THE JOURNAL OF PHYSICAL CHEMISTRY Letters

pubs.acs.org/JPCL

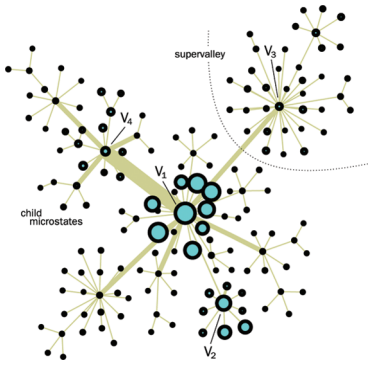
Local Transition Gradients Indicating the Global Attributes of Protein Energy Landscapes

Francesco Rao*

Laboratoire de Chimie Biophysique/ISIS, Universite de Strasbourg, 8 alle G. Monge, Strasbourg, France

ABSTRACT The relationship between local atomic fluctuations and global protein rearrangements is elusive. Here, molecular dynamics simulations of a peptide concerted with complex network analysis show that local transition gradients provide a quantitative description of the free-energy landscape in terms of its valleys. The latter, compared with established methods, are found with the correct populations. Moreover, the iterative analysis of the fastest intervalley relaxations provides an intuitive tree-like representation of the system dynamics. These results indicate that knowledge of the fast local relaxations is sufficient to identify the general properties of free-energy surfaces in terms of basins of attraction and the hierarchy of transitions.

SECTION Macromolecules, Soft Matter



In complex systems, the behavior of the whole is hardly predictable from the fundamental laws of interactions of the single components.[1] As a consequence, it is hard to reveal the intimate connection between local properties and the global behavior of such systems. This problem has been formalized in the context of complex networks[2,3] as a question of "navigability", that is, the mechanism for the efficient flow of information when the single network nodes do not have a global view of the overall topology.

The study of protein dynamics involves a similar problem—the coupling between the fast atomic fluctuations, which are local, and the slow conformational changes, which are global.[4] Those dynamical aspects have been recently recognized to be crucial for protein function, playing an important role in signaling, allosteric pathways, and enzymatic reactions.[5−7] Molecular dynamics (MD) simulations are playing an increasing role in complementing the experimental results[8−10] which supply useful, but limited, information to this question. Most descriptions of the free-energy surface governing protein dynamics have been rather qualitative because of the lack of proper order parameters and the intrinsic multidimensionality of the problem.[11−14] These limitations have triggered the development of a completely new arsenal of tools inspired by network theory.[15−19] A useful approach maps the protein trajectory, obtained by computer simulations or experiments, on a conformation space network (CSN), whose nodes represent the different microstates and whose links correspond to direct transitions between them.[17,20,21] This approach has been successfully applied to the study of peptide folding and structural transitions,[17,18,21−24] as well as to interpret electron-transfer experiments[25] and time-resolved IR measurements.[26,27]

In this Letter, the relation between the local properties of the free-energy landscape and its global architecture is investigated by MD simulations and complex network analysis of a four residue peptide, $(GlySer)_2$. In particular, when the CSN is reduced to the subgraph containing only one link per microstate pointing toward the most probable transition (i.e., following the transition gradient), the presence of energy valleys and subvalleys and their equilibrium populations is naturally extracted as well as the hierarchy of transitions between them. Hence, the fast local motions build up the backbone of the global interconversion pathways. The observed coupling between local and global dynamical properties is expected to occur in a large class of complex systems.

