



Project no. 12835

EMBIO

## **Emergent organisation in complex biomolecular systems**

Instrument type FP6-2003-NEST-Path

Activity name NEST-2003-1

### **Final activity report**

Period covered: from 1 Mar 2005 to 31 Aug 2008

Date of preparation: 1 Oct 2008

Start date of project: 1 Mar 2005

Duration: 42 months

Project coordinator name

Prof. Robert C. Glen

Project coordinator organisation name

Cambridge University

## 1. Project execution

Complexity and self-organisation are critical yet poorly understood phenomena. This project aimed to develop and apply mathematical and computational approaches that identify principles governing the emergent organisation of self-organising biomolecular systems. Computational methods for characterising the dynamics of these intrinsically complex processes are developed and applied to protein folding and molecular self-assembly. The methodologies focus on the complexity of the system's dynamics thus advancing fundamental knowledge concerning the role of complexity in biological systems.

A composite approach has been pursued, focusing on critical aspects of the problems defining the route from sequence (chemical formula) to complex functionality of native structures. This involves determination of the underlying potential energy and fitness surfaces together with their topological, statistical and dynamic properties for models of polypeptides and RNA self-organization; finding the features of the energy funnel for simplified protein in water models in order to distinguish between and understand “good” and “bad” folders; characterising biomolecular dynamics in terms of information flows between different time- and length-scales (for which the aggregation of spatial information can be used to detect emergent structures); reconstructing dynamic hierarchies in model biopolymer systems thus directly detecting the emergence of the dynamic forms and information flow on different spatiotemporal scales in the system; and calculating the statistical complexity of model biomolecular systems and the key parts of realistic biomolecules. The data for these investigations are obtained from sophisticated all-atom simulations of realistic biomolecular systems and experimental mechanical stretching of giant single molecular proteins.

The improved understanding that is obtained of the dynamical complexity of native and folding biomolecular systems provides a basis for the accurate calculation of biomolecular folding and function and is therefore of fundamental relevance to biology, medicine and biotechnology.

### Main ideas

Bio-molecular systems possess all the properties of complex systems described above. They consist of a large number of small parts that have relatively simple interactions. They exhibit sophisticated non-linear dynamics and form “native structures” that are the manifestations of persisting dynamic forms. Bio-molecular self-organisation also lends itself naturally to the use of a composite approach for the study of complexity by focusing on different aspects of the problem on the route from sequence (chemical formula) to complex functionality (native structures):

**sequence → potential surface → dynamic forms → structure formation → complexity**

We aimed to address complexity of molecular self-organisation using a range of state of the art approaches and techniques for quantifying dynamic complexity. As part of this endeavour we generated data on the dynamics of self organising molecular systems, performed complexity analysis using a diverse set of approaches and applied the resulting conclusions to practical problems from bio-chemo- and medical science. We also note that although the focus of the project is on understanding complexity in the context of bio-molecular self-organization, methods and approaches are general and applicable to a wide range of systems which give rise to complex behaviour. Our specific focus on bio-molecular self-organisation has been chosen not only because it is a critical problem in biology but also because there already exist a large body of experimental and simulation data on which method development can be based.

### The objectives

The goal of the project was to **quantify the complexity associated with self-organization in bio-molecular systems as a means to understand complex phenomena in systems that exhibit spontaneous emergence**. Specific objectives were:

- to monitor the process of emergent *complexity* by performing all-atom simulations of peptide/protein folding and lipid membrane self-assembly in explicit water;
- to obtain detailed, all-atom data on representative regions of the free energy landscape (*folding funnel*) by simulating biopolymers (proteins and/or RNA) in their denatured and native forms;
- to investigate the *sequence space* of RNA molecular models and its effect on the *kinetics of folding*;
- to study topological, statistical and dynamic properties of *generic* potential energy and fitness *surfaces* in models of polypeptides and in RNA during self-organization;
- to find and characterize the features of the free energy *funnel* for simplified protein-in-water *dynamic models* in order to distinguish between “good” and “bad” folders;

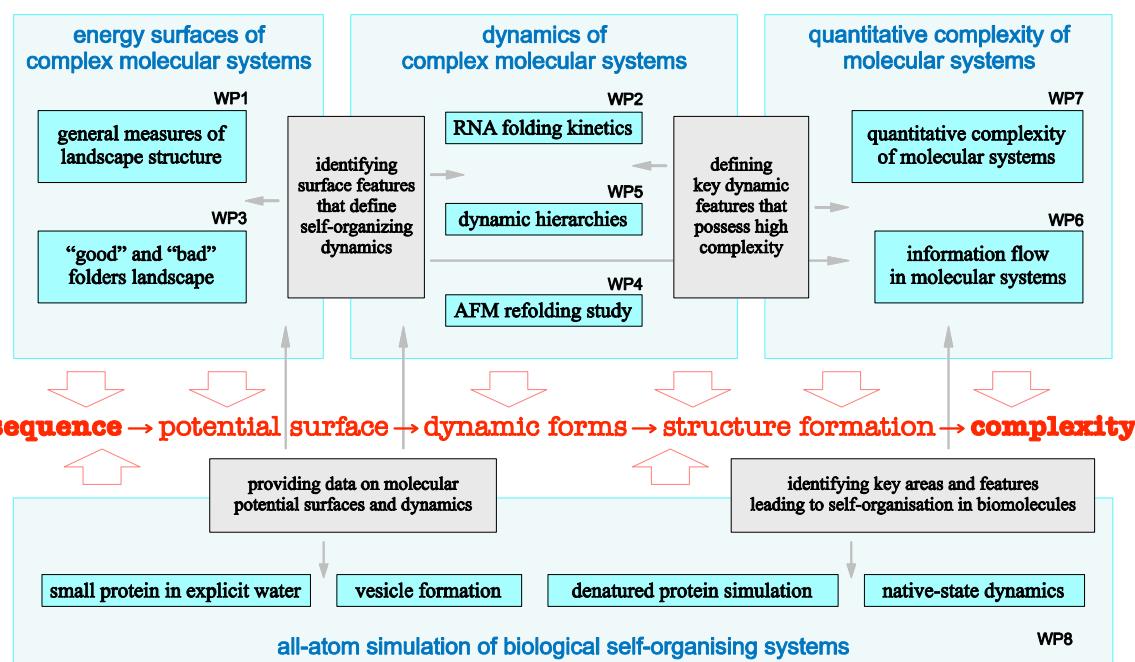
- to define and describe parameters of the folding-unfolding *pathways* on the free energy *funnel* through the experimental mechanical stretching of giant single molecular proteins;
- to reconstruct *dynamic hierarchies* in a model bio-polymer system thus directly detecting the *emergence* of the *dynamic forms* and *information flow* at different scales in the system;
- to calculate the *dynamic complexity* of the system's trajectories in different regions of the energy funnel as well as the folding process as a whole.

Due to required parallel work in achieving these objectives, their fundamental character, and tight interconnection (see the workplan) most of the deliverables were produced by the end of the project. During the course of the project the work was reviewed and the progress assessed at the planned annual workshops in Vienna, Leipzig, and Venice. One of the objectives of the project was to collect the approaches and algorithms for quantitative estimation of complexity of a general multidimensional dynamic system. These can be applied to a large variety of complex systems in any branch of natural and social sciences.

The common thread in the workplan was the route from biopolymer sequence to complex functionality. Accordingly the workpackages were aligned to this route and contributed to various parts of it. Conceptually the work was divided into four main areas:

- 1) energy surfaces of complex molecular systems;
- 2) dynamics of complex molecular systems;
- 3) dynamic complexity of molecular systems;
- 4) all-atom simulation of biological self-organising systems.

The first three areas were linked in a chain-like manner so that the whole structure-complexity route was encompassed and interlinked. The fourth area was distributed among, and connected to, each of the first three areas:



## The Consortium

The consortium consisted of eight Universities from six countries and was coordinated by Cambridge University:

- University of Cambridge (UK) - Coordinator
- University of Groningen (The Netherlands)
- University of Florence (Italy)
- Chalmers University of Technology (Sweden)
- University of Vienna (Austria)
- University of Heidelberg (Germany)
- University of Leipzig (Germany)
- University of Freiburg (Germany)

Prof. Robert Glen was responsible for scientific and administrative management of the project:

Unilever Centre for Molecular Informatics, Chemistry Department, University of Cambridge,  
Lensfield Road, Cambridge, CB2 1EW, UK



## Summary of the work results

The project has successfully accomplished the planned activities in both scientific and

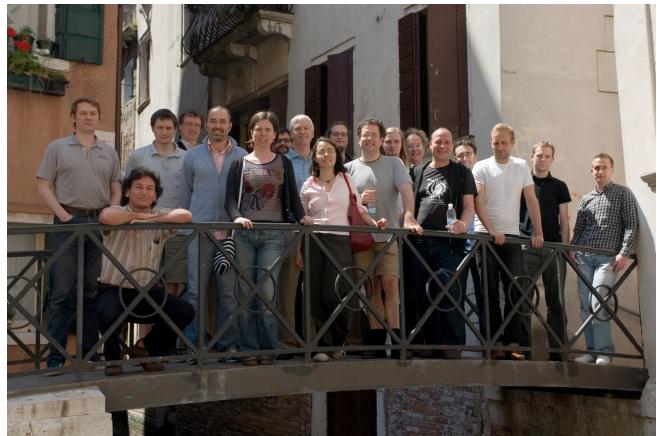
organisational aspects. Four meetings were held:

1. A kick-off meeting was held in Cambridge in July 2005 where from 2 to 5 participants from each member of the Consortium were present. The most recent results were discussed and specific plans for collaborative work, visits between the partners, and the Consortium meetings were made.
2. Another meeting summarising the work during the first year of the project was held in Vienna in May 2006. Nineteen presentations from all the participants made an excellent

foundation for the discussions that resulted in new links and collaborative lines of research. In addition two invited lectures, one outside the Consortium, Dr Cosma Shalizi, and one from the Consortium, Prof. Roberto Livi, presented two lectures that both educated the participants in common areas of non-linear dynamics and complexity theory and outlined the forefront of the research in these areas.

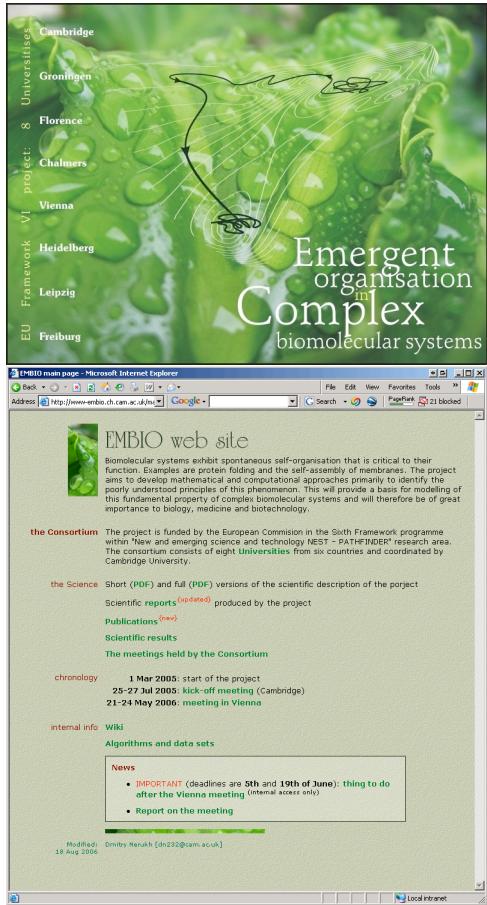
3. The meeting in Leipzig in May 2007 was devoted to reporting the progress of the project, reporting the scientific results and planning the work for the rest of the project. Two invited speakers gave lectures on modern aspects of mathematics of complexity: Dr. Roumen Dimitrov and Dr Nihat Ay.

4. The last meeting of the project took place in Venice at the European Center for Living Technology (ECLT). The meeting has summarised the project's results and discussed the organisational issues of reporting to the EC.



The detailed information on the meeting including all the presentations and lectures are available at the project's website:

[www-embio.ch.cam.ac.uk](http://www-embio.ch.cam.ac.uk)



The project's website was the main communication point for the members of the Consortium. Since the information was organised as freely accessible or password protected the website also served as a means of communicating the results of the work to the general public. The contents of the site included the sections about the members of the Consortium, the reports from the members, the publications, the summaries of the meetings together with all the presentations given, current news and announcements, a Wiki for quick exchange of information between the group members, etc.

All forty three deliverables were achieved. They can be found at [http://www-embio.ch.cam.ac.uk/reports/index.html - deliverables](http://www-embio.ch.cam.ac.uk/reports/index.html)

77 publications have been published or accepted from the Consortium. The references together with the links to the texts are available at <http://www-embio.ch.cam.ac.uk/publications/index.html>. The members of the Consortium have presented 64 lectures and posters at various Conferences where they have made the scientific results produced within EMBIO available to

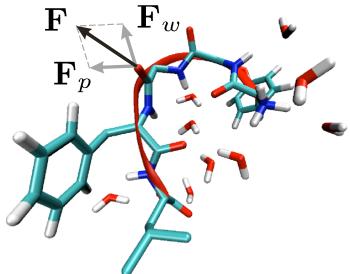
the public. There were over 30 reciprocal visits between the groups, mostly bidirectional, but also including three or more participants.

