclic effect on the stability of copper(II) tetramine complexes. J Am Chem Soc 91(23):6540–6541
151. Sokol LSWL, Ochrymowycz LA, Rorabacher DB (1981) Macrocyclic, ring size, and anion effects as manifested in the equilibrium constants and thermodynamic parameters of copper(II)-cyclic polythia ether complexes. Inorg Chem 20:3189–3195
152. Mandolini L (1986) Intramolecular reactions of chain molecules. Adv Phys Org Chem 22:1–111
153. Illuminati G, Mandolini L, Masci B (1977) Ring-closure reactions. 9. Kinetics of ring formation from *o*-ω-bromoalkoxy phenoxides and *o*-ω-bromoalkyl phenoxides in the range of 11- to 24-membered rings. A Comparison with related cyclization series. J Am Chem Soc 99(19):6308–6312
154. Smith GF, Margerum DW (1975) Diminution of the macrocyclic effect for nickel(II) complexes of thioethers in nonaqueous solvents. J C S Chem Commun 807–808
155. Schwarzenbach G (1952) Der chelateffect. Helv Chim Acta 35(7):2344–2359
156. Rubinson KA (2014) Small-angle neutron scattering of aqueous SrI_2 suggests a mechanism for ion transport in molecular water. J Solut Chem 43:453–464
157. Evstigneev MP, Lantushenkoa AO, Golovchenkoa IV (2016) Hidden entropic contribution in the thermodynamics of molecular complexation. Phys Chem Chem Phys 18:7617–7626
158. Faver JC, Benson ML, He X, Roberts BP, Wang B, Marshall MS, Sherrill CD, Merz KMJ (2011) The energy computation paradox and ab initio protein folding. PLoS ONE 6(4):e18868

159. Rabinowitch E (1937) Collision, co-ordination, diffusion and reaction velocity in condensed systems. Trans Farad Soc 33:1225–1233
160. Laidler KJ (1987) Chemical kinetics. Harper & Row Publishers, New York
161. Feller W (1968) An introduction to probability theory and its applications, vol. I, 3rd edn. Wiler, New York, pp 67–97
162. Owusu-Apenten RK (1995) A three-state heat-denaturation of bovine α-lactalbumin. Food Chem 52:131–133
163. Wong K-B, Freund SMV, Fersht AR (1996) Cold denaturation of barstar: ^{1}H, ^{15}N and ^{13}C NMR assignment and characterisation of residual structure. J Mol Biol 259:805–818
164. Shortle D (1996) The denatured state (the other half of the folding equation) and its role in protein stability. FASEB J 10(1):27–34
165. Zaidi FN, Nath U, Udgaonkar JB (1997) Multiple intermediates and transition states during protein unfolding. Nat Struct Biol 4(12):1016–1024
166. Zocchi G (1997) Proteins unfold in steps. Proc Natl Acad Sci USA 94:10647–10651
167. Bowler BE (2007) Thermodynamics of protein denatured states. Mol BioSyst 3:88–99
168. Jensen MR, Markwick PRL, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernado P, Blackledge M (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. Structure 17:1169–1185
169. Stefanowicz P, Petry-Podgorska I, Kowelewska K, Jaremko L, Jreemko M, Szewczuk Z (2010) Electrospray ionization mass spectrometry as a method for studying the high-pressure denaturation of proteins. Biosci Rep 30:91–99
170. Receveur-Bréchot V, Durand D (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. Curr Prot Peptide Sci 13:55–75
171. Lumry R, Biltonen R (1966) Validity of the "two-state" hypothesis for conformational transitions of proteins. Biopolymers 4:917–944
172. Amor BRC, Schaub MT, Yaliraki SN, Barahona M (2016) Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. Nat Commun 7:12477
173. Rajasekaran N, Suresh S, Gopi S, Raman K, Naganathan AN (2017) A general mechanism for the propagation of mutational effects in proteins. Biochemistry 56:294–305
174. Ayaz P, Munyoki S, Geyer EA, Piedra1 F-A, Vu1 ES, Bromberg R, Otwinowski Z, Grishin NV, Brautigam CA, Rice LM (2014) A tethered delivery mechanism explains the catalytic action of a microtubule polymerase. eLife 3:03069
