Chapter 1

Protein folding and stability

1.1 The protein folding problem

One of the great challenges in protein science is the folding problem, that is the task to predict the protein native structure solely from the amino acid sequence. This has become a major research field as a consequence of the increasing amount of sequence data acquired by DNA analysis and various genome projects.

The acquisition of sequence information is a fast and straightforward procedure. On the other hand the experimental determination of the three dimensional structure assumed by a protein is slow and limited to proteins that can be crystallized in a suitable form or are sufficiently small to be solved by solution or solid state NMR. Therefore, algorithms capable to translate the linear information given by the protein sequence into the spatial information defining the structure are required.

Besides structure prediction, the other fundamental issue pursued by protein science is the *de novo* design of proteins and enzymes with specified activities. This design problem is related to the folding problem. Both tasks are based on finding the most stable fold of a sequence, natural or designed. Furthermore, this fold must be kinetically accessible. If an enzyme is designed, the active site must possess all features required for specific binding, without compromising stability.

The so called Levinthal paradox [1] is traditionally invoked to explain why the protein folding problem is so challenging. Considering, according to Anfinsen's hypothesis [2] (see section 1.3), that the native state of a protein corresponds to a global free energy minimum, folding amounts to find this minimum among a huge number of conformations. One can estimate that for a protein given by 100 amino acids, if each amino acid can assume for instance three different conformations, there are 3^{100} or 10^{47} states available. If the protein should find its native state by random search, visiting one state each picosecond, this search could take longer than the age of universe. Proteins usually fold in a time range from microseconds to seconds.

Theory is still far from solving the protein folding problem in general, because of an intrinsic

limit beyond computational capability. One could formulate the Levinthal paradox in terms of computing power. If the computation of the energy of one single conformation took one picosecond, the native state of a protein could not be calculated by random search. However, there has been a significant progress in protein science since the pioneering work of the early 70s and the current view of the folding problem suggests a solution of the Levinthal paradox [3]. The theoretical approach to protein folding aims at finding simplified descriptions of proteins which take into account only the important degrees of freedom. On the other hand protein engineering offers an effective experimental route to describe factors which govern protein stability. In this chapter the basic concepts both from the experimental and theoretical point of view will be outlined.

1.2 Native and denatured state

The native state of a globular protein is defined as the small ensemble of compact conformations, reached under folding conditions *in vivo* or *in vitro*, in which the protein is stable and performs its action. This state is characterized by a low amount of entropy, because all protein atoms are kept by mutual interaction in a well defined geometry. Intramolecular enthalpy is also relatively low because many attractive interactions are satisfied. Solvent molecules surrounding the folded protein in its native state take instead advantage of a large amount of entropy, since they are not involved in interactions with the hydrophobic protein interior.

By perturbing the folding conditions in some way, like increasing temperature, changing the solvent pH or adding a chemical denaturant, the protein unfolds and reaches a much more complex and heterogeneous ensemble of conformations, which is called the denatured or unfolded state. It is difficult to characterize the structural properties of the denatured state, since it strongly depends on the unfolding conditions. One can describe the ensemble by means of parameters like the radius of gyration or the hydrodynamic volume, which both define the average compactness. Experimental techniques like infrared and circular dichroism spectroscopy provide information on secondary structure elements. When a high percentage of secondary structure is lost, which is the case under strong unfolding conditions, the unfolded state assumes a *random coil* structure.

The denatured state is characterized by a high conformational entropy, because native interactions are lost and residues are free to assume a large set of arrangements. Intramolecular enthalpy may be higher because of loss of native interactions. The solvent molecules get in contact with hydrophobic parts of the protein chain which were buried in the native state. This fact causes water molecules to assume more ordered conformations in order to minimize the contact with nonpolar groups, what reduces the entropy of the solvent relatively to the folded state. Therefore, one can understand the entropy increase as a driving force for folding [4].

When restoring the folding conditions *in vitro*, the spontaneous refolding of a denatured protein is usually initiated. This means that the molecule is able to find the native state spontaneously.

However, this may not be the case for some large multidomain proteins, which cannot refold without assistance.

Many proteins show folding intermediates, which are non native in structure, but rather compact and partly folded. They are characterized by a reduced amount of tertiary interactions, by a rather large content of secondary structure, and a fluctuating hydrophobic core. Such states are termed *molten globule* states and are formed rather early during refolding, for instance in the dead time (< 1 ms) of a stopped-flow mixing experiment under refolding conditions.

1.3 Thermodynamics of protein folding

In 1973 Anfinsen [2] formulated the *thermodynamic hypothesis* for protein folding, stating that under physiological conditions the native state of a protein is stable since it is a global minimum of the free energy. Under denaturing conditions of increasing strength, this is not true anymore and the native state is destabilized. The concentration of native protein decreases until a point is reached, where the unfolded state dominates and is more populated and more stable.

Both states, unfolded and folded, can be characterized at equilibrium at temperature T by their Gibbs free energy (see section 3.6):

$$G_i = H_i - TS_i + pV_i \tag{1.1}$$

where i can be the native state N or the unfolded state D. The free energy is given by three terms: enthalpy H, the product of entropy S and temperature T and the product of pressure and volume. The last term is usually constant under experimental conditions. Therefore, it can be neglected when calculating differences. The *free energy of folding* is defined as difference of the Gibbs free energies between the denatured and the native state

$$\Delta G_{folding}(T) = \Delta H_{D-N}(T) - T\Delta S_{D-N}(T) \tag{1.2}$$

If this difference is positive, the native state is favored, else the unfolded state is more populated. One should keep in mind that not only the polypeptide chain, but also the surrounding solvent molecules contribute to the Gibbs free energy (see section 3.6). For instance the stabilizing entropy contribution due to the solvent is larger in the folded than in the unfolded state, as already mentioned. Another effect due to water is the increase upon unfolding of the specific heat at constant pressure, C_p. To a rough approximation, the value of ΔC_p of unfolding is about 12 cal/K mol per residue [5]. This increase is again due to the interaction between water molecules and non polar protein residues in the unfolded state, which causes the formation of ordered water structures like ice. The specific heat change ΔC_p rules the temperature dependence of both enthalpy and entropy of folding. Namely, given $\Delta H_{D-N}(T_1)$ and $\Delta S_{D-N}(T_1)$,

changing the temperature to the value T_2 one can write:

$$\Delta H_{D-N}(T_2) = \Delta H_{D-N}(T_1) + \Delta C_p(T_2 - T_1) \tag{1.3}$$

$$\Delta S_{D-N}(T_2) = \Delta S_{D-N}(T_1) + \Delta C_p \ln \frac{T_2}{T_1} \tag{1.4}$$

such that the free energy of folding at temperature T_2 becomes:

$$\Delta G_{folding}(T_2) = \Delta H_{D-N}(T_1) + \Delta C_p(T_2 - T_1) - T_2 \left[\Delta S_{D-N}(T_1) + \Delta C_p \ln \frac{T_2}{T_1} \right] \tag{1.5}$$

1.3.1 The unfolding transition

Protein denaturation is a process in which, starting from physiological folding conditions, the environment is progressively altered, for instance by increasing temperature, changing pH or adding a chemical denaturant like urea to the solution. This leads to a transition from the native state to a denatured state, dependent on the denaturation procedure. As mentioned above, the denaturation is usually reversible, that is the protein spontaneously returns to the original native structure when the folding conditions are restored. However, it can be irreversible for large multidomain proteins.

A first order transition from state A to state B is defined such that at each point of the transition both species are present at changing concentrations. The interconversion from one state to the other happens cooperatively, that is all parts contribute at the same time to the reaction. If concentration of one species is plotted to show the progress of the transition with changing conditions, the shape of the plot is sigmoidal.

Unfolding and refolding of small proteins are typically first order transitions, with A and B being the native and the unfolded state. This statement is made under the assumption that the reaction between native and unfolded state is a simple two-state equilibrium, for which no intermediates are present. A two-state transition can be identified in experiments if all spectroscopic probes, like circular dichroism, fluorescence and UV spectra, change simultaneously [4] as the equilibrium changes.

Calorimetric measurements can detect intermediates. There are also cases of apparent two-state transitions, where stable intermediates exist at low concentration of denaturant and are not observed by spectroscopic methods [6].

Some typical unfolding techniques are listed and briefly discussed here.

Thermal unfolding The free energy of folding in eq. (1.2) is under folding conditions positive, such that the native state is favored. If temperature increases, according to eq. (1.5), the negative part, proportional to entropy, will dominate when T_2 is sufficiently high, such that $\Delta G_{folding}$ becomes negative, leading to unfolding. The melting temperature T_m

is the temperature at which the free energy of folding, eq. (1.2), becomes zero:

$$T_m = \frac{\Delta H_{D-N}(T_m)}{\Delta S_{D-N}(T_m)} \tag{1.6}$$

Thermal unfolding can be measured by differential scanning calorimetry.

Solvent denaturation The effect of denaturants like urea or GdmCl is that they solubilize the protein chain, backbone and hydrophobic side chains. This is due to a double action of the denaturant. On one hand, it directly interacts with the protein groups; on the other hand, denaturant molecules also perturb the water solvent, which in turn alters the water-protein interaction [7]. To a first approximation the free energy of transfer of side chains and polypeptide backbone from water to a denaturating solvent is linear with respect to the concentration of denaturant [8]:

$$\Delta G_{D-N} = \Delta G_{D-N}^{H_2O} - m_{D-N}[\text{denaturant}] \tag{1.7}$$

The denatured state is more exposed to solvent than the native state. Therefore, the former is stabilized by the presence of denaturant, that is, the denaturant energy coefficient m_{D-N} is positive. Each residue of the protein contributes to the value of m_{D-N} according to its fraction of solvent exposure gained by denaturation. Thus, the coefficient is an extensive quantity and is correspondingly larger for larger proteins with more residues. A small value of m_{D-N} indicates that the surface area of the protein is not strongly changing upon denaturation, which means that the protein is not fully unfolded.

pH induced denaturation Acidic and basic residues buried in proteins have often altered pK_a's values, due to the strong interactions with other charges in a low dielectric medium like the protein interior (see chapter 3). For instance an acidic side chain involved in a salt bridge with a basic residue might have a pK_a far below the standard solution value. Nevertheless, when the pH of the solution decreases, at some point the acidic group becomes protonated, which may favor unfolding. Given a single titratable group in the protein, the relationship between its dissociation constant and the folding equilibrium can be described by a thermodynamic cycle (see fig. (1.1)). If the proton dissociation constant of the titratable group in the native state is K_a^N and in the denatured state is K_a^D, and the equilibrium constant for folding in the unprotonated state is K_{N-D} and in the protonated state is $K_{N-D(H^+)}$, then the equilibrium constants are related by:

$$\frac{K_{N-D(H^+)}}{K_{N-D}} = \frac{K_a^N}{K_a^D} \tag{1.8}$$

The thermodynamic cycle must be expanded if more titratable groups are present. The contribution to the free energy of folding due to changes in the protonation state of titratable groups can be written as follows, if the mutual interaction of the titratable residues

can be ignored :

$$\frac{\partial \Delta G_{D-N}}{\partial pH} = 2.3RT[Q_D(pH) - Q_N(pH)] \tag{1.9}$$

where $Q_i(pH)$ is the number of bound protons in the i-th state of the protein [8].

Destabilization at low pH also occurs because the protein increases its net positive charge, which may provide additional repulsive electrostatic interactions.

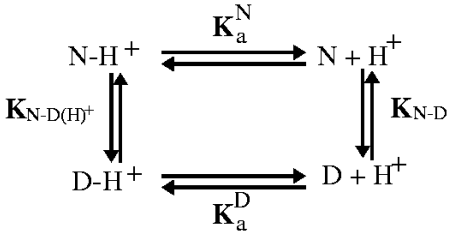

Figure 1.1: Thermodynamic cycle explaining pH induced unfolding. From [4].

1.4 Folding kinetics

When no folding intermediates are present, or if they are not detectable, a protein folds showing a monoexponential time evolution, typical for a two state kinetics. This is usually the case for small globular proteins with less than 100 amino acids [9, 10]. The equilibrium constant K_D for the folding-unfolding reaction:

$$N \leftrightarrow D \qquad K_D = \frac{[N]}{[D]} \tag{1.10}$$

is equal to the ratio of the microscopic rate constants for unfolding (k_{ND}) and for refolding (k_{DN}):

$$K_D = \frac{[N]}{[D]} = \frac{k_{ND}}{k_{DN}} \tag{1.11}$$

Therefore, protein stability is related to the folding rate:

$$\Delta G_{\text{unfolding}} = RT \ln K_D = RT \ln \frac{k_{ND}}{k_{DN}} \tag{1.12}$$

and the observed rate constant, which is inverse of the decay time τ, is defined:

$$k_{obs} = \frac{1}{\tau} = k_{ND} + k_{DN} \tag{1.13}$$

Denaturants affect the rate constants for folding and unfolding. The logarithm of the rate

constant for unfolding k_{ND} is found to change linearly with the concentration of denaturant at $[\text{denaturant}] > [\text{denaturant}]_{50\%}$ at concentration higher than 50% :

$$\ln k_{ND} = \ln k_{ND}^{H_2O} + m_D[\text{denaturant}] \tag{1.14}$$

while for some proteins [4] the rate constant for folding k_{DN} follows a similar relationship for low concentration $[\text{denaturant}] < [\text{denaturant}]_{50\%}$:

$$\ln k_{DN} = \ln k_{DN}^{H_2O} - m_N[\text{denaturant}] \tag{1.15}$$

These two equations form a V-shaped kinetic curve, called *chevron plot*, constructed by combining the two rate constants:

$$\ln k_{obs} = \ln(k_{ND} + k_{DN}) = \ln[k_{DN}^{H_2O} \exp(-m_N[\text{denaturant}]) + \mathrm{k}_{\mathrm{ND}}^{\mathrm{H_2O}} \exp(+\mathrm{m_D}[\text{denaturant}])] \tag{1.16}$$

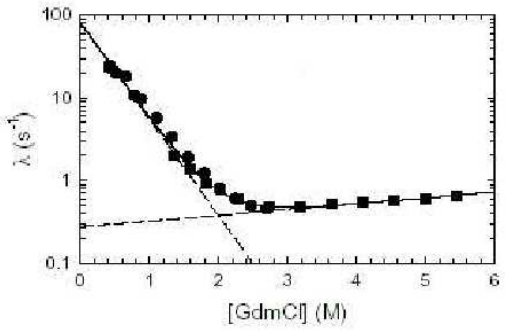

Figure 1.2: A chevron plot, showing the linear dependence of observed rate of folding or unfolding on denaturant concentration. The data refer to cold shock protein CspB from *Therotoga maritima* in GdmCl. From [11].

1.5 Transition state and Φ-value analysis

Given a monoexponential reaction scheme, the reactants must overcome a barrier to be transformed into the products. All treatments of chemical kinetics base on the hypothesis that the reaction rate depends on the height of this free energy or enthalpy barrier E_A, and on a probability factor that accounts for the number of accessible states for the molecule in the ground state versus the transition state. Since thermal motion is required to cross the energy barrier, the reaction rate is proportional to a Boltzmann factor.

According to the *transition state theory* the activated state is a quasi-thermodynamic state, named transition state, which is in a sort of fast equilibrium with the ground state [12]. The rate constant is given by the corresponding virtual equilibrium constant, multiplied by the frequency (ν) of the ground state and by the transmission coefficient (κ) of passing over the energy barrier. If the activated or transition state X has a free energy which is higher than the ground state by an amount ΔG_{X-D}, the rate constant is given by:

$$k_{DN} = \kappa\nu \exp(-\Delta G_{X-D}/RT) \tag{1.17}$$

Transition state theory was originally developed for reactions of small molecules in the gas phase, and it has been questioned whether it is well suited for describing reactions that occur in solution and involve the simultaneous formation and breakage of many weak interactions, as in protein folding [13]. Kramers' theory [14] provides an alternative formalism specifically for reactions in solution. A chemical reaction from a reactant D over the activated state X to the product N is modeled as a diffusive passage over a barrier:

$$k_{DN} = \tau_{NX}^{-1} \exp(-\Delta V/RT) \tag{1.18}$$

The Boltzmann factor in the rate expression in (1.18) depends on the height of the potential energy barrier and not on the free energy. The prefactor τ_{NX} depends on the frequency of the system in the ground state N and the frequency of escape from the activated state X, which reflects the local mobility of the system in X. This mobility depends on the friction with the solvent, represented by the macroscopic viscosity η . Therefore, if a reaction is modeled according to Kramers' theory, the rate depends on solvent viscosity. In the simplest case the dependence is inversely linear:

$$k^{-1} \propto \eta \tag{1.19}$$

It is a matter of debate whether the folding reaction is diffusion controlled, that is rate-limited by the diffusion of the polypeptide parts [12]. In such a case the folding rate should depend on the solvent viscosity. The folding of the bacterial cold shock protein CspB was proposed to be a diffusional process [13], due to the observed inverse dependence of folding rate on viscosity. On the other hand such a dependence may arise also from frictional effects at the transition state instead of being due to the global chain collapse [15].

The transition state theory provides a direct relationship between rate constant for folding and free energy differences between the ground and the activated state (see eq. 1.17). The main use of this formalism is related to the comparison of folding rates between wild type proteins and mutants. When calculating activation free energy changes upon mutation the prefactor $\kappa\nu$ cancels. If a protein folds with a rate k_f and a mutant with a rate k'_f, then the change in

activation free energy is given by:

$$\Delta\Delta G_{X-D} = \Delta G_{X-D} - \Delta G'_{X-D} = RT \ln \frac{k'_f}{k_f} \tag{1.20}$$

The method of Φ-value analysis, introduced by Fersht [16] and currently used in a wide range of experiments, is based on eq. (1.20). This procedure is aimed at obtaining the transition state structure for folding at the level of individual residues, by comparing the activation energies and the free energies of folding of different mutants of a protein. A single mutation of a protein that reduces its stability leads to a decrease -in absolute value- of the free energy of folding. Also the folding kinetics of the mutant protein can be monitored to determine the folding and unfolding rates and to provide activation energies. An increase of activation energy of folding upon mutation is related to a destabilization of the transition state. In fact, if the native state and the transition state are destabilized by the same energy amount, one can suppose that the mutated residue in the wild type protein is involved in native interactions which are already formed at the transition state. This can be confirmed if also the neighboring residues show upon mutation a similar behavior. If instead no change in transition state stability is produced by a mutation, which affects the stability of the native state, the native contacts of that residues are likely to be formed after passing the activation barrier during folding. This analysis is formalized in terms of Φ-values [4]. Let us suppose that the mutation destabilizes the folded structure by $\Delta\Delta G_{N-D}$, measured relative to the unfolded state. Then, if the free energy of a transition state, measured relative to the unfolded state, changes by $\Delta\Delta G_{X-D}$, Φ_F is defined as:

$$\Phi_F = \frac{\Delta\Delta G_{X-D}}{\Delta\Delta G_{N-D}} = \frac{RT \ln \frac{k'_f}{k_f}}{\Delta\Delta G_{N-D}} \tag{1.21}$$

with the expression on the right derived from eq. (1.20). Let us suppose that upon mutation the energy of native, denatured and transition state changes in the following way:

$$\begin{aligned} G_D &\rightarrow G'_D = G_D + \alpha \\ G_N &\rightarrow G'_N = G_N + \beta \\ G_X &\rightarrow G'_X = G_X + \gamma \end{aligned} \tag{1.22}$$

Then one has:

$$\Phi_F = \frac{G'_D - G'_X - G_D + G_X}{G'_D - G'_N - G_D + G_N} = \frac{\alpha - \gamma}{\alpha - \beta} \tag{1.23}$$

A Φ_F value of 0 means that the energy of the transition state is affected by the mutation by the same amount as the denatured state ($\alpha = \gamma$). This situation is described in the left diagram in fig. (1.3). One can infer that in this case the protein structure in the neighborhood of the mutated residue is unfolded at the transition state. On the other hand, a Φ_F value of 1 means that the energy of the transition state is affected by the same amount as the native state (or $\gamma = \beta$,

see right diagram in fig. (1.3)). This means that the structure made by the mutated residue is folded at the transition state as much as it is in the native state.

Fractional Φ_F values, and even values larger than 1 or smaller than 0 may also occur and are usually more difficult to interpret. Fractional Φ_F values between 0 and 1 can result either from specific residue interactions that are weaker in the mutant at the transition state or from the presence of a mixture of states, some with the interactions fully formed, others with the interactions fully broken. A number of point mutations is then required in order to discriminate between these two possibilities.

Another problem is that, in general, there is no linear relationship between a Φ_F value and the extent of structure formation. From a fractional Φ_F value, one can only conclude that the native structure is basically present, although weakened.

Despite these aspects, the method is in practice very efficient in determining structural properties of the transition state of folding proteins.

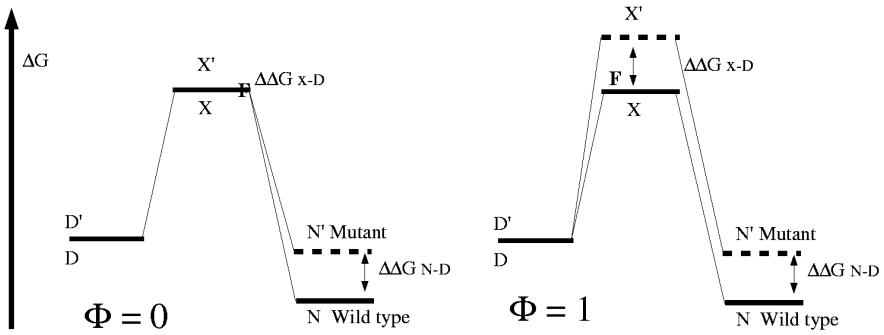

Figure 1.3: Φ-value analysis. From [4]. Free energy diagrams showing two possible scenarios for the effect of a single mutation on a folding protein. The energies of the denatured state for wild type and mutant are superimposed in the plot, but this does not affect the analysis [4]. On the left the mutation affects the transition state as much as it affects the denatured state (this is shown as no change at the transition state due to the superposition of energies for wild type and mutant denatured state). This leads to a Φ_F value of 0. On the right, the mutation affects transition and native state by nearly the same amount, which gives a Φ_F value of 1.

1.6 Folding mechanisms

The Levinthal paradox suggested that an unbiased search for the native conformation through an astronomical number of states would never be successful. Therefore, it was argued that there must be specific pathways to reduce the large number of possibilities during folding.
The main problem in folding theories was how to reconcile the simultaneous formation of stable secondary and tertiary structure. Three mechanisms were proposed to explain how the route to the native state proceeds. All of them simplify the scheme by decoupling the building of secondary structure elements from the collapse to the three dimensional structure. They are presented here:

The nucleation model [17] A group of neighboring residues starts forming native secondary structure, which propagates to the whole chain in a stepwise manner. Tertiary structure arises then as a consequence of secondary structure.

The diffusion-collision model [18] The secondary structure elements form early during folding and independently from tertiary structure. These elements then diffuse until they collide and successively coalesce to give the tertiary structure.

The hydrophobic collapse model [19, 20] The first event during folding is the collapse of the polypeptide chain, driven by the hydrophobic effect. After collapsing, the chain rearranges locally to form the native secondary structure.

First and second approach are both framework models [21], in which the formation of the correct secondary structure drives the subsequent arrangement of the tertiary conformation. The hydrophobic collapse model is at the other extreme: secondary structure forms as a consequence of the spatial collapse. A synthesis between these two classical views is provided by the **nucleation-condensation model**, which was proposed as folding mechanism for many small proteins after experiments on chymotrypsin inhibitor CI2. This was the second protein, after barnase, for which extensive investigation using Φ values analysis was done [21, 4]. The transition state of CI2 turned out to be an expanded structure in which secondary and tertiary structure are formed in parallel and there are no fully formed secondary structure elements yet. Molecular dynamics calculations also confirmed this picture [21].
The nucleation-condensation approach involves a nucleus consisting of neighboring residues, which forms a pre-native structure, subsequently stabilized by non-local interactions, i.e. interactions among residues that are far away from each other in sequence. This nucleus may be not present in the denatured state and does not need be completely formed in the transition state. The residues involved in the nucleus are identified by their Φ values, which are close to 1.
Since proteins largely differ in structure and size, folding mechanisms are likely to vary among different protein classes. Nevertheless, many small proteins, like for instance SH3 spectrin domain [12], seem to fold according to the nucleation-condensation mechanism. Also in more

complex multidomain systems secondary and tertiary structure may form in a concerted way via multiple transition states [22].

1.7 Energy landscape and funnel theory

The new theoretical approach to protein folding, which eventually leads to a solution of the Levinthal paradox, was developed in analogy to spin glass theory and polymer physics [23, 24, 25].

A folding protein is characterized by the coexistence of many competing interactions among amino acids. This leads to a situation called *frustration*, because the protein is not able to satisfy all interactions at the same time. The most stable state is then the structure maximizing the fraction of attractive interactions, which is the minimally frustrated structure. A useful concept related to this view is the *energy landscape* description. Given a set of coordinates describing the dependence on energy of a conformation, the free energy of the solvated protein as function of these coordinates defines a hypersurface, or energy landscape. A conformational transition of the protein, like the folding event, is then given by a specific trajectory or bundle of trajectories on this energy landscape. A smooth energy landscape allows rapid transitions, whereas a rugged energy surface has kinetic traps which slow down folding. Frustration is associated with a rugged energy landscape.

The picture offered by the Levinthal paradox is, in terms of landscape, a flat energy surface with a single narrow absolute minimum representing the native state, similarly to a golf course with a single hole. A random search through the flat surface, without any bias, would be successful in finite time only with an extremely low probability. A different extreme is represented by a rugged landscape with many maxima and minima without a preferred conformation: this would be the case of a random heteropolymer.

The energy landscape of a folding protein is described as a rugged surface sloped down from all directions into a global minimum, representing the native state (see fig. 1.4). It is explicitly assumed that there is a kinetic flow through a series of states that are progressively lower in energy.

At the top of the funnel the protein exists in a number of random states that have relatively high entropy and high enthalpy. Progress down the funnel is given by collapse and reconfiguration. The reconfiguration occurs as a motion through adjacent, thus geometrically similar, conformations, by a Brownian-like motion. The acquisition of native structure reduces the free energy and drives the protein towards the bottom of the funnel, while enthalpy is progressively reduced. The driving force is working against entropy, which decreases simultaneously with entropy.

The progress of folding can be described by the parameter Q_i, the fraction of native contacts in the state i:

$$Q_i = \frac{C(i)}{C(n)} \tag{1.24}$$

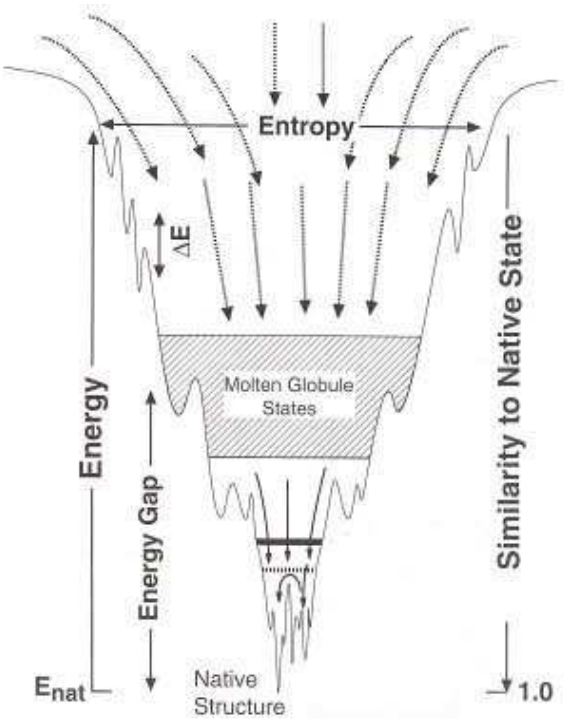

Figure 1.4: The folding funnel, from [23]. The height of the funnel represents enthalpy, which decreases going downhill (here referred to as energy, left arrow). The width of the funnel represents entropy. On the right the decrease of the fraction of native contacts Q is given by an arrow. The molten globule (see section 1.2) represents a compact denatured state. It is separated from the native state by an energy gap which guarantees the stability of the native structure.

where $C(n)$ is the number of contacts formed by residue pairs in the native state and $C(i)$ is the number of such contacts conserved in state i.

The molten globule state is rather compact but still highly non-native. The transition state is located at a higher value of Q_i, thus closer to the native structure. If Q does not increase uniformly, but in few large steps, then the protein folding occurs through nucleation.

After passing through the transition state, there is a single low energy state corresponding to the native conformation. This drives the protein to fold completely, while the absence of other relevant minima prevents from misfolding.