Quantitative theory of hydrophobic effect as a driving force of protein structure

Nikolay Perunov[1] and Jeremy L. England[*1]

[1]Department of Physics, Massachusetts Institute of Technology

June 29, 2013

Running title: Conformational motion in globular proteins.

Total number of

- manuscript pages: 18,
- supplementary material pages: NaN,
- tables: 1,
- figures: 4.

1. Title page

- full article title;
- names and affiliations of all authors (matched by superscript numbers);
- name, mailing address, telephone number, fax number, and E-mail address of the corresponding author;
- running title of 50 characters or less;
- list of total number of manuscript pages, supplementary material pages, tables, and figures; and
- a description of supplementary material intended for publication in the Electronic Edition, including filenames.

[*]Corresponding Author: J. L. England, Building 6C, 77 Massachusetts Avenue, Cambridge, MA, 02139. Phone/Fax: 617-253-0063. Email: jengland@mit.edu.

Abstract

Various studies suggest that the hydrophobic effect plays a major role in folding of proteins. However, quantifying the hydrophobic effect and analysing how it promotes assembly of proteins into a native structure has been challenging. Here, we extend a phenomenological model of protein folding that considers the hydrophobic effect to be a major force of protein structure and provides a mechanistic explanation of structural rearrangements in globular protein domains. We also show how the parameters of this model can be computed from crystal structures and apply the model to predict positions of ligand binding sites and to explain conformational changes in real proteins.

Keywords: hydrophobicity scale, protein structure, conformational fluctuations, structural rearrangements, ligand binding sites, mutations.

Statement for a broader audience: The functions performed by proteins depend on their shape and ability to change shape during interaction with the environment. Therefore, to design drugs, that would promote or inhibit specific processes in the cells, it is crucial to understand why proteins fold in particular shapes. Here, we develop a physical theory of protein folding which provides a general approach to studying spatial motion of proteins and which can be used as a tool to study large collections of proteins.

2. Abstract and keywords . Include (a) an abstract of no more than 250 words, followed by (b) four to ten keywords or short phrases for indexing that reflect the content and major thrust of the paper, and (c) a 50-75-word statement, written for a broader audience, outlining the importance and/or impact of the work presented in the manuscript. The abstract should succinctly describe the objectives of the research, the experimental approach, and the major results and their significance. It must be self-explanatory and suitable for abstracting services such as Chemical Abstracts, Biosis, etc. Reference ci-

tations in the abstract should be avoided whenever possible and, if necessary, given in full. Avoid the use of abbreviations and acronyms in the abstract unless they are defined therein.

1 Introduction

4. Introduction. The text of the paper begins on a new page. The Introduction should state the purpose of the investigation, the hypotheses tested, and the relationship to other work in the field. Avoid lengthy reviews of the literature.

Since the experiments by Anfinsen [2], the basis of structural biology rested on the idea that a shape of a protein is completely determined by its sequence. However, for a long time scientists have not addressed this question directly. For instance, the most common experimental techniques, X-ray crystallography and NMR-spectroscopy, have focused only on finding three dimensional structure of proteins with full atomic resolution, whereas the primary goal of theoretical studies has been to explain how a protein reaches a single native state structure and how it can do it fast [19]. Simulations of protein folding with all atom resolution have been challenging because of the large size of proteins [23, 26]. Only recently, when large collections of sequence homologs have become available, have there appeared computational methods that use evolutionary sequence variation to predict the native structure of proteins more than 60-residue long [15]. Despite the success of computational methods in predicting protein structure and advances in computer simulations (breaking 1 millisecond timescale [23]), neither computational methods nor simulations reveal the physical mechanisms of protein folding. Therefore, there is a necessity for a physical theory which will provide a general quantitative approach to the protein folding problem and might be used as a tool to study large collections of protein sequences.

Various studies suggest that the hydrophobic effect plays a major role in folding of proteins [22, 4]. Although the hydrophobic effect is well understood qualitatively – non-

polar amino acid residues tend to be buried in the core of the protein, and the polar residues are more likely to be on the surface – a quantitative theory of the hydrophobic effect is difficult to construct [4]. To measure the relative hydrophobicities of amino acids, a number of experimental and numerical methods have been developed. Experimental hydrophobicity scales are based on the measurements of the free energy of solvation of single amino acids or short peptides in water and ethanol [13, 27, 18], while numerical methods look at the partitioning of amino acid residues between the core and the surface in proteins with known 3−D structures [9, 22]. These scales are widely used to compute hydrophobicity profiles or window averaged sequence hydrophobicity which can be used to find out information about secondary and tertiary structure of a protein from its sequence. However, some studies point out that the physical properties of amino acid residues can change depending on a local environment inside a protein so that hydrophobicity scales measured in solvation experiments and used to compute hydrophobicity profiles might not indicate well which residues are buried or exposed [14, 5]. In addition, hydrophobicity profiles depend on the size of a window or the number of Fourier harmonics that are used to smooth sequence hydrophobicity, which is the parameter that is not known, and they do not provide any information about conformational changes [24].

Previously, we introduced a phenomelogical model of protein folding that considers the hydrophobic effect, steric repulsion, and bending of a protein backbone to be the driving forces of protein structure [7]. Using only the sequence of a protein, this model allows us to compute not only the minimum energy conformational state of a protein, but also an ensemble of low energy excited states, which can be used for studying coupled motion of different parts of a protein, called allostery. In this study, we show how the parameters of the model can be found from a collection of 3-D protein structures, ~~define the area of applicability of the model,~~ discuss how the model can be improved by taking

into account interdomain interactions, and finally demostrate how the model can be used to analyze conformational changes in proteins.

