ELSEVIER

Biophysical Chemistry 100 (2003) 367–395

Biophysical Chemistry

www.elsevier.com/locate/bpc

# Thermodynamics of protein folding: a microscopic view

Themis Lazaridis[a,*], Martin Karplus[b,c,1]

[a]Department of Chemistry, City College of New York, Convent Ave & 138th Street, New York, NY 10031, USA
[b]Department of Chemistry & Chemical Biology, Harvard University, Cambridge, MA 02138, USA
[c]Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France

Received 15 April 2002; accepted 18 June 2002

**Abstract**

Statistical thermodynamics provides a powerful theoretical framework for analyzing, understanding and predicting the conformational properties of biomolecules. The central quantity is the potential of mean force or effective energy as a function of conformation, which consists of the intramolecular energy and the solvation free energy. The intramolecular energy can be reasonably described by molecular mechanics-type functions. While the solvation free energy is more difficult to model, useful results can be obtained with simple approximations. Such functions have been used to estimate the intramolecular energy contribution to protein stability and obtain insights into the origin of thermodynamic functions of protein folding, such as the heat capacity. With reasonable decompositions of the various energy terms, one can obtain meaningful values for the contribution of one type of interaction or one chemical group to stability. Future developments will allow the thermodynamic characterization of ever more complex biological processes.

*Keywords:* Energy functions; Solvation; Protein stability; Denaturation

## 1. Prologue (M. Karplus)

Although I had been aware of John Edsall when I was a Harvard undergraduate (He was the biochemistry tutor of my classmate, Gary Felsenfeld), I first ‘met’ John through the wonderful volume ‘*Biophysical Chemistry*’, published by him and Jeffries Wyman in 1958 [1]. It was one of several books which played a formative role in my view of chemistry and the realization that chemistry has an essential part to play in understanding living systems. I became interested in biology as a teenager, and when I went to Harvard as an undergraduate in Chemistry and Physics, I eagerly wanted to believe that the concepts of physical chemistry could be used to understand biological systems. An important book in my education was Linus Pauling’s ‘*The Nature of the Chemical Bond*’ [2], which showed me that chemistry, particularly structural chemistry, could be understood and not just memorized. Schrödinger’s ‘*What is Life*’ [3] was inspirational in that it proposed a rational view of living processes, although much of it is no longer valid. Edsall and Wyman’s

*Corresponding author. Tel.: +1-212-650-8364; fax: +1-212-650-6107.
*E-mail address:* themis@sci.ccny.cuny.edu (T. Lazaridis).

[1] Also corresponding author. Tel.: +1-617-495-4081; fax: +1-617-496-3204; e-mail: marci@tammy.harvard.edu (M. Karplus).

0301-4622/03/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0301-4622(02)00293-4

'*Biophysical Chemistry*', which I first encountered when I was an Instructor at the University of Illinois, is quite a different book. It was important when it was published because it defined the field and most of what it says is as valid today as when it was written, so that it still serves as a useful reference in the field. What is missing, of course, are specific applications to biological macromolecules, which were to be in the never-finished Volume II. (We must remember that Volume I was written before the first protein structure was solved.) In the main body of this birthday article, we shall describe some of the modern applications of thermodynamics to proteins that could have been in Volume II.

Before doing so, let me describe my first actual meeting with John. It occurred shortly after I had heard Max Perutz describe his mechanism [4,5] for hemoglobin cooperativity in lectures at MIT. After some discussion with him, Attila Szabo and I began to formulate the 'Perutz mechanism' in mathematical form. This was my first attempt at interpreting measured biophysical parameters, such as binding constants, rates of reactions, and their dependence on the environmental conditions, like pH. John and subsequently Guido Giddotti, to whom I was referred by John, became our counselors in trying to untangle the confusing and often contradictory literature in the immense hemoglobin field. Again and again, when I was stuck with inconsistent measurements for what was apparently the same system, they came to my rescue by telling me which article I could believe. I am not sure their advice was always correct, but without it Attila and I would never have completed our model for cooperativity in hemoglobin nor its applications to interpreting a large body of thermodynamic data in structural terms [6–8].

After the hemoglobin work, John and I mainly had contact when we met each other in the courtyard of the biology building as I was walking to work. We would always stop for a short chat (longer if the weather was good) and I was amazed by John's insightful questions on some of my own recently published work or his comment on something in the literature that I might find of interest. Most recently, John and I interacted more intensely through his editorial role in the *Advances in Protein Chemistry*. It concerned the final *Advances* that John edited, the last of 47 volumes which appeared between 1944 and 1995, and it turned out to be a difficult one. A pair of articles, which dealt with thermodynamics of protein folding, were to be published in the same volume of *Advances*, one by G.I. Makhatadze and P.L. Privalov [9] and the other by T. Lazaridis, G. Archontis and myself [10]. Although the data we used were the same (i.e. those measured by Privalov and co-workers), the methods of analysis and conclusions were very different. It required the best of John's diplomacy and wisdom, with some aid from David Eisenberg, to make possible a publication that presented the disagreement in a constructive way.

In what follows, we shall review some of our recent work on the thermodynamics of proteins. This is a subject that has always been of interest to John (see his commentary [11] on the 1931 article on protein denaturation by H. Wu [12], republished in the *Advances* of 1995) and we make use of some of the ideas in Edsall's and Wyman's '*Biophysical Chemistry*'.

## 2. Introduction: the relevance of equilibrium thermodynamics to biological phenomena

It has long been assumed that biological systems obey the natural laws of physics and chemistry. This assumption continues to be supported as more details are learned concerning how living organisms function and the highly complex interactions involved in many essential processes. There is now an enormous amount of information on the events that take place in living systems at the molecular level. However, much of this information is qualitative and descriptive, even when the components involved are known and the structures of many of them (proteins and nucleic acids) have been determined. Many ingenious experiments have been done to establish which phenomena take place, but most of them do not address the question of why things happen the way they do. This is where the physical sciences, including thermodynamics, can make an essential contribution to biology.

Biology has not always had a comfortable relation with thermodynamics because of seemingly irreconcilable differences between the simple, physical processes traditionally described by thermodynamics and the complex processes that occur in biology. Simple systems rapidly evolve towards equilibrium, which for an isolated system is characterized by maximum entropy (maximum disorder). In contrast, living systems never reach equilibrium and in many cases evolve towards states of increasing order. For example, biological development involves the growth of a complex multicellular organism out of a single cell. A complete thermodynamic analysis of such a complex order-generating process is a challenge to thermodynamicists [13]. It is easy to argue that such phenomena do not contradict the laws of thermodynamics, simply by virtue of the fact that biological systems are open systems far from equilibrium and, although order is generated in the biological system, the entropy of the universe as a whole may still increase [13]. However, this argument shows only that biological ordering can arise (which we already know by experience); it does not explain why it occurs. Why out of all possible ways of increasing the entropy of the universe nature chooses one that involves creation of order in a few of its subsystems? A complete analysis of biological phenomena in terms of thermodynamics does not yet exist. Some interesting studies of evolutionary mechanisms and the appearance of order in biological systems have been made both experimentally and theoretically [14–16].

The question of how biological systems evade equilibrium was taken up by Schrödinger in his influential treatise '*What is Life*' [3], already mentioned in the Prologue. He proposed that living organisms feed on negative entropy which they import from their surroundings. Actually, import of energy in an appropriate form (other than thermal) can also be used to maintain a system far from equilibrium. The absorption of photons by the photosynthetic apparatus of plants provides the motive power of life on earth. This energy gradually 'trickles' down to other parts of the plant and from plants to animals through the food cycle. The efficiency of the whole process is high because the energy is maintained in the form of chemical energy, i.e. it does not generate much unrecoverable thermal energy. At a more microscopic level, it is likely that all individual processes in the cell evolve towards equilibrium but equilibrium is not reached because the external conditions change and shift the equilibrium point to some other state. For example, the folded conformation of a protein may be the lowest free energy state at neutral pH and indeed the protein evolves towards the folded state at neutral pH. If at some future point the pH is changed, the protein will evolve towards its new equilibrium state, which may be the unfolded state. In fact, differences in pH in different parts of an organism play an important role in biological processes, including the entrance of viruses into cells [17]. That biological macromolecules have marginal stability is in part due to their need to respond readily to changes in the environmental conditions. It should be noted that the lifetime of most protein molecules is generally much shorter than that of the organism, although there are some interesting exceptions. (The eye len crystallins apparently do not turn over.) Apparently, processes in the cell are dynamically coupled in such a way that overall equilibrium is never reached in a living system. This subject is under active investigation today, as we learn more about cellular networks and the coupling of reactions that used to be treated in isolation [18].

Despite the fact that biological processes do not occur under equilibrium conditions, equilibrium thermodynamics has been extensively used in biochemistry. In the usual reductionistic fashion, biochemists have isolated specific processes and studied them in vitro under equilibrium conditions. One reason for this is that the measurements are much easier under equilibrium conditions. One example of a process that has been subjected to detailed biophysical characterization is protein folding. It was shown by Anfinsen [19] that ribonuclease A folds spontaneously to its native state in vitro. Already in 1934, Anson and Mirsky [20] demonstrated that a protein precipitate could be redissolved and the function of the protein regenerated, presaging the work of Anfinsen; a popular description of this experiment was given by Perutz under the title 'Unboiling an Egg' [21]. It might also be mentioned that an educational film was made by Robert Karplus of 'unfrying' an egg, by simply running the actual cooking event backwards; this would be scientifically correct if

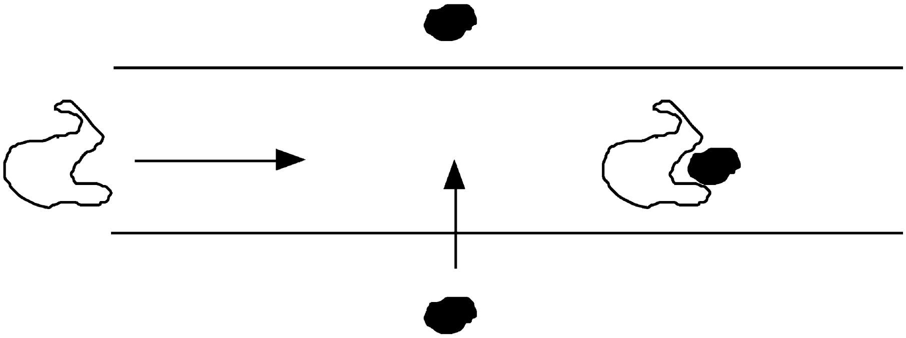

Fig. 1. Non-equilibrium ligand binding.

irreversibility did not enter so reversing the time could recreate what had existed earlier. Experiments suggest that many processes (or at least the process of protein folding) do evolve towards equilibrium in the cell and are similar, therefore, to the physical processes studied by traditional thermodynamics in simple media. Changes inside the cell which are uphill in free energy are coupled to others which are downhill in free energy (usually hydrolysis of ATP) so that the overall process is spontaneous [22].

Equilibrium thermodynamic measurements can be very useful and relevant to the situation in vivo because despite the lack of overall equilibrium in the cell, there can exist partial equilibrium, either temporal (within a certain time scale) or spatial (within a certain region). Nevertheless, one has to keep in mind the limitations of such measurements. One hypothetical example is depicted in Fig. 1. A protein flows in a channel and a ligand crosses the channel and has the opportunity to bind to the protein. It is of interest to know what percentage of ligand is taken up by the protein and the percentage of the protein that binds the ligand. Measurement of the equilibrium binding constant will be useful only if the kinetics of binding is fast relative to the rate of transport of the protein and the ligand. Otherwise, it will be completely irrelevant and a kinetic analysis would be required.

Classical thermodynamics is a macroscopic science which is not concerned with the microscopic nature of matter. Since biological phenomena depend critically on the nature of the molecules involved, thermodynamics is most useful in biology when combined with a microscopic description. Molecular models that describe the interactions within and between molecules are critically important. The link between microscopic interactions and macroscopic properties is provided by statistical mechanics. Statistical mechanics considers an ensemble of microscopic systems consistent with a given set of macroscopic conditions. The thermodynamic internal energy is obtained as a Boltzmann-weighted average of the energy of all these microscopic states. The entropy and free energy are obtained from the partition function (the sum of the Boltzmann factors of the microscopic states). Direct estimation of the partition function is possible only for very simple systems. For realistic systems, approximations need to be made. In this paper, we illustrate the microscopic approaches by focusing on the thermodynamics of protein folding and its interpretation in terms of the interactions involved. We start with a theoretical framework for the stability of macromolecules that provides information concerning the interactions that contribute to the free energy of folding. These are analyzed in subsequent sections.

## 3. Stability of macromolecular conformations

Biopolymers, such as proteins and nucleic acids, adopt essentially unique conformations under physiological conditions. One central question is whether this conformation is under thermodynamic or kinetic control; i.e. whether the native protein conformation corresponds to the most stable (thermodynamic control) or to the kinetically most accessible (kinetic control) conformation; for a discussion, see for example Wetlaufer and Ristow [23] and Anfinsen and Scheraga [24]. The first situation would arise if the barriers for interconversion of conformations were small enough to be traversed within experimental time scales. Thus, equilibrium is established and the conformations are populated according to the Boltzmann distribution. On the other hand, if the barriers between conformations are too high, the system is no longer ergodic (i.e. it does not sample all possible states in accord with the Boltzmann distribution) and the macromolecule ends up in the lowest local minimum it can find within the available time. There has been no proof of either of the two situations. The original experiments of Anfinsen seem to support the thermodynamic hypothesis. Recent experiments, however, have suggested that there may be exceptions, especially for larger and more complex proteins [25–27]; care has to be taken in analyzing such experiments to avoid the complications of autolysis and covalent modifications. It is also of interest to mention that such metastability has been found in a highly simplified, but detailed, simulation model of protein folding [28]. It seems likely that kinetic control is more prevalent in complex cellular processes. For example, the migration of lipids from one leaflet of a lipid bilayer to the other can take days, so that a non-equilibrium composition in biological membranes can be maintained throughout the lifetime of a cell [29].