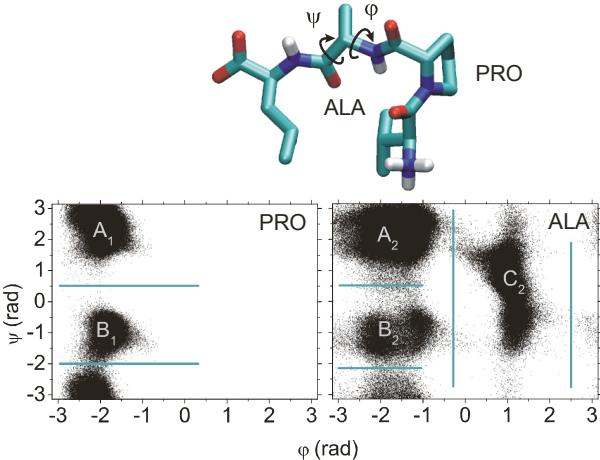
## Scientific results

### Statistical Complexity of molecular dynamics

Very interesting and unexpected results for very long time phenomena in molecular liquids have been obtained in Cambridge using the Statistical Complexity measure. Computer simulated trajectories of bulk water molecules form complex spatio-temporal structures at the picosecond time scale. This intrinsic complexity, which underlies the formation of molecular structures at longer time scales, has been quantified using a measure of Statistical Complexity. The method estimates the information contained in the molecular trajectory by detecting and quantifying temporal patterns present in the simulated data (velocity time series). Two types of temporal patterns are found. The first, defined by the short-time correlations corresponding to the velocity autocorrelation decay times ( $<0.1$  ps), remains asymptotically stable for time intervals longer than several tens of nanoseconds. The second is caused by previously unknown longer-time correlations (found at longer than the nanoseconds time scales) leading to a value of Statistical Complexity that slowly increases with time. A direct measure based on the notion of Statistical Complexity that describes how the trajectory explores the phase space and independent from the particular molecular signal used as the observed time series is introduced.

Also, the dynamics of peptides and proteins generated by classical MD is described using a Markov model. The model is built by clustering the trajectory into conformational states and estimating transition probabilities between the states. Assuming that it is possible to influence





the dynamics of the system by varying simulation parameters, we show how to use the Markov model to determine the parameter values that preserve the folded state of the protein and at the same time reduce the folding time in the simulation. We investigate this by applying the method to two systems. The first system is an imaginary peptide described by given transition probabilities with a total folding time of  $1\mu\text{s}$ . We find that only small changes in the transition probabilities are needed to accelerate (or decelerate) the folding. This implies that

folding times for slowly folding peptides and proteins calculated using MD cannot be meaningfully compared to experimental results. The second system is a four residue peptide VPAL (Valine - Proline - Alanine - Leucine) in water. We control the dynamics of the transitions by varying the temperature and the atom masses. The simulation results show that it is possible to find the combinations of parameter values that accelerate the dynamics and at the same time preserve the native state of the peptide. A method for accelerating larger systems without performing simulations for the whole folding process is outlined.

### Coarse graining meso-scale simulations

Modeling self-assembly in chemical systems is only possible if simplifying assumptions that reduce the overall complexity of the dynamics are possible. Molecular dynamics is in itself a Newtonian approximation of a quantum system, but this simplification is often not coarse grained enough to deal with the length and time scales that are most interesting for self-assembling materials. In recent years a number of effective particle models on the meso-scale have been developed. One of the most popular is the Dissipative Particle Dynamics (DPD) method and the United Atoms (UA) methods. Both DPD and UA have been established as standard methods for mesoscopic simulation. Despite its popularity, the question of how accurately DPD represents the underlying molecular system has still not been fully resolved. In standard DPD the functional form of the interactions between the mesoscopic particles are chosen more or less heuristically (subject to the fluctuation-dissipation constraint), with the interaction strengths calibrated to fit macroscopic observables, usually the compressibility and the diffusion rate. Several authors, e.g. [1,2], have established bottom-up connections between the micro- and mesoscale models. While of great theoretical and conceptual value, no practical method has yet been presented for calculating the functional form of the effective interactions in the coarse-grained models.

To obtain a well-defined bottom-up scheme, the dynamics of the coarse-grained DPD particles must be defined through a projection of the microscopic trajectories. The problem is to find a self-contained representation of the system at the coarse-grained level, i.e. to determine all interactions in the DPD model. Recently, we have presented a method for doing this. The method uses data from simulation of the microscopic system. All necessary calculations are easily carried out in real applications. Our results are important for two reasons: They can provide a solid base for evaluating the applicability of DPD coarse graining schemes; and they specify how to adapt the DPD method to model different systems. Our results show that the functional form of the stochastic interaction in DPD differs significantly from the standard heuristic function used.

The new method for determining the effective stochastic interactions in coarse-grained molecular dynamics uses Mori-Zwanzig projection operators. The reduced dynamics can reproduce e.g. the transport properties correctly, which previous methods typically have failed to do. This is very

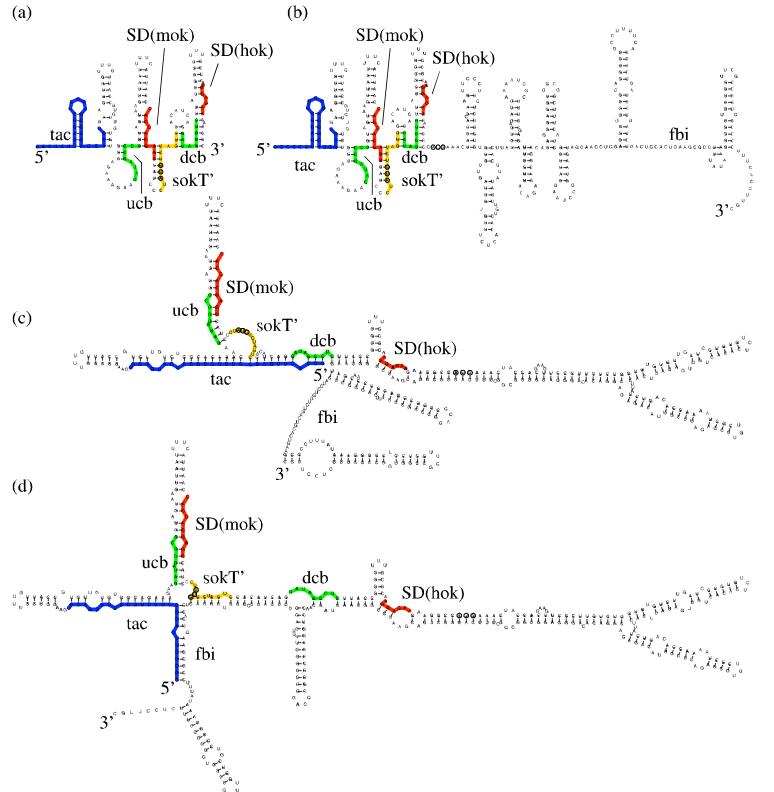
important since transport is very important in biomolecular systems. This work aims to strengthen the objective regarding physically realistic simulations of self-assembly processes. The implication of the work is not limited to constructing coarse-grained molecular simulations with the correct transport coefficients. The direct construction of the dissipative interaction shows that the standard thermostats, e.g. Berendsen or the DPD thermostat, used when modeling united atoms should be changed significantly. One of the main conclusions is that the thermostat should be considered as an integral part of the dynamics, rather than an add-on primarily used for equilibration. This perspective can possibly have much wider applications than the systems we are currently investigating.

We have used the method for determining the stochastic interactions in DPD representation of coarse-grained MD [3,4]. We use detailed molecular dynamics simulations of small subsystems of SPC water. An explicit projection operator is then applied. Depending on the choice of projection, we get either a UA or a DPD dynamics on the coarse-grained level. It is well known in the literature that the pairwise conservative interaction in UA can be determined directly from the radial distribution function [5]. What is new is that we also determine the stochastic interaction by using Mori-Zwanzig theory, and a generalization of the expressions for the structure of the noise found [6] (also closely related to the famous Green-Kubo relations).

### Hierarchical dynamics

Reduction of complexity is of central importance for the ability to model large scale dynamical systems, such as large bio-molecules for example. Inferring the hierarchical structure of a dynamical process also provides increased theoretical understanding of the underlying process, i.e. insights into mechanisms behind emergence in systems that are, at least in principle, reductionistic. Emergence is one of the central themes in complex systems science. Methods and ideas used for analyzing hierarchical dynamics are likely to play a central role in any theoretical framework addressing emergence. Finally, many of the techniques that we discuss in the context of hierarchical dynamics originated in the related fields of model reduction and multi-scale simulation. In model reduction one typically seeks systematic methods for reducing the complexity of a specific model, or a class of models, thereby making them more manageable in terms of simulation or analytic analysis. As we will see, this is a special case of hierarchical dynamics (a case of great practical and conceptual importance).

There are mainly two types of hierarchies in dynamical systems: structural and temporal hierarchies. Temporal hierarchies are defined by separation of time scales between the different levels. The local nature of physical interactions connects time scales to length scales. As a consequence, the levels in temporal hierarchies are also often associated with separation in length scales. One may turn the argument around and claim that our choice of metric is a reflection of how the interactions between objects in the universe behave. In any case the result is the same, there is a tight coupling



between time and length scales. Structural hierarchies are derived from geometric properties of the dynamics. These geometric properties stems from decomposability, or skew-product decomposability (to be defined later), of the vector field defining the flow. Simple examples of structural hierarchies are non-interacting subsystems or systems with constants of motion.

Hierarchy in a dynamical system is a more general concept than dimensional reduction. To define a hierarchy in a more general setting, we must focus on projections of the dynamics that constitutes a new "self-contained" (Markovian) dynamical system. The levels in the hierarchy do not necessarily evolve on different time scales. The general idea of autonomy naturally leads to the concept of preserved fibrations and foliations. This ideas has been explored in detail in two publications [7,8].

Within EMBIO the work on reduction has also been focused on Markov chains, which is often used to model transitions between conformation states in a folding process. Our first approach to this problem was to use information theory to search for reductions that respect the Markov property [9]. Recently we have developed a much more efficient method based on aggregation of variables, or lumping, to solve this problem [10,11]. The method includes several traditional techniques such as spectral graph partitioning [12] and spectral identification of meta-stable states [13], as special cases.

### **Information-theoretic approaches**

#### *Information dynamics*

Two information-theoretic frameworks for spatio-temporal dynamics have been developed in the EMBIO project, reported in deliverable D32, and here we summarise the results and the conclusions. The first framework deals with the analysis of pattern formation in chemical reaction-diffusion dynamics. Here, the ambition has been to connect the thermodynamic constraints of the chemical processes with an information characterisation of the patterns given by concentration profiles across the system. We have succeeded in establishing information flow quantities that via a continuity equation shows how an inflow of information capacity (from the free energy of the fuel) aggregate into and maintain information in patterns [14, 15]. The inflow balances in this way a destruction of information due to entropy production from diffusion and chemical reactions.

The second framework deals with information in microscopic patterns. If the microscopic dynamics is reversible, the information in total is conserved, and we have established a continuity equation for information also on this level [16, 17]. This means that, if a chemical system is modelled as a microscopically reversible system, there is no loss of information. The entropy production observed can in this situation be understood as information being increasingly more difficult to detect as correlation information in the microscopic configurations involve an increasing number of particles and stretches over increasing distances. In principle, the entropy production that is an important constraint in the first framework could be captured in detail in the second framework. In that case it would not be seen as an information loss, but as a distribution of correlation information over the large number of microscopic degrees of freedom. It turns out that, even if the two frameworks are useful as separate tools for investigating complex dynamics, the connection between the two is difficult to exploit. It is demonstrated that the information quantities are directly related when they are aggregated to the whole system, but the local quantities, on the macro and on the micro level, have different interpretations. The reason is that an information characterisation of spatial structure critically depends on *how* one is searching for that structure. In the two frameworks we have used different approaches, based on what is relevant on the different scales. Whether there is a reasonable approach that would be simultaneously applicable on both the micro and the macro scales is still an open question.

#### *Information theory, computational mechanics, and $\varepsilon$ -machines*

We have investigated how computational mechanics (see, e.g., refs [18, 19]) can be used to characterize dynamical systems and what practical considerations this may require. Although computational mechanics applies to discrete systems, it is possible, as we will see, to utilize the framework on continuous systems (both with respect to time and space), subject to carefully discretizing these. We have investigated how various possible approaches for dividing state-space when making a discrete dynamics from a continuous system may affect the computational mechanics quantities characterising the discrete dynamics. Some connections between computational mechanics and information theory has been exploited in this analysis.

The main conclusion of our study of how the choice of partition influences the resulting  $\varepsilon$ -machines is that if the partition is accurate, then it is relatively straight-forward to find an  $\varepsilon$ -machine. However, using the reconstruction method to search for a partition seems fruitless; since the  $\varepsilon$ -machines are so sensitive to violations of the generativeness of the partition, which makes measures based on the resulting computational complexity unsuitable for this purpose. The problem is severely compounded by the huge amount of data needed in order to achieve reliable statistics, and that in order to represent  $\varepsilon$ -machines corresponding to partitions that are not Markov, it may be necessary to capture dependencies far back in time.