2 Results

5. Results . The results should be presented in a clear and concise manner, mentioning figures and tables that summarize or illustrate important findings.

2.1 Burial Mode Model

In burial mode model a globular protein domain is represented as a linear chain of N residues which are indexed by the number s and have position $\mathbf{r}(s)$ relative to the center of mass of the globule (Figure 1). The hydrophobic effect and polymeric bonds are incorporated into the Hamiltonian

$$\mathcal{H} = \int \mathrm{d}s \left[k \left| \frac{\mathrm{d}\mathbf{r}(s)}{\mathrm{d}s} \right|^2 + \varphi(s) |\mathbf{r}(s)|^2 \right], \tag{1}$$

where relative hydropathy $\varphi(s)$ is obtained by converting amino acid sequence into numbers using the standard Kyte-Doolittle hydrophobicity scale [13]. The steric repulsion between different parts of a chain is taken into account as a global constraint on the ratio of mean-square distance to the maximum distance to the center of mass R

$$\int \mathrm{d}s |\mathbf{r}(s)|^2 = \frac{3}{5} ? \alpha N R^2. \tag{2}$$

The goal of this constraint is to prevent residues from collapsing into the core a globule and, thus, to account for the limited space in the core. The burial traces, which are the

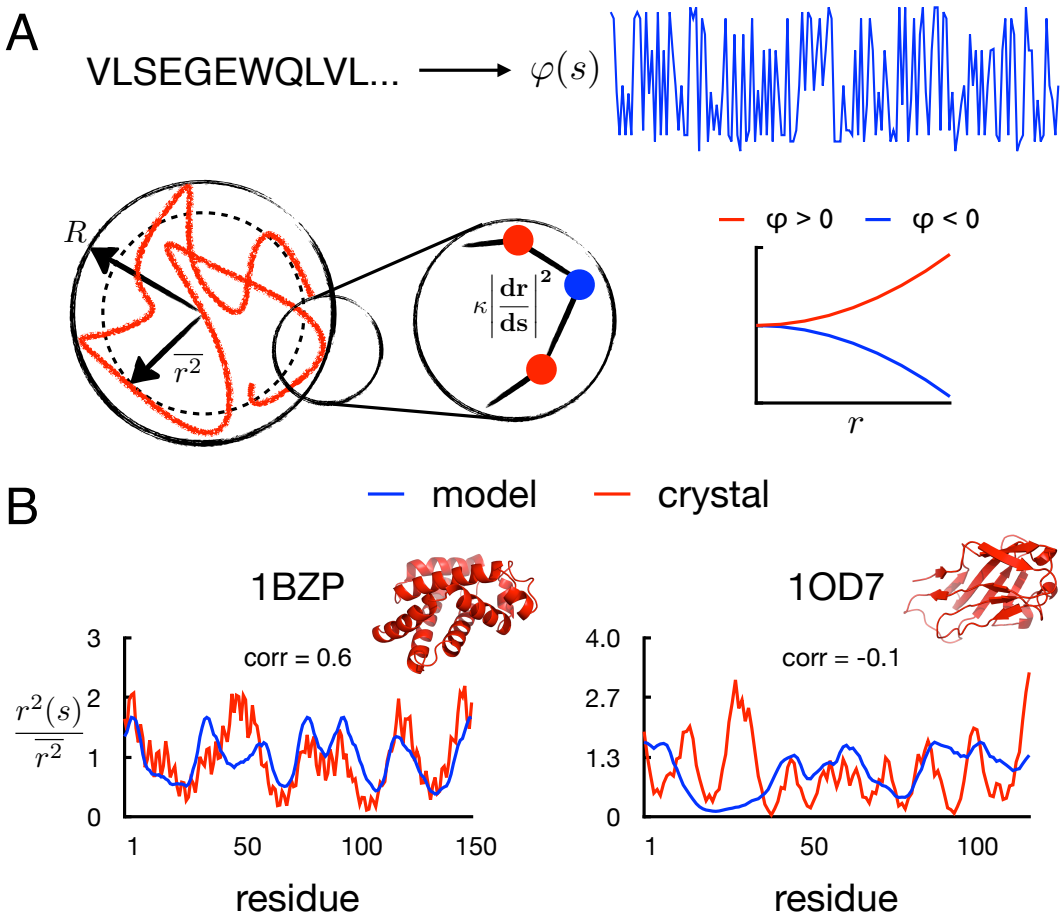

Figure 1: Computation of burial traces in the model.
A. Constructing burial mode model Hamiltonian.
B. Examples of burial traces with different Pearson correlation with the crystal.

distances from each residue to the center of mass $r(s)$, of the lowest energy and excited conformations are computed by finding combinations of the eigenmodes of the Hamiltonian (1), called burial modes, that satisfy steric constraint (2). As a result, allostery naturally arises in the model as an interplay between different burial modes. Should we say something about linear programming here? and how conformational changes arise from interplay between burial modes?

