Dynamic complexity of protein folding: computational mechanics on classical molecular trajectories

Dmitry Nerukh**[*]**, ***George Karvounis***, *and* ***Robert C. Glen***,
Unilever Centre for Molecular Informatics, Department of Chemistry,
Cambridge University, Cambridge CB2 1EW, UK

Self-organisation and complex behaviour are ubiquitous in molecular systems. A well-known and vitally important example is the ability of many proteins to spontaneously fold into a consistent and reproducible three-dimensional shape fundamental to their biological function. By a process that is reproducible and yet chaotic, the dynamics of the system is such that non-trivial structures, such as protein secondary structure spontaneously emerge with time. Although this type of behaviour ('emergence') has been known for many years, current analytical tools are inadequate and indeed, the mathematical framework necessary to investigate such systems is still being developed.

Our ongoing research consists of developing this methodology; specifically for the quantitative estimation of the complexity of real physical systems whose dynamics can be measured or computed. In this way we can understand the underlying physical processes that generate this behaviour to obtain radically new insights and predict the emergence and evolution of complexity.

We have designed a fundamentally new approach, based on a complexity measure, to reveal the underlying structure of the dynamics of the inter-particle interactions in molecular systems that lead to self-organising, structuring behaviour. For our purposes we adopted the approach by Crutchfield ***et. al.*** termed "computational mechanics"[1]. This approach combines and implements the ideas from Shannon entropy and Kolmogorov-Chaitin algorithmic complexity theories.

We show that the *general mathematical apparatus* of computational mechanics can be modified to cover *classical Hamiltonian systems*. The system's trajectories in phase space posses various values of complexity depending on the degree of chaoticity of different regions and can be link to local Lyapunov exponents of the system. This opens a fundamentally new route to quantify the emergence in such systems since the instantaneous characteristics of the dynamics can now be explored and the whole method acquires predicting power. This general formulation indicates that large regions of the phase space can be excluded from the analysis based on the general characteristics without recourse to expensive simulations.

The latter constitutes the areas of research that is focused on the use of the general approach applied to the specificity of molecular systems in order to estimate the *complexity of the full-dimensional dynamics*. The complexity is a direct measure of self-organisation, thus, these investigations reveal the potential mechanisms that can lead to emergent behaviour in molecular systems.

From the other hand the complexity of *low-dimensional projections* of the system's trajectories can be analysed based on the general theory described above. The particular approximations inherent to molecular systems' dynamics are used to project the full-dimensional phase space trajectory into low-dimensional subspaces. This is vitally important for realistic systems' applications since it is only feasible to analyse the trajectories in a very limited dimensionality subspaces of the phase space.

Connecting the above two areas the multivariate probability theory concept called "copulas" is used to establish the *link between the whole-dimensional dynamics and its low-dimensional projections*. We show that this approach can be used to calculate the complexity of the former using the latter. It is demonstrated that the dynamics of relatively small clusters of the atoms can be used to reduce the 6N-variate probabilities to a product of only few degrees of freedom -variate probabilities. This result forms a basis for extremely important techniques of estimating the complexity of high-dimensional molecular system utilising the information of a low-dimensional (computationally feasible) subsystems' dynamics.

[*] author for correspondence, email: dn232@cam.ac.uk

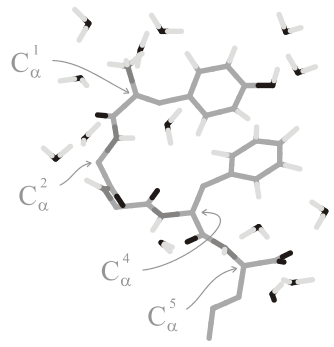

Fig. 1 Leu-Enkephalin molecule with hydrogen bonded waters at the moment of transition. The colours represent (from grey to black): H, C, N, O atoms

We have demonstrated that our approach provides new information about molecular dynamics. The reorientation of water molecules in the liquid phase has been analysed focusing specifically on estimating complexity of their molecular dynamics. We found that the motion of water molecules in the bulk and close to the ion possess dynamics having considerably different complexity[2]. Another class of molecular systems, zwitterions, demonstrates a simplified "folding" behaviour by forming a loop-like structure. The analysis of molecular motions provides evidence of different values of complexity in different dynamic states of the system. We found that precisely at the moment of "folding" the complexity rises to a higher level when the molecule stabilizes in the "folded" conformation[3]. The analysis is extracting information on the emergence of a new property of the system that is unavailable from any other analysis technique.

Finally our research has focused on applying the approach to small peptides (enkephalins) in explicit water (Fig. 1). We have concentrated on one of the key elementary events in protein folding – a β-turn formation. A decisively important role of the water network emerges and we observe that the complexity of the water dynamics around the peptide before, at, and after the moment of turn formation has different values. Analysis of various characteristics (dipole reorientation, lifetimes, etc.) of the water dynamics supports the hypothesis that water network around the peptide "freezes" before the transition takes place and becomes very agile at the moment of transition[4]. We now can explain and link this behaviour to a general theory of transition dynamics in Hamiltonian systems. It turns out that complexity measure uniquely indicates the less chaotic character of the system's dynamics (for all the peptides atoms and the majority of the solvation shell waters) specifically at the very short moment of β-turn formation[5] (Fig. 2). This is the first evidence supporting the very general hypothesis for realistic high-dimensional molecular systems.

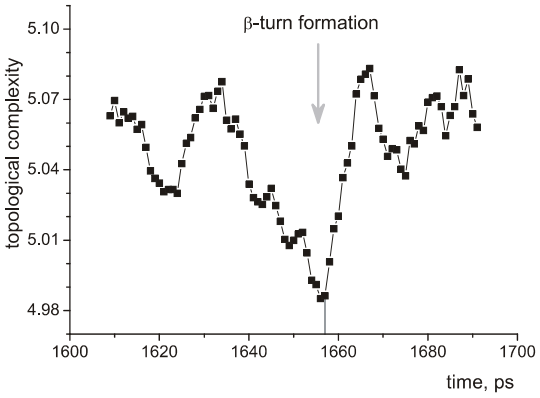

Fig. 2 Topological complexities of the waters' atoms. The symbolisation alphabet of size 32 and history length of 4 ps were used. The β-turn transition is at 1657 ps

These results will have profound implications for the understanding of protein folding, in particular, the driving forces behind the folding processes. Additionally, this opens up the possibility of designing new experiments (e.g. NMR experiments on bound water) that could add weight to the analyses.

[1] J.P. Crutchfield, D.P. Feldman, and C.R. Shalizi, *Phys. Rev. E*, **62**, 2996 (2000)

[2] D. Nerukh, G. Karvounis, and R.C. Glen, *J. Chem. Phys.*, **117**(21), 9611-9617 (2002); D. Nerukh, G. Karvounis, and R.C. Glen, *J. Chem. Phys.*, **117**(21), 9618-9622 (2002)

[3] D. Nerukh, G. Karvounis, and R.C. Glen, *Complexity*, **10**(2), (2005)

[4] G. Karvounis, D. Nerukh, and R.C. Glen, *J. Chem. Phys.*, **121**(10), 4925-4935 (2004); D. Nerukh and A.V. Luzanov, **Chirality measure of Leu-Enkephalin's conformational transition during β-turn formation**, submitted for publication in *PCCP*

[5] D. Nerukh, G. Karvounis, and R.C. Glen, **Dynamic complexity of chaotic transitions in high-dimensional classical dynamics: Leu-Enkephalin folding**, submitted for publication in *J. Chem. Phys.*