ELSEVIER

Biophysical Chemistry 64 (1997) 69–79

Biophysical Chemistry

Structure-based statistical thermodynamic analysis of T4 lysozyme mutants: structural mapping of cooperative interactions [1]

Vincent J. Hilser, Benjamin D. Townsend, Ernesto Freire *

Department of Biology and Biocalorimetry Center, The Johns Hopkins University, Baltimore, MD 21218, USA

Received 7 June 1996; revised 21 August 1996; accepted 21 August 1996

Abstract

The recent development of a structural parameterization of the energetics of protein folding has permitted the incorporation of the functions that describe the enthalpy, entropy and heat capacity changes, i.e. the individual components of the Gibbs energy, into a statistical thermodynamic formalism that describes the distribution of conformational states under equilibrium conditions. The goal of this approach is to construct with the computer a large ensemble of conformational states, and then to derive the most probable population distribution, i.e. the distribution of states that best accounts for a wide array of experimental observables. This analysis has been applied to four different mutants of T4 lysozyme (S44A, S44G, V131A, V131G). It is shown that the structural parameterization predicts well the stability of the protein and the effects of the mutations. The entire set of folding constants per residue has been calculated for the four mutants. In all cases, the effect of the mutations propagates beyond the mutation site itself through sequence and three-dimensional space. This phenomenon occurs despite the fact that the mutations are at solvent-exposed locations and do not directly affect other interactions in the protein. These results suggest that single amino acid mutations at solvent-exposed locations, or other locations that cause a minimal perturbation, can be used to identify the extent of cooperative interactions. The magnitude and extent of these effects and the accuracy of the algorithm can be tested by means of NMR-detected hydrogen exchange.

Keywords: Cooperative interactions; Structural mapping; Structure-based statistical thermodynamic analysis; T4 lysozyme mutants

1. Introduction

Proteins cannot be considered as an equilibrium between two or a few discrete conformational states since the number of conformations that are available to them is astronomically large. To understand protein equilibrium it is necessary to understand *why* the vast majority of states never become populated and to identify *which* conformational states have the highest probability of being populated. It is clear that simple two-state, three-state or other models that only include a few discrete states cannot account for observables like hydrogen exchange protection factors even under conditions in which folding/unfolding transitions are well accounted for by these models. Recently, three developments have provided the

* Corresponding author. Tel. (410) 516-7743; fax (410) 516-6469; e-mail bcc@biocal2.bio.jhu.edu.

[1] Supported by grants from the National Institutes of Health (RR04328 and GM51362).

PII S0301-4622(96)02220-X

necessary foundation for an experimentally based statistical thermodynamic formulation of the conformational equilibrium of proteins: 1) The development of a structural parameterization of protein energetics that accurately accounts for the Gibbs energy if the high resolution structure of a protein conformation is known [1–6]; 2) The development of a computer algorithm that generates a large ensemble of conformations starting from the high resolution structure of the native state [7–9]; and, 3) The technique of NMR-detected hydrogen exchange with its ability to provide structural and energetic information at the residue level (see for example [10–20]) and, therefore, the possibility of empirically testing statistical thermodynamic models. Together, these developments permit the evaluation of the most probable distribution for the ensemble of states generated by the computer. Specifically, the probability of any given state, P_i, is equal to

$$P_i = \frac{\exp(-\Delta G_i/RT)}{Q} \quad (1)$$

where the statistical weights or Boltzmann exponents ($\exp(-\Delta G_i/RT)$) are defined in terms of the relative Gibbs energies ΔG_i for each state (R is the gas constant and T the absolute temperature) and Q is the conformational partition function defined as the sum of the statistical weights of all the states in the ensemble

$$Q = \sum_{i=0}^{n} \exp(-\Delta G_i/RT) \quad (2)$$

Having access to the partition function and the probability distribution of states allows prediction of experimental properties using standard statistical thermodynamic methods. The accuracy of the approach is evaluated by comparison of predicted and experimental physical observables. In general, the larger the number and type of physical observables that can be accounted for, the more closely the computer-generated ensemble will be expected to match the actual distribution of protein conformations in solution.

In this paper we have applied the above formalism to four different Ala and Gly single amino acid mutations at solvent-exposed locations in T4 lysozyme. It is shown that the structural parameterization accurately accounts for the differences in stability and energetics of the mutants. In addition, the stability constants and apparent Gibbs energies per residue have been evaluated. It is shown that the differences in Gibbs energies per residue between mutants provide a unique way of evaluating the extent and propagation of cooperative interactions in the molecule. Hence, these results suggest that there exist structural determinants of cooperativity which can be predicted from the crystallographic or NMR structure of a protein, and which can be experimentally verified by NMR-detected hydrogen exchange experiments.

2. General theory

Three elements are required in the statistical thermodynamic modelling of conformational equilibrium: 1) A way of calculating the Gibbs energy for an arbitrary structure; 2) A way of generating a large number of conformational structures; and, 3) A set of statistical descriptors that describes in a concise way the main features of the most probable distribution of states in the ensemble. These elements are summarized below.

2.1. Calculation of Gibbs energies from structure

It has been shown before that the free energy of folding ($\Delta G \equiv \Delta H - T \cdot \Delta S$) can be parameterized in terms of structural parameters [1–6]. This parameterization has been derived from the analysis of protein data and involves calculation of the relative heat capacity (ΔC_p), enthalpy (ΔH) and entropy (ΔS) of each state at the desired temperature.

