Final Report of the Workshop

Complex Free Energy Landscapes:
Computational and Statistical Methods for Soft Matter

EPSD - Science Meeting 2811

The following final report intends to review the main aspects of the workshop "Complex Energy Landscapes" held in Zaragoza the days 2,3 and 4 of June of 2010. The report is divided in the following two parts:

(A) Scientific Report
(B) Financial Report

(A) Scientific report

Summary

The workshop “Complex Energy Landscapes: Computational and Statistical methods of soft matter” took place in Zaragoza from June 2 to June 4 of 2010. It was hosted by both the Institute for Biocomputation and Physics of Complex Systems” and the Zaragoza node of CECAM (ZCAM).
The information of the workshop was published in:
http://bifi.es/events/cel2010/

The organizing committee was composed by:

- Pierpaolo Bruscolini (U. Zaragoza)
- Fernando Falo (U. Zaragoza)
- Jesus Gómez-Gardeñes (U. Rey Juan Carlos and U. Zaragoza)
- Diego Prada-Gracia (U. Zaragoza)

Besides, the conference secretary was Beatriz Antolí (from the BIFI and ZCAM) who was in charge of the management of the logistic of the conference and the communication via the workshop e-mail complexlandscapes@gmail.com.

The workshop participants were divided in two categories keynote and invited speakers. The talks of keynote speakers had a duration of one hour while those of the invited participants were about 35 minutes. The full list of participants composed by:

Keynote speakers:

- R. Best (University of Cambridge)
- P. Bolhuis (Univ. Amsterdam)
- L. Bongini (Univ. Barcelona)
- P. De Los Rios (EPFL)
- G. Franzese (Univ. Barcelona)
- S. Krivov (Univ. Leeds)
- Y. Moreno (Univ. Zaragoza)
- F. Noe (Freie Univ. Berlin)
- M. Palassini (University of Barcelona)
- F. Rao (Univ. Strasbourg)
- F. Ritort (Univ. Barcelona)
- D. Wales (Univ. Cambridge)

Invited speakers:

- F. Cao (Univ. Complutense de Madrid)
- A. N. Naganathan (Institute for Research in Biomedicine of Barcelona)
- M. Seeber (Univ. Zurich)
- B. Seoane (Univ. Complutense de Madrid)

In addition, the workshop was open to the interested audience (about 10 people).

Description of the scientific content and discussion at the event

Introduction and motivation

The thermodynamics and kinetics of soft matter are usually described in terms of the topological properties of Potential Energy Surfaces (PES) and Free-Energy Landscapes (FEL). PES and FEL theories have been a topic of interest since 40 years spanning fields as diverse as supercooled liquids, glass-forming systems, biomolecules, molecular clusters, etc. A correct coarse-grained description of these landscapes requires the identification of the set of minima, maxima and saddle points of the landscape, as well as the kinetic connections between them. This overall description is seen necessary to characterize the phenomena of interest. However, landscapes of multibody systems are multidimensional and often very complex surfaces and therefore the main concern is to bridge the gap between the computational data (i.e. the microscopic information) and the experimental results (i.e. the macroscopic characterization) by means of a mesoscopic description of the landscapes.

One of the most active fronts of research in this direction is to fully describe the landscape of complex polymers, such as proteins. Knowledge of the landscape topology is the essential key to understand protein structure and dynamics. The determination of the conformers of a protein and their basins of attraction takes a central role for studying molecular isomerization reactions. In particular, it is of utmost importance to answer the following questions: What are the structural conformations possible? Is there any relevant hierarchy among these conformers? What are the transition paths between them? With this knowledge, several problems about biomolecular reactions, such as enzymatic activity, protein folding, protein deposition diseases, peptide immunogenicity, etc, can be tackled.

State of the art

Up to date different approaches have been used to determine the topography of energy surfaces:

(i) Conformational networks have proved to be a powerful tool by analyzing the conformational space of polymers making use of network representations [1]–[4], and trying to decompose the network in modules corresponding to the free energy basins.

(ii) Markovian state models let us treat the information of one or several trajectories of molecular dynamics constructing a transition matrix offering global observables as relaxation times and modes [6]-[7];

(iii) Finally, Disconnectivity graphs [8]–[10] study the PES by assuming the provisos of classical transition-state, i.e. the knowledge of the minima and saddles of the PES is used to infer the behavior of the system at nonzero temperature.