Equilibrium thermodynamic concepts have an important role even if the thermodynamic hypothesis is not valid in all cases because partial equilibrium can exist with respect to certain degrees of freedom, even when overall equilibrium is not established. One example concerns the solvent degrees of freedom. For most changes of macromolecular conformations the solvent equilibrates on a picosecond timescale [30]. For configurations where water needs to equilibrate between the bulk and an internal cavity, equilibration can take much longer [31,32], but rarely longer than the experimental time scales of seconds or minutes. Because in most cases (but see the recent work by P. Rein ten Wolde and D. Chandler, private communication) equilibration of the solvent is rapid compared to movements of the macromolecule, the solvent degrees of freedom can be integrated out to give the equilibrium solvation free energy. This quantity is then added to the internal macromolecular energy to give the 'effective energy' or 'potential of mean force' for each macromolecular conformation. The effective energy defines a hypersurface in the conformational space of the molecule (the 'energy landscape' [33]) whose shape determines the conformational properties of the macromolecule, independent of whether it is under thermodynamic control with respect to the macromolecular degrees of freedom.

Formal integration of the solvent degrees of freedom can be accomplished using standard statistical mechanical methods [34]. We briefly review what is involved, using the notation of Lazaridis and Karplus [35]. Consider a macromolecule consisting of $M$ atoms with Cartesian coordinates $(\mathbf{R}_i)=(X_i,Y_i,Z_i)$, $i=1,M$ and internal coordinates $\mathbf{q}=(q_i)$, $i=1,\ldots,3M-6$. The macromolecule is immersed in a bath of $N$ rigid solvent molecules with coordinates $r_i=(x_i,y_i,z_i,\omega_i,\phi_i,\chi_i)$, $i=1,N$ where $x,y,z$ are Cartesian coordinates of the center of mass and $\omega,\phi,\chi$ are the Euler angles specifying the orientation. For simplicity we assume constant temperature and volume conditions corresponding to the canonical ensemble. The canonical partition function is

$$Q=\frac{Z}{N!\Lambda^{3M}\Lambda^{3N}} \tag{1}$$

where $Z$ is the classical configurational integral

$$Z=\int\exp(-\beta H)\,\mathrm{d}\mathbf{r}^N\mathrm{d}\mathbf{R}^M \tag{2}$$

with $H$ the Hamiltonian and $\beta=1/kT$. The Helmholtz free energy is given by

$$A=-kT\ln Q=-kT\ln Z+kT\ln(N!\Lambda^{3M}\Lambda^{3N}) \tag{3}$$

One can formally perform the integration over the solvent coordinates in Eq. (2) by defining the potential of mean force, $W$, as

$$\exp(-\beta W)=Z_{\mathrm{ww}}^{-1}\int\exp(-\beta H)\,\mathrm{d}\mathbf{r}^N \tag{4}$$

where

$$Z_{ww}=\int \exp(-\beta H_{ww})\,d\mathbf{r}^N, \tag{5}$$

is the pure solvent configurational integral. Introducing $W$, the configurational integral can be written

$$Z=Z_{ww}\int \exp(-\beta W)\,d\mathbf{R}^M \tag{6}$$

Thus, the integral in Eq. (6) depends explicitly only on the macromolecule degrees of freedom $R$. If the Hamiltonian is additive, as it is in most molecular mechanics force fields [36], further explicit simplifications are possible. We can write $H$ as $H=H_{mm}+H_{mw}+H_{ww}$, where the three components are the intra-macromolecule, macromolecule–solvent, and solvent–solvent interactions, respectively. We then obtain

$$\exp(-\beta W)=\exp(-\beta H_{mm})Z_{ww}^{-1}\int \exp(-\beta H_{mw}-\beta H_{ww})\,d\mathbf{r}^N \tag{7}$$

so that

$$W=H_{mm}+X \tag{8}$$

where

$$\exp(-\beta X)=Z_{ww}^{-1}\int \exp(-\beta H_{mw}-\beta H_{ww})\,d\mathbf{r}^N \tag{9}$$

Eq. (9) for $X$ can be written as

$$X=-kT\ln\langle\exp(-\beta H_{mw})\rangle_o\equiv\Delta G^{slv} \tag{10}$$

where the ensemble average $(o)$ is taken over the pure solvent. Eq. (10) is the familiar expression for the excess chemical potential, or standard solvation free energy [37]. Thus, the potential of mean force consists of two terms: the intra-macromolecular energy and the solvation free energy. Instead of the term 'potential of mean force', we can use the more intuitive term '*effective energy*'. The function $W$ defines a hypersurface in the conformation space of the macromolecule in the presence of equilibrated solvent and, therefore, includes the solvation entropy. This hypersurface is now often called an 'energy landscape'. It determines the thermodynamics and kinetics of macromolecular conformational transitions.

This separation of the effective energy can be accomplished formally even in the case of non-pairwise additive potentials. For example, in the presence of three-body forces the Hamiltonian can be written

$$H=(H_{mm}+H_{mmm})+(H_{mw}+H_{ww}+H_{www}+H_{wwm}+H_{mmw}) \tag{11}$$

and the effective energy would be

$$W=(H_{mm}+H_{mmm})-kT\ln\langle\exp(-\beta H_{mw}-\beta H_{mmw}-\beta H_{mww})\rangle_o \tag{12}$$

However, for this case the actual evaluation of $W$ is much more difficult.

For a formal description of the macromolecule it is often more convenient to use the internal coordinates, $\mathbf{q}$. The Jacobian for the transformation depends only on bond lengths and bond angles and is, therefore, approximately constant for all conformations and can be taken out of the integral in Eq. (6). Here we include it in the notation $d\mathbf{q}$. The integration over the six external coordinates can be performed since the system is homogeneous, to give $V8\pi^2$, so that

$$Z=Z_{ww}V8\pi^2\int \exp(-\beta W)\,d\mathbf{q} \tag{13}$$

It can be shown that the probability of finding the system at the configuration $(\mathbf{q})$ is [35]

$$p(\mathbf{q})=\frac{\exp(-\beta W(\mathbf{q}))}{\int \exp(-\beta W(\mathbf{q}))\,d\mathbf{q}} \tag{14}$$

Consequently,

$$\begin{aligned}\int p(\mathbf{q})\ln p(\mathbf{q})d\mathbf{q}&=\int p(\mathbf{q})\{-\beta W(\mathbf{q})-\ln Z+\ln Z_{ww}+\ln V8\pi^2\}\,d\mathbf{q}\\&=-\ln Z+\ln Z_{ww}+\ln V8\pi^2-\beta\int p(\mathbf{q})W(\mathbf{q})d\mathbf{q}\end{aligned} \tag{15}$$

and from Eq. (3)

$$
\begin{aligned}
A &= A^{\circ} + kT\ln\left(\frac{\Lambda^{3M}}{V8\pi^2}\right) + \int p(\mathbf{q})\{H_{\mathrm{mm}}(\mathbf{q}) \\
&\quad + \Delta G^{\mathrm{slv}}(\mathbf{q})\}\mathrm{d}\mathbf{q}\; kT\int p(\mathbf{q})\ln p(\mathbf{q})\mathrm{d}\mathbf{q} \\
&= A^{\circ} + kT\ln\left(\frac{\Lambda^{3M}}{V8\pi^2}\right) + \langle W\rangle - TS^{\mathrm{conf}}
\end{aligned}
\tag{16}
$$

where $A^{\circ}$ is the free energy of the pure solvent and the second term is the ideal contribution from macromolecular translation and rotation. The third term in Eq. (16) is the average effective energy, which is equal to the average intramolecular energy plus the average solvation free energy. The last term is the contribution of the configurational entropy of the macromolecule to the free energy. The solvent entropy is contained in $\Delta G^{\mathrm{slv}}(\mathbf{q})$ and in $p(\mathbf{q})$. The Gibbs free energy is equal to the Helmholtz free energy plus the *PV* term. Under ambient conditions, the *PV* term is negligible and Gibbs and Helmholtz free energies can be used interchangeably; we do so in this chapter.

Expressions can also be obtained for the energy and entropy of the system. The energy has the form

$$
E = kT^2\left(\frac{\partial \ln Q}{\partial T}\right)_{N,V}
\tag{17}
$$

which gives

$$
E = \frac{3}{2}kT(M+N) + \int p(\mathbf{q},\mathbf{r}^N)H(\mathbf{q},\mathbf{r}^N)\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{r}^N
\tag{18}
$$

where the first term is the kinetic energy and the second the potential energy. The conditional probability distribution $p(\mathbf{r}^N|\mathbf{q})$ of finding the solvent configuration $\mathbf{r}^N$, given that the macromolecule is in conformation $\mathbf{q}$, is defined by

$$
p(\mathbf{q},\mathbf{r}^N) + p(\mathbf{q})p(\mathbf{r}^N|\mathbf{q})
\tag{19}
$$

With Eqs. (19) and (6) we obtain for the potential energy

$$
\begin{aligned}
&\int p(\mathbf{q})H_{\mathrm{mm}}(\mathbf{q})\mathrm{d}\mathbf{q} + \int p(\mathbf{q})\mathrm{d}\mathbf{q}\int p(\mathbf{r}^N|\mathbf{q}) \\
&\quad \times\left[H_{\mathrm{mw}}(\mathbf{q},\mathbf{r}^N) + H_{\mathrm{ww}}(\mathbf{r}^N)\right]d\mathbf{r}^N
\end{aligned}
\tag{20}
$$

The first term in Eq. (20) is the average intramolecular energy and the second the average solute–solvent and solvent–solvent energy.

The entropy is given by

$$
S = k\ln Q + kT\left(\frac{\partial \ln Q}{\partial T}\right)_{N,V}
\tag{21}
$$

which yields

$$
\begin{aligned}
S &= -k\int p(\mathbf{q},\mathbf{r}^N)\ln p(\mathbf{q},\mathbf{r}^N)\mathrm{d}p\mathrm{d}\mathbf{r}^N \\
&\quad - k\ln(N!\Lambda^{3M}\Lambda^{3N}) + \frac{3}{2}k(M+N) \\
&= -k\int p(\mathbf{q})\ln p(\mathbf{q})\mathrm{d}\mathbf{q} \\
&\quad - k\int p(\mathbf{q})\mathrm{d}\mathbf{q}\int p(\mathbf{r}^N|\mathbf{q})\ln p(\mathbf{r}^N|\mathbf{q})\mathrm{d}\mathbf{r}^N \\
&\quad - k\ln(N!\Lambda^{3M}\Lambda^{3N}) + \frac{3}{2}k(M+N)
\end{aligned}
\tag{22}
$$

where the first term is the configurational entropy of the macromolecule and the second term is the average solvent entropy; i.e. the entropy that arises from solute–solvent and solvent–solvent correlations.

By use of the above analysis, it is possible to provide a more detailed microscopic description of the native state of a protein, given that it is under thermodynamic control. The probability distribution $p(\mathbf{q})$ completely specifies the conformational 'state' of a protein. The native state could be defined as the distribution of configurations of the macromolecule, $p(\mathbf{q})$, which minimizes the free energy functional, Eq. (16) under physiological conditions; a variational minimization of Eq. (16) gives Eq. (14). This definition has the advantage of including protein flexibility and accounting for possible disorder in the native state. For convenience, the native state is usually defined as including only certain values of $\mathbf{q}$ (see below). The average effective energy term tends to localize the macromolecule in the deepest wells of the multidimensional effective energy surface [38], but the configurational entropy term tends to make $p(\mathbf{q})$ as uniform as possible. As a result, the native state, which consists of the conformations of lowest free energy, need not be the ones of lowest effective energy, since some deep wells may be so 'narrow' that the vibrational entropy of a protein in those wells would be very small.

Integration of the solvent degrees of freedom greatly simplifies the treatment of conformational equilibria. However, this remains only a formal exercise unless a model for the solvation free energy as a function of macromolecular conformation is available. The derivation of such models is an active area of research. It can, in principle, be done by using statistical thermodynamics, but it has so far been based more on intuitive arguments because of theoretical difficulties. The statistical thermodynamics of solvation is considered in Section 5 and empirical solvation models are discussed in Section 6. In Section 7 we present a simplified theoretical model that is parameterized based on experimental data.

The thermodynamic stability of a macromolecular native state can be expressed in terms of the standard free energy of folding, $\Delta G$. Given the equilibrium constant $K$ for the folding reaction, we have

$$\Delta G = -RT\ln K, \qquad K = \frac{[\text{native}]}{[\text{denatured}]} \tag{23}$$

This equilibrium constant cannot be measured under physiological conditions because the concentration of the denatured state is vanishingly small. Consequently, $K$ and $\Delta G$ are usually determined either at high temperature or at high denaturant concentration and the results are extrapolated to physiological conditions (room temperature, or zero denaturant concentration). Under usual physiological conditions the Gibbs free energy is essentially equal to the Helmholtz free energy; the $P\Delta V$ term is negligible. We use $\Delta G$ because most measurements are made under constant pressure conditions.