The heat capacity change is a weak function of temperature and has been parameterized in terms of changes in solvent accessible surface areas (ΔASA) since it originates mainly from changes in hydration [3,21,22]:

$$\Delta C_p = \Delta C_{p,\mathrm{ap}} + \Delta C_{p,\mathrm{pol}} \quad (3a)$$

$$\Delta C_p = a_C(T) \cdot \Delta\mathrm{ASA}_{\mathrm{ap}} + b_C(T) \cdot \Delta\mathrm{ASA}_{\mathrm{pol}} \quad (3b)$$

where the coefficients $a_C(T) = 0.45 + 2.63 \times 10^{-4} \cdot (T - 25) - 4.2 \times 10^{-5} \cdot (T - 25)^2$ and $b_C(T) = -0.26 + 2.85 \times 10^{-4} \cdot (T - 25) + 4.31 \times 10^{-5} \cdot (T - 25)^2$. The hydration of the hydroxyl group in

aliphatic hydroxyl side chains (Ser and Thr) appears to contribute positively and not negatively to ΔC_p as previously assumed (0.17 cal K^{-1} mol^{-1} $Å^{-2}$ at 25°C) [22,23]. In the equation above, ΔASA changes are in $Å^2$ and the heat capacity in cal K^{-1} mol^{-1}. In general, for low temperature calculations ($T < 80$°C) the temperature-independent coefficients are sufficient [3,22]. Specific effects like heat capacity changes associated with changes in protonation, differential binding of ligands or denaturants, etc. need to be considered individually [3,22].

The bulk of the enthalpy change, the so-called generic enthalpy change (ΔH_{gen}) also scales in terms of ΔASA changes and at the reference temperature of 60°C it can be written as

$$\Delta H_{gen}(60) = a_H(60) \cdot \Delta ASA_{ap} + b_H(60) \cdot \Delta ASA_{pol} \quad (4)$$

where $a_H(60) = -8.44$ and $b_H(60) = 31.4$ and the enthalpy is in cal mol^{-1} [1,2,5].

In the calculation of the entropy change, two primary contributions are included, one due to changes in solvation and the other due to changes in conformational degrees of freedom ($\Delta S = \Delta S_{solv} + \Delta S_{conf}$). The entropy of solvation can be written in terms of the heat capacity if the temperatures at which the apolar and polar hydration entropies are zero ($T^*_{S,ap}$ and $T^*_{S,pol}$) are used as reference temperatures

$$\Delta S_{solv} = \Delta S_{solv,ap} + \Delta S_{solv,pol} \quad (5a)$$

$$\Delta S_{solv} = \Delta C_{p,ap} \ln(T/T^*_{S,ap}) + \Delta C_{p,pol} \ln(T/T^*_{S,pol}) \quad (5b)$$

$T^*_{S,ap}$ is known to be equal to 385.15 K [24,25] and $T^*_{S,pol}$ has been recently found to be close to 335.15 K [4]. The entropy of apolar hydration appears to be additive in the sense that elementary coefficients obtained from small molecules or proteins are similar. The situation is not the same for polar hydration which appears to be largely nonadditive [4,5]. The coefficients presented here were obtained from an analysis of the protein database; therefore, they reflect the average environment found in proteins and do not necessarily extrapolate to small molecules.

Conformational entropies are evaluated by explicitly considering the following three contributions for each amino acid: 1) $\Delta S_{bu \rightarrow ex}$, the entropy change associated with the transfer of a side chain that is buried in the interior of the protein to its surface; 2) $\Delta S_{ex \rightarrow u}$, the entropy change gained by a surface-exposed side chain when the peptide backbone unfolds; and, 3) ΔS_{bb}, the entropy change gained by the backbone itself upon unfolding. The magnitude of these terms for each amino acid has been estimated by computational analysis of the probability of different conformers as a function of the dihedral and torsional angles [4,26]. Additional entropic contributions due to the presence of disulfide bridges are estimated as described by Ref. [27].

Additional contributions to the Gibbs energy such as those due to protonation, specific ligands, etc. are calculated on an individual basis from knowledge of the pKs, protonation enthalpies and experimental pH, or knowledge of the binding constant, binding enthalpy and ligand concentration [5,22].

Solvent-accessible surface areas are calculated from structure by using the Lee and Richards algorithm [28] using a new set of C language libraries written in this laboratory. This mathematical abstraction layer permits a user to concentrate on structural coordinate manipulation while the accessible surface area calculation is performed by independent libraries linked to the program. The libraries require only three-dimensional coordinate data, atomic radii, a Z-plane integration slice width and the radius of a rolling ball solvent to operate. In all calculations, a slice width of 0.25 Å and a solvent radius of 1.4 Å are used.

Changes in polar and apolar solvent accessible areas upon unfolding (ΔASA_{ap} and ΔASA_{pol}) are calculated by using a set of optimized ASA_{ap} and ASA_{pol} values for each amino acid in the unfolded state [6]. Calculated ΔASA values are then used to evaluate ΔH, ΔC_p and ΔS_{solv} values. In addition, for each residue the state of the side chain (buried, exposed in a folded region, exposed in an unfolded region) and the backbone (folded or unfolded) are determined in order to evaluate conformational entropies. The structural parameterization has been incorporated in a computer program (Virtual Differential Scanning Calorimetry, VDSC) that generates the expected heat capacity function from a PDB data file containing the atomic coordinates of a protein; and a second program (CoreFHT) that predicts the NMR-

derived hydrogen exchange protection factors for individual amino acids.

2.2. Accuracy of structural parameterization.

In previous publications, we have evaluated the accuracy with which the structural parameterization accounts for the magnitude of the enthalpy and heat capacity changes associated with protein folding/unfolding [3,5]. Only recently, the basic elements for the parameterization of the entropy were developed [4,26] and therefore an evaluation of the Gibbs energy from structure has become possible. It has recently been shown [6] that the structural parameterization is able to account for the intrinsic helix propensities of each amino acid. Application of the analysis to four different systems: T4 lysozyme [29], Barnase [30], a synthetic leucine zipper [31], and a synthetic peptide [32] gave excellent results. For T4 lysozyme, the average value of the absolute difference between predicted and experimental ΔG values was 0.09 kcal mol^{-1}, for barnase 0.14 kcal mol^{-1}, for the synthetic coiled coil 0.11 kcal mol^{-1} and for the synthetic peptide 0.08 kcal mol^{-1}. Furthermore, for nineteen amino acids a correlation analysis between the results obtained with the structural parameterization and the empirical helix propensity scale developed by Muñoz and Serrano [33] yielded a slope of 1.089 ± 0.09 and a correlation coefficient of 0.94.