The thermodynamics and kinetics of soft matter are usually described in terms of the topological properties of Potential Energy Surfaces (PES) and Free-Energy Landscapes (FEL). PES and FEL theories have been a topic of interest since 40 years spanning fields as diverse as supercooled liquids, glass-forming systems, biomolecules, molecular clusters, etc. A correct coarse-grained description of these landscapes requires the identification of the set of minima, maxima and saddle points of the landscape, as well as the kinetic connections between them. This overall description is seen necessary to characterize the phenomena of interest. However, landscapes of multibody systems are multidimensional and often very complex surfaces and therefore the main concern is to bridge the gap between the computational data (i.e. the microscopic information) and the experimental results (i.e. the macroscopic characterization) by means of a mesoscopic description of the landscapes.

One of the most active fronts of research in this direction is to fully describe the landscape of complex polymers, such as proteins. Knowledge of the landscape topology is the essential key to understand protein structure and dynamics. The determination of the conformers of a protein and their basins of attraction takes a central role for studying molecular isomerization reactions. In particular, it is of utmost importance to answer the following questions: What are the structural conformations possible? Is there any relevant hierarchy among these conformers? What are the transition paths between them? With this knowledge, several problems about biomolecular reactions, such as enzymatic activity, protein folding, protein deposition diseases, peptide immunogenicity, etc, can be tackled.

Up to date different approaches have been used to determine the topography of energy surfaces: (i) Conformational networks have proved to be a powerful tool by analyzing the conformational space of polymers making use of network representations [1]–[4], and trying to decompose the network in modules corresponding to the free energy basins; (ii) Markovian state models let us treat the information of one or several trajectories of molecular dynamics (MD) constructing a transition matrix offering global observables as relaxation times and modes [6]-[7]; (iii) Finally, Disconnectivity graphs [8]–[10] study the PES by assuming the provisos of classical transition-state, i.e. the knowledge of the minima and saddles of the PES is used to infer the behavior of the system at nonzero temperature.

The above approaches, although providing with an acceptable coarse graining of the landscapes, still involves extremely large computational costs and they only (successfully) apply to systems with moderate dimensionality. Therefore, there is a great need to find tools that overcome the computational limitations, allowing us to apply the theoretical concepts behind each of the methods.

Motivation and Goals

The main goal of the workshop is to join the knowledge and tools of the diverse researchers working in the development of algorithms aimed at describing the

topology of the FEL. Given the divergence of the techniques used to analyze FELs and the broad range of the fields where this problem is being tackled, a number of numerical algorithms have been developed to overcome the initial computational limitations. However, these numerical techniques are specific of particular dynamics of complex systems rather than being of general use for the broad community working in FEL theory. It is therefore timely and necessary to share the point of view and the latest achievements of each approach in order to converge to more general computational methods.

References:

[1] Noé F, Horenko I, Schütte C, Smith JC (2007) Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. J Chem Phys 126: 155102.

[2] Rao F, Caflisch A (2004) The protein folding network. J Mol Biol 342: 299–306.

[3] Caflisch A (2006) Network and graph analyses of folding free energy surfaces. Curr Opin Struct Biol 16: 71–78.

[4] Noé F, Fischer S (2008) Transition networks for modeling the kinetics of conformational change in macromolecules. Curr Opin Struct Biol 18: 154–162.

[6] Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. J Chem Phys 126: 155101.

[7] Buchete NV, Hummer G (2008) Coarse master equations for peptide folding dynamics. J Phys Chem B 112: 6057–6069.

[8] Wales DJ (2006) Energy landscapes: calculating pathways and rates. Int Rev Phys Chem 25: 237–282.

[9] Evans DA, Wales DJ (2003) The free energy landscape and dynamics of met-enkephalin. J Chem Phys 119: 9947–9955.

[10] Wales DJ, Doye JPK, Miller MA, Mortenson PN, Walsh TR (2000) Energy landscapes: from clusters to biomolecules. Adv Chem Phys 115: 1–111.