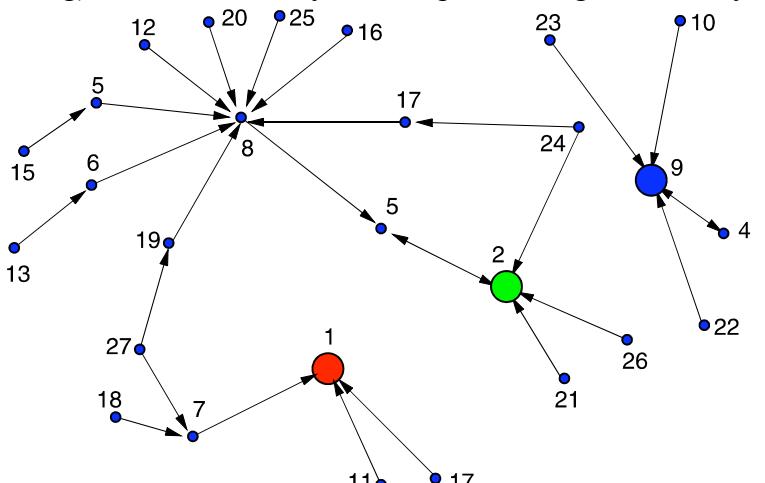
An alternative route is to extend the recently developed methods for identifying generating partitions directly from the iterated maps (see, e.g., refs [20, 21]). In this approach, one assumes a generating partition with a given number of entries. Each data point is then assigned to one of the partition elements such that when following the sequence of labels of two points initially close in the parameter space, the (Hamming) distance in the symbolic space diverges as slowly as possible. While promising, this method in itself does not give any guidance to how many elements the partition need to have in order to be generating. Second, in order to characterise the dynamics one would usually attempt to find a Markov representation; this is a highly non-trivial task and is one that the  $\varepsilon$ -machine reconstruction techniques are aimed at. Thus, combining the two methods in a self-consistent manner rather than sequentially (first

finding an approximately generating partition, and then using that partition as the alphabet in which one finds an  $\varepsilon$ -machine), i.e. a method for directly finding a Markov partition, could strengthen both approaches. For details of this work, see Deliverable D31.

### References

- References

  - [1] Flekkøy, E., Coveney, P. and Fabritiis, G.D., “ Foundations of Dissipative Particle Dynamics”, Physical Review E, 62(2), p. 2140-2157, 2000.
  - [2] Espanol, P., “ Statistical Mechanics of Coarse Graining”, in Novel Methods in Soft Matter Simulations, Springer, p. 69-115, 2003.
  - [3] Eriksson, A, Nilsson Jacobi, M, Nyström, J. and Tunstrøm, K., “Using force covariance to derive effective stochastic interactions in dissipative particle dynamics”, Physical Review E 77, 016707, 2008.
  - [4] Eriksson, A, Nilsson Jacobi, M, Nyström, J. and Tunstrøm, K., “Effective thermostat induced by coarse-graining of SPC water”, to appear in Journal of Chemical Physics.
  - [5] Lyubartsev, A. and Karttunen, M. and Vattulainen, I. and Laaksonen, A., “On Coarse-



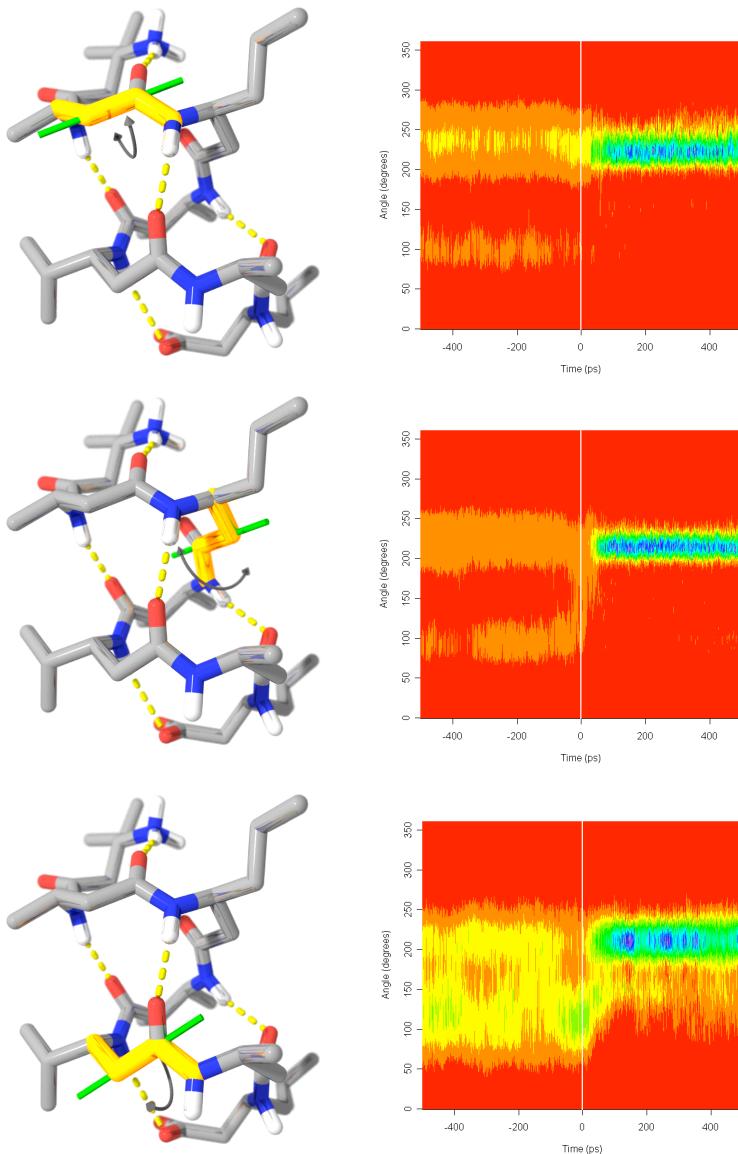
- Graining by the Inverse Monte Carlo Method: Dissipative Particle Dynamics Simulations Made to a Precise Tool in Soft Matter Modeling”, Soft Materials, 1, 1, p. 121-137, 2003.
- [6] Just, W. and Gelfert, K. and Baba, N. and Riegert, A. and Kantz, H., “Elimination of fast chaotic degrees of freedom: On the accuracy of the Born approximation”, Journal of Statistical Physics, 112, p. 277-292, 2003.
- [7] Nilsson Jacobi, M., “Quotient Manifold Projections and Hierarchical Dynamics”, in proceeding of ECCS, 2006.
- [8] Nilsson Jacobi, M., “Hierarchical Dynamics”, to appear in Springer's encyclopedia on complex systems,
- [9] Görnerup, O. and Nilsson Jacobi, M. “A Method for Inferring Hierarchical Dynamics in Stochastic Processes”, in Advances in Complex Systems, 11 (1), p. 1-16, 2008.
- [10] Nilsson Jacobi, M. and Görnerup, O. “A dual eigenvector condition for strong lumpability in Markov chains”, submitted to Journal of Linear algebra and its applications.
- [11] Nilsson Jacobi, M. and Görnerup, O. “Aggregation of variables in linear dynamical systems”, submitted for publication in Advances in complex systems.
- [12] Pothen, A., Simon, H., and Liou, K.“Partitioning sparse matrices with eigenvectors of graphs”, SIAM journal of matrix analysis and its applications, 11(3), p. 430-452, 1990.
- [13] Deuflhard, P., Huisenga, W., Fischer, A. and Schütte,C., “Identification of Almost Invariant Aggregates in Reversible Nearly Uncoupled Markov Chains”, Linear algebra and its applications, 315(1-3) p. 39-59, 2000.
- [14] Lindgren, K., Eriksson, A., and Eriksson, K.-E., “Flows of information in spatially extended chemical dynamics.” *Artificial Life IX Proceedings*, Edited by J. Pollack, M. Bedau, P. Husbands, T. Ikegami and R. A. Watson (MIT Press), 2004.
- [15] Lindgren, K., *Information Theory for Complex Systems – Lecture Notes* (Chalmers University of Technology), 2008.
- [16] Helvik, T., K. Lindgren, and M.G. Nordahl, ”Local information in one-dimensional cellular automata,” in proceedings for ACRI-2004: *From individual to collective behaviour*, Amsterdam, October, 2004. Published in *Springer Lecture Notes in Computer Science*, Vol. 3305, pp. 121-130 (Springer), 2004.
- [17] Helvik, T., K. Lindgren, and M.G. Nordahl, “Continuity of Information Transport in Surjective Cellular Automata,” *Communications in Mathematical Physics* **272**, 53-74, 2007.
- [18] Crutchfield, J.P., and Young, K., “Inferring statistical complexity”, *Physics Review Letters* **63**, 105, 1989.
- [19] Shalizi, C., “Causal architecture, complexity and self-organization in time series and cellular automata,” PhD thesis, University of Wisconsin, Madison, 2001.
- [20] Hirata, Y., Judd, K., Kilminster, D., “Estimating a generating partition from observed time series: Symbolic shadowing,” *Phys. Rev. E* **70**, 016215, 2004.
- [21] Buhl, M., Kennel, M. B., “Statistically relaxing to generating partitions for observed time-series data,” *Phys. Rev. E* **71**, 046213, 2005.

The Florence EMBIO group is composed of researchers from the Physics Department of the University of Florence as well as from the Florence-based Institute Of Complex Systems of the Italian National Research Council (CNR); all of them are affiliated with the Center for the Study of Complex Dynamics (CSDC) of the Florence University. In addition to the faculty and staff members (L. Casetti - local principal investigator, F. Bagnoli, R. Livi, A. Politi, B. Tiribilli, A. Torcini, M. Vassalli) who devoted part of their activity to the project, the activity of the group has benefited from the contribution of a PhD student funded by EMBIO, S. Luccioli, of a research scientist funded by EMBIO, M. Baiesi, of post-doc researchers partly/mostly funded by

EMBIO (L. Bongini, F. Sbrana, and R. Franzosi) and of post-docs and students on other grants (C. Guardiani and L. N. Mazzoni).

The activity of the Florence group was mainly devoted to the problem of protein folding. More precisely, the issue we focused on is the difference between proteins and all the other polymers. Not all polypeptides are proteins: only a very small subset of all the possible sequences of the twenty naturally occurring aminoacids have been selected by evolution. According to our present knowledge, all the naturally selected proteins fold to a uniquely determined native state, but a generic polypeptide does not (it rather has a glassy-like behavior). Then the following question naturally arises: what makes a protein different from a generic polypeptide? or, more precisely, which are the properties a polypeptide must have to behave like a protein, i.e., to fold into a unique native state regardless of the initial conditions, when the environment is the correct one? This question is, obviously, much less general than the whole folding problem, nonetheless if one could give it an answer it would surely help in the quest of a solution to the folding problem. To approach this issue we adopted a strategy which is based on the concept of "energy landscape". Energy landscape, or more precisely potential energy landscape, is the name commonly given to the graph of the potential energy of interaction between the microscopic degrees of freedom of the system [1]; the latter is a high-dimensional surface, but one can also speak of a free energy landscape when only its projection on a small set of collective variables (with a suitable average over all the other degrees of freedom) is considered [1]. Before having been applied to biomolecules, this concept has proven useful in the study of other complex

systems, especially of supercooled liquids and of the glass transition [2]. The basic idea is very simple, yet powerful: if a system has a rugged, complex energy landscape, with many minima and valleys separated by barriers of different height, its dynamics will experience a variety of time scales, with oscillations in the valleys and jumps from one valley to another. Then one can try to link special features of the behavior of the system (i.e., the presence of a glass transition, the separation of time scales, and so on) to special properties of the landscape, like the topography of the basins around minima, the energy distribution of minima and saddles connecting them and so on. Anyway, a complex landscape yields a complex dynamics, where the system is very likely to remain trapped in different valleys when the temperature is not so high. This is consistent with a glassy behavior, but a protein does not show a glassy behavior, it rather has relatively low frustration. This means that there must be some



property of the landscape such as to avoid too much frustration. This property is commonly referred to as the "folding funnel" [3]: though locally rugged, the low-energy part of the energy landscape is supposed to have an overall funnel shape so that most initial conditions are driven towards the correct native state. The dynamics must then be such as to make this happen in a reasonably fast and reliable way, i.e., non-native minima must be efficiently connected to the native state is that trapping in the wrong configuration is unlikely (although as our results showed -- see below -- that this picture is oversimplified).

Our main focus has then been to characterize the energy landscape of proteins and to understand what makes it different from that of other polymers that do not fold. To this end, we followed three main lines of research: (1) a general theoretical/computational one, (2) an experimental one, and a (3) computational one devoted to the modelization of the experimental geometry. Research line (1) forms the core of Workpackage 2 of the EMBIO project, while lines (2) and (3) together contribute to Workpackage 3. In the theoretical and computational studies we considered the potential energy/free energy landscape of minimalistic models of proteins (with the exception of the line (3) where all-atom models were also considered). Minimalistic models are those where the polymer is described at a coarse-grained level, as a chain of N beads where N is the number of aminoacids; no explicit water molecules are considered and the solvent is taken into account only by means of effective interactions among the monomers. Minimalistic models can be relatively simple, yet in some cases yield very accurate results which compare well with experiments [4].

