How does the first water shell fold proteins so fast ?

Olivier Collet

How does the first water shell fold proteins so fast ?

Olivier Collet
Institut Jean Lamour, Département 1, CNRS, Nancy-Université, UPV-Metz,
Boulevard des Aiguillettes BP 239, F-54506 Vandoeuvre-lès-Nancy
(Dated: January 28, 2011)

First shells of hydration and bulk solvent plays a crucial role in the folding of proteins. Here, the role of water in the dynamics of proteins has been investigated using a theoretical protein-solvent model and a statistical physics approach. We formulate a hydration model where the hydrogen bonds between water molecules pertaining to the first shell of the protein conformation may be either mainly formed or broken. At thermal equilibrium, hydrogen bonds are formed at low temperature and are broken at high temperature. To explore the solvent effect, we follow the folding of a large sampling of protein chains, using a master-equation evolution. The dynamics shows a clear mechanism. Above the glass-transition temperature, a large ratio of chains fold very rapidly into the native structure irrespective of the temperature, following pathways of high transition rates through structures surrounded by the solvent with broken hydrogen bonds. Although these states have an infinitesimal probability, they act as strong dynamical attractors and fast folding proceeds along these routes rather than pathways with small transition rates between configurations of much higher equilibrium probabilities. At a given low temperature, a broad jump in the folding times is observed. Below this glass temperature, the pathways where hydrogen bonds are mainly formed become those of highest rates although with conformational changes of huge relaxation times. The present results reveal that folding obeys a double-funnel mechanism.

PACS numbers: 87.14.et

To this date, the three-stranded β sheet is the faster folder finding its native structure in the amazingly short time of 140 nano-seconds [1]. The protein folding problem is still considered as one of the major unsolved problems of science [2] and the answer to the Levinthal question [3] "how a protein can fold so fast ?" remains a "grand challenge" [4].

Protein folding is the process whereby a protein folds into its native structure. The slowest folding proteins may require a few minutes due among other factors to proline isomerization [5]. They fold passing through many intermediate states. On the other hand, many small single-domain proteins fold very rapidly over time scales of a few microseconds [6, 7]. For many of these proteins, the folding process is a single exponential function of time [5, 6] and is modeled by a two-state mass action model and an Arrhenius diagram on which the free energy of some ensembles of chain conformations is plotted as a function of reaction coordinate, usually not known. This diagram exhibits a transition state between the unfolded and the native states [8].

Moreover, some ultra fast folders exhibit more complex kinetics with non-Arrhenius behavior [10] (i.e. non-linear dependence of the logarithm of the folding rate on the inverse of the temperature). Some results show that the activation energy is positive at room temperature, decreases as the temperature increases and may become negative at high temperature [11]. It has been suggested that this could arise from the temperature dependence of the hydrophobic effect [12, 13].

An alternative to this transition-state view is the concept of folding funnel [14]. This energy-landscape picture is based on the idea of minimal frustration [15], which states that the evolutionary mechanism has retained those protein sequences that have a funnel-like energy landscape. In that concept, the height of the funnel represents the conformational energy and its width represents the entropy of the subset of chain conformations of a given energy [16–20]. The top of the funnel is populated by the huge number of denaturated configurations with a large energy and entropy and the bottom with the unique native structure of very low energy and quasi-nil entropy. Each protein chains folds from the top of the funnel towards the bottom.

The transition-state theory and the folding-funnel picture are two different approaches. The first one describes well the two-state kinetics, but does not explain why folding is so fast. The second one explains well why folding is so fast, and the thermodynamic free energy barrier, that gives rise to two-state kinetics and makes transition-state theory applicable to the folding process, is essentially of entropic nature.

Lattice models of proteins are among the favorite tools for the theoretical study of folding. The microscopic representation of the proteins is simplified to allow large sampling of the configurational space. Proteins are modeled as self-avoiding-walk chains of beads, which are located on a two or three-dimensional square or cubic lattice. For small-length chains, a full enumeration of the conformations allows the exact calculation of the partition function, and, thus of statistical averages [21–24]. The Hamiltonian of a given conformation of the chain results from the interaction of the first neighbors of the beads on the lattice, but not in the sequence. The more popular model of couplings between monomers are known as the Gõ[22], the HP [21, 25–27] and the random-energy model(REM) [28–33]. In the Gõ model, the interaction between the beads of a given compact structure, chosen as the native conformation, is set to -1 and all the other couplings to 0. In the HP model, the sequence of the protein is given in terms of a series of hydrophobic (H) or polar (P) residues and the coupling between two hydrophobic beads is set to -1 and that involving at least one polar monomers to 0. In the REM, a matrix of couplings between all pairs of residues is constructed by drawing random numbers from a Gaussian distribution.

Despite the numerous results obtained with such models, they fail to reproduce a fundamental feature of a protein : its cold denaturation. This mechanism is associated to the loss of stability of the native structure upon cooling down the system [34–36]. For half a century, it has been well known that water plays a crucial role in the mechanism of folding [37–39] and an understanding of cold denaturation requires a finer model of the solvent than the temperature-independent attractive parameters used in Gõ, HP or REM. Some recent refinements of the models, including temperature dependence of the hydrophobic effect, have allowed to model the cold transition [24, 40–42] to be described.

As the physics of protein folding proceeds in water and the interactions of the protein chemical groups are solvent mediated, the representation of the solvent is of great importance [43]. The hydrophobic effect is an active field of research by itself [44–48]. It represents the tendency of water to exclude nonpolar solutes. It results from a disruption of the network of hydrogen bonds between water molecules caused by the transfer of an nonpolar solute into water. The energy variation of this process is favorable at room temperature, whereas the entropy cost leads to a large positive free energy of transfer. In addition, the physics of the solvent-mediated interactions of the protein may be captured by studying the interaction of two nonpolar solutes immersed in water [43]. This prototypical interaction can be handled by averaging e.g. the degrees of freedom of the solvent in the free-energy function of two methane solutes in explicit water [49, 50]. The profile of the free energy as a function of the distance between the two methane solutes shows a deep minimum for the contact distance between the two molecules and another one, less pronounced when they are separated by a distance slightly smaller than one solvent molecule diameter. A maximum, higher than the free energy of the two isolated solutes arises between these two configurations. This maximum is known as the desolvation barrier. As the temperature increases, the contact between the nonpolar solutes becomes more favorable as the desolvation barrier is reduced. This barrier tends to favor a high thermodynamic cooperativity of the model in contrast with a model without desolvation barrier [51, 52]. It has also been shown that the physics behind this barrier is responsible for the large diversity in the folding rates, similar to what is observed experimentally [53]

Moreover, recent experimental work has shown that structural fluctuations of the solvent may control structural fluctuations of the protein [54–58]. It has also been observed that motion of hydration water drives protein dynamics [59]. This could be responsible for the protein-solvent dynamical transition connected with the liquid glass transition of hydration water [60]. These results show the importance of the degrees of freedom of the hydration first shell for the dynamics of the proteins. A theoretical model of the hydrophobic effect introduced by Muller [61] and extended by Lee and Graziano [62] allows to separate the contribution of the hydrogen bonds of the first shell and that of the bulk water. The basic idea stems from the result that the hydration of nonpolar solutes presents a large entropy cost and a small favorable energy. The hydrogen bond breakage in the bulk is considered as a two-state equilibrium between the formed and the broken hydrogen bonds. The equilibrium constant between the two states is related to the fraction of formed hydrogen bonds and to the difference in enthalpy and the ratio of degeneracy resulting from the breaking of one hydrogen bond. A similar description is used for to the water molecules of the first shell. It considers that the thermodynamics of a broken hydrogen bond in the first shell is the same than that in the bulk and gives a picture of the hydrophobic effect based on the enthalpy gain and entropy cost arising from the creation of a bond in both situations. A little bit later, Lee and Graziano [62] pointed out that the energy associated to a broken hydrogen bond is not the same for a water molecule of the first shell and for one of the bulk. The presence of an nonpolar solute induces the breakage of a hydrogen bond of the first shell, leading to a more unfavorable energy than the same event in the bulk. The two-state model of the hydrophobic effect has been applied [63] to the two-dimensional Mercedes-Benz model of water [64, 65]. As a result, they give a spectrum where the non-degenerated ground state is for the formed hydrogen bond in the first shell and the highly degenerated states for the broken hydrogen bond in the first shell corresponds to the larger value of the spectrum. The energies and the degeneracies of the formed or broken bonds in the bulk are found between the two previously described.

In this paper, this picture has been reduced further by gathering together the two close energy levels associated to the broken and formed hydrogen bonds of bulk water [40] and has been applied to a lattice model of protein to study the effect of the first shell on the protein dynamics. Aside from the hydrophobic model itself, how the solvent is simulated has a significant impact on the energy landscape of the protein. Explicit solvent models are very computationally expensive [66]. Implicit solvent models have been developed to take into account the solvent as a mean field effect [67–73]. Yet, results obtained from explicit simulations do not always agree with those from implicit models [74, 75]. Up until now, the strategy to follow the kinetics of the proteins consisted in averaging the degrees of freedom of the solvent by calculating the free energy of solvation of each protein structure and the transition rates between two protein conformations. The system evolves along effective routes made of conformations surrounded by an averaged solvent.

Here, as the solvent model allows it, we have *calculated the rate between two protein-solvent microscopic configurations and grouped together some equivalent transitions.* The physical pathways are microscopic routes in the protein and the solvent configurational space, not "mean" routes in the conformational space of the protein surrounded by an effective solvent.

The dynamics of a large set of chains in the solvent is calculated using a master equation evolution. In the spirit of the concept of folding funnel [14, 16, 17, 19], a picture of the folding in terms of two surfaces, depending on the state of the hydrogen bonds of the first shell solvent, drawn in an entropy-energy plot, is given. The mechanisms responsible for the fast folding and the glass transition are detailed in the body of the paper. In the first part, the protein and the solvent models are described. In the second part, the equations of the evolution are established. In the third part, the mechanism, which occurs during the fast overcoming of kinetic barriers is explained, then the mechanism responsible for the glass transition is revealed.

I. MODEL.

The microscopic Hamiltonian of the chain in conformation m, surrounded by a first shell of solvent molecules in configuration β and bulk solvent molecules in structure α is denoted :

$$\mathcal{H}^{\rm mic}_{m\alpha\beta} = \mathcal{H}^{\rm ch}_{m} + \mathcal{H}^{\rm shell}_{m\beta} + \mathcal{H}^{\rm bulk}_{m\alpha}$$

The first term results from the intrachain interactions, the second one from the contribution of the molecules of the first shell solvent in interaction with the protein and the last one from that of the bulk water.

A. Protein Model

The proteins are represented as self-avoiding walk strings of N monomer beads located on a two-dimensional lattice [21, 23, 33] (here $N = 12$). This length of the chain is short enough to allow analytical calculations for the dynamics and long enough to give interesting results. The compactness of a structure m is the number of intrachain contacts of the chain conformation m : $C_m = \sum_{i\geq j+3} \Delta^{(m)}_{ij}$ where $\Delta^{(m)}_{ij} = 1$ if the monomers i and j are first neighbors on the lattice and 0 otherwise. The accessible surface area of the conformation m to the solvent is defined as the number of links between the chain beads and the empty sites of the lattice : $A_m = 2N + 2 - 2C_m$. The intrachain Hamiltonian of the peptide structure m is :

$$\mathcal{H}^{\rm ch}_{m} = \sum_{i\geq j+3} B_{ij}\ \Delta^{(m)}_{ij}$$

To model the heterogeneity of the sequence of amino-acids of the chain, the couplings B_{ij} between monomers i and j are drawn at random from a Gaussian distribution centered on -2 with standard deviation equal to 1 [29]. Such a way of designing the sequences leads to create a configurational space with small energy gaps between the structures of bottom of the energy spectrum. A particular compact (native) structure does not emerge as a stable conformation of the sequence with a large energy gap with other compact conformations. To increase the stability of the native conformation of the sequence[76, 77], we select a compact conformation for the native structure of the sequence and we rank the couplings such that the minimum ones are associated with the native contacts.

B. Solvent Model

For each chain conformation, the empty nodes of the lattice models the solvent effect[24, 40]. We do not attempt to introduce a fine description of the structural properties of solvent around proteins itself, but we describe a realistic solvent effect on the weights of the chain conformations.

The bulk water contribution is simply modelled by an extensive negative free energy term which guides chains towards compact structures. The microscopic structures of the first shell solvent around any given chain conformation are separated into two groups depending on whether most of the hydrogen bonds are formed or not. Hence, each protein conformation has two possible values for its energy depending on the structure of the first shell: one for a ground state (GS) associated to a rather organized first shell and another one for an excited state (ES) with mainly broken hydrogen bonds.

For each chain conformation m, the first shell interaction is extensive with respect to A_m. The links between a solvent node and the chain beads account for the first shell contribution. The non-degenerated ground state denoted by $\beta = 0$ models the first shell water molecules with formed hydrogen bonds around the protein. It is taken as the energy reference which equals 0. The excited states, corresponding to $\beta \geq 1$, are for the $g_{\rm sh}^{A_m}$ first shell structures with broken hydrogen bonds. Their energy is $A_m\ \varepsilon_{\rm sh}$. The Hamiltonian of the first shell solvent for the chain structure m is written :

$$\mathcal{H}^{\rm shell}_{m\beta} = A_m \varepsilon_{\rm sh} \sigma(\beta) \quad \text{with} \begin{cases} \sigma(0) = 0 \\ \sigma(\beta) = 1 \text{ if } 1 \leq \beta \leq g_{\rm sh}^{A_m} \end{cases}$$

When one intrachain contact is formed, two monomers-solvent bonds are broken. Then, after removing the constant term, the pure solvent contribution is a simple function of $2C_m$[24, 42]. The factor of 2 guarantees that the solvent volume does not depend on the chain structure. For each chain structure m, the $g_{\rm bk}^{2C_m}$-fold degenerated Hamiltonian of the pure solvent is independent of the bulk micro-state :

$$\mathcal{H}^{\rm bulk}_{m\alpha} = 2C_m \varepsilon_{\rm bk} \qquad \text{for } 1 \leq \alpha \leq g_{\rm bk}^{2C_m}$$

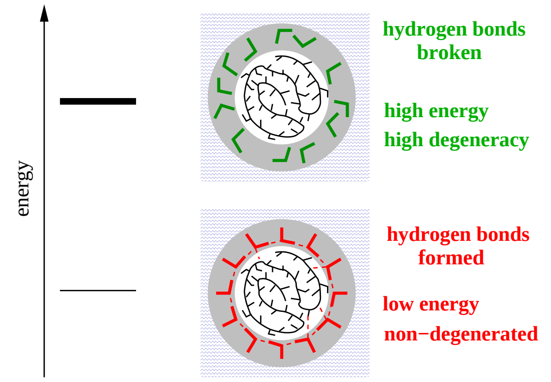

FIG. 1: The solvent around a chain conformation chosen at random. The unique solvent configuration where the hydrogen bonds between water molecules and between water and the chain are formed (shown in the left picture) is the non-degenerated ground state (GS) of the first shell and the other structures (one is shown in the right picture), where the hydrogen bonds are mostly broken, are grouped together in the highly degenerated excited state (ES). Only, the highly organized and highly disordered solvent configuration are taken into account. All the other cases are not considered in this two-state picture which only takes into account the lowest energy and largest entropy macro-states which are the most important contribution for the statistical physics approach.

The results given in the paper hold while the parameters are ranked as follow : $0 < \varepsilon_{\rm bk} < \varepsilon_{\rm sh}$ and $g_{\rm bk} < g_{\rm sh}$. In the spirit of the results obtained by Silverstein *et al*[63], the values of the solvent parameters are ranked as follow $\varepsilon_{\rm bk} = 0.2$, $\varepsilon_{\rm sh} = 0.6$, $g_{\rm bk} = 2$ and $g_{\rm sh} = 3$. The reported results in this paper are quite robust with regards to change the parameters holding the above ranking equation. However, for some technical reasons (discussed below) if $g_{\rm bk}$ or/and $g_{\rm sh}$ are chosen too large, the computational times of the calculation becomes too prohibitive.

The energy and the degeneracy of the ground and excited macro-state of each peptide conformation are :

$$\left.\begin{array}{l}\mathcal{H}^{\rm mac}_{m\sigma} = \mathcal{H}^{\rm ch}_{m} + 2C_m\varepsilon_{\rm bk} + \sigma A_m\varepsilon_{\rm sh} \\ g_{m\sigma} = g_{\rm sh}^{\sigma A_m}\; g_{\rm bk}^{2C_m}\end{array}\right\} \text{ with } \sigma = 0 \text{ or } 1$$

C. Results for the chain in interaction with the solvent

The thermal equilibrium probability of each macro-state is given by : $\mathcal{P}^{\rm eq}_{m\sigma} = g_{m\sigma}\exp(-\mathcal{H}^{\rm mac}_{m\sigma}/T)/Z(T)$ and that of a chain structure is $P^{\rm eq}_m = \sum_{\sigma=0}^{1}\mathcal{P}^{\rm eq}_{m\sigma}$. Here the Boltzmann constant is set to 1 and the partition function is:

$$\begin{aligned} Z(T) &= \sum_m \sum_{\alpha=1}^{g_{\rm bk}^{2C_m}} \sum_{\beta=0}^{g_{\rm sh}^{A_m}} \exp\left(-\frac{\mathcal{H}^{\rm mic}_{m\alpha\beta}}{T}\right) \\ &= \sum_m g_{\rm bk}^{2C_m}\exp\left(-\frac{\mathcal{H}^{\rm ch}_m + \mathcal{H}^{\rm bulk}_{m\alpha}}{T}\right) \\ &\quad \left[1 + g_{\rm sh}^{A_m}\exp\left(-\frac{A_m\varepsilon_{\rm sh}}{T}\right)\right] \end{aligned} \tag{1}$$

The native conformation is the structure of largest value of $P^{\rm eq}_m$ determined by a full enumeration of the conformational space of the chain at low temperature. For the set of couplings B_{ij} and solvent parameters used here, the native conformation is a compact structure of intrachain energy -9.895. The folding transition temperature is defined as, the melting temperature T_m of the experimental literature[78, 79] at which the equilibrium probability of the native structure equals that of all the other denatured conformations One specific chain structure (the native conformation) has a larger equilibrium probability to occur than all the other conformations for $T < T_m$ or in others words $P^{\rm eq}_{\rm Nat}(T_m) = 0.5$ (here $T_m = 0.90$). Fig

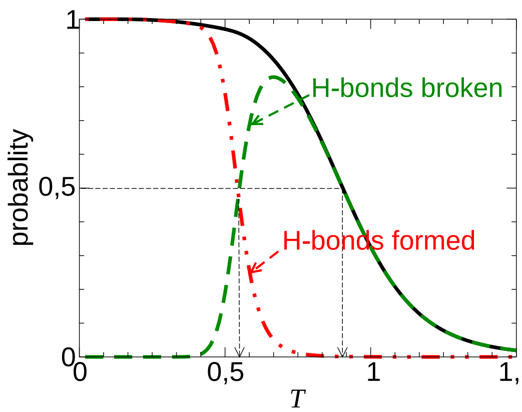

FIG. 2: Equilibrium probabilities of occurrence of the native structure (black solid line) as functions of the temperature (given in arbitrary units) with the contribution of the ground state (red long dashed line) and of the excited state (green dashed line). At a temperature T_0, the probability that the hydrogen bonds are broken becomes significant and at T^*, the probability to observe hydrogen bonds around the native conformation is the same as that to observe water with broken hydrogen bonds. The probability of occurrence of the native structure equals 1/2 at T_m.

rence of this native structure (Nat) surrounded by water with formed hydrogen bonds reaches one at low temperature. Other solvent configurations around Nat become relevant for $T > T_0$ (here $T_0 = 0.45$). Last, the probability of occurence of the native structure with of the first shell solvent in ES equals that in GS at a specific temperature denoted T^* (here $T^* = 0.54$).

We note in passing, as the solvent parameter of the GS of the first shell is lower than that of the bulk, it could be possible to observe cold denaturation of the chains, if the GS of the extended chains would be the state of lowest energy among the whole configurational space [24, 42]. Here, however, the parameterization chosen avoids this possibility in order to study only the folding mechanism.

Figure 3 shows the conformational distribution as function of the number of native intrachain contacts calculated as follow :

$$F_T(Q) = -T \ln \sum_m \delta(Q - Q_m) \sum_\sigma g_{m\sigma} \exp(-\mathcal{H}^{\rm mac}_{m\sigma}/T)$$

where $\delta(0 = 1$ and 0 otherwise and Q_m is the number of native contact of the chain structure m. Under dena-

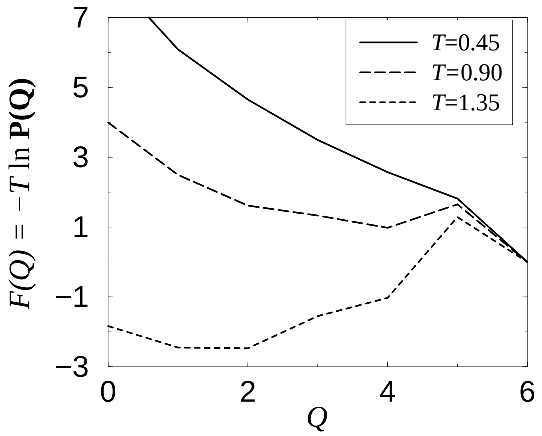

FIG. 3: Free energy profile for the lattice model at the three temperatures indicated. $P(Q)$ is the probability of occurrence over Q at thermal equilibrium.

tured conditions ($T > T_m$), the free energy profile show a barrier between the native and non-native conditions. The sets of conformation with one or two intrachain contacts, surrounded by solvent in ES (see fig. 4) have the largest probability of occurrence. At the melting temperature, the same barrier still separates the equiprobable native and non-native populations. Under strong native condition (low temperature), the shape of the free energy profile is similar to that observe for a downhill folding.

Indeed, as the temperature decreases, the more probable non-native population shifts from the sets with few native contacts to that with the maximal native contacts

D. Results for the solvent at equilibrium

At thermal equilibrium with a bath at temperature T, the probability of occurrence of the conformation m with the solvent in micro-state $(\alpha\beta)$ tend towards : $p^{\rm eq}_{m\alpha\beta} = \exp\left(-\mathcal{H}^{\rm mic}_{m\alpha\beta}/T\right)/Z(T)$.

The mean energy of the first shell solvent around chain conformation m is :

$$U^{\rm shell}_m(T) = \frac{\sum_{\alpha\beta} p^{\rm eq}_{m\alpha\beta} \mathcal{H}^{\rm shell}_{m\beta}}{\sum_{\alpha\beta} p^{\rm eq}_{m\alpha\beta}} = \frac{A_m \varepsilon_{\rm sh} g_{\rm sh}^{A_m} \exp(-A_m\varepsilon_{\rm sh}/T)}{1 + g_{\rm sh}^{A_m} \exp(-A_m\varepsilon_{\rm sh}/T)}$$

and the heat capacity is :

$$c^{\rm shell}_m(T) = \frac{{\rm d}U^{\rm shell}_m}{{\rm d}T} = \frac{{A_m}^2 \varepsilon_{\rm sh}^2 g_{\rm sh}^{A_m} \exp(-A_m\varepsilon_{\rm sh}/T)}{T^2[1 + g_{\rm sh}^{A_m} \exp(-A_m\varepsilon_{\rm sh}/T)]^2}$$

These curves as function of the temperature only depends on the chain exposure to the solvent, *i.e.* on the compactness (fig.4). They exhibit a maximum at the same tem-

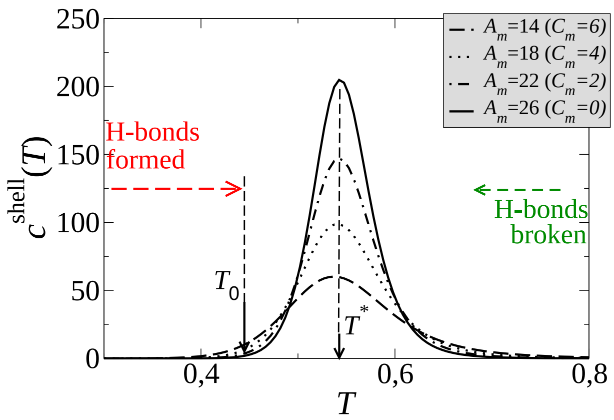

FIG. 4: The heat capacity of the solvent around the chain structures depends on the compactness of the conformation. At low temperature, the chain are surrounded by rigid cages of water molecules. The higher the compactness, the higher is the temperature of occurrence of the chain with broken hydrogen bonds of the solvent.

perature: $T^* = 0.54$. It is the temperature of equiprobability of occurrence of the two phases: broken and formed hydrogen bonds of the first shell around the peptide. For $T_0 < T < T^*$, the solvent configurations with a water cage around the peptides is preferred to the broken hydrogen bonds. For $T < T_0$, it is the only relevant state. Above T^*, the solvent occurs in the excited macro-state. This is in good agreement with the result obtained for the thermodynamics of the chains presented above.

II. TIME EVOLUTION.

To explore all the possible routes from the non-native structures to the native conformation, the probabilities of each micro-state, composed of one protein structure in interaction with one solvent configuration, evolve using a continuous time Markov process applied to a large sampling of peptides.

Master equation approach to protein folding has already be used in lattice model with an effective solvent [80]. It is shown in appendix A, that a master equation of the macro-states may be deduced from the master equation of the micro-states. In a finite time approach, the probabilities of the macro-states evolve following the Euler algorithm[81] :

$$\mathcal{P}^{\rm mac}_{m\sigma}(t + \delta t) = \mathcal{P}^{\rm mac}_{m\sigma}(t) + \delta t \sum_{m'} \sum_{\sigma'} Y_{m\sigma,m'\sigma'} \mathcal{P}^{\rm mac}_{m'\sigma'}(t)$$

where

$$Y_{m\sigma,m'\sigma'} = g_{m\sigma}\,\frac{V^{(0)}_{mm'}}{\tau^{\rm mic}_{m,m'}}(1+\exp((\mathcal{H}^{\rm mac}_{m\sigma}-\mathcal{H}^{\rm mac}_{m'\sigma'})/T)^{-1}$$

for $m \neq m'$ or $\sigma \neq \sigma'$ and

$$Y_{m\sigma,m\sigma} = -\sum_{m'\sigma'\neq m\sigma} Y_{m'\sigma',m\sigma}$$

As explained in appendix A, $V^{(0)}_{mm'} = 1$ if structures m and m' are connected by a one monomer move (either a corner flip or a tail move) or if $m = m'$. The characteristic time associated to a chain move ($\tau^{\rm mic}_{m,m'} = \tau_c$ if $m \neq m'$) and to a solvent move ($\tau^{\rm mic}_{m,m'} = \tau_s$ if $m = m'$) are set to $\tau_c = 1$ and $\tau_s = 0.001 \ll \tau_c$.

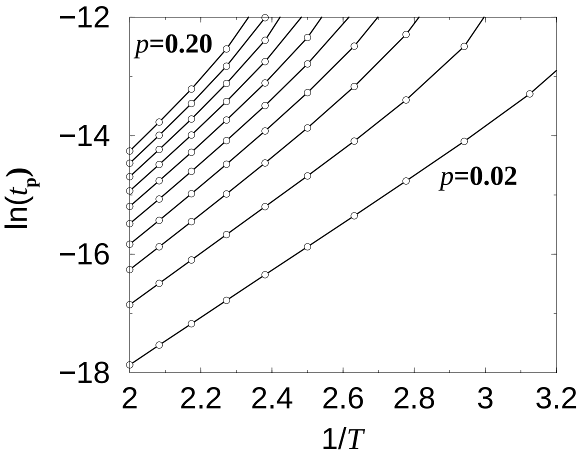

FIG. 5: Logarithm of the waiting time t_p to observe the native structure with a probability equal to p plotted as a function of the inverse of the temperature for $p = 0.02$ to $p = 0.20$ by steps of 0.02.

A sufficient condition to conserve the norm of the probability vector (i.e. $\sum_{m\sigma}\mathcal{P}^{\rm mac}_{m\sigma}(t) = 1$) is satisfied by fixing $\delta t = 1/\max_{m\sigma}\{Y_{m\sigma,m\sigma}\}$. As the values of $g_m\sigma$ and then those of $Y_{m\sigma,m'\sigma'}$ may be huge, the value of δt is tiny. Then, the evolution of the probability vector is very slow.

The simulations of folding start with the initial condition : $\mathcal{P}^{\rm mac}_{m\sigma}(0) = 1/2N_{\rm conf}$ where $N_{\rm conf} = 15019$ is the number of chain structures. The probability of occurrence of conformation m after time t is $P_m(t) = \mathcal{P}^{\rm mac}_{m;0}(t) + \mathcal{P}^{\rm mac}_{m;1}(t)$. The waiting time to observe the native structure with a probability p is noted t_p. The main results of the calculation are summarized in fig.5 and 6. They exhibit some findings on the out of equilibrium folding of a large sampling of chains at different temperatures.

Figure 5 shows a non-Arrhenius behavior of the model. Because of the tiny value of δt, the early events of the folding are shown, here. The remaining part of this plot will be deduced from the results given below. It will be shown that even if the curves may be well fitted by a Vogel-Fulcher-Tamman function ($t_p(T) \propto \exp(-E_a/(T-T_0))$), which is the signature of an α-relaxation with a dynamical temperature $T_0 = T_0(p)$, the remaining part of the curves exhibits a more complex shape which can not be capture by effective solvent models.

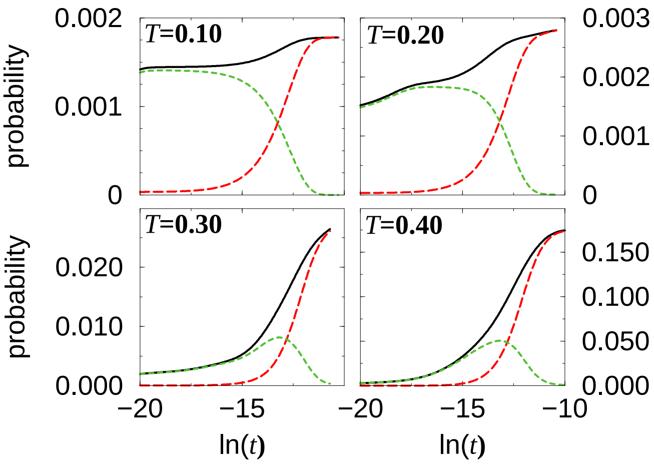

FIG. 6: Evolution of the probabilities of occurrence of native structure (back solid line), the contribution of excited state (green dashed line) and of the ground state (red long dashed line) as functions of the time at different temperatures. For temperature of 0.10 and 0.20, some chains topologically close to the native structure reach it very fast because no activation barrier occurs in their pathway. Then, the probability of the ES decreases monotonically to zero and that of the native structure reach a plateau and the kinetics is frozen. As the number of chain structures of the basin do not depend on the temperature, the plots have very similar shapes and the values of the probabilities are very close to each other showing that a local phenomenon is taking place. For temperature of 0.30 and 0.40, the ES of the native conformation (which have a null equilibrium probability) acts as an attractor, in the early events, and reach very rapidly a maximum of its probability of occurrence, much larger than its equilibrium probability. A fast transition takes place towards the GS of the native structure. Then, a slower dynamic regime guides the other chains towards the native structure. Now, the values of the probabilities depend on the temperatures showing that a global mechanism dominates.

Figure 6 shows that, even if its equilibrium probability is infinitesimal, the excited state of the solvent acts as a strong dynamical attractor above a particular temperature to be discussed later. Below this temperature the kinetics is the same whatever the temperature.

III. RELAXATION TIMES OF THE MOVES.

Only two connected macro-states $(m\sigma)$ and $(m'\sigma')$ are considered for a while. The equations 6 and 7 show that the rate of the transition $(m'\sigma') \to (m\sigma)$ depends on the degeneracy of the macro-state $(m\sigma)$ and not of that of $(m'\sigma')$. In other words, the increase of the probability of $(m\sigma)$ and the decrease of that of $(m'\sigma')$ is due to the degeneracy of $(m\sigma)$ and to the difference of energy between the states with $(m'\sigma')$. After thermal equilibration, the probability of $(m\sigma)$ would go towards $\mathcal{P}_{m\sigma}^{(\infty)} = \mathcal{P}_{m\sigma}^{\rm eq}/(\mathcal{P}_{m\sigma}^{\rm eq} + \mathcal{P}_{m'\sigma'}^{\rm eq})$ The solution of the system of equations 6, for only two states is :

$$\mathcal{P}_{m\sigma}^{\rm mic}(t) = \mathcal{P}_{m\sigma}^{(\infty)} + [\mathcal{P}_{m\sigma}(0) - \mathcal{P}_{m\sigma}^{(\infty)}]\ \exp(-t/\overline{\tau}_{m\sigma,m'\sigma'})]$$

with a very small relaxation time :

$$\overline{\tau}_{m\sigma,m'\sigma'} = (Y_{m\sigma,m'\sigma'} + Y_{m\sigma,m'\sigma'})^{-1} \ll \tau_{m,m'}^{\rm mic} \qquad (2)$$

which now depends on the energies and degeneracies of the macro-states. While numerous chains initially in state $(m\sigma)$ move into state $(m'\sigma')$, some of them may go back to $(m\sigma)$. As a consequence, the relaxation times depend on forward and backward rates of a move. As an aside, if the move from $(m\sigma)$ to $(m'\sigma')$ were allowed and not the backward transition, then the relaxation time would simplify to $1/Y_{m\sigma,m'\sigma'}$.

As the degeneracy of the excited state is always larger than that of the ground state, the value of the rates between two excited states of the chain is larger than that between two ground states (if the energy difference is of the same order) As a consequence, the relaxation times of the former transitions are always very small at not too low temperature. They are supposed to simulate peptide evolving in a fluid solvent. In contrast, the relaxation times of the latter transitions are larger. This models chains evolving in a viscous medium. Thus, the dynamics of the chains surrounded by water with broken hydrogen bonds is faster than that of low degenerated structures in interaction with solvent with formed hydrogen bonds.

In addition, the more extended the structure, the larger the degeneracy of the excited states and the smaller is the relaxation time. This is because a more extended chain has a larger exposure to the solvent and thus a lot of hydrogen bonds are broken in the first shell and as a consequence the dynamics is faster.

Last the connexion between the excited and ground states of the same chain conformation ($m \equiv m'$) have a small relaxation time because $\tau_s \ll \tau_c$. The difference between that case and the fluid connexion depends on the ratio τ_c/τ_s.

IV. FAST FOLDING MECHANISM

To understand the mechanism underlying the fast folding, a possible pathway leading to the native structure, shown in fig.7, is considered as a first approach of the problem. This pathway consists in five chain structures connected by one monomer moves. In contrast with the whole conformational space where there exist a great number of routes from a given conformation to the native structure, in this section we first consider a simplified trajectory where there is a unique pathway (via TS$_1$, Tp$_2$ and TS$_2$) from Tp$_1$ to Nat. Structures Tp$_1$ and Tp$_2$ have five intrachain contacts and thus, smaller energy than TS$_1$ and TS$_2$ which have three intrachain contacts. Then, the GS of the structures Tp$_1$ and Tp$_2$ should act as kinetics traps and that of TS$_1$ and TS$_2$ as transition states towards the native conformation. At $t = 0$, the initial probability of each macro-state is set to $P(t = 0) = 0.1$. An Euler algorithm is applied to simulate the evolution of the probabilities of the ten macro-states of this subsystem composed by the ES and the GS of this five chain structures. The waiting time t_p to observe the native structure with a probability equal to p is plotted for $p = 0.30$ to $p = 0.65$ by steps of 0.05 as function of the temperature. At $t = 0$, the probability of occurrence of Nat is already 0.20 and for $p < 0.30$, the waiting time is tiny since a many chains in the TS$_2$ states fold instantaneously in Nat. For $T < 0.20$, the waiting time to reach a ratio $p = 0.30$ becomes constant, but that to observe a larger ratio become huge.

For $p > 0.30$, the waiting times increase continuously as the temperature is decreased until a temperature, depending on p and denoted by $T_g(p)$, where a broad dynamical transition occurs. The kinetics is fluid above T_g and glassy below.

Figure 8 shows the very early events of the folding

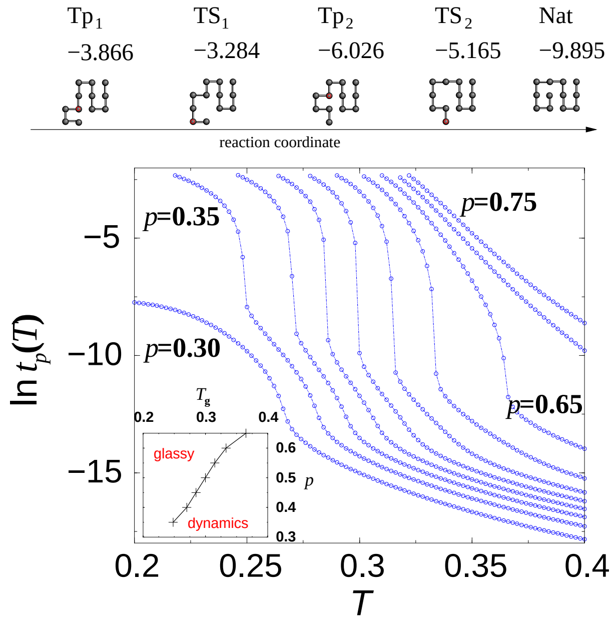

FIG. 7: Top: A possible pathway with five chain structures connected by one monomer moves ending in Nat taken as an example to capture the role play by the solvent in the folding kinetics. The energy of the GS is given above of each conformation. A single route from Tp$_1$ to Nat is considered, to simplify the study of this first approach, and the reaction coordinate is simply a function (not given) of the number of conformational changes necessary to reach Nat. Bottom: the waiting times to observe a ratio p of proteins in Nat, starting from an equiprobability of all the states, show a glass transition at temperatures depending on p (shown in the inset).

of this small system. The ES of Nat relax into the GS via a, very fast, solvent transition. At a temperature independent time, t_0, the probability of the ES of Nat becomes very small.

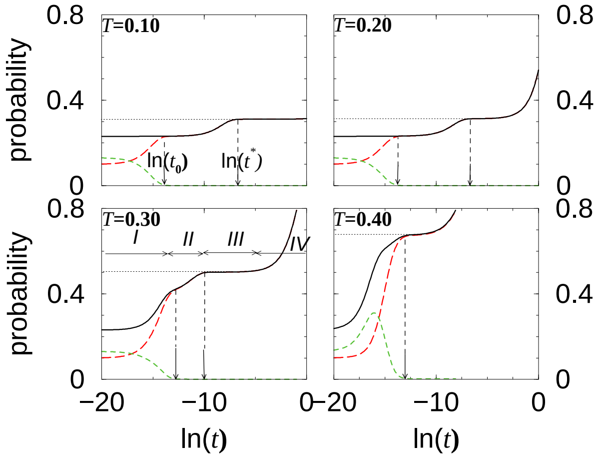

FIG. 8: . Evolution of the probabilities of the native structure (black solid lines) and the ES (green dashed lines) and the GS (red long dashed lines) contributions as functions of the time for different temperatures.

At low temperature ($T = 0.10$ or $T = 0.20$), half of the chains in conformation TS_2 fold in Nat and the other half goes to Tp_2 between the time t_0 and a temperature dependent time, t^* after which the probability of occurrence of the native structure becomes constant for a while. Then, a long plateau with $p \approx 0.3$ occurs where the dynamics is glassy.

At higher temperature, the probability of occurrence of Nat at t_0 becomes higher and the length of the plateau smaller. Many chains in the Tp_1, TS_1 or Tp_2 jump to Nat. At $T = 0.40$, the probability of the ES becomes very high (≈ 0.3) at a time denoted by t_M in the early events and the probability of Nat equals 0.8 very fast. Between t_M and t^*, the solvent of the chains in ES relax in GS. Here, the ES of Nat acts as a strong dynamical attractor.

The instantaneous flux from $(m'\sigma')$ to $(m\sigma)$, given by $k_{m'\sigma'\to m\sigma} = Y_{m\sigma,m'\sigma'}\mathcal{P}_{m'\sigma'}(t) - Y_{m'\sigma',m\sigma}\mathcal{P}_{m\sigma}(t)$ can also be calculated. Figure 9 shows that, at $T = 0.10$, the kinetics is only guided by the difference of energy of the micro-states.

The solvated chains have a very large probability to move towards a micro-state of lower energy. Between the initial time and t_0, the ES of Nat relax very fast to the GS of Nat. During this period, the ES of all structures relax to the GS. Between the times t_0 and t^*, TS_1 and TS_2 relax to Tp_1, Tp_2 and Nat. The probabilities of GS of Tp_1, Tp_2 and Nat increase until all other state have a infinitesimal probability and the probability of the Nat increases to one half of that of TS_2. Then, all the fluxes become tiny and the dynamics is frozen.

At $T = 0.40$, the equilibrium probabilities of the ES

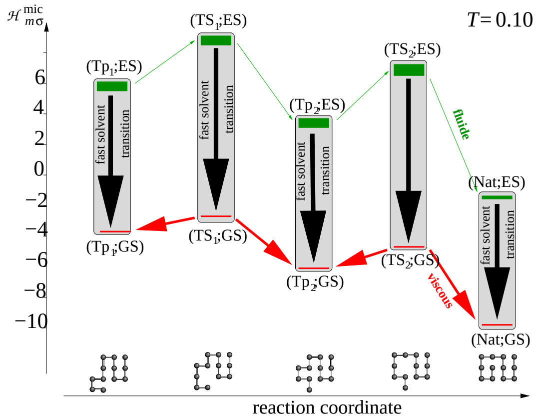

FIG. 9: The direction and magnitudes of the flux of the connexion of the pathway shown in fig.7 at t_0 and $T = 0.10$. The larger the flux, the wider is the line. Connections with tiny flux are not drawn.

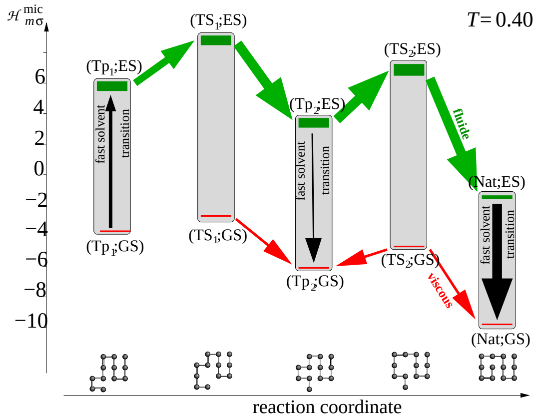

FIG. 10: Same than fig.9 at $T = 0.40$. Flux in the GS pathway is the same for both simulations.

are infinitesimal. Then, we could also expect a kinetic which would only pass through the ground states, as for $T = 0.10$. Instead of this, the sampling of chains finds another strategy to overcome the barrier very fast. At time t_0, the probability of the GS of Nat is already of 0.6. The flux between the pairs of states, drawn in fig.10, show that the chains in GS of Tp_1 do not fold via the GS of the other structures of the pathway. First, they reach the ES of Tp_1 and then they pass through the ES of the other structures and lastly they relax to the GS of Nat. This is a consequence of the larger transition rates between the ES than between the GS. In addition, the probabilities of the ES remain very small.

The chains follow a pathway with large transition rates between improbable states for which the in going flux equals the out going keeping their probabilities small.

At time t_0, the probability of Nat is around 0.7. The probabilities of the ES are quasi-null. The flux via the ES pathway is the same as via the GS pathway. The kinetics become very slow and then a plateau occurs in the curves of the probability of Nat as a function of the time shown previously.

V. HOW WATER LUBRICATES OR FREEZES THE FOLDING. A PHYSICAL PICTURE.

Curves of figs. 6 and 8 are not exactly the same but are similar. The basic mechanisms found for the direct fold in the case of the five-chain conformations pathway may be extended to the whole configurational space of the protein. Put together, the results of this paper allow us to give the picture of the folding given in fig.11. To draw one of the envelop in the energy-entropy plot, the entropy, noted $S_\sigma(E)$ for $\sigma = 0$, associated to a given microscopic energy, is calculated from the number of chain and solvent configurations, with formed hydrogen bonds. The same calculation, of the entropy noted $S_\sigma(E)$ for $\sigma = 1$, is done for the solvent without hydrogen bonds. The two functions $S_\sigma(E)$ are related to the total number of protein-solvent configurations whose total energy matches the energy values E:

$$S_\sigma(E) = \ln[\sum_m \sum_{\alpha,\beta} \delta(\sigma - \sigma(\beta)) \quad \delta^{(\varepsilon)}(E - \mathcal{H}^{\rm mic}_{m\alpha\beta})]$$

where $\delta^{(\varepsilon)}(x) = 1$ if $-\varepsilon/2 < x < \varepsilon/2$ and 0 otherwise. They may be written as functions of the degeneracy of the macro-states:

$$S_\sigma(E) = \ln[\sum_m g_{m\sigma} \ \delta^{(\varepsilon)}(E - \mathcal{H}^{\rm mac}_{m\sigma})]$$

Moreover a parameter θ_m allows one to distinguish between the structures of the bottom of the configurational valleys, which may be kinetics traps in the folding. One defines $\theta_m = 1$ for the macro-states, only connected to macro-states of higher energies and 0 otherwise. Although the native conformation satisfies to this definition, it can not be considered as a trap and the value of $\theta_{\rm Nat}$ is set to 0. Considering the five structures of fig. 7, 9 and 10 as an example, one has $\theta_{\rm Tp_1} = \theta_{\rm Tp_2} = 1$ and $\theta_{\rm TS_1} = \theta_{\rm TS_2} = \theta_{\rm Nat} = 0$. Thus, the envelop of the trap region in the energy-entropy plot is a subset of the GS barrel, given by :

$$S_{\rm Tp}(E) = \ln[\sum_m \ \theta_m g_{m;0} \ \delta^{(\varepsilon)}(E - \mathcal{H}^{\rm mac}_{m;0})]$$

where $g_{m;0}$ and $\mathcal{H}^{\rm mac}_{m;0}$ are for the degeneracy and energy of the GS of the structure m.

Figure 11 shows the three surfaces drawn on the same plot. The GS-barrel (respectively ES-funnel) is associated to the GS (respectively ES) of the chain structures. A part of the surface of the GS-barrel is populated with the trap conformations We have already shown, in this paper, that the connections between two chain structures in the ES-funnel have small relaxation times and that in the GS-barrel long relaxation times. This results from the following mechanism. A single protein and its solvent is in a given configuration, at a given time, which belongs to a corresponding macro-state, and not in all the configurations of a the macro-state. As a consequence, the transition rates between one protein-solvent configuration of the macro-state $(m\sigma)$ and those of a macro-state $(m'\sigma')$ depends on the energy difference between these states and also on the degeneracy of $(m'\sigma')$ but not on the degeneracy of $(m\sigma)$.

In particular, the connexion from a given configuration

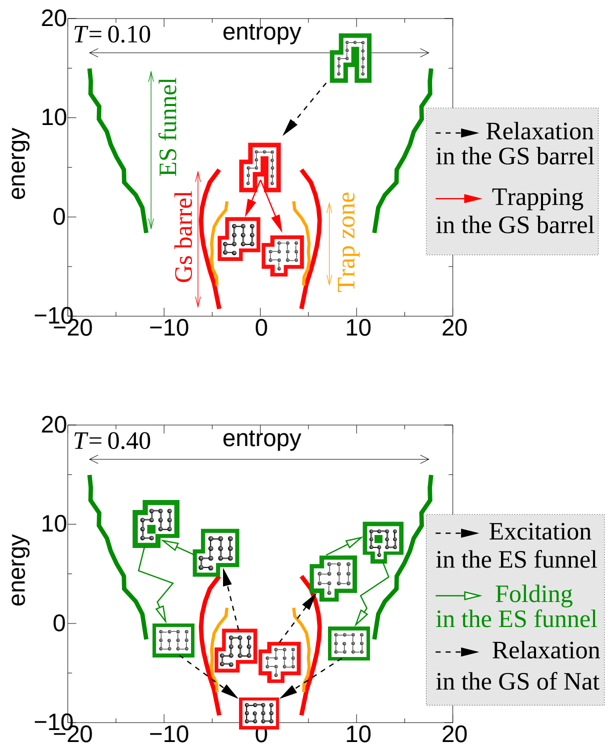

FIG. 11: Top: At $T = 0.10$, the chains in any energy level of the ES-funnel relax very fast in the GS-barrel. Then, they move down slowly in the barrel, they reach the trap zone and the dynamics is frozen. Only a tiny fraction of the chains, synthesized in chain structures very close to Nat, reach it in a reasonable time. Bottom: The picture of the folding at higher temperature but below T_0 where the ES of the chains still has an infinitesimal equilibrium probability, depicts a different mechanism. At $T = 0.40$, the chains in the GS-barrel move to the ES-funnel. Then, most of them fold very fast towards the bottom of the ES-funnel (in the ES of the native structure) and they relax in the bottom of the GS barrel (in the GS of the native structure). They get round the trap zone by passing in the ES-funnel where the transition rates are very high. A few chains relax in the GS-barrel during their descent in the ES-funnel and then they have slow dynamics. The temperature where the folding mechanism switches from one scenario to the other one is the glass temperature of the system.

of the GS to those of the ES of the same chain conformation involves a huge number of routes which increase the energy On the opposite, a few connexions allow the backward transition which decreases the energy.

At "very low" temperature, the proteins and solvent follow pathways which minimize the microscopic energy. Then, they very slowly evolve, along the few routes of the GS barrel and fall rapidly in some trap conformations.

At "not too low" temperature, proteins and solvent follow pathways where the number of routes is maximized. Then, they quickly evolve along the vast possibilities of routes of the ES funnel without traps and reach very fast the native structure.

As it has previously been shown, the temperature of glass transition which separates the two mechanisms is not clearly defined because it depends on the ratio of folding proteins

In addition, we mention that, the calculation of the partition function is independent on the informations concerning the network of connexions. As a consequence, the glass temperature is not related to the temperatures (T_0, T^* and T_m) defined above. This explains why the excited states of the first shell lubricate the folding under conditions where only the ground states have non-nil equilibrium probabilities and why the folding is frozen at lower temperature.

VI. CONCLUSION.

We have shown that a model of protein-solvent which takes into account the difference of degeneracies of the bulk solvent and the first shell solvent with mainly formed or mainly broken hydrogen bonds permits an understanding of the mechanism which may lead to quasi-instantaneous folding of a sufficiently significant ratio of the proteins in solution. Figure 11, shows that two distinct folding mechanisms exist. In the first one, the folding times are very large and in the second one very short.

Acknowledgments

Acknowledgment to Christophe Chatelain and Bertrand Berche for helpful discussions and to Rosemary Harris and Chris Chipot for critical reading of the manuscript.

Appendix A.

The probability of occurrence of the conformation m and the solvent in micro-state (α, β) at time t is denoted by $p^{\rm mic}_{m\alpha\beta}(t)$. The master equation of the system is written :

$$\frac{\mathrm{d}p^{\rm mic}_{m\alpha\beta}}{\mathrm{d}t} = \sum_{m'\alpha'\beta'} X_{m\alpha\beta;m'\alpha'\beta'}\ p^{\rm mic}_{m'\alpha'\beta'} \qquad (3)$$

where $X_{m\alpha\beta;m'\alpha'\beta'} = X(m'\alpha'\beta' \to m\alpha\beta)$ is the transition rate from configuration $(m'\alpha'\beta')$ to $(m\alpha\beta)$. The diagonal terms which take into account of the transition from $(m\alpha\beta)$ to the other configurations, are :

$$X_{m\alpha\beta;m\alpha\beta} = -\sum_{(m'\alpha'\beta')\neq(m\alpha\beta)} X_{m'\alpha'\beta';m\alpha\beta}$$

The detailed balance conditions,

$$X_{m\alpha\beta;m'\alpha'\beta'}p^{\rm eq}_{m'\alpha'\beta'} = X_{m'\alpha'\beta';m\alpha\beta}p^{\rm eq}_{m\alpha\beta} \qquad (4)$$

allow to write the rate of transition as :

$$X_{m\alpha\beta;m'\alpha'\beta'} = \frac{V^{(0)}_{mm'}}{\tau^{\rm mic}_{m,m'}}\ a_T(\mathcal{H}^{\rm mic}_{m\alpha\beta};\mathcal{H}^{\rm mic}_{m'\alpha'\beta'}) \qquad (5)$$

where $V^{(0)}_{mm'} = V^{(0)}_{m'm} = 1$ if the two structures are connected by a corner flip and tail moves[82] (see fig.12)whatever the solvent configurations and 0 otherwise. The acceptance function is $a_T(x;x') = [1+\exp((x-x')/T)]^{-1}$ and $\tau^{\rm mic}_{m,m'}$ is a symmetric function : $\tau^{\rm mic}_{m,m'} = \tau_c$ if $m \neq m'$ and $\tau^{\rm mic}_{m,m'} = \tau_s$ if $m = m'$.

The probability of occurrence of the macro-state $(m\sigma)$ at time t denoted

$$\mathcal{P}^{\rm mac}_{m\sigma}(t) = \sum_{\alpha,\beta} p^{\rm mic}_{m\alpha\beta}(t)\ \delta(\sigma-\sigma(\beta))$$

where $\delta(0) = 1$ and $\delta(n) = 0$ if $n \neq 0$ and the following relations :

$$\begin{aligned}\sum_{\alpha'\beta'} p^{\rm mic}_{m'\alpha'\beta'} &= \sum_{\sigma'}\sum_{\alpha'\beta'} p^{\rm mic}_{m'\alpha'\beta'}\delta(\sigma'-\sigma(\beta')) \\ &= \sum_{\sigma'}\mathcal{P}^{\rm mac}_{m\sigma'}\end{aligned}$$

$$\sum_{\alpha\beta}\delta(\sigma-\sigma(\beta)) = g_{m\sigma}$$

will be used below.

Now, we rewrite the master equation 3 as :

$$\sum_{\alpha\beta}\frac{\mathrm{d}p^{\rm mic}_{m\alpha\beta}}{\mathrm{d}t}\delta(\sigma-\sigma(\beta)) = \sum_{\alpha\beta}\sum_{m'\alpha'\beta'}\delta(\sigma-\sigma(\beta))X_{m\alpha\beta m'\alpha'\beta'}\ p^{\rm mic}_{m'\alpha'\beta'}$$

yields :

$$\sum_{\alpha\beta}\frac{\mathrm{d}p^{\rm mic}_{m\alpha\beta}}{\mathrm{d}t}\delta(\sigma-\sigma(\beta)) = \sum_{m'}\sum_{\alpha\beta}\delta(\sigma-\sigma(\beta))\sum_{\sigma'}\sum_{\alpha'\beta'}\cdot \delta(\sigma'-\sigma(\beta'))\frac{V^{(0)}_{mm'}}{\tau^{\rm mic}_{m,m'}}a_T(\mathcal{H}^{\rm mic}_{m\alpha\beta};\mathcal{H}^{\rm mic}_{m'\alpha'\beta'})\ p^{\rm mic}_{m'\alpha'\beta'}$$

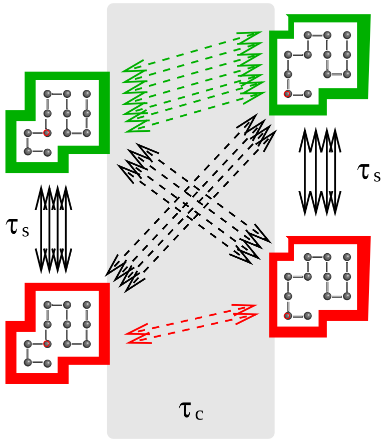

FIG. 12: Two types of moves, with different characteristic times, are considered in the dynamics: the protein monomer move (dashed arrows) or the solvent configuration transition (other arrows). For this last event, the connection between solvent configurations of the same macroscopic level do not affect the kinetics. The solvent transition is supposed to have smaller relaxation time than the monomer move. The relaxation times of the connection are defined as follows: we solve eqs.3 and 5 for an isolated connexion between two states $(m\sigma)$ and $(m'\sigma')$ leading to: $p^{\mathrm{mic}}_{m\alpha\beta}(t) = p^{(\infty)}_{m\alpha\beta} + [p_{m\alpha\beta}(0) - p^{(\infty)}_{m\alpha\beta}]\exp(-t/\tau^{\mathrm{mic}}_{m,m'})$. $\tau^{\mathrm{mic}}_{m,m'}$ is chosen as τ_c if the connexion is between two different protein structures and is chosen as τ_s if only water moves ($\tau_s \ll \tau_c$). The set of one monomer moves considered here, is the corner flip (shown in this figure) and the tail move (not shown).

The evolution equation is rewritten :

$$\frac{\mathrm{d}\mathcal{P}^{\mathrm{mac}}_{m\sigma}(t)}{\mathrm{d}t} = \sum_{m'}\sum_{\sigma'} Y_{m\sigma,m'\sigma'}\ \mathcal{P}^{\mathrm{mac}}_{m'\sigma'} \qquad (6)$$

with

$$Y_{m\sigma,m'\sigma'} = g_{m\sigma}\,\frac{V^{(0)}_{mm'}}{\tau^{\mathrm{mic}}_{m,m'}}\ a_T(\mathcal{H}^{\mathrm{mac}}_{m\sigma};\mathcal{H}^{\mathrm{mac}}_{m'\sigma'}) \qquad (7)$$

In addition, we mention that this result also satisfies the following balance equation: $Y_{m\sigma,m'\sigma'}\mathcal{P}^{\mathrm{eq}}_{m'\sigma'} = Y_{m'\sigma',m\sigma}\mathcal{P}^{\mathrm{eq}}_{m\sigma}$.

[1] Y. Xu, P. Purkayastha, and F. Gai, J. Am. Chem. Soc. **128**, 15836 (2006).
[2] Science **309**, 86 (2005).
[3] C. Levinthal, J. Chim. Phys. **65**, 44 (1968).
[4] K. A. Dill, S. B. Ozkan, T. R. Weikl, J. D. Chodera, and V. A. Voelz, Current Opinion in Structural Biology **17**, 342 (2007).
[5] P. Kim and R. Baldwin, Ann. Rev. Biochem. **59**, 631 (1990).
[6] S. Jackson, Folding Design **3**, 81 (1998).
[7] J. Kubelka and W. A. Hofrichter, J. andEaton, Curr. Opin. Struct. Biol. **14**, 76 (2004).
[8] J. Schonbrun and K. Dill, Proc. Natl. Acad. Sci. USA **100**, 12678 (2003).
[9] A. Matouschek, K. Jr., J.T., L. Serrano, and A. Fersht, Nature **340**, 122 (1989).
[10] K. Ghosh, S. Ozkan, and K. Dill, J.A.C.S. **129**, 11920 (2007).
[11] Y. Zhu, D. Alonso, K. Maki, C.-Y. Huang, S. Lahr, V. Daggett, H. Roder, W. Delgado, and F. Gai, Proc. Natl. Acad. Sci. USA **100**, 15486 (2003).
[12] H. Chan and K. Dill, Proteins: Struct. Funct. and Genet. **30**, 2 (1998).
[13] A. Akmal and V. Muoz, Proteins: Struct. Funct. and Genet. **57**, 142 (2004).
[14] P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992).
[15] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins Struct. Funct. Genet. **21**, 167 (1995).
[16] J. N. Onuchic, N. D. Socci, Z. Luthey-Schulten, and P. G. Wolynes, Folding & Design **1**, 441 (1996).
[17] P. Wolynes, Proc. Natl. Acad. Sci. **94**, 6170 (1997).
[18] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, Proteins Struct. Funct. Genet. **32**, 1136 (1998).
[19] J. N. Onuchic and W. P. G., Current Opinion in Structural Biology **14**, 70 (2004).
[20] P. Clark, Trends in Biochemical Sciences **29**, 527 (2004).
[21] K. F. Lau and K. A. Dill, Macromolecules **22**, 3986 (1989).
[22] N. Gõ, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[23] A. Dinner, A. Šali, M. Karplus, and E. Shakhnovich, J. Chem. Phys. **101**, 1444 (1994).
[24] O. Collet, Europhys. Letters **53**, 93 (2001).
[25] H. S. Chan and K. A. Dill, Annu. Rev. Biophys. Chem. **20**, 447 (1991).
[26] H. S. Chan and K. A. Dill, J. Chem. Phys. **100**, 9238 (1994).
[27] H. S. Chan and K. A. Dill, Proteins Struct. Funct. Genet. **30**, 2 (1998).
[28] B. Derrida, Phys. Rev. B **24**, 2613 (1981).
[29] E. I. Shakhnovich and A. M. Gutin, Nature **346**, 773 (1990).
[30] A. Šali, E. Shakhnovich, and M. Karplus, J. Mol. Biol. **235**, 1614 (1994).

[31] A. Šali, E. Shakhnovich, and M. Karplus, Nature **369**, 248 (1994).
[32] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. USA **92**, 1282 (1995).
[33] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, Biochemistry **34**, 3066 (1995).
[34] P. L. Privalov, Crit. Rev. Biochem. Mol. Biol. **25**, 281 (1990).
[35] I. Nishi, N. Kataoka, F. Tokunaga, and Y. Goto, Biochemistry **33**, 4903 (1994).
[36] R. Kumar, A. Prabhu, D. Rao, and A. Bhuyan, J. Mol. Biol. **364**, 483 (2006).
[37] W. Kauzmann, Adv. Protein Chem. **14**, 1 (1959).
[38] A. Warshel and S. Lifson, J. Chem. Phys. **53**, 582 (1970).
[39] K. Dill, Biochemistry **29**, 7133 (1990).
[40] P. De Los Rios and G. Caldarelli, Phys. Rev. E **62**, 8449 (2000).
[41] P. De Los Rios and G. Caldarelli, Journal of Biological Physics **27**, 229 (2001).
[42] O. Collet, Europhys. Letters **72**, 301 (2005).
[43] H. Chan, La Physique au Canada **60**, 195 (2004).
[44] S. Garde, G. Hummer, A. E. Garcia, M. E. Paulaitis, and L. R. Pratt, Phys. Rev. Lett. **77**, 4966 (1996).
[45] G. Hummer, S. Garde, A. E. Garcia, M. E. Paulaitis, and L. R. Pratt, J. Phys. Chem. B **102**, 10469 (1998).
[46] S. Garde, A. E. Garcia, L. R. Pratt, and H. Hummer, Biophysical Chemistry **78**, 21 (1999).
[47] M. A. Gomez, L. R. Pratt, G. Hummer, and S. Garde, J. Phys. Chem. B **103**, 3520 (1999).
[48] R. M. Lyndell-Bell and J. C. Rasaiah, J. Chem. Phys. **107**, 1981 (1997).
[49] L. Pratt and D. Chandler, J. Chem. Phys. **67**, 3683 (1997).
[50] S. Shimizu and H. Chan, J. Chem. Phys. **113**, 4683 (2000).
[51] H. Kaya and H. Chan, J. Mol. Biol. **326**, 911 (2003).
[52] H. Kaya, Z. Liu, and H. Chan, Biophysical Journal **89**, 520 (2005).
[53] A. Ferguson, Z. Liu, and H. Chan, J. Mol. Biol. **389**, 619 (2009).
[54] P. Fenimore, H. Frauenfelder, B. McMahon, and R. Young, Proc. Natl. Acad. Sci. USA **101**, 14408 (2004).
[55] H. Frauenfelder, P. Fenimore, G. Chen, and B. McMahon, Proc., Natl., Acad., Sci., USA **103**, 15469 (2006).
[56] V. Lubchenko, P. Wolynes, and H. Frauenfelder, J. Phys. Chem. **109**, 7488 (2005).
[57] N. Shenogina, P. Keblinski, and S. Garde, J. Chem. Phys. **129**, 155105 (2008).
[58] H. Frauenfelder, G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. MacMahon, I. R. Stroe, J. Swenson, and R. D. Young, PNAS **106**, 5129 (2009).
[59] J. M. Zanotti, G. Gibrat, and M. C. Bellisent-Funel, Phys. Chem. Chem. Phys. **10**, 4865 (2008).
[60] W. Doster, S. Cusack, and W. Petry, Nature **337**, 754 (1989).
[61] N. Muller, Acc. Chem. Res. **23**, 23 (1990).
[62] B. Lee and G. Graziano, J. Am. Chem. Soc. **118**, 5163 (1996).
[63] K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill, J. Chem. Phys. **111**, 8000 (1999).
[64] A. Ben-Naïm, J. Chem. Phys. **54**, 3682 (1970).
[65] J.-P. Becker and O. Collet, Journal of Molecular Structure. Theochem **774**, 23 (2006).
[66] C. Chipot, B. Maigret, and A. Pohorille, Proteins Struct. Funct. Genet **36**, 383 (1999).
[67] B. Lee and F. M. Richards, J. Mol. Biol. **55**, 379 (1971).
[68] D. Eisenberg and A. McLachlan, Nature **319**, 199 (1986).
[69] L. Wesson and D. Eisenberg, Protein Science **1**, 227 (1992).
[70] T. Ooi, M. Ootabake, G. Nemethy, and H. Scheraga, P.N.A.S. **84**, 3086 (1993).
[71] O. Collet and S. Premilat, International Journal of Peptide and Protein Research **47**, 239 (1996).
[72] S. Premilat and O. Collet, Europhysics Letters **39**, 575 (1997).
[73] O. Collet, J. Chem. Phys. **129**, 155101 (2008).
[74] O. Collet, S. Premilat, B. Maigret, and H. A. Scheraga, Biopolymers **42**, 363 (1997).
[75] Z. Liu and H. Chan, Physical Biology **2**, S75 (2005).
[76] C. Hardin, M. P. Eastwood, M. Prentiss, Z. Luthey-Schulten, and P. G. Wolynes, J. Comp. Chem. **23**, 138 (2002).
[77] C. Hardin, T. V. Pogorelov, and Z. Luthey-Schulten, Current Opinion in Structural Biology **12**, 176 (2002).
[78] G. Privalov and P. Privalov, Methods Enzymol. **323**, 31 (2000).
[79] P. Faisca and K. Plaxco, Protein Sci. **15**, 1608 (2006).
[80] M. Cieplak, M. Henkel, and J. Banavar, cond-mat/9806313 (1999).
[81] M. Cieplak, M. Henkel, J. Karbowski, and J. Banavar, Phys. Rev. Lett. **80**, 3654 (1998).
[82] O. Collet, Phys. Rev. E **67**, 061912 (2003).