Assesment of the results and impact of the event on the future direction of the field

The main conclusion of the workshop is the strongly convergence in the last years of the three different approaches used to analyze FELs. This conclusion, specific scientific questions aside, is probably the best achievement at the end of the workshop. Giving the chance to share knowledge and experiences among researchers from these, apparently, different sub-communities has been timely and necessary (and this was the main conviction to decide to organize this workshop).

In this sense, the workshop organized with the ESF fundings has left its own hallmark in the community. The invited speakers concluded at the end of the workshop that it is necessary to consider this event as the first one of a biannual sequence of scientific meetings on this topic.

Several different techniques of analysis of soft matter FEL's have been developed in the recent years. These techniques, are currently tested in real systems by means of the confrontation of the results from the study of Molecular Dynamics Simulation at a high degree of fidelity (with all-atom forcefields) and experiments. At this point, the three different approaches represented by the invited speakers find some particular limitations that can be overcome with the help of the other approaches. In this sense:

-From the PES: The description of a system given by the PES yields an extremely large amount of different metastable states with different degrees of stability. This information is usually represented by means of dendograms giving a global picture of the landscape, although the data-base of minima and transition-states (disconnectivity-graph) must be lumped in a coarse-grained preserving the physical magnitudes. The Conformational networks analysis and Markov Models have developed strategies to lump conformational states, with its transition rates, keeping the slowest, most relevant, global relaxation processes (Talks by P. de los Rios, L. Bongini and F. Noe) so that new data-bases of macro-states obtained from the PES can be built up to make the analysis of the global behavior of the system.

-From the Conformational Kinetic Networks: The metastable conformations unveiled by this approach comes straight-forward from a MD trajectory encoded into a graph. The accuracy of the description obtained from the topology of the network depends on the discretization of the conformational space and time. The work with Markov state models offers answers to questions as: What is the precision and the accuracy predicting long-time kinetics? How should a network be generated in order to achieve maximum precision? How can we compute transition pathways from Markov models? (Talk by F. Noe). While the analysis from the PES provides with a

useful approach to test the validity of the conformational states found, as well as shows the best description including the transition states (Talk by D. Wales).

-From Markov Models: This approach can successfully distinguish the different time scales of the dynamical processes which are present in the description of the system. It has been showed, mainly by F. Noe, how can metastable-states can be extracted to build kinetic models. On the other hand, the approach is computationally expensive and not very accurate defining the transition states. In this way, the models can take advantage in some points of the information obtained from the PES to refine and check the tools developed.

Since these sub-communities have been walking in the same way and with the same goals, some common conclusions, which will drive the projects in the near future, have been found:

- The description of soft-matter systems with these tools offers a natural way to analyze the landscape avoiding the projection over a few reaction coordinates. This fact constitutes a decisive point to address experimental problems with new perspectives avoiding relevant misleading interpretations. For instance, a growing number of experiments and models show that, when projected onto a reaction coordinate, protein dynamics is subdiffusive, but this is just a wrong interpretation. (Talks by S. Krivov and R. Best). Thus, it is necessary to dessign optimum reaction coordinates which exhibit diffusive dynamics to describe the kinetic behavior of proteins . (Talk by S. Krivov)

- It is necessary to keep on working on complex network analysis. Systems with higher complexity and dimensionality will be tackled in the future, and handling the information with networks will make the analysis feasible. Together with the advance in this issue, new techniques of visualization of the coarse-grained landscape must be developed. The high number of relevant states to characterize the landscape (and low connectivity) makes the analysis with master equations and kinetic models hard if we want to extract global features. (Talk by F. Rao)

- The battery of tools, have been giving new insight on the physics of bio-molecules (most on the protein folding problem). It is time to jump towards the study of the interaction of proteins or to revisit complex systems as fluids and water dynamics (Talk by F. Rao and G. Francese)

- Since It is necessary to include in most of this models the transition-states, it is essential to keep on developing physical strategies to characterize them as for instance transition sampling methods. With special relevance if the analysis aims to describe rare events (or need to include them). (Talk by P. Bolhuis)

- It is necessary to study how the conformational landscape can be reduced preserving the relevant physical features. We have understood

how the problem of the protein folding cannot be fully described in terms of two states (folded-unfolded) but: How many reaction coordinates, or degrees of freedom, are needed to study the complexity of the system? The analysis made with Markov transition models (talk by F. Noe), which is able to decompose in low and fast relaxation processes can provide useful results that will need to be checked with the other approaches.