The most important test in the accuracy of any structure-based stability calculation is the prediction of the protein transition temperature (T_m) for thermal unfolding. Fig. 1 presents the predicted T_m values for thirteen different proteins at their pH of maximal stability. Depending on the protein, the pH of maximal stability usually lies anywhere between 5 and 7 as indicated in the figure. With the current level of refinement of the structural parameterization (as implemented in VDSC version 1.0), the mean value of the difference between predicted and experimental T_m values is -0.2 with a standard deviation of ± 9.5°C. This is the error in the prediction of the absolute value of T_m; the error is much smaller in the prediction of the effect of single amino acid mutations on the stability of a protein, as discussed in this paper.

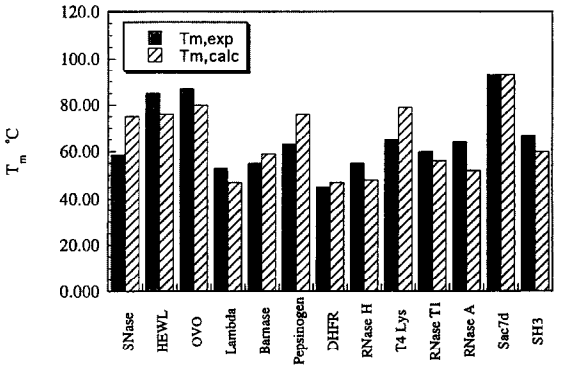

Fig. 1. Comparison between predicted and experimental T_m values for thirteen proteins at their pH of maximal stability. Staphyloccocal nuclease, pH 7 [2,34]; hen egg white lysozyme, pH 5 [35]; turkey ovomucoid third domain, pH 5 [20]; lambda repressor, pH 7 (T. Oas, personal communication); barnase, pH 6 [36,37]; pepsinogen, pH 6 [38]; cysteine-free dihydrofolate reductase, pH 7 (S. Pucciarelli, unpublished data from this laboratory); cysteine free ribonuclease H, pH 8 [39]; T4 lysozyme, pH 5 [40]; ribonuclease T1, pH 5 [41]; ribonuclease A, pH 6 [42]; Sac7d, pH 6 [43]; SH3 domain, pH 5 [44].

2.3. The COREX algorithm

The ensemble of partially folded states of a protein is approximated with the computer by using the high resolution structure as a template. In the COREX algorithm [8,9] the entire protein is considered as being composed of different folding units. Partially folded states are generated by folding and unfolding these units in all possible combinations.

The division of the protein into a given number of folding units is called a partition. In order to maximize the number of distinct partially folded states, different partitions are included in the analysis. Each partition is defined by placing a block of windows over the entire sequence of the protein. The folding units are defined by the location of the windows irrespective of whether or not they coincide with specific secondary structure elements. By sliding the entire block of windows one residue at a time, different partitions of the protein are obtained. For two consecutive partitions the first and last amino acids of each folding unit are shifted by one residue. This procedure is repeated until the entire set of partitions have been exhausted. For a protein like T4 lysozyme of 162 residues, a window size of 12

residues was used, resulting in a total of 188,394 different states ranging from the completely unfolded to the native.

Each of the states generated by the COREX algorithm is characterized by having some regions folded and some other regions unfolded. There are two basic assumptions in this algorithm: 1) the folded regions in partially folded states are native-like; and, 2) the unfolded regions are assumed to be devoid of structure. The thermodynamic quantities (ΔH, ΔS, ΔC_p, and ΔG) for each state as well as the partition function and probability (Eqs. (1) and (2)) are calculated using the structural parameterization of the folding energetics described above.

2.4. Statistical descriptors of conformational equilibrium

Experimental observables are by definition statistical averages over the ensemble of conformations accessible to a protein. For an arbitrary observable, X, the ensemble average $\langle X \rangle$ is defined by the contributions of all states accessible to the protein weighted according to their probabilities

$$\langle X \rangle = \sum_{i=0}^{n} X_i P_i \tag{6}$$

Global averages like the excess enthalpy function or its temperature derivative, the excess heat capacity function, which are measured by differential scanning calorimetry [45,46] are derived from expressions similar to Eq. (6). In general, these global averages provide low resolution information about the nature of the ensemble of protein conformations under a particular set of conditions. It has been shown, for example, that the widely used ratio of van't Hoff to calorimetric enthalpies ($\Delta H_{vh}/\Delta H_{cal}$) is an extremely poor indicator of the number of states that become populated in a folding/unfolding transition [47].

Recently, significant experimental advances that allow determination of ensemble averages for individual residues have been made. In particular, the technique of NMR-detected hydrogen exchange has been shown to be able to provide structural and thermodynamic information at the residue level [10–20]. The experimental results, usually expressed as hydrogen exchange protection factors per residue, contain information with which to probe the nature of the ensemble of protein conformations under a particular set of conditions [8,9]. Recently, different statistical descriptors aimed at describing the state of individual residues in terms of the ensemble of protein conformations have been introduced [8,9]. Two particularly important ones are described below.

2.4.1. Folding probabilities per residue

The probability that a given residue j is folded, $P_{f,j}$, is equal to the sum of the probabilities of all the conformational states of the protein in which that particular residue is folded

$$P_{f,j} = \sum_{\text{(states with residue } j \text{ folded)}} P_i \tag{7}$$

2.4.2. The apparent folding constant per residue

The apparent folding constant or apparent stability constant per residue, $\kappa_{f,j}$, is defined as the ratio of the probabilities of all states in which residue j is folded ($P_{f,j}$) to the probabilities of the states in which residue j is not folded

$$\kappa_{f,j} = \frac{P_{f,j}}{P_{nf,j}} \tag{8}$$

The corresponding apparent free energy is simply $\Delta G_{f,j} = -RT\ln\ \kappa_{f,j}$. The apparent folding constant per residue, $\kappa_{f,j}$, is the quantity that one would measure if it were possible to determine experimentally the stability of the protein by monitoring each individual residue. This quantity is important because it can be correlated with experimental data, especially hydrogen exchange protection factors measured by NMR, which under certain conditions are able to measure $\kappa_{f,j}$.

