

The inherent structure landscape of a protein

Naoko Nakagawa, and Michel Peyrard

PNAS 2006;103;5279-5284; originally published online Mar 27, 2006; doi:10.1073/pnas.0600102103

This information is current as of November 2006.

Online Information & Services	High-resolution figures, a citation map, links to PubMed and Google Scholar, etc., can be found at: www.pnas.org/cgi/content/full/103/14/5279
References	This article cites 22 articles, 7 of which you can access for free at: www.pnas.org/cgi/content/full/103/14/5279#BIBL
	This article has been cited by other articles: www.pnas.org/cgi/content/full/103/14/5279#otherarticles
E-mail Alerts	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.
Rights & Permissions	To reproduce this article in part (figures, tables) or in entirety, see: www.pnas.org/misc/rightperm.shtml
Reprints	To order reprints, see: www.pnas.org/misc/reprints.shtml

Notes:

The inherent structure landscape of a protein

Naoko Nakagawa*^{†‡} and Michel Peyrard[†]

*Department of Mathematical Sciences, Ibaraki University, Mito, Ibaraki 310-8512, Japan; and [†]Laboratoire de Physique, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

Communicated by Hans Frauenfelder, Los Alamos National Laboratory, Los Alamos, NM, January 8, 2006 (received for review April 14, 2005)

Using the Gō model of a real protein, we explore the landscape of its metastable structures. First, we show how the inherent structure energy density can be obtained from the probability density determined by sampling molecular dynamics trajectories and quenching. The analysis of the inherent structure landscape can characterize the folding transition. Then we show how thermodynamics of the inherent states can be established to study the equilibrium properties of proteins. Our work brings some elements into the current discussion about the protein dynamical transition. The study uses a simplified model to illustrate the ideas, but, as the inherent structure landscape is much simpler than the free energy surface of the protein, it appears to be accessible for an all-atom model of a small protein, at the expense of much longer calculations.

biomolecule | protein dynamical transition | protein folding | statistical physics

he concept of energy landscape (1, 2) is very important for proteins because it provides a unifying viewpoint on the structure of the proteins, their ability to fold, and their fluctuations, which are essential to function. The energy landscape of a protein is itself a very complex object so that, except for simple lattice models of proteins (3), it cannot be explicitly obtained. But, do we really need to know this landscape? In this article we argue that the answer is no, or more precisely that a partial view of this landscape can already tell us a lot about a protein. The example of glass-forming materials, which are also very complex, shows that many structural and thermodynamic properties can be deduced from the knowledge of a very small subset of the configuration space, the inherent structures, i.e., the local minima of the energy landscape (4, 5). Here we show how the inherent structure landscape (ISL) can be determined for the model of a real protein, and that it provides a fairly complete description of the thermodynamics of the protein.

The key result needed to establish reduced thermodynamics from the ISL is the corresponding density of states. Previous investigations have pointed out the importance of the general features of the density of energy states in determining the properties of proteins (6, 7). These studies show that the most significant properties of proteins can be obtained from a suitably built random energy model, based on a stochastic Hamiltonian. In this context protein folding has been formulated in terms of a discrete set of microstates, which could be taken as the minima of a continuously defined potential energy surface (7). The spirit of ISL-based thermodynamics is similar, but, in addition we show here how to build a density of microstates that is based on the structure of an actual protein. The framework that we introduce with a simple model to reduce computation time could be used to analyze any model, including models with an atomic description. The calculations would, of course, become much heavier, but they appear feasible with the computing facilities available today.

Exploring the ISL of a protein is not a new idea, but, to our knowledge, up to now this exploration has been restricted to its structural properties, for a model heteropolymer with 46 beads (8). We consider here a 56-aa real protein, and even with a very simple model, it implies that the number of inherent structures is increased by more than four orders of magnitude. An exhaustive search is not possible, but we show that a statistical physics analysis can be used to extract the necessary information from the phase space trajectories of the thermalized protein. Then we compare the thermodynamics over the ISL to the full thermodynamics of the protein.

We selected protein G (9) for this study because it is rather small (56 residues) but nevertheless includes typical secondary structure elements of proteins, one α -helix and two β -strands. We use molecular dynamics at constrained temperature to let the protein explore its free energy landscape, and applying rapid quenchings, i.e., calculating a steepest descent path from different points of this trajectory, we sample the underlying inherent structures. Although the characterization of the ISL is much simpler than a full determination of the energy landscape, calculations with an all-atom model of the protein would require a large computer power. Here, to illustrate the ideas, we use an off-lattice Gō model (10), which is rather simple but nevertheless accurate enough to reproduce experimental properties of the transition state for several proteins (11).