### 1. Energy landscape of minimalistic models of proteins (WP 2)

A direct visualization of the energy landscape is impossible due to its high dimensionality, and its detailed properties must be inferred indirectly. A possible strategy is a local one: one searches for the minima of the landscape and then for the saddles connecting different minima. This is practically unfeasible for accurate all-atom potential energies, but may become accessible for minimalistic potentials; with a considerable numerical effort a reasonable sampling of these points can be achieved.

A first results that our group obtained while studying minimalistic two-dimensional models is that a funnel-like structure is present also in homopolymers and non-protein-like heteropolymers, but what makes a big difference is that in protein-like systems jumps between minima corresponding to distant configurations and minima in the native state are much more favoured dynamically, i.e., the transition rate towards the native state is considerably higher for protein-like sequences [L. Bongini, R. Livi, A. Politi, and A. Torcini, "Exploring the energy landscape of model proteins: a metric criterion for the determination of dynamical connectivity", *Physical Review E* 72, 051929 (2005)].

#### 1.1. The connectivity graph: dynamics and topology

The aim of this approach is to replace the "real" protein dynamics with a fictive dynamics on a graph whose nodes are the minima of the energy landscape and whose edges are the saddles between the minima; each link is weighted by a transition probability given by a Langer estimate involving the energy height of the saddles, the harmonic frequencies in the minimum and in the saddle, and temperature. This graph is called the connectivity graph. Numerical strategies to efficiently search the landscape in order to define this graph have been described in [L. Bongini, R. Livi, A. Politi, and A. Torcini, "Exploring the energy landscape of model proteins: a metric criterion for the determination of dynamical connectivity", *Physical Review E* 72, 051929 (2005); see also Deliverable 18 on the EMBIO website].

However, dynamics a connectivity graph may realistically mimick the "true" dynamics only if the nodes of the graph correspond to actual metastable states of the system: to this end, one should consider an effective graph where many minima are grouped together and considered as a single node if the transition rates between them are sufficiently high. Moreover, since transition rates depend on the temperature, at different temperatures one may get different effective graphs. One of our main results has been to define and implement a reliable renormalization procedure to define the effective connectivity graph. We checked this procedure against molecular dynamics

in the case of two-dimensional minimalistic models finding a very good agreement [L. Bongini, L. Casetti, R. Livi, A. Politi, and A. Torcini, "A graph theoretical analysis of the energy landscape of a model protein", paper in preparation; see also Deliverable 19 on the EMBIO website]. We also applied the same procedure to more realistic three-dimensional minimalistic models (introduced by C. Clementi and co-workers, see [4]); to this end we also developed an alternative procedure to explore the landscape [M. Baiesi, L. Bongini, and L. Casetti, "Enumeration of minima and saddles in the energy landscape of peptide models", paper in preparation; see also Deliverable 20 on the EMBIO website].

We also studied the topology of the connectivity graph, mainly for two-dimensional models (but some results have been obtained also in the case of three-dimensional models). The two topological properties studied were degree distribution and spectral dimension. As to the degree distribution, the renormalized graph shows a power-law distribution typical of scale-free graphs, in all cases. Moreover the renormalization procedure makes the native minimum a highly connected hub while at the other (clusters of) minima lose connectivity. This effect is strong for the heteropolymers and much weaker for the homopolymer. Another informative topological quantity when studying the diffusion over a graph is the spectral dimension of the graph itself. It has long been known that recurrence and first passage times in Markov chains on regular lattices diverge with the system size for lattice dimensions higher than 2. These results have been recently extended to general graphs by making use of the spectral dimension [5], that is the exponent describing how the integrated spectral density of the graph vanishes at low eigenvalues. It is important to notice that the spectral dimension is a very robust quantity, being invariant for topological rescaling such as the renormalizing procedure we introduced, local link redirection, decimation or creation. This property, which is rigorously true only in case of infinite graphs, still holds for sufficiently large graphs. Such topological invariance is very useful when dealing with graphs that are a sample of other graphs, because it implies that the spectral dimension will not change dramatically if some connections are lost or wrongly assigned. Before thermal renormalization we measured essentially the same spectral dimension for homopolymers, fast-folding and slow-folding heteropolymers. We then considered the effective renormalized graphs: the main result is that in the case of the homo-polymer the measured spectral dimension remains the same as before, while the hetero-polymers show a substantial decrease in this quantity for increasing temperatures. As a consequence, the configuration space of heteropolymers appears more compact to a random exploration than that of an homopolymer. The overall scenario emerging from this analysis of the spectral properties of connectivity graphs is that the zero-temperature connectivity graph, which contains no information at all about saddle height, is not able to discern any difference between the kind of amino-acidic sequence of the system. As soon as some information about saddles is introduced in the graph description in the form of a finite temperature connectivity graph, the distinction between homo- and hetero-polymers becomes possible and it is highlighted by a smaller spectral dimension for hetero-polymers. However, in order to be able to determine the actual folding propensity of a sequence, the full kinetic characterization of each saddle must be taken into account [L. Bongini, L. Casetti, R. Livi, A. Politi, and A. Torcini, "A graph theoretical analysis of the energy landscape of a model protein", paper in preparation; see also Deliverables 19 and 21 on the EMBIO website].

## 1.2. Global geometric properties of the energy landscape

Exploring energy landscapes searching for minima and saddles requires however a huge computational effort if one wants to obtain a good sampling. So the following question arises: is there some global property of the energy landscape which can be easily computed numerically as an average along dynamical trajectories and which is able to identify polymers having a protein-like behavior? One of our main results is that we were able to show that such a quantity indeed exists and is of a geometric nature. In particular, the fluctuations of a suitably defined curvature of the energy landscape clearly mark the folding transition while do not show any remarkable feature when the polymer undergoes a hydrophobic collapse without a preferred native state.

This is at variance with thermodynamic global observables, like the specific heat, which show a very similar behavior in the case of a folding transition and of a simple hydrophobic collapse. The global geometry of the energy landscape appears then to contain relevant information on the behavior of the various sequences and allow the proteinlike sequences to be distinguished from the others, because in the former case the fluctuations of the curvature as a function of the temperature show a pronounced peak close to the folding transition, while the others do not. We interpreted this behavior as due to an effective two-state behaviour of the system (this idea is supported also by an effective description of a 22-mer good folder sequence in terms of a single degree of freedom [L. N. Mazzoni and L. Casetti, "Effective description of folding in a simple model of a protein", paper in preparation]), because the native valley appears to be associated with a curvature considerably higher than that of the unfolded basin, so that when the systems frequently switches between the two basins, as it happens close to the transition, anomalously large fluctuations occur [L. N. Mazzoni and L. Casetti: "Curvature of the energy landscape and folding of model proteins", Physical Review Letters 97, 218104 (2006) and "Geometry of the energy landscape and folding in a simple model of a protein", Physical Review E 77, 051917 (2008)].

This analysis has been repeated on more refined minimalistic models able to describe real proteins, in particular S6 and SHC proteins [4]. The results obtained completely confirmed the above scenario [L. N. Mazzoni, R. Franzosi, L. Casetti, and C. Clementi, paper in preparation].

### 1.3 Good and bad folders

What picture emerges from these results? Can we say something definite about the differences in the energy landscapes of protein-like polymers (good folders) and of polymers that do not fold into a unique structure (bad folders, including homopolymers)? The answer can be split in three parts.

-- kinetic properties of the connectivity graph. The study of the energy distribution of minima and saddles shows that all the energy landscapes have a funnel-like structure, regardless of their folding propensity. The funnel shape is not sufficient to provide good foldability: what distinguishes good folders from bad folders is that the transitions to native state from distant minima are much more favoured kinetically for good folders [L. Bongini, R. Livi, A. Politi, and A. Torcini, Phys. Rev. E 72, 051929 (2005)].

-- topological properties of the connectivity graph. Here a clear-cut distinction between good and bad folders is not available yet: homopolymers show a markedly different behavior in the clustering and spectral dimension of their connectivity graphs with respect to heteropolymers, but among heteropolymers it is not clear how to discriminate between good and bad folders [L. Bongini, L. Casetti, R. Livi, A. Politi, and A. Torcini, in preparation]. This is coherent with the findings referred to above as to the kinetic properties.

-- global geometric properties of the energy landscape. These properties, and more specifically the fluctuations of a suitably defined curvature of the energy landscape, do discriminate between good and bad folders, in that a sharp peak of the fluctuations shows up in the good folder case. This has been interpreted as indicating an effective two-state behavior of good folders [L. N. Mazzoni and L. Casetti, Physical Review Letters 97, 218104 (2006), Physical Review E 77, 051917 (2008), and in preparation; L. N. Mazzoni, R. Franzosi, L. Casetti, and C. Clementi, in preparation].

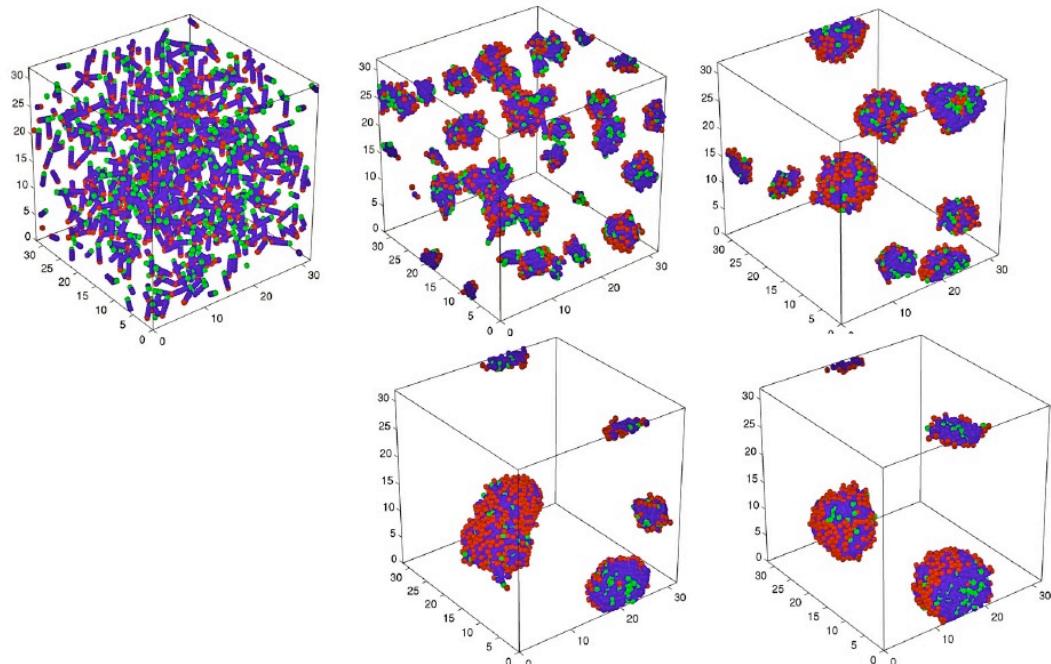
## 2. Atomic Force Microscopy experiments on single proteins (WP 3)

Exploiting the Atomic Force Microscopy potentialities we studied the unfolding process of a single real protein. Performing single molecule stretching experiments, it was possible to force the protein to unfold along a path defined by the experimental protocol, thus gaining information on structural and dynamics parameters that characterize the unfolding. In particular we strategically modified the experimental set up in order to extract information on the protein with high throughput and efficiency, of fundamental importance to reconstruct the free energy landscape of a protein by applying Jarzynsky's equality.

The Atomic Force Microscopy is mostly known for its imaging capabilities, but it is also a very sensitive tool for quantitatively measuring forces on a single molecule level. Hence the AFM is an alternative technique for the investigation of energy landscapes of proteins under native conditions. The core of an AFM is composed of a small tip attached to a cantilever. In the experiment the tip is driven to move up and down above a single point on the surface with a constant pulling speed (linear protocol), and once the tip of the cantilever is brought into contact with the surface, the molecules adsorbed on the substrate can interact with the tip and adhere to it; then, as the cantilever is retracted, the molecules resist the extension, causing the deflection of the cantilever. The force so exerted rises until a domain suddenly unfolds. When the protein is a multi-domain one further retraction causes the extension of successive unfolded domain until all domains are unfolded and the protein-tip interaction is broken. The corresponding force-distance curve has a typical saw-tooth pattern whose peaks correspond to the consecutive mechanical unfolding of each individual protein domain.

## 2.1. Experimental set-up

A first series of experiments was performed to optimize the experimental setup using a commercial AFM (Pico SPM, Agilent). The sample was an engineered polypeptide extracted



from Titin protein, a giant globular protein responsible for the contraction and the elasticity of the muscles, and it is composed by 300 domains among fibronectin III and immunoglobulin and also by a region called PEVK, distributed along two band A and I. We were interested in the I band that is directly responsible for the elasticity and it is made of immunoglobulin-like modules and the PEVK region. In particular we engineered two different molecules composed by 4 and 8 Ig-like domain starting from Ig27 module. Two cysteine residues inserted at the C terminus facilitated the absorption to the gold substrate. Some AFM hardware parts were improved during this procedure and first encouraging results were obtained and in good accordance with the literature.

