Emergent organisation in complex biomolecular systems

EMBIO

Activity code addressed: *NEST-2003-Path-1*

Abstract

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

A composite approach will be 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 will be obtained from sophisticated all-atom simulations of realistic biomolecular systems and experimental mechanical stretching of giant single molecular proteins.

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

1. Project objective(s)

Introduction

Examples of emergence of complex behaviour in many-bodied systems that have inherently simple interactions span all areas of science and technology including molecular sciences (e.g., crystallisation), biology (e.g., ant colonies), meteorology and social interactions (e.g., traffic control). Although the phenomenon of emergent complexity has been recognized for many years, a general mathematical framework that can be used to describe such systems is still being actively developed and the analytical tools that are currently available are inadequate for describing many important systems. One of the fundamental difficulties in describing the behaviour of "complex" systems is that they are dynamic: non-trivial structures emerge only with time. In addition, in most cases complex systems are chaotic, or sensitive to fluctuations. The precise configuration of the final state is often poorly defined and is dependent on the precise configuration of the initial state.

To develop a general framework in which to describe complex systems there is a need to focus on problems involving well-defined systems - that is, systems which are well studied, in which the end points can be well defined and for which information is readily available. These well-defined systems are provided by life at the sub-cellular level. Sub-cellular life is dependent on the inherent ability of all bio-molecules such as DNA, RNA, lipids, sugars and proteins to reproducibly form sophisticated multi-scale structures without a centralised, governing "blueprint". Understanding these processes remains a major challenge and there is an urgent need for better tools to describe and predict how such emergent behaviour is encoded in bio-molecular systems.

A particularly important and striking example of emergent complex behaviour is protein folding. Most proteins spontaneously and reproducibly fold from an arbitrary initial configuration into a specific three-dimensional structure required for their biological function. Protein folding is one example of the way bio-molecular systems self-organize. Using modern biochemical techniques it is possible to elucidate the chemical composition (amino acid, nucleotide sequence, bio-polymer composition) of most bio-molecules. Moreover, the Protein Data Bank, contains the 3-dimensional structures of twenty-five thousand molecules whose structures have been determined by X-ray crystallography, Nuclear Magnetic Resonance (NMR) and other experimental techniques. NMR and other techniques are revealing the pathways and timescales of folding in ever more detail. Such work can also be combined with simulation studies to provide an atomic picture of the folding process.

Although significant effort has been expended processes such as protein folding are still poorly understood and their understanding remains one of the grand challenges in modern biology. It is currently not possible to predict or explain the mechanism by which the transition from an apparently random chain to a folded structure takes place. Neither are there tools to predict why this process, by means of 'emergent' complexity (structure), develops from the chaotic dynamic motions of individual atoms to a folded structure.

Part of the difficulty is that bio-molecular systems are high-dimensional non-linear systems. They involve many degrees of freedom and many factors influence the time evolution the system. Despite the apparent chaotic nature of the dynamics, self-organisation occurs and qualitatively new features spontaneously emerge. Thermodynamic considerations have lead to the proposal of a "folding funnel". There is the assumption that the free energy surface of bio-molecular systems must be arranged such that higher energy regions occupy large areas in phase space and low energy regions just a small region of phase space. As the energy decreases, the available phase space volume reduces and eventually leads to a single energy minimum corresponding to the native state of the molecule. The landscape of the funnel is assumed to be rugged, with many local minima and different pathways from the top to the native state giving rise to the complex dynamics of the systems. The existence of a "funnel" is generally accepted, and much effort has been devoted to the investigation of the energy landscape in general. However, the funnel landscape does not give a complete picture and how the domain (argument space) in this picture should be rationalised is uncertain. The whole phase space has little use in this context because it does not in itself resolve Levinthal's paradox: that the folding is much faster than a random search in phase space. Approaches that reduce dimensionality of the problem by extracting only the most important degrees of freedom have encountered problems and suggest that it is difficult to even identify a "reaction coordinate" for folding in a simple way.

The solution to this problem lies in an accurate description of the relevant dynamics: the intricacies of the system's trajectories in the whole-dimensional phase space produce the desired self-organisation. The reason for this is that the difficulty in understanding and explaining this kind of complex phenomena is ultimately connected to the 'crisis' of the traditional "analytic" approach, where the problem is divided into elementary "simples" followed by "synthesis" in an attempt to explain its features and functionality. As pointed out by Gunter [1], nonlinearity does not make this type of analysis possible when broken down into ultimate simples. In nonlinear systems the processes develop as a whole and the overall result cannot be derived from the isolated parts. Starting from Prigogine's work, the conceptual importance of dynamic form has been recognised [2]. Through the dynamic form order, information and complexity appear to be the result of a spontaneous, holistic process emerging from the chaotic dynamics of the system. Discovering the dynamic forms that lead to the emergence of native structures in bio-molecular systems is one of the most promising directions towards understanding complex systems. This overcomes the problems outlined above and provides a fundamentally new approach to explain complex phenomena. In addition to being biologically very important self-organizing bio-molecular systems thus provide a unique possibility to study emergent behaviour in complex systems in general. For multidimensional systems such as bio-molecules, understanding the dynamics is tightly connected to determination of the relevant energy surfaces.

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** (1)

We aim 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 will generate data on the dynamics of self organising molecular systems, perform complexity analysis using a diverse set of approaches and apply the resulting conclusions to practical problems from bio-chemo- and medical science. We also note that although the focus of the project will be on understanding complexity in the context of bio-molecular self-organization, methods and approaches will be general and will be 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 exists a large body of experimental and simulation data on which method development can be based.

The objectives

The goal of the project is 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 are:

- 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 graphical presentation of the workplan) the specific deliverables are expected to be produced by the end of the project. During the course of the project the work will be reviewed and the progress assessed at the planned annual workshops (see "Project management"). One of the objectives of the project is 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.

2. Participant list

1	University of Cambridge	UK
2	University of Groningen	The Netherlands
3	University of Florence	Italy
4	Chalmers University of Technology	Sweden
5	University of Vienna	Austria
6	University of Heidelberg	Germany
7	University of Leipzig	Germany
8	Friedrich-Schiller-University Jena	Germany

3. Relevance to the objectives of the specific programme and/or thematic priority

The project addresses all major objectives of NEST and, especially, the specific ones of the NEST-PATHFINDER initiative. Focusing directly on complexity itself the project does not fit within the existing Thematic Priorities of the Framework Programme.

The project is highly interdisciplinary. Having complexity as a main subject it involves research from mathematics, statistical physics, chemistry, information theory, biology, and computing. The methods from these fields used in the project are at the cutting edge and the ideas proposed are highly innovative. Despite the wide span of the methodologies involved the main subject is clearly defined and serves as a single focal point for all the approaches –**self organisation in molecular systems**. This enables an effective transfer of techniques between these diverse fields of science with a potential to even more remote areas such as social science, forest fires or brain activities.

One of the specific practical problems addressed by the project, protein folding, poses one of the greatest challenges for several decades. Advances along this route will have a significant long-term impact. By acknowledging the difficulties in solving this problem, the authors propose a fundamentally new point of view, namely, complexity analysis of the protein folding problem.

The objectives of the project and the majority of the techniques directly address the issue of emergence. We believe that this is the key notion that will provide a significant advance in understanding self-organisation and in particular, protein folding.

The consortium proposed consists of an interdisciplinary team of researchers from synthetic and computational chemistry, theoretical statistical physics, experimental physics, computing and information theory, large systems analysis, and computational geometry.

State-of-the art and proposed research

Current research in complexity and self-organisation spans many disciplines and sciences. The fundamental objectives for this project are to develop a robust underpinning in both mathematics and applications software that will allow the use of these new techniques to any complex phenomena. The utilisation of the biopolymer examples will allow realistic evaluation of these methods and suggest additional approaches to researchers in other areas of science and engineering. Because this is a relatively new area of research, significant advances in the underlying methodology will have widespread application.

We have outlined current interests of the collaboration group that directly contribute to the studies of complexity.

Dynamic complexity

At the Unilever Centre in Cambridge we have designed a fundamentally new approach, based on a complexity measure to reveal the underlying structure of the dynamics of the interparticle interactions in molecular systems that lead to self-organising, structuring behaviour. A number of different methodologies for the estimation of complexity are being developed. For our purposes we initially adopted the approach by Crutchfield *et. al.* termed “computational mechanics” [3]. This approach combines and implements the ideas from Shannon entropy and Kolmogorov-Chaitin algorithmic complexity theories.

