Statistical complexity of classical molecular dynamics systems: a measure of phase space exploration and non-Markovian behaviour of molecular conformations

Christian H. Jensen, Dmitry Nerukh and Robert C. Glen

Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge

Introduction

Phase Space Exploration

Molecular Dynamics simulated trajectories of molecular systems form complex spatio-temporal structures at the picoseconds time scale. Traditional methods of correlation analysis, as well as more advanced approaches based on Takens embedding theorem and surrogate time series analysis [1] used in dynamical systems theory for detecting hidden dynamic correlations in experimental data indicate the purely stochastic nature of molecular trajectories. However, we have shown that the intrinsic complexity, which is the basis of the formation of molecular structures on much longer time scales, can be quantified using a measure of Statistical Complexity [2-4]. The method estimates the information contained in the molecular trajectory by detecting and quantifying temporal patterns present in the analysed time series. Two types of temporal patterns are found for bulk water molecules. The first is defined by the short-time correlations corresponding to the velocity autocorrelation decay times (≈0.1 ps) and remains unchanged for the time intervals up to the order of tens of nanoseconds. The second is caused by previously unknown long-time correlations (found at longer than the nanoseconds time scale) and has complicated dynamics that slowly change with time. Also a direct measure is introduced that describes how the trajectory explores the phase space. The measure is based on Statistical Complexity and does not depend on the properties of the particular molecular system used.

Non-Markovian Behaviour

A Molecular Dynamics simulation of the four-residue peptide VPAL (Valine - Proline - Alanine - Leucine) has also been performed. The trajectory is subjected to the Statistical Complexity analysis. This is based on clustering the conformational space into states. The results show non-Markovian memory effects during folding transitions.

Methods

Molecular Dynamics

In Molecular Dynamics, the trajectories of the different atoms are found by solving Newton's second law for the system:

$$m\vec{a}_i = \vec{F}_i$$

Here *i* runs over all the atoms. m_i is the mass of atom *i*, a_i is the acceleration and F_i is the force exerted on the atom. The force is derived from the potential energy, $V(r_1 \ldots r_n)$, in the following way:

$$\vec{F}_i = -\nabla_{\vec{r}_i} V$$

To keep the temperature constant during simulation, the system is allowed to exchange energy with an infinite heat bath. This will cause the probabilities for states with energy *E* to occur with the Boltzmann probability $exp(-E/k_BT)$ in the simulation. Sampling conducted in this way will be in the canonical ensemble (or *NVT*). There are several numerical schemes which implement this ensemble in Molecular Dynamics and for obvious reasons these are called thermostats

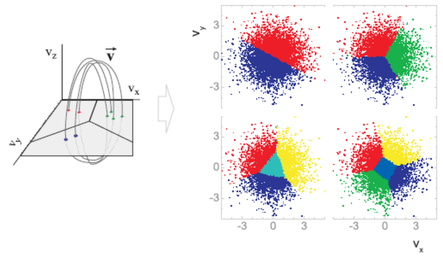
Complexity Analysis

Computational Mechanics is a method which predicts a signal based on its past. The foundation of Computational Mechanics was laid in 1989 with the work of J. P. Crutchfield and K. Young [2]. In Computational Mechanics, a pattern basis is generated for the signal which enables its optimal prediction. The pattern basis contains several elements and these are termed causal states. The prediction obtained from the pattern basis is achieved by an ε-machine which describes the dynamics of the system. Using the ε-machine, it is possible to calculate the statistical complexity. This is defined as the Shannon entropy of the causal states:

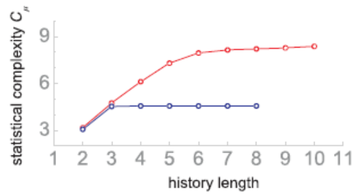
$$C_\mu(S) = H[S]$$

This determines the minimum average amount of memory required to optimally predict the future of a sequence.

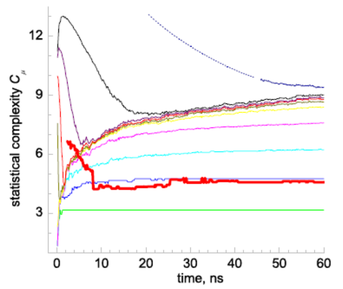
Phase Space Exploration



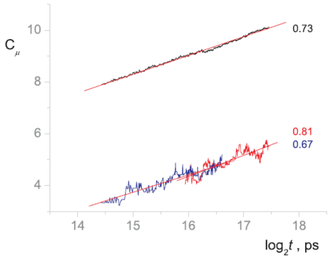
The system under investigation is bulk water. Simulations were performed by Molecular Dynamics at different temperatures. The figure shows approximations for generating partitions obtained using the method by Buhl and Kennel for the cross section of the hydrogen velocity space for the partitions corresponding to 2, 3, 4, and 5-symbols alphabet.