The relation between fast local modes and global chain rearrangements is investigated by constructing from the full CSN a new network with a reduced number of links. For each network node, the transition with the highest probability (excluding self-interactions) is kept, and all of the others are deleted (Figure 1, inset). These transitions define a gradient in the network dynamics and result in a partitioning of the network into several disconnected minimum-spanning trees,[28] called here gradient-clusters for convenience. (Gradient networks were originally introduced by Toroczkai and Bassler to study jamming,[29] though, in their case, the gradient is defined on the nodes as a quenched scalar field.) The physical justification for this approach is that following the pathways defined by the most probable transitions leads to microstates lower and lower in free energy, resulting in a kind of "steepest descent pathway" on the free-energy surface. High-energy microstates, in the neighborhood of a free-energy barrier, would connect either to one valley or another. As a matter of fact, the network characterizing the free-energy surface is split into a set of disconnected minimum-spanning trees representing the local attractors of the system dynamics.

Received Date: March 30, 2010
Accepted Date: April 27, 2010
Published on Web Date: April 29, 2010

DOI: 10.1021/jz1004158 | *J. Phys. Chem. Lett.* **2010**, *1*, 1580–1583

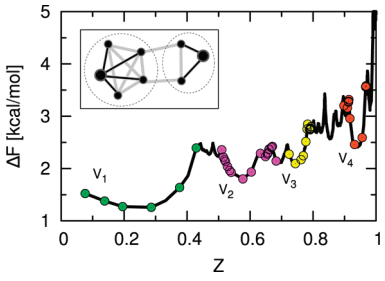


Figure 1. Cut-based free-energy profile for the $(GlySer)_2$ peptide. Microstates belonging to the four gradient clusters are shown on the profile with different colors. In the inset, a schematic representation of the gradient approach is presented. The underlying network is depicted in gray, while the most probable transition for each node is shown in black.

To better understand the nature of the gradient clusters (161 in total), a cut-based free-energy profile (CFEP)[30] is calculated on the CSN and compared to the output of the gradient partition. The CFEP is a well-established technique.[30−33] It results in the projection of the free energy on a partition-function-based reaction coordinate Z. The method is based on a network flux analysis following the idea that network regions of minimum flow correspond to transition states. In Figure 1, the calculated CFEP is shown. The profile reveals that the peptide free-energy landscape is rugged. Remarkably, the obtained gradient clusters coincide with the valleys shown by the CFEP. In Figure 1, the four most populated gradient clusters (called V_1, V_2, V_3, and V_4) do overlap with single minima of the free-energy profile. Hence, gradient clusters correspond to the valleys of the free-energy landscape. It is interesting to note that the four valleys include both compact as well as extended structures, like the ones in V_3 (end-to-end distance, $d_{ee} \approx 10.9$ Å) that are entropically stabilized. A quantitative comparison of the populations calculated using the gradient approach with one of the most accurate methods for this type of analysis[18,20] — the minimum cut method — shows very good agreement (Table 1). These results indicate that the gradient approach effectively reproduces the correct thermodynamics of the system. This is particularly relevant since the approach does not use any global property of the system, either in terms of barrier height or microstates energy. On the other hand, the CFEP and the minimum cut method do perform a global analysis of the network. These observations suggest that (i) the information stored in the most probable (i.e., the fastest) transition is sufficient to provide an accurate partition of the free-energy surface into its valleys (the latter representing the metastable states of the dynamical system) and (ii) the gradient approach can be applied as an automatic, computationally inexpensive method to find the metastable states characterizing the system dynamics[30,34] (in fact, it scales linearly with the number of microstates, and it is faster with respect to other approaches which require the analysis of the full transition matrix).

It is worth noting that the gradient approach is based on a strong approximation and cannot be used to accurately estimate barrier heights, that is, the transition probabilities, between the detected metastable states. Those probabilities are sensible for a few wrong assignments of the microstates lying on top of the barrier. Nevertheless, an efficient procedure to estimate intervalley transition probabilities is to count, directly along the trajectory, the number of transitions between the microstates lying at the bottom of the found valleys.[31,33] This technique represents an accurate estimation of the transition probabilities and has shown to be quite successful even when the definition of the microstates was not optimal.[31] Indeed, combination of the gradient approach with this technique will provide accurate Markov models to describe the system dynamics at a coarse level.