Studying complexity is becoming an active field of research; nevertheless its application to real physical systems is still quite scarce. While Kolmogorov-Chaitin algorithmic complexity and Shannon entropy have been applied to physical systems, there are few applications of computational mechanics and its measure of complexity. However it is very advantageous to focus on this particular approach as it clearly provides better opportunities to understand complex behaviour. To the best of our knowledge there is no application of either algorithmic complexity or computational mechanics to the dynamics of real molecular systems, including protein folding.

Studying complexity in its numerous manifestations has recently become a very active subject. One of the reasons for this is the emergence of the necessary mathematical apparatus. There is a wide diversity of definitions of complexity (for the most recent reviews on the subject see, for example, [4,5,6,7]), which indicates, from one side, a substantial interest in the scientific community, and from the other, an active stage of development of this field when many points of view on the same subject concurrently exist. The Santa Fe Institute was founded in 1984 and specifically devoted to studying complexity. We believe that this application has the potential to stimulate and develop research in complexity theory and its wider applications in the EU.

The problem of complexity is being increasingly studied in physical [8,9], biological [10], chemical [11] sciences. There is a surprising absence of applications in molecular sciences. However, since molecular systems possess all the necessary features, they are an obvious candidate for this sort of investigations. There are publications proving the chaotic nature of molecular dynamics [12,13,14,15]. What we suggest here is to take this much further and consider the problem in much greater detail, not just to appreciate the chaotic dynamic nature, but to specify to what degree chaotic it is and what implications does it have to the emergence of qualitatively new structure and behaviour. We believe that our approach in applying computational mechanics and statistical complexity will lead to fundamental discoveries in the role of emergence in molecular structure.

We have demonstrated that this approach can be applied to molecular systems and provides additional novel information about their dynamics. The reorientation of water molecules in the liquid phase has been analysed focusing specifically on estimating complexity of the elementary molecular signals. The statistical complexity of water molecules at different locations of diluted electrolyte solution has been calculated. It has been found that the motion of water molecules in the bulk and close to the ion possess considerably different complexity [16]. At the same time we investigated another class of molecular system, a simple zwitterion which is made up of two opposite charged groups separated by an aliphatic chain. A distinctive feature of this system is that in vacuum it has a specific “loop”-like conformation that the molecule, if allowed to dynamically evolve, takes regardless of the initial configuration. The system, being simple, nevertheless demonstrates a “folding” behaviour and elements of self-organization. Again, the analysis of various molecular parameters provides evidences of different values of complexity in different dynamic

states of the system. Precisely at the moment of "folding" the complexity shows a considerable drop, and then rises to a higher level when the molecule stabilizes in the "folded" conformation.

Recently our research has focused on applying the approach to small peptides in explicit water. We have concentrated on one of the key elementary events in protein folding – β-turn formation. A decisively important role of the water network was found and attention is now being concentrated on the complexity of the water dynamics around the peptide before, at, and after the moment of turn formation. Analysis of various characteristics of the water dynamics support the hypothesis that a simplification of water dynamics takes place in the second solvation shell of the peptide at the moment of the turn formation. These results are being prepared for publication. What we would really like to find out is this: is the system driven by changes in the complexity of motions in the water i.e. is the driving force for folding changes in the complexity of water networks adjacent to the protein? Is this a common phenomenon of wide-ranging implication for the self organisation of biomolecular systems?

We plan to proceed along two general routes. The first one will include fundamental research on calculating complexity of full-dimensional Hamiltonian system with application to molecular systems. The second one will mainly consist of further, deeper elaboration of our present work: further developing the existing algorithm of complexity estimation of various molecular signals and its application to a realistic protein system. At the final stage of the project the results from these two routes will be combined in the investigation of dynamic complexity of realistic protein in explicit water in order to get an insight into overall picture from the complexity point of view.

The objectives for the first part are: develop the theory for estimating the complexity of a full 6N-dimensional trajectory of molecular system; implement this approach in numerical algorithms; test the theory on simpler molecular-like Hamiltonian systems; and, finally, apply it to a more realistic self-organizing molecular system.

Once the theory is formulated the numerical implementation should be devised. This will include modification of the standard molecular dynamics packages to incorporate the theory. We envisage the need for rigorous testing at this stage because our highly non-trivial approach will require the modification of the very basic steps in classical MD as well as time-consuming computer runs. We plan to incorporate the algorithm into a simple "textbook"-like MD program having clear manageable code with the application to simple liquids. We expect to obtain novel interesting results even at this early stage since this will be a fundamentally new description of the liquid state, to the best of our knowledge never investigated before.

The second part objectives include: investigating different characteristics of the water network (such as hydrogen bonding, orientational and translational diffusion, etc.) in terms of their dynamic complexity; elucidating the details of elementary folding events and contribution of various protein quantities into the complexity of overall dynamics; probing the change in complexity of the water network around the protein during the folding process.

We will then proceed with applying the complexity analysis to a realistic protein system and its surrounding water network. We expect to discover a significant quantity of valuable new information: our preliminary results strongly suggest existence of new unexpected features in these processes when considered from the complexity point of view.

In order to pursue this direction, much more data on the atomic level dynamics of proteins in water is needed. These data can be obtained from molecular dynamics simulations of real-size proteins in explicit water and within the proposed consortium this will be provided by the Groningen and Heidelberg groups.

Simulation of molecular systems

In Heidelberg work is being done on the interpretation of small-angle neutron scattering data on strongly-denatured proteins using random polymer theory. Initially this involved the application of a mesoscale model of a freely-jointed chain of spheres of excess scattering density. However, this was later refined to atomic detail modelling and it was shown that there is a change in the protein backbone angle distribution on folding of phosphoglycerate kinase in 4M guanidinium chloride solution. The importance of this work for the proposed consortium is that it provides an estimation of the width of the rim of the protein folding funnel, i.e. the distribution of configurations of a protein in the strongly-denatured state [17,18,19,20]. These results provide a starting point for simulations of the initial stages of protein folding, which will be performed with all-atom models and implicit solvent and using lattice models. The results of these simulations will be analysed using the hierarchical dynamical descriptions of the other contributing groups. A further application that we will pursue is the roughness of the energy surface at points determining rate-limiting factors in protein folding [21,22]. Calculations performed recently indicate that salt-bridge formation is of particular importance in this respect. We propose to continue this work so as to provide further input on the physical basis of folding funnel topologies. Configurations of the strongly-denatured protein will be used as input for protein –folding simulations by the Firenze group. Although present-day computer power is not sufficient for a complete simulation of protein folding at atomic detail, it is possible to reach the 100-nanosecond timescale and thus to obtain trajectories covering several decades of dynamical relaxation times. The resulting dynamics will be interpreted using, for example, the models of statistical complexity examined by the Cambridge group.

Also, protein dynamics in the native state is being investigated. The group is particularly interested in characterizing the internal motions in proteins and in relating the results to experimental and functional protein steps. In particular, we are interested in the protein glass transition, as manifested by a nonlinear increase in the atomic mean-square

displacements of a protein as a function of temperature. Initial work demonstrated that the motions involved can be described in terms of liquid-like dynamics of protein side-chains. How the dynamical transition is perceived by dynamic neutron scattering was recently investigated [23-27]. This glass transition appears to be driven by a similar transition in the surrounding aqueous solvent. Finally, we have performed a principal component analysis of the motions activated at the glass transition and shown that they can be described in terms of a very small number of global, collective motions. Again, models of statistical complexity developed by the Cambridge group will be used to interpret these results. In this way a description of the dynamics of functional, native proteins will be obtained. Subsequently we will apply the models to describing native proteins in different functional states e.g., with and without ligands bound, and for determining the properties of large-scale conformational transitions in folded proteins.

The focus of the group in Groningen (Partner 2) is on the application of molecular dynamics techniques to the understanding of macroscopic properties of biomolecular systems. In recent years we have concentrated on simulating the folding and assembly of peptides and proteins as well as simulating the self-organization of membrane components. In terms of the network we already have atomic level simulations of a wide range of systems that either self-assemble or undergo reversible folding. Partner 2 is interested in measures to identify when systems begin sampling new regions of phase space and how we might identify when a system is fully folded.