Statistical complexity vs. the length of histories (dimension of the phase space) for the original data (red) and the surrogate (blue), t = 30 ns.



Statistical complexity against time for the hydrogen velocity signal and the surrogate. The curves, from bottom to top, correspond to the values of the history length l from 2 to 11. The l = 11 curve does not settle on the logarithmic part within the shown area but seems to follow the same trend. Heavy red line is the C_μ values for the phase-shuffled surrogate signal (l = 9). For all curves the alphabet size K is equal to 3.



h_Q values (indicated on the right) for various observables: black - the hydrogen velocity, red - the oxygen velocity, and blue - the instantaneous temperature.

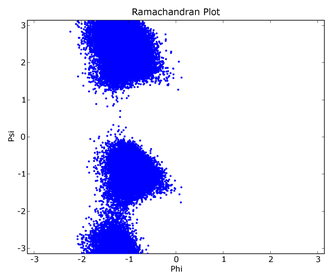
Since the growth of C_μ has a clear logarithmic character, we propose to introduce a coefficient that can measure the growth rate:

$$C_\mu = a + h_Q \log_2 t$$

We would like to emphasize that the coefficient h_Q can be used as a robust and universal characteristic of Statistical Complexity of molecular trajectories since it appears not to depend on the particular numerical model, the details of computational procedure, the size of the molecular ensemble and the type of the test atom (hydrogen or oxygen).

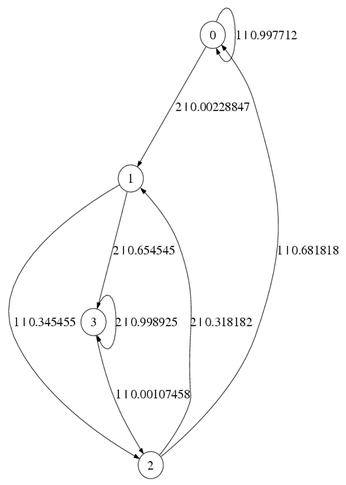
Non-Markovian Behaviour

The system under investigation is the four residue peptide VPAL (Valine – Proline – Alanine – Leucine). The system was simulated using Molecular Dynamics for 500ns. The Ramachandran plot for the Proline residue is shown. By clustering this into two states a symbol sequence is obtained. This is analysed using Computational Mechanics and the resulting ε-machine is also shown.



Had the dynamics been Markovian in respect of the conformational states, the ε-machine would consist of only two causal states. These would have been the two conformational states of the Proline residue. The ε-machine, however, consists of four causal states, leading to the conclusion that the dynamics are non-Markovian.

In the case under review we believe that the four states occur because we are investigating the system on a very small time scale, i.e. the time between recording conformations in the Molecular Dynamics simulation is short. As a result, each conformational state is divided into two parts. From one part it is possible to access the other conformational state and from the other part it is not possible. This therefore results in a total of four causal states, as shown in the ε-machine.



Conclusions

- It is shown that arbitrary long memories (much longer than one can expect from a statistical or correlation analysis) are present in the recorded time series, manifesting themselves as stationary patterns of causal states in the velocity-defined phase space.
- The exponent h_Q represents a universal physical constant characterizing water, since it does not depend on the specific macroscopic observable analysed, parameters of the system or simulation model, such as temperature, number of molecules or numerical model employed
- The conformational transitions for the peptide VPAL show non-Markovian behaviour on a small time scale.

References

1. J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Physica D 58, 77 (1992)
2. J. P. Crutchfield and K. Young, Phys. Rev. Lett. 63, 105 (1989)
3. J. P. Crutchfield and K. Young, in *Entropy, Complexity, and Physics of Information, SFI Studies in the Sciences of Complexity, VIII*, edited by W. Zurek (Addison-Wesley, Reading, Massachusetts, 1990)
4. J. P. Crutchfield, Physica D 75, 11 (1994)

Acknowledgments

The work is supported by Unilever and the European Commission (EC Contract Number 012835 - EMBIO).