To study allosteric motion, conformational changes, ligand binding, and misfolding of

real proteins using burial mode model one should be confident that at least the burial traces predicted by the model agree well with that of computed from crystal structures. Thus, to understand how the model works for different classes of proteins, we selected protein domains with unique sequences of length between 100 and 300 a.a. from SCOP-database [17]. For each domain we computed a pair of burial traces (the one from a sequence, and the other from crystal structure) and then we calculated Pearson's correlation coefficient (PCC) between burial traces in each pair. The examples of burial traces with different PCC are shown on Figure 1B. In addition, for these domains we calculated maximum PCC between burial trace from crystal structure and a moving average of sequence hydrophobicity $\varphi(s)$. The results of these calculations are summarized in Table 1. As one can see from this table, with any hydrophobicity scale the model does not work well for a wide range of proteins. The later suggests that the parameters of the model are not optimal, or the model works only for a narrow class of proteins. To examine these hypothesis, we first tried to optimize parameters of the model for a large set of proteins, and then to identify the origin of disagreement between model prediction and crystal structures for specific groups of proteins.

2.2 Parameter Optimization

There are 21 independent parameters in burial mode model of protein folding: the bond stiffness κ, the ratio of the mean-square radius to maximum size of a protein α, and 19 relative hydrophobicities of amino acid residues. However, not all of these parameters can be changed in the framework of the model. Indeed, the bond stiffness κ fixes the units of length in the model and was chosen so that corresponding mean-square distance between neighbouring C_α atoms is equal to one; the parameter α ranges from 0.4 to 0.6

Hydrophobicity scale	Protein class			
	α	β	$\alpha+\beta$	α/β
Kyte-Doolittle	0.25 ± 0.2 0.3 ± 0.2	0.25 ± 0.2 0.35 ± 0.2	0.25 ± 0.2 0.3 ± 0.2	0.25 ± 0.2 0.3 ± 0.2
Wimley-White	0.27 ± 0.20 0.3 ± 0.2	0.25 ± 0.2 0.35 ± 0.2	0.25 ± 0.2 0.3 ± 0.2	0.25 ± 0.2 0.3 ± 0.2
Janin	0.25 ± 0.2 0.3 ± 0.2	0.25 ± 0.2 0.35 ± 0.2	0.25 ± 0.2 0.3 ± 0.2	0.25 ± 0.2 0.3 ± 0.2

Table 1: Performance of the model with different hydrophobicity scales for different classes of proteins. First line corresponds to the mean and the standard deviation of PCC between burial traces computed from sequence and extracted from crystal structure, while the second line corresponds to those between hydrophobicity profiles and burial traces extracted from crystal structures.TODO: enter correct values.

in real proteins and was set to 3/5 assuming that globular proteins are roughly spherical and have uniform density. As a consequence, to improve the performance of the model on a large number of proteins one can vary only hydrophobicity scale of amino acid residues.

For a given protein one can easily find a set of parameters such that burial traces computed from the crystal structure and from the model will match almost perfectly (PCC > 0.9), but the parameters obtained by fitting model burial trace into a given structure are unreasonable, i.e. they are in bad agreement with all existing (standard) hydrophobicity scales and, when used with this parameters, the model does not work well for other proteins. On the other hand, finding parameters by fitting into burial traces for a large of number of protein is computationally costly. Therefore, instead of running a brute force parameter optimization we first checked how the model performs with standard hydrophobicity scales different from Kyte-Doolittle [18, 9, 27], and then developed a method to compute hydrophobicity scale from known protein structures.

In our previous study, the relative hydrophobicities of amino acid residues were taken from Kyte-Doolittle (KD) scale and multiplied by a constant factor so that the energy

change associated with motion of glutamine from core to the surface of the globule is equal to 0.5 k_BT. To compare the performance of the model with different hydrophobicity scales we normalized these scales so that the difference between the maximum and the minimum hydrophobicities was the same as in KD scale. Table 1 shows the mean and the variance of the distribution of PCC between the burial traces from crystal structure and burial traces computed with the model using different hydrophobicity scales. As one can see, with all hydrophobicity scales the model works alike for large set of proteins (SCOP class). However, there are some families of proteins (such as globins a.1.1.2) for which the model with KD scale succeeds in predicting burial traces much better than with the other scales. This brings us back to the idea that the hydrophobicities of amino acid residues might depend on local environment in a protein [14], and maybe it is possible to find an optimal hydrophobicity scale for protein domains with similar structures (SCOP superfamily/family).