Partner 2 is investigating folding and assembly at multiple levels. This dates back to 1998 when they published the first simulations of reversible peptide folding in a realistic environment. This involved simulating the folding of a beta-heptapeptide peptide in methanol. For this system it is possible to simulate folding from an arbitrary configuration to the native conformation with experimental accuracy. This and a number of related systems are extremely well behaved and have been analyzed in detail. Extended folding trajectories are available for a range of different temperatures as well as for sequences which show different conformational preferences. This system is a good model system to compare different complexity measures as it is one of the very few peptide systems for which a large proportion of the accessible conformational space has been enumerated. Also far from being random, the denatured state of these peptides is best characterized in terms of a small number of well-defined states that rapidly interconvert. By extending the existing simulations it should be possible to for the first time obtain a complete description of the folding landscape for a peptide in atomic detail. Other systems that have been studied intensely by the Groningen group include a very small tetra-peptide. This system has been simulated for more than a micro-second in explicit solvent and shows very well defined transitions. Partner 2 also has extensive trajectories on the 20 a.a. Betanova and the EPO mimetic peptide. The EPO mimetic peptide forms simple dimers and may be an easy system to develop descriptors. In regard to larger systems, in conjunction with S. Berger and B. Koch (University of Leipzig) they are studying a set of peptides that have been engineered to undergo specific structural transitions in response to changes in the environment. These peptides are unstructured in isolation, associate at higher concentration into α-helical coiled-coils at neutral pH but under extreme pH (depending on the sequence) assemble into extended β-sheet like structures (fibrils). This system can be used to study both folding and aggregation, in particular to help understand the role of cooperatively in secondary and tertiary structure formation.

Lipid membranes are another archetypical self-organizing system critical to cell function which are well suited to application of complexity measures. Partner 2 has recently showed that under the appropriate conditions random mixtures of lipid in water will spontaneously assemble in simulations into micelles, equilibrated bilayers and even small vesicles. It is also possible to spontaneously assemble systems incorporating several different lipids and even peptides. Such simulations are also allowing Partner 2 to study collective processes such as phase transitions and the formation of lipid microdomains or “rafts” which are functional aggregates that result from microscopic phase separation and are thought to play a role in intracellular transport of lipid and proteins, as well as processes such as membrane fusion.

Dynamics of protein folding

The general issues of the dynamics of the protein during the folding process are being addressed by the Florence (Firenze) group. It involves both theoretical and experimental work.

On the theory side, the Firenze group has developed in the past years a dynamical approach to the problem of protein folding [28,29]. Within this approach, the dynamics of a simplified model of a protein, i.e., a chain of point like residues interacting via short- and long-range forces is studied *via* molecular-dynamics simulations.

A typical problem that cannot be understood within the static approach is the characterization of good and bad folders: it is known that only a limited number of amino acid sequences behave like a protein, i.e., converge systematically towards a unique native configuration regardless of the initial conditions, whereas other sequences remain in a coil-like state or end up in different compact conformations depending on the initial state. The former sequences are called "good folders"; the latter, "bad folders". Using the dynamical approach it has been possible to identify three different temperatures: the θ-temperature, analogous to the θ-transition temperature of homopolymers, the folding temperature, where the polymer spends a large fraction of the time in the native configuration, and the glassy temperature, below which the dynamics of the polymer becomes glass-like. It turns out that the ratio of the folding temperature to the glassy temperature of good folders is larger than the same ratio of bad folders [28].

Another aspect that has been investigated from the point of view of dynamics is the structure of the energy landscape of model proteins in 2d close to the native configuration. A connectivity graph has been constructed, showing the connections of the native state with its neighbouring minima in the energy landscape. Bad folders show a poor connectivity, i.e. there are few routes from the neighbouring minima to the native state, while in good folders there are many paths connecting "excited" minima with the native configuration. The knowledge of the connectivity graph also allows a study of the protein folding dynamics as a thermally activated process [29].

Within the present project we plan to attack the following issue: which are the main characteristic features of a free energy funnel (typical of a "good folder")? In particular, we will examine for the BPN model (introduced in [30]) how the energy landscapes of short sequences of length L changes upon increasing L. The landscapes will be characterized in terms of connectivity, saddle and minima distribution, degree of clustering, etc.

A final theoretical issue that we will pursue is the possibility - yet unexplored – to apply to the study of biological molecules a set of ideas and tools developed within the Firenze group which are of a geometrical and topological nature [32]. The geometrical and topological approach has been very successful when applied to the study of chaotic dynamics and is now being used to understand some aspects of phase transitions in terms of the topological structure of the energy landscape. These ideas will complement the analysis of the landscape carried out *via* the construction of the connectivity graph.

As far as the experimental side is concerned, the group in Firenze will study the mechanically induced folding and refolding of single proteins by means of an Atomic Force Microscope (AFM). This kind of experiment is particularly suited to reconstruct the folding-unfolding pathways as well as the free energy landscape of the examined protein [33-35].

The Firenze group is planning a series of experiments on the muscle protein titin. This giant protein has been engineered to obtain tetramers and octamers of Ig-like modules. The part of titin responsible for the passive elasticity of the muscles contains several tandem modules. The stretching of engineered proteins made of several repeats of the same module yields a saw-tooth pattern due to the sequential unfolding-refolding of modules. Our aim is to extract from such non-equilibrium measurements information about the equilibrium thermodynamic properties of the protein. Recently, in [36] has been suggested a possible experiment able to reproduce in detail the equilibrium free energy profile of the protein by considering a periodic loading of the AFM cantilever (instead of the usual linear motion of the cantilever [34]). We would like to verify the limits and capability of this new procedure.

These experiments will be complemented by theoretical and numerical approaches. In particular, the Firenze group plans to develop a kinetic model able to reproduce the experimental results. Molecular dynamics simulation of simple models [30,37] will then allow us to examine in detail the different experimental procedure (periodic loading versus linear ramp) proposed to reproduce equilibrium energy landscapes. Moreover, we would also like to investigate the limit of applicability of Jarzynski’s equality used to extract equilibrium results from non equilibrium measurements [35].

Summarizing, the specific objectives for the Centre for the Study of Complex Dynamics (CSDC), Firenze, Italy are as follows:

- It is planned to determine the main characteristic features characterizing the free energy landscape of a protein. The problem will be addressed by considering simple off-lattice coarse grained models able to capture the relevant aspects of protein folding. In particular it will be examined, for the well known BPN model and for related models, how the energy landscape of a protein builds up by considering sub-chains of the examined protein of increasing length. It will be characterized how the connectivity and the degree of clustering of the landscape as well as the distribution of saddles and minima changes upon increasing the number of amino acids in the sub-chains. Some geometrical and topological properties of the energy landscapes will also be measured.

- It is planned to examine simplified models of protein taking into account the polymer-solvent interaction. In particular, by using a model where the role of water molecules is considered (even if in a very simplified way) and that is capable of reproducing both "cold" and "warm" unfolding of a hydrophobic polymer, the dynamical time scales involved in the unfolding will be studied. Moreover, the two different unfolding transitions will be characterized and compared from the dynamical point of view; by simulating the dynamics of the above mentioned model, the time evolution of some configurational distance from the native state will be monitored during the unfolding of the chain. Using these data whether the unfolding dynamics, for the warm and cold transitions, can be characterized as activation or a diffusive process will be investigated.

- As a final point, it is planned to study experimentally the mechanically induced folding and refolding of single proteins by means of an Atomic Force Microscope (AFM). These experiments are particularly well-suited to reconstructing the folding-unfolding pathways as well as the free energy landscape of the examined protein. In particular the free energy profile associated with titin and elastin, by considering a periodic loading of the AFM cantilever (instead of the usual linear ramp) will be measured. These experiments will be complemented by theoretical and numerical studies. Molecular dynamics simulation of simple models will then allow us to examine in detail the different experimental procedures (periodic loading versus linear ramp) proposed to reproduce equilibrium energy

landscapes. Moreover, the limit of applicability of Jarzynski's equality used to extract equilibrium results from non equilibrium measurements will also be investigated.

Energy surfaces and folding kinetics

As has been pointed out earlier, the fundamental importance of the energy surface on which the systems dynamics evolves makes research in this direction an absolute prerequisite for the successful development of the complexity studies of molecular systems. The problem, being particularly complicated when applied to bio-molecules, is under investigation in three groups of the consortium: Leipzig, Jena and Vienna.

The theory of landscapes at present does not provide a comprehensive set of descriptors that would allow us to classify all landscapes and to decide which search or optimization algorithm to use based on this classification. While measures for ruggedness (correlation length, number of local optima, and distribution of basin sizes) and neutrality give useful hints, they are far from a sufficient description.

The Leipzig group will therefore explore the properties of numerical measures of other aspects of landscape structure. Anisotropies, i.e. the fact that the energy landscape may look very different for different classes of spatial conformations, will take centre-stage. It seems clear that there is not a single measure of anisotropy but most likely a suite of different aspects depending on the size of regions with different properties and on their embedding in the underlying configuration space. At present there are no mathematically sound definitions for "ridges", "valleys" or different shapes of "mountains" in discrete landscapes. As part of the project we propose to develop a local geometric theory of discrete functions for this purpose.