## 2.2. Measurements on Titin

Successively a recombinant poly-protein composed of eight repeats of the Ig27 domain of Titin was also considered and investigated. From the analysis of the force-distance curve it was possible to extract structural information of the protein: the Worm Like Chain (WLC) model, that describes a peptide as a semiflexible chain in terms of contour length and of persistence length, was applied to each sawtooth profile in the force-distance curve. The variation in contour length between two consecutive peaks yielded the lengths of the single domains. From the

analysis of the force-distance curves it was also possible to extract kinetic parameters as the distance between native state and transition state and the unfolding rate at zero force. This allowed us to estimate the most probable unfolding force so that it was possible to extract two dynamical parameters: the unfolding length, usually interpreted as the deformation after which the molecule unfolds, and the typical unfolding rate at zero force. Mechanical unfolding experiments were performed on fragments of Titin with velocities ranging from 200 to 2000 nm/s at room temperature.

The mechanical unfolding of a protein is typically an out-of-equilibrium process. This prevents the possibility of performing such experiments in quasi-equilibrium conditions and thus of obtaining direct measurements of thermodynamic variables. However, this problem can be overcome by using the remarkable equality introduced by Jarzynski [6], which relates the free energy difference between the folded and the unfolded state of a biomolecule to the thermodynamic work done on the system along non-equilibrium trajectories. At least in principle one should then be able to extract thermodynamic information from repeated non-equilibrium pulling experiments. In the typical AFM pulling experiments the system includes the cantilever spring and the molecule so that to correctly calculate the free energy as a function of the molecular end-to-end distance an extended version of the Jarzynski equality that connects the nonequilibrium fluctuations of work to the free energy  $F(l)$ , where  $l$  is the molecular elongation, was considered. The free energy landscape as function of  $l$  was evaluated for different values of the pulling rate on the same unfolding trajectories acquired on a polypeptide of (Ig27)8. A collapse of the curves for the two smallest velocities could be observed, while the agreement with the curves obtained with larger values of pulling velocity worsened as the coordinate  $l$  increased. That is in agreement with simulation data (see section 3 below): a reliable estimate of the free energy can be obtained only if the estimated curves  $F(l)$  collapses onto a single curve as the pulling velocity is decreased [A. Imparato, F. Sbrana, and M. Vassalli, "Reconstructing the free energy landscape of a polyprotein by single-molecule experiments", *Europhysics Letters* 82, 58006 (2008).]

A new experimental set-up is in progress in order to verify these deductions through force clamp spectroscopy experiments.

## 2.2. Measurements on Elastin

We used the technique described above also to study the dynamical behaviour of the exon28 of the Elastin protein. Elastin is an insoluble extracellular matrix protein that provides elasticity for many tissues and organs in vertebrates as skin, lung and large blood vessels. It has a peculiar amino acid composition, being characterized by repetitive sequences rich in hydrophobic residues that can be considered responsible for its elasticity. Of paramount interest is the fact that many synthetic monomeric and polymeric sequences belonging to the putatively elastomeric regions of elastin are able to assemble themselves in a way similar to that of elastin itself.

Reasonable studies on isolated domains corresponding to the exons of elastin could give useful insights into the structural properties of the protein.

Single molecule force spectroscopy analysis was performed on exon28 of the human Elastin protein. The peptide were deposited on different substrate to asses the stability of the bond substrate and the peptide. Silanized mica and gold substrates were the best candidate.

Considering the peptide as a semiflexible rod, WLC model were applied to the data, allowing to estimate the structural parameters, persistence and contour length. The first results are encouraging and we expect to be able to reconstruct the free energy landscape also for this protein, though further work is needed.

## 3. Numerical simulations of protein stretching experiments (WP 3)

This research line aimed at computing the free-energy profiles of protein models when stretched via mechanical manipulation and to compare the obtained results with experimental measurements. The research line started with the study of a toy model by F. Bagnoli and C. Guardiani [C. Guardiani and F. Bagnoli, "A toy model of polymer stretching", *J. Chem. Phys.* 125, 084908 (2006)] and went on since January 2006, when S. Luccioli started the EMBIO

funded PhD on "Nonlinear Dynamics and Complex Systems" at the University of Florence under the supervision of A. Torcini, to the end of the project.

### 3.1. Minimalistic models

The first step has been the development of MD codes to study the model protein pulled at constant velocity and constant force. The model chosen is a so-called "3-color, 46-bead", in the specific the heteropolymer is made of a chain of 46 beads linked by spring-like bonds and with three kinds of residues: hydrophilic, hydrophobic and neutral. Furthermore, we have chosen to study the folding properties under mechanical stress of two specific sequences: namely, a sequence known in literature to give rise to a reasonably fast-folder; and a random sequence. These sequence had been also studied from the geometric point of view, but in equilibrium conditions, by L. N. Mazzoni and L. Casetti (see 1.2).

We aimed at reconstructing the equilibrium free energy landscape from out-of-equilibrium steered molecular dynamics data as a function of an internal coordinate, namely the end-to-end distance. This task has been accomplished via two independent methods: by employing an extended version of the Jarzynski equality (EJE) and the inherent structure (IS) formalism. For a sequence with good foldability properties, a quantitative agreement between the free energies obtained with the two schemes is observed in a range of temperatures around the "folding transition". Moreover, for this sequence the mechanically induced structural transitions can be related to thermodynamical aspects of folding. However, for a random sequence the IS reconstruction is unable to reproduce the free energy landscape, since in this case the main contribution to entropic terms cannot be ascribed to vibrational fluctuations around the ISs, being due to large conformational rearrangements instead [A. Imparato, S. Luccioli, and A. Torcini, "Reconstructing the free energy landscape of a mechanically unfolded model protein", Phys. Rev. Lett. 99 (2007) 168101 (erratum in 100, 159903(E) (2008)); S. Luccioli, A. Imparato, and A. Torcini, "Free energy landscape of mechanically unfolded model proteins: extended Jarzinsky versus inherent structure reconstruction", submitted to Phys. Rev. E, (2008).]

It is important to notice that the same approach has been successfully applied by the Florence EMBIO experimental group, as well as by other groups, to reconstruct the free energy landscape of an experimentally manipulated sequence of titin I27 domains ([A. Imparato, F. Sbrana and M. Vassalli, Europhys. Lett. 82, 58006 (2008) and Ref. 7].

### 3.2. All-atom models

Thanks to a collaboration with A. Irbaeck (Lund University, Sweden), we studied also mechanical unfolding of all-atoms models. In particular, we have investigated the mechanical unfolding of the tenth type III domain from fibronectin, both at constant force and at constant pulling velocity, by all-atom Monte Carlo simulations. We observed both apparent two-state unfolding and several unfolding pathways involving one of three major, mutually exclusive intermediate states. All the three major intermediates lack two of seven native beta-strands, and share a quite similar extension. The unfolding behavior is found to depend strongly on the pulling conditions. In particular, we observe large variations in the relative frequencies of occurrence for the intermediates. Using the extended Jarzynski equality, we also estimate the equilibrium free-energy landscape, calculated as a function of chain extension. The application of a constant pulling force leads to a free-energy profile with three local minima, one of which can be associated with the unfolding intermediates [S. Mittmann, S. Luccioli, A. Torcini, A. Imparato, and A. Irbaeck, "Changing the mechanical unfolding pathway of FnIII\_10 by tuning the pulling strength", submitted to Biophysical Journal (2008)].

Given the free energy landscape at equilibrium as a function of an internal reaction coordinate, which kind of information about the folding dynamics can be extracted from the knowledge of the free-energy profile itself? We plan to address this problem by examining the unfolding kinetics of a coarse-grained off-lattice model as well as all-atom models subjected to an external constant pulling force  $F$ . In particular, we would like to examine up to which extent a relationship between the average unfolding times and a generalization of the Kramers expression, which includes free-energy barriers, still holds. We expect to observe a cross-over

from a unfolding process reproducible in terms of activation mechanisms to a scenario where the unfolding times are essentially ruled by a Markovian process plus a drift proportional to the external force F.

#### 4. Other research lines

Researchers in the EMBIO Florence group were also involved in other research lines that although not directly included in the workpackages of the project shared with them conceptual issues and/or theoretical, computational and experimental techniques.

##### 4.1. Minimalistic modeling of real proteins

The use of minimalistic models as those referred to above is not limited to the study of general issues: Florence EMBIO researchers were able to show that they can be fruitful also to study some properties of real proteins.

C. Guardiani and R. Livi, in collaboration with F. Cecconi (ISC-CNR, Rome, Italy) first considered a coarse-grained model based upon the chemistry of (Sorenson/Head-Gordon model, SHG), benchmarking it on the WW domain of the hPin1 protein, due to the abundance of experimental data for this system. The model yielded correct predictions on the native topology of the WW domain, but failed in describing cooperativity and reversibility of the folding process, which is at variance well captured by a Go model. Hence they considered the latter model to address a problem of utmost biomedical relevance: the study of the molecular impact of a series of mutations in the C5 domain of the Myosin Binding Protein C (MyBP-C) involved in a serious cardiac genetic disease. Their results have shown that the manifestation of the disease is correlated with the thermodynamic destabilization as well as with the speeding up of the folding process due to the mutations [F. Cecconi, C. Guardiani, and R. Livi, "Stability and kinetic properties of C5-domain of Myosin binding protein C and its mutants" *Biophysical Journal* 94, 1403-1411, (2007); "Computational analysis of folding and mutation properties of C5 domain of Myosin binding protein C", *Proteins: Structure, Function, and Bioinformatics* 70, 1313-1322, (2008); "Analyzing pathogenic mutations of C5 domain from cardiac myosin binding protein C through MD simulations", *European Biophysics Journal* (2008, to appear)].

##### 4.2. Lattice models of biopolymers with vibrations

The simplest description of a polymer is a self-avoiding walk on a lattice. M. Baiesi and L. Casetti, in collaboration with the group of P. De Los Rios and F. Piazza (EPFL Lausanne, Switzerland) have developed an efficient code for the simulation of a new lattice model that incorporates vibrational effects into the weight of each configuration. The model is supposed to mimick the minima of proteins. Its statistical properties so far have been investigated mainly for the homopolymer case. It has novel features compared with the standard self-avoiding walk. Moreover, the study of some simple motifs resembling alpha-helices and beta-sheets is currently under study [M. Baiesi, L. Casetti, P. De Los Rios, C. Maffi, and F. Piazza, "A new lattice model of polymers with vibrational entropy", in preparation].

##### 4.3. Minimalistic models of proteins and protein aggregation

The shape of the energy landscape in the neighbourhood of the global energy minimum -- and most precisely its curvature -- has been analysed by L. Bongini and L. Casetti in collaboration with the group of P. De Los Rios and F. Piazza (EPFL Lausanne, Switzerland) as concerns the problem of protein aggregation. This very common phenomenon in all sorts of polypeptide chains leads to the formation of large supramolecular aggregates that share a unique beta-spine architecture. The lack of clear signatures of the propensity of aggregation at the sequence level suggests that the phenomenon suggests that the phenomenon is governed by a quite general mechanism, possibly underpinning a thermodynamic origin. In order to investigate this hypothesis we have analyzed, using atomistic molecular models, the dependence of vibrational entropy from the content of secondary structure in a large database of protein structures. The origin of the observed correlations was then addressed by means of a simple coarse-grained model [L. Bongini, F. Piazza, L. Casetti, and P. De Los Rios, The role of vibrational entropy in the structural organization of protein aggregates, submitted to *Physical Biology* (2008)].

##### 4.4. Frequency of knots in swollen and collapsed polymers

One of the main topological features of biopolymers is the presence of knots, whose role and effects are not completely understood yet.

In collaboration with A. L. Stella and E. Orlandini (University of Padua), M. Baiesi has studied the frequency of knots in interacting self-avoiding walks, which are the classical model of homopolymers in solution. The presence of knots, their kind and their frequency have been numerically and theoretically studied in two regimes: high and low temperature (swollen and collapsed phase, respectively). It turns out that the statistics of knots in these regimes is quite different. At high temperature knots are localized along the polymer, and they act as independent entities. This fact suggests a prediction for the dependence of the knot frequency on the polymer length and on the presence of other knots, which works remarkably well when compared to numerical data [M. Baiesi, E. Orlandini, and A. L. Stella, "Ranking knots of random, globular polymer rings", *Physical Review Letters* 99, 058301 (2007)]. At low temperature (collapsed phase) knots instead are delocalized in the globule, and they appear with much more frequency than at high temperature [ M. Baiesi, E. Orlandini, and A. L. Stella, "The frequency of knots in swollen polymers", in preparation]. Numerical data show that the number of different possible knots grows at least exponentially with the chain length.

#### 4.5. Dynamics of DNA denaturation

Minimalistic models are useful not only for proteins but also for other biopolymers, especially DNA. M. Baiesi and R. Livi have put forward a new model to describe the dynamics of a long double-stranded DNA that melts in a high temperature regime. The main novelty in the model is that it correctly includes the effect of the entanglement of the two strands. While in previous models this was not considered, the twist around each other strand due to the double-helical structure is a constraint relevant for the dynamical rules of the model. The model is of the Poland-Scheraga type, and we have shown that our version predicts long denaturation times, scaling with the length of the macromolecule. Moreover, the denaturation is characterized by two regimes: in the first one, the double-helix opens by untwisting portions of DNA close to its ends and redistributes the twisted parts along the sequence. When a critical density of bubbles is reached, a second slower regime takes place, with bubbles forming entropic barriers for residual helical segments, which are in fact almost trapped inside the DNA. The final denaturation is reached in times scaling roughly as the cube of DNA length [M. Baiesi and R. Livi, "Multiple time scales in a model for DNA denaturation dynamics", submitted to *Phys. Rev. Lett.* (2007)].

