

# Chapter 66

## Inquiring Protein Thermostability: Is Resistance to Temperature Stress a Rigidity/Flexibility Trade-off?

Maria Kalimeri, Simone Melchionna, and Fabio Sterpone

**Abstract** In this work, we are studying the behavior of two homologous hyperthermophilic and mesophilic proteins via molecular dynamics simulations at different temperatures and at timescales that reach up to hundreds of nanoseconds. A multi-disciplinary conformational analysis on the resulting trajectories, supports the idea that thermostability is induced by an interesting partition of flexible and rigid parts along the protein matrix and points out, to that end, the crucial role of electrostatic interactions.

**Keywords** Thermal stability · Protein flexibility · Molecular dynamics simulations

### 66.1 Introduction

Understanding how proteins maintain a stable fold in different thermodynamic conditions and how their motion at different length and time scales correlates to biological functions are of principal importance in biophysics. Here we are interested in a special class of proteins: thermophiles. These proteins can resist thermal stress being able to function at temperatures as high as 100 °C. The molecular origin of such special stability is still unknown and extremely appealing for technological applications, e.g. biotechnology or industrial chemical catalysis.

From the point of view of physics, thermophiles are a privileged case study for gaining insight on the main forces that keep a protein folded and functional. In particular it is key to inquire whether or not thermal resistance of thermophilic proteins

---

M. Kalimeri (✉) · F. Sterpone

Sorbonne Paris Cité, Laboratoire de Biochimie Théorique, Univ. Paris Diderot, UPR9080 CNRS, Paris, France

e-mail: [maria.kalimeri@ibpc.fr](mailto:maria.kalimeri@ibpc.fr)

F. Sterpone

e-mail: [fabio.sterpone@ibpc.fr](mailto:fabio.sterpone@ibpc.fr)

S. Melchionna

CNR-IPCF, Consiglio Nazionale delle Ricerche, Rome, Italy

e-mail: [simone.melchionna@roma1.infn.it](mailto:simone.melchionna@roma1.infn.it)

is caused by a special rigidity of the protein matrix as assumed by common belief. In order to tackle the complexity of this challenge we have designed a multi-scale strategy based on atomistic and coarse-grained Molecular Dynamics simulations as well as on the use of advanced techniques for sampling rare events and enhanced protein conformational changes.

Here we report the results of a preliminary, multidisciplinary analysis that supports the idea that thermostability is induced by an interesting partition of flexible and rigid parts along the protein matrix and points out, to that end, the crucial role of electrostatic interactions. We note that we quantify flexibility here in terms of atomistic deviations around a mean position.

### 66.1.1 Theoretical Background

Thermodynamically, protein stability relates to the energetics of the transition from the native state  $F$  to the unfolded state  $U$  :

$$\Delta G^{f \rightarrow u} = G^u - G^f = -k_b T \ln \left( \frac{\langle U \rangle}{\langle F \rangle} \right) \quad (66.1)$$

where  $G$  stands for the free energy, and  $\langle U \rangle$  and  $\langle F \rangle$  are the sizes of the populations occupying the unfolded and folded states, respectively. By increasing temperature the relative population of the unfolded state is favored. Thermophilic proteins are characterized by a high melting temperature or in other words the unfolded state is favored only at very high temperatures (80–90 °C).<sup>1</sup>

The molecular origin of such stability shift is not clear since, thermodynamically, several scenarios are plausible. In the simple two-state model [2] the higher melting temperature may result from (i) a larger  $\Delta G^{f \rightarrow u}$  difference characterizing the  $\langle U \rangle$  and  $\langle F \rangle$  populations at a comparable temperature, (ii) a slow variation of such difference as temperature increases, and finally (iii) from a shift of the temperature associated to the stability state (where the folded state is preferential) [13].

At the same time the free energy of unfolding is a result of a fine interplay between enthalpic and entropic forces:

$$\Delta G^{f \rightarrow u} = \Delta H^{f \rightarrow u} - T \Delta S^{f \rightarrow u} \quad (66.2)$$

The enhanced stability of thermophiles can be rationalized either by considering a favoring enthalpic contribution, or by being entropic in nature. In the former case the special packing of residues in the protein matrix or the higher internal connectivity ( $H$ -bonds, salt-bridges) have being invoked, while for the latter residual secondary structure in the unfolded state or enhanced flexibility of the folded state have both been proposed.

Molecular Dynamics is a powerful technique to explore the protein behavior in atomistic resolution via the direct integration of the classical equations of motion

---

<sup>1</sup>The melting temperature of mesophilic proteins is around 40 °C.

[5, 10, 11, 13–15]. Moreover it can be used in combination with advanced techniques or simplified coarse-grained models in order to explore the conformational many-fold landscape, the folding/unfolding process and gather information on the kinetics and thermodynamics of the system. In this respect it is also worth mentioning the strategic use of tools borrowed from the theory complex networks and systems to analyze protein dynamics and conformational landscape, e.g. protein internal contacts and h-bonds networks, Markov State Model for conformational transitions [4, 8, 10, 11, 15].

