Nonlinear dynamics of secondary protein folding

Natalia G. Berloff
Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Cambridge, CB3 0WA

January 16, 2005

Abstract

We propose a simple phenomenological model that describes the conformational dynamics of proteins from the primary to the secondary structure. The folding pathway is determined by the local potential energy of individual amino acids in a protein chain, by the spring tension of the protein representing the internal hydrogen bonds that hold the helices and sheets together, and by the strength of nonlinear excitations that propagate through the protein backbone. Two opposite cases are considered with the length scale of an excitation being much larger or much smaller than the characteristic scale on which the conformational energy field varies.

Pacs: 87.15.-v, 87.15By, 05.45.-a, 41.20Jb
Keywords: Folding pathway, protein folding, nonlinear excitation, soliton, nonlinear Schrödinger equation.

1 Introduction

Proteins are made up from a set of 20 amino acids whose diversity derives from the R group. Given a protein's amino acid sequence, the pathway prediction problem is to determine the time ordered sequence of folding events that leads from the primary to the secondary and to the tertiary structure. Secondary structure is the shape assumed by each segment of a polypeptide and is promoted by hydrogen bonds, van Der Waals forces, electrostatic interaction and hydrophobic effects. The most common shapes of the protein folding are alpha (α) and beta (β). The α-helix is a right-handed helix averaging 3.6 amino acids per turn. The $\beta-$ribbon is a flat section of amino acids. Typically two or more $\beta-$ribbons associate into a $\beta-$sheet, formed by hydrogen bonds. Because of the nature of their R groups, certain amino acids are more commonly found in α-helices while others have a predisposition for β-sheets. The scenario of the secondary protein folding is believed to consist or several stages. Initially a secondary structure is formed around a group of amino acids

for which this is a ground (preferred) state and then extends to include adjacent amino acids that either favour the structure or have no strong disinclination towards it. This process terminates when the blocking amino acids are reached. This process is repeated along the polypeptide until the chain as a whole adopted its preferred secondary structure. The algorithms for predicting the secondary structure are based on statistical analysis of identifying the probability with which amino acids are located in each secondary structure and by studying the folding of small polypeptides. A prediction accuracy of such algorithms nevertheless do not exceed 75%. Moreover, the main mechanisms responsible for a structured folding pathway have not yet been identified. Understanding of such mechanisms is important, since protein misfolding has been identified as the cause of several diseases, such as Creutzfeld-Jacob disease, cystic fibrosis, hereditary emphysema and some cancers [1].

Nonlinear excitation such as breathers, solitons, intrinsic localised modes have been drawing increasing attention over recent years as the tools to transfer heat, charge, and energy in molecular chains. In their early works [2] Davydov and Kislukha investigated pulses of coupled intermolecular and lattice vibrations in biopolymers and showed that they can propagate as a solitary wave conserving form and velocity. Davydov's theory [3] described a single electronic or vibrionic excitation that propagates along a deformable molecular chain. In α-helical proteins the leading role in the energy transfer is played by amide-I vibrations (C=O) of atoms in the peptide groups (H–N–C=O). The α-helical protein structure consists of peptide groups bonded in periodic arrays in three spiralling chains forming a helical chain in which neighbouring peptide groups in each chain are linked by hydrogen bonds. The propagation of amide-I vibration induces longitudinal sound waves which provide a potential well, so that the coupled excitation propagates as a soliton. Davydov's theory is based on the Fröhlich Hamiltonian [4], which in a long-wavelength approximation or in continuum limit becomes the focusing nonlinear Schrödinger equation extensively studies in nonlinear optics. This equation is completely integrable in one dimension and possesses soliton solutions which propagate with the velocity that depends on the wave amplitude.

The question of whether nonlinear charge transport in DNA is possible has been a subject of a heated debate over the last few years. There is mounting experimental and theoretical evidence that DNA behaves as a conducting one-dimensional wire [5]. The DNA transport has been explained as meditated by polarons [6], electrons or holes [7], and solitons [8]. The theoretical predictions of Davydov's theory have received experimental confirmations. In particular, the measurements of vibrational relaxation rates in high intensity IR pump-probe experiments in myoglobin indicated that nonlinear excitations are long-lived and play a significant role in energy transfer in biomolecules [9]. It has also been demonstrated that the cooperation between the vibrations in the protein chain and the electron dynamics mediates long-lived localised excitations travelling along the β-sheet proteins [10]. Recently it has been shown that nonlinear excitations play an

important role in conformational dynamics by decreasing the effective bending rigidity of a biopolymer chain leading to a buckling instability of the chain [11]. This instability may induce the change in the ground state conformation resulting in conformational dynamics of biopolymers.