Landscapes can be studied from two distinct points of view: either one considers a particular landscape, e.g. the potential energy surface of the phenylalanine-tRNA of E.coli, or one is concerned with an ensemble of landscapes, as is usually the case in statistical physics approaches to optimization.

In the first case the task is to (efficiently) measure features of the landscape such as the number of local minima, the size distribution of the basins of attraction, the barrier tree, etc. A number of techniques for this purpose have already been developed for the special case of RNA secondary structure [46, 47]. Most of them depend crucially on the possibility to enumerate the near-ground state part of the energy landscape. The structure of the barrier tree itself can be used to extract characteristic quantifiers. A simple approach [48] will be extended to more sophisticated tree descriptors.

We propose here to develop similar tools that can be applied to lattice models of protein folding. In [38,39,40], the Jena group has designed a global optimization technique for HP-kind lattice models on the cubic and face-centred cubic lattices based on constraint programming approaches. The HP-model is a simplified protein model introduced by Ken Dill. It distinguishes between polar (P) and hydrophobic (H) amino acids, and searches for a conformation with a maximal packing of the hydrophobic (H) amino-acids. Using the constraint-based approach, we can successfully fold sequences up to length 300, thus greatly improving on the results of competing groups.

We want to use and extend this method to generate also the near-ground state part. One aim is to enlarge the sequence length which can be accessed using this method. But still we have to cope with an exponentially increasing number of conformations. For this purpose, we want to combine stochastic search methods with constraint-based approaches. This "constraint-based sampling" allows us a more directed (targeted) investigation of the landscape, thus improving both quality and efficiency of the sampling.

An information theoretic framework that does not assume an *a priori* random structure of a landscape is currently being developed (I.Erb & PF. Stadler, unpublished). The objective is to derive measures of ruggedness that are canonical from a Shannon-Entropy perspective rather than in terms of correlation measures. The relationship of such information-theoretic quantities to the Fourier decomposition of landscapes [49] will be studied in detail.

A second line of investigation aims at developing methods for measuring quantities such as basin sizes and saddle-point energies from sampling on the landscape rather than exhaustive enumeration. In addition to specialized MC simulation techniques, the applicability of population based search procedure for this purpose will be explored.

We intend to develop new swarm intelligence methods which differ from the conventional methods in that they do not search for a single optimal point in the landscape but instead search for a set of points that allow the characterization of the landscapes with respect to different measures. These methods use dynamic scenarios where the search for new points depends on the information that has been gained from the points found so far. We already started to extend the particle swarm optimization meta-heuristic so that the behaviour of the swarm depends on a neighbourhood relation that reflects local characteristics of the landscape around each particle and can be applied to dynamic optimization [42,43].

A relatively unexplored method of coordination between search agents in swarm optimization that we will use in our methods is communication by changing the environment (stigmergy). This method is attractive for the study of landscapes because the changes of the agents themselves can give hints on properties of the landscape. So far only ant colony optimization uses stigmergy and more deeper studies on the influences between ant behaviour, use of stigmergy, and landscape properties have just started [44]. Since we search for sets of points in the landscape, we will investigate multiple swarm methods [45].

Ensembles of landscapes can be studied in terms of probability spaces allowing us to consider generic properties of certain classes of landscapes. They include the "typical" number of metastable states, the typical distribution of saddle points heights and consequently the typical behaviour of trajectories of search procedures (or MD simulations). In the same case, one can derive simplified models. For example, the typical structure of neutral networks (sets of configurations with the same energy/fitness value in a degenerate landscape) can be explored in terms of a random subgraph model [50,51]. This approach will be extended to explore random directed acyclic graphs as models of energy landscapes that retain only the order relation of the fitness values.

A particularly interesting point is the question of which dynamical features of the search process are robust against a re-scaling of the energy function. Clearly such features are already determined by the directed acyclic graphs underlying the landscape.

Beside the stochastic methods described above, the Jena group wants also to employ global optimization techniques to analyze landscapes of lattice proteins (see e.g. [52]). Whereas stochastic methods can be applied to investigate many properties of landscape, there are special cases where this approach is not well suited. To give an example, the number of conformations having minimal energies (e.g. in the HP-model) cannot be well estimated from Monte-Carlo approaches, but have been determined using global optimization techniques [53].

We have designed a global optimization technique for HP-kind lattice models on the cubic and face-centred cubic lattices [40,39,38]. The method is based on constraint optimization and can successfully fold sequences up to length 300. In addition, we have applied the method to enumerate all minimum energy conformations for sequences up to length 48 [40].

In this project, we want to use and extend this technique to determine important properties of the landscape. The technique can be used to optimize parameters for the stochastic algorithms, since it provides the means for a dense sampling of low energy conformations. In addition, such a dense sampling allows us to compare different stochastic algorithms, as well as different lattices (for an extension to greater alphabets, see [41]). Furthermore, we want to extend our method to combine stochastic search methods with constraint-based approaches. This "constraint-based sampling" allows a more directed (targeted) investigation of the landscape, thus improving both quality and efficiency of the sampling.

Another bio-molecular system's energy landscape (and its kinetics of folding) is being investigated by the Vienna group. RNA secondary structures are listings of paired nucleotides that determine molecular structures in a kind of coarse-grained manner. They fulfil three criteria which are important for the understanding of biopolymer structure formation: (i) They are sufficiently simple in order to allow mathematical modelling and computer simulation, (ii) the conformation space is discrete by definition and thus appropriate for simple statistical sampling and information theoretical analysis, and (iii) the structures are well defined and allow for experimental testing of predictions.

A detailed analysis of RNA folding landscapes has become possible by means of an algorithm developed by Wuchty et al. [54], which allows us to generate the ground state and all suboptimal structures within a predefined energy range. A neighbourhood in the space of possible RNA structures can be defined in terms of a move set to inter-convert structures. A minimal move set comprises e.g. insertion and deletion of individual base pairs. Based on these elementary steps in conformation space we have developed a stochastic algorithm for the simulation of kinetic folding of RNA [55]. Furthermore, local minima, saddle points, and barriers between minima, can be extracted from the list of suboptimal structures. This leads to a hierarchic order of conformations that can be represented compactly in so-called barrier trees [56]. An extended Arrhenius-type kinetics can be formulated on the barrier tree and the calculated conformational kinetics compares favourably with the results of the stochastic simulations. Deviations can be interpreted by inspection of the details of the barrier trees, in particular through computation of the influence of multiple paths between metastable conformations [57]. This two step strategy consisting of (i) the construction of a barrier tree and (ii) modelling the reaction dynamics on the tree. This can be carried over to any kind of discrete landscape, in particular also to lattice proteins.

RNA secondary structures are proposed as a model of biopolymer folding in general. Lattice proteins, for example, are known to exhibit a number of features of RNA secondary structures. Kinetic folding will be considered as a process of information retrieval in the sense that information encoded in the nucleotide sequence is unfolded to yield a defined structure. Recently it has been shown [58], that it is possible to design RNA sequences with predefined structural and thermodynamic properties, e.g. two distinct local minima, which dominate the folding landscape. Extensions of our current investigations are planned: (i) Simulation of the folding of the growing RNA chain during transcription, (ii) cofolding of two RNA sequences into a common complex exhibiting intermolecular base pairs, (iii) studies on kinetic folding of sequences that are neutral with respect to structure in order to detect frequently used nucleation structures of the folding process(to search for and to identify folding modules that are formed independently of sequence details), and (iv) investigations of kinetic folding patterns of sequences, which are neighbours in sequence space and form different structures in order to be able to understand the influence of mutation on the folding behaviour of RNA molecules. The latter two extensions aim at the exploration of folding dynamics as a property in sequence space. Previous studies on the distribution of minimum free energy structures in sequence space revealed the existence of extended neutral networks as a generic property of common secondary structures [59] that has been directly verified

experimentally [60]. In the case of kinetic folding as a process in sequence space we aim for the detection of common dynamical patterns that show robustness against mutation. Eventually, a simplified folding kinetics can be detected that allows for classifications of families of sequences exhibiting certain kinds of characteristic patterns.

Dynamical hierarchies and information theory

The Chalmers group contributes with two sub-projects to EMBIO: (1) Dynamical Hierarchies and (2) Information dynamics. The objective is to develop both of these areas and to unify them into a theoretical framework that can be applied to problems in molecular dynamics, in particular to folding kinetics.