2.5. Hydrogen exchange protection factors

The apparent folding constants per residue, $\kappa_{f,j}$, are the quantities that one would measure if it were possible to study the folding/unfolding equilibrium using every single residue in the protein as the experimental probe. Of all the techniques available today, NMR-detected hydrogen exchange is the one that comes the closest to that implementation. Hy-

drogen exchange experiments under the so-called EX2 regime measure the equilibrium constant for the following reaction (see for example Refs. [17,48])

$$NH_j(\text{closed}) \overset{K_{op,j}}{\leftrightarrow} NH_j(\text{open}) \quad (9)$$

The protection factors, PF_j, usually reported in the literature are equal to the inverse of the equilibrium constants $K_{op,j}$. The standard interpretation of reaction Eq. (9) is that slowly exchanging protons exchange with the solvent only after becoming exposed to it as a result of local, partial or global unfolding. For this reason, protection factors provide a local probe of the folding/unfolding equilibrium. It must be noted, however, that in partially folded states not only those residues that are unfolded become exposed to the solvent. Those residues located in the so-called complementary regions also become exposed, i.e. those residues located in portions of the protein that remain folded but were structurally complementary to the regions of the protein that became unfolded [49].

While the residue stability constants are purely thermodynamic quantities defined for all residues, the protection factors also contain non-thermodynamic contributions and are defined only for a subset of residues. From a statistical standpoint, the protection factor for any given residue j can be defined as the ratio of the sum of the probabilities of the states in which residue j is closed, to the sum of the probabilities of the states in which residue j is open

$$PF_j = \frac{\sum_{(\text{states with residue } j \text{ closed})} P_i}{\sum_{(\text{states with residue } j \text{ open})} P_i} = \frac{P_{\text{closed},j}}{P_{\text{open},j}} \quad (10)$$

The statistical definition of the protection factors has the same form as that of the stability constants and can be expressed in terms of the folding probabilities, as follows

$$PF_j = \frac{P_{f,j} - P_{f,xc,j}}{P_{nf,j} + P_{f,xc,j}} \quad (11)$$

where the correction term $P_{f,xc,j}$ is the sum of the probabilities of all states in which residue j is folded, yet exchange-competent. It is evident that the hydrogen exchange protection factors, PF_j, are equal to the stability constants per residue, $\kappa_{f,j}$, only when the $P_{f,xc,j}$ terms are small. The most common situations in which a residue is folded but exposed to the solvent occurs when: 1) the amide group of the residue is exposed in the native state; and, 2) the amide group of the residue becomes exposed by being located in a region of the protein that is structurally complementary to an unfolded region. Proline residues lack an amide group and therefore are not included in the exchange reaction.

2.6. Cooperativity

The protein folding reaction is cooperative, i.e. the vast majority of partially folded conformations never becomes populated; and, under most conditions, the population of those partially folded conformations that do become populated is very small. Traditionally, the cooperativity of folding/unfolding reactions has been evaluated in terms of the so-called van't Hoff to calorimetric enthalpy ratio ($\Delta H_{vH}/\Delta H_{cal}$); however this test is very insensitive [47] and lacks structural resolution. A better alternative involves the use of hydrogen exchange protection factors and their dependence on a global variable like temperature or denaturants [17,20]. The ideal alternative, however, will involve the apparent folding constants per residue and their response to a local rather than a global perturbation. Such a local perturbation can be a single amino acid mutation and the response, the propagation of the effect to the stability constants of other residues. Ideally, the mutation should not perturb the structure of the protein or the interactions among other residues, i.e. the more localized the perturbation the better. In some cases, this type of local perturbation can also be induced by a specific ligand. The scenario presented here is experimentally feasible in the sense that hydrogen exchange measurements should provide experimental estimation of critical residue stability constants.

For the computational analysis in this paper we have chosen two well characterized pairs of Ala and Gly mutations at solvent-exposed α-helix locations in T4 lysozyme (S44A, S44G, V131A, V131G). These mutations have been shown to cause a negligible structural perturbation [50] and to affect the global stability of T4 lysozyme by an amount that

largely corresponds to the intrinsic α-helix propensities of the amino acids under consideration [6,26,50].

3. Results and discussion

3.1. Virtual differential scanning calorimetry (VDSC) of alanine and glycine mutants of T4 lysozyme

Fig. 2 shows the molar heat capacity as a function of temperature generated by the program VDSC under conditions in which the thermal transition obeys two-state behavior for the two pairs of Ala → Gly mutants S44A, S44G, and V131A, V131G of T4 Lysozyme. These mutants differ by one amino acid substitution at a central position in solvent-exposed α-helices [50]. As expected, the Ala mutants are predicted to be more stable than the Gly mutants. The coincidence of the heat capacity changes in the VDSC scans of each mutant is an indication that the two sets of molecules exhibit similar changes in solvent accessibilities upon unfolding as expected for an amino acid substitution at a solvent-exposed location.

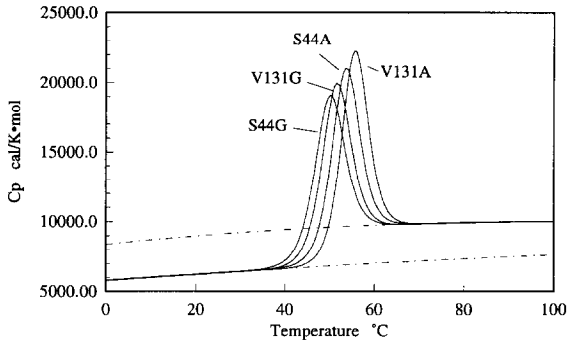

Fig. 2. The heat capacity function versus temperature for four mutants of T4 lysozyme under conditions in which the folding/unfolding obeys two-state behavior. Shown as dotted lines are the heat capacities of the native and unfolded states. The curves were generated by the computer program VDSC which implements the structural parameterization of the energetics discussed here. VDSC directly reads a PDB file and generates the structure-based thermodynamic analysis. The simulated conditions in the figure correspond to pH 3.2. For these calculations, the structures of the mutants were derived from that of the pseudo-wild-type C54T/C97A structure which is the template used for the experimental Ala and Gly mutations. These mutations are at solvent-exposed locations (see Fig. 4) and have been shown to have essentially superimposable structures [50].

At 25°C, the differences in Gibbs energies between the Ala and Gly mutants are 1.0 and 1.2 kcal mol^{-1} which is close to what is expected from the difference in backbone conformational entropies between Gly and Ala mutants (0.7 kcal mol^{-1}) indicating that most of the differences can be attributed to this effect. The reported experimental differences in Gibbs energies between the Ala and Gly mutants are 0.96 and 0.94 kcal mol^{-1}, in close agreement with the values calculated from structure [50]. Inspection of the values in Table 1 reveals the existence of small differences in ΔH and in the entropy of solvation. However, these effects cancel each other to a large extent and contribute little to the final ΔG. This observation illustrates the phenomenon of enthalpy–entropy compensation reported previously for this protein [52].

The structure-based thermodynamic parameters are summarized in Table 1 and agree well with experimentally determined values. Experimental ΔC_p values are about 2.5 kcal K^{-1} mol^{-1} [52,53] compared to the calculated value of about 2.65 kcal K^{-1} mol^{-1}. The bulk of the enthalpy change, excluding protonation effects, is about 120 kcal mol^{-1} at 60°C as indicated in Table 1. The experimental values, which include protonation effects, are about 130 kcal mol^{-1}, which is close to the predicted value.

It has been shown before that protonation effects contribute significantly to the stability of T4 lysozyme. In particular, a single salt bridge between His31 and Asp70 has been shown to contribute around 3–5 kcal mol^{-1} to the stability of the molecule [54]. His31 exhibits an anomalous pK of 9.1 and Asp70 a pK of 0.5 in the native state. At pH 3.2, the overall excess protonation contribution to the Gibbs energy is destabilizing and close to -9.8 kcal mol^{-1} (calculated by Dr. B. Garcia-Moreno, using the methods described in Ref. [51]). At this pH, four additional protons bind to the protein molecule upon unfolding [52].

3.2. Apparent folding constants per residue

Fig. 3, panels A and B, show the residue stability constants for the Ala and Gly mutations at position 44 and 131 respectively. Several important features can be distinguished in these figures. In general, the overall pattern of residue stability constants is simi-

lar for the four mutants. Under the conditions of these calculations, it is clear that the C terminal domain (residues 1–12, 72–164) has higher stability constants than the N terminal domain (residues 13–71). This situation might vary under different conditions since a number of electrostatic interactions (e.g. His31, Asp70) contribute significantly to the stability of the N domain and to the cooperative coupling of the stability between the two domains. According to the predicted stability map, several peaks and valleys are present. The highest stability constants correspond to residues at the beginning of helix 5 (residues 93–106) which incidentally are also among the residues that show the highest protection in early folding intermediates of T4 lysozyme [55].

A notable feature of the pattern of stability is that in most cases the stabilility constants for residues within a given helix are not the same. Helices may show fraying at the ends and in the case of long helices, they might exhibit regional variations. This is especially evident for helix 3 which is the largest of the helices in T4 lysozyme (residues 60–80). Helix 3 is unique in that it spans across both structural domains; the first 12 residues belonging to the N terminal domain and the last 8 residue belonging to the C terminal domain. Fig. 3 shows that the N terminal half of helix 3 is significantly less stable than the C terminal half and underscores the notion that individual secondary structural elements do not obligatorily correspond to cooperative folding units.

Table 1
Summary of thermodynamic parameters for alanine to glycine mutants of T4 lysozyme

	S44A	S44G	V131A	V131G
T_m [a]	53.5	50.0	55.5	51.5
$\Delta H_{gen}(60)$/kcal mol^{-1}	121.1	121.4	121.5	121.7
$\Delta H_{gen}(25)$/kcal mol^{-1}	28.6	29.4	28.7	29.4
ΔC_p/cal K^{-1} mol^{-1}	2.6	2.6	2.67	2.6
ΔS_{conf}/cal K^{-1} mol^{-1}	930.9	933.3	931.9	934.1
$\Delta S_{solv}(25)$/cal K^{-1} mol^{-1}	−886.7	−883.2	−889.4	−885.2
$\Delta S(25)$/cal K^{-1} mol^{-1}	44.2	50.1	42.5	48.9
ΔG^a_{prot}/kcal mol^{-1}	−9.8	−9.8	−9.8	−9.8
$\Delta G(25)$/kcal mol^{-1}	5.6	4.6	6.2	5.0

[a] Predicted data are for pH 3.2. The excess protonation free energy was calculated by Dr. García-Moreno using the methods described in Ref. [51]. All thermodynamic values have been rounded to one decimal place.