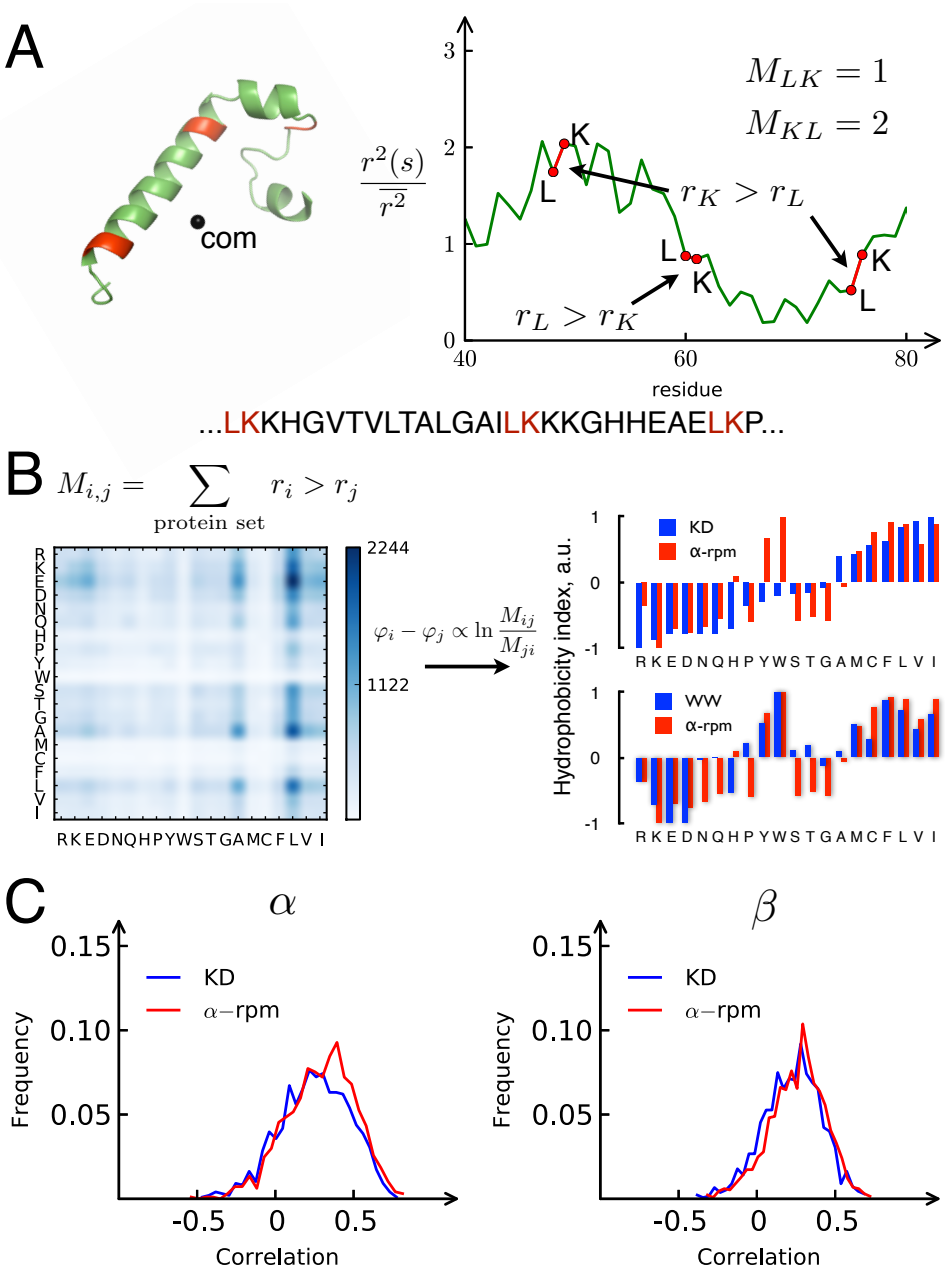

Figure 2: Extracting model parameters from crystal structures.
A. Construction of the matrix with relative positions of amino acids M_{ij}.
B. Calculation of hydrophobicity scale (α-rpm) from the matrix M_{ij} and comparison of this scale with Kyte-Doolittle and Wimley-White hydrophobicity scales. The matrix M_{ij} was constructed using α domains from SCOP database with unique sequences and lengths between 100 and 300 a.a..
C. Distribution of Pearson correlation coefficient between burial traces computed from the crystal structures and predicted by the model using KD and α-rpm scales.

To check this idea we developed a method of extracting relative hydrophobicities from a collection of proteins with known structures. In particularly, we looked at the distribution of amino acid positions inside globular protein domains with unique sequences and constructed a matrix M_{ij}, elements of which equal to the number of times that residue of type i is further from the center of the globule than residue of type j, given that these residues are the nearest neighbors on a chain (Figure 2). Assuming that the probability of amino acid of type i being closer to the center of the globule than amino acid of type j is given by Boltzmann weight, we find that the relative hydrophobicity of these amino acids is given by

$$\varphi_i - \varphi_j \propto \ln \frac{M_{ij}}{M_{ji}}.$$

Figure 3B shows the matrix of relative positions of amino acid residues M_{ij} and the hydrophobicity indices computed for a set of α-helical protein domains with unique sequences of length between 100 and 300 a.a. from SCOP database (970 total). To compute this matrix we used only the residues that are far from the center of a domain ($r^2 > 0.5R^2$). As one can see, the computed hydrophobicity scale ($\alpha-$rpm) agrees well with KD scale with the exception of histidine (H) and aromatic amino acids (Y, W). The difference in hydrophobicity of histidine might be due to the statistical error, whereas for the aromatic amino acids the source of difference with KD scale is unknown. However, our values for Y and W agree well with Whimley-White hydrophobicity scale. This scale was generated by measuring free energy of solvation (water-octanol) of polypeptides of the from AcWL-X-LL, where X is variable region composed of 2-4 amino acids. It should be noted, that WW scale includes the effects of the peptide bonds, while KD scale is a weighted average of free energy of solvation water-vapor and water-ethanol of single amino-acids.

Finally, we tested how the model works with the new hydrophobicity scale. As one can see from Figure 3C, new parameters do not significantly improve the performance of the model on a large set of proteins. This points out to the limitations of the model which come from neglect by intrachain and interdomian interactions and burail trace approximation. Therefore, for the rest of the paper we decided to use KD scale and apply the model to the protein domains for which it succeeds in predicting burial traces.