Density of States for the Inherent Structures

Molecular dynamics simulations show that the model has the expected qualitative properties of a good folder: at low temperature it folds to the global minimum of the potential energy, whereas at high temperature it denaturates. If the model is cooled down quickly from a high-temperature unfolded state to a temperature well below its folding temperature (T_f) , the folding process is, however, very slow, and may never reach a shape close to the native shape within the time scale accessible in a simulation. This finding suggests that this model protein exhibits the glassy properties observed in actual proteins. The folding transition can be detected from the temperature variation of the specific heat C_{ν} . As shown in Fig. 1, a characteristic peak of C_{ν} appears at a temperature that we define as T_{f} . The ISL is determined by the basins of the potential energy surface, labeled by an index α , which are defined as the set of conformations that are connected to the same local energy minimum, called the inherent structure, which has an energy e_{α} .

The rapid change of the shape of the curve showing the probability density of inherent structures P_{IS} versus their energy e_{α} in the vicinity of T_f confirms this interpretation: below T_f , $P_{IS}(e_{\alpha}, T)$ is higher for the low values of e_{α} . At T_f it shows two maxima, one for low energies corresponding to folded states, and one for high-energy, unfolded states. Above T_f , $P_{IS}(e_{\alpha}, T)$ is dominated by the high-energy, unfolded states. To characterize the ISL we need not only the values of the energies e_{α} but also their density of states $\Omega_{IS}(e_{\alpha})$. A realistic protein has a huge number of inherent states so that $\Omega_{IS}(e_{\alpha})$ cannot be determined from a systematic search. Instead it can be deduced from the sampling of the inherent structure α , the potential energy of the protein can be written as $V(\mathbf{r}) = e_{\alpha} + \Delta V(\mathbf{r})$, where \mathbf{r} designates all of the degrees of freedom of the protein. Let us assume that

Conflict of interest statement: No conflicts declared.

Abbreviations: ISL, inherent structure landscape; *T*_f, folding temperature.

[‡]To whom correspondence should be addressed. E-mail: nah@mx.ibaraki.ac.jp.

^{© 2006} by The National Academy of Sciences of the USA



Fig. 1. Temperature dependence of specific heat C_v and probability distribution of the inherent-state energies, $P_{IS}(e_{\alpha}, T)$ at three temperatures in the vicinity of the T_f .

the inherent structures can be split into a set of discrete states α_0 , α_1 , ..., α_K (which includes at least one state, the ground state) and a continuum of higher-energy states. With this notation, the configurational contribution of the partition function of the protein is

$$Z(T) = \sum_{\alpha_i = \alpha_0}^{\alpha_k} e^{-\beta e_{\alpha_i}} e^{-\beta F_{\nu}(\alpha_i, T)} + \int_{e_{\alpha_k}}^{e_{\max}} \Omega_{IS}(e_{\alpha}) e^{-\beta e_{\alpha}} e^{-\beta F_{\nu}(\alpha, T)} de_{\alpha},$$
[1]

where e_{\max} is the highest inherent structure energy. Moreover we have introduced

$$e^{-\beta F_{\nu}(\alpha,T)} = \int_{B(\alpha)} e^{-\beta \Delta V(\mathbf{r})} d\mathbf{r},$$
 [2]

which is the vibrational free energy in the basin of attraction $B(\alpha)$ of the inherent structure α . The probability to be in the basin of attraction of a discrete inherent structure is

$$p_{\alpha_i}(T) = \frac{1}{Z(T)} \exp(-\beta e_{\alpha_i}) \exp[-\beta F_{\nu}(\alpha, T)], \qquad [3]$$

and, for the continuum of inherent structures, the probability density of inherent structures is

$$P_{IS}(e_{\alpha}, T) = \frac{1}{Z(T)} \Omega_{IS}(e_{\alpha}) \exp(-\beta e_{\alpha}) \exp[-\beta F_{\nu}(\alpha, T)].$$
[4]

