John von Neumann Institute for Computing

Constraint Network Analysis: Exploiting the Link Between Protein Rigidity and Thermostability

S. Radestock, H. Gohlke

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 357-360, 2008.

http://www.fz-juelich.de/nic-series/volume40

Constraint Network Analysis: Exploiting the Link Between Protein Rigidity and Thermostability

Sebastian Radestock[1,2] **and Holger Gohlke**[1,2]

[1] Fachbereich Biowissenschaften, Molekulare Bioinformatik,
Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany

[2] Mathematisch-Naturwissenschaftliche Fakultät, Pharmazeutisches Institut,
Christian-Albrechts-Universität, Kiel, Germany
E-mail: {*radestock, gohlke*}*@pharmazie.uni-kiel.de*

Understanding the relationship between microscopic structure and macroscopic stability is important for developing strategies to improve protein stability in the reaction media used in industrial processes, *e.g.*, at high temperatures. Protein thermostability has been repeatedly linked to an enhanced structural rigidity of the folded state. Here, we used constraint network analysis for directly probing the rigidity of protein structures from mesophilic and thermophilic organisms along a thermal unfolding trajectory. The approach allowed for identifying structural features from which a destabilization of the structure originates upon thermal unfolding. These predictions showed a good agreement with experiment. The information might thus be exploited in data-driven protein engineering by pointing to residues that should be varied to obtain a protein with higher thermostability.

1 Introduction

Stable proteins are important for broadening the industrial applicability of enzymes.[1] Naturally occurring enzymes have usually not evolved to be tolerant to the presence of organic solvents, extremes of pH or high temperatures that might occur in industrial processes. The identification or the development of enzymes with higher stability will thus increase the adoption of biocatalytic syntheses in industrial production. Understanding the relationship between microscopic structure and macroscopic stability is essential for this. In this context, computational approaches that allow for identifying structural features from which a destabilization of the structure originates should provide valuable guidance.[2]

Of all potentially destabilizing factors that might occur in industrial production, temperature is the best studied.[1] As an approach to understand the determinants of thermostability, proteins from thermophilic organisms with optimum growth temperatures of more than 60°C have been investigated. These proteins show a substantially higher intrinsic thermostability than their counterparts from mesophilic organisms, while retaining the basic fold characteristics of the particular protein family.[3] Protein thermostability has been repeatedly linked to an enhanced structural rigidity of the folded state.[3]

2 Materials and Methods

Crystallographic models of 20 homologous pairs of mesophilic and thermophilic protein structures were collected from the Protein Data Bank (PDB).[4] Protein structures were modeled as constraint networks, where vertices represent atoms and edges represent covalent

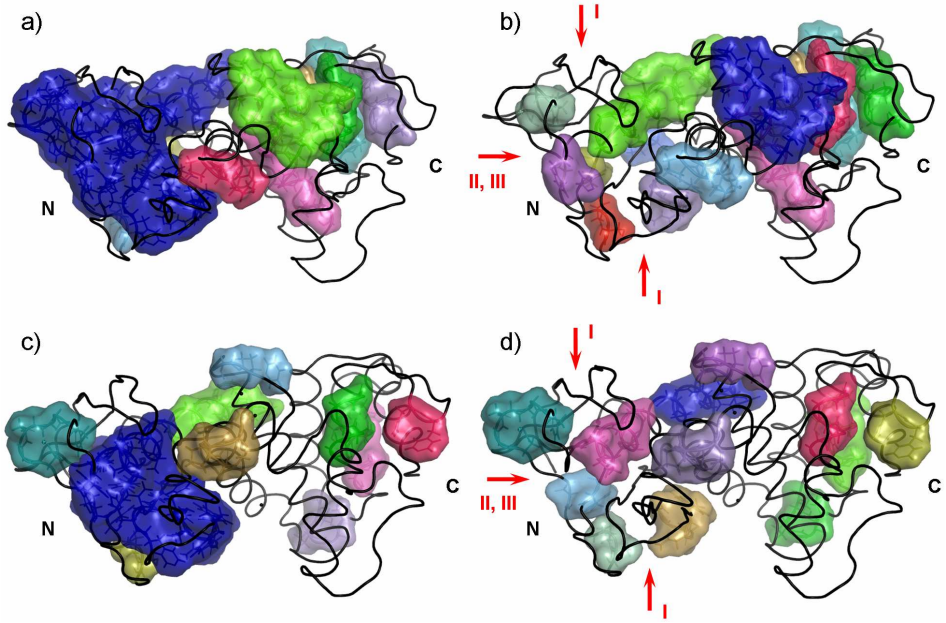

Figure 1. Mesophilic (a, b) and thermophilic (c, d) thermolysin-like protease (TLP) directly before (a, c) and after (b, d) the phase transition. Rigid clusters are depicted as uniformly colored bodies. The blue body in (a) and (c) represents the giant cluster. Arrows in (b) and (d) indicate potential unfolding nuclei. Roman numbers refer to the numbering of the unfolding nuclei in Table 1. The N- and C-termini are marked.

and non-covalent bond constraints as well as angular constraints. The network was constructued using the FIRST software (version 6.2).[5] A fast combinatorial algorithm can be applied to determine the number and spatial distribution of bond-rotational degrees of freedom in the network and, hence, the local network rigidity. Such a rigidity analysis is available with FIRST.[5]

By diluting non-covalent constraints in the protein structure network starting from the native state, FIRST has been applied to simulate thermal unfolding of proteins.[6] Here, heating was simulated by removing hydrogen bonds from the network in the order of increasing interaction energy. The energy of a hydrogen bond relates to the temperature at which the bond breaks. In going from a rigid to a flexible network, a phase transition can be observed that defines the rigidity percolation threshold. To identify the temperature of the phase transition T_p, concepts from percolation theory and network science were applied.[7,8]

3 Results and Discussion

In a first step, the general percolation behavior of the constraint networks was analyzed. The phase transition can be viewed as a rigid to flexible transition of the kind observed in network glasses. It is characterized by the decay of a large rigid cluster (the giant cluster) in the network.[4] The temperature of the phase transition relates to the melting temperature of the protein. A higher phase transition temperature was observed for approximately two-

thirds of the proteins from thermophilic organisms from our data set compared to their mesophilic counterparts (data not shown).

In a second step, the microscopic structure of the networks was related to their observed macroscopic behavior, in order to characterize stability features of the protein structures. For this, networks directly before and after the phase transition were compared. In Figure 1, the rigid cluster decomposition of mesophilic and thermophilic thermolysin-like protein (TLP) is shown. Figure 1a and c show the networks from mesophilic and thermophilic TLP directly before the phase transition, respectively. Apparently, the giant cluster dominates the system in both cases. Moreover, the giant cluster is located in the same region of the proteins: It extends over the N-terminal domain and comprises the β-sheet region and an α-helix in the N-terminal domain. After the phase transition, the giant cluster decays into smaller rigid clusters and regions that are flexible (Figure 1b and d). The close correspondence of the rigid cluster distribution in the networks of the homologous proteins before and after the phase transition is an intriguing result of our analysis.

Unfolding nucleus		**Predicted sites**	**Experimentally verified sites**[9]
I	β-sheet region in the N-terminal domain	21-24, 29, 31-34, 39-42, 44, 101-107, 114-115, 122-123	–
II	N-terminus of the α-helix in the N-terminal domain	68-70	69
III	Region around F63 in the N-terminal domain	54, 56-62	4, 56, 58, 63, 65

Table 1. Comparison of predicted with experimentally verified unfolding sites[9] for thermolysin-like protein (TLP).

In analogy to experimental protein unfolding, where initial unfolding of local regions precedes the denaturation of the entire protein, the loss of rigidity in certain regions is considered to precede the phase transition. These regions were identified as parts of the giant cluster that become flexible upon the phase transition, each representing an unfolding nucleus. In case of TLP, three unfolding nuclei could be found (Table 1, Figure 1). The predicted unfolding nuclei were compared with experimental data (Table 1). Notably, the predicted unfolding nuclei are in good agreement with sites where stabilizing mutations have successfully been introduced into TLP.[9] A likewise good agreement between our predictions and experimental data was found for many other proteins from our data set.

4 Concluding Remarks

Our findings strongly support the notion that the stability of thermophilic proteins is in general linked to an enhanced structural rigidity of the folded native state.[3] Furthermore, direct support is found for the corresponding states concept which states that homologous proteins exist in corresponding states of similar flexibility at their respective optimal temperature.[3] To the best of our knowledge, this is the first theoretical study addressing this

issue by directly probing the rigidity of protein structures along a thermal unfolding trajectory for a comprehensive dataset. Regarding the identification of regions that become flexible when approaching the phase transition (unfolding nuclei), we were encouraged to see the good (albeit not perfect) agreement between predicted sites and experimental mutations that led to higher structural stability. The result demonstrates that our approach will indeed be helpful to guide data-driven protein engineering to regions where mutations most likely will have a notable effect on thermostability.

References

1. D. C. Demirjian, F. Moris-Varas, C. S. Cassidy, *Enzymes from extremophiles*, Curr. Opin. Chem. Biol. **5**, 144–151, 2001.
2. J. F. Chaparro-Riggers, K. M. Polizzi, A. S. Bommarius, *Better library design: Data-driven protein engineering*, Biotechnol. J. **2**, 180-191, 2007.
3. R. Jaenicke, G. Böhm, *The stability of proteins in extreme environments*, Curr. Opin. Struct. Biol. **8**, 738-748, 1998.
4. S. Radestock, H. Gohlke, *Exploiting the link between protein rigidity and thermostability for data-driven protein engineering*, submitted.
5. D. J. Jacobs, A. J. Rader, L. A. Kuhn, M. F. Thorpe, *Protein flexibility predictions using graph theory*, Proteins **4**, 150–165, 2001.
6. A. J. Rader, B. M. Hespenheide, L. A. Kuhn, M. F. Thorpe, *Protein unfolding: Rigidity lost*, Proc. Natl. Acad. Sci. U.S.A. **99**, 3540–3545, 2002.
7. D. Stauffer, A. Aharony, *Introduction to percolation theory*, 2nd ed., Taylor and Francis, London, 1994.
8. I. R. Tsang, I. J. Tsang, *Cluster size diversity, percolation, and complex systems*, Phys. Rev. E **60**, 2684-2698, 1999.
9. B. van den Burg, B. W. Dijkstra, G. Vriend, B. Vandervinne, G. Venema, V. G. H. Eijsink, *Protein stabilization by hydrophobic interactions at the surface*, Eur. J. Biochem. **220**, 981-985, 1994.