175. Hamilton CL, Niemann C, Hammond GS (1966) A quantitative analysis of the binding of *N*-acyl derivatives of α-aminoamides by α-chymotrypsin. Proc Natl Acad Sci USA 55(3):664–669
176. Niemann C (1964) Alpha-chymotrypsin and the nature of enzyme catalysis. Science 143:1287–1296
177. Vitagliano L, Merlino A, Zagari A, Mazzarella L (2000) Productive and nonproductive binding to ribonuclease A: X-ray structure of two complexes with uridylyl(2′,5′)guanosine. Protein Sci 9:1217–1225
178. Freire E (1999) The propagation of binding interactions to remote sites in proteins: analysis of the binding of the monoclonal antibody D1.3 to lysozyme. Proc Natl Acad Sci USA 96:10118–10122
179. Pan H, Lee JC, Hilser VJ (2000) Binding sites in *Escherichia coli* dihydrofolate reductase communicate by modulating the conformational ensemble. Proc Natl Acad Sci USA 97(22):12020–20125
180. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of *all* dynamic proteins? Proteins 57:433–443
181. Long D, Brüschweiler R (2011) Atomistic kinetic model for population shift and allostery in biomolecules. J Am Chem Soc 133:18999–19005
182. Wei G, Xi W, Nussinov R, Ma B (2016) Protein ensembles: how does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell. Chem Rev 116:6516–6551
183. Dagliyan O, Tarnawski M, Chu P-H, Shirvanyants D, Schlichting I, Dokholyan NV, Hahn KM (2016) Engineering extrinsic disorder to control protein activity in living cells. Science 354(6318):1441–1444
184. Segel IH (1993) Enzyme kinetics. Wiley, New York
185. Neher E, Sakmann B (1976) Single-channel currents recorded from membrane of denervated frog muscle fibres. Nature 760(April 29):799–802
186. Keller BU, Hartshorne RP, Talvenheimo JA, Catterall WA, Montal M (1986) Sodium channels in planar lipid bilayers: channel gating kinetics of purified sodium channels modified by batrachotoxin. J Gen Physiol 88:1–23
187. Rubinson KA (1992) Steady-state kinetics of solitary bachtrachotoxin-treated sodium channels. Kinetics on a bounded continuum of polymer conformations. Biophys J 61:463–479
188. Eggeling C, Fries JR, Brand L, Günther R, Seidel CAM (1998) Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy. Proc Natl Acad Sci USA 95:1556–1561
189. Xie XS, Lu HP (1999) Single-molecule enzymology. J Biol Chem 274:15967–15970
190. English BP, Min W, van Oijen AM, Lee KT, Luo G, Sun H, Cherayil BJ, Kou SC, Xie S (2006) Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. Nat Chem Biol 2(2):87–94
191. Volkman BF, Lipson D, Wemmer DE, Kern D (2001) Two-state allosteric behavior in a single-domain signaling protein. Science 291(23 March):2429–2433
192. Lisi GP, Loria JP (2016) Solution NMR spectroscopy for the study of enzyme allostery. Chem Rev 116:6323–6369
193. Otrusinová O, Demo G, Padrta P, Jaseňáková Z, Pekárová B, Gelová Z, Szmitkowska A, Kadeřávek P, Jansen S, Zachrdla M, Klumpler T, Marek J, Hritz J, Janda L, Iwaï H, Wimmerová M, Hejátko J, Žídek L (2017) Conformational dynamics are a key factor in signaling mediated by the receiver domain of a sensor histidine kinase from *Arabidopsis thaliana*. J Biol Chem 292(42):17525–17540
194. Koshland DE (1994) The key-lock theory and the induced fit theory. Angew Chem Int Ed Eng 33:2375–2378
195. Warshel A (1978) Energetics of enzyme catalysis. Proc Natl Acad Sci USA 75(11):5250–5254
196. Liu H, Warshel A (2007) The catalytic effect of dihydrofolate reductase and its mutants is determined by reorganization energies. Biochemistry 46:601–6025
197. Fried SD, Bagchi S, Goxer SG (2014) Extreme electric fields power catalysis in the active site of ketosteroid isomerase. Science 346(6216):1510–1514
198. Fried SD, Boxer SG (2017) Electric fields and enzyme catalysis. Annu Rev Biochem 86:387–415