If we can assume that the free energy $F_{\nu}(\alpha, T)$ does not depend on the inherent structure, in Z the term $\exp[-\beta F_{\nu}(T)]$ can be factorized so that p_{α_i} and P_{IS} simplify into

$$p_{\alpha_i}(T) = \frac{e^{-\beta e_{\alpha_i}}}{Z_{IS}(T)}, \quad P_{IS}(e_{\alpha}, T) = \frac{1}{Z_{IS}(T)} \Omega_{IS}(e_{\alpha}) e^{-\beta e_{\alpha}}$$
 [5]

with

$$Z_{IS}(T) = \sum_{\alpha_i = \alpha_0}^{\alpha_K} e^{-\beta e_{\alpha_i}} + \int_{e_{\alpha_K}}^{e_{\max}} \Omega_{IS}(e_{\alpha}) e^{-\beta e_{\alpha}} de_{\alpha}.$$
 [6]

 Z_{IS} can be viewed as an inherent structure partition function. It can be expressed in terms of the probability $p_{\alpha_0}(T)$ that the protein is in the basin of attraction of the ground state, which is henceforth denoted as $p_0(T)$. This allows us to rewrite $P_{IS}(e_{\alpha}, T)$ as

$$P_{IS}(e_{\alpha}, T) = p_0(T)\Omega_{IS}(e_{\alpha})e^{-\beta e_{\alpha}},$$
[7]

if we chose the ground-state energy as the reference state of the energies $(e_{\alpha_0} = 0)$. Eq. 7 provides a method to compute $\Omega_{IS}(e_{\alpha})$ from the probability density $P_{IS}(e_{\alpha}, T)$ deduced from the sampling of molecular dynamics trajectories. The validity of the method relies on the assumption that the vibrational free energy in a basin of attraction $F_{\nu}(\alpha, T)$ does not vary significantly from a basin to another. For harmonic basins of attraction, we have

$$F_{\nu}(\alpha, T) = k_{\rm B}T \sum_{q} \ln\left(\frac{\hbar \omega_{q}}{k_{\rm B}T}\right),$$
[8]

where the values ω_q are the frequencies of the vibrational modes of the protein when it is in the configuration corresponding to the inherent structure α . Of course, these frequencies change individually from one inherent state to another, but they keep the same order of magnitude, and the sum has an averaging effect that keeps $F_v(\alpha, T)$ only weakly dependent of the conformation.

This conjecture is verified by our calculations because $\Omega_{IS}(e_{\alpha})$ deduced from $P_{IS}(e_{\alpha}, T)$ at different temperatures gives the same result in all of the ranges of inherent structure energies that are accessed with a sufficient probability to allow a correct sampling at the temperature considered. Fig. 2 shows the density of states $\Omega_{IS}(e_{\alpha})$ computed from the sampling of molecular dynamics trajectories at three temperatures. Fig. 2's results are remarkably similar, showing that the assumption that $F_{\nu}(\alpha, T)$ can be eliminated in the calculation is a good approximation; however, the results are obviously not identical. Fig. 2 Right, deduced from the calculation at the lowest temperature, shows some values of e_{α} where the density of states appears to vanish, whereas it does not vanish if $\Omega_{IS}(e_{\alpha})$ is computed from a trajectory at higher T. Such inconsistency is particularly noticeable for high values of e_{α} , and it points out the limit of the numerical calculation. As shown from Eq. 7, the probability to be in the basin of attraction of an inherent structure decreases as $\exp(-e_{\alpha}/k_{\rm B}T)$. Therefore, if we use a molecular dynamics trajectory at a low temperature to sample the basins of attraction of the inherent structures, it may happen that we completely miss some inherent structures because of insufficient sampling. This outcome is even more likely if the basin of attraction of a given inherent structure is narrow because, besides the temperature factor, its small size also reduces the chance to sample it properly. Studies performed at higher temperatures reduce this problem and $T = T_f$ is particularly appropriate because we expect a molecular dynamics trajectory to sample both structures corresponding to folded and unfolded states. Once $\Omega_{IS}(e_{\alpha})$ has been determined from Eq. 7 at a temperature where the sampling is particularly efficient, $Z_{IS}(T)$ can be computed at any temperature with Eq. 6.



Fig. 2. Comparison of the density of states $\Omega_{IS}(e_{\alpha})$ computed from the probability densities $P_{IS}(e_{\alpha}, T)$ deduced from molecular dynamics trajectories at three different temperatures: $T = T_f$ (Left), T = 0.83 T_f (Center), and T = 0.69 T_f (Right).