#### 4.6. AFM imaging of Titin and Elastin

As stated above in section 3, the atomic force microscope is best known as an imaging tool. Although the Florence group mainly used it a force spectrometer, its imaging capabilities were also exploited to perform morphological investigation on the two proteins studied with force spectroscopy, namely Titin and Elastin.

AFM imaging investigations on two fragments of titin protein composed of four domains from Ig27 to Ig30 (tetramer) and eight domains from Ig27 to Ig34 (octamer) were combined to Dynamic Light Scattering (DLS) analysis turning out to be complementary for size measurements of the fragments and fragment aggregates. An unexpected result was that the octamer folds into a smaller structure than the tetramer and the unfolded octamer is also smaller than the unfolded tetramer. This feature could be related to the significance of the hydrophobic interactions between domains of the fragment. The longer the fragment, the more easily the hydrophobic parts of the domains interact with each other. The fragment aggregation behavior, in particular conditions, was also revealed by both DLS and AFM as a process that is parallel to the folding-unfolding transition [S. Marchetti, F. Sbrana, R. Raccis, L. Lanzi, C.M.C. Gambi, M. Vassalli, B. Tiribilli, A. Pacini, and A. Toscano, "Dynamic light scattering and atomic force microscopy imaging on fragments of beta-connectin from human cardiac muscle", *Physical Review E* 77, 021910 (2008)].

Morphological AFM investigation were performed also on exon28 of the human Elastin protein.

#### 5. Collaborations and visits

Many exchanges occurred with EMBIO partners, especially with the Cambridge group (D. Nerukh) and with the Groningen group (A. Wassenaar, A. Rampioni). A number of external collaborations were ongoing or started during the project: the most important were with C. Clementi (Rice Univ., Houston, USA), P. Bruscolini (BiFi Zaragoza, Spain), P. De Los Rios and F. Piazza (EPFL Lausanne), F. Cecconi (ISC-CNR, Rome, Italy), A. Imparato (Politecnico, Turin, Italy), A. Irbaeck (Lund University, Sweden).

#### References

- [1] D. J. Wales, "Energy landscapes" (Cambridge University Press, Cambridge, 2004).
- [2] P. G. Debenedetti and F. H. Stillinger, Nature 410, 259 (2001).
- [3] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, and N. D. Soccia, Proc. Natl. Acad. Sci. USA 92, 3626 (1995).
- [4] C. Clementi, Curr. Opin. Struct. Biol. 18, 10 (2008).
- [5] R. Burioni and D. Cassi, Phys. Rev. Lett. 76 1091 (1996).
- [6] C. Jarzynski, Phys. Rev. Lett. 78, 2690 (1997).
- [7] N. C. Harris, et al., Phys. Rev. Lett. 99, 068101 (2007).

In Freiburg the work was performed to enable energy landscape studies of HP-lattice proteins (WP 1) a detailed investigation of the lowest energy structures is of high importance. The constraint programming based CPS-P approach (Backofen and Will, 2006) is the only exact method that allows the complete calculation of all optimal (minimal energy) structures of sequences in the HP-model. The method is tailored for complex 3D-lattices as the cubic and face centered cubic lattice. We improved the implementation to allow for high throughput experiments and set up a package of tools that covers the main tasks needed for lattice protein studies (Mann et al., 2008). We investigated new constraint methods to speedup the degeneracy calculation and structure enumeration (deliverable 9). Furthermore, we extended the approach to enable a statistical studies of these lowest energy structures by enumerating only sufficiently different structures according to defined measures (deliverable 10). A combination of the basic constraint approach and stochastic methods enabled a randomized sampling of optimal structures that was not possible before (deliverable 11).

In addition, a generic C++ library, the Energy Landscape Library (ELL), was developed to be the base of our studies. This library allows for the development and implementation of landscape model independent algorithms (as the flooding algorithm by Wolfinger et al.) and the modular application of them to separately defined landscape models. This platform was well suited to implement the newly defined measures and algorithms within the project (deliverable 5). For example, we have integrated the funnel partitioning of an energy landscape as defined by the Leipzig group in deliverable 6. Or we developed advanced methods for an efficient landscape flooding or landscape structure sampling. We are in close collaboration with the groups from Leipzig and Vienna to investigate the features of discrete energy landscapes for RNA, lattice proteins, or spin glass models. Finally we made some progress in the application of the curvature measure by the Florence group to discrete energy landscapes. The measure allows for a folding sampling based classification of sequences into "good" and "bad" folders (WP 3) and was previously developed and applied to off-lattice proteins. We have been working on an equivalent definition of the measure in the discrete case and have implemented the relevant tools based on the ELL.

The study of HP-like lattice protein models faces the problem of the enormous degeneracy of the system. Therefore, most of the molecules show extreme numbers of optimal structures (e.g. several million) even for short sequence lengths. The CPS-P approach, developed by our group (Backofen and Will, 2006) is the only method that allows for a complete enumeration of these structures and of the minimal energy a sequence can adopt. We have increased the power of the method by developing a new search strategy that resulted in significant speedups (delivery 9). To allow for a statistical investigation of such optimal structures, a reasonable subset has to be

derived. We extended the exact CPSP-approach for optimal structure prediction in the HP-model to allow for an enumeration of such a subset. The new feature enables the calculation of all optimal structures that show a defined structural distance between each other. This provides a way to characterise differences of the lowest energy structures. This extension constitutes deliverable 10. Furthermore, we introduced the possibility to sample optimal structures using a combination of the exact constraint approach with stochastic methods (deliverable 11). A major extension of the CPSP-approach was its application to side chain protein lattice models. We developed the first method that is able to predict all optimal structures within this model and to calculate the minimal energy reachable. We investigated the differences of the energy landscapes between the 'standard' backbone model with this extended side chain model.

The Energy Landscape Library (ELL) was developed to serve as a platform for generic algorithm for discrete energy landscapes. The C++ library focuses on a strict separation of the algorithms and the landscape models without loosing efficiency. It is therefore well suited to test newly developed measures for energy landscape properties and as a test platform for new algorithms that are the goals of WP 1. For instance, the new landscape measure of the funnel partitioning was defined in deliverable 6 by the Leipzig group. It describes the partitioning of the landscape into the basins of attractions of minima and their relations among each other. The resulting directed graph representation represents basins (nodes) and the minimal escape paths from a basin to another basin with lower energy. The graph shows therefore a decomposition into several sub-graphs that define folding funnels in the energy landscape. To enable detailed studies on these funnel partitions, we integrated their basic data structures and the generic algorithms needed for their production into the Energy Landscape Library (ELL). The resulting methods are highly model independent and allow therefore for their application on a variety of discrete energy landscape models. We are currently investigating the folding funnel partitions of lattice proteins to identify good and bad folding sequences. The implementation and application resulted in deliverable 5.

Another application of the C++ Energy Landscape Library (ELL), developed within the project (see WP 1), was the classification of sequences according to their folding behavior in the HP-lattice protein model. This was done by using a recently published structural move set, the pull-moves, that are better suited to model the local behavior of folding and are therefore more realistic than the previously applied pivot-moves. We studied sequences with non-degenerated ground states, i.e. unique optimal structure, that were identified using CPSP-based method from the CPSP-tools package (Mann et al., 2008) (see WP 1). We encountered significant differences in the folding behavior and were able to classify the sequences. Only very few were able to reach their unique ground state and were classified as "good" folders. Many sequences never found their optimum and were defined as "bad" folders. The resulting sequence sets were used to adopt and test the curvature measure (for off-lattice proteins) of the Florence group to the discrete lattice model.

## 2. Dissemination and use

**Overview table**

Planned /actual Dates	Type	Type of audience	Countries addressed	Size of audience	Partner responsible /involved
	<i>Press release(press/radio/TV)</i>	<i>General public</i>			
	<i>Media briefing</i>	<i>Higher education</i>			
	<i>Conference</i>	<i>Research</i>			<i>All partners</i>
	<i>Exhibition</i>	<i>Industry ( sector x)</i>			
	<i>Publications</i>				<i>All partners</i>
	<i>Project web-site</i>				
	<i>Posters</i>				
	<i>Flyers</i>				
	<i>Direct e-mailing</i>				<i>All partners</i>
	<i>Film/video</i>				

**Project's website:**

<http://www-embio.ch.cam.ac.uk/>

**Publications of the Consortium:**

1. Neusius T., Daidone I., Sokolov I.M. and Smith J.C., **Subdiffusion in peptides originates from fractal-like structure of configuration space**, *Phys. Rev. Lett.* 100, 188103 (2008).
2. Noé F., Daidone I., Di Nola A., Smith J.C., Di Nola A. and Amadei, A. **Solvent electrostriction driven peptide folding revealed by Quasi-Gaussian entropy theory and molecular dynamics simulation**, *J. Phys. Chem.* (2008), in press.
3. Andreas R. Gruber, Ronny Lorenz, Stephan H Bernhart, Richard Neubock, and Ivo L. Hofacker (2008). **The Vienna RNA websuite**. , *Nucl. Acids Res.*,36:W70-W74.
4. Christoph Flamm and Ivo L. Hofacker (2008). **Beyond energy minimization: Approaches to the kinetic folding of RNA**. , *Monatsh. f. Chemie*, 139:447- 457.
5. Muckstein, Ulrike and Tafer, Hakim and Bernhard, Stephan H. and Hernandez-Rosales, Maribel and Vogel, Jorg and Stadler, Peter F. and Hofacker, Ivo L., **Translational Control by RNA-RNA Interaction: Improved Computation of RNA-RNA Binding Thermodynamics**, *BioInformatics Research and Development --- BIRD 2008*, Springer, Berlin, 2008, in press,
6. Klemm, Konstantin and Flamm, Christoph and Stadler, Peter F., **Funnels in Energy Landscapes**,, 2008, *Eur. Phys. J. B*, 63, 387-391, ECCS 2007 contribution
7. Geis, Michael and Flamm, Christian and Wolfinger, Michael T. and Tanzer, Andrea and Hofacker, Ivo L. and Middendorf, Martin and Mandl, Christian and Stadler, Peter F. and Thurner, Caroline, **Folding kinetics of large RNAs**, *J. Mol. Biol.*, 379, 160-173, 2008,

8. Geis, Michael and Middendorf, Martin,  
**A Particle Swarm Optimizer for Finding Minimum Free Energy RNA Secondary Structures,,**  
*Proceedings of the IEEE Swarm Intelligence Symposium 2007*, 1-8, 2007
9. Heine, Christian and Scheuermann, Gerik and Flamm, Christoph and Hofacker, Ivo L. and Stadler, Peter F.,  
**Visualization of Barrier Tree Sequences,,**  
*IEEE Trans. Vis. Comp. Graphics*, 12, 781-788, 2006, Info-Vis 2006 contribution,
10. Christoph Flamm, Ivo L. Hofacker, Barbel M. R. Stadler, and Peter F. Stadler,  
**Saddles and Barrier in Landscapes of Generalized Search Operators,,**  
*Foundations of Genetic Algorithms IX*, Stephens, C. R. and Toussaint, M. and Whitley, D. and Stadler, P. F., Springer, Berlin, Heidelberg, 2007, 194-212, Lecture Notes Comp. Sci., 4436,
11. Kei Moritsugu and Jeremy C. Smith,  
**REACH: Realistic Extension Algorithm via Covariance Hessian,,**  
*Biophys. J.* 93, 3460-3469 (2007).
12. Kei Moritsugu and Jeremy C. Smith,  
**REACH Coarse-Grained Biomolecular Simulation: Transferability between Different Protein Structural Classes,,**  
*Biophys. J.* In press.
13. B. Tarus and J. C. Smith.  
**A study of the folding of the A?21-30-peptide under confining conditions.,**  
*in preparation*
14. Periole, X. and Mark, A. E.  
**(2007) Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent,,**  
*J. Chem. Phys.* 126, 014903.
15. Periole, X., Rampioni, A., Vendruscolo, M. and Mark, A.E.  
**(2007) Molecular dynamics simulations of a cross-? filament of the GNNQQNY peptide reveal a combination of factors that determine the degree of twist of ?-sheets.,**  
*Biophysical Journal* (submitted)
16. Xavier Periole, Lucy R. Allen, Kamil Tamiola, Alan E. Mark, Emanuele Paci  
**Probing the free energy landscape of the FBP28 WW domain by multiple techniques.,**  
*Journal Molecular Biology* (submitted)
17. Fuchslin, R. M., A. Eriksson, H. Fellermann, H.-J. Ziock,  
**"Coarse-graining and scaling in dissipative particle dynamics,",**  
*Submitted to Journal of Chemical Physics*. Preprint availabel at arXiv:cond-mat/0703.682, 2007.
18. Eriksson, A, Nilsson Jacobi, M, Nystrom, J. and Tunstrom, K.,  
**"Using force covariance to derive effective stochastic interactions in dissipative particle dynamics".,**  
*Physical Review E* 77, 016707, 2008.
19. Eriksson, A, Nilsson Jacobi, M, Nystrom, J. and Tunstrom, K.,  
**"Effective thermostat induced by coarse-graining of SPC water".,**  
*to appear in Journal of Chemical Physics*.

20. Gornerup, O., and J. P. Crutchfield,  
**Primordial Evolution in the Finitary Process Soup,,**  
*Electronic Journal of Theoretical Physics Special Issue 2007: Physics of Emergence and Organization*, 4:16 I (2007). In press.
21. Gornerup, O. and Nilsson Jacobi, M.  
**"A Method for Inferring Hierarchical Dynamics in Stochastic Processes",,**  
*in Advances in Complex Systems*, 11 (1), p. 1-16, 2008.
22. Helvik, T., K. Lindgren, and M. G. Nordahl,  
**"Continuity of information transport in surjective cellular automata,",**  
*Communications in Mathematical Physics* 272, 53-74 (2007).
23. Lindgren, K.,  
**Information Theory for Complex Systems - Lecture Notes.,**  
*Chalmers University of Technology*, 2008. (To be revised and exteneded for a book publication.)
24. Nilsson Jacobi, M.,  
**"Quotient Manifold Projections and Hierarchical Dynamics",,**  
*in proceeding of ECCS*, 2006.
25. Nilsson Jacobi, M.,  
**"Hierarchical Dynamics",,**  
*to appear in Springer's encyclopedia on complex systems*,
26. Nilsson Jacobi, M. and Gornerup, O.  
**"A dual eigenvector condition for strong lumpability in Markov chains",,**  
*submitted to Journal of Linear algebra and its applications*, arXiv:0710.1986
27. Nilsson Jacobi, M. and Gornerup, O.  
**"Aggregation of variables in linear dynamical systems",,**  
*submitted for publication in Advances in complex systems*.
28. Nilsson Jacobi, M. and Gornerup, O.  
**"An algorithm for aggregating variables in linear dynamical systems",,**  
*submitted for publication in Advances in complex systems*.
29. Christian H. Jensen, Dmitry Nerukh, and Robert C. Glen,  
**Controlling protein molecular dynamics: how to accelerate folding while preserving the native state,**  
accepted for publication in *J. Chem. Phys.*
30. Dmitry Nerukh,  
**Computational Mechanics reveals nanosecond time correlations in molecular dynamics of liquid systems,**  
[Chem. Phys. Lett.](#), 457(4-6), 439-443 (2008) Erratum to "Computational mechanics reveals nanosecond time correlations in molecular dynamics of liquid systems"  
[[Chem. Phys. Lett. 457 \(2008\) 439](#)], [Chem. Phys. Lett.](#), 459(1-6), 203 (2008)
31. Dmitry Nerukh,  
**Dynamical frustration of protein's environment at the nanoseconds time scale,**  
accepted for publication in *J. Mol. Liq.*
32. V.A. Buts, A.G. Nerukh, N.N. Ruzhytska, and D.A. Nerukh,  
**Wave Chaotic Behaviour Generated by Linear Systems,**  
accepted for publication in [Optical and Quantum Electronics](#)
33. Dmitry Nerukh, Vladimir Ryabov, and Robert C. Glen,  
**Complex temporal patterns in molecular dynamics: a direct measure of the phase**

- space exploration by the trajectory at macroscopic time scales,**  
*Phys. Rev. E*, **77**, 036225 (2008) (reproduced in *Virtual Journal of Biological Physics Research*, **15**(7), (2008))
34. Christian H. Jensen, Dmitry Nerukh, and Robert C. Glen,  
**Sensitivity of peptide conformational dynamics on clustering of a classical molecular dynamics trajectory,**  
*J. Chem. Phys.*, **128**, 115107 (2008) (reproduced in *Virtual Journal of Biological Physics Research*, **15**(7), (2008))
35. F. Cecconi, C. Guardiani, and R. Livi,  
**"Stability and kinetic properties of C5-domain of Myosin binding protein C and its mutants",**  
*Biophysical Journal*, **94**, 1403-1411, (2007).
36. C. Guardiani, F. Cecconi, and R. Livi,  
**"Computational analysis of folding and mutation properties of C5 domain of Myosin binding protein C",**  
*Proteins: Structure, Function, and Bioinformatics* **70**, 1313-1322, (2008).
37. F. Cecconi, C. Guardiani and R. Livi,  
**"Analyzing pathogenic mutations of C5 domain from cardiac myosin binding protein C through MD simulations",**  
*European Biophysics Journal* (2008, to appear).
38. M. Baiesi, E. Orlandini, and A. L. Stella,  
**"Ranking knots of random, globular polymer rings"**,  
*Physical Review Letters* **99**, 058301 (2007).
39. M. Baiesi and R. Livi,  
**"Multiple time scales in a model for DNA denaturation dynamics"** ,  
*submitted for publication* (2007).
40. L. N. Mazzoni and L. Caselli,  
**"Geometry of the energy landscape and folding in a simple model of a protein"**,  
*Physical Review E* **77**, 051917 (2008).  
[selected for the Virtual Journal of Biological Physics Research]
41. A. Imparato, S. Luccioli, and A. Torcini,  
**"Reconstructing the free energy landscape of a mechanically unfolded model protein"**,  
*Physical Review Letters* **99**, 168101 (2007).  
[selected for the Virtual Journal of Biological Physics Research]
42. A. Imparato, S. Luccioli, and A. Torcini,  
**"Erratum: Reconstructing the Free-Energy Landscape of a Mechanically Unfolded Model Protein"**,  
*Physical Review Letters* **100**, 159903(E) (2008).
43. S. Luccioli, A. Imparato, and A. Torcini,  
**"Free energy landscape of mechanically unfolded model proteins: extended Jarzinsky versus inherent structure reconstruction"**,  
*submitted for publication* (2008).
44. S. Mitternacht, S. Luccioli, A. Torcini, A. Imparato, and A. Irbaeck,  
**"Changing the mechanical unfolding pathway of FnIII\_10 by tuning the pulling strength"**,  
*submitted for publication* (2008).

45. S. Marchetti, F. Sbrana, R. Raccis, L. Lanzi, C.M.C. Gambi, M. Vassalli, B.Tiribilli, A. Pacini, and A. Toscano,  
**"Dynamic light scattering and atomic force microscopy imaging on fragments of beta-connectin from human cardiac muscle",**,  
*Physical Review E* 77, 021910 (2008).  
[selected for the Virtual Journal of Biological Physics Research]
46. A. Imparato, F. Sbrana, and M. Vassalli,  
**"Reconstructing the free energy landscape of a polyprotein by single molecule experiments",**,  
*Europhysics Letters* 82, 58006 (2008).
47. F. Sbrana, C. Sassoli, D. Nosi, R. Squecco, F. Paternostro, B. Tiribilli, S. Zecchi Orlandini, F. Francini, and L. Formigli,  
**"Role for stress fiber contraction on membrane tension and stretch activated channel activation",**,  
*submitted for publication* (2008).
48. Martin Mann, Sebastian Will, and Rolf Backofen,  
**The energy landscape library - a platform for generic algorithms,**  
*In BIRD'07* volume 217, pages 83-86. Oesterreichische Computer Gesellschaft, 2007
49. Martin Mann, Sebastian Will, and Rolf Backofen,  
**CPSP-tools - exact and complete algorithms for high-throughput 3D lattice protein studies,**  
*BMC Bioinformatics* 9 pp. 230, 2008. doi:10.1186/1471-2105-9-230
50. Daidone I, Ulmschneider MB, Di Nola A, Amadei A, Smith JC,  
**Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding,**  
*Proc Natl Acad Sci U S A.* 2007 Sep 25;104(39):15230-5
51. T. Thalheim, D. Merkle, M. Middendorf,  
**A Hybrid Population based ACO Algorithm for Protein Folding,**  
*Proc. 2008 IAENG International Conference on Bioinformatics* 6 pp. (2008)
52. A. Imparato, S. Luccioli, and A. Torcini,  
**Reconstructing the free energy landscape of a mechanically unfolded model protein,**  
*Phys. Rev. Lett.* to appear (2007)
53. Peter F. Stadler and Barbel M. R. Stadler,  
**Genotype Phenotype Maps,**  
*Biological Theory* 3, 268-279 (2006)
54. Matthias Kruspe and Peter F. Stadler,  
**Progressive Multiple Sequence Alignments from Triplets,**  
*BMC Bioinformatics* in press (2007)
55. S. Janson, D. Merkle,  
**A New Multi-Objective Particle Swarm Optimization Algorithm Using Clustering Applied to Automated Docking,**  
in *Proc. Second Int. Workshop on Hybrid Metaheuristics (HM 2005)*, Springer, LNCS 3636, 128-140, (2005)
56. S. Janson, D. Merkle, M. Middendorf,  
**Solving molecular docking with particle swarm optimization,**  
in *Abstracts of 5th Biotechnology Symposium 2006, Center of Biotechnology and Biomedicine, University of Leipzig*, 77, (2006)

57. S. Janson, M. Middendorf,  
**A Hierarchical Particle Swarm Optimizer for Noisy and Dynamic Environments**,  
Accepted for *Genetic Programming and Evolvable Machines*
58. Kei Moritsugu and Jeremy C. Smith,  
**Langevin Model of the Temperature and Hydration Dependence of Protein Vibrational Dynamics**,  
*J. Phys. Chem. B*, 109, 12182-12194 (2005)
59. Kei Moritsugu and Jeremy C. Smith,  
**Temperature Dependent Protein Dynamics: A Simulation-Based Probabilistic Diffusion-Vibration Langevin Description**,  
*J. Phys. Chem. B*, 110, 5807-5816 (2006)
60. F. Noe', D. Krachtus, J. C. Smith and S. Fischer,  
**Transition Networks for the Comprehensive Characterization of Complex Conformational Change in Proteins**,  
*J. Chem. Theory and Comput.*, 2, 840-857 (2006)
61. F. Noe', M. Oswald, G. Reinelt, S. Fischer and J. C. Smith,  
**Computing Best Transition Pathways in High-Dimensional Dynamical Systems: Application to the L - R Transitions in Octaalanine**,  
*Multiscale Model. Sim.*, 5, 393-419 (2006)
62. O. Gornerup and J. P. Crutchfield,  
**Hierarchical Self-Organization in the Finitary Process Soup**,  
To appear in *Proceedings of ALife X*, 2006
63. M. Nilsson,  
**Hierarchical Structures in Smooth Dynamical Systems**,  
*Artificial Life* 11, 493-512 (2005)
64. Sebastian Will and Martin Mann,  
**Counting protein structures by dfs with dynamic decomposition**,  
in *Proc. of the Workshop on Constraint Based Methods for Bioinformatics*, page 6, 2006
65. L. Bongini, R. Livi, A. Politi, and A. Torcini,  
**Exploring the energy landscape of model proteins: a metric criterion for the determination of dynamical connectivity**,  
*Physical Review E* 72, 051929 (2005).  
[selected for the December 1, 2005 issue of Virtual Journal of Biological Physics Research (<http://www.vjbio.org>)]
66. L. N. Mazzoni and L. Casetti,  
**Curvature of the energy landscape and folding of model proteins**,  
*Physical Review Letters* 97, 218104 (2006).  
[Featured in Physics News Update] [selected for the Virtual Journal of Biological Physics Research]
67. C. Guardiani and F. Bagnoli,  
**A toy model of polymer stretching**,  
*Journal of Chemical Physics* 125, 084908 (2006).
68. Dmitry Nerukh, George Karvounis, and Robert C. Glen,  
**Dynamic complexity of chaotic transitions in high-dimensional classical dynamics: Leu-Enkephalin folding**,  
accepted for publication in *Lecture Notes in Bioinformatics (LNBI)*

69. S. Potzsch, G. Scheuermann, M. T. Wolfinger, C. Flamm, P. F. Stadler,  
**Visualization of Lattice-Based Protein Folding Simulations**,  
*10th International Conference on Information Visualization* London, July 2006
70. Konstantin Klemm, Peter F. Stadler,  
**Statistics of cycles in large networks**,  
*Phys. Rev. E* 73, 025101(R) (2006)
71. Michael T. Wolfinger, Sebastian Will, Ivo L. Hofacker, Rolf Backofen, Peter F. Stadler,  
**Exploring the Lower Part of Discrete Polymer Model Energy Landscapes**,  
*Europhysics Letters* 74(4), 726-732 (2006)
72. Dmitry Nerukh, Anatoliy V. Luzanov,  
**Simple One-Electron Invariants of Molecular Chirality**,  
accepted for publication in *J. Math. Chem.*
73. Soto, P., Cladera, J., Mark, A.E. and Daura, X.  
**Stability of SIV gp32 Fusion Peptide Single Layer Prototibrils as Monitored by Molecular Dynamics Simulations**,  
*Angewandte Chemie* 17, 1089-1091 (2005).
74. Fan H, Mark A.E., Zhu J, Honig, B  
**Comparative study of generalized Born models: Protein dynamics**,  
*Proceedings of the National Academy of Sciences of the United States of America*, 102, 19, 6760-6764 (2005)
75. Marrink S.J., Risselada J, Mark A.E.  
**Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model**,  
*Chem and Phys of Lipids*, 135, 2, 223-244 (2005)
76. de Vries A.H., Yefimov S, Mark A.E., Marrink S.J.  
**Molecular structure of the lecithin ripple phase**,  
*Proceedings of the National Academy of Sciences of the United States of America*, 102, 15, 5392-5396 (2005)
77. Knecht, V., Mueller, M., Bonn, M., Marrink, S.J. and Mark, A.E.  
**Simulation studies of pore and domain formation in a phospholipid monolayer**,  
*J.Chem.Phys.* 122, 024704 1-9 (2005)