2.3 Multidomain proteins. Globins.

We looked at the distribution of PCC for SCOP family of globins (a.1.1.2). This family consists of two groups of proteins: myoglobin (single domain protein) and hemoglobin (heterotetramer). As one can see from Figure 3A, the model predicts burial traces significantly better for single domain than for multidomain proteins. This might be due to interaction between domains in hemoglobin that is not included in the model. To account for this interaction we modified original burial mode model. In particularly, we successively pinned each residue to the surface of the globule by setting its hydrophobictiy index to a large negative number and calculated the burial trace. This procedure is suppose to mimick hydrophobic patches on the surface by preventing these patches from burying. The PCC as a function of pinning position is shown on Figure 3B. As one see from this figure, both in $\alpha-$ and $\beta-$chains of hemoglobin there are two regions of points that provide high PCC when they are pinned to the surface: residues 74-79 and C-terminus region. However, as one can see from Figure 3, none of these region corresponds to the inter-domain interfaces.

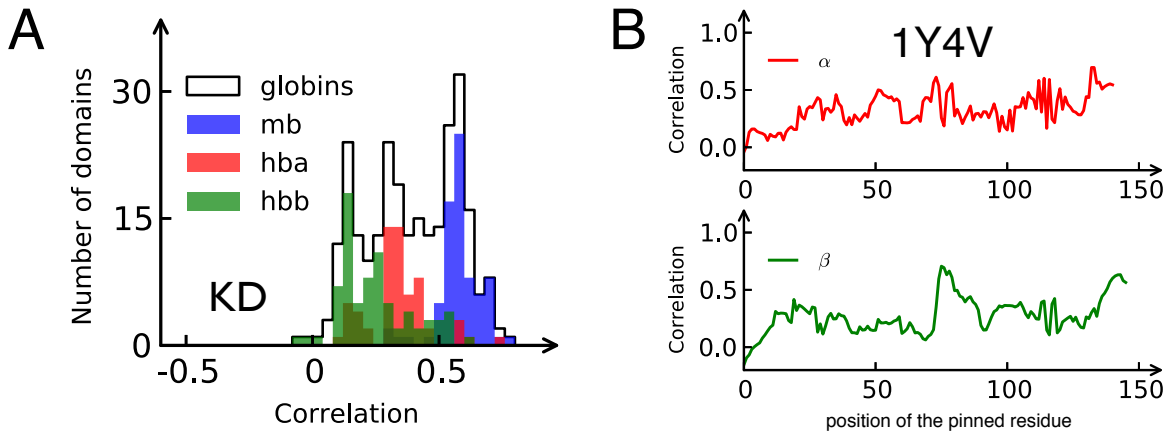

Figure 3: Accounting for inter-domain interaction in haemoglobin.
A. Distribution of Pearson correlation coefficient between burial traces predicted from the sequence using KD scale and extracted from crystal structures for a.1.1.2 SCOP group.
B. Pearson correlation coefficient between burial traces extracted from crystal structures of α and β chains of hemoglobin (1Y4V) and computed using the model when one of the residues is pinned to the surface of the globule. Should be use gray bar to highlight interdomain contacts?

2.4 Conformational changes

Should we say about burial mode analysis: steric repulsion vs. burial? In addition to prediction of the lowest energy structure, the burial mode model can be used to study conformational changes in single domain proteins. By generating an ensemble of burial traces with energy $\Delta E = 1 - 5k_BT$ above the ground energy, one can compute the covariance matrix $\mathrm{cov}[r^2(s), r^2(s')]$ for a given protein. This procedure for the whale sperm myoglobin (1BZP) is illustrated on Figure 4A and Supp. Info.. The off-diagonal elements of the covariance matrix show how the motion of different parts of the protein is correlated: when $\mathrm{cov}[r^2(s), r^2(s')]$ is greater than zero then residues s and s' tend to be buried in the core or exposed to the solvent together. On the other hand, the diagonal elements of the covariance matrix are equal to the variance of squared radial position/distance

$r^2(s)$ and, thus, measure the magnitude of fluctuations in radial position. This information can be used to compute the response to ligand binding and to access the stability of a protein.

For the case of allosteric response, i.e. when the region that undergoes the largest rearrangement is different from the ligand binding site, one can take the sum of rows of the covariance matrix which correspond to ligand binding region and look at the positions of the peaks in the this sum. Previously, on the example of LFA, H-ras, and CheY it was demonstrated that these peaks correspond to the regions of strongest/allosteric response [7]. However, the drawback of this method is that the allosteric reponse can be computed only when the ligand binding site are known in advance. Fortunately, the ligand binding events quite often lead to large conformational changes in the region of binding, and in these cases the active sites are the most variable regions of the protein. Therefore, by examining the variances of squared radial distance var$[r^2(s)]$ as a function of the residue position along the chain one can predict the positions of ligand binding sites. For example, the subplots at the bottom of the Figure 4A show the var$[r^2(s)]$ and the 3D structure of the myoglobin (1BZP) colored according to this function. As one can see from these subplots, the peak in var$[r^2(s)]$ corresponds to the to the heme binding site.