With statistical mechanics we can derive an expression for the free energy of folding in terms of the interactions and the distributions of microscopic states. We divide the configurational space into subsets A consisting of different configurations for the macromolecule in analogy to the treatment of the equilibrium between different isomers. The free energy of conformational set A is

$$A_\mathrm{A} = -kT\ln Z_\mathrm{A} + kT\ln(N!\Lambda^{3M}\Lambda^{3N}) \tag{24}$$

with

$$Z_\mathrm{A} = Z_\mathrm{ww} V 8\pi^2 \int_\mathrm{A} \exp(-\beta W)\,\mathrm{d}\mathbf{q} \tag{25}$$

where the integration is carried out over the configurations in set A. By definition,

$$\sum_\mathrm{A} Z_\mathrm{A} = Z \tag{26}$$

The free energy of the conformational set A can then be written:

$$\begin{aligned} A_\mathrm{A} &= A^\mathrm{o} + kT\ln\left(\frac{\Lambda^{3M}}{V8\pi^2}\right) + \int_\mathrm{A} p_\mathrm{A}(\mathbf{q})W(\mathbf{q})\mathrm{d}\mathbf{q} \\ &\quad + kT\int_\mathrm{A} p_\mathrm{A}(\mathbf{q})\ln p_\mathrm{A}(\mathbf{q})\mathrm{d}\mathbf{q} \\ &= A^\mathrm{o} + A^\mathrm{id} + \langle W\rangle_\mathrm{A} - TS_\mathrm{A}^\mathrm{conf} \end{aligned} \tag{27}$$

where $p_\mathrm{A}$ is a probability distribution normalized within the set A; i.e.

$$p_\mathrm{A}(\mathbf{q}) = \int_\mathrm{A} \frac{\exp(-\beta W(\mathbf{q}))}{\exp(-\beta W(\mathbf{q}))\mathrm{d}\mathbf{q}} \tag{28}$$

The first two terms in Eq. (27) are, as in Eq. (16), the free energy of pure solvent and the ideal translational and rotational free energy of the macromolecule. These are the same for all conformational states. The third term in Eq. (27) is the average effective energy of state A and the last term is the configurational entropy of state A. The free energy difference between the two sets A and B is then

$$\begin{aligned} \Delta A &= A_\mathrm{B} - A_\mathrm{A} = A[p_\mathrm{B}(\mathbf{q})] - A[p_\mathrm{A}(\mathbf{q})] \\ &= \langle W\rangle_\mathrm{B} - \langle W\rangle_\mathrm{A} - T[S_\mathrm{B}^\mathrm{conf} - S_\mathrm{A}^\mathrm{conf}] \\ &= \Delta\langle W\rangle - T\Delta S^\mathrm{conf} \\ &= \Delta\langle H_\mathrm{mm}\rangle + \Delta\langle\Delta G^\mathrm{slv}\rangle - T\Delta S^\mathrm{conf} \end{aligned} \tag{29}$$

where the notation $A[p(\mathbf{q})]$ denotes that the free energy is a functional of the distribution function. One can also use $A[p_\mathrm{B}(\mathbf{q})] - A[p_\mathrm{A}(\mathbf{q})]$ with $p_\mathrm{A}$ and $p_\mathrm{B}$ defined as the conformational distributions under different external conditions, for example, one under physiological conditions and another for high denaturant concentrations. This is more consistent with the definition of the native state given above and does not require an arbitrary separation of the conformational space into 'native' and 'denatured' regions. The Gibbs free energy difference between A and B is obtained by adding the $P\Delta V$ term; that is,

$$\Delta G=\Delta\langle H_{\mathrm{mm}}\rangle+\Delta\langle\Delta G^{\mathrm{slv}}\rangle-T\Delta S^{\mathrm{conf}}+P\Delta V \qquad (30)$$

If A is the denatured state and B the native state, both of which have to be defined in some way and both of which include many configurations, Eq. (29) gives the free energy of folding. This equation expresses the intuitive idea that protein stability is a result of a balance between the effective energy, which favors the native state, and the configurational entropy, which favors the denatured state. The change in average effective energy is related to the depth of the native state well on the effective energy hypersurface, though there may be a barrier between the two states. The entropic cost of localizing the protein in this well was estimated to be of the order of a few hundred kcal/mol [78]. Since the overall free energy change upon folding is usually between 5 and 15 kcal/mol [9], the depth of the native state well on the effective energy surface is also of the order of a few hundred kcal/mol [10].

## 4. Energy functions

An essential element in developing a microscopic description for the interpretation of protein thermodynamics is the potential energy function, which makes it possible to calculate the potential energy of the system as a function of the atomic coordinates. The potential energy can be used directly to determine the relative stabilities of the different possible structures of the system. To obtain the forces acting on the atoms of the system, the first derivatives of the potential with respect to the atom positions are calculated. These forces can be used to determine dynamic and thermodynamic properties of the system; e.g. by solving Newton's equations of motion to describe how the atomic positions change with respect to time and by calculating average properties, such as the enthalpy, from these positions [39,40]. From the second derivatives of the potential surface, the force constants for small displacements can be evaluated and used to find the normal modes. The normal modes provide an alternative approach to the dynamics in the harmonic limit. They are very useful also for introducing quantum corrections to the vibrational contributions to thermodynamic quantities.

To obtain potential energy surfaces for proteins with the required accuracy and speed, it is necessary to introduce a simple model which is calibrated by fitting it to experimental or quantum mechanical information. When working with macromolecules, there is a need to have available a reliable method for calculating interaction energies many times ($10^4$–$10^6$ energy calculations) for systems of hundreds to thousands of atoms. Such a method is supplied by empirical energy functions. However, there is a price to pay for introducing this type of model for the calculation. Empirical energy functions do not have the generality of quantum mechanical calculations. They are at best limited to the systems for which they were designed.

The potential energy, $U(\mathbf{R}^M)$ of the macromolecule as a function of the atomic coordinate, $\mathbf{R}^M$, has the form

$$\begin{aligned}U(\mathbf{R}^M)= &\sum_{\text{bonds}} K_{\mathrm{b}}(b-b_o)^2+\sum_{\text{UB}} K_{\mathrm{UB}}(S-S_o)^2\\ &+\sum_{\text{angle}} K_\theta(\theta-\theta_o)^2+\sum_{\text{dihedrals}} K_\chi(1\\ &+\cos(n\chi-\delta))\\ &+\sum_{\text{impropers}} K_{\mathrm{imp}}(\varphi-\varphi_o)^2\\ &+\sum_{\text{nonbond}} \varepsilon_{ij}\left[\left(\frac{R\mathrm{min}_{ij}}{r_{ij}}\right)^{12}\right.\\ &\left.-2\left(\frac{R\mathrm{min}_{ij}}{r_{ij}}\right)^{6}\right]+\frac{q_iq_j}{\varepsilon\, r_{ij}}\end{aligned} \qquad (31)$$

where $K_{\mathrm{b}}$, $K_{\mathrm{UB}}$, $K_\theta$, $K_\chi$, and $K_{\mathrm{imp}}$ are the bond, Urey–Bradley, angle, dihedral angle and improper dihedral angle force constants, respectively; $b$, $S$, $\theta$, $\chi$ and $\varphi$ are the bond length, Urey–Bradley 1,3 distance, bond angle, dihedral angle and improper torsion angle, respectively (all the internal coordinates are expressed as functions of $\mathbf{R}^M$); the subscript zero represents the values for which the individual terms have their minima. The dihedral term depends on the parameters $n$ and $\delta$, which define the multiplicity or periodicity and phase, respectively. Coulomb and Lennard–Jones 6–12 terms make up the external or non-bonded interactions; $\varepsilon_{ij}$ is the Lennard–Jones well-depth and $R\text{min}_{ij}$ is the distance at the Lennard–Jones minimum, $q_i$ is the partial atomic charge, $\varepsilon$ is the effective dielectric constant and $r_{ij}$ is the distance between atoms $i$ and $j$, respectively. The Lennard–Jones parameters between pairs of different atoms are obtained from the Lorentz–Berthelot combination rules, in which $\varepsilon_{ij}$ and $R\text{min}_{ij}$ are derived based on the geometric mean and the arithmetic mean, respectively, of the parameters for atoms $i$ and $j$.

As is evident from Eq. (31) the potential energy function has a form that makes it simple to calculate. This simplicity is achieved with little sacrifice in accuracy for the properties of primary interest for the macromolecules for which the potential function is designed. The bond and angle energies are treated with harmonic terms, as is the Urey–Bradley term. Thus, bond making and breaking cannot be treated directly using the standard potential energy function. However, it is possible to study such processes (e.g. chemical reactions) by use of quantum mechanical/molecular mechanical (QM/MM) potential energy functions [41]. Use of the harmonic terms is satisfactory for the majority of condensed phase simulations, which are performed close to or below room temperature, such that configurations with large deviations from the minimum energy bond lengths and angles are usually unimportant. Even high temperature simulations (e.g. protein unfolding at 600 K) have been shown to yield meaningful results without modification of the potential function [42]. Rotations about bonds are treated with a sinusoidal function that may involve a Fourier series of the length needed for the accurate treatment of torsional surfaces. In most cases only one or two terms in the series are required to obtain results of sufficient accuracy. For each pair of bonded atoms, terms for all possible definitions of the dihedral angle are used. Improper dihedral terms, traditionally used to maintain chirality in extended atom potential functions, are included to allow for fine tuning of specific properties, such as ring torsional deformations and out-of-plane bending of aromatic hydrogens. Simple Coulombic and Lennard–Jones terms are used for the non-bonded interactions. The latter are important because of the central role of non-bonded interactions in macromolecular structure and dynamics and the computational costs associated with the calculation of these terms.

The parameters were determined by optimizing the reproduction of experimental pure liquid, solution and crystal data, as well as the appropriate ab initio results. Details are given in MacKerell et al. [36].

## 5. Statistical thermodynamics of solvation

In molecular dynamics simulations solvation is usually treated by surrounding the macromolecule with a large number of explicit water molecules. This approach has two main limitations. First, the computational expense is exceedingly high; most CPU time in the simulation is expended calculating the motions of the solvent molecules, which often are of no direct interest. The second limitation is that in explicit solvent simulations the effective energy of a macromolecular conformation is not known, only the intramolecular energy is known. The solute–solvent and solvent–solvent energies can be calculated as well, but they are not directly related to the solvation free energy. The alternative to explicit solvation is to include in the energy function a model for the solvation free energy; i.e. to perform simulations with an *effective* energy function. This approach is referred to as implicit solvation and is approximately two orders of magnitude faster than corresponding simulations with explicit solvent.

The solvation free energy in Eq. (10) is more traditionally known as the excess chemical potential. It is the part of the chemical potential that depends on the interactions between the solute and the solvent (it is zero for ideal gas particles). It can be thought of as the free energy of transferring the solute from a fixed point in the gas phase to a fixed point in the solution [37]. Progress in the theoretical study of the conformational properties of macromolecules depends critically on development of quantitative models for the solvation free energy for these systems.

Statistical thermodynamics has been used to develop several different approaches for calculating the solvation free energy. Analytical theories are applicable only to simple fluids, such as hard spheres and Lennard–Jones particles. For example, the thermodynamic properties of hard sphere fluids are quite well predicted by scaled particle theory [43] and by integral equation theories (particularly, the Percus–Yevick approximation) [44]. For aqueous solutions the XRISM (X-reference interaction site model) integral equation [45–47] has been used to obtain the thermodynamic solvation properties of small solutes. The predictions of XRISM are qualitatively correct but suffer from quantitative deficiencies [47,48]. Solvation free energies for complex systems, or more specifically small changes in complex systems (e.g. the effect of an amino acid mutation of a protein), can be calculated from molecular dynamics and Monte Carlo simulations, combined together with techniques, such as free energy perturbation theory and thermodynamic integration methods [49,50]. Such methods are exact, in principle, for given intermolecular potentials, but in practice, they usually require very long configurational sampling times to converge.

Other than quantitative deficiencies, the major limitation of these methods is that they do not provide a strategy for going from small molecules to macromolecules. A possible exception is the theory for the hydrophobic solvation recently developed by Lum et al. [51]; its range of practical applicability has still to be determined. Performing solvation free energy calculations for every single conformation of a macromolecule that we want to consider is out of the question. Even integral equation theories, which are computationally much more efficient that free energy simulations, require the numerical solution of an integro-differential equation. This is *quite* expensive to do at every step of a molecular dynamics run. What is needed is simple semi-analytical models that describe the solvation free energy of macromolecules as a function of conformation. Ideally, the results of the rigorous solvation free energy calculations should be used as the basis for building such models for macromolecules. Lacking such results, experimental data can serve as a basis for constructing models for the solvation free energy.

One approach that holds promise in this regard is the 'inhomogeneous fluid' solvation theory, so named because it views the solutes as inhomogeneities in the solvent and treats the solution as an inhomogeneous system [52,53]. This approach considers the energetic and entropic contributions separately; i.e.

$$\Delta G^{\mathrm{slv}} = \Delta E^{\mathrm{slv}} - T\Delta S^{\mathrm{slv}} + P\Delta V^{\mathrm{slv}} \tag{32}$$

and gives the solvation energy and entropy as a sum of a solute–solvent and a solvent–solvent term; that is

$$\Delta E^{\mathrm{slv}} = E_{\mathrm{sw}} + \Delta E_{\mathrm{ww}} \tag{33}$$

$$\Delta S^{\mathrm{slv}} = S_{\mathrm{sw}} + \Delta S_{\mathrm{ww}} \tag{34}$$

For solute insertion at a fixed point in the solvent [37], we have [53]

$$E_{\mathrm{sw}} = \rho \int g^{(1)}(\mathbf{r}) u_{\mathrm{sw}}(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{35}$$

$$\Delta E_{\mathrm{ww}} = \frac{1}{2}\rho^2 \int g^{(1)}(\mathbf{r})\left[g^{(1)}(\mathbf{r}') - 1\right] \times g^{(2)}(\mathbf{r},\mathbf{r}') u_{\mathrm{ww}}(\mathbf{r},\mathbf{r}')\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' \tag{36}$$

$$S_{\mathrm{sw}} = -k\rho \int g^{(1)}(\mathbf{r}) \ln g^{(1)}(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{37}$$

$$\Delta S_{\mathrm{ww}} = -\frac{1}{2}k\rho^2 \int g^{(1)}(\mathbf{r})\left[g^{(1)}(\mathbf{r}') - 1\right] \{g^{(2)}(\mathbf{r},\mathbf{r}') \ln g^{(2)}(\mathbf{r},\mathbf{r}') - g^{(2)}(\mathbf{r},\mathbf{r}') + 1\} \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' \tag{38}$$

where $\rho$ is the solvent number density, $\rho g^{(1)}(\mathbf{r})$ is the local density of the solvent located at $\mathbf{r}$ [with $g^{(1)}(\mathbf{r})$, the pair correlation function between the solute and the solvent], $u_{\mathrm{sw}}(\mathbf{r})$ is the interaction potential between the macromolecule and a solvent molecule, $g^{(2)}(\mathbf{r},\mathbf{r}')$ and $u_{\mathrm{ww}}(\mathbf{r},\mathbf{r}')$ are, respectively, the pair correlation function and the interaction potential between two solvent molecules, one at $\mathbf{r}$ and the other at $\mathbf{r}'$. Eqs. (35)–(38) involve certain approximations, in particular, the assumption that correlations involving more than two particles can be neglected; the details are given in Lazaridis [53]. In Eq. (32), $\Delta V^{\mathrm{slv}}$ is the excess partial molar volume of the solute; that is,

$$\Delta V^{\text{slv}} = \int (1 - g^{(1)})\,\mathrm{d}\mathbf{r} \tag{39}$$

The $P\Delta V^{\text{slv}}$ term can be neglected under ambient conditions. Eqs. (35)–(38) are written here for monatomic solvent particles but can be generalized to polyatomic molecules by adding integrations over orientational degrees of freedom [54].

The expression for the solute–solvent energy, $E_{\text{sw}}$, is a straightforward extension of the energy equation of statistical thermodynamics. The solute–solvent entropy, $S_{\text{sw}}$, arises from correlations between the solute and the solvent. These include positional correlations, described by the density oscillations around the solute, and orientational correlations (i.e. the fact that solvent preferentially adopts certain orientations with respect to the solute). The solvent reorganization energy ($\Delta E_{\text{ww}}$) and entropy ($\Delta S_{\text{ww}}$) account for changes in solvent–solvent interactions and correlations upon solute insertion. All the terms in Eqs. (33) and (34) are largest close to the inhomogeneity (i.e. the solute) and decay to zero far from the solute.