1. Dynamical Hierarchies.

The relaxation process whereby macromolecules (such as proteins) are folded into three dimensional structures display highly complex dynamics. Traditional techniques from molecular dynamics, where the equations of motions for each molecule in the macromolecule and in the solvent are integrated explicitly, are computationally too intensive for many practical applications. The folding process has also been shown to be sensitive to small perturbations in the dynamics and it has been difficult to systematically derive reduced models with maintained predictive power. In this field there is a clear need for novel model reduction techniques.

Analysis of nonlinear dynamic systems is a challenging task and it undoubtedly remains an area of active research effort. Among the most notable research endeavours and objectives in nonlinear dynamic analysis is the existence of invariant manifolds and the associated problem of finding (computing) them. The problem under consideration is of great importance and it has been traditionally motivated by efforts to develop systematic methods for simplifying the analysis of nonlinear systems by effectively reducing their dimensionality [61]. In particular, the study of invariant manifolds has been conducted in connection with the problem of stable, unstable and centre manifolds for nonlinear dynamical systems, their long-term asymptotic behaviour and their associated stability characteristics, as well as bifurcation analysis. Slow and fast degrees of freedom, corresponding to the unstable respective stable invariant manifolds, are also a cornerstone in modern nonequilibrium statistical mechanics [62].

In a recent study [63] we showed how invariant manifolds relate directly to the hierarchical organization in dynamic systems, i.e., systems where dynamics occur on multiple time- and length- scales simultaneously. Higher (more course grained) levels of descriptions are defined through projective maps of the degrees of freedom (the configuration space) from lower levels, as well as an induced map of the dynamics. Whether or not a projective map constitutes a new level of description can be determined by studying the induced dynamics on the higher level. To clarify the concept, we state the following requirements for a dynamical hierarchy:

- Each level in the hierarchy should be deterministic when described in isolation. At each level the degrees of freedom should carry sufficient information to determine the time evolution on that level.
- A higher level in the hierarchy should be derived from a lower through a smooth projective map i.e. there should be upward causality in the hierarchy.
- Arbitrary nonlinear projective maps should be allowed, and thereby allow for highly heterogeneous (or “functional”) course graining. Note that this implies that the levels in the hierarchy need not have any relation to specific dimension carrying parameters, such as length scale. This allows for levels to be functionally defined with respect to the dynamics rather than physical properties of the system.

In a recent paper we showed that a necessary and sufficient condition for a projective map to describe a transition between description levels is that the kernel from the differential of the map is tangent to an invariant manifold with respect to the flow. The implications of this are explored in detail.

Practical implementation in relation to model reduction in high dimensional complex systems is currently the work in progress. The main focus has been on simplifying the dynamics in many-particle systems, such as molecular dynamics models of molecular self-assembly processes. In the current proposal we intend to apply this approach to protein folding. We expect that the complexity of large heterogeneous molecular systems can be analyzed and greatly reduced by identifying a hierarchical organization of the interrelated description levels.

The ideas behind the dynamical hierarchies framework is related to computation mechanics (see discussion above on "Dynamic complexity"), even though the techniques used differ significantly. In computational mechanics one seeks a dynamical description of a system that maximizes the mutual information between present and future states. In dynamical hierarchies the vector field that generates the dynamics is analyzed using methods from differential geometry. Understanding the exact correspondence between these two approaches will be an important part of the project. The "Information Dynamics" concept discussed in the next section provides a starting point for bridging the two frameworks into a unified model reduction technique.

2. Information Dynamics

We have also developed an information-theoretic formalism that at present can be applied to the macroscopic description (in terms of concentrations) of chemical systems [64]. This formalism is based on a double decomposition

of information in terms of length scale as well as position [65,66]. This makes it possible to detect structure at different (macroscopic) length scales. Through the information-theoretic formulation of statistical mechanics, the information content in structure can be related to thermodynamic properties and constraints of the system studied. The in-flow of free energy driving the system is directly related to an information flow that can be aggregated into spatial structure in the form of concentration variations. At the finest length scales, there is an information flow leaving the system, corresponding to the entropy being produced as information disappears into microscopic degrees of freedom due to diffusion and chemical reactions.

The challenge would be to extend the information-theoretic analysis so that it extends below the macroscopic scales to capture the structure and dynamics in microscopic configurations. We shall also extend the formalism in order to capture the information flow in the opposite direction; the dynamical systems entropy, characterizing an information transfer from microscopic to macroscopic length scales. The connection between microscopic and macroscopic structure has been analysed before [67], and at present we continue to develop these ideas in collaboration with Torbjørn Helvik (Trondheim University).

4. Potential Impact

The impact of the work will be in two main areas. First of all, the fundamental nature of the project will contribute to recent, very active research of complex systems. Practical impact of the latter spans from dealing with forest fires to brain activities. It will definitely have a big impact in the field of protein folding – a very long-standing, hard problem – by providing an alternative view on the problem. The benefits will go to the large community of molecular and structural biologists and in the pharmaceutical industry, medicinal chemists and protein engineers who are involved in research of proteins and other biomolecules, in particular, their folding and misfolding. Besides a direct impact connected with revealing the details of protein folding, a significant body of data is expected to be collected and analysed e.g. the role of water, dynamic evolution of particular biomolecules in explicit water (this valuable data takes much computer power to obtain).

Second, through developing the algorithms of complexity estimation and adapting these to multidimensional continuous signals, a substantial impact is expected in the areas where complicated, chaotic dynamics is the main feature of the studied subject. Examples of the latter can be periodic heart rhythms or impulse-shaped electromagnetic signals. Heart rhythms are known to have a chaotic nature and subtle variations in (still largely unknown) dynamic rules indicate dramatic changes in the health of the patient. Thus, discovering those rules, or internal structure of heart beats (ECG, blood pressure, etc.) through the mechanisms of complexity that we can study would lead to qualitatively new knowledge and potentially, new diagnosis tools. As another example, transformation of an electromagnetic signal propagating in photonic devices leads to extremely complicated optical signals and estimation of their complexity brings important knowledge in understanding this process. This is especially important because these devices are often used for storing and transmitting information – the direct subject of complexity measures.

The strategic benefits of the proposed research will be manifested in drug design and biotechnology. Despite the genomics revolution there has been a decrease in the rate of discovery of new drugs in the last few years. This indicates that new innovation is required. This new innovation will proceed partly through the structure-based understanding of protein folding and binding. This in turn will require detailed computer simulations which will have to be interpreted using the various descriptions of complexity that will be investigated in this proposal.

The proposed research will significantly deepen our understanding of function and evolution of biopolymers. The kinetics of complex formation is of direct technological relevance, since hybridization reactions are the basis of technologies like DNA-chips and regulation of gene expression through small interfering RNAs (siRNA). Likewise, the ability to rationally design molecules with desired structural and dynamical properties may be exploited in a variety of settings.

If properly understood, the spontaneous emergence of coarse grained descriptive variables could for example be used for efficient simulation of complex dynamical systems. More generally, identifying generic features in heterogeneous systems with multi-scale dynamics is of central importance to the complex systems community. Protein folding is one of the most striking examples of problems exhibiting these characteristics, and is therefore well suited to serve as a testbed for novel computational techniques. Progress in this area can be expected to have wide applicability in complex systems and elsewhere.

The project will put specific efforts into disseminating the results and spreading awareness in the research community and beyond. Particularly, besides the publications in the specialized journals it is expected to collect the methodologies, algorithms and software developed in the course of the project that quantify the complexity measures and that is of interest to wide community of researchers. It is also planned, as a result of the annual workshops, to compile the documents that summarise the progress in estimating the complexity of dynamic systems and make them available online at the project's webpage (see WP9).

5. Participants and consortium

The consortium consists of 8 partners, each representing a research group in an EU national University. The research areas of the partners share two main themes: complexity analysis and molecular self-organisation. This provides a unified and complementary expertise for achieving the project's goals.

Since the project is mostly theoretical-computational, the foreseen resources will be used for employing the research personal to implement the project. Each partner plans to employ a postdoctoral researcher and train a PhD student. The other category of the expenditure will be travel of partners working within a common workpackage. This will facilitate an effective exchange of results and discussions during the implementation of the project. It will also culminate with a common workshop that is expected to result in the summary of the complexity measures and emergent characteristics of molecular systems developed within the project framework. In addition, Partner 3 has allocated resources for purchasing the controller for Atomic Force Microscope.