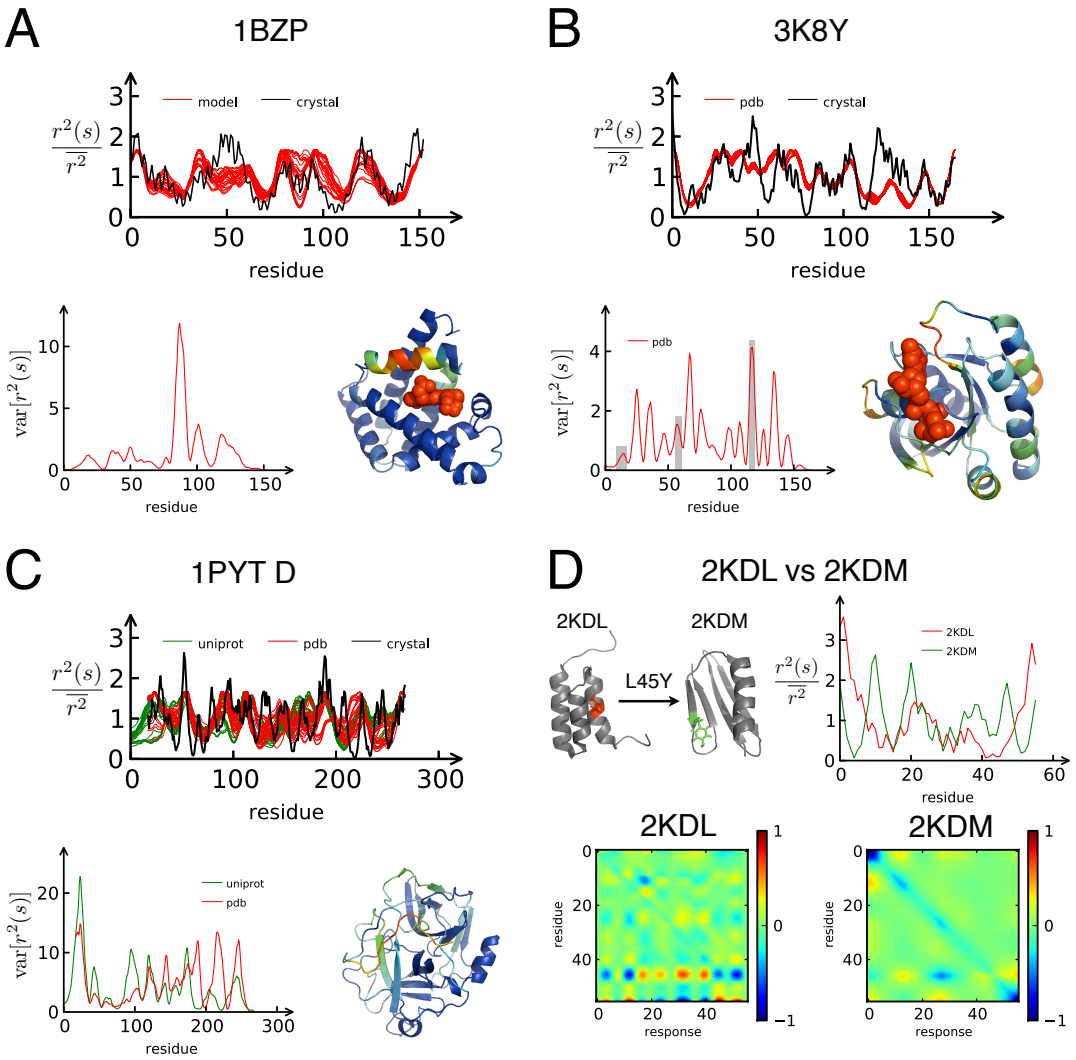

Figure 4: Conformational changes in proteins.
A. The most variable region in myoglobin corresponds to heme binding site. Crystal structure of 1BZP is colored according to the variance of $r^2(s)$ computed from low-energy excitations above the ground state (4 k_BT). Add gray bar?
B. Low-energy excitations of H-Ras computed for uniprot and pdb sequences. The most variable region according to the model is not present in PDB file.
C. Signal region of trypsinogen, which is cut in activation process, is the most variable region.
D. Mutation L45Y leads to the transformation of a 3-α fold into a 4β+α fold. Linear response matrices $\partial r^2(s)/\partial h(s')$. Residies 43-47 correspond to the region of the strongest response.

Following the approach described above we studied low-energy structural fluctuations in H-ras protein (3K8Y) and trypsinogen (1PYT, D chain). It is worth noting, that neither of these two proteins has been crystallized with the whole/uniprot sequence: 17 residues at the C-terminus of H-ras are not present at the crystal structure of the H-ras, and trypsinogen is crystallized only in its active form (trypsin) when first 16 residues at the N-terminus are cut. Figure 4B shows the the burial traces of low-energy excited conformations and the variance var$[r^2(s)]$ computed for the uniprot and pdb sequences of H-ras. As one can see from this figure, the 16 residues at the C-terminus which are not present at the in the crystal structure are the most variable region of H-ras according to the/our model. Furthermore, the variance var$[r^2(s)]$ computed for uniprot sequence is at least ten times larger than that of for pdb sequence. This fact might indicate that the structure corresponding to the uniprot sequence of H-ras is less stable than.../unstable and explain why H-ras is crystallizable only when the last 17 residues at the C-terminus are cut. In addition, from the variance var$[r^2(s)]$ computed from pdb sequence and the the 3D crystal structure of H-ras one can see that GTP binding sites (10-17, 57-61, 116-119) are highly fluctuating regions.

another way to generate covariance matrix is to add random noise to the hydrophobic potential $\phi(s)$?