Table 1. Summary of Thermodynamical and Dynamical Properties of the Four Most Important Gradient Clusters/Valleys of the $(GlySer)_2$ Peptide

valley	gradient populations[a]	minimum cut populations	MFPT [ps][b]	RMSD [Å][c]	$\langle d_{ee} \rangle$ [Å][d]
V_1	0.429	0.428		0.0	4.167
V_2	0.120	0.120	30.74	1.78	4.253
V_3	0.068	0.064	34.16	2.40	10.960
V_4	0.037	0.051	59.96	1.61	4.160

[a] To take into account entropic effects, gradient clusters separated by free-energy barriers smaller than $k_BT/10$ have been merged. [b] The mean first passage time MFPT was calculated on the MD trajectory including the relaxations from all of the "child" microstates of the V_2, V_3, and V_4 supervalleys (Figure 2), respectively. [c] The RMSD is calculated between the most populated microstate belonging to V_1 and the most populated microstate belonging to the gradient cluster under consideration. [d] Average value of the end-to-end distance d_{ee} in the considered gradient cluster.

The gradient approach can be applied in an iterative, hierarchical fashion when the gradient clusters are considered themselves as nodes of a higher-level CSN. In this case, the nodes and the links of the network are the gradient clusters and the connections between them, respectively. The iteration can be applied recursively until all of the microstates are merged into one cluster, which is represented as a minimum-spanning forest. This tree-like structure is shown in Figure 2 for the $(GlySer)_2$ peptide. Link widths represent the iteration step at which the edge was introduced. At each iteration, the newly obtained clusters represent a partition of the free-energy landscape into a set of so-called supervalleys. The later the appearance of a supervalley in the iteration process, the slower the relaxation to the global root of the tree (i.e., the most populated microstate, in this case, V_1). For example, V_1 and V_2 valleys are merged at the first iteration, indicating that they interconvert rapidly. On the other hand, V_4 is merged to V_1 at the fourth iteration because the microstates belonging to the V_4 supervalley have slower relaxations to V_1. These observations are in agreement with the values calculated along the trajectory of the mean first passage times to reach V_1 (MFPT) reported in Table 1. The gradient minimum-spanning forest represents, in an intrinsically multidimensional fashion, the valleys of the free-energy landscape and their dynamical organization in fast (star-like structures) and slow relaxations (larger link widths). This scenario is in agreement with the early works of Frauenfelder on Myoglobin.[4]

Concluding, we found that for an all-atom peptide MD simulation, the fast local relaxations produce a partition of the conformational space into disconnected minimum-spanning trees. This partition reproduces the organization of the free-energy

DOI: 10.1021/jz1004158 | J. Phys. Chem. Lett. 2010, 1, 1580–1583

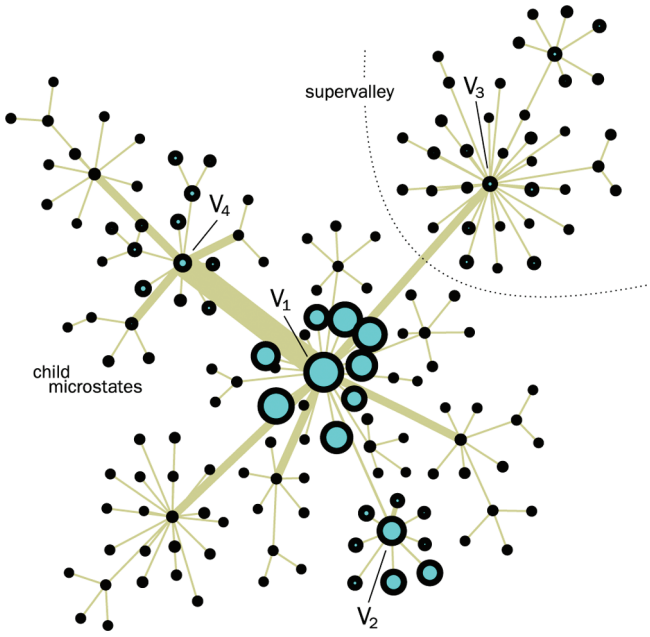


Figure 2. Gradient minimum-spanning forest of the $(GlySer)_2$ peptide energy landscape. Larger link widths indicate slower relaxations (see main text for details). The size of the nodes is proportional to the microstate population (to avoid overcrowding, only nodes with populations larger than 10^{-3} are shown; this subset of nodes accounts for 83% of the total population).