## 66.2 Results

In the following we present the preliminary results of our research on the flexibility/rigidity response of protein to thermal stress at different levels of spatial resolution (See also methodology section at the end).

**Stability vs. Unfolding** We first stress that by performing simulations in the hundred-nanosecond timescale and longer, in a range of physical temperature (25–100 °C) we verify a lower stability for a mesophilic versus a hyperthermophilic protein. At the working temperature of the hyperthermophilic homologue (85 °C) the former explores the early steps of the unfolding process while the latter maintains its fold stable (e.g. high secondary structure conservation and low deviation from crystallographic native state). This finding shows that the present Force Field for biomolecular simulation contains all the ingredients necessary to distinguish the different temperature-related stability of proteins.

**Atomistic Fluctuations** A first insight on how the flexibility/rigidity of the protein matrix changes upon thermal stress is recovered by a detailed and rigorous analysis of atomistic fluctuations. This can be performed routinely by computing the root mean square fluctuation (RMSF) of atomic positions along a trajectory after removing rigid body motions,

$$\text{RMSF}_i = \sqrt{\frac{1}{T} \sum_{t_j=1}^T (x_i(t_j) - \tilde{x}_i)^2} \quad (66.3)$$

where  $T$  is the total time and  $\tilde{x}$  is a reference position for particle  $i$ , usually the time-averaged one. In general such analysis is performed rather blindly without special care on the effective meaning of the observable. In particular RMSF measures the second moment of the distribution of atomic positions; this parameter is meaningful only if this distribution is approximately unimodal. At a long time-scale simulation since the protein experiences large conformational changes the above condition breaks down. It is then necessary to use a precise procedure for individuating the maximal length scale that—in an average sense—allows to compute correctly the atomistic fluctuations.

To that end, we follow the rigorous procedure introduced by Maragliano et al. [5]. The time window on which we perform our block sampling is about 350 ps. At longer time scales mean atomic positions start to experience many-fold localization. We compute the RMSF for backbone C-alpha atoms and perform block averages on several fragments of the trajectory. We find that, despite the fact that the magnitude of the observable is comparable for both proteins, there is a very intriguing difference in the partitioning of flexible (high RMSF) and rigid (low RMSF) fragments along the sequence among the two systems. In particular the latter shows a remarkable anti-correlating behavior in the RMSF between groups of neighboring residues that seems to be independent of the temperature, a sort of caging effect borrowing a concept from liquid state theory. For the hyperthermophilic protein flexible and rigid parts of the sequence alternate more frequently and regularly than in its mesophilic homologue. Thus a more regular distribution of rigid fragments possibly stops the energy flow along the protein matrix preventing progressive unfolding.

**Electrostatics** The above results are complemented by an analysis of electrostatics interactions which due to a surplus of charged amino acids for the hyperthermophile are considered to play a crucial role for thermal stability. We begin by observing, as expected, a substantially larger number of salt-bridges for the aforementioned system as well as a higher number of possible ionic pair combinations and thus extended salt-bridge clusters.

Furthermore, a big fraction of salt-bridges present in the crystal structure of the hyperthermophile show an exceptional stability which is not the case for the mesophile. These salt-bridges have been verified to be related with the less flexible parts of the matrix, thus it is possible that they act as clamps or stopping points, that way enhancing if not organizing the anti-correlating behavior of atomistic fluctuations.

We further continue with the calculation of the electrostatic characteristic path length (CPL) [17] which is defined as the average number of contacts needed to connect, along the shortest path, two randomly chosen nodes [15]. We follow the same technique as in Ref. [15]. The nodes of our system are the Ca-atoms and a connection exists if there exists an attractive electrostatic interaction between the respective residues (salt-bridge or hydrogen bond). Our results agree with the respective ones therein. Namely, the hyperthermophile reacts to the temperature increase by decreasing its CPL. The same is not true for the mesophilic protein. CPL is inversely proportional to the degree of electrostatic connectivity of the fold, thus this finding suggests a positive correlation between electrostatic connectivity and thermal stability.

**Collective Variables** The flexibility/rigidity of the protein matrix is also investigated via the construction of 2D free-energy landscapes representations over a set of collective variables such as the radius of gyration, the fraction of native contacts or the deviation of instantaneous protein configurations from the native state. For the folded state both proteins show a rather harmonic basin with comparable width. This finding supports again the idea that the thermophiles do not show a special rigidity.

Clearly a strong deviation is observed at high temperature when the thermophile is rather stable and the mesophile starts to unfold.

**Compressibility** Previous work has drawn the attention to an existing correlation between protein compressibility and stability [1]. Low compressibility generally correlates to an increased enthalpic stability and a suggested uniform core-to-surface distribution of charged amino-acids. We rigorously compute the compressibility of the protein as a function of temperature [6] and find that while at ambient temperature the two proteins show similar compressibility this is not true at higher temperature where the hyperthermophile shows a smaller one. We stress here that since compressibility relates to the fluctuations of protein volume with respect to that of the simulation cell, we take caution in extending the calculation only to the steady part of the trajectory. In other words at higher temperature we exclude the unfolding process of the mesophile. The volume of protein and its fluctuation are computed via the Voronoi tessellation of the space. We also point out that this precise evaluation of the atomistic volume allow us to check whether or not the mesophile and hyperthermophile are characterized by a different packing behavior, and this is not the case.