Recently [12] Caspi and Ben-Jacob have suggested that Davydov solitons, propagating through the backbone of the protein, can mediate the transition of a protein from a metastable conformation to its ground state conformation. The process in which conformational changes are induced by a propagation of a soliton along the molecular chain was termed as a soliton mediated conformational transition (SMCT). The thermally or otherwise induced soliton carries the energy that is transferred to the conformational field causing the transition from a metastable to globally stable state. It was suggested that for reversing this transition external interactions should be added.

The most reliable theoretical framework to tackle the conformational dynamics of biomolecules would be the *ab initio* quantum chemistry approach. However, numerical intensity makes these calculations prohibitive for realistic biological systems. The classical theories of protein folding describe this process as a nearly sequential series of discrete intermediates. Newly emerged the energy landscape theory considers folding as the diffusion of an ensemble of protein configurations over a low-dimensional free-energy surface [13]. Since quantum effects are excluded from these theories, they are not able to describe quantum tunnelling of the excitons, so a SMCT will not be generated in the simulations [14]. One may also note that that the notion of a SMCT is consistent with the concept of minimal frustration in which the interactions among the subunits of a biomolecule are mutually supportive and cooperatively lead to a low-energy structure [15].

The role of dynamics in allosteric regulation has been recently investigated [16]. In allosteric biomolecules the binding of one ligand in one part of the molecule affects the affinity of another ligand at a distant site. This includes single-domain proteins where phosphorylation at one site changes the structure of the protein in a remote site [17], allosteric connection between protein and ATP that bind to distant domains [18], and the effect tryptophan binding has on RNA bindings at a remote site [19]. It seems that nonlinear excitations, such as solitons, are able to propagate through biomolecule and act as entropic carriers of free energy of allostery and, therefore, can explain the actual pathways of allosteric change.

In this Letter we suggest a simple phenomenological model that describes the folding pathway, where the folding is the result of the interplay between the energy transfer from a solitary solution that travels along the backbone of a protein and string tension. We show that the addition of string tension allows the conformational structure to go from the globally stable state to metastable state.

2 The models

In what follows we consider both continuous and discrete models. In the continuous case the blocks of amino acids are modelled using the segments of length $\ell_i = x_i - x_{i-1}$ along the continuous line of length $L = \sum \ell_i$ with x being the position measured along the polypeptide. In the discrete representation each amino acid is represented by a point in conformational space.

We start with the continuous model. To describe the local conformation of the protein we adopt the representation suggested in [12], where function $\phi(x)$ is the local curvature of the conformation at position x. We assume that $\phi(x) = 1$ for $\alpha-$helix and $\phi = 0$ for $\beta-$sheet. Each protein subunit carries an internal excitation which can be characterised [2, 3, 11, 20, 21] by the complex amplitude $\psi(x)$ that represents a vibrational (amide-I or base-pair) or a polaron state. We therefore model these excitations following other authors by the nonlinear Schrödinger equation. The interactions between the soliton and the conformation field is taken in its simplest quadratic form [12] as

$$U(|\psi|, \phi) = \Lambda |\psi|^2 (\phi - \tfrac{1}{2})^2, \tag{1}$$

where Λ is a positive parameter that represents the intensity of the coupling between two fields. If Λ is too large, the solitary wave will loose its energy very quickly and stop. If the coupling constant, Λ, is too small, the soliton will not be able to transfer enough energy to take ϕ from one state to another over the energy barrier.

Similar to [12], the local potential energy of the conformational field ϕ will be modelled by a ϕ^4 double well potential

$$V(\phi) = C\Big(\phi - \gamma(x)\Big)^2 \Big(\gamma(x)^2 + 2\gamma(x)(\phi - 1) + \phi(3\phi - 4)\Big), \tag{2}$$

where C is a positive parameter that characterises the depth of the potential well and the energy difference between the minima of ϕ which is $\Delta V = C|2\gamma(x) - 1|$. In our model $\gamma(x)$ takes on values between 0 and 1 and varies along the polypeptide. This function represents the predisposition of the amino acid located at x to take the $\beta-$sheet conformation. For $\gamma(x) \in (\frac{1}{2}, 1)$ the stable conformation is $\beta-$sheet, for $\gamma(x) \in (0, \frac{1}{2})$ the stable conformation is $\alpha-$helix. For $\gamma(x) = \frac{1}{2}$ the amino acid is equally likely to take any conformational state. We shall model $\gamma(x)$ as a piecewise constant function so that

$$\gamma(x) = \gamma_i = \text{const}, \qquad \text{if} \quad x_{i-1} \le x < x_i, \tag{3}$$

where $\gamma_i \in [0, 1]$.