landscape into valleys with the correct populations. The iterative connection of those valleys into a minimum-spanning forest recovers the global architecture of the dynamics occurring on the free-energy landscape. A similar coupling between local and global dynamics is expected to take place in other complex systems.

This work enhances our understanding of the relationship between local atomic fluctuations and global macromolecular rearrangements. The extrapolation of the reported results to proteins may not be straightforward, but the quantitative understanding of simpler, yet complex, systems is a necessary step toward this goal.

METHODS

GlySer peptides have been used for quite some time as flexible linkers (and are known to show poor secondary structure) for polypeptide dynamics.[35,36] MD simulations, using the Langevin algorithm with a friction coefficient equal to 0.6 ps^{-1}, were calculated with the CHARMM program,[37] and the polar hydrogen energy function (PARAM19) was used. The effects of water were included using the generalized Born FACTS implicit solvation model.[38] SHAKE was employed so that an integration step of 2 fs could be used. A trajectory of 280 ns at 340 K was obtained for a total of 10^6 saved conformations. During the simulation, the peptide visited several different conformations characterized by an end-to-end distance d_{ee} between 3.1 and 12.7 Å, indicating large structural fluctuations. The length of the trajectory is long enough to provide accurate sampling. For example, the transition from valley V_1 to V_2 is seen 1695 times in one way and 1730 in the other way. A less visited transition such as V_1 to V_4 is still visited 151 times in one way and 159 in the other way. These data show that equilibrium is reached because the forward and backward transitions are visited with similar populations. Consequently detailed balance is satisfied at the intervalley level.

The peptide microstates are defined as the inherent structures (IS), that is, the potential energy minima of the system.[39,40] The IS are natural, physically meaningful partitions into microstates.[33,41−43] They were calculated minimizing all of the 10^6 snapshots along the trajectory, resulting in 3044 different IS. This calculation is as computationally expensive as other procedures to define the system microstates. Moreover, it is embarrassingly parallel and, in the present case, requires a fraction of a second per snapshot. The conformation space network (CSN) is built on top of this microstate definition: the nodes and the links are the microstates (i.e., the IS) and their direct transitions observed during the MD trajectory, respectively. The obtained network is weighted, and it is equivalent to a classical transition matrix when the columns of the adjacenct matrix (i.e., the network links) are appropriately normalized to one.

AUTHOR INFORMATION

Corresponding Author:

*To whom correspondence should be addressed. E-mail: francesco.rao@gmail.com.

ACKNOWLEDGMENT The author wish to thank Prof. A. Caflisch, Prof. M. Karplus, Dr. S. Muff, Dr. S. Krivov, and Dr. M. Cecchini for useful comments and discussions and Prof. P. Hamm and Dr. S. Garrett-Roe for critically reading this manuscript. CFEPs were calculated using the program WORDOM.[44] This research was partially funded by the European Molecular Biology Organization (EMBO).