Participant 1

University of Cambridge
Chemistry Department, Lensfield Road, Cambridge, CB2 1EW, UK
Principal Investigator: Prof. Robert C. Glen

Qualifications, expertise and intellectual resources

The Unilever Centre was established in 2001 and has grown to nearly forty personnel studying a diverse range of science in molecular informatics. We have been involved in developing the fundamental theory of complexity applied to molecular systems for two years. The research has attracted strong interest and was recently abstracted in a review journal as a significant novel approach. Members of the group have considerable experience in simulation and in the interpretation of computational and experimental data. The implementation of various techniques that we have developed through the use of numerical methods and applied math in end-user computer software is an important part of the groups work e.g. the Gold and GASP computer programs widely used in the Pharmaceutical industry, Chemical Markup Language, the Bleep protein/ligand scoring function, aspects of Macromodel.

Role and main tasks

The role of Cambridge group will be in studying the dynamic complexity of molecular systems. The latter will be either simplified models (Participant 3), or the dynamic characteristics obtained from molecular dynamic simulations (Participants 2 and 6). The results of complexity analysis will be directly compared to the dynamic hierarchies and information flows in molecular systems (Participant 4). Also, the Cambridge group will be the coordinator of the Consortium with all implied coordination and organisational responsibilities.

Personnel

Professor Robert Glen

Prof. Robert Glen is Director of the new Unilever Centre for Molecular Informatics, University of Cambridge, a $20m addition to the Chemistry department, which he has expanded to forty people in the last three years. Funding has recently been extended by $5.2M for a further five years. He obtained his Ph.D. in X-ray Crystallography and organic synthesis from the University of Stirling. One of the highlights was the first co-crystallisation of a reactant and product of a chemical reaction in a single crystal. At the Wellcome Foundation he created one of the first computer-aided molecular design groups which included protein crystallography, computational chemistry, molecular transport properties and Electrochemistry. He co-developed the GASP and GOLD computer programs which are used extensively in the pharmaceutical industry, is a co-inventor of Zomig (AstraZeneca) for migraine and invented two other compounds that have entered Phase-2 development. He then became Vice President for Collaborative Research at Tripos Inc., assisted in setting up three biotechnology companies, obtained academic research grants of $3.8M and directed collaborative and contract research in drug discovery. Two of the research programs have compounds in early stage studies in Schizophrenia and Artherosclerosis. He has published over 80 papers and has numerous patents. He is on the SAB of a number of institutes, biotechnology and pharmaceutical companies, is a fellow of the RSC, an honorary member of the AACR, serves on the Netherlands genomics and bioinformatics initiatives and is an editor of the Encyclopaedia of Computational Chemistry, the Journal of Molecular Structure and the ACS JCICS and on the RSC publications board.

Selected Publications:
1) A. Bender, H. Y. Mussa and R. C. Glen. Molecular Similarity Searching using Atom Environments, Information-Based Feature Selection and a Naïve Bayesian Classifier. *J. Chem. Inf. Comput. Sci.*, **44**(1); 170-178 (2004).
2) R. C. Glen and S. Xing L.; Glen R. C. and Clark, R. D. Predicting pKa by Molecular Tree Structured Fingerprints and PLS. *J. Chem. Inf. Comput. Sci.*, **43**(3), 870 (2003).
3) R. C. Glen and S. C. Allen. Ligand-protein docking: cancer research at the interface between biology and chemistry. *Current Medicinal Chemistry*, **10** (9), 759-774 (2003).

4) R. C. Glen, S. Aldridge. Developing tools and standards in molecular informatics. *Chem. Comm.*, 2745-2747 (2002).
5) R. C. Glen and Li Xing. Novel methods for the prediction of pKa, logP and logD. *J. Chem. Inf. Comput. Sci*., **42**(4); 796-805 (2002).
6) P. Murray-Rust, R. C. Glen, Y. Zhang and J. Harter. The World Wide Molecular Matrix - a peer-to-peer XML repository for molecules and properties, pages 163-164. "EuroWeb2002, The Web and the GRID: from e-science to e-business", Editors: B. Matthews, B. Hopgood, M. Wilson, The British Computer Society (2002).

Dr Dmitry Nerukh is Research Associate at the Unilever Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, UK. He was: Senior member of Wolfson College; 2002-2004, Research Associate at the University of Nevada, Reno, USA; 1998 – 2000, Royal Society - NATO Postdoctoral Fellowship at the University of Leeds, School of Chemistry; 1997 – 1998, and Research assistant at the Kharkov State University, 1996 – 1997. He received his PhD from the Kharkov State University in 1996.

Selected Publications:

1) D. Nerukh and T. R. Griffiths, Real and imaginary parts of the vibrational relaxation of acetonitrile in its electrolyte solutions: new results for the dynamics of solvent molecules *Journal of Molecular Liquids*, **109**(1/2), (2003).
2) D. Nerukh, G. Karvounis, and R. Glen, Complexity of classical dynamics of molecular systems. Part I: methodology and Part II: finite statistical complexity of a water-Na+ system *J. Chem. Phys*., ***117***(21), 9611-9617 and 9618-9622 (2002). (reproduced in Virtual Journal of Biological *Physics Research*, **4**(10), (2002).
3) D. Nerukh and T. R. Griffiths, Complex vibrational correlation functions extracted from the resolved 2 band of liquid acetonitrile *Phys. Chem. Chem. Phys*., ***3***(10), 1799 - 1805 (2001).
4) D. Nerukh and J. H. Frederick, Multidimensional quantum dynamics with trajectories: a novel numerical implementation of Bohmian mechanics. *Chem. Phys. Lett*., **332**(1-2), 145-153 (2000).
5) O. N. Kalugin, D. A. Nerukh, I. N. Vyunnik, E. G. Otlejkina, Y. N. Surov, and N. S. Pivnenko, IR and NMR studies of hydrogen bonding in hexan-1-ol-tetrabutilammonium iodide solutions in the temperature range 28-145°C and in tetrachloromethane. *J. Chem. Soc. Faraday Trans*., ***90***(2), 297-303 (1994).

Participant 2

Rijksuniversiteit Groningen
Postbus 72, 9700 AB Groningen, The Netherlands
Principal Investigator: Prof. Alan E. Mark

Qualifications, expertise and intellectual resources

The MD Group in Groningen concentrates on dynamical simulation of biopolymers and lipid aggregates. Their aim is to understand and predict macroscopic behaviour of complex biomolecular systems on the basis of interactions between atoms using different levels of approach. The group is associated with the development of two internationally recognized molecular dynamics simulation packages Gromos and Gromacs which have hundreds of licensed users worldwide. The group has extensive experience in the simulation and analysis of peptide/ protein folding and aggregation as well as the analysis of the dynamic and thermodynamic properties of biological macromolecules in general. The group has very extensive computational facilities in house as well as access to University and National facilities.

Role and main tasks

The role of the Groningen MD group (partner 2) will be to provide an atomic level description of the folding of small peptides and aggregation of lipids into various phases. We will provide a complete description of the folding landscape of a number of systems which will form that basis to investigate the effectiveness of various measure of complexity. This work on elucidating the complete landscape for simple systems will be complementary to the simulations of larger systems conducted by partner 6 (Heidelberg). In initial phase of the project we will provide access to our pre-existing simulations which give a partial description of the folding landscape to members of the network. During the project we will attempt to generate a complete landscape for a number of small systems in atomic detail which will involve extending our current ns simulations into the microsecond range.

Personnel

Prof. Alan E. Mark heads the molecular dynamics group in Groningen. He is the author of over 80 scientific articles is peer reviewed journals and co-author of the GROMOS96 simulation package. He is also coordinator of an EC TMR network Protein folding misfolding, aggregation and disease.

Dr. Siewert-Jan Marrink (KNAW fellow) has extensive experience in the atomic simulation of lipid aggregates. In particular formed some of the first simulation of spontaneous bilayer and vesicle formation in atomic detail.

Dr. Xavier Periole (postdoc) has experience in simulation of peptide folding.

Dr. Ruud Scheek (senior lecturer) expert in the analysis of peptide and protein interactions by NMR.

Prof. Herman Berendsen (emeritus professor at the University of Groningen) will be a scientific advisor to the project. He was professor of physical chemistry from 1963 to 1999 and is (co)author of 260 papers, mainly on methodological aspects of molecular simulation. He has taught courses on information theory for 20 years.

Selected publications of the group:
1) S. J. Marrink and A. E. Mark. Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *J. Am. Chem. Soc.*, ***125,*** 15233-15242 (2003)
2) P. Soto and A. E. Mark. The effect of the neglect of electronic polarization in peptide folding simulations. *J. Phys. Chem. B*, ***106***, 12830-12833 (2002).
3) X. Daura, W. F. van Gunsteren, and A. E. Mark. Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations. *Proteins*, ***3,:*** 269-280 (1999).
4) W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A*, ***103***, 3596-3607 (1999).
5) X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark. Reversible Peptide folding in solution by molecular dynamics simulation. *J. Mol. Biol.***, *28***, 925-932 (1998).