Fig. 3 displays $\Omega_{IS}(e_{\alpha})$ in a large range of energies. It shows an isolated minimum energy state, followed by a few states very close in energy and then a large gap, of the order of $k_{\rm B}T_f$, before the next state. A plot of the structure shows that the ground state and states immediately above it in energy have very similar shapes. Thus the few states near the minimum can be viewed as conformational substates (12) of the ground state. The gap between these states and the next states indicates that the model protein is a good folder (3), as observed in molecular dynamics simulations.

For larger energies $\Omega_{IS}(e_{\alpha})$ shows an exponential growth, with two different slopes depending on the energy range:

$$\Omega_{IS}(e_{\alpha}) = \Omega_0 \exp[g(e_{\alpha})]$$
[9]

with

$$g(e_{\alpha}) = \begin{cases} \frac{e_{\alpha}}{k_{\rm B}T_s} + A & (e_{\alpha} < 45k_{\rm B}T_f), \\ \frac{e_{\alpha}}{k_{\rm B}T_u} & (e_{\alpha} > 50k_{\rm B}T_f), \end{cases}$$
[10]

where Ω_0 and A are positive constants, and $T_s = 1.1 T_f$ and $T_u =$ 0.9 T_f . It is interesting to notice that a similar exponential dependence of the density of metastable state was found for spin glasses (13) although the energy dependence of $g(e_{\alpha})$ is more complex for such systems.

Thermodynamics in the Inherent Structure Space

The density of states in the ISL can be used to analyze the folding of the protein. Substituting the expression of $\Omega_{IS}(e_{\alpha})$ into Eq. 7 gives



Fig. 3. Density of state Ω_{IS} (e_{α}) for the inherent structures, which is obtained from $P_{IS}(e_{\alpha}, T_f)$ according to Eq. 7. The dotted lines are two distinct exponential functions with exponent corresponding to Eq. 10. (Inset) A magnification of the curve for the energies close to the ground state.

$$P_{IS}(e_{\alpha}, T) = \frac{1}{Z_{IS}(T)} \Omega_0 \exp\left[\frac{e_{\alpha}}{k_{\rm B}} \left(\frac{1}{T_s} - \frac{1}{T}\right) + A\right]$$

for $e_{\alpha} < 45k_{\rm B}T_f$, [11]

$$P_{IS}(e_{\alpha}, T) = \frac{1}{Z_{IS}(T)} \Omega_0 \exp\left[\frac{e_{\alpha}}{k_{\rm B}} \left(\frac{1}{T_u} - \frac{1}{T}\right)\right]$$

for $e_{\alpha} > 50k_{\rm B}T_f$. [12]

But Eq. 11 is also the expression that gives the behavior of $P_{IS}(e_{\alpha},$ T) at low temperature $(T < T_u)$ because, in this temperature range, the equations giving $P_{IS}(e_{\alpha}, T)$ show that low-energy inherent structures dominate. On the contrary, Eq. 12 should be used at high temperature $(T > T_s)$, because in this range high-energy inherent structures dominate. As shown in Fig. 1, in the range $T_u < T \approx T_f < T_s$ none of the energy ranges dominate, and $P_{IS}(e_{\alpha}, T)$ has two humps, one of them at low e_{α} corresponding to folded structures, the other one at high e_{α} , corresponding to unfolded structures. Therefore, when we take into account the expression of $\Omega_{IS}(e_{\alpha})$, we find that the shape of $P_{IS}(e_{\alpha}, T)$ changes qualitatively in a narrow temperature range around T_f . This picture is consistent with a first-order transition, $T_u < T < T_s$ being the coexistence region, so that the two temperatures exhibited by the analysis of $\Omega_{IS}(e_{\alpha})$ appear as the limit of the spinodals. These characteristics are consistent to the behavior of specific heat C_v . As is indicated in Fig. 1, T_s and T_u are close to the temperatures at the start and the end of the peak for $C_{\nu_{1}}$

As noticed above, the exponential growth for the density of state Ω_{IS} of the inherent structures is not a special feature of the protein model but was also obtained in spin glasses (13). However, in the present case, the exponent $g(e_{\alpha})$ is connected to the folding transition, which is specific to proteins. We also observed a similar behavior in a standard Go model, without interaction through dihedral angles.