The components of the solvation free energy can be written as integrals over the space around the solute; that is

$$\Delta G^{\text{slv}} = \int f(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{40}$$

where $f(\mathbf{r})$ is the solvation free energy density. Neglecting the $P\Delta V^{\text{slv}}$ term, we obtain from Eqs. (33) and (34) that

$$\begin{aligned} f(\mathbf{r}) = {}& \rho g^{(1)}(\mathbf{r}) u_{\text{sw}}(\mathbf{r}) + \frac{1}{2}\rho^2 g^{(1)}(\mathbf{r}) \\ & \times \int [g^{(1)}(\mathbf{r}') - 1] g^{(2)}(\mathbf{r},\mathbf{r}') u_{\text{ww}}(\mathbf{r},\mathbf{r}')\, d\mathbf{r}' \\ & + kT\ \rho\ g^{(1)}(\mathbf{r}) \ln g^{(1)}(\mathbf{r}) \\ & + \frac{1}{2} k \rho^2 g^{(1)}(\mathbf{r}) \int [g^{(1)}(\mathbf{r}') - 1] \\ & \{g^{(2)}(\mathbf{r},\mathbf{r}') \ln g^{(2)}(\mathbf{r},\mathbf{r}') - g^{(2)}(\mathbf{r},\mathbf{r}') + 1\}\,\mathrm{d}\mathbf{r}' \end{aligned} \tag{41}$$

This approach has a number of advantages over the traditional solvation theories mentioned above. First, it provides an explicit connection between solvation thermodynamics and solvent structure around the solute and gives a detailed decomposition of the solvation free energy. It has been used to analyze the thermodynamics of hydrophobic hydration [54–56] as well as the solvation in simple fluids [57,58]. This clearly is useful for the physical understanding of the solvation process. Because it gives the solvation free energy components as integrals over space, it leads to a 'modular' approach to the solvation free energy of large molecules. To calculate the difference in solvation free energy between two molecules that differ in one group, one needs only to calculate the integrals in the region around that group. Moreover, this concept can be used to transfer small molecule information to the study of macromolecular solvation. For example, consider an isolated methyl group and a methyl group in a macromolecule. To the extent that the solvent structure next to the methyl group is the same as that next to an isolated methyl group, the solvation free energy coming from that region of space will be the same. In actuality this is not exact and so corrections may have to be introduced [51,54–58]. Nevertheless, this concept is used in Section 7 to develop a simple analytical model for the solvation free energy and other solvation properties of proteins.

## 6. Empirical solvation models

Because the development of solvation models for macromolecules based purely on statistical mechanics has proven difficult, a number of empirical approaches have been proposed. The simplest is the Atomic Solvation Parameter (ASP) model [59]. In this model, the solvation free energy is given as a sum of atomic contributions. The solvation free energy of a group is assumed to be proportional to its accessible surface area, $A_i$,

$$\Delta G^{\text{slv}} = \sum_i \sigma_i A_i \tag{42}$$

and the proportionality coefficients $\sigma_i$ depend on the type of atom and are determined by fitting experimental data. We use the Gibbs free energy symbol as is customary, but $\Delta G^{\text{slv}}$ is equal, for all practical purposes, to the corresponding Helmholtz free energy, as pointed out earlier. There are two types of models that differ in the type of data that are used to determine the parameters. Models which view the protein interior as a non-polar solvation medium use data for the transfer of molecules from non-polar liquids to water [59]. The second type uses data from the transfer of molecules from the gas phase to water [60,61]. The latter solvation models can be added to molecular mechanics force fields. Fraternali and van Gunsteren [62] used a very simple surface-area based model for molecular dynamics simulations and Caflisch et al. [63] combined a simplified version of the model of Lazaridis and Karplus [35] and Section 7 with the Fraternali and van Gunsteren [62] for studies of peptide and protein folding.

Other empirical solvation models do not use the accessible surface area. For example, the hydration shell model [64] assumes that the hydration free energy of a group arises from the first hydration shell and that it is proportional to the volume of the hydration shell that is accessible to the solvent (i.e. that is not occupied by other solute atoms). Another type of solvation model is based on the contacts each group makes with other solute atoms [65]. The more contacts there are, the smaller is the magnitude of the solvation free energy of the group and the contacts are weighted according to some function that depends on their distance from the group. This model is much faster to use because counting the number of contacts takes considerably less time than calculating the surface area. A version of this model was parametrized based on solvation free energies of small molecules [66]. The model assumes a linear relationship between the solvation free energy and a weighted sum of the contacts that the group makes with other solute atoms. A solvation parameter is assigned to each group.

These models can be used not only for the solvation free energy but also for the enthalpy and the heat capacity of solvation. Makhatadze and Privalov [67] developed such a model assuming that the solvation enthalpy of both polar and non-polar groups is proportional to the surface area. Evidence for the breakdown of this assumption was obtained by theoretical methods [10]. The first test was based on the RISM integral equation theory applied to *N*-methyl acetamide, the alanine dipeptide, and the alanine tetrapeptide. The solvation enthalpy was calculated for a large number of conformations. It was found that for non-extended conformations of the alanine tetrapeptide the CONH group had solvation enthalpies lower than the surface area proportionality assumption would predict. The second test involved the calculation of CO group–solvent interactions from a MD simulation of a protein in solution. While this is only one component of the solvation enthalpy (the other is the solvent reorganization energy [47]), it gives an indication of the validity of this assumption. The plot of the CO interaction energies versus ASA exhibited considerable scatter. As with the integral equation results, it was found that ASA underestimated the interactions of buried groups with the solvent. Even groups with zero ASA can interact significantly with the solvent. This may result in overestimation of the solvation enthalpy change upon denaturation.

Another set of solvation models treats the entire protein at once and is based on continuum electrostatics and the linearized Poisson–Boltzmann equation [68]. This method evaluates only the electrostatic component of solvation. The solute is treated as a low dielectric cavity in a high dielectric medium. This approach assumes that the electrostatic component of solvation free energy is described adequately by continuum electrostatics and that the laws of continuum electrostatics hold down to the atomic scale. To obtain a complete solvation model, the non-polar component of solvation is added, usually a term equal to the accessible surface times a surface tension-like coefficient. The electrostatic treatment involves the numerical solution of the linearized Poisson–Boltzmann equation which can be applied to realistic solute geometries (e.g. X-ray or NMR structure of a protein) [68,69].

Although the PB approach is computationally more efficient than free energy simulations, it still is too computationally demanding to use in a molecular dynamics simulations. Semi-analytical or analytical approximations that are much faster to evaluate have been proposed. Still et al. [70] introduced a simple generalization of the Born formula to polyatomic molecules. More recently, the generalized Born equation was combined with

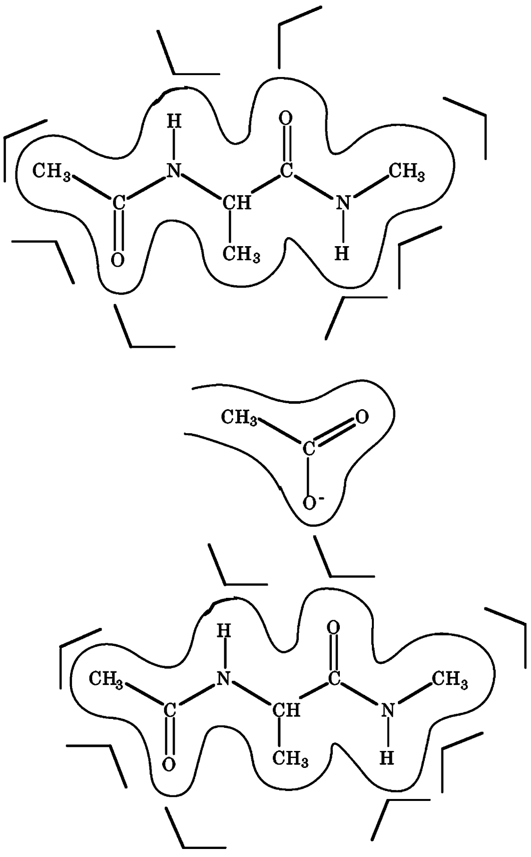

Fig. 2. Perturbation of solvent structure by neighboring groups.

an integrated field method for self-energies to give a completely analytical treatment of electrostatic energies and forces [71]; successful applications have been made to conformational equilibria of peptides [72].

## 7. An effective energy function

In Section 5 it was shown that the solvation free energy can be written as an integral over the space around the solute [see Eq. (40)]. This can be used as a basis for developing analytical solvation models. By the use of theory or simulations one can estimate how much of the solvation free energy comes from the first solvation shell, the second solvation shell, etc. Such calculations for specific components of the solvation free energy have been performed for a few simple systems [55–58,64,73]. These studies have shown that a large portion of the solvation entropy or solvation energy for non-polar and polar solutes (70–90%) arises from the first solvation shell. On the basis of these results, it is reasonable to assume that $f(\mathbf{r})$ for uncharged groups is short-ranged; i.e. it decays to zero within the second solvation shell. A simple Gaussian function, with the appropriate correlation length was found to exhibit such a behavior and has been used in the Gaussian exclusion model for solvation thermodynamics [35].

When a solute $j$ approaches solute $i$ the solvation free energy of $i$ is modified because $j$ excludes solvent from the volume it occupies. If this was the only effect of $j$, the solvation free energy of $i$ when $j$ is close by would be

$$\Delta G_i^{\mathrm{slv}} = \Delta G_i^{\mathrm{ref}} - \int_{V_j} f_i(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{43}$$

where $\Delta G_i^{\mathrm{ref}}$ is the solvation free energy of isolated $i$ and the integration is over $V_j$, the volume excluded by $j$; the variable $\mathbf{r}$ here stands for position of $j$ relative to $i$. This is the excluded volume effect, in which the presence of solute excludes solvent from a certain region. The second effect is that the structure of the solvent in the region not occupied by $j$ is modified by $j$ and this may affect the interactions of $i$ with the solvent, especially if $i$ and $j$ are polar (Fig. 2). This is the 'solvent perturbation' effect.

Based on this physical principle, we have developed a model for the solvation free energy and

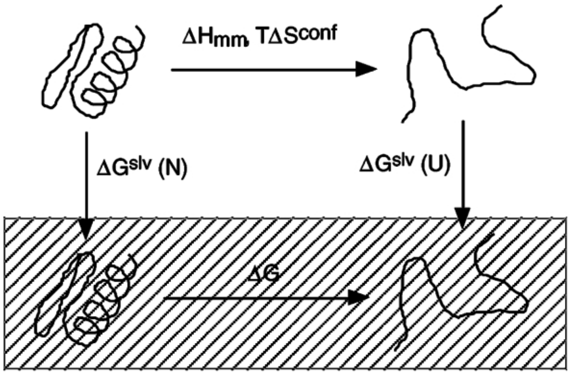

Fig. 3. Thermodynamic cycle for protein folding.

Table 1
Calculated vacuum enthalpy of unfolding[a]

| Protein | $N_{\text{res}}$ | MW | $\Delta H_{\text{N}}^{\text{U}}$(vac) | $\Delta H_{\text{N}}^{\text{U}}$(vdW) | $\Delta H_{\text{N}}^{\text{U}}$(elec) | $\Delta H_{\text{N}}^{\text{U}}$(bond) |
|---|---|---|---|---|---|---|
| Cytochrome *c* | 103 | 12 300 | 943 | 688 | 225 | 30 |
| RNAse A | 124 | 13 700 | 1068 | 654 | 398 | 16 |
| Lysozyme | 129 | 14 300 | 1116 | 738 | 351 | 27 |
| Myoglobin | 153 | 17 800 | 1492 | 1020 | 421 | 51 |

[a] All values in kcal/mol calculated as described in text (see also Lazaridis et al. [10], Table V). For myoglobin and cytochrome *c* it is assumed that the free heme will have the same self (intra-heme) energy as in the protein.

combined it with the CHARMM polar hydrogen energy function to obtain an approximation to the effective energy function (EEF1) [35]. The model assumes that for a polyatomic solute we can write the solvation free energy as a sum over group contributions (see Section 10); that is

$$\Delta G^{\text{slv}} = \sum_i \Delta G_i^{\text{slv}} \tag{44}$$

where $\Delta G_i^{\text{slv}}$ is the solvation free energy of group *i*. Taking into account only the solvent exclusion effect, we can write

$$\Delta G_i^{\text{slv}} = \Delta G_i^{\text{ref}} - \sum_j \int_{V_j} f_i(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{45}$$

where $\Delta G_i^{\text{ref}}$ is the solvation free energy of group *i* in a suitably chosen small molecule in which group *i* is essentially fully solvent exposed. The integral in Eq. (44) is over the volume $V_j$ of group *j* and the summation is over all groups *j* around *i*. Although the formalism provides the basis for a more exact treatment, in the simplest approximation the integral over $f_i(\mathbf{r})$ is assumed to be the product $f_i(\mathbf{r}_{ij})V_j$, so that

$$\Delta G_i^{\text{slv}} = \Delta G_i^{\text{ref}} - \sum_{j \neq i} f_i(\mathbf{r}_{ij}) V_j \tag{46}$$

where $r_{ij}$ is the distance between *i* and *j*. The solvation free energy density is assumed to be given by the Gaussian function

$$f_i(\mathbf{r}) 4\pi r^2 = \alpha_i \exp(-x_i^2), \qquad x_i = \frac{r - R_i}{\lambda_i} \tag{47}$$

where $R_i$ is the van der Waals radius of *i* (1/2 of the distance to the energy minimum in the Lennard–Jones potential), $\lambda_i$ is a correlation length, and $\alpha_i$ is a proportionality coefficient given by

$$\alpha_i = 2\Delta G_i^{\text{free}} / \sqrt{\pi}\lambda_i \tag{48}$$

where $\Delta G_i^{\text{free}}$ is the solvation free energy of the free (isolated) group *i*; $\Delta G_i^{\text{free}}$ is close to $\Delta G_i^{\text{ref}}$ but not identical to it and is determined empirically by requiring that the solvation free energy of deeply buried groups be zero.