#### **Lectures and posters at the Conferences attended by the members of the Consortium**

1. Isabella Daidone, “About Protein Folding”, Attendance. Accademia dei Lincei, Rome, Italy, 9-10 June 2008.
2. Isabella Daidone, “7th workshop on molecular theories and simulations”, Invited speaker, Gaeta, Italy, 16-18 May 2008.
3. Christian H. Jensen, Dmitry Nerukh and Robert C. Glen, Statistical complexity of classical molecular dynamics systems: a measure of phase space exploration and non-Markovian behaviour of molecular conformations, Physics of Living Matter, Symposium II (Cambridge, UK, 2007)
4. Dmitry Nerukh and Makoto Taiji, Peptide's dynamics between structural transitions, Joint Conference of JMLG/EMLG Meeting 2007 and 30th Symposium on Solution Chemistry of Japan Molecular Approaches to Complex Liquids System (Fukuoka, Japan, 2007)
5. Christian H. Jensen, Dmitry Nerukh, and Robert C. Glen, Calculating Mean First Passage Times from Markov Models of Proteins, CompLife 2007 (Utrecht, the Netherlands, 2007)

6. Dmitry Nerukh, Maxim Fedorov, Christian Jensen, and Robert C Glen, Physical chemistry of aqueous solutions of macromolecules: the emergence of self-organised structures, *Modern Physical Chemistry for Advanced Materials* (Kharkov, Ukraine, 2007)
7. Dmitry Nerukh, Christian Jensen, and Vladimir Ryabov, Statistical complexity of classical molecular dynamics systems: a measure of phase space exploration and non-Markovian behaviour of molecular conformation, XIV Annual Seminar Nonlinear Penomena in Complex Systems (Minsk, Belorussia, 2007)
8. Dmitry Nerukh, George Karvounis, and Robert Glen, The dynamic complexity of biomolecules: quantifying emergent behaviour, *Computational Life Science* (Cambridge, UK, 2006)
9. Kei Moritsugu and Jeremy C. Smith, "Langevin Model of Protein Dynamics". Workshop on computer simulations of soft matter and biosystems, Heidelberg, March 2007.
10. Kei Moritsugu and Jeremy C. Smith, "REACH: Realistic Extension Algorithm via Covariance Hessian". Tennessee Structural Biology Symposium, USA, June 2007.
11. Eriksson, A., Information Theory and Multi-scale Simulations. Presentation at the University of Cambridge, department of Chemistry, Nov. 2006.
12. Eriksson, A., Dissipative particle dynamics for protein folding. EMBIO conference in Leipzig, May 16, 2007.
13. Eriksson, A., Dynamics of mesoscopic biomolecular systems. Bio Center seminar at Chalmers University of Technology, Oct. 5, 2007.
14. Görnerup, O., A Method for Inferring Hierarchical Dynamics in Stochastic Processes. ECLT workshop, March, 2008.
15. Görnerup, O. "Aggregation of variables in linear dynamical systems", 2008 Embio meeting, Venice 12-15 May 2008.
16. Görnerup, O. "A method for coarse graining Markov models with applications to biological systems", European Conference on Mathematical and Theoretical Biology, Edinburgh, 29 June - 4 July 2008.
17. Lindgren, K., Flows of information in chemical pattern formation, talk at *Modelling Complex Biological Systems in the Context of Genomics*, Evry, 3 May, 2007.
18. Lindgren, K., Quantifying spatial structure in dynamical systems — from microscopic reversibility to macroscopic irreversibility, talk at Max Planck Institute for Mathematics in the Sciences, Leipzig, Nov. 2007.
19. Lindgren, K., Self-replicating patterns: dynamics, energy, and information. Talk at *Joint Nordic Astrobiology and Swedish Astrobiology Network Meeting*, Stockholm, Nov. 6-7, 2007.
20. Lindgren, K., Self-replicating patterns and information dynamics. AlbaNova and Nordita Colloquium, Stockholm, Febr. 7, 2008.
21. Lindgren, K., A dynamical systems route to complexity, Lecture at ECLT Summer School *Blueprint for an artificial cell*, ECLT, 8 May, 2008.
22. Lindgren, K., Quantifying spatial structure in symbol sequences — identifying correlations and local information. Talk at *Statistical Semantics of Genomes: From Sequence to Function*, Evry, May 27-29, 2008.
23. Nilsson Jacobi, M. "Dynamical hierarchies", ECCS06, Paris 2006. (The talk was on

- hierarchies and reduction in continuous dynamical systems.)
24. Nilsson Jacobi, M. "Dynamical hierarchies, a Lie algebra approach", ECCS07, Cambridge, 2007. (The talk was on hierarchies and reduction in continuous dynamical systems using methods from the theory of Lie algebra.)
  25. Olof Gornerup, "Container growth and replicator dynamics in pre-biotic chemistry". European Conference on Complex Systems, Paris, France, November 2005.
  26. Olof Gornerup and Martin Nilsson Jacobi, Container Growth and Replicator Dynamics in Prebiotic Chemistry, Life Matters Conversations in Bionanotechnology, poster presentation at Chalmers, Goteborg, Sweden, October 2005.
  27. Kristian Lindgren, Flows of information as a driving force behind chemical pattern formation. Complex systems workshop organized by Cris Moore and Michel Morvain at Santa Fe Institute, July 2005.
  28. Kristian Lindgren and Anders Eriksson, Flows of information as a driving force behind chemical pattern formation. Life Matters Conversations in Bionanotechnology, poster presentation at Chalmers, Goteborg, Sweden, October 2005.
  29. Kristian Lindgren, Flows of information as a driving force behind chemical pattern formation.. European Conference on Complex Systems, Paris, France, November 2005.
  30. Kristian Lindgren, Flows of information in spatially extended systems. Mini Symposium on Dynamical Systems, Norwegian University of Science and Technology, Trondheim, 12 January 2006.
  31. Martin Nilsson Jacobi, Hierarchical dynamics. Complex systems workshop organized by Cris Moore and Michel Morvain at Santa Fe Institute, July 2005.
  32. Martin Nilsson Jacobi, Hierarchical dynamics. European Conference on Complex Systems, Paris, France, November 2005.
  33. Martin Nilsson Jacobi, Quasi-Species and Aggregate Dynamics. ECLT workshop: Evolution and Self-Assembly, March 2006.
  34. Martin Nilsson Jacobi, Aggregate and replicator dynamics in prebiotic systems, Artificial Life conference, Bloomington, Indiana, 2006.
  35. Anders Eriksson, Olof Gornerup, Martin Nilsson Jacobi, Johan Nystrom. Methods for simulating and analysing the dynamics of self-organising molecular systems. Presentation at the EMBO meeting in Vienna, 2006.
  36. Yuriy Chesnokov, Dmitry Nerukh, Robert Glen, Individually Adaptable Automatic QT Detector, Computers in Cardiology (Valencia, Spain, 2006)
  37. Alexander Nerukh, Nataliya Ruzhytska, Dmitry Nerukh, Hurst's Index and Complexity of Wave in Modulated Dielectric Medium, 8th International Conference on Transparent Optical Networks (ICTON 2006) (Nottingham, UK, 2006)
  38. A. Nerukh, N. Ruzhytska, and D. Nerukh, Quasy-intermittency of waves and their complexity in modulated dielectric medium, International Conference on Mathematical Methods in Electromagnetic Theory (MMET\*06) (Kharkov, Ukraine, 2006)
  39. Yuriy Chesnokov, Dmitry Neruh, Robert Glen, HRV analysis of patients prone to atrial fibrillation using a Neural Network approach, Measurement & Modelling of Autonomic Function (London, UK, 2006)
  40. Robert Glen and Dmitry Nerukh, Why do proteins fold?, The First European Conference On Chemistry For Life Sciences (Rimini, Italy, 2005)

41. F. Fedotov, N. Ruzhytska, D. Nerukh, Complexity of Electromagnetic Pulse Passing a Layer of Nonlinear Medium, 7th International Conference on Transparent Optical Networks (ICTON 2005) (Barcelona, Spain, 2005)
42. Dmitry Nerukh, George Karvounis, and Robert C Glen, Dynamic complexity of protein folding: computational mechanics on classical molecular trajectories, 2nd International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA 2005) and the 11th International Conference on Information Systems Analysis and Synthesis (ISAS 2005) (Orlando, USA, 2005) (invited lecture)
43. Dmitry Nerukh, George Karvounis, and Robert C Glen, Loss of chaoticity during beta-turn transition: dynamic complexity of protein folding, International Workshop on Theoretical and Computational Chemistry of Complex Systems (Hong Kong, China, 2005)
44. "Energy Landscapes and Dynamics of Biopolymers" (M.T. Wolfinger, W.A. Svrcek-Seiler, Ch. Flamm, I.L. Hofacker, P.F. Stadler) "Evolution of Biomolecular Structure", Vienna, May 25-27, 2006
45. "Do SS+RNA Viruses Depend on Host miRNAs: First Steps" (C. Thurner) "Evolution of Biomolecular Structure", Vienna, May 25-27, 2006
46. "A Graph-Based Artificial Chemistry Model" (G. Benkoe, Ch. Flamm, P.F. Stadler) "Evolution of Biomolecular Structure", Vienna, May 25-27, 2006
47. "Minicellsim - A Generics System for Studying the Evolution of Cellular Networks" (L. Endler, S. Widder, Ch. Flamm, S. Mueller, P. Schuster) "Evolution of Biomolecular Structure", Vienna, May 25-27, 2006
48. "SOSLIB: A tool for Multi-Scale Evolutionary Studies of Biological Reaction Networks" (R. Machne, Ch. Flamm) "Evolution of Biomolecular Structure", Vienna, May 25-27, 2006
49. Peter Schuster and Ivo Hofacker, "Computational approaches to functional and regulatory RNAs", Benasque, Spain, July 16-28, 2006
50. Kei Moritsugu and Jeremy C. Smith, "Langevin Model of Protein Dynamics", METHOD OF MOLECULAR SIMULATIONS 2005, Heidelberg (Germany), Jul 25-27, 2005
51. Kei Moritsugu and Jeremy C. Smith, "Langevin Model of Protein Dynamics", COMPUTER SIMULATION AND THEORY OF MACROMOLECULES 2005, Huenfeld (Germany), Apr 22-24, 2005.
52. Jeremy C. Smith and Frank Noe' and Isabella Daidone, "Molecular Dynamics for Biomolecular Mechanism", BIOMOLECULAR SIMULATION: FROM PHYSICAL PRINCIPLES TO BIOLOGICAL FUNCTION, Bad Honnef (Germany), May 22-24, 2006.
53. L. Bongini, L. Casetti, R. Livi, A. Politi, and A. Torcini, "Dynamical and topological properties of model proteins" IUPAB-EBSA International Biophysics Congress in Montpellier, France, August 27 September 1, 2005
54. C. Guardiani, F. Cecconi, and R. Livi, "Kinetic bottlenecks identification in different folding models" IUPAB-EBSA International Biophysics Congress in Montpellier, France, August 27 September 1, 2005
55. F. Sbrana, M. Vassalli, B. Tiribilli, L. Casetti, A. Torcini, A. Toscano, and A. Pacini, "Exploring the free energy landscape of a folded protein by means of AFM stretching experiments" IUPAB-EBSA International Biophysics Congress in Montpellier, France, August 27 September 1, 2005

56. Carlo Guardiani F. Cecconi and R. Livi "A molecular dynamics investigation of the kinetic
57. bottlenecks of the hPin1 WW domain" the 2006 WSEAS International Conference on Mathematical Biology and Ecology, Miami, Florida, USA, January 18-20, 2006.
58. Lapo Casetti, L. Mazzoni "Geometry of the energy landscape and folding of simple models of proteins" BIFI-2006 "From Physics to Biology: the interface between experiment and computation", organized by Institute for Biocomputation and Physics of Complex Systems, BIFI at the University of Zaragoza, 8-11 February 2006 in Zaragoza, Spain.
59. Carlo Guardiani "A study of folding of hPin1 WW domain through simplified protein models" International School of Physics "Enrico Fermi", course CLXV, organized by Societa Italiana di Fisica on "Protein folding and drug design", Varenna, Italy, July 4-14 2006
60. Matthias Kruspe poster at German conference on Bioinformatics Hamburg, October 5-7, 2005
61. P. F. Stadler, M. Middendorf lecture at Seminar on Theory of Evolutionary Algorithms Dagstuhl, February 5-10, 2006
62. K. Klemm, lecture at Annual meeting of the German Physical Society, Dresden, March 27-31, 2005
63. S. Janson, D. Merkle lecture at Second International Workshop on Hybrid Metaheuristics (HM 2005), Barcelona, August 29-30, 2005
64. P. F. Stadler lecture at Computational Approaches to Noncoding RNAs Benasque, July 16-28, 2006