The density of inherent structures energies and the associated partition function $Z_{IS}(T)$ are not only useful for characterizing the folding transition. They can also tell us a lot about the thermodynamics of the protein. We can build thermodynamics based only on the configurations of the inherent structures by the usual scheme of statistical mechanics,

$$U_{IS}(T) = \langle e_{\alpha} \rangle = \int e_{\alpha} P_{IS}(e_{\alpha}, T) de_{\alpha}, \qquad [13]$$

$$F_{IS}(T) = -k_{\rm B}T \log(Z_{IS}(T)),$$
 [14]

$$S_{IS}(T) = \frac{U_{IS} - F_{IS}}{T},$$
 [15]

BIOPHYSICS



Fig. 4. Thermodynamic quantities defined from the inherent structures. (*Top*) Conformational specific heat C_{IS}/k_B . (*Middle*) Magnification of the variation of the specific heat in the low-temperature range, also showing the contribution $C_W = C_v - C_{IS}$. which is expected to reflect the properties of thermalization inside each well. Notice that the value shown is C_W/N_r , i.e., the vibrational specific heat per residue, while C_{IS} , which is expected to reflect a global property of the protein, is shown for the full molecule. The dashed line indicates the harmonic limit $C_W/N_r = 3 k_B/2$. (*Bottom*) Conformation entropy S_{IS} and conformation energy U_{IS} . (*Inset*) Conformational entropy in logarithmic scale.

$$C_{IS}(T) = \frac{\langle \Delta e_{\alpha}^{2} \rangle}{k_{\rm B} T^{2}} = \frac{\langle e_{\alpha}^{2} \rangle - \langle e_{\alpha} \rangle^{2}}{k_{\rm B} T^{2}},$$
[16]

where U_{IS} , F_{IS} , S_{IS} , or C_{IS} are the internal energy, Helmholtz free energy, entropy, or specific heat, respectively, deduced only from the thermal fluctuations between conformations, so that they can be called conformational energy, conformational free energy, etc. They are shown in Fig. 4.

The conformational specific heat is remarkably similar to the full specific heat of Fig. 1. The detailed analysis of Fig. 4 *Middle*, which displays the vibrational specific heat per residue $C_W/N_r = (C_v - C_{IS})/N_r$ and conformational specific heat C_{IS} , shows that the vibrational contribution of the specific heat is very close to the value $3 k_B/2$ expected for harmonic oscillators. This finding suggests that the thermalization in the basin of attraction of each inherent structure allows a good separation between a vibrational part and a conformational part of the specific heat. When comparing the values of C_W and C_{IS} one should notice an essential difference between the two. C_W is a vibrational contribution, which comes from all of the residues and, as it is well approximated by the harmonic value, indicates that the contributions of all vibrational degrees of freedom are independent, so that C_W/N_r has an actual physical meaning. On the contrary, the

structural fluctuations entering in C_{IS} may be either a flip of a residue, for instance, and thus can be very local, or a slow structural change extending to a significant part of the protein, such as the motion of the whole α -helix. This is why the quantity C_{IS}/N_r would not have a physical meaning because a protein is not a self-similar structure. A meaningful comparison must compare C_{IS} with C_W/N_r : structural fluctuation will start to play a role in the physics of the protein when $C_{IS} > C_W/N_r$. As shown in Fig. 4, C_{IS} starts to rise significantly $\approx 0.3 T_f$ and quickly takes over for $T > 0.4 T_f$, whereas the fluctuations are dominated by the vibrational contribution in the low temperature range.

The behaviors of the conformational entropy and energy (Fig. 4 *Bottom*) show a similar rise clearly visible for $T > 0.4 T_f$. The increase of entropy that appears to accelerate and become exponential for $T > 0.4 T_f$ (see Fig. 4 *Bottom Inset*) can be understood as coming from the structural variety observed in the thermal fluctuations. This idea would be consistent with the behavior of C_{IS} and C_{W} .