Participant 3

Centro interdipartimentale per lo Studio delle Dinamiche Complesse (CSDC) FIRENZE
Dipartimento di Fisica, Universita' di Firenze, Via G. Sansone, 1, 50019 Sesto Fiorentino (FI), Italy
Principal Investigator: Dr Lapo Casetti

Qualifications, expertise and intellectual resources

The Interdepartmental Centre for the Study of Complex Dynamics (CSDC) is an organization coordinating interdisciplinary research in the field of complex systems with special reference to non-linear systems with many degrees of freedom. In the last years the research activity was focused mainly on biological physics with studies concerning DNA and heteropolymer dynamics. The folding dynamics of short heteropolymers has been studied by considering simplified but not trivial off-lattice and lattice models. In particular, statistical and dynamical methods have been applied to the characterization of good and bad folders. Moreover, models for the interaction of a polymer with the solvent, able to describe both the warm and the cold unfolding transitions, were introduced. Recently the theoretical activity of the group has focused also on the description of the properties of energy landscapes of physical systems, in the context of protein folding as well as in the more general perspective of a geometrical-topological description of collective phenomena like phase transitions.

The experimental activity of the CSDC is currently focused on biomolecule stretching, in close collaboration with the National Institute of Applied Optics (INOA), where is located a laboratory equipped with advanced microscopy facilities including a scanning probe microscope and a pulsed force SPM add-on for elasticity measurements. Experiments are performed by researchers affiliated both to INOA and CSDC.

Role and main tasks

Elucidation of the general statistical mechanics properties of model biopolymers in terms of the energy landscapes and folding abilities is the role of this participant that contributes and links together two main activities of the project: the energy landscapes and dynamics of molecular systems. The collaboration with Participants 5, 6, and 4 is planned in this framework. Also, the participants will perform the experimental investigation of the giant protein titin thus providing a direct link with the detailed simulation of protein in water (Participant 2).

Personnel

Dr. Lapo Casetti, born in 1967, is assistant professor of theoretical physics at the University of Florence, Italy. He received his PhD in Physics from the Scuola Normale Superiore in Pisa, Italy, in 1997. His research interests cover basic issues as well as applications of statistical physics: classical and quantum complex systems, Hamiltonian dynamics, numerical simulation, geometric and topologic approaches to dynamics and statistical mechanics, the theory of phase transitions, the modelling of biomolecules, and protein folding. He is author of more than 30 scientific publications and has organized an international workshop and two schools. He will act as coordinator of the CSDC Firenze group.

Prof. Roberto Livi, born in 1953, professor of Statistical Mechanics at the University of Florence, Italy. He received his degree in Physics in 1977. His research interests are: statistical mechanics, dynamical systems, complex systems and biophysics. He is author of more then hundred publications in the fields of statistical mechanics, statistical field theory, dynamical, complex and quantum systems. He managed several international and national research projects, and organized several international congresses, workshops and schools.

Dr. Antonio Politi. Born in Firenze in 1955, he is currently director of research at the National Institute of Applied Optics (INOA), where he is Head of the Quantum Optics Division; he is also associate editor of Physical Review E, as well as scientific advisor of several other journals. His current research interests cover nonlinear dynamics, fractals,

nonequilibrium statistical mechanics, transport phenomena, application of information theory to DNA sequences, proteing folding. He has organized various international conferences and Workshop, and is author of more than 110 papers on international journals.

Dr. Alessandro Torcini, born in 1961, is researcher at the National Institute of Applied Optics (INOA). He obtained the PhD degree in physics in 1994. He has actively worked in nonlinear dynamics, theory and molecular dynamics simulations of simple and complex liquids, phase transitions in systems out of equilibrium, and simulations of protein off-lattice models. He has published more than 60 papers on international peer-reviewed journals. He has organized several local workshops and 4 international conferences.

Dr. Massimo Vassalli holds a permanent position at the National Institute of Applied Optics, in the context of the interdisciplinary Biophotonics laboratory. He received a degree in physics in 1999 and recently completed his PhD on non-linear dynamics and complex systems. His core skills are on Atomic Force Microscopy (AFM), and, in particular, he is expert in force spectroscopy experiments carried out on biological microsystems. Being also interested in technological applications, he collaborates with industrial partners in technology transfer programs. Due to his experience, he will be the tutor of the early-stage researcher working with the AFM.

Also Dr. Marco Pettini, Prof. Stefano Ruffo and Dr. Bruno Tiribilli will participate in the project

Selected publications of the group:
1) A. Torcini, R. Livi, and A. Politi. A dynamical approach to protein folding. *Journal of Biological Physics,* **27,** 181 (2001).
2) L. Bongini, R. Livi, A. Politi, and A. Torcini. Thermally activated processes in polymer dynamics. *Physical Review E,* **68**, 061111 (2003).
3) L. Casetti, M. Pettini, and E. G. D. Cohen. Geometric approach to Hamiltonian dynamics and statistical mechanics. *Physics Reports,* **337**, 237 (2000).
4) P. Bruscolini and L. Casetti. Modeling hydration water and its role in polymer folding. *Journal of Biological Physics,* **27**, 243 (2001).
5) L.Formigli, E.Meacci, M. Vassalli, D. Nosi, F.Quercioli, B. Tiribilli, A. Tani, R. Squecco, F. Francini, P. Bruni, and S. Zecchi-Orlandini. Sphingosine 1-Phosphate induces cell contraction via Ca2+-independent/Rho-dependent pathway in undifferentiated skeletal muscle cells. *Journal of Cell Physiology,* **198,** 1 (2004).

Participant 4

Chalmers/Goteborg University
Department of Physical Resource Theory, School of Physics and Engineering Physics, 41296 Goteborg, Sweden
Principal Investigator: Prof Kristian Lindgren

Qualifications, expertise and intellectual resources

Complex systems has been part of the Department of Physical Resource Theory at Chalmers and Göteborg University since the mid 1980s. The research started with a basis in information theory and statistical mechanics, first applied to self-organising systems and cellular automata. The current projects in the field includes: (1) Evolutionary models and game theory, (2) Information theory, complexity measures, and cellular automata, (3) Artificial Life (EU funded integrated project PACE), (4) Geographic networks and urban growth, (5) Agent-based modelling in environmental science.

The group initiated the masters programme in Complex Adaptive Systems in 2000 and the PhD education in Complex Systems in 1999. Annually 25 students enter the masters programme, and currently 5 students are enrolled in the PhD education.

Selected publications of the group:
1) K. Lindgren, A. Eriksson and K.-E. Eriksson, "Flows of information in spatially extended chemical dynamics," to be published in Proceedings for Artificial Life 9 (Boston conference, September, 2004).
2) C. Andersson, A. Hellervik, K. Lindgren, A. Hagsson, and J. Tornberg, "Urban economy as a scale-free network," *Physical Review* **E 68** 036124 (2003).
3) C. Andersson, K. Lindgren, S. Rasmussen, and R. White, "Urban growth simulation from 'first principles'," *Physical Review* **E 66** 026204 (2002).
4) K. Lindgren, C. Moore and M. G. Nordahl, "Complexity of two-dimensional patterns", *Journal of Statistical Physics* **91**, 909-951 (1998).
5) K. Lindgren and M. G. Nordahl, "Evolutionary dynamics of spatial games", *Physica* **D 75**, 292-309 (1994).
6) K. Lindgren, "Evolutionary phenomena in simple dynamics", pp. 295-312 in Artificial Life II, C. Langton et al (eds.), (Addison-Wesley, Redwood City, 1992).
7) K. Lindgren and M. G. Nordahl, "Complexity measures and cellular automata", *Complex Systems* **2**, 409-440 (1988).
8) K. Lindgren, "Microscopic and macroscopic entropy", *Physical Review* **A38**, 4794-4798 (1988).
9) K.-E. Eriksson, K. Lindgren, and B. Å. Månsson, Structure, Context, Complexity, and Organization (World Scientific, Singapore, 1987).

10) K. Lindgren, "Correlations and random information in cellular automata", *Complex Systems* **1**, 529-543 (1987).
11) K.-E. Eriksson and K. Lindgren, "Structural information in self-organizing systems," *Physica Scripta* **35**, 388-397 (1987).

Role and main tasks

The role of the Chalmers group will be to develop a unified theoretical framework for dynamical hierarchies and information dynamics (WP5 and WP6). The methodology will then be applied to folding kinetics in collaboration with other partners. We shall also contribute with our experience in information theory and computation theory to develop quantitative measure of complexity relevant for folding kinetics.