- Since most of the experiments physics can perform are out of equilibrium, it is necessary to learn how to build these models from non-equilibrium trajectories. In particular, we have to focus in the near future on understanding how this tools (the characterization of the landscape in equilibrium) can help to understand a non-equilibrium process as the stretching of proteins and DNA. (Talk by F. Ritort)

In conclusion, this field has been recently developed from three different approaches, according to their origins, which are converging nowadays. The same problems are being tackled to answer similar questions and thereby finding similar conceptual barriers. The cross-fertilization found in this meeting will help to keep on developing the analysis of the FEL in an effective way and overcome all the conceptual problems. However, this field has to make a bigger effort to offer its potential capability to experimental analysis, as it has been shown by solving the problem of the apparently sub-diffusive dynamics of proteins.

Final programme of the meeting

The detailed program of the workshop was the following:

Day 1: June 2nd, 2010

09:50-10:00 Registration

10:00-10:15 Wellcome

10:15-11:15 **D. J. Wales**: "Energy Landscapes for Soft Matter"

11:15-11:35 Coffe Break.

11:35-12:35 **M. Palassini**: "Complex free energy landscapes in disordered systems. Examples from condensed matter physics and computer science"

12:35-13:35 **G. Franzese**: "Free Energy Landscape of Hydration Water: Theory and Experiments"

13:35-15:30 Lunch

15:30-16:30 **P. de los Rios**: "Coarse-graining of configuration-space networks"

16:30-17:30 **L. Bongini**: "A graph theoretical analysis of the energy landscape of model proteins"

17:30-17:45 Break

17:45-18:20 **B. Seoane**: "Separation and fractionation of order and disorder in highly polydisperse systems."

Day 2: June 3rd, 2010

09:30-10:30 **S. Krivov**: "Is protein folding sub-diffusive?"

10:30-11:30 **F. Noe**: "Markov models of molecular kinetics: Generation, Validity and Analysis"

11:30-11:50 Coffe Break.

11:50-12:50 **P. Bolhuis**: "Extraction of reaction coordinates from the reweighted path ensemble. "

12:50-13:25 **F. J. Cao**: "How occasional backstepping can speed up a processive motor protein."

13:30-15:30 Lunch

15:30-16:30 **R. Best**: "Mapping protein folding dynamics to low-dimensional coordinates"

16:30-17:30 **F.Rao**: "Complex network analysis of protein free-energy landscapes"

17:30-17:45 Break.

17:45-18:20 **A. N. Naganathan**: "The Curious Case of Protein Folding Transition States"

Day 3: June 4th, 2010

09:30-10:30 **Y. Moreno**: "Use and Misuse of Networks in Biology"

10:30-11:05 **M. Seeber**: "Exploring mutational effects on the energy landscape of a beta-hairpin"

11:05-11:25 Coffe Break.

11:25-12:25 **F.Ritort**: "Molecular misfolding investigated by mechanically unzipping nucleic acids"

12:25-12:40 **M.Mareschal** (Z-CAM Director): Concluding Remarks

12:40-13:30 Round Table

(B) Financial report

Detailed list of expenses

These are the expenses covered with the help of the ESF grant:

- Hotel accommodation of invited and plenary speakers: 2768,89 Euros

- Trip expenses: 3741,34 Euros:

  R. Best: 372,95 Euros
  P. Bolhuis 643,53 Euros
  L. Bongini: 134,50 Euros
  G. Franzese: 415,36 Euros
  S. Krivov: 420,07 Euros
  F. Noe: 909,19 Euros
  M. Palasini: 79,35 Euros
  F. Rao: 453,22 Euros
  F. Ritort: 169,65 Euros
  D. Wales: 162,62 Euros
  A. Natarahan: 101,9 Euros

- Meals and Dinners: 1483,29 Euros

TOTAL AUMOUNT: 7993,52 Euros

We have not included other expenses that will be covered by local sponsors (BIFI and ZCAM)