199. Morgenstern A, Jaszai M, Eberhart ME, Alexandrova AN (2017) Quantified electrostatic preorganization in enzymes using the geometry of the electron charge density. Chem Sci 8:5010–5018
200. Blomberg MRA, Siegbahn PEM (2010) Quantum chemistry as a tool in bioenergetics. Biochim Biophys Acta 1797:129–142
201. Kim KH, Kim JG, Nozawa S, Sato T, Oang KY, Kim TW, Jo HKJ, Park S, Song C, Sato T, Ogawa K, Togashi T, Tono K, Yabashi M, Ishikawa T, Kim J, Ryoo R, Kim J, Ihee H, S-i

Adachi (2015) Direct observation of bond formation in solution with femtosecond X-ray scattering. Nature 518(7539):385–389
202. Shiró G, Natali F, Cupane A (2012) Physical origin of anharmonic dynamics in proteins: new insights from resolution-dependent neutron scattering on homomeric polypeptides. Phys Rev Lett 109:128102
203. Morresi A, Mariani L, Distefano MR, Giorgini MG (1995) Vibrational relaxation processes in isotropic molecular liquids. A critical comparison. J Raman Spec 26:179–216
204. Lambert FL (2002) Entropy is simple, qualitatively. J Chem Educ 79(10):1241–1246
205. Takeuchi H, Allen G, Suzuki S, Dianoux AJ (1980) Low frequency vibraions in solid *n*-butane and *n*-hexane by incoherent inelastic neutron scattering. Chem Phys 51:197–203
206. Giraud G, Karolin J, Wynne K (2003) Low-frequency modes of peptides and globular proteins in solution observed by ultrafast OHD-RIKES spectroscopy. Biophys J 85:1903–1913
207. Levitt M, Sander C, Stern PS (1985) Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 181:423–447
208. Silverstein RM, Bassler GC, Morrill TC (1974) Spectrometric identification of organic compounds, 3rd edn. Wiley, Ch, p 3
209. Schachtschneider JH, Snyder RG (1963) Vibrational analysis of the n-paraffins-II. Normal co-ordinate calculations. Spectrochim Acta 19:117–168, esp. Table 114
210. Scott DW, El. Sabban MZ (1969) A valence force field for aliphatic sulfur compounds: alkanethiols and thioalkanes. J Mol Spectrosc 30:317–337, esp. Table II
211. Herzberg G (1950) Spectra of diatomic molecules, 2nd edn. Van Nostrand, Princeton, NJ, p 91
212. Ginn SGW, Wood JL (1967) The intermolecular stretching vibration of some hydrogen-bonded complexes. Spectrochim Acta 23A:611–625
213. Miyazawa T, Pitzer KS (1959) low frequency vibrations, polarizability and entropy of carboxylic acid dimers. J Am Chem Soc 81:74–79
214. Matsushima N, Hikichi K, Tsutsumi A, Kaneko M (1976) X-ray scattering of synthetic poly(α-amino acid)s in the solid state. III. Temperature dependence of the 1.5 Å-meridional reflection of the α-helix. Polymer J 8:88–95
215. Bustamante C, Smith SB, Liphardt J, Smith D (2000) Single-molecule studies of DNA mechanics. Curr Opin Struct Biol 10:279–285
216. Efimova YM, Haemers S, Wierczinski B, Norde W, van Well AA (2006) Stability of globular proteins in H_2O and D_2O. Biopolymers 85(3):264–273
217. Lassalle MW, Yamada H, Akasaka K (2000) The pressure-temperature free energy-landscape of staphylococcal nuclease monitored by ^1H NMR. J Mol Biol 298:293–302
218. Klotz IM (1996) Equilibrium constants and free energies in unfolding of proteins in urea solutions. Proc Natl Acad Sci USA 93:14411–14415
219. Ahmad F, Bigelow CC (1982) Estimation of the free energy of stabilization of ribonuclease A, lysozyme, α-lactalbumin, and myoglobin. J Biol Chem 257(21):12935–12938
220. Nojima H, Ikai A, Oshima T, Noda H (1977) Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero enthalpy change. J Mol Biol 116:429–442
221. Privalov PL, Khechinashvili NN (1974) A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. J Mol Biol 86:665–684
222. Beadle BM, McGovern SL, Patera A, Shoichet BK (1999) Functional analyses of AmpC β-lactamase through differential stability. Prot Sci 8:1816–1824