In his study [1] Alexander investigated how the fold of the protein changes upon mutations. In particular, he demonstrated that it is possible to design the protein such that a single point mutation (L45Y) leads to switching from 3α to $4\beta{+}\alpha$ fold. Furthermore, he obtained a high-resolution NMR structures for two proteins different by three mutations (20, 30, 45). These 3D structures and the corresponding burial traces of these proteins (2KDL, 2KDM) are shown at the top panel of Figure 4D. Using the burial mode model,

we contructed the response matrix

$$\chi_{s,s'} = \frac{\delta r^2(s)}{\delta \varphi(s')},$$

where $\delta r^2(s)$ is the change in optimal burial trace at position s upon a small change in hydrophobicity $\delta\varphi(s')$ at position s' along the chain. Thus, rows/columns of this matrix shows how sensitive is the optimal structure of the protein to mutations. The bottom panel of Figure 4D shows the linear response matrices computed from the sequences of 2KDL and 2KDM. As one can see, for both proteins small changes in hydrophobicity in the region 43-47 produce large changes in optimal burial traces. The latter is consistent with Alexander's finding.

!!! The pearson correlation between the crystal and the model is about 0.2...

3 Discussion

Briefly interpret the results and relate them to existing knowledge in the field, but do not merely restate the results or present reviews of the literature.

The problem of protein structure prediction from amino acid sequence has a long history... In this study, we presented the model of protein folding which considers the hydrophobic effect, the polymeric bonds, and the steric repulsion to be the major factors that determine protein structure. The crutial parameter of this model is the hydrophobicity scale by means of which the amino acid sequence is mapped into the sequence of relative hydrophobicities. Thus, our initial goal was to find a hydrophobicity scale that would improve the prediction of protein structure by burial mode model.

Inspired by Miazawa-Jernigan (MJ) and Sippl's statistical potentials, we have developed method to infer relative hydrophobicities of amino acid residues from the analysis of known protein structures [16, 25]. However, unlike Miyazawa and Jernigan, who derived pairwise interaction matrix from the frequencies of all contacts between amino acid residues, we considered only local interactions along the chain. In this respect, our method is similar to Sippl's statistical potential, which is based on the probability distribution of pairwise distances between amino acids separated by k ($k < 5$) residues along the chain; yet our method is different from Sippl's approach since it focuses on the relative positions of amino acid residues with respect to the center of mass of the protein rather than pairwise distances. On one hand, the results of our calculation are encouraging because the hydrophobicity scale we obtained agrees well with KD and WW hydrophobicity scales which are based on the measurements of the solvation free energy. On the other hand, new hydrophobicity scale does not significantly improve the performance of the model on the large set of proteins. The later indicates that the model has limitations

which come from burial trace approximation and neglect by intrachain and interdomain interactions. It should also be noted, that the matrix of relative positions that we used to compute hydrophobicity scale contains more information about amino acids residue than just hydrophobicity scale. There are 190 parameters in this matrix that correspond to relative hydrophobicitiy of a pair of amino acids, so one can develop a model similar to the burial trace model where instead of the distance to the center of the globule one would look at the relative positions of two neighbouring residues.

In this study, we also addressed the question how function emerges from primary sequence in proteins. In particularly, we used the method of burial mode analysis to predict ligand binding regions and to study conformational changes in proteins. Earlier studies suggested the 'conformational selection' paradigm to describe the binding events [6]. In this paradigm, a protein undergoes conformational fluctuations and a ligand is assumed to select one conformational state which is compatible with binding. This process is accompanied by large structural rearrangements if there is an energy exchange between protein regions with 'discrete breathers' (localized excitations) [20, 21, 11, 12]. Thus, these regions are often located close to ligand binding sites. The later implies that one can predict the position of ligand binding or catalytic sites from the structural variability/flexibility of the protein. It should be noted, that unlike methods which use the normal mode analysis to compute structural variability and mechanical response [3, 10, 8], burial mode analysis relies on the knowledge of neither initial nor final conformational states of a protein. Explaination of burial mode analysis method?

To conclude, we presented a phenomenological model of protein folding which allows to compute information about protein structure directly from its sequence. We showed how the parameters of this model can be found from a set of proteins with known structure. Finally, we demonstrated that burial mode model captures the conformational changes

in proteins correctly and used the model to predict regions most sensitive to mutations. Because of the high speed, the model can be used to study large collections of homologous sequences to access structural stability of different mutants.

4 Materials and methods

7. Materials and methods. Describe materials and methods briefly but in sufficient detail to allow others to repeat the experiments. Novel procedures should be described in detail, but published procedures should be referenced by a literature citation. If hazardous materials or dangerous procedures are employed, necessary precautions must be stated.

References

[1] Patrick A. Alexander, Yanan He, Yihong Chen, John Orban, and Philip N. Bryan. A minimal sequence code for switching protein structure and function. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES 106(50):21149–21154, DEC 15 2009.

[2] Anfinsen Christian B. Principles that govern the folding of protein chains. Science, 181(4096):223–230, July 1973.

[3] Mark Bathe. A finite element framework for computation of protein normal modes and mechanical response. PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS, 70(4):1595–1609, MAR 2008.

[4] D Chandler. Interfaces and the driving force of hydrophobic assembly. NATURE, 437(7059):640–647, SEP 29 2005.

[5] C CHOTHIA. NATURE OF ACCESSIBLE AND BURIED SURFACES IN PROTEINS. JOURNAL OF MOLECULAR BIOLOGY, 105(1):1–14, 1976.

[6] Peter Csermely, Robin Palotai, and Ruth Nussinov. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. TRENDS IN BIOCHEMICAL SCIENCES, 35(10):539–546, OCT 2010.