Discussion

The results of the thermodynamics of the inherent structure states shed an interesting light on the current debate about the existence and observation of a protein dynamical transition (14). For our model protein two temperatures appear to play a particular role. The lowest one is $0.05 T_f$, above which energy and entropy show a rise (Fig. 4). With the resolution of our calculation, $\leq 0.05 T_f$, fluctuations among inherent states seem completely suppressed, so that this temperature could be viewed as the Kauzmann temperature for this model (15). One should, however, notice that Fig. 4 shows a gradual decrease of the entropy $\approx 0.05 T_f$, suggesting that, as previously noticed for another off-lattice protein model (16), a true thermodynamics glass transition does not exist for a finite system such as a protein. The other characteristic temperature is $0.4 T_{f}$. It should correspond to a qualitative change in some observations on the protein because, around this temperature, conformational changes among the different inherent structures start to dominate the fluctuations, as shown by the specific heat. Thus a glassy behavior could be expected when the temperature is lowered to values in the range $0.05 T_f < T < 0.4 T_f$ or below.

However, as far as we deal with equilibrium aspects, the characteristics of a glass phase are not detected in our study because they only show up in out-of-equilibrium situations. From the viewpoint of equilibrium thermodynamics, we cannot find a characteristic change caused by the "glass transition." In the temperature range $0.05 T_f < T < 0.4 T_f$, in which we could expect the glass transition, the specific heat C_v only shows a smooth evolution without any drop when temperature is lowered. In the thermodynamic sense there is no transition. However, the picture could become different if nonequilibrium aspects were probed. For instance, when cooling the protein rapidly from high temperature identified in the range $0.05 T_f < T < 0.4 T_f$, but it should depend on the cooling rate.

Our results raise a question about the nature of the fluctuations that we observe: do they correspond to the α or β relaxations observed in proteins (14)? Experiments (17, 18) and molecular dynamics simulations in which the temperature of the solvent and the protein are constrained to be different (19) show that the solvent mobility is an important factor in determining the large-scale fluctuations of a protein. The α relaxation is "slaved" to the solvent fluctuations, which is not the case for the glassy aspects of our results because the solvent only enters into

[§]The results from thermodynamics in the inherent structure space are not affected by a possible lack of ergodicity at low temperature because they are established from a density of states deduced from simulations around the T_{f} . Their accuracy is determined by the quality of the sampling of the phase space in the vicinity of the T_{f} .

our study through its contribution to the interaction energies not through its dynamics because the model uses effective potentials to include hydrophobic effects. The fluctuations must be attributed to the protein energy landscape, which suggests that the fluctuations that are observed in our simulations correspond to the β relaxation, i.e., to fluctuations that do not involve large conformational changes that could be hindered by a frozen solvent in an actual experiment. This makes sense if we think of the peculiarities of the Gō model, which strongly favors the native structure by the selection of the interactions that are introduced in the model. It is likely that such a bias toward the native shape prevents such a simplified model from exhibiting an α relaxation.

In the studies of glassy systems with inherent structures, it is customary to introduce the configurational entropy

$$s_c(e_\alpha) = k_{\rm B} \log(\Omega_{IS}(e_\alpha)).$$
[17]

In the context of a thermodynamic description based on the inherent structures, s_c can be viewed as the microcanonical entropy, whereas $S_{IS}(T)$ defined by Eq. 15 is the canonical entropy. Therefore, in the thermodynamic limit, we could expect that $s_c(\tilde{e}_\alpha) = S_{IS}(T)$, where the most probable value \tilde{e}_α of the inherent structure energy would also tend toward its mean value $\langle e_\alpha \rangle$. However, this is not true for a protein, because of the finiteness and heterogeneity of the system, but also because, below T_{f_5} the most probable value of e_α is the ground state of the protein. In this regime the density of states $\Omega_{IS}(e_\alpha)$ is dominated by a few isolated states, and one cannot build thermodynamics on the configurational entropy, whereas $S_{IS}(T)$ defined by Eq. 15 has the expected properties of a thermodynamic entropy, such as $dS_{IS}(T)/dU_{IS}(T) = 1/T$.

In this article we have shown that the ISL of a protein can be analyzed because its density of states can be deduced from the probability density of inherent states extracted from molecular dynamics simulations. The ISL contains considerably less information than the free-energy landscape of the protein, and this is what makes it accessible. At the expense of much longer calculations, the ISL appears to be accessible for an all-atom model of a simple protein because our scheme to obtain it does not require the full determination of every local minimum because it takes advantage of the usual ability of statistical physics to provide very accurate results from a simplified description of the system. An important aspect is that, in this framework, all thermodynamics properties are deduced from the density of inherent states that is obtained from molecular dynamics simulations in the vicinity of T_f to sample folded and unfolded states. As a result the difficult issue of the ergodicity of simulations at very low temperature can be avoided.