We shall also introduce the tension potential between neighbouring regions

$$T(\phi(x)) = \xi[(\phi(x) - \phi(x - \ell_i))^2 + (\phi(x) - \phi(x + \ell_i))^2], \qquad \text{if} \quad x_{i-1} \le x < x_i, \tag{4}$$

where ξ is a positive constant. $T(\phi)$ represents the potential energy arising from two adjacent amino acids being in two different conformational states.

The full Lagrangian density becomes

$$\mathcal{L} = \mathrm{i}\psi^*\partial_t\psi - |\partial_x\psi|^2 + |\psi|^4 + \tfrac{1}{2}m(\partial_t\phi)^2 - V(\phi) - U(|\psi|,\phi) - T(\phi) - (\partial_x\phi)^2. \qquad (5)$$

The first three terms in (5) represent a Lagrangian of the nonlinear Schrödinger equation. The last term of (5) represents the (local) tension of the protein and introduces the penalty to the local discontinuities in conformation angle. The Euler-Lagrange equations written for the Lagrangian (5) have the form

$$\begin{aligned} \mathrm{i}\partial_t\psi &= -\psi_{xx} + (\Lambda(\phi - \tfrac{1}{2})^2 - 2|\psi|^2)\psi \qquad (6)\\ m\partial_{tt}\phi &= -12C\phi(\phi-1)(\phi-\gamma(x)) - 2\Lambda|\psi|^2(\phi-\tfrac{1}{2}) + \phi_{xx} \qquad (7)\\ &\quad -2\xi(\phi-\phi(x-\ell_i)) - 2\xi(\phi-\phi(x+\ell_i)) - \Gamma\partial_t\phi, \end{aligned}$$

where the last term represents dissipation also included in [12]. The boundary conditions are free to allow the solitary wave to reflect from the protein end points

$$\psi_x(x=0) = \psi_x(x=L), \quad \phi_x(x=0) = \phi_x(x=L). \qquad (8)$$

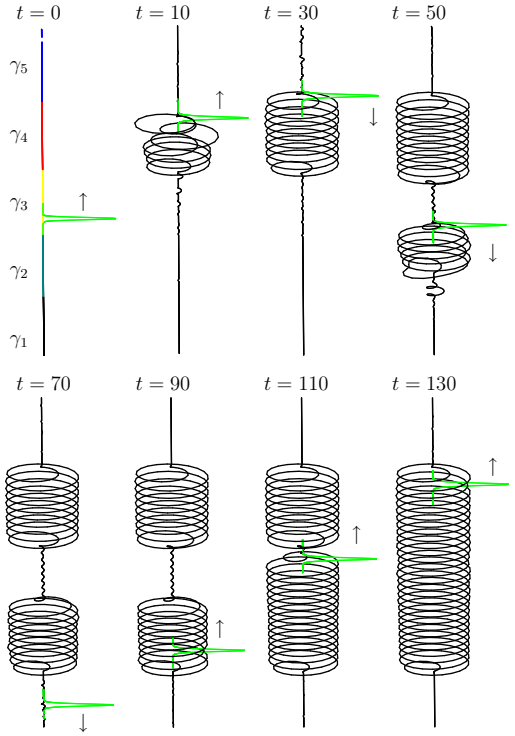
The solitary wave solution of the nonlinear Schrödinger equation (6) in the absence of the coupling with ϕ field ($\Lambda = 0$) is well-known [22]

$$\psi_s(x,t;\lambda) = \sqrt{\lambda}\,\mathrm{sech}(\sqrt{\lambda}x)\exp(i\lambda t), \qquad (9)$$

where λ is an arbitrary constant. This solution is immobile. By means of a Galilean transformation one directly finds the corresponding soliton solution of the nonlinear Schrödinger equation which moves with an arbitrary velocity.

Firstly, to elucidate the stages of the secondary folding we consider a sequence of five amino acid regions each of length $\ell_i = \ell = 200$ in our nondimensional units. We take $\gamma_1 = \gamma_5 = 0.9$, $\gamma_2 = \gamma_4 = 0.1$ and $\gamma_3 = 0.55$, so that the amino acids at the both end of the protein chain have a strong predisposition for $\beta-$sheet. The $\beta-$sheet is the stable configuration for the amino acids in the middle of the chain, but the energy barrier to the metastable $\alpha-$helix is rather small and their neighbours have a strong predisposition to $\alpha-$ helices. The initial solitary wave is (thermally or otherwise) excited inside the middle region. Figure 1 illustrates the folding pathway initiated by the excitation that the solitary wave transfers to the protein through a sequence of time snapshots. Soliton propagates upward transferring energy to the region #4 allowing its transition to the ground state (folding to $\alpha-$helix; see Fig. 1 at $t = 10, 30$). The solitary wave is reflected from the chain end points and traverses the entire polypeptide this time raising the region #2 to its ground (stable) state (Fig. 1 at $t = 50, 70$). When the soliton reached the bottom end

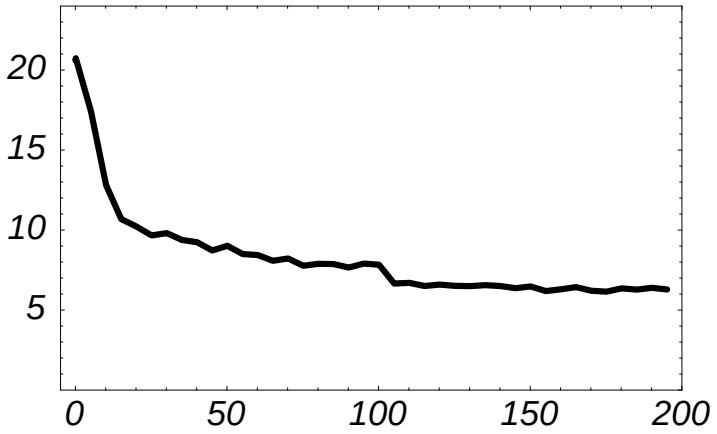
Figure 1: (Color online) Time snapshots of the secondary folding of a toy protein consisting of five regions where the local potential energy functional is constant with $\gamma_1 = \gamma_5 = 0.9$, $\gamma_2 = \gamma_4 = 0.1$ and $\gamma_3 = 0.55$. The regions with different values of assigned γ_i are shown in different colours(shadings) along the initial ($t = 0$) state. Initially the solitary wave is $\psi(t = 0) = 2\ \mathrm{sech}(2(x - 40))\exp[i(x - 40)]$ and $\phi = 0$. The coefficients in (6-8) are $\xi = 0.1, \Gamma = 0.1, m = 0.5, C = 2, \Lambda = 0.5$. The position of the solitary wave is shown in green (light grey) and the arrows indicate the direction of its motion.



point ($x = 0$) all five amino acid regions are in their ground states, but as the soliton passes through the protein domain for the third time, the middle region, that is now possesses a large string tension energy $T(\phi)$, is moved to its metastable state to minimise the penalty in the Lagrangian (5) for the deviation from the neighbouring conformations (Fig. 1 at $t = 110, 130$).

The effective energy of the solitary wave is given by the part of (5) that corresponds to

Figure 2: Kinetic energy $\mathcal{E}_{\rm kin} = \int |\partial_x \psi|^2 dx$ of the solitary wave as a function of time obtained by numerical integration of the continuous model (6-8) with the initial condition $\psi(t=0) = 2\ \mathrm{sech}(2(x-40))\exp[i(x-40)]$ and $\phi(t=0) = 0$. The coefficients in (6-8) are $\xi = 0.1, \Gamma = 0.1, m = 0.5, C = 2, \Lambda = 0.5$.



the nonlinear Schrödinger equation

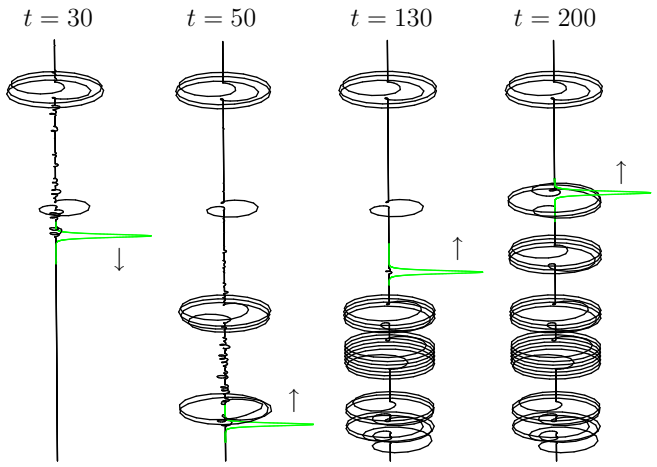
$$\mathcal{E}_{NLS} = \int |\partial_x \psi|^2 dx - \int |\psi|^4 dx, \tag{10}$$

where the first and the second terms represent kinetic and potential energy correspondingly. Figure 2 shows the plot of the kinetic energy of the solitary wave as a function of time. Four quantitatively different regimes could be identified. Initially, during $t \in [0, 15]$ soliton's energy is converted into the vibrational energy of the upper half of the protein chain. The rate of change of the kinetic energy slows down for $t \in [15, 95]$ as soliton traverses the entire chain. The rate of change increases again around $t \approx 100$ when the soliton reaches the middle region; this large energy loss is needed to overcome the energy barrier from stable to metastable conformation. The kinetic energy is almost a constant in the following times.

Next we notice that the system of continuous equations (6-8) can be replaced by the discrete one, where each state (ψ_j, ϕ_j) is written for a particular amino acid

$$\begin{aligned} \mathrm{i}\partial_t \psi_j &= -Q(\psi_{j+1} - 2\psi_j + \psi_{j-1}) + (\Lambda(\phi_j - \tfrac{1}{2})^2 - 2|\psi_j|^2)\psi_j & (11) \\ m\partial_{tt}\phi_j &= -12C\phi_j(\phi_j - 1)(\phi_j - \gamma_j) - 2\Lambda|\psi_j|^2(\phi_j - \tfrac{1}{2}) \\ &\quad + P(\phi_{j+1} - 2\phi_j + \phi_{j-1}) - \Gamma\partial_t\phi_j, & (12) \end{aligned}$$

Figure 3: (Color online) Time snapshots of the secondary folding of a toy protein consisting of 1000 amino acids. An initial solitary wave (13) is moving in the upward direction and is reflected from the end points of the protein chain. The coefficients in (11-12) are $Q = 100, P = 20, \Gamma = 0.1, m = 0.5, C = 2, \Lambda = 0.3$. The position of the solitary wave is shown in green (light grey) and the arrows indicate the direction of its motion.

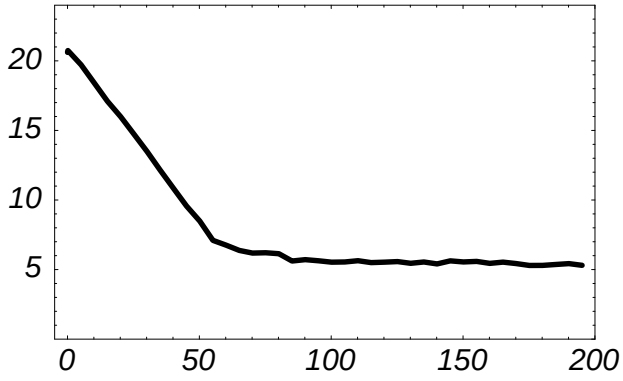


We consider a sequence of 1000 amino acids. For each amino acid the value of γ_j is uniformly randomly chosen from $[0, 1]$. In the previously considered case, the length scale of the soliton is small in comparison with the length scale on which the local potential energy varies. Here we consider an opposite case, when the width of the soliton is much larger than the distance on which the energy functional changes, so that we take $Q = 100$ and the initial form of the soliton as

$$\psi_j(t = 0) = 2 \text{ sech}[\tfrac{1}{5}(x_j - 400)] \exp(i\tfrac{1}{10}(x_j - 400)). \tag{13}$$

The solitary wave passes through the protein backbone three times until the final configuration is reached that consists of nine $\beta-$sheets and eight $\alpha-$helices. Each $\alpha-$helix contains on average 28 amino acids, although these regions vary in size from 8 to 63 amino acids. So in spite of the large fluctuations in the stable ground state from one amino acid to the next, the final conformation is macroscopic. Figure 4 shows the plot of the kinetic energy of the solitary wave as a function of time that exhibits a clear bimodal behaviour. The fastest decay of the soliton energy is until $t \sim 50$ at which point the soliton traversed the entire chain at least once. After that, in spite of a large change in conformational structure, the soliton barely looses any energy, as the main source of the energy that enables the transition comes from the potential energy of the spring.

Figure 4: Kinetic energy $\mathcal{E}_{\rm kin} = \sum(\psi_{j+1} - \psi_j)^2$ of the solitary wave as a function of time obtained by numerical integration of the discrete model (11-12) with the initial condition (13) and $\phi_j = 0$ for all j. The coefficients in (11-12) are $Q = 100, P = 20, \Gamma = 0.1, m = 0.5, C = 2, \Lambda = 0.3$. The bimodal character of the energy decay corresponds to two different sources of the protein folding. The value of $\mathcal{E}_{\rm kin}$ is scaled by the factor of 10, so that a direct comparison with Fig. kin can be made.



3 Conclusions

We have demonstrated, using simple phenomenological models, that the transition from the primary to the secondary conformation in a protein chain could, in principal, result from an interplay between the energy supplied by a solitary wave moving through the protein backbone and the potential energy of the chain tension. The described folding pathway follows the scenario of that for known proteins [23]. Of course, in order to accurately model the dynamics of protein folding, more information about amino acids should be taken into account such as the biochemical structure of individual amino acids, a hydrophobic propensity or hydopathy of the side chains or whether it is a globular or membrane protein etc. Simple models as these presented here are too simplistic to describe the actual folding process of a complex biomolecule. Nevertheless, they show that the important processes of the folding pathways can be identified and presented by completely deterministic models.

Acknowledgements

The author is indebted to Professor Ben-Jacob for stimulating discussions. This research was supported by NIH grant HG02411.

References

[1] C. M. Dobson, Nature, 426 (2003) 884

[2] A.S.Davydov and N.I.Kisluka, Zh. Eksp. Teor. Fiz. 71 (1976) 1090 [Sov. Phys. JETP 44 (1976) 571]

[3] A.S. Davydov, Biology and Quantum Mechanics, Pergamon Press 1982

[4] H. Fröhlich, Proc. R.Soc.London, Ser. A 215 (1952) 291

[5] D.B. Hall, R.E. Holmkin, J.K. Barton, Nature 382 (1996) 731; D.B.Hall, J.K.Barton, J. Am. Chem. Soc. 119 (1997) 5045; H.-W. Fink, C. Schönenberger, Nature 398 (1999) 407.

[6] E.M. Conwell and S. V. Rakhmanova, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 4556

[7] D.N. Beratan, S. Priyadarshy, and S.M. Risser, Chem. Biol. 4 (1997) 3

[8] Z. Hermon, S. Caspi, and E. Ben-Jacob Europhys. Lett. 43 (1998) 482

[9] A. Xie, L. van der Meer, W. Hoff, R. H. Austin Phys. Rev. Lett. 84 (2000) 5435

[10] N.K. Voulgarakis, D. Hennig, H. Gabriel, and G.P. Tsironis, J. Phys.: Condens. Matter 13 (2001) 9821

[11] S.F.Mingaleev, Y.B.Gaididei, P.L.Christiansen and Y.S.Kivshar, Europhys. Lett, 59 (2002) 403

[12] S. Caspi and E. Ben-Jacob, Phys. Lett. A 272 (2000) 124

[13] J. N. Onuchic and P. G. Wolynes, Curr. Opin. Struct. Biol. 14 (2004) 70; A.G. García and J. N. Onuchic, Proc. Nat. Acad Sci. USA 100 (2003) 13898

[14] E. Ben-Jacob, private communication.

[15] J.D.Bryngelson and P. G. Wolynes, Proc. Nat. Acad Sci. 84 (1987) 7524

[16] D. Kern and E. R. Zuiderweg, Curr. Opin. Struct. Biol. 13 (2003) 748

[17] L.N.Johnson, Nat. Struc. Biol. 1 (1994) 657; B.F.Volkman, D. Lipson, D.E.Wemmer and D. Kern, Science 291 (2001) 2429.

[18] M.P.Mayer, H. Schroder, S. Rudiger, K. Paal, T. Laufen, and B. Bukau, Nat. Struc. Biol. 7 (2000) 586

[19] C. McElroy, A. Manfredo, A. Wendt, P. Gollnick, M. Foster, J. Mol. Biol. 323 (2002) 463

[20] M.Peyrard, Nonlinear Excitations in Biomolecules, Springer, Berlin, 1995

[21] A.J.Heeger, S. Kivelson, J.R.Schrieffer and W.-P.Su, Rev. Mod.Phys., 60 (1988) 781

[22] E.A.Kuznetsov, A.M.Rubenchik, V.E.Zakharov, Phys. Rep. 142 (1986) 103

[23] T.A.Brown, Genomes, John Wiley & Sons, Inc., 1999