Personnel

Kristian Lindgren, born 1960, is professor in Physical Resource Theory at Chalmers, doing research in both complex systems and environmental science. He initiated the international masters programme in Complex Adaptive Systems at Chalmers for which he has been director since 2000, and he is the examiner for the PhD education in Complex Systems. In complex systems, Lindgren has done research in the areas of dynamical systems (non-linear dynamical systems, cellular automata, self-organising systems), information theory (applied to the analysis of complex systems and its connections to statistical mechanics and thermodynamics), evolutionary modelling and game theory (mainly dealing with the evolution of cooperation), urban growth and economic networks, and computation theory (universal computation in cellular automata and formal language theory for two-dimensional patterns). Since April 1st, Lindgren is participating in the European FET integrated project PACE (Programmable Artificial Cell Evolution).

Martin Nilsson has a background in theoretical physics and complex systems. He did his undergraduate and graduate studies at Chalmers University, Gothenburg Sweden. His PhD thesis was in theoretical biology, with the title "Mathematical Models of Evolution". After the dissertation he spent three years as a post doctoral fellow at the Los Alamos National laboratory, working on self-organizing systems. The main project involved theoretical and computational studies of self-assembling molecular systems. Nilsson is currently pursuing his second post doc at the Nordic Institute of Theoretical Physics (Nordita) in Copenhagen, where he is working in the biophysics group. In September he will join with Lindgren's group in Göteborg as an assistant professor. An overarching goal with Nilssons's research is to understand the interplay between self-organization and natural selection in biological systems, especially in relation to the transition from nonliving to living matter.

Anders Eriksson is a PhD student in Complex Systems who will present his thesis on evolutionary models in September 2004. He has done work on game theory and evolution as well as on population genetics.

Jon Klein is a PhD student in Complex Systems, currently working at Hampshire College, Massachusetts, USA. He is the author behind the simulation platform for artificial life, modelling 3d physical interactions between objects, freely available at www.spiderland.org/breve.

Participant 5

University of Vienna
University of Vienna, Institut fur Theoretische Chemie und Molekulare Strukturbiologie, Wahringerstrasse 17, A-1090 Wien, Austria
Principal Investigator: Prof. Peter Schuster

Qualifications, expertise and intellectual resources

The Vienna group has a long experience in the computational biology of RNA which culminated in the development of a freely available software package, the Vienna RNA Package, consisting of a C code library and several stand-alone programs for the prediction and comparison of RNA secondary structures. In the last years our research activity was focused on the analysis of fitness and folding landscapes. The research staff will benefit from the interaction with the other participating groups by acquire new working techniques and sharing important ideas. The collaboration will also have a positive impact on the career opportunities of the research staff.

Role and main tasks

This participant’s role will consist of developing and investigating the models of RNA. The significant portion of this work will be accomplished in collaboration with the Leipzig and Jena group (Participants 7 and 8) with the purpose of developing the energy surfaces of the model RNA. This part of the project includes Participant 3 who will be involved in the energy landscape investigations together with Participant 5. Also, the research of this participant will contribute to the dynamics of molecular systems through investigating the folding kinetics of RNA.

Personnel

Prof. Peter Schuster is the Chair in Theoretical Chemistry at the University of Vienna, Institute of Theoretical Chemistry and Structural Biology. He studied chemistry and physics and received his Ph.D. degree with "sub auspiciis praesidentis" in 1967. After Postdoctoral studies with Prof. Manfred Eigen at the MPI in Göttingen he received his habilitation from the University of Vienna in 1971. He was a founding director of the IMP in Jena (Germany). He received several scientific awards and published about 280 papers in journals and 9 books.

Selected publications:
1) C. Reidys, C. Forst, and P. Schuster. Replication and mutation on neutral networks. *Bull. Math. Biol*., **63**, 57-94, (2001).
2) C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster. RNA folding at elementary step resolution. *RNA*, **6**, 325-338, (2000).
3) S. Wuchty, W. Fontana, I. L. Hofacker, and P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145-165 (1999).
4) W. Fontana and P. Schuster. Continuity in evolution: On the nature of transitions. *Science*, **280**, 1451-1455 (1998).

Dr. Ivo Ludwig Hofacker is currently University assistant at the Institute for Theoretical Chemistry and Structural Biology at the University of Vienna. He studied physics at the Technical University of Munich and received his Ph.D. degree for studies on the Statistical Characterization of the Sequence to Structure Mapping in RNA from the University of Vienna. After Postdoctoral researcher at the Institute for Theoretical Chemistry he became Postdoctoral Assistant with Prof. Klaus Schulten from the University of Illinois.

Selected publications:
1) I. L. Hofacker, M. Fekete, and P. F. Stadler. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, **319**, 1059-1066 (2002).
2) I. L. Hofacker. The Vienna RNA secondary structure server. *Nucl. Acids Res*., **31**, 3429-3431 (2003).
3) I. L. Hofacker. RNA secondary structure prediction. In D. Cooper (editor), *Encyclopedia of the Human Genome.* Nature publishing group, London (2003).
4) I. L. Hofacker, B. Priwitzer, and P. F. Stadler. Prediction of locally stable RNA secondary structures for genome-wide surveys. *Bioinformatics*, **20**, 191-198 (2004).

Dr. Christoph Flamm is currently University assistant at the Institute for Theoretical Chemistry and Structural Biology at the University of Vienna. He studied organic chemistry at the University of Vienna and received his Ph.D. degree for studies on the Kinetic Folding of RNA from the University of Vienna. After being Postdoctoral Research Assistant at the Institute for Theoretical Chemistry he got his current position.

Selected publications:
1) M. T. Wolfinger, W. A. Svrcek-Seiler, C. Flamm, I. L. Hofacker, and P. F. Stadler. Exact folding dynamics of RNA secondary structures. *J. Phys. A: Math. Gen*., in press (2004).
2) C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger. Barrier trees of degenerate landscapes. *Z. Phys. Chem*., **216**, 155-173 (2002).
3) C. Flamm, I. L. Hofacker, S. Maurer-Stroh, P. F. Stadler, and M. Zehl. Design of multi-stable RNA molecules. *RNA*, **7**, 254-265 (2001).
4) C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster. RNA folding at elementary step resolution. *RNA*, **6**, 325-338 (2000).

Participant 6

University of Heidelberg
The Computational Molecular Biophysics Group, IWR, Uni-Heidelberg, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany
Principal Investigator: Prof. Jeremy C. Smith

Qualifications, expertise and intellectual resources

The Computational Molecular Biophysics Group has extensive experience in molecular dynamics simulation of realistic proteins both in explicit and explicit solvent. The group has 30members and is a part of the Interdisciplinary Centre for Scientific Computing (IWR). The group concentrates on understanding protein dynamics, conformational transitions and reactions and has a significant computational facilities in house and access to outside computing power.

Role and main tasks

The role of the Computational Molecular Biophysics Group will be to provide atomic-detail models of protein dynamics that can then be interpreted using the various models of the other competing groups. In this respect the work of the Heidelberg group will complement that of the Groeningen group which will perform extensive molecular dynamics (MD) simulations of peptides and membranes/vesicles. The Heidelberg group will supply simulations of the dynamics of folded and denatured globular proteins.

Native state: Interaction with the other groups on this point will be at the level of the interpretation of MD simulations of a native protein at different temperatures using landscape and complexity theories. The MD trajectories provide a large amount of data that needs to be interpreted efficiently while taking into account the natural structural units of a protein. Success in this endeavour will lead to a deepened understanding of the internal motions required for protein function.

Personnel

Prof. Jeremy C. Smith holds the Chair of Computational Molecular Biophysics at the University of Heidelberg. In Heidelberg he is a member of the Physics and Biosciences departments and has built up a 30-strong group in the Interdisciplinary Center for Scientific Computing (IWR). He obtained his Ph.D. in protein dynamics from the University of London in 1985. After four years as a postdoctoral fellow and lecturer in the Chemistry Department at Harvard University he set up a research group in the Biology Department of the Commissariat a l'Energie Atomique in Saclay, France, before he moved to Heidelberg. The group in Heidelberg concentrates on understanding protein dynamics, conformational transitions and reactions. He has published over 150 papers.

Selected publications:
1) N. Bondar, M. Elstner, S. Fischer, J.C. Smith and S. Suhai. Mechanism of primary proton transfer in bacteriorhodopsin. *Structure* 12 1281-1288 (2004).
2) E. Balog, T. Becker, M. Oettl, R. Lechner, R. Daