Research Article

Journal of Physical Organic Chemistry

Received: 4 October 2007, Revised: 30 November 2007, Accepted: 1 December 2007, Published online in Wiley InterScience: 24 January 2008

(www.interscience.wiley.com) DOI 10.1002/poc.1315

Proteins multi-funnel energy landscape and misfolding diseases

L. Cruzeiro[a]*

The problem of how a given α-amino acid sequence, in cells, most of the times, assumes the native structure, is a fundamental problem in Biology, known as the protein folding problem. Here, evidence is presented that suggests that the same α-amino acid sequence can assume several, very different, structures that have the same Gibbs energy as the native structure, in the same thermodynamic conditions. These results lend support to a multi-funnel Gibbs energy landscape for proteins in which Anfinsen's thermodynamic hypothesis alone cannot explain protein folding. How then do proteins fold? In a multi-funnel picture, transient deterministic forces are needed to select the native funnel from all other funnels that the protein can potentially fall into. The suggestion here is that such transient, deterministic forces arise from specific vibrational excited states (VES) that constitute the first step in protein folding and function (the VES hypothesis). An application of the VES hypothesis to calculate the energy absorbed, from water molecules, by a protein α-helix with extensive glutamine (gln) repeats is made, which can explain the structural instability of these proteins and their association with many misfolding diseases. Copyright © 2008 John Wiley & Sons, Ltd.

Keywords: energy landscape; protein folding; VES hypothesis; Davydov/Scott model

PACS numbers: 87.14.Ee; 87.15.He; 87.15.-v

INTRODUCTION

Proteins are the machines of life, they mediate most of the processes that take place in living cells. In order to work they must first acquire a well-defined average structure, the native structure. A fundamental question in Biology and Medicine, known as the protein folding problem, is how a given sequence of α-amino acids, in cells, most of the times assumes the native structure.[1,2] Since Anfinsen's renaturing experiments, which seem to indicate that the native structure is solely specified by the α-amino acid sequence, one answer to the protein folding problem has been the thermodynamics hypothesis, that is, the idea that the native structure of a protein is that which minimizes its Gibbs energy.[2] More recent theories of protein folding further assume that the Gibbs energy landscape of proteins is funnel-shaped[3–6] and that the native structure is that which is found at the bottom of the funnel. Since all spontaneous processes occur in a direction in which the Gibbs energy is minimized, according to the thermodynamic hypothesis, the folding of a given α-amino acid sequence into a protein native structure is simply driven by thermal agitation. In this view, the protein folding problem is essentially solved[6] and finding a protein structure from its sequence, in a computer, depends only on having an accurate potential to describe the interactions of its atoms with one another and on the availability of sufficient computer power.

Although the funnel hypothesis has been applied to numerous studies on protein folding, molecular dynamics simulations indicate the possibility of a multi-funnel Gibbs energy landscape (refer to References [7] and [8] especially note 22) and doubts remain about whether the interactions that stabilize protein structure are compatible with a funnel-shaped Gibbs energy landscape.[9] While a few systematic studies of the shapes of the energy landscape of small molecules and clusters have been attempted, which lead to both funnelled and multi-funnelled landscapes,[10–13] with the local topography of the energy landscape being related to the conformation of the molecule,[11] the conformational space and the Gibbs energy landscape of proteins continue to be too large to be probed in such a systematic manner, even with the most powerful computers. Instead, in Section 'The Energy Landscape of Proteins', a cursory study of the energy landscape of four proteins is made and evidence is presented for that the same α-amino acid sequence can assume several, very different, structures, with Gibbs energies that can be comparable to or indeed be lower than that of the corresponding native structure. The implication is that the Gibbs energy landscape of proteins is not funnel-shaped, but rather, multi-funnel-shaped and that the native structure is not acquired by the simple thermodynamic principle of Gibbs energy minimization. An alternative mechanism for protein folding is proposed which can explain why the proteins involved in misfolding diseases are less stable than other proteins (refer to Section 'The Ves Hypothesis and Misfolding Diseases'). The paper ends with a discussion of the experimental evidence for the mechanism proposed.

* *CCMAR and FCT, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal.*
E-mail: lhansson@ualg.pt

a *L. Cruzeiro*
FCT, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

Copyright © 2008 John Wiley & Sons, Ltd.

THE ENERGY LANDSCAPE OF PROTEINS

Four proteins from the Protein Data Bank[14] were selected, namely, 1QLX (104 α-amino acids),[15] 1I0S (161 α-amino acids),[16] 1AAP (56 α-amino acids)[17] and 1IGD (61 α-amino acids).[18] These proteins have different sizes, different biological origins and different functions. While the first is a fragment of the human prion,[15] the second is an oxireductase from archae,[16] the third is the protease inhibitor domain of Alzheimer's amyloid β-protein[17] and the fourth is a immunoglobulin binding domain of streptococcal protein G.[18] The main criterion for their selection was to have one representative of each of the four main classes of proteins, according to the CATH classification Scheme;[19] mainly α (1QLX), mainly β (1I0S), essentially structureless (1AAP) and α/β (1IGD). To probe their energy landscape, for each protein, three alternative structures were built by forcing it to assume the fold or part of the fold, of the other three. With this protocol, for each protein, three non-native alternative structures were defined and a total of 16 different systems was studied.

Initial coordinates for the 16 structures were obtained as follows. The experimental coordinates for the native structures of the four proteins selected were taken from the protein data bank[14] and energy minimized with the AMBER force field.[20] The energy minimized structures were then used as templates to build the three alternative structures by making each protein adopt a backbone fold that is found in one of the others. For example, in one case, the *first* 61 α-amino acids of 1QLX were forced to have the native backbone fold of *all* 61 α-amino acids of 1IGD and, in a second case, the 61 α-amino acids of 1IGD were forced to adopt the native backbone fold of the first 61 α-amino acids of 1QLX. In the first case, protein 1QLX is the structural template and part of its backbone fold is imposed on all of the protein 1IGD and, in the second case, the structural template, protein 1IGD, has fewer α-amino acids and only part of the backbone fold of 1QLX changes to that of 1IGD, with the rest remaining as it is in the native fold of 1QLX.

The 12 alternative structures generated in the manner described above were relaxed in order to eliminate all the steric interactions that such a protocol inevitably leads to, and, after relaxation, the structures was energy minimized in the same way as the native folds, using the AMBER force field.[20] All 16 minimum energy structures, 4 of which correspond to the native folds of the 4 proteins and 12 of which are alternative, hybrid, folds, can be seen in Fig. 1 and are available, in PDB format, upon request to this author. The native structures for the four proteins are found along the diagonal of Fig. 1 and each row in that Figure includes the native plus its three alternative structures, all in the same colour. On the other hand, all proteins in the same column have a common structure, taken from the fold of the native structure in that column.

In Fig. 1, the protein structures in each row are local energy minima of the same α-amino acid sequence. If protein structure is solely specified by a given α-amino acid sequence,[2] one might expect the energies of the non-native structures to be significantly higher than the energy of the corresponding native structure. In Table 1 are displayed the energies, calculated in kcal mol^{-1} with the AMBER force field,[20] of all the structures, ordered in the same manner as in Fig. 1. Table 1 shows that the alien protein structures have energies that are comparable to the energy of the corresponding native structure, and that can, in many cases, even be lower than that of the native structure. The

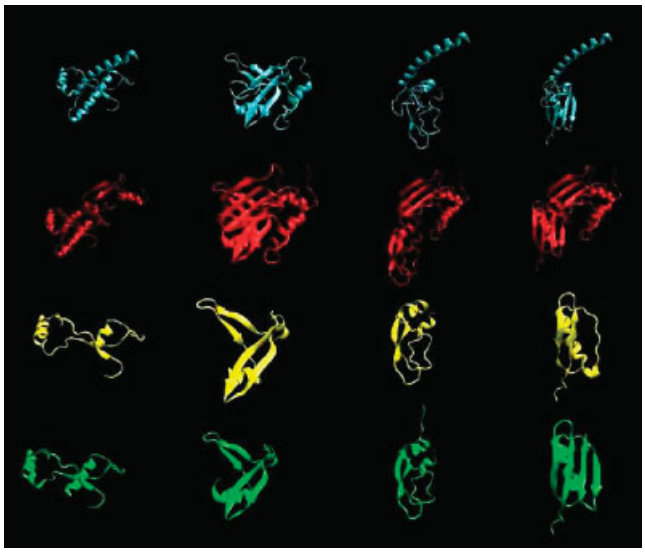

Figure 1. Native and hybrid protein folds after energy minimization with AMBER.[20] All proteins drawn in the same colour have the same α-amino acid sequence. The four native folds are displayed along the diagonal. The first row has the conformations for protein 1QLX (cyan), the second for 1I0S (red), the third for 1AAP (yellow) and the fourth is for 1IGD (green). Along each column, the non-native conformations are obtained by imposing the backbone fold or part of the backbone fold, of the native structure in that column on to the backbone of the other proteins. This figure was made with VMD.[44]

numbers in Table 1 represent only the potential energies of the different conformations and to have the corresponding relative Gibbs energies, we need to know the entropies associated with each structure. In order to have an insight into such entropic factors, all the structures were immersed in an explicit water bath, and using the molecular dynamics package NAMD,[21] they were heated to 300 K and equilibrated for 0.8 ns at that temperature.

Figure 2 shows the cumulative root mean square deviation per atom (RMSD) of the last 0.2 ns with respect to the initial structure. The plots show that an initial fast change takes place in approximately 10 ps, which is followed by slower changes over the next 0.2 ns. The plots also show that the native structures fluctuate less than the non-native structures, that is, that the entropy of the non-native structures is larger and that their Gibbs energy is lower. One reason for the consistently higher entropy of the alternative structures studied here is that they are less compact, more open structures, than the corresponding native structures. This is particularly obvious in two of the alternative

Table 1. Conformational energies of minimized structures (kcal mol^{-1})

Seq	Mainly α Str 1QLX	Mainly β Str 1I0S	Disordered Str 1AAP	α/β Str 1IGD
1QLX	−1900	−2008	−1948	−1989
1I0S	−2129	−2070	−2161	−2243
1AAP	−895	−875	−850	−1007
1IGD	−860	−846	−849	−895

Copyright © 2008 John Wiley & Sons, Ltd.

Journal of Physical Organic Chemistry

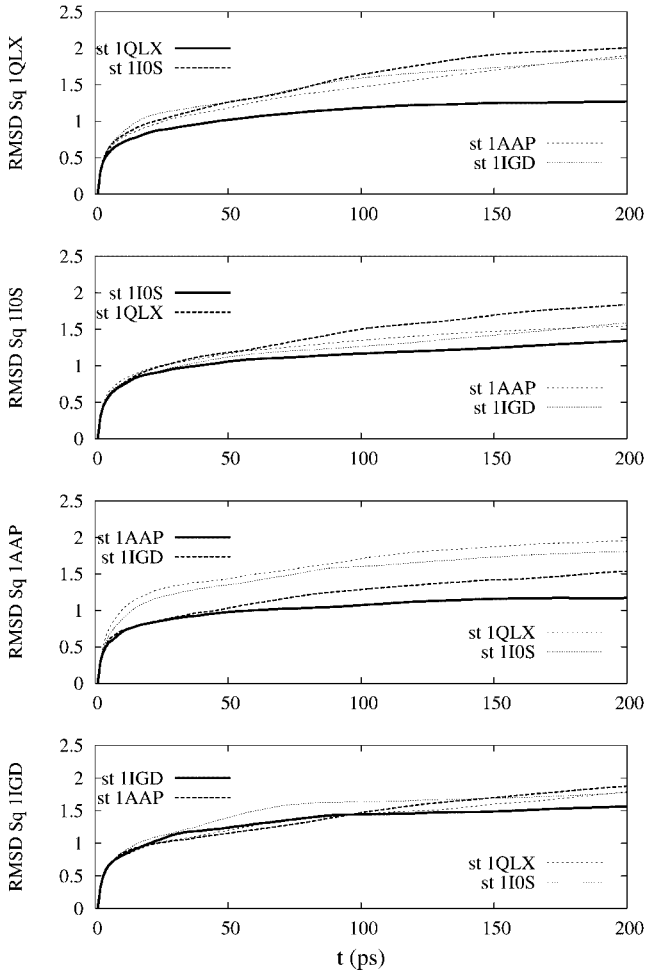

Figure 2. Variation of the cumulative RMSD with respect to the corresponding initial structure in the last 0.2 ns. Each plot is for a different protein and the order from top to bottom is as in Fig. 1. In each plot, the solid lines are for the native structures and the remaining three lines are for the proteins indicated by the keys. The RMSD values are in Å. Time is in picoseconds

structures of protein 1QLX (Fig. 1, last two structures in the first row) which have a helix protruding away from the main part of the protein, in the alternative structures of protein 1I0S (Fig. 1, off-diagonal structures in the second row), which are organized in two domains and in two of the alternative structures of proteins 1AAP and 1IGD (Fig. 1, first two structures in the third and fourth rows) whose residues are largely solvent exposed. Although it is possible to construct more compact non-native structures, the point here is to show that even in the absence of any specific fine tuning it is possible to get very different structures with a thermodynamic stability comparable to that of the native structure. This point is strengthened by inspection of all the structures at the end of the simulations which shows that their overall backbone folds have not changed after heating to 300 K and throughout the 0.8 ns simulation at that temperature. Although the simulations are too short for definitive conclusions to be made, they do indicate that the non-native structures studied here are at least metastable in normal conditions of temperature and pressure.

While longer simulations are being pursued, the simulations so far do favour a multi-funnel shaped Gibbs energy landscape, that is, each of the three alternative structures for each native structure displayed in Fig. 1 can be thought as associated with a funnel, distinct and widely separate from that of the native structure. In this multi-funnel, Anfinsen's thermodynamic principle *alone* leads with equal probability to each of the alternative structures, and is therefore incapable of explaining why the native structure of proteins is statistically preferred. Thus, in a multi-funnel energy landscape, a mechanism is needed to explain how, most of the times, the native funnel is selected, among all the thermodynamically equivalent funnels that are available. The suggestion of this author is that the transient forces needed to push a given α-amino acid sequence into the native funnel come from vibrational excited states (the VES hypothesis[22]). The next section presents the foundations of the VES hypothesis and also an application of this hypothesis to the structural instability of proteins associated with misfolding diseases.

THE VES HYPOTHESIS AND MISFOLDING DISEASES

The possibility that VES have a role in protein function was first proposed more than 30 years ago by McClare[23] in the context of a crisis in bioenergetics. This idea was taken up by Davydov[24,25] who was interested in muscle contraction, where the trigger and the energy donating reaction is the hydrolysis of adenosine-triphosphate (ATP). Davydov's assumption is that the first event after the hydrolysis of ATP is the storing of the energy released in the chemical reaction in a vibrational mode of the peptide group, known as amide I, which consists essentially of the stretching of the C═O bond. Davydov's studies of amide I energy storage and propagation were made in the context of a continuum model for the α-helix, for which it was possible to find analytical solutions.[24,25] Numerical studies of more realistic discrete models were done by Scott[26] and co-workers and a review of the field up to 1992 can be found in[26]. While Davydov and Scott's investigations were focussed essentially on the low temperature regime, numerical simulations at biological temperatures showed that, when the amide I vibration is properly treated as a quantum particle, its wavefunction is localized and the amide I vibration jumps from carbonyl to carbonyl as a Brownian particle.[27,28] In this stochastic manner, the energy that is stored at the active site can travel to other regions of the protein where it can be used for work. *The suggestion here is that the transient forces needed for the selection of the native funnel come from the release of the energy stored initially in the amide I modes,*[9] *that is, the amide I modes are the functionally important modes for folding.*

While a detailed description of how amide I vibrations can lead to conformational changes requires the development of models that include the decay of such vibrational states,[29,30] something that the Davydov/Scott model cannot do, in a previous publication,[22] the VES hypothesis has been used to explain why prions are structurally less stable than other proteins. Prions are the proteins whose misfolding lead to nervous system degeneration in cows (bovine spongiform encephalopathy) and in humans (variant Creutzfeldt–Jakob disease). In terms of their α-amino acid sequence, prions differ from other proteins in that they possess a much greater number of the α-amino acids glutamine (gln) and asparagine (asn).[31] As was pointed out before,[22] while all α-amino acids can have amide I excitations in their main chain, gln and asn are the only two α-amino acids that can have amide I excitations in their side chains.[32] This means

 Copyright © 2008 John Wiley & Sons, Ltd.

that amide I excitations that are travelling in the protein backbone can be diverted to the side chains of gln and asn and vice-versa. It was also pointed out before that the bending mode of water is nearly resonant with the amide I mode so that one quantum of excitation in the bending mode of a water molecule that is close to a protein carbonyl can jump to that carbonyl.[22] Huntington's is a misfolding disease that is correlated with greater than average numbers of gln α-amino acids. In particular, gln repeats greater than 37 are known to lead to the onset of the disease.[33] In this section, in order to further motivate the VES hypothesis, the ideas already applied to explain the structural instability of prions[22] are tested on a protein α-helix with 46 gln α-amino acids, that is, the Davydov/Scott Hamiltonian, extended to take into account the full three-dimensional structure of a protein α-helix and of its interactions with water, is used to calculate the energy transferred from the water to the α-helix in the presence and in the absence of gln. Proceeding as before,[22] the α-helix is immersed in an explicit water bath and equilibrated, for 2 ns at 300 K, with the AMBER force field.[20] After equilibration, 1000 snapshots of the helix plus water system are stored from a further 100 ps simulation, sampled at 0.1 ps intervals. Amide I vibrational energy transfer is modelled in the system constituted by the helix plus the 300 water molecules that were closest to the helix in the 1000 snapshots collected. A total of 391 sites, 45 of which are the backbone carbonyls, another 46 being the carbonyl groups in the gln side chains and another 300 being the closest water molecules, are available for the vibrational excitations in this system.

The extended Davydov/Scott Hamiltonian used to describe the storage and propagation of vibrational excitations (either amide I or bending mode) in the helix-water system is;[22]

$$\hat{H} = \hat{H}_{ex} + H_{at} + \hat{H}_{int} \tag{1}$$

where $\hat{H}_{ex}$, the exciton Hamiltonian, describes the transfer of the amide I excitation between carbonyls in the protein and of the bending mode of water between water molecules, as well as the exchange of an amide I excitation in the protein with a bending excitation in a water molecule; H_{at}, the atomic Hamiltonian, describes the fluctuations of all atoms in the system, as predicted by the AMBER force field[20] and $\hat{H}_{int}$, the interaction Hamiltonian as proposed by Davydov (Davydov, 1991), describes the interaction of the amide I excitation in a given C═O group with the deviations from equilibrium lengths of the hydrogen bond connected to it (if such a hydrogen bond exists, as explained below).

With the appropriate changes, the exciton Hamiltonian is essentially as before,[22] namely:

$$\hat{H}_{ex} = \sum_{n=1}^{N} \varepsilon_n \hat{a}_n^{\dagger} \hat{a}_n + \sum_{n<m=1}^{N} [V_{nm}(\hat{a}_n^{\dagger} \hat{a}_m + \hat{a}_m^{\dagger} \hat{a}_n)] \tag{2}$$

where $\hat{a}_n^{\dagger}(\hat{a}_n)$ are the creation(annihilation) operators for an vibrational excitation at site n and $N = 391$ is the total number of sites where a vibrational excitation can be found. In the first term, ε_n, for $n = 1, \ldots, 45$, is the energy of a non-interacting amide I vibration in the backbone of the gln helix, taken to be 1660 cm^{-1}, as in previous simulations.[22,24–26,28,34] For $n = 46, \ldots, 91$, ε_n is the energy of a non-interacting amide I vibration in the gln side chains, which is taken to be 1650 cm^{-1} [32] and, for $n = 92, \ldots, 391$, it is the energy of the bending mode of water, that is, 1640 cm^{-1}.[35] In the second term, $V_{nm} = \frac{1}{4\pi\varepsilon_0 k} \frac{|\vec{\mu}_n||\vec{\mu}_n|}{R_{nm}^3}[\vec{e}_n \cdot \vec{e}_m - 3(\vec{u} \cdot \vec{e}_n)(\vec{u} \cdot \vec{e}_m)]$ is the dipole–dipole interaction between sites n and m, $\vec{\mu}_n$ being the transition dipole moment associated with the vibrational excitation in site n, $\varepsilon_0 = 8.8542 \times 10^{-12}$ F m^{-1} being the permittivity of vacuum and k being the dielectric constant of the medium. As previously,[22] for the amide I excitation this dielectric constant was taken to be 10 and for the hydration waters it was taken to be 40. For the dipole–dipole couplings that correspond to an energy transfer between the helix and water this dielectric constant was taken to be the geometric average of these two values, namely, 20.

H_{at}, the atomic Hamiltonian, is the AMBER force field,[20] the same classical energy functional that has been used to determine the energy minimized structures in Fig. 1. This energy functional includes harmonic potentials for bond stretching and angle bending, a truncated Fourier series to represent torsions, a Lennard–Jones potential to represent non-bonded interactions and, most importantly, electrostatic interactions between atoms that are more than two covalent bonds away from each other. It is this potential that is used to select the 1000 snapshots of the helix-water system, mentioned above.

Finally, the interaction Hamiltonian, $\hat{H}_{int}$, includes the interaction already proposed by Davydov[24,25] and Scott,[26] namely, the change of the amide I energy with the length of the hydrogen bond connected with it, when such a bond exists (explained below):

$$\hat{H}_{int} = \chi \sum_{n=1}^{91} \left[\left(|\vec{R}_n^O - \vec{R}_m^N| - d_{eq} \right) cos(\theta_{nm}) \hat{a}_n^{\dagger} \hat{a}_n \right] + \sum_{n=92}^{N} \xi_n \hat{a}_n^{\dagger} \hat{a}_n \tag{3}$$

where $\chi = 62$ pN is the value determined from experimental data[26] and used in previous investigations[22,24–26,28] and $(|\vec{R}_n^O - \vec{R}_m^N| - d_{eq})$ is the deviation of the hydrogen bond length between the oxygen of the carbonyl group of the amide bond n with the nitrogen of the amide bond m, where $m \neq n$, from its equilibrium value, d_{eq}. For each carbonyl oxygen, n, there was at most one such hydrogen bond to another α-amino acid, and when the distance $|\vec{R}_n^O - \vec{R}_m^N|$, between the oxygen and the nitrogen atoms was greater than 3.3 Å, or the angle between C═O and H—N groups deviated more than 30° from a straight line, the corresponding hydrogen bond was assumed to be broken and the corresponding term in Equation (3) was zero. The factor $cos(\theta_{nm})$, where θ_{nm} is the angle between the C═O bond of the carbonyl and the H—N bond of the amine, was included to take into account the strong directionality of the hydrogen bond.

The summation in the first term of Equation (3) covers only the carbonyl sites. The water sites are taken care of by the second term, where ξ_n is a Gaussian random variable with a standard deviation of 50 cm^{-1} that represents the influence of the local hydrogen connections on the energy of the bending mode of the water molecules.

Vibrational energy transfer from the water to the 46 gln helix and vice versa is considered in two different systems. One is the system constituted by the helix plus the 300 closest waters and the second system is a helix with the same backbone structure but with no gln side chains. The first system has a total of 391 vibrational sites, as explained above, and the second one has a total of 345 vibrational sites (45 amide I sites in the backbone and 300 bending mode in the water molecules). The purpose is to compare the probability of vibrational energy transfer from the water to the protein α-helix when gln side chains are present and when they are absent. Two different initial conditions are

Copyright © 2008 John Wiley & Sons, Ltd.

Journal of Physical Organic Chemistry

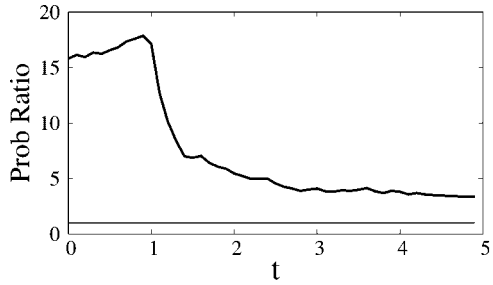

Figure 3. Enhancement by gln of the probability of finding vibrational energy in the protein α-helix when the energy is initially in the bending mode of water (thick line) and when the energy is initially in the helical backbone, as an amide I vibration (thin line). Time is in picoseconds

considered: in one the vibrational energy is initially in one of the waters closest to the helix, in the form of a bending mode, and in the other the vibrational energy is initially in one of the helix sites, in the form of an amide I vibration. For both initial conditions, we are interested in finding out how much of this vibrational energy is in the helix after a time *t*. Averages are made over 45 different initial sites in each of the 1000 helix-water conformations collected. Figure 3 displays the enhancement, due to the presence of gln side chains, in the energy that, on average, remains in the protein α-helix when the starting position was a water molecule (thick line) and when the starting position was a helix backbone carbonyl (thin line). A value of 1 means that the presence of gln does not influence the flow of energy from water to the helix, or vice versa, very much and a value greater than 1 means that more energy is found in the helix in the presence of gln side chains, than in their absence. Figure 3 shows that, in the presence of gln side chains, there is a marked increase in the amount of energy that is transferred from the bending mode of water to the amide I vibration in the protein α-helix, while the amount that goes from the helix backbone to the water is practically not influenced by the presence of gln. Most of the energy transferred initially from the water to protein is stored in the gln side chains, where it stays for about 1 ps. Then, most of this energy flows back to water again, but roughly one fifth of the energy initially absorbed, in the presence of gln, goes to the protein backbone, where it remains. The result is that, in the presence of gln residues, after 3 ps, there is approximately three times more energy in the protein than in the absence of gln. It is curious to note that while for the prion[22] the enhancement was, on average, less than 2, for the poly-glutamine helix studied here the enhancement varies between 3 and 17. According to the VES hypothesis, such a poly-gln helix is much more likely to undergo a conformational change than the prion. This may explain why Huntington's disease for individuals with long gln repeats sets in faster than prion diseases.

DISCUSSION

One aim of this work is to investigate whether protein folding can be understood by a Gibbs energy bias to the native structure.[2–6] To test the underlying idea that the native structure of proteins is a global minimum of their Gibbs energy, three alternative conformations to the native structure of four proteins were built and their energies compared to that of the native structure. Although evidence exists for a multi-funnel energy landscape,[7,8,10–13] it is the first time that the stability of structures of small proteins with completely alien backbone folds have been studied. If the α-amino acid sequence *alone* specifies protein structure,[2] one would not expect a sequence that folds to a given class of proteins to have an energy minimum when it is forced to assume the fold of another class, and even less would one expect this energy minimum to be as pronounced as that of its native structure. But Fig. 1 and Table 1 show that this happens for all the three alien folds of each of the four proteins studied. Furthermore, the alien folds studied here were maintained when the structures were heated to 300 K and also during a further 0.8 ns simulation at that temperature with an explicit water bath. Therefore, the results presented here favour a multi-funnel shaped Gibbs energy landscape for proteins in which each funnel is associated with an average structure that can be very different from the native, and yet be as thermodynamically viable as the native structure.

In spite of the theoretical prevalence of the thermodynamic hypothesis,[2–6] experimental evidence for a multi-funnel Gibbs energy landscape in the case of proteins was obtained already in 1968 by Levinthal[1] who found two forms of an alkaline phosphatase at 317 K, one active and the other inactive, synthesized at different temperatures, in mutants of *E. coli*. More recently, other cases have been found of proteins that can assume more than one structure in the same thermodynamic conditions.[36–40] While within a funnel-shaped Gibbs energy landscape it is difficult to explain the misfolding cases mentioned, within a multi-funnel Gibbs energy landscape they can be understood as cases in which a non-native funnel was selected. On the other hand, within a multi-funnel Gibbs energy landscape protein folding cannot take place solely by Gibbs energy minimization. Indeed, in a multi-funnel, the difficulty in determining the native structure from a given α-amino acid sequence is not just due to the size of the conformational space and the lack of computer power, but, more essentially, to the fact that the native structure is not a well-defined global Gibbs energy minimum.

How can proteins fold in a multi-funnel Gibbs energy landscape? The proposal here is that, before the energy minimization step, there must be a first step in which a specific funnel is selected.[9] This first step is a non-equilibrium kinetic mechanism for which there is *direct* experimental evidence in a few cases[38–40] and which was already proposed by Levinthal[1] who suggested that there are specific pathways for folding. Considering that such pathways can be characterized by intermediates one can say that the experimental evidence for a kinetic mechanism is indeed substantial[41,42] and may even include proteins that apparently follow a two-state process.[42,43]

But while there is much experimental evidence for a kinetic step in protein folding, there is as yet neither theoretical nor experimental insight about the drivers of this kinetic step. In the same way that the interactions that stabilize protein structure, which are represented in potentials such as AMBER,[20] lead to a multi-funnel shaped Gibbs energy landscape, they also lead, not to one, but to many possible pathways to each of these funnels. So, the forces that push proteins into the native funnel cannot arise from those interactions alone. The proposal by this author, follows McClare,[23] Davydov[24,25] and Scott,[26] in saying that the forces that lead to the selection of a specific funnel, in most cases that funnel being the native funnel, come from VES[9,22] and, in particular, from amide I vibrations. The idea is that protein folding and function involves a first step in which energy is stored in the

 Copyright © 2008 John Wiley & Sons, Ltd.

form of amide I vibrations (the VES hypothesis[22]) and that the release of this energy leads to the transient forces needed to select a funnel. According to the VES hypothesis, conformational changes are driven by amide I vibrations and, the more energy proteins store in this form, the more prone they are to undergo conformational changes, which in some cases can become harmful. This idea was successfully tested in a fragment of the human prion[22] and also here in a poly-gln helix. Indeed, as for the prion,[22] also the calculations for the poly-gln helix in Section 'The Ves Hypothesis and Misfolding Diseases' showed that the presence of gln side chains does lead to a marked increase in the energy stored in the form of amide I vibrations, which correlates positively with a higher probability of undergoing conformational changes. In the case of the poly-gln helix, these conformational changes can lead to the onset of Huntington's disease.

A complete picture of protein folding within a multi-funnel Gibbs energy landscape must include the nature of the transient forces responsible for the selection of the native funnel. Here, only a source of the energy for these forces, namely, VES, is put forward. While models capable of explaining the generation of these forces are being developed,[29,30] one conclusion of this work is that to understand how a given α-amino acid sequence, in a cell, folds, most of the times, to the native structure of the corresponding protein or indeed why it sometimes misfolds, one must know the nature of these forces.

Acknowledgements

This work was funded in part by the Foundation for Science and Technology (FCT, Portugal) and by POCI 2010 and the European Community fund, FEDER. Most of the computer simulations were performed at the Laboratory for Advanced Computing, (LCA), University of Coimbra, Portugal.

REFERENCES

[1] C. Levinthal, *J. Chim. Phys.* **1968**, *65*, 44–45.
[2] C. B. Anfinsen, *Science* **1973**, *181*, 223–233.
[3] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, *Proteins* **1995**, *21*, 167–195.
[4] J. N. Onuchic, Z. Luthey-Schulten, P. G. Wolynes, *Ann. Rev. Phys. Chem.* **1997**, *48*, 545–600.
[5] P. G. Wolynes, *Quart. Revs. Biophys.* **2005**, *38*, 405–410.
[6] M. Karplus, J. Kuriyan, *Proc. Natl. Acad. Sci. U S A* **2005**, *102*, 6679–6685.
[7] S. V. Krivov, M. Karplus, *Proc. Natl. Acad. Sci. U S A* **2004**, *101*, 14766–14770.
[8] T. Lazaridis, M. Karplus, *Science* **1997**, *278*, 1928–1931.
[9] L. Cruzeiro-Hansson, P. A. S. Silva, *J. Biol. Phys.* **2001**, *27*, S6–S9.
[10] O. M. Becker, M. Karplus, *J. Chem. Phys.* **1997**, *106*, 1495–1517.
[11] Y. Levy, O. M. Becker, *Phys. Rev. Lett.* **1998**, *81*, 1126–1129.
[12] D. J. Wales, M. A. Miller, T. R. Walsh, *Nature* **1998**, *394*, 758–760.
[13] P. N. Mortenson, D. J. Wales, *J. Chem. Phys.* **2001**, *114*, 6443–6454.
[14] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.
[15] R. Zahn, A. Liu, T. Luhrs, R. Riek, C. Von Schroetter, F. L. Garcia, M. Billeter, L. Calzolai, G. Wider, K. Wuthrich, *Proc. Natl. Acad. Sci. U S A* **2000**, *97*, 145–150.
[16] H.-J. Chiu, E. Johnson, I. Schroder, D. C. Rees, *Structure* **2001**, *9*, 311–319.
[17] T. R. Hynes, M. Randal, L. A. Kennedy, C. Eigenbrot, A. Kossiakoff, *Biochemistry* **1990**, *29*, 10018–10022.
[18] T. Gallagher, P. Alexander, P. Bryan, G. L. Gillilan, *Biochemistry* **1994**, *33*, 4721–4729.
[19] C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, J. M. Thornton, *Structure* **1997**, *5*, 1093–1108.
[20] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, P. A. Kollman, AMBER 6, University of California, San Francisco, **1999**.
[21] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comp. Chem.* **2005**, *26*, 1781–1802.
[22] L. Cruzeiro, *J. Phys.: Condens. Matter* **2005**, *17*, 7833–7844.
[23] C. W. F. McClare, *Ann. N. Y. Acad. Sci.* **1974**, *227*, 74–97.
[24] A. S. Davydov, *J. Theor. Biol.* **1973**, *217*, 559–569.
[25] A. S. Davydov, *Solitons in Molecular Systems*, 2nd edition, Kluwer Academic Publ., Dordrecht, **1991**.
[26] A. Scott, *Phys. Rep.* **1992**, *217*, 1–67.
[27] L. Cruzeiro-Hansson, *Europhys. Lett.* **1996**, *33*, 655–659.
[28] L. Cruzeiro-Hansson, S. Takeno, *Phys. Rev. E* **1997**, *56*, 894–906.
[29] P. A. S. Silva, L. Cruzeiro-Hansson, *Phys. Lett. A* **2003**, *315/6*, 447–451.
[30] P. A. S. Silva, L. Cruzeiro, *Phys. Rev. E* **2006**, *74*, 021920.
[31] M. D. Michelitsch, J. S. Weissman, *Proc. Natl. Acad. Sci. U S A* **2000**, *97*, 11910–11915.
[32] S. Krimm, J. Bandekar, *Adv. Prot. Chem.* **1986**, *38*, 181–364.
[33] J. F. Gusella, M. E. Macdonald, *Nat. Rev. Neurosci.* **2000**, *1*, 109–115.
[34] L. Cruzeiro-Hansson, *Phys. Lett. A* **1998**, *249*, 465–473.
[35] G. Sieler, R. Schweitzer-Stenner, *J. Am. Chem. Soc.* **1997**, *119*, 1720–1726.
[36] S. B. Prusiner, *Science* **1982**, *216*, 136–144.
[37] S. B. Prusiner, M. McCarty, *Annu. Rev. Gen.* **2006**, *40*, 25–45.
[38] D. Baker, J. L. Sohl, D. A. Agard, *Nature* **1992**, *356*, 263–265.
[39] J. L. Sohl, S. S. Jaswal, D. A. Agard, *Nature* **1998**, *395*, 817–819.
[40] Y. Tsutsui, L. Liu, A. Gershenson, P. L. Wintrode, *Biochemistry* **2006**, *45*, 6561–6569.
[41] S. W. Englander, *Annu. Revs. Biophys. Biomol. Struct.* **2000**, *29*, 213–238.
[42] D. J. Brockwell, S. E. Radford, *Curr. Opin. Struct. Biol.* **2007**, *17*, 30–37.
[43] H. Roder, W. Colón, *Curr. Opin. Struct. Biol.* **1997**, *7*, 15–28.
[44] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38.

Copyright © 2008 John Wiley & Sons, Ltd.