When a complicated model, such as an all-atom description is considered, the landscape becomes immense and one may be concerned by the validity of the sampling of the phase space to get the inherent structure density of states. But it should be noted that, even for a frustrated Go model, the landscape is rugged and complex. For our protein model, it is multifunneled. Kinetic studies of folding detect the existence of a second folding funnel, leading to a kinetic trap (unpublished work). The density of inherent states in the kinetic trap can be obtained and it is very low. Therefore, although the second funnel plays a significant role in the kinetics of the folding, it does not show up in the thermodynamics of the protein. This is an interesting aspect of the canonical sampling that we use. It introduces some selfconsistency in the method because states that are important for the reduced thermodynamics are also those that are more likely to be sampled. Another test of the validity of this approach is provided by the possibility to get $\Omega_{IS}(e_{\alpha})$ by sampling at different temperatures in the vicinity of T_{f} . They should lead to the same structure for the density of states. This is why, despite the obvious difficulty of carrying long molecular dynamics trajectories at different temperatures for a complex protein model, the thermodynamics in the ISL seems reachable in this case.

The $G\bar{o}$ model of a real protein is sufficient to allow the investigation of the properties of the ISL. It shows that, despite the loss of information with respect to the full free energy landscape, the ISL still contains a lot of data on the protein. It can be used to study the folding transition because P_{IS} exhibits a qualitative change across T_f and can even detect the associated spinodals. The density of inherent states shows an exponential scaling, which can be related to the folding transition. This finding suggests that it could be a general property of proteins, and it will be interesting to test this hypothesis on other protein models, including all-atom descriptions. The ISL can also be used to analyze the thermodynamics of the protein. It cannot observe the vibrational contributions, but detects the most important features of proteins, their conformational fluctuations that are responsible for their function.

Materials and Methods

The potential energy of our Go model of protein G is

$$V = \sum_{i=1}^{N_r - 1} \frac{K_b}{2} (b_i - b_{0i})^2 + \sum_{i=1}^{N_r - 2} \frac{K_a}{2} (\theta_i - \theta_{0i})^2 + \sum_{i=1}^{N_r - 3} \frac{K_d}{2} \left(1 - \cos\left(2\phi_i - \frac{\pi}{2}\right) \right) + \sum_{i>j-3}^{\text{native}} \varepsilon \left\{ 5 \left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6 \left(\frac{r_{0ij}}{r_{ij}}\right)^{10} \right\} + \sum_{i>j-3}^{\text{non-native}} \varepsilon \left(\frac{C}{r_{ij}}\right)^{12},$$
[18]

where N_r is the number of residue, and b_i , θ_i , and ϕ_i are the *i*th bond length, *i*th bond angle, and *i*th dihedral angle, respectively. r_{ij} is the distance between *i*th and *j*th residues. The index 0 denotes the parameters determined from the native structure. Such a model includes local and nonlocal interactions. The local interactions are the bonds that define the C^{α} chain through stiff elastic potentials describing the covalent bonds, angular potentials between adjacent bonds, and dihedral potentials. The natural bond length or bond angles are determined from the native structure. The nonlocal bonds connect C^{α} carbons that are not adjacent along the peptide chain, but nevertheless close in space in the folded geometry of the protein. Such carbons are said to form a "native contact" and interact with a Lennard-Jones potential in the model. They are determined from the experimental structure of the protein in its folded state: two C^{α} carbons are said to form a native contact if the distance between them or between atoms of the side chains attached to them is less than a critical distance, here chosen as 5.5 Å. A repulsive interaction is introduced between all C^a carbons that do not form a native contact to describe the steric repulsion. Such a $G\bar{o}$ model is specifically designed to ensure that the system reaches the correct folded structure and these models are generally good folders, i.e., their folding transition is sharp and leads to a well defined structure, because they are minimally frustrated. A comparison of the folding of an off-lattice Go model and a frustrated model based on hydrophobic (B), hydrophilic (P), and neutral (N) residues (BPN model) showed that the Go model leads to a sharp transition where collapse and folding occur simultaneously, whereas the highly frustrated model shows a very broad collapse and even at the lowest temperature the number of contacts is far from maximal (20), pointing out that a realistic model of protein should be minimally frustrated (2). In our model a slight frustration is introduced (10) by assuming that the dihedral angles have the ideal values 45° or 135° of the α -helices or β -sheets instead of the angles deduced from the native structure. The native structure of protein G has been obtained from the Protein Data Bank (21). The dimensionless parameters of the model are $K_b = 200.0$, $K_a = 40.0$, $K_d = 0.3$, ε = 0.18, and C = 4.0. The experimental structure lies in the basin of the global minimum of the model but does not correspond exactly to the ground-state structure of our model because of the frustration introduced on the dihedral angles.

Thermalized phase space trajectories are generated for this model, using underdamped Langevin simulations, where the masses of all of the residues are assumed to be equal. Our goal here is not to explore out-of-equilibrium states, but rather to analyze the equilibrium properties of a protein. Therefore, to avoid long nonequilibrium transients that occur on cooling, except for the general tests of the model properties, all of our simulations started from the ground state and heated up to the

- 1. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. (1991) Science 254, 1598-1603.
- 2. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) Proteins Struct. Funct. Genet. 21, 167-195.
- 3. Sali, A., Shakhnovich, E. & Karplus, M. (1994) Nature 369, 248-251.
- 4. Stillinger, F. H. & Weber, T. A. (1982) Phys. Rev. A 25, 978-989. 5. Sciortino, F., Kob, W. & Tartaglia, P. (1999) Phys. Rev. Lett. 83, 3214-3217.
- 6. Bryngelson, J. D. & Wolynes, P. G. (1987) Proc. Natl. Acad. Sci. USA 84, 7524-7528
- 7. Bryngelson, J. D. & Wolynes, P. G. (1989) J. Phys. Chem. 93, 6902-6915.
- 8. Miller, M. A. & Wales, D. J. (1999) (1999) J. Chem. Phys. 111, 6610-6616. 9. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M.,
- Wingfield, P. T. & Clore, G. M. (1991) Science 253, 657-661.
- 10. Karanicolas, J. & Brooks, C. L., III (2003) J. Mol. Biol. 334, 309-325.
- 11. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000) J. Mol. Biol. 298, 937-953.
- 12. Frauenfelder, H., Parak, F. & Young, R. D. (1988) Annu. Rev. Biophys. Biophys. Chem. 17, 451-479.

temperature of interest. To collect statistics over a given equilibrium trajectory, a typical simulation included 3×10^8 steps after the desired temperature was reached. For each temperature 20 simulations were performed with different sets of random numbers in the Langevin equations.

Equilibrium trajectories were used, on one hand, to analyze the thermodynamic properties of the model, and, on the other hand, to characterize its ISL. The basins of the potential energy surface were determined by quenching 6×10^4 instantaneous configurations separated by 10⁵ steps, i.e., calculating a steepestdescent path from these points of the trajectory. A similar algorithm had been used earlier for the BPN model of a four-helix bundle protein (22) to detect states that could be long-lived intermediates in the folding.

We thank Fumiko Takagi for helpful advice. This work was supported by the Japan Society for the Promotion of Science and the Centre National de la Recherche Scientifique.

- 13. Bray, A. J. & Moore, M. A. (1981) J. Phys. C Solid State Phys. 14, 1313-1327.
- 14. Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Young, R. D. (2004)
- Proc. Natl. Acad. Sci. USA 101, 14408-14413.
- 15. Kirkpatrick, T. R. & Wolynes, P. G. (1987) Phys. Rev. B 36, 8552-8564.
- 16. Guo, Z. & Thirumalai, D. (1995) Biopolymers 36, 83-102.
- 17. Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Young, R. D. (2005) Physica A 351, 1-13.
- 18. Ansari, A., Jones, C. M., Henry, E. R., Hofrichter, J. & Eaton, W. A. (1992) Science 256, 1796-1798.
- 19. Vitkup, D., Ringe, D., Petsko, G. A. & Karplus, M. (2000) Nat. Struct. Biol. 7, 34 - 38
- 20. Nymeyer, H., Garcia, A. E. & Onuchic, J. N. (1998) Proc. Natl. Acad. Sci. USA 95, 5921-5928.
- 21. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindvalov, I. N. & Bourne, P. E. (2000) Nucleic Acids Res. 28, 235-242.
- 22. Guo, Z. & Thirumalai, D. (1996) J. Mol. Biol. 263, 323-343.