### 66.3 Conclusions

In this work we present a case for which a hyperthermophilic protein exhibits, in general, a comparable degree of flexibility in comparison to its mesophilic counterpart. The difference between the two systems however, concerns how flexibility/rigidity is partitioned in the protein matrix at the atomistic fluctuations timescale and how it relates to the distribution and number of key interactions (e.g. salt-bridge between charged amino acids). Moreover we report a clear difference in the response to thermal stress (as expected due to the different thermal stability), with the hyperthermophilic variant showing a systematic lower compressibility and increased electrostatic connectivity. In the coming months we will explore in more detail the configurational landscape of the two proteins considering the kinetics between local stable conformational states, hence gaining information on the relative distribution of free energy barriers separating local clusters of similar configuration.

Our present results are based on atomistic simulations. We are now refining a coarse-grain model in order to account in an effective way the specific ion-pair interactions that are key for thermostability. The coarse-grained potential has been extracted from atomistic simulations of charged amino acid pairs in dilute solution. The iterative Boltzmann inversion procedure allows the construction of an effective interaction that has been merged in the existing coarse-grained model for protein simulation OPEP [7]. After concluding tests on a small reference system, we will soon be able to investigate in detail the unfolding/folding process of thermophiles via this simplified, hence low-time consuming model.

**Methodology** We realize molecular dynamics simulations of two homologue proteins, hyperthermophile and mesophile gdomains of elongation factor Tu using the NAMD package [9]. The employed force field is Charmm22 [3] with the TIP3P model for water molecules. Initial coordinates for both systems were obtained from the crystal structures found in Protein Data Bank (PDB) after isolating the amino acid stretches of each protein's gdomain. The crystallographic PDB codes for the mesophilic and the hyperthermophilic species are 1EFC and 1SKQ, respectively [12, 16].

**Acknowledgements** The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 258748).

## References

1. Dadarlat VM, Post CB (2003) Adhesive-cohesive model for protein compressibility: an alternative perspective on stability. *Proc Natl Acad Sci USA* 100:14778–14783
2. Feler G (2010) Protein stability and enzyme activity at extreme biological temperatures. *J Phys Condens Matter* 22:323101
3. MacKerell AD Jr., Brooks B, Brooks CL III, Nilsson L, Roux B, Won Y, Karplus M (1998) CHARMM: the energy function and its parameterization with an overview of the program. *Encycl Comput Chem* 1:271–277
4. Mann M, Klemm K (2011) Efficient exploration of discrete energy landscapes. *Phys Rev E* 83:011113
5. Maragliano L, Cottone G, Cordone L, Ciccotti G (2004) Atomic mean-square displacements in proteins by molecular dynamics: a case for analysis of variance. *Biophys J* 86(5):2765–2772
6. Marchi M (2003) Compressibility of cavities and biological water from Voronoi volumes in hydrated proteins. *J Phys Chem B* 107(27):6598–6602
7. Maupetit J, Tuffery P, Derreumaux P (2007) A coarse-grained protein force field for folding and structure prediction. *Proteins* 69(2):394–408
8. Numata J, Wan M, Knapp EW (2007) Conformational entropy of biomolecules: beyond the quasi-harmonic approximation. *Genome Inf* 1:192–205
9. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802
10. Prada-Gracia D, Gómez-Gardeñes J, Echenique P, Falo F (2009) Exploring the free energy landscape: from dynamics to networks and back. *PLoS Comput Biol* 5(6):e1000415
11. Rao F, Garrett-Roe S, Hamm P (2010) Structural inhomogeneity of water by complex network analysis. *J Phys Chem B* 114:15598–15604
12. Song H, Parsons MR, Rowsell S, Leonard G, Phillips SE (1999) Crystal structure of intact elongation factor EF-Tu from *Escherichia coli* in GDP conformation at 2.05 Å resolution. *J Mol Biol* 285:1245–1256
13. Sterpone F, Melchionna S (2012) Thermophilic proteins: insight and perspective from in silico experiments. *Chem Soc Rev* 41:1665–1676
14. Sterpone F, Bertonati C, Briganti G, Melchionna S (2009) Key role of proximal water in regulating thermostable proteins. *J Phys Chem B* 113(1):131–137

15. Tavernelli I, Cotesta S, Di Iorio EE (2003) Protein dynamics, thermal stability, and free-energy landscapes: a molecular dynamics investigation. *Biophys J* 85(4):2641–2649
16. Vitagliano L, Ruggiero A, Masullo M, Cantiello P, Arcari P, Zagari A (2004) The crystal structure of *Sulfolobus solfataricus* elongation factor 1alpha in complex with magnesium and GDP. *Biochemistry* 43:6630–6636
17. Watts DJ, Strogatz SH (1998) Collective dynamics of “small world” networks. *Nature (London)* 393:440–442