Eq. (45) accounts for the solvent–exclusion effect. The solvent perturbation effect, which is most important for polar and charged groups, is approximately taken into account in the model by using neutralized forms of the ionic sidechains and a linear distance-dependent dielectric constant ($\varepsilon = r$). Both van der Waals and electrostatic interactions are cut off at 9 Å with a switching function between 7 and 9 Å. Electrostatic interactions are calculated on a group by group basis. The value of $\lambda_i$ was taken to be the thickness of one hydration shell (3.5 Å) except for the neutralized ionic groups, for which a larger value (6 Å) was used. The same model can be employed to calculate other solvation properties, such as the solvation enthalpy, entropy, and heat capacity, by introducing the appropriate reference solvation values for the enthalpy and heat capacity; the solvation entropy is obtained by difference from the solvation free energy and enthalpy.

EEF1 has been tested extensively. It gives stable structures for native proteins during molecular dynamics simulations at room temperature with modest deviations from the crystal structure [35], it discriminates native from misfolded conformations [74] and gives unfolding pathways at high temperatures in agreement with explicit water simulations [75]. Nevertheless, the model has a number of deficiencies: it does not account for the directionality of polar group–solvent interactions, assumes that all empty space is occupied by solvent, and neglects solvent orientational polarization effects. These deficiencies can be rectified by refinements (work in progress), inevitably at the expense of simplicity and computational efficiency.

Unlike many of the knowledge-based energy functions, which appear to be limited to the evaluation of protein conformations obtained by threading and related procedures, EEF1 is applicable to protein folding and unfolding studies by molecular dynamics simulations, for example. The formulation proposed is only 50% slower than a vacuum simulation and thus makes possible many studies for which simulations in explicit water are prohibitively expensive. Given its physical basis and decomposability, the effective energy function can be used for approximate thermodynamic analysis of contributions to protein stability.

## 8. Contribution of intramolecular interactions to protein stability

The intramolecular energy term in Eq. (29) includes contributions from bonded terms (bond stretching and bending, torsional potentials, deviations from planarity of aromatic rings, etc.), and non-bonded terms (dispersion and electrostatic interactions, including hydrogen bonding). The bonded terms are not expected to be substantially different in the folded and unfolded conformations. The largest contribution to $\Delta\langle H_{\mathrm{mm}}\rangle$ is expected to come from the non-bonded interactions. Nonbonded interactions, of course, exist between the solute and the solvent, but in the present analysis these are included within the solvation free energy term.

A schematic construct that can be used to visualize the partitioning of the free energy in Eq. (29) is the thermodynamic cycle of Fig. 3[10]. The unfolding reaction is imagined to occur in the gas phase in the same way as it occurs in solution; i.e. the same conformational ensembles correspond to the native and the unfolded state. The terms in Eq. (29) relevant to the gas phase reaction are $\Delta\langle H_{\mathrm{mm}}\rangle$ and $T\Delta S^{\mathrm{conf}}$. The vertical processes correspond to inserting the folded and unfolded ensembles into the solvent. They give rise to the $\Delta G^{\mathrm{slv}}$ terms for the native (N) and unfolded (U) configuration states. The total $\Delta G$ in Eq. (29) is obtained by completing the thermodynamic cycle in Fig. 3. The same cycle can be used to analyze the enthalpy of unfolding. In that case, $\Delta\langle H_{\mathrm{mm}}\rangle$ is the relevant quantity for the gas phase and $\Delta\Delta H^{\mathrm{slv}}$ for the vertical reactions. The enthalpy of unfolding in solution is

$$\Delta H = \Delta\langle H_{\mathrm{mm}}\rangle + \Delta\langle \Delta H^{\mathrm{slv}}\rangle \tag{49}$$

Several functions (Section 4) that describe the intramolecular energy of the protein are available [76,77] and can be used, in principle, to evaluate the intramolecular contribution to the enthalpy and free energy of unfolding. Since the native state energy can be evaluated by molecular dynamics simulations starting with the known structure, it is necessary only to construct a model for the conformational ensemble corresponding to the denatured state. In previous work, it has been customary to approximate the properties of the unfolded state by summing the properties of individual amino acid residues. A better model, though still approximate, is to use a fully extended polypeptide chain to represent for the unfolded state.

Using the latter model for the unfolded state, the $\Delta\langle H_{\mathrm{mm}}\rangle$ term was calculated with the CHARMM polar hydrogen energy function and neutralized sidechains [10]. Table 1 shows the results of this calculation as well as the decomposition of this term into van der Waals, electrostatic, and bonded contributions. The calculated values for $\Delta\langle H_{\mathrm{mm}}\rangle$ ranged from 943 kcal/mol for cytochrome *c* to 1492 kcal/mol for myoglobin. It was indeed found that the bonded terms make a very small contribution to $\Delta\langle H_{\mathrm{mm}}\rangle$. $\Delta\langle H_{\mathrm{mm}}\rangle$ was

Table 2
Van der Waals contributions to the unfolding enthalpy, $\Delta H_{\mathrm{N}}^{\mathrm{U}}(\mathrm{vac})$[a]

| | $\Delta H_{\mathrm{N}}^{\mathrm{U}}(\mathrm{vdW})$ | | | |
|---|---|---|---|---|
| | np–np | np–p | p–p | Total |
| Cyt *c* | 211 | 372 | 113 | 688 |
| RNase A | 148 | 366 | 140 | 654 |
| Lyso | 205 | 396 | 137 | 738 |
| Mb | 306 | 573 | 144 | 1020 |

[a] All values in kcal/mol; see Table VII of Lazaridis et al. [10].

Table 3
Estimate of the enthalpy of denaturation (25 °C)

| Protein | $\Delta H_N^U$(vac) | $\Delta H_N^U$(sol,np) | $\Delta H_N^U$(sol,p) | $\Delta H$(calc) | $\Delta H$(exp) |
|---|---|---|---|---|---|
| Cytochrome *c* | 943 | −172 | −1067 (−750) | −296 | 21 |
| RNAse A | 1068 | −159 | −1127 (−838) | −218 | 71 |
| Lysozyme | 1116 | −192 | −1224 (−866) | −300 | 58 |
| Myoglobin | 1492 | −276 | −1541 (−1214) | −325 | 1.4 |

All values in kcal/mol; see also Table 1.

shown to consist 60–70% of van der Waals interactions and 25–35% of electrostatic interactions; hydrogen bonding arises mainly from the electrostatic term in the energy function [Eq. (31)].

The van der Waals term was further decomposed into contributions from non-polar–non-polar, polar–polar and polar–non-polar interactions (Table 2). Non-polar atoms are all carbons except the backbone carbonyl carbon and the polar carbon atoms of the Asp, Glu, Asn and Gln sidechains. All other atoms are considered to be polar. What is perhaps surprising in Table 2 is the large contribution from polar–non-polar interactions. This shows that the interactions in the protein interior are rather complex and cannot be represented as a simple sum of hydrophobic interactions and hydrogen bonds.

The estimates for the change in intramolecular enthalpy upon folding can be combined with solvation enthalpies calculated from the accessible surface area [66] to obtain estimates of the enthalpy of denaturation. Unlike intramolecular enthalpies, solvation enthalpies are highly temperature dependent. (There is a large solvation heat capacity.) The results of the calculation for four proteins are shown in Table 3 and refer to 25 °C [10]. The values for $\Delta H$ are large and negative, whereas the experimental values are small and positive. Part of the discrepancy arises from the assumption of the proportionality of solvation enthalpy to the accessible surface area, particularly for the polar terms (see Section 6). If one assumes the major error arises from the polar term, the reduced values in parentheses are required to obtain agreement with experiment; see also Table E-IV of Lazaridis et al. [10]. More important is the fact that the fully extended model used for the denatured state is unrealistic. These issues are further explored in the following section.

The Poisson–Boltzmann approach has also been used to estimate the electrostatic solvation free energy for the native and unfolded forms of these four proteins and for a 20 residue polyalaline helix [10]. The results are shown in Table 4. The first two columns in Table 4 are the Poisson–Boltzmann electrostatic solvation free energy of the native and unfolded protein. The third column is the difference of the two, i.e. the change in solvation free energy upon unfolding. The electrostatic solvation energies are more negative for the unfolded conformations primarily because backbone hydrogen bonding groups that are buried in the native state become exposed to the solvent. The fourth column is the change in intramolecular

Table 4
Electrostatic solvation (free) energy differences of folded and unfolded proteins in the Poisson–Boltzmann approximation[a]

| | $\Delta G_{\mathrm{vac}}^{\mathrm{sol}}$(N) | $\Delta G_{\mathrm{vac}}^{\mathrm{sol}}$(U) | $\Delta G_N^U$(sol,p) | $\Delta H_N^U$(vac,elec) | $\Delta G_N^U$(sol,elec)[b] |
|---|---|---|---|---|---|
| Cyt c | −441 | −789 | −348 | +225 | −123 |
| RNAse A | −539 | −958 | −419 | +398 | −21 |
| Lyso | −546 | −1020 | −474 | +351 | −123 |
| Myo | −558 | −1087 | −529 | +421 | −108 |
| ALA20 | −45 | −89 | −44 | +60 | +16 |

[a] All values in kcal/mol.

[b] $\Delta G_N^U(\mathrm{sol,elec}) = \Delta G_N^U(\mathrm{sol,p}) + \Delta H_N^U(\mathrm{vac,elec})$.

Table 5
Proposed contributions of polar and non-polar groups to $\Delta H_{\mathrm{N}}^{\mathrm{U}}$(sol) at 25 °C[a]

| | Polar | Non-polar |
|---|---|---|
| Cytocrome *c* | −204 | 225 |
| RNAse A | −101 | 171 |
| Lysozyme | −152 | 210 |
| Myoglobin | −315 | 316 |

[a] All values in kcal/mol; see Table E-I of Lazaridis et al. [10].

electrostatic enthalpy upon unfolding. The last column is the sum of the third and fourth columns and gives the estimated contribution of electrostatic interactions to unfolding in solution. The latter is found to be negative for the four proteins (i.e. it stabilizes the unfolded state) and positive for the polyalanine helix.

If we use the results for the electrostatic solvation free energy as an estimate of the change in solvation enthalpy to close the cycle in Fig. 3, the resulting unfolding enthalpies would be too positive. This is probably due to the fact that these values correspond to solvation free energies, not enthalpies, and include a positive entropic contribution (the solvation entropy is negative). Also, the non-electrostatic component of the solvation enthalpy is not included in these values. Finally, the problem with the fully extended model of the unfolded state is still present.

The intramolecular energy terms [see Eq. (31)] and their decomposition into polar and non-polar contributions together with the empirical estimates of solvation enthalpy and entropy contributions can be used to address the question of the relative contribution of polar and non-polar groups to protein stability. The contribution of the non-polar groups to the enthalpy is defined as (a) the non-polar–non-polar van der Waals term; (b) one half of the non-polar–polar van der Waals term; and (c) the non-polar solvation enthalpy change upon unfolding. The contribution of the polar groups is equal to (a) the polar–polar van der Waals term; (b) the electrostatic energy term; (c) one half of the polar–non-polar van der Waals term; and (d) the polar solvation enthalpy change upon unfolding. The results are shown in Table 5. The polar groups are seen to make a negative contribution and the non-polar groups a positive contribution to the enthalpy of unfolding. If we add the solvation entropy estimates (see Table 6), the non-polar groups are seen to make the dominant contribution to the free energy of unfolding while the polar groups make a smaller or zero contribution. This is consistent with the traditional idea [78] that the hydrophobic interaction provides the major driving force for folding. The role of the polar groups is to make the protein soluble in water and to introduce specificity in the low free energy conformations that make up the native protein; i.e. the many compact structures that exist are not competitive in free energy with the native structure due to poor polar interactions.

## 9. The denatured state of proteins

In past work the model adopted for the denatured state was either the sum of individual amino acids [66,79] or an extended, completely solvent-exposed polypeptide chain [9,10]. This assumption is at odds with experimental evidence showing that the denatured state in the absence of denaturants is rather compact [80,81]. Residual structure has been found in the heat denatured states of many proteins [82,83]. Theoretical work on simplified protein models [84,85] and molecular dynamics simulations in explicit solvent [86,87] also suggest a relatively compact denatured state. Residual structure and compactness implies that a significant fraction of the protein residues are interacting with each other and are, at least in part,

Table 6
Contribution of polar and non-polar groups to the entropy, $-T\Delta S_{\mathrm{N}}^{\mathrm{U}}$(sol), and free energy, $\Delta G_{\mathrm{N}}^{\mathrm{U}}$(sol), of protein unfolding[a]

| | $-T\Delta S_{\mathrm{N}}^{\mathrm{U}}$(sol)[b] | | $\Delta G_{\mathrm{N}}^{\mathrm{U}}$(sol) | |
|---|---|---|---|---|
| | Polar | Non-polar | Polar | Non-polar |
| Cyt *c* | 202 | 208 | −2 | 433 |
| RNAse A | 225 | 200 | +124 | 371 |
| Lyso | 239 | 240 | +87 | 450 |
| Myo | 289 | 340 | −26 | 656 |

[a] All values in kcal/mol at 25 °C; see Table E-VI of Lazaridis et al. [10].

[b] Values from Makhatadze and Privalov [9].

sequestered from solvent, although the fluctuations in the structures are such that little hydrogen exchange protection is present.

The model of a completely unfolded denatured state has arisen in part from calorimetric measurements of the heat capacity change of protein denaturation. It is generally believed that the positive $\Delta C_p$ values for unfolding arise primarily from the exposure of non-polar groups to water [88–92]. The basis for this conclusion is that the transfer of non-polar groups from the gas phase or non-polar liquids into water is also accompanied by a large positive heat capacity change [93]. It is of interest in this regard to mention the very early paper of John Edsall on apparent molar heat capacities of amino acid and other organic molecules [94]. More recently, it was realized that polar groups also make a contribution to $\Delta C_p$ that is smaller and of the opposite sign to that for non-polar groups [87,92,95]. It has been found that $\Delta C_p$ can be well reproduced assuming a fully unfolded denatured state (all residues fully exposed to solvent [96]) if $\Delta C_p$ is taken to be proportional to the change in the exposed surface area on unfolding. In such calculations the proportionality constants appropriate for the amino acids are obtained from small model compound transfer experiments [92,95,97,98]. Thus, there is an apparent discrepancy between the realistic description of the denatured state as being rather compact and the semi-empirical $\Delta C_p$ models which yield good agreement with experiment if a fully extended denatured state is used.

To resolve this discrepancy we have used EEF1 to generate models for the denatured state by molecular dynamics simulations [99]. The system chosen for this analysis was the 64-residue truncated version of the small protein CI2, for which experimental data and unfolding simulations are available. The heat capacity was calculated by performing simulations of the native (N) and denatured (D) conformations at 280 and 320 K, calculating the enthalpy at these temperatures using the solvation enthalpy parameters in EEF1, and taking the finite difference:

$$\Delta C_p = \left(\frac{\partial \Delta H}{\partial T}\right)_P = \left(\frac{\Delta H(320\text{ K}) - \Delta H(280\text{ K})}{40\text{ K}}\right) \tag{50}$$

Although the absolute partial molar heat capacity of a protein cannot be calculated quantitatively by classical mechanics [100], quantum effects cancel out approximately when the difference in heat capacity between two protein states is taken. Since the enthalpy can be decomposed into intramolecular and solvation terms, the heat capacity can likewise be decomposed. The calculation of the heat capacity takes into account both the intrinsic temperature dependence of the solvation enthalpy of the protein groups and the contribution from the temperature dependence of the protein conformational distribution $p(\mathbf{q})$ (see Section 3). The latter contribution has been neglected in empirical models that use the fully extended chain as a model for the denatured state. An additional small contribution, that from redistribution of the protein population among different local minima in the denatured state, is neglected because it would require extensive sampling of the conformational space.

Three molecular dynamics simulations were run for 1.2 ns at 300 K, starting from a fully extended chain, to generate models for the denatured state. The resulting systems all had relatively compact structures. The effective energies of the three denatured states were between 38 and 65 kcal/mol higher than that of the native state. This is a reasonable difference, considering the experimental protein stability and the estimated change in conformational entropy [35]. The radius of gyration ($R_g$) of the denatured conformations was only 12–18% greater than that of the N state, in agreement with experimental data. The 'hydrophobic collapse' from the extended state found in the molecular dynamics simulations was accompanied by formation of a large number of protein–protein hydrogen bonds; the number of hydrogen bonds in the denatured forms was only slightly smaller than that in the native state but very few were native-like. Moreover, the three denatured conformations had almost no native contacts (pairs of atoms more than three residues apart in sequence that are within 4 Å from each other in the crystal structure [101]).

It was found that the average value for $\Delta C_p$ of unfolding is $0.53 \pm 0.15$ kcal/mol K, compared to the experimental value of 0.79 kcal/mol K [102]; the experimental value pertains to the untruncated form of CI2, but the contribution of the disordered region to $\Delta C_p$ is expected to be small. The underestimation of $\Delta C_p$ may be due to deficiencies of the hydration model (for example the group additivity assumption), to the obviously limited sampling of the denatured state or to the neglect of transitions on a longer time scale (see above). Covalent contributions are approximately equal in the denatured and native conformations and thus do not make a significant contribution to the unfolding heat capacity, as expected [98,103]. In one of the three denatured models, the solvation contribution was found to be negative despite the fact that exposure of non-polar groups is higher in the denatured state. This was due to the fact that the rise in temperature increased the exposure of polar groups in the denatured state and thus decreased the solvation enthalpy difference between the denatured and native conformation. For one of the denatured conformations the intramolecular contribution was very large (0.95 kcal/mol K); for the other two it was smaller (0.05 and 0.13 kcal/mol K), but still significant.

The major conclusion from this simulation study is that a significant contribution to the heat capacity of denaturation comes from protein–protein non-bonded interactions and their changing contribution as a function of temperature. The denatured state D is relatively compact but more labile than the native state so that temperature can break interactions in D more easily than in N. This means that the enthalpy of D increases with temperature more than that of N and contributes significantly to $\Delta C_p$. Moreover, the results demonstrate that relatively compact denatured states are qualitatively consistent with the calorimetric data for $\Delta C_p$. The smaller contribution of hydrophobic group exposure to the heat capacity in this model for the denatured state is compensated by the contribution of non-covalent protein interactions.

## 10. The contributions to protein stability

Analyses of protein stability attempt to provide an understanding of the contributions of different types of interactions. This is exactly what was done in the previous section, where the unfolding enthalpy was decomposed into intraprotein bonded, van der Waals and electrostatic terms and solvation terms. If one wishes to decompose the intraprotein van der Waals terms further into polar and non-polar contributions, as is often done, the large polar–non-polar van der Waals interactions have to be divided between the two; see Table XXII of Lazaridis et al. [10] and the associated discussion. This is a simple example of the fact that some assumptions are required in decomposing the enthalpy, and also the free energy into contributions that are useful for understanding and for predictions. The complexity of the problem is such that there has been considerable discussion recently concerning the validity of such decompositions, as described below. It has become clear that the decompositions are meaningful, though it is necessary to interpret them carefully and understand the assumptions involved. Even the authors who most strongly criticized the decompositions of the free energy have begun to claim them as their own in studying, for example, ligand protein interactions by perturbation models [104].

Two types of decompositions are in common use. The first is a decomposition of the total free energy of folding into types of interactions; the second is a decomposition of the free energy into contributions from constituent groups. The latter is particularly useful for the interpretation of site-directed mutagenesis experiments, where one residue is replaced by another and the effect of the substitution on protein stability or the value of a ligand binding constant is determined. Such decompositions have often been performed on a somewhat intuitive basis, although careful discussions of the assumptions involved have been given in some of the experimental analyses [105]. As mentioned above, it has been argued that such decompositions are not possible because entropy is a global property and cannot be dissected [106,107]. This is incorrect as we show by providing a theoretical basis for such decompositions and

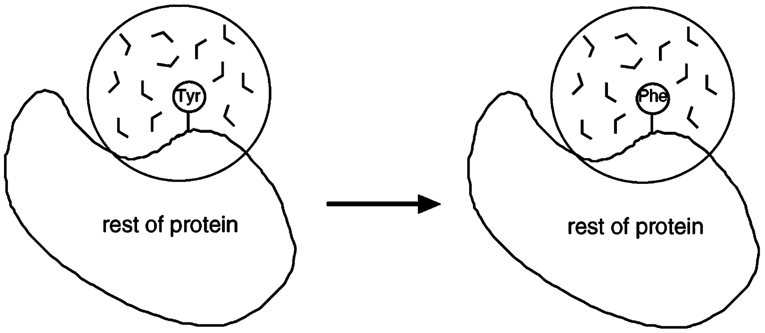

Fig. 4. Free energy simulation transforming the wild type protein with Tyr in the binding site to a mutant with Phe in the binding site. The dashed line indicates the spherical region that is studied in the free energy simulations and the (<) indicates water molecules solvating the binding site.

$$\begin{tikzcd}
\text{Tyr 169 + Tyr} \arrow[r, "\Delta G_f"] \arrow[d, "\Delta G_w"'] & \text{Phe 169 + Tyr} \arrow[d, "\Delta G_m"] \\
\text{Tyr 169 + Tyr} \arrow[r, "\Delta G_b"'] & \text{Phe 169 + Tyr}
\end{tikzcd}$$

Fig. 5. Thermodynamic cycle used in the Tyr 196 to Phe mutation.

by presenting a clear definition of group contributions to stability.

Thermodynamic integration offers an exact way of decomposing a free energy differences into contributions from interactions or groups [108]. The basic formula for thermodynamic integration is

$$\Delta A=\int_0^1 \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \qquad (51)$$

where $U(\lambda)$ is an empirical energy function [e.g. Eq. (31)] that describes the initial state for $\lambda=0$ (e.g. the wild type protein) and the final state for $\lambda=1$ (e.g. the mutant protein); i.e.

$$U(\lambda)=(1-\lambda)U_{\text{wild type}}+\lambda\, U_{\text{mutant}} \qquad (52)$$

The subscript $\lambda$ on the brackets in Eq. (51) indicates that the simulation is done with $U(\lambda)$. Calculations with Eq. (51) have been referred to as computer alchemy [108,109] because they involve transforming one molecule into another on the computer. It would be easy to satisfy the alchemist's dream of transforming lead into gold by this technique.

Since $U(\lambda)$ can be decomposed into contributions from various types of interactions and Eq. (51) is linear in $U(\lambda)$, the free energy itself can be decomposed in a corresponding fashion. To obtain the effect of a point mutation on protein stability, for example, the protein is simulated in both the native and denatured states starting with the wild type sequence and reversibly changing the energy function to that of the mutant (see Fig. 4). It has been realized that this type of decomposition is dependent on the integration path taken in the transformation of one group into another [110,111]. Despite the path dependence, these decompositions are useful [112,113]. Formal analyses of these decompositions have been presented [114,116].

One example of such a decomposition that offered useful insights is the analysis of the binding of Tyrosine to tyrosyl-tRNA synthetase, where the difference in free energy of binding of the substrate to the wild type and the Tyr169 → Phe mutant of the enzyme was calculated by thermodynamic integration following Eq. (51) [117]. The experi-

Table 7
Free energy decomposition of non-covalent interactions in Tyr → Phe simulations (kcal/mol)

| Term | $\Delta G_b$ | $\Delta G_f$ | $\Delta\Delta G$ |
|---|---|---|---|
| Protein | 5.3 | 6.1 | −0.83 |
| Solvent | 0.3 | 5.6 | −5.2 |
| Ligand | 8.6 | – | 8.6 |
| Total | 14.2 | 11.7 | 2.6 |

mental analysis of a series of tyrosyl-tRNA mutants, including the Tyr 169 → Phe mutant is given in Wells and Fersht [118]. (An excellent overview of 'protein engineering' methods, as such mutation studies are now called, is given in Ch. 15 of Fersht [119].) The calculations showed that the mutation reduces the binding affinity by approximately 3 kcal/mol, a result in good agreement with experiment. The path chosen in the study was the linear transformation of the hydroxyl group of Tyr 169 to a hydrogen atom in the bound system (enzyme plus substrate) and the free system (see Fig. 5). From the diagram, the difference $\Delta\Delta G$ in binding free energy between the mutant ($\Delta G_m$) and the wild type ($\Delta G_w$) is equal by Hess's Law (see Tembe and McCammon [120]) to

$$\Delta\Delta G = \Delta G_b - \Delta G_m = \Delta G_m - \Delta G_w \tag{53}$$

where $\Delta G_b$ is the free energy difference between the mutant and the wild type in the bound form and $\Delta G_f$ is the corresponding free energy difference for the free enzyme. The experiments are done following the vertical paths for the wild type and the mutant, while the calculations are done most easily along the horizontal (alchemical) path. It would, of course, be possible to follow the vertical paths, which would provide the potential of mean force curve for binding the tyrosine ligand to the wild type and mutant protein.

The free energy difference was decomposed in several different ways. First, it was decomposed into electrostatic, van der Waals, and covalent interactions. The dominant contribution was found to arise from the electrostatic term. The free energy difference was also decomposed into interactions of the mutated group with the protein, the ligand, and the solvent (see Table 7). As expected, the largest contributions were made by the interactions with the ligand and the solvent. The protein contribution was further analyzed into contributions from the neighboring amino acids. It was found that, although the total protein contribution was small, it contained large compensating contributions from individual amino acids. This is probably due to the fact that the position and orientation of the mutated hydroxyl shifts between the free and bound states, forming favorable interactions with some residues in one case and with other residues in the other. The solvent term was also broken down into contributions from water molecules at different distances. It was found that significant contributions arise from relatively distant water molecules.

Although the analysis just described for decomposing the free energy is very useful, one can ask whether it is possible to define contributions to stability that are path-independent. To better understand the type of decomposition we are interested in, consider a hydrogen bond in the interior of a protein. How could one define precisely the 'contribution' of this hydrogen bond to protein stability? Obviously, it is not the interaction energy between the CO and NH groups because, first, the direct CO–NH interaction is only a small part of the interactions in which these two groups are involved in the protein interior [10]. Moreover, this neglects the fact that these two groups in the unfolded state interact favorably with the solvent. Then the free energy contribution to stability is determined both by the folded and unfolded state, as was already pointed out for the free energy simulations above. Clearly, any definition of the 'contribution' of protein groups to stability must consider (a) the multitude of interactions made by these protein groups; and (b) the effects of changes in the solvation of these groups.

The starting point for the present discussion is Eq. (29). The decomposition presented in Eq. (29) is rigorous as long as the intramolecular interaction terms are separable from the solute–solvent and solvent–solvent interaction terms. This is the case for most energy functions in use today [e.g. Eq. (31)]. Actually, Eq. (29) can be generalized to take account of the presence of three-body forces. For the intramolecular energy of the macromolecule, $H_{mm}$, we assume that the empirical energy function is valid (see Section 4). The non-bonded interactions can naturally, though arbitrarily be assigned to the interacting atoms, half of the interaction to each partner. Usually, the bonded terms ('strain energy') are treated separately. However, they too can be assigned to the atoms which are involved: two atoms for bonds, three for angles, and four for dihedral angles. In this model, each atom in the macromolecule is assigned one half of its non-bonded binding energy with all other atoms, one-half of the covalent bond energy of the bonds, one-third of the bond angle energy, and one-fourth of the dihedral energy in which it participates.

Decomposition of the solvation free energy is more complicated. Eq. (10) for the solvation free energy is not very convenient for application to a macromolecule because it does not easily lend itself to a decomposition of the solvation free energy into contributions from groups. The inhomogeneous theory of Section 5 is more convenient for this purpose. Once group solvation free energies are defined, one can define the effective free energy (GFE) of group $i$ as

$$W_i = H_i + \frac{1}{2}\sum_j H_{ij} + \Delta G_i^{\mathrm{slv}} \tag{54}$$

where $H_i$ corresponds to the internal group energy (including the bonded terms within group $i$ and any intra-group non-bonded interactions), $H_{ij}$ corresponds to the interactions between groups $i$ and $j$, and $\Delta G_i^{\mathrm{slv}}$ is the solvation free energy of group $i$ (see Section 5). Based on these results, the free energy of unfolding can be written

$$\Delta G = \sum_i \langle \Delta W_i \rangle - T\Delta S^{\mathrm{conf}} \tag{55}$$

where $\langle \Delta W_i \rangle$ is equal to the contribution of group $i$ to stability (the brackets indicate a configurational average)

$$\langle \Delta W_i \rangle = \frac{1}{2}\sum_j \langle \Delta H_{ij} \rangle + \langle \Delta H_i \rangle + \langle \Delta\Delta G_i^{\mathrm{slv}} \rangle \tag{56}$$

where the extra $\Delta$ symbols in Eq. (56), relative to Eq. (54), denotes the the difference between the folded and unfolded state. Although $\langle \Delta W_i \rangle$ includes the group's share of the bonded energy, $\langle \Delta H_i^{\mathrm{bond}} \rangle$ is not expected to make a large contribution when differences between GFEs are considered. As is indicated in Eq. (55), $\langle \Delta W_i \rangle$ does not include any conformational entropy. The sum of all group contributions is not the free energy of unfolding, $\Delta G$, but the change in effective energy upon unfolding, $\Delta W$.

This clear definition of group contributions can be of use in interpreting mutation experiments. Among site-directed mutagenesis experiments the easiest to interpret are the so called 'non-disruptive deletions' [121], where an amino acid sidechain is truncated to eliminate a particular interaction without disrupting the protein structure or introducing additional interactions. For example, the mutation of an ILE to a VAL should give information about van der Waals and hydrophobic interactions, whereas the mutation of a SER to an ALA should give information about hydrogen bonding interactions. However, it is important to note [105,119] that elimination of a group by mutation does not measure the contribution of that group to protein stability. Deletion of a group leads to loss of the contribution of the deleted group, and modification of the contributions of the neighboring groups:

$$\Delta\Delta G(i) = \langle \Delta W_i \rangle + \sum_{j\neq i} \Delta\langle \Delta W_j \rangle, \quad j \text{ close to } i \tag{57}$$

Thus, in protein engineering the sum of the changes in stability or binding affinity upon mutation of a series of groups does not give the total stability or binding affinity, even if the protein structure $[p(\mathbf{q})]$ is not affected. This is true because the modification of the contributions of interacting groups is counted twice [115].

In several protein engineering studies a hydrophobic sidechain, partly or fully buried, was truncated to directly estimate the contribution of hydrophobic interactions to protein stability [122–127]. Simulation studies of these mutations were also undertaken [128,129]. Perhaps the most surprising result was that the magnitude of the destabilization observed was higher than that expected from water to octanol transfer free energies of model compounds. Crystallographic studies showed that the magnitude of the destabilization correlated with the size of the cavity left in the mutant when the group was deleted [126]. This finding suggested that the extra destabilization free energy can be attributed to the free energy cost of having a cavity in the protein interior [126,130,131]. GFEs offer a particularly simple way of analyzing the problem. As mentioned above, deletion of a group will eliminate its contribution but also modify the GFEs of its neighboring groups. Consider for example the ILE → VAL mutation, in which one $CH_2$ group is removed. In the unfolded state, where the mutated residue is most likely to be mainly exposed to solvent, the deletion eliminates the GFE of a $CH_2$ group, which is dominated by the solvation free energy of that group, $\Delta G_i^{\mathrm{solv}}$. The GFEs of the surrounding groups may change somewhat because the elimination of the $C\delta$ group increases the solvent exposure of the $C\gamma$ group. In the folded state, the sidechain is buried in the cases studied, so that the GFE of the deleted group will be approximately equal to one-half of the van der Waals interaction of the group with its surroundings, $E_{\mathrm{vdW}}$. When the group is deleted, assuming that the structure remains constant, what is lost is the whole $E_{\mathrm{vdw}}$, only half of which 'belongs' to the deleted group. The rest belongs to the GFEs of the surrounding groups. If the protein relaxes to completely fill the gap, the van der Waals energies of the surrounding groups will approximately return to their original value. Thus, the measured difference in protein stability between the mutant and the wild type corresponds to the intrinsic contribution of the deleted group only in the limit of complete protein relaxation. The van der Waals interaction energy for a methylene group in the protein interior is approximately $-3.5$ kcal/mol, relative to approximately $-2$ kcal/mol in liquid-like environments [10]. Therefore, the maximum extra destabilization one can expect for the deletion of a $CH_2$ group is one-half of that or approximately 1.75 kcal/mol. The experimental values for Ile to Val mutations are in the range of 0.3–1.8 kcal/mol [127].

Another area where analysis on the thermodynamics of an atomic basis is of interest concerns the magnitude of contribution of hydrogen bonds to protein stability [10,132,133]. The classical view has been that upon folding hydrogen bonds between protein atoms replace similar hydrogen bonds with the solvent in the unfolded state and thus lead to little net stabilization of the native structure. Their major role is then presumed to be that they restrict the number of possible folded structures and thus contribute to the stabilization of a unique structure. Mutagenesis experiments have been interpreted as indicating that hydrogen bonds make a favorable contribution to stability or binding [121,134]. This also emerged in an empirical treatment of protein thermodynamics based on model compound data [52,65]. To analyze these results in the present framework, we have to define the term 'contribution of a hydrogen bond' to protein stability. This corresponds to the free energy gain from the interaction of the hydrogen bonding groups in the protein minus the cost of desolvating these groups (i.e. having them interacting with solvent in the denatured state). One problem that arises in evaluating this quantity is that it is not well defined because polar groups in the interior of proteins are involved in a multitude of interactions with other polar groups, some of which are hydrogen bonding and some of which are not [10]. In addition, hydrogen bonding groups have significant van der Waals interactions with nearby non-polar groups. The cost of desolvating the two hydrogen bonding groups cannot be assigned to their pairwise interaction alone, but includes all the interactions in which these groups are involved. Consequently, there is a conceptual difficulty in isolating the 'hydrogen bonding' energy [10]. For these reasons, it is preferable to speak of contribution of groups, GFEs, rather than of interactions, which, as shown above, are well defined quantities. Calculation of the GFE of a hydrogen bonding group in the folded and unfolded protein would give the contribution of this group to stability. If the group is buried, then its stability contribution will be favorable if half of the sum of all of its interactions in the protein interior is more negative than its solvation free energy in the unfolded protein. That will be the case, for example, if the group forms multiple hydrogen bonds and possibly other favorable dipole–dipole interactions [10] in the protein interior.

## 11. Future directions

Equilibrium thermodynamics is an established tool for the study of biological processes in vitro. Knowing the equilibrium properties of macromolecules is an essential first step for the characterization of the processes in which they participate. In combination with molecular models and statistical mechanics, it promises to provide a microscopic understanding of the mechanisms of many of the phenomena operating in living systems.

The cornerstone of a microscopic analysis of macromolecular thermodynamics is the energy function. The size and complexity of biological molecules necessitates the use of approximate energy functions. Although molecular mechanics force fields have deficiencies, the source of the largest errors at this time is in the calculation of the solvation term. Significant efforts are being made in developing quantitative models for the solvation free energy and decomposing them into the enthalpy and the entropy of solvation. Modeling of the solvent should not be restricted to pure water but should be extended to more complex solvent media that contain ions, non-neutral pH, cosolvents, cosolutes etc., to be able to understand the behavior in the complex cellular medium.

We have used protein folding as the example to illustrate the microscopic analysis of macroscopic thermodynamic data. This is, of course, only one area, albeit an important one, where thermodynamics can be used to study elementary biological process. Another area that is justifiably receiving considerable attention is the process of binding between macromolecules or between a macromolecule and a ligand. Many more complex events occur in the living cell; they include protein translocation through membranes, large conformational changes, active diffusion, DNA unwinding, vesicle budding, membrane fusion, and virus assembly. At an even larger scale, there are processes like cell division, cell differentiation, development, growth, and aging. The role of thermodynamics is to provide an approach for understanding the driving forces responsible for all these processes and for rationalizing the observed pathways. As more detailed data become available concerning the specific molecules involved, it will become possible to utilize techniques like the ones described here to achieve a detailed description of these events at the atomic level. Such an understanding may aid in learning how to control these cellular events so as to be able to alter them in desirable directions. This offers hope of aiding in the development of methods leading to the demise of cancerous cells or to an increase in the resistance to viral infections.

## Acknowledgments

Partial support of this work was provided by a grant from the National Science Foundation (T.L.) and by a grant from the National Institutes of Health to Harvard University (M.K.). This paper is based on Lazaridis et al. [10] and Lazaridis and Karplus [135].

## References

[1] J.T. Edsall, J. Wyman, Biophysical Chemistry, Acad. Press, New York, 1958.

[2] L. Pauling, The Nature of the Chemical Bond, Cornell University Press, Ithaca, New York, 1939, p. 1960.

[3] E. Schrödinger, What is Life?, Cambridge University Press, Cambridge, 1945.

[4] M.F. Perutz, Stereochemistry of cooperative effects in haemoglobin, Nature (London) 228 (1970) 726.

[5] M.F. Perutz, Stereochemistry of cooperative effects in haemoglobin, Nature (London) 228 (1970) 734.

[6] A. Szabo, M. Karplus, A mathematical model for structure–function relations in hemoglobin, J. Mol. Biol. 72 (1972) 163–197.

[7] A. Szabo, M. Karplus, Analysis of cooperativity in hemoglobin. Valency hybrids, oxidations, and methemoglobin replacement reactions, Biochemistry 14 (1975) 931–940.

[8] W.A. Eaton, E.R. Henry, J. Hofrichter, A. Mozzarelli, Is cooperative oxygen binding by hemoglobin really understood, Nat. Struc. Biol. 6 (1999) 351–358.

[9] G.I. Makhatadze, P.L. Privalov, Energetics of protein structure, Adv. Protein Chem. 47 (1995) 307–425.

[10] T. Lazaridis, G. Archontis, M. Karplus, Enthalpic contribution to protein stability: atom-based calculations and statistical mechanics, Adv. Protein Chem. 47 (1995) 231–306, (Academic Press, Inc.).

[11] J.T. Edsall, H. Wu, The First Theory of Protein Denaturation, Adv. Protein Chem. 46 (1931) 1–5 (Academic Press, Inc.).

[12] H. Wu, Studies on denaturation of proteins XIV. A theory of denaturation, Chin. J. Physiol. 5 (1931) 321–

344; reprinted in Adv. Protein Chem. 46 (1995) 6–26 (Academic Press, Inc.).

[13] A.I. Zotin, Thermodynamic Aspects of Developmental Biology, S. Karger, Basel, 1972.

[14] M. Eigen, The physics of molecular evolution, Chem. Scr. 26B (1986) 13–26.

[15] S.A. Kauffman, The Origins of Order, Oxford University Press, 1993.

[16] F.H. Arnold, Design by directed evolution, Acc. Chem. Res. 31 (1998) 125–131.

[17] D.C. Wiley, J.J. Skehel, The structure and function of the hemagglutinin membrane glycoprotein of influenza virus, Annu. Rev. Biochem. 56 (1987) 365–394.

[18] E.M. Marcotte, M. Pellegrini, H.L. Ng, D.W. Rice, T.O. Yeates, D. Eisenberg, Detecting protein function and protein–protein interactions from genome sequences, Science 285 (1999) 751–753.

[19] C.B. Anfinsen, Principles that govern folding of protein chains, Science 181 (1973) 223–230.

[20] M.L. Anson, A.E. Mirsky, J. Gen. Physiol. 17 (1934) 393–408.

[21] M. Perutz, 'Unboiling an Egg', Discovery, March 1940.

[22] D.S. Eisenberg, D.M. Crothers, Physical Chemistry: with Applications to the Life Sciences, Benjamin/ Cummings Pub. Co, Menlo Park, CA, 1979.

[23] D.B. Wetlaufer, S.S. Ristow, Acquisition of three-dimensional structure of proteins, Annu. Rev. Biochem. 42 (1973) 135–158.

[24] C.B. Anfinsen, H.A. Scheraga, Adv. Protein Chem. 29 (1975) 205–300.

[25] M.E. Goldberg, The second translation of the genetic message: protein folding and assembly, TIBS 10 (1985) 388–391.

[26] D. Baker, D.A. Agard, Kinetics versus thermodynamics in protein folding, Biochemistry 33 (1994) 7505–7509.

[27] D. Baker, Metastable states and folding free energy barriers, Nat. Struct. Biol. 5 (1998) 1021–1024.

[28] A.R. Dinner, M. Karplus, A metastable state in folding simulations of a protein model, Nat. Struct. Biol. 5 (1998) 236–241.

[29] M.K. Jain, Introduction to Biological Membranes, Wiley-Interscience, New York, 1988.

[30] B. Halle, T. Andersson, S. Forsen, B. Lindman, Protein hydration from water oxygen-17 magnetic relaxation, J. Am. Chem. Soc. 103 (1981) 500–508.

[31] G. Otting, E. Liepinsh, K. Wüthrich, Protein hydration in aqueous solution, Science 254 (1991) 974–980.

[32] J.A. Ernst, R.T. Clubb, H.-Z. Zhou, A.M. Gronenborn, G.M. Clore, Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR, Science 267 (1995) 1813–1816.

[33] H. Frauenfelder, S.G. Sligar, P.G. Wolynes, The energy landscape and motions of proteins, Science 254 (1991) 1598–1603.

[34] D.A. McQuarrie, Statistical Mechanics, Harper & Row, New York, 1976.

[35] T. Lazaridis, M. Karplus, Effective energy function for proteins in solution, Proteins 35 (1999) 133–152.

[36] A.D. MacKerell, D. Bashford, M. Bellott, et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.

[37] A. Ben-Naim, Standard thermodynamics of transfer. Uses and misuses, J. Phys. Chem. 82 (1978) 792–803.

[38] R. Elber, M. Karplus, Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin, Science 235 (1987) 318–321.

[39] J.A. McCammon, S. Harvey, Dynamics of Proteins and Nucleic Acids, Cambridge University Press, Cambridge, 1987.

[40] C.L. Brooks III, M. Karplus, B.M. Pettitt, Proteins: A Theoretical Perspective of Dynamics, Structure, & Thermodynamics, Adv. Chem. Phys. LXXI, John Wiley & Sons, 1988.

[41] M.J. Field, P.A. Bash, M. Karplus, A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations, J. Comp. Chem. 11 (1990) 700–733.

[42] A. Caflisch, M. Karplus, Acid and thermal denaturation of barnase investigated by molecular dynamics simulations, J. Mol. Biol. 252 (1995) 672–708.

[43] H. Reiss, Scaled particle methods in the statistical thermodynamics of fluids, Adv. Chem. Phys. 9 (1965) 1–84.

[44] J.P. Hansen, I.R. McDonald, Theory of Simple Liquids, Academic Press, London, 1986.

[45] F. Hirata, P.J. Rossky, An extended RISM equation for molecular polar fluids, Chem. Phys. Lett. 83 (1981) 329–334.

[46] B.M. Pettitt, P.J. Rossky, Alkali halides in water: ion–solvent correlations and ion–ion potentials of mean force at infinite dilution, J. Chem. Phys. 84 (1986) 5836–5844.

[47] H.-A. Yu, M. Karplus, A thermodynamic analysis of solvation, J. Chem. Phys. 89 (1988) 2366–2379.

[48] T. Ichiye, D. Chandler, Hypernetted chain closure reference interaction site method theory of structure and thermodynamics for alkanes in water, J. Phys. Chem. 92 (1988) 5257–5261.

[49] D.L. Beveridge, F.M. DiCapua, Free energy via molecular simulation, Ann. Rev. Biophys. Chem. 18 (1989) 431–492.

[50] P.A. Kollman, Free energy calculations: applications to chemical and biochemical phenomena, Chem. Rev. 93 (1993) 2395–2418.

[51] K. Lum, D. Chandler, J.D. Weeks, Hydrophobicity at small and large length scales, J. Phys. Chem. B 103 (1999) 4570–4577.

[52] N. Matubayasi, L.H. Reed, R.M. Levy, Thermodynamics of the hydration shell 1. Excess energy of a hydrophobic solute, J. Phys. Chem. 98 (1994) 10640–10649.

[53] T. Lazaridis, Inhomogeneous fluid approach to solvation thermodynamics 1. Theory, J. Phys. Chem. 102 (1998) 3531–3541.

[54] T. Lazaridis, Solvent reorganization energy and entropy in hydrophobic hydration, J. Phys. Chem. B 104 (2000) 4964–4979.

[55] T. Lazaridis, M.E. Paulaitis, Entropy of hydrophobic hydration: a new statistical mechanical formulation, J. Phys. Chem. 96 (1992) 3847–3855.

[56] T. Lazaridis, M.E. Paulaitis, Simulation studies of the hydration entropy of simple hydrophobic solutes, J. Phys. Chem. 98 (1994) 635–642.

[57] T. Lazaridis, Solvent size vs cohesive energy density as the origin of hydrophobicity, Acc. Chem. Res. 34 (2001) 931–937.

[58] T. Lazaridis, Inhomogeneous fluid approach to solvation thermodynamics 2. Applications to simple fluids, J. Phys. Chem. 102 (1998) 3542–3550.

[59] D. Eisenberg, A.D. McLachlan, Solvation energy in protein folding and binding, Nature 319 (1986) 199–203.

[60] T. Ooi, M. Oobatake, G. Nemethy, H.A. Scheraga, Accessible surface-areas as a measure of the thermodynamic parameters of hydration of peptides, Proc. Natl. Acad. Sci. 84 (1987) 3086–3090.

[61] L. Wesson, D. Eisenberg, Atomic solvation parameters applied to molecular dynamics of proteins in solution, Protein Sci. 1 (1992) 227–235.

[62] F. Fraternali, W.F. van Gunsteren, Conformational transitions of a dipeptide in water, J. Mol. Biol. 256 (1996) 939–948.

[63] A. Cavalli, P. Ferrara, A. Caflisch, Weak temperature dependence of the free energy surface and folding pathways of structured peptides, Proteins: Struct. Funct. Genet. 47 (2002) 305–314.

[64] Y.K. Kang, G. Nemethy, H.A. Scheraga, Free energies of hydration of solute molecules 1. Improvement of the hydration shell model by exact computations of overlap volumes, J. Chem. Phys. 91 (1987) 4105–4109.

[65] F. Colonna-Cesari, C. Sander, Excluded volume approximation to protein−solvent interaction. The solvent contact model, Biophys. J. 57 (1990) 1103–1107.

[66] P.F.W. Stouten, C. Frommel, H. Nakamura, C. Sander, An effective solvation term based on atomic occupancies for use in protein simulations, Mol. Simul. 10 (1993) 97–120.

[67] G.I. Makhatadze, P.L. Privalov, Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration, J. Mol. Biol. 232 (1993) 639–659.

[68] K. Sharp, B. Honig, Electrostatic interactions in macromolecules: theory and applications, Ann. Rev. Biophys. Biophys. Chem. 19 (1990) 301–332.

[69] B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry, Science 268 (1995) 1144–1149.

[70] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics, J. Am. Chem. Soc. 112 (1990) 6127–6129.

[71] M. Schaefer, M. Karplus, A comprehensive analytical treatment of continuum electrostatics, J. Phys. Chem. 100 (1996) 1578–1599.

[72] M. Schaefer, C. Bartels, M. Karplus, Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model, J. Mol. Biol. 284 (1998) 835–847.

[73] T. Lazaridis, M. Karplus, Orientational correlations and entropy in liquid water, J. Chem. Phys. 105 (1996) 4294–4316.

[74] T. Lazaridis, M. Karplus, Discrimination of the native from misfolded protein models with an energy function including implicit solvation, J. Mol. Biol. 288 (1999) 477–487.

[75] T. Lazaridis, I. Lee, M. Karplus, Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin, Protein Sci. 6 (1997) 2589–2605.

[76] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comp. Chem. 4 (1983) 187–217.

[77] S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case, An all atom force field for simulations of proteins and nucleic acids, J. Comput. Chem. 7 (1986) 230–252.

[78] W. Kauzmann, Some factors in the interpretation of protein denaturation, Adv. Protein Chem. 14 (1959) 1–63.

[79] S. Miller, J. Janin, A.M. Lesk, C. Chothia, Interior and surface of monomeric proteins, J. Mol. Biol. 196 (1987) 641–656.

[80] K.A. Dill, D. Shortle, Denatured states of proteins, Ann. Rev. Biochem. 60 (1991) 795.

[81] D. Shortle, The denatured state (the other half of the folding equation) and its role in protein stability, FASEB J. 10 (1996) 27–34.

[82] O.B. Ptitsyn, Molten globule and protein folding, Adv. Protein Chem. 47 (1995) 83–229.

[83] B. Nölting, R. Golbik, A.S. Soler-González, A.R. Fersht, Circular dichroism of denatured barstar suggests residual structure, Biochemistry 36 (1997) 9899–9905.

[84] D. Shortle, H.S. Chan, K.A. Dill, Modeling the effects of mutations on the denatured states of proteins, Protein Sci. 1 (1992) 201–215.

[85] E.E. Lattman, K.M. Fiebig, K.A. Dill, Modeling compact denatured states of proteins, Biochemistry 33 (1994) 6158–6166.

[86] C.J. Bond, K.-B. Wong, J. Clarke, A.R. Fersht, V. Daggett, Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 13409–13413.

[87] S.L. Kazmirsky, V. Daggett, Simulation of the structural and dynamic properties of unfolded proteins: the 'molten coil' state of bovine pancreatic trypsin inhibitor, J. Mol. Biol. 277 (1998) 487–506.

[88] J.F. Brandts, Thermodynamics of protein denaturation. 2. Model of reversible denaturation + interpretations regarding stability of chymotrypsinogen, J. Am. Chem. Soc. 86 (1964) 4302–4314.

[89] J.M. Sturtevant, Heat capacity and entropy changes in processes involving proteins, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 2236–2240.

[90] R.L. Baldwin, Temperature-dependence of the hydrophobic interaction in protein folding, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 8069–8072.

[91] R.S. Spolar, J.-H. Ha, M.T. Record, Hydrophobic effect in protein folding and other noncovalent processes involving proteins, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 8382–8385.

[92] G.I. Makhatadze, P.L. Privalov, Heat capacity of proteins. 1. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect, J. Mol. Biol. 213 (1990) 375–384.

[93] K.P. Murphy, S.J. Gill, Solid model compounds and the thermodynamics of protein unfolding, J. Mol. Biol. 222 (1991) 699–709.

[94] J.T. Edsall, Apparent molar heat capacities of amino acids and other organic compounds, J. Am. Chem. Soc. 57 (1935) 1506–1510.

[95] R.S. Spolar, J.R. Livingstone, M.T. Record, Use of liquid-hydrocarbon and amide transfer data to estimate contributions to thermodynamics functions of protein folding from the removal of nonpolar and polar surface from water, Biochemistry 31 (1992) 3947–3955.

[96] P.L. Privalov, E.I. Tiktopoulo, S.Y. Venyaminov, Y.V. Griko, G.I. Makhatadze, N.N. Khechinashvili, Heat capacity and conformation of proteins in the denatured state, J. Mol. Biol. 205 (1989) 737–750.

[97] K.P. Murphy, E. Freire, Thermodynamics of structural stability and cooperative folding behavior in proteins, Adv. Protein Chem. 43 (1992) 313–361.

[98] J. Gomez, V.J. Hilser, D. Xie, E. Freire, The heat capacity of proteins, Proteins 22 (1995) 404–412.

[99] T. Lazaridis, M. Karplus, Heat capacity and compactness of denatured proteins, Biophys. Chem. 78 (1999) 207–217.

[100] B.R. Brooks, M. Karplus, Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor, Proc. Natl. Acad. Sci. U.S.A. 80 (1983) 6571–6575.

[101] T. Lazaridis, M. Karplus, 'New view' of protein folding reconciled with the old through multiple unfolding simulations, Science 278 (1997) 1928–1931.

[102] S.E. Jackson, A.R. Fersht, Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, Biochemistry 30 (1991) 10428–10435.

[103] M. Karplus, T. Ichiye, B.M. Pettitt, Configurational entropy of native proteins, Biophys. J. 52 (1987) 1083–1085.

[104] P.R. Gerber, A.E. Mark, W.F. van Gunsteren, An approximate but efficient method to calculate free energy trends by computer simulation. Application to dihydrofolate reductase inhibitor complexes, J. CAMD 7 (1993) 305–323.

[105] A.R. Fersht, Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis, Biochemistry 27 (1988) 1577–1580.

[106] A.E. Mark, W.F. van Gunsteren, Decomposition of the free energy of a system in terms of specific interactions, J. Mol. Biol. 240 (1994) 167–176.

[107] K.A. Dill, Additivity principles in biochemistry, J. Biol. Chem. 272 (1997) 701–704.

[108] J. Gao, K. Kuczera, B. Tidor, M. Karplus, Hidden thermodynamics of mutant proteins: a molecular dynamics analysis, Science 244 (1989) 1069–1072.

[109] T.P. Straatsma, J.A. McCammon, Computational alchemy', Ann. Rev. Phys. Chem. 43 (1992) 407–435.

[110] C.L. Brooks III, Molecular simulations of protein structure, dynamics and thermodynamics, in: C.R.A. Catlow, S. Parker, M. Allen (Eds.), Computer Modeling of Fluids Polymers and Solids, Kluwer Academic Publishers, Dordecht, 1990.

[111] T. Simonson, A.T. Brünger, Thermodynamics of protein−peptide interactions in the ribonuclease-S system studied by molecular dynamics and free energy calculations, Biochemistry 31 (1992) 8661–8674.

[112] S. Boresch, M. Karplus, The meaning of component analysis: decomposition of the free energy in terms of specific interactions, J. Mol. Biol. 254 (1995) 801–807.

[113] T. Simonson, G. Archontis, M. Karplus, Free energy simulations come of age: protein–ligand recognition, Acc. Chem. Res. (in press, 2002).

[114] G.P. Brady, K.A. Sharp, Decomposition of interaction free energies in proteins and other complex-systems, J. Mol. Biol. 254 (1995) 77–85.

[115] G. Archontis, M. Karplus, Cumulant expansion of the free energy: application to free energy derivatives and component analysis, J. Chem. Phys. 105 (1996) 11246–11260.

[116] G.P. Brady, A. Szabo, K.A. Sharp, On the decomposition of free energies, J. Mol. Biol. 263 (1996) 123–125.

[117] F.T.K. Lau, M. Karplus, Molecular recognition in proteins: simulation analysis of substrate binding by tyrosyl-tRNA synthetase mutants, J. Mol. Biol. 236 (1994) 1049–1066.

[118] T.N.C. Wells, A.R. Fersht, Use of binding energy in catalysis analyzed by mutagenesis of the tyrosyl-tRNA synthetase, Biochemistry 25 (1986) 1881–1886.

[119] A.R. Fersht, Structure and Mechanism in Protein Science: A guide to Enzyme Catalyses and Protein Folding, W.H. Freeman, 1999.

[120] B.L. Tembe, J.A. McCammon, Ligand–receptor interactions, Comput. Chem. 8 (1984) 281–283.

[121] A.R. Fersht, The hydrogen bond in molecular recognition, TIBS 12 (1987) 301–304.

[122] M. Matsumura, W.J. Becktel, B.W. Matthews, Hydrophobic stabilization in T4 lysozyme determined directly

by multiple substitutions of Ile-3, Nature 334 (1988) 406–410.

[123] J.T. Kellis, K. Nyberg, A.R. Fersht, Energetics of complementary side-chain packing in a protein hydrophobic core, Biochemistry 28 (1989) 4914–4922.

[124] D. Shortle, W.E. Stites, A.K. Meeker, Contributions of the large hydrophobic amino-acids to the stability of staphylococcal nuclease, Biochemistry 29 (1990) 8033–8041.

[125] W.S. Sandberg, T.C. Terwilliger, Energetics of repacking a protein interior, Proc. Natl. Acad. Sci. U.S.A. 88 (1991) 1706–1710.

[126] A.E. Eriksson, W.A. Baase, X.-J. Zhang, D.W. Heinz, M. Blaber, E.P. Baldwin, B.W. Matthews, Response of a protein-structure to cavity-creating mutations and its relation to the hydrophobic effect, Science 255 (1992) 178–183.

[127] L. Serrano, J.T. Kellis, P. Cann, A. Matouschek, A.R. Fersht, The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability, J. Mol. Biol. 224 (1992) 783–804.

[128] M. Prévost, S.J. Wodak, B. Tidor, M. Karplus, Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile 96 → Ala mutation in barnase, Proc. Natl. Acad. Sci. U.S.A. 88 (1991) 10880–10884.

[129] S.F. Sneddon, D.J. Tobias, The role of packing interactions in stabilizing folded proteins, Biochemistry 31 (1992) 2842–2846.

[130] B. Lee, Estimation of the maximum change in stability of globular proteins upon mutation of a hydrophobic residue to another of smaller size, Protein Sci. 2 (1993) 733–738.

[131] A.A. Rashin, Aspects of protein energetics and dynamics, Prog. Biophys. Mol. Biol. 60 (1993) 73–200.

[132] B. Honig, A.-S. Yang, Free energy balance in protein-folding, Adv. Protein Chem. 46 (1995) 27–58.

[133] J.K. Myers, C.N. Pace, Hydrogen bonding stabilizes globular proteins, Biophys. J. 71 (1996) 2033–2039.

[134] B.A. Shirley, P. Stanssens, U. Hahn, C.N. Pace, Contribution of hydrogen bonding to the conformational stability of ribonuclease-T1, Biochemistry 31 (1992) 725–732.

[135] T. Lazaridis, M. Karplus, Microscopic basis of macromolecular thermodynamics, in: E. Di Cera (Ed.), Thermodynamics in Biology, Oxford University Press, 2000.