REFERENCES

(1) Anderson, P. More is Different. *Science* **1972,** *177*, 393–396.
(2) Boguñá, M.; Krioukov, D.; Claffy, K. Navigability of Complex Networks. *Nat. Phys.* **2008,** *5*, 74–80.
(3) Boguñá, M.; Krioukov, D. Navigating Ultrasmall Worlds in Ultrashort Time. *Phys. Rev. Lett.* **2009,** *102*, 58701.
(4) Frauenfelder, H.; Sligar, S.; Wolynes, P. The Energy Landscapes and Motions of Proteins. *Science* **1991,** *254*, 1598–1603.
(5) Kern, D.; Zuiderweg, E. R. The Role of Dynamics in Allosteric Regulation. *Curr. Opin. Struct. Biol.* **2003,** *13*, 748–577.
(6) Eisenmesser, E. Z.; et al. Intrinsic Dynamics of an Enzyme Underlies Catalysis. *Nature* **2005,** *438*, 117–121.
(7) Benkovic, S. J.; Hammes, G. G.; Hammes-Schiffer, S. Free-Energy Landscape of Enzyme Catalysis. *Biochemistry* **2008,** *47*, 3317–3321.
(8) Boehr, D. D.; McElheny, D; Dyson, H. J.; Wright, P. E. The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* **2006,** *313*, 1638–1642.
(9) Schuler, B; Eaton, W. A. Protein Folding Studied by Single-Molecule FRET. *Curr. Opin. Struct. Biol.* **2008,** *18*, 16–26.
(10) Colletier, J. P.; et al. Shoot-and-Trap: Use of Specific X-ray Damage to Study Structural Protein Dynamics by Temperature-Controlled Cryo-crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 11742–11747.