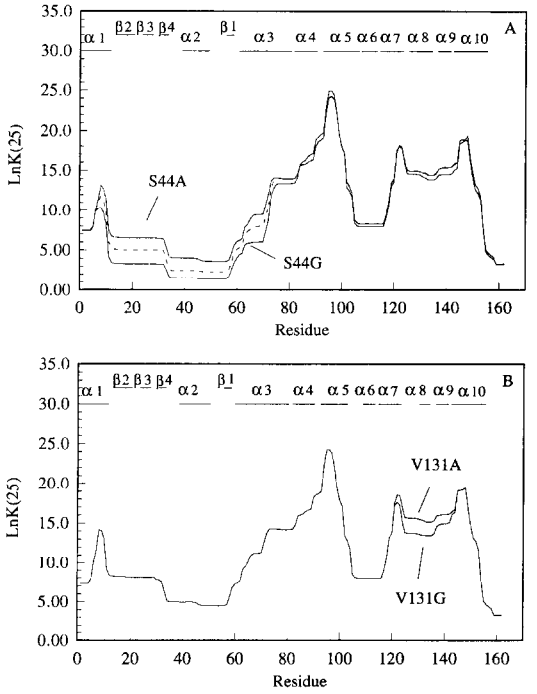

Fig. 3. Panel A. The natural logarithm of the predicted folding constant per residue (ln κ_f) for each amino acid residue of the mutants at position 44 (S44A and S44G). The solid lines represent the results of the calculations on the experimentally determined crystallographic structures of the two mutants (PDB files 1L68 and 107L). The dotted line represents the results of the calculations on the structure of S44G obtained by replacing Ala by Gly in the structure of S44A. Panel B. The natural logarithm of the predicted folding constant per residue (ln κ_f) for each amino acid residue of the mutants at position 131 (V131A and V131G). Since crystallographic structures are not available for these mutants, the structures were generated from the structure of the pseudo-wild-type protein (PDB file 1L63). The calculations were made under simulated conditions that maximize stability in order to optimize the resolution in the pattern of folding constants.

A second observation is that the C terminal domain (residues 1–12, 72–164) appears to have two distinct groupings of residues with high stability constants (72–100 and 119–164) which are separated by a region of lesser stability (101–118). These three groupings of residues correspond to contiguous clusters within the C terminal domain. This feature becomes important when the effects of the Ala to Gly mutations are analyzed.

3.3. Propagation and extent of cooperative interactions

It is clear in Fig. 3A and B that the effects of the mutations are not restricted to the mutated sites. The effect of each mutation propagates along the sequence and along tertiary interactions. This is better appreciated in Fig. 4 (panels A and B) in which the extent of the effect of each mutation is shown on the structure of the protein. The mutation at position 44 in particular, affects the N terminal domain (residues 13–71) in its entirety. This effect reflects an intrinsic property of this domain and not a structural perturbation elicited by the mutation, as illustrated by the calculations performed with the crystallographic structure and with the structure obtained by simply replacing Ala 44 by a Gly with the computer. The Gly and Ala mutants at position 44 show extremely small structural differences (0.14 Å backbone RMS) and are practically superimposable [50]. It is also evident that the effect of the mutation is localized to the N terminal domain and is not transmitted to the C terminal domain (residues 1–12, 72–164). The mutation at position 131, however, induces a more restricted effect limited primarily to a contiguous region between residues 120 and 146. These results indicate that the intrinsic cooperativity of the C terminal domain is not as complete as that of the N terminal domain.

Since the hydrogen exchange protection factors track the stability constants per residue, it is expected that experimentally determined protection factors for those residues highlighted in Fig. 4A and B will be affected by the mutations at position 44 and 131 respectively. In this respect, Anderson et al. [57] measured the base-catalyzed hydrogen exchange for the mutants A98V, A146T, Y25G and T157I and found that the effects of the mutations extended beyond the mutation site in sequence and space but were restricted to their respective domains.

Fig. 5 illustrates the effects of the Ala and Gly mutations at position 44 by showing the predicted hydrogen exchange protection factors for the N terminal domain of S44A and S44G. The predicted protection factors were calculated according to Eq. (11). The protection factors in the C terminal domain remain essentially unchanged and are not shown in the figure. It is clear that the protection factors for

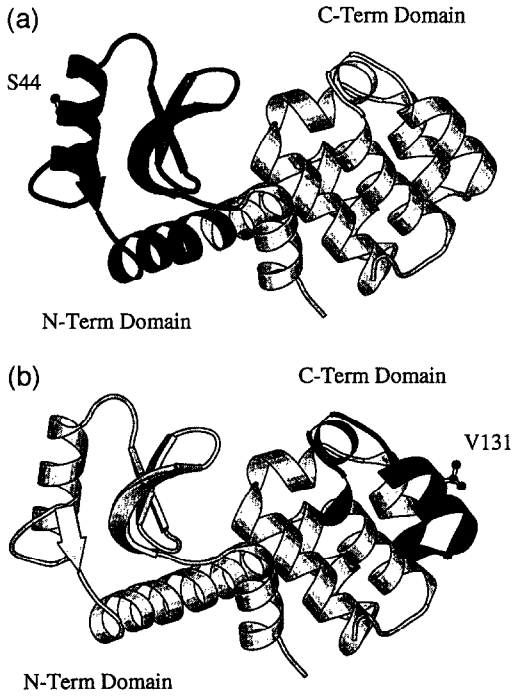

Fig. 4. Schematic representation of the extent of the effects of Ala and Gly mutations at position 44 (panel A) and at position 131 (panel B). Regions shaded in black are expected to have higher residue stability constants and hydrogen exchange protection factors in the Ala mutants. This figure illustrates that the effects of mutations are not restricted to the mutated amino acid even in the absence of structural perturbations, and that they can be used to map the extent of cooperative interactions. The figures were generated by the program MOLSCRIPT [56].

S44G are lower than those for S44A throughout the entire N terminal domain. As discussed before, not all amino acids exhibit protection from exchange even if they are characterized by significant stability constants [8,9]. Lack of protection is observed if the amide group of a residue is exposed to the solvent in the native state. Proline residues do not exhibit protection. The amide groups of some residues exhibit a finite probability of becoming solvent-exposed by being adjacent to regions with high unfolding probabilities. In addition, and depending on the technique and conditions under which protection is measured, exchange rates faster than a certain limiting value cannot be measured and their protection factors cannot be evaluated.

3.4. Structural mapping of cooperative effect

The origin of the propagation of the Gibbs energy along sequence and three-dimensional space can be illustrated with the following example. Consider the situation in Fig. 6 in which two regions of a protein (1 and 2) interact with each other. The sub-partition function for this region can be written as

$$Q_{sub} = 1 + \phi K_1 + \phi K_2 + \phi K_1 K_2 \quad (12)$$

where K_1 and K_2 are the intrinsic statistical weights of regions 1 and 2, and ϕ is the interaction term [58]. Let us consider a residue in region 1 (residue A) and a residue in region 2 (residue B). The apparent folding constants for these residues are obtained by applying Eq. (8)

$$\kappa_{f,A} = \frac{1 + \phi K_2}{\phi K_1 + \phi K_1 K_2} \quad (13a)$$

$$\kappa_{f,B} = \frac{1 + \phi K_1}{\phi K_2 + \phi K_1 K_2} \quad (13b)$$

It is clear from the above equations that a stabilizing or destabilizing mutation in region 2 will affect $\kappa_{f,A}$ even if the mutation does not directly affect region 1 or the interaction term ϕ. A similar situation occurs with $\kappa_{f,B}$ and a mutation in region 1. The propagation of the effect is mediated through the cooperative interaction parameter ϕ. If the two regions were independent of each other, i.e. $\phi = 1$, then the value of $\kappa_{f,A}$ will be completely independent of mutations affecting region 2, and similarly, the value of $\kappa_{f,B}$ will be independent of mutations in region 1. It follows directly from Eq. (13a) and Eq. (13b) that if $\phi = 1$ then $\kappa_{f,A} = K_1^{-1}$ and $\kappa_{f,B} = K_2^{-1}$. If cooperative interactions are present ($\phi \neq 1$) then the extent of the propagation of the protein response to the mutation effectively maps the cooperative region regardless of whether the mutation results in a structural perturbation of the protein.

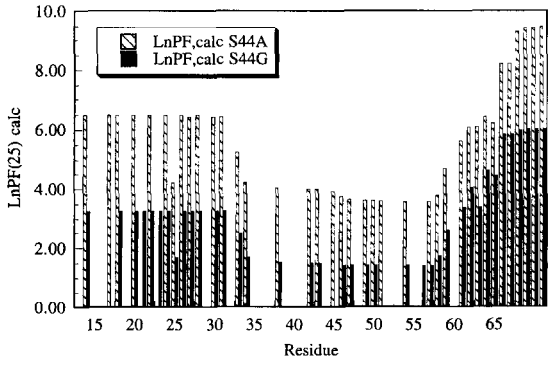

Fig. 5. Predicted hydrogen exchange protection factors for the N terminal domain (residues 13-71) of S44A (open bars) and S44G (dark bars). The protection factors for the Ala mutant are higher than those of the Gly mutant throughout the entire domain, consistent with the existence of cooperative interactions that involve the entire domain.

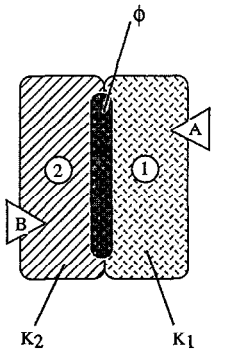

Fig. 6. Schematic illustration of two regions (1 and 2) of a hypothetical protein. Shown in the figure are two residues (A and B) far away from the interface region. It is demonstrated in the text that a stabilizing or destabilizing mutation in residue A will affect the apparent folding constants or hydrogen exchange protection factors of residues in region 2 provided that the cooperative interaction energy between regions ($-RT\ln\phi$) is not equal to zero. A similar situation occurs for residue B and region 1. This analysis demonstrates that the propagation of the effect of a mutation is mediated by the existence of cooperative interactions suggesting that mutations can be used to map the extent of cooperativity along the sequence and three-dimensional space.

4. Conclusions

The formalism discussed in this and other papers of this series [8,9] provides a link between experimental observables and statistical thermodynamics that permits interpretation and analysis of protein folding within a unified framework. In the past, statistical thermodynamic models of protein folding have provided oversimplified pictures of reality and have lacked explicit connections to experiment. Hence, data analysis has been usually performed in terms of simple chemical models. The results presented here suggest that the high resolution structure of a protein can be used to probe the nature of the ensemble of conformations found under equilibrium conditions, to evaluate and map cooperative interac-

tions within a protein, and to verify experimentally the analysis by NMR-detected hydrogen exchange experiments.

References

[1] D. Xie and E. Freire, Proteins: Struct., Func. and Genetics, 19 (1994) 291–301.
[2] D. Xie, R. Fox and E. Freire, Protein Sci., 3 (1994) 2175–2184.
[3] J. Gomez, J.V. Hilser, D. Xie and E. Freire, Proteins: Structure, Function and Genetics, 22 (1995) 404–412.
[4] J.A. D'Aquino, J. Gómez, V.J. Hilser, K.H. Lee, L.M. Amzel and E. Freire, Proteins, 25 (1996) 143–156.
[5] V.J. Hilser, J., Gomez and E. Freire, Proteins, 26 (1996) 123–133.
[6] I. Luque, O. Mayorga and E. Freire, Biochemistry, 35 (1996) 13681–13688.
[7] D. Xie and E. Freire, J. Mol. Biol., 242 (1994) 62–80.
[8] V.J. Hilser and E. Freire, J. Mol. Biol., (1996), in press.
[9] V.J. Hilser and E. Freire, Proteins, 262 (1996) 756–772.
[10] J.B. Udgaonkar and R.L. Baldwin, Nature, 335 (1988) 694–699.
[11] H. Roder, G.A. Elove and S.W. Englander, Nature, 335 (1988) 700–704.
[12] S.N. Loh, K.E. Prehoda, J. Wang and J.L. Markley, Biochemistry, 32 (1993) 11022–11028.
[13] K.-S. Kim, J.A. Fuchs and C.K. Woodward, Biochemistry, 32 (1993) 9600–9608.
[14] K.-S. Kim and C. Woodward, Biochemistry, 32 (1993) 9609–9613.
[15] P.A. Jennings and P.E. Wright, Science, 262 (1993) 892–896.
[16] M.D. Jacobs and R.O. Fox, Proc. Natl. Acad. Sci. USA, 91 (1994) 449–453.
[17] Y. Bai, T.R. Sosnick, L. Mayne and S.W. Englander, Science, 269 (1995) 192–197.
[18] L.A. Morozova, D.T. Haynie, C. Arico-Muendel, H. Van Dael and C.M. Dobson, Nature Structural Biology, 2 (1995) 871–875.
[19] B.A. Schulman, C. Redfield, Z. Peng, C.M. Dobson and P.S. Kim, J. Mol. Biol., 253 (1995) 651–657.
[20] L. Swint-Kruse and A.D. Robertson, Biochemistry, 35 (1996) 171–180.
[21] K.P. Murphy, V. Bhakuni, D. Xie and E. Freire, J. Mol. Biol., 227 (1992) 293–306.
[22] J. Gomez and E. Freire, J. Mol. Biol., 252 (1995) 337–350.
[23] S.M. Habermann and K.P. Murphy, Protien Sci., 5 (1996) 1229–1239.
[24] R.L. Baldwin, Proc. Natl. Acad. Sci. USA, 83 (1986) 8069–8072.
[25] K.P. Murphy and E. Freire, Adv. Protein Chem., 43 (1992) 313–361.
[26] K.H. Lee, D. Xie, E. Freire and L.M. Amzel, Proteins Struct. Func. and Genetics, 20 (1994) 68–84.
[27] C.N. Pace, G.R. Grinsley, J.A. Thomson and B. Barnett, J. Biol. Chem., 263 (1988) 11820–25.
[28] B. Lee and F.M. Richards, J. Mol. Biol., 55 (1971) 379–400.
[29] M. Blaber, X.-J. Zhang and B.W. Matthews, Science, 60 (1993) 1637–1640.
[30] A. Horovitz, J.M. Matthews and A.R. Fersht, J. Mol. Biol., 227 (1992) 560–568.
[31] K. O'Neil, and W. DeGrado, Science, 250 (1990) 646–651.
[32] P.C. Lyu, M.I. Liff, L.A. Marky and N.R. Kallenbach, Science, 250 (1990) 669–673.
[33] V. Muñoz and L. Serrano, J. Mol. Biol., 245 (1995) 275–296.
[34] J.H. Carra, E.A. Anderson and P.L. Privalov, Protein Sci., 3 (1994) 944–951.
[35] W. Pfeil and P.L. Privalov, Biophys. Chem., 4 (1976) 23–32.
[36] J.C. Martinez, M.E. Harrous, V.V. Filimonov, P.L. Mateo and A.R. Fersht, Biochemistry, 33 (1994) 3919–3926.
[37] Y. Griko, G.I. Makhatadze, P.L. Privalov and R.W. Hartley, Protein Sci., 3 (1994) 669–676.
[38] P.L. Mateo and P.L. Privalov, FEBS Lett., 123 (1981) 189–192.
[39] J.M. Dabora and S. Marqusee, Protein Sci., 3 (1994) 1401–1408.
[40] T. Alber, S. Dao-Pin, K. Wilson, J.A. Wozniak, S.P. Cook and B.W. Matthews, Nature, 330 (1987) 41–46.
[41] I.M. Plaza del Pino, C.N. Pace and E. Freire, Biochemistry, 31 (1992) 11196–202.
[42] P.L. Privalov, E.I. Tiktopulo and N.N. Khechinashvili, Int. J. Pept. Protein Res., 5 (1973) 229–237.
[43] J.G. McAfee, S.P. Edmondson, P.K. Datta, J.W. Shriver and R. Gupta, Biochemistry, 34 (1995) 10063–77.
[44] A.R. Viguera, J.C. Martinez, V.V. Filimonov, P.L. Mateo and L. Serrano, Biochemistry, 33 (1994) 2142–2150.
[45] E. Freire, Methods in Enzymology, 240 (1994) 502–568.
[46] E. Freire, Methods in Enzymology, 259 (1995) 144–168.
[47] E. Freire, in B. Shirley (Ed.), Protein Stability and Folding, Vol. 40, 1995, pp. 191–218.
[48] A. Hvidt and S.O. Nielsen, Adv. Prot. Chem., 21 (1966) 287–386.
[49] E. Freire, D.T. Haynie and D. Xie, Proteins: Structure, Function and Genetics, 17 (1993) 111–123.
[50] M. Blaber, X.-J. Zhang, J.L. Lindstrom, S.D. Pepiot, W.A. Baase and B.W. Matthews, J. Mol. Biol., 235 (1994) 600–624.
[51] B.E. García-Moreno, Methods Enzymol., 259 (1995) 512–538.
[52] P. Connelly, L. Ghosaini, C.Q. Hu, S. Kitamura, A. Tanaka and J.M. Sturtevant, Biochemistry, 30 (1991) 1887–1891.
[53] J.E. Ladbury, C.Q. Hu and J.M. Sturtevant, Biochemistry, 31 (1992) 10699–702.
[54] D.E. Anderson, W.J. Becktel and F.W. Dahlquist, Biochemistry, 29 (1990) 2403–2408.
[55] J. Lu and F.W. Dahlquist, Biochemistry, 31 (1992) 4749–4756.
[56] P.J. Kraulis, J. Appl. Crystallogr., 24 ((1991) 946–950.
[57] D.E. Anderson, J. Lu, L. McIntosh and F.W. Dahlquist, in M.C.a.A. Gronenborn (Ed.), NMR of Proteins, Macmillan Press, New York, pp. 258–304.
[58] E. Freire and K.P. Murphy, J. Mol. Biol., 222 (1991) 687–698.