[7] Jeremy L. England. Allostery in protein domains reflects a balance of steric and hydrophobic effects. Structure, 19(7):967 – 975, 2011.

[8] RJ Hawkins and TCB McLeish. Coarse-grained model of entropic allostery. PHYSICAL REVIEW LETTERS, 93(9), AUG 27 2004.

[9] J JANIN. SURFACE AND INSIDE VOLUMES IN GLOBULAR PROTEINS. NATURE, 277(5696):491–492, 1979.

[10] Do-Nyun Kim, Reza Sharifi Sedeh, Cong Tri Nguyen, and Mark Bathe. FINITE ELEMENT FRAMEWORK FOR MECHANICS AND DYNAMICS OF SUPRAMOLECULAR PROTEIN ASSEMBLIES. In NEMB2010: PROCEEDINGS OF THE ASME FIRST GLOBAL CONGRESS ON NANOENGIN pages 315–316, THREE PARK AVENUE, NEW YORK, NY 10016-5990 USA, 2010. ASME Nanotechnol Inst, AMER SOC MECHANICAL ENGINEERS. ASME 1st Global Congress on Nanoengineering for Medicine and Biology, Houston, TX, FEB 07-10, 2010.

[11] G Kopidakis and S Aubry. Intraband discrete breathers in disordered nonlinear systems. I. Delocalization. PHYSICA D, 130(3-4):155–186, JUN 15 1999.

[12] G Kopidakis, S Aubry, and GP Tsironis. Targeted energy transfer through discrete breathers in nonlinear systems. PHYSICAL REVIEW LETTERS, 87(16), OCT 15 2001.

[13] Jack Kyte and Russell F. Doolittle. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology, 157(1):105 – 132, 1982.

[14] P MANAVALAN and PK PONNUSWAMY. HYDROPHOBIC CHARACTER OF AMINO-ACID RESIDUES IN GLOBULAR PROTEINS. NATURE, 275(5681):673–674, 1978.

[15] Debora S. Marks, Lucy J. Colwell, Robert Sheridan, Thomas A. Hopf, Andrea Pagnani, Riccardo Zecchina, and Chris Sander. Protein 3d structure computed from evolutionary sequence variation. PLoS ONE, 6(12):e28766, 12 2011.

[16] S MIYAZAWA and RL JERNIGAN. ESTIMATION OF EFFECTIVE INTER-RESIDUE CONTACT ENERGIES FROM PROTEIN CRYSTAL-STRUCTURES - QUASI-CHEMICAL APPROXIMATION. MACROMOLECULES, 18(3):534–552, 1985.

[17] Alexey G. Murzin, Steven E. Brenner, Tim Hubbard, and Cyrus Chothia. Scop: A structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology, 247(4):536 – 540, 1995.

[18] Y NOZAKI and C TANFORD. SOLUBILITY OF AMINO ACIDS AND 2 GLYCINE PEPTIDES IN AQUEOUS ETHANOL AND DIOX-

ANE SOLUTIONS - ESTABLISHMENT OF A HYDROPHOBICITY SCALE. JOURNAL OF BIOLOGICAL CHEMISTRY, 246(7):2211–&, 1971.

[19] Vijay S. Pande, Alexander Yu. Grosberg, and Toyoichi Tanaka. Heteropolymer freezing and design: Towards physical models of protein folding. Rev. Mod. Phys., 72:259–314, Jan 2000.

[20] Francesco Piazza and Yves-Henri Sanejouand. Discrete breathers in protein structures. PHYSICAL BIOLOGY, 5(2), JUN 2008.

[21] Francesco Piazza and Yves-Henri Sanejouand. Long-range energy transfer in proteins. PHYSICAL BIOLOGY, 6(4), DEC 2009.

[22] GD ROSE, AR GESELOWITZ, GJ LESSER, RH LEE, and MH ZEHFUS. HYDROPHOBICITY OF AMINO-ACID RESIDUES IN GLOBULAR-PROTEINS. SCIENCE, 229(4716):834–838, 1985.

[23] David E. Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, Michael P. Eastwood, Joseph A. Bank, John M. Jumper, John K. Salmon, Yibing Shan, and Willy Wriggers. Atomic-level characterization of the structural dynamics of proteins. Science, 330(6002):341–346, 2010.

[24] BD Silverman. Underlying hydrophobic sequence periodicity of protein tertiary structure. JOURNAL OF BIOMOLECULAR STRUCTURE & DYNAMICS, 22(4):411–423, FEB 2005.

[25] MJ SIPPL. CALCULATION OF CONFORMATIONAL ENSEMBLES FROM POTENTIALS OF MEAN FORCE - AN APPROACH TO THE KNOWLEDGE-

BASED PREDICTION OF LOCAL STRUCTURES IN GLOBULAR-PROTEINS. JOURNAL OF MOLECULAR BIOLOGY, 213(4):859–883, JUN 20 1990.

[26] Vincent A. Voelz, Marcus Jäger, Shuhuai Yao, Yujie Chen, Li Zhu, Steven A. Waldauer, Gregory R. Bowman, Mark Friedrichs, Olgica Bakajin, Lisa J. Lapidus, Shimon Weiss, and Vijay S. Pande. Slow unfolded-state structuring in acyl-coa binding protein folding revealed by simulation and experiment. Journal of the American Chemical Society, 134(30):12565–12577, 2012.

[27] WC Wimley, TP Creamer, and SH White. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. BIOCHEMISTRY, 35(16):5109–5124, APR 23 1996.