DOI: 10.1021/jz1004158 | *J. Phys. Chem. Lett.* **2010,** *1*, 1580–1583

(11) Du, R.; Pande, V.; Grosberg, A.; Tanaka, T.; Shakhnovich, E. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998,** *108*, 334–350.
(12) Pande, V.; Grosberg, A.; Tanaka, T.; Rokhsar, D. Pathways for Protein Folding: Is a New View Needed? *Curr. Opin. Struct. Biol.* **1998,** *8*, 68–79.
(13) Hegger, R.; Altis, A.; Nguyen, P.; Stock, G. How Complex Is the Dynamics of Peptide Folding? *Phys. Rev. Lett.* **2007,** *98*, 28102.
(14) Maisuradze, G.; Liwo, A.; Scheraga, H. How Adequate Are One-and Two-Dimensional Free Energy Landscapes for Protein Folding Dynamics? *Phys. Rev. Lett.* **2009,** *102*, 238102.
(15) Doye, J. Network Topology of a Potential Energy Landscape: A Static Scale-Free Network. *Phys. Rev. Lett.* **2002,** *88*, 238701.
(16) Evans, D.; Wales, D. the Free Energy Landscape and Dynamics of Met-enkephalin. *J. Chem. Phys.* **2003,** *119*, 9947.
(17) Rao, F.; Caflisch, A. The Protein Folding Network. *J. Mol. Biol.* **2004,** *342*, 299–306.
(18) Krivov, S.; Karplus, M. Hidden Complexity of Free Energy Surfaces for Peptide (Protein) Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101*, 14766–14770.
(19) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. Hierarchical Analysis of Conformational Dynamics in Biomolecules: Transition Networks of Metastable States. *J. Chem. Phys.* **2007,** *126*, 155102.
(20) Caflisch, A. Network and Graph Analyses of Folding Free Energy Surfaces. *Curr. Opin. Struct. Biol.* **2006,** *16*, 71–78.
(21) Gfeller, D.; De Los Rios, P.; Caflisch, A.; Rao, F. Complex Network Analysis of Free-Energy Landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104*, 1817–1822.
(22) Settanni, G.; Fersht, A. R. High Temperature Unfolding Simulations of the TRPZ1 Peptide. *Biophys. J.* **2008,** *94*, 4444–4453.
(23) Yang, S.; Roux, B. Src Kinase Conformational Activation: Thermodynamics, Pathways, and Mechanisms. *PLoS Comput. Biol.* **2008,** 4.
(24) Prada-Gracia, D.; Gómez-Gardeñes, J.; Echenique, P.; Falo, F. Exploring the Free Energy Landscape: From Dynamics to Networks and Back. *PLoS Comput. Biol.* **2009,** *5*, e1000415.
(25) Li, C. B.; Yang, H.; Komatsuzaki, T. Multiscale Complex Network of Protein Conformational Fluctuations in Single-Molecule Time Series. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 536–541.
(26) Ihalainen, J. A.; et al. Folding and Unfolding of a Photoswitchable Peptide from Picoseconds to Microseconds. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104*, 5383–5388.
(27) Ihalainen, J. A.; et al. Alpha-Helix Folding in the Presence of Structural Constraints. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 9588–9593.
(28) Carmi, S.; Krapivsky, P.; Ben-Avraham, D. Partition of Networks into Basins of Attraction. *Phys. Rev. E* **2008,** *78*, 066111.
(29) Toroczkai, Z.; Bassler, K. Jamming is Limited in Scale-Free Systems. *Nature* **2004,** 716.
(30) Krivov, S. V.; Karplus, M. One-Dimensional Free-Energy Profiles of Complex Systems: Progress Variables That Preserve the Barriers. *J. Phys. Chem. B* **2006,** *110*, 12689–12698.
(31) Krivov, S. V.; Muff, S.; Caflisch, A.; Karplus, M. One-Dimensional Barrier-Preserving Free-Energy Projections of a Beta-Sheet Miniprotein: New Insights into the Folding Process. *J. Phys. Chem. B* **2008,** *112*, 8701–8714.
(32) Muff, S.; Caflisch, A. Identification of the Protein Folding Transition State from Molecular Dynamics Trajectories. *J. Chem. Phys.* **2009,** *130*, 125104.
(33) Rao F., Karplus M. Protein Dynamics by Inherent Structure Analysis. To be published.
(34) Chodera, J.; Singhal, N.; Pande, V.; Dill, K.; Swope, W. Automatic Discovery of Metastable States for the Construction of Markov Models of Macromolecular Conformational Dynamics. *J. Chem. Phys.* **2007,** *126*, 155101.
(35) Bieri, O.; et al. the Speed Limit for Protein Folding Measured by Triplet−Triplet Energy Transfer. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96*, 9597.
(36) Möglich, A.; Joder, K.; Kiefhaber, T. End-to-End Distance Distributions and Intrachain Diffusion Constants in Unfolded Polypeptide Chains Indicate Intramolecular Hydrogen Bond Formation. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 12394.
(37) Brooks, B. R.; et al. CHARMM: a Program for Macromolecular Energy, Minimization, And Dynamics Calculations. *J. Comput. Chem.* **1983,** *4*, 187–217.
(38) Haberthur, U.; Caflisch, A. FACTS: Fast Analytical Continuum Treatment of Solvation. *J. Comput. Chem.* **2008,** *29*, 701–715.
(39) Stillinger, F.; Weber, T. Hidden Structure in Liquids. *Phys. Rev. A* **1982,** *25*, 978–989.
(40) Stillinger, F.; Weber, T. Dynamics of Structural Transitions in Liquids. *Phys. Rev. A* **1983,** *28*, 2408–2416.
(41) Baumketner, A.; Shea, J. E.; Hiwatari, Y. Glass Transition in an off-Lattice Protein Model Studied by Molecular Dynamics Simulations. *Phys. Rev. E* **2003,** *67*, 011912.
(42) Kim, J.; Keyes, T. Inherent Structure Analysis of Protein Folding. *J. Phys. Chem. B* **2007,** *111*, 2647–2657.
(43) Nakagawa, N.; Peyrard, M. The Inherent Structure Landscape of a Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 5279–5284.
(44) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. WORDOM: A Program for Efficient Analysis of Molecular Dynamics Simulations. *Bioinformatics* **2007,** *23*, 2625–2627.

DOI: 10.1021/jz1004158 | *J. Phys. Chem. Lett.* **2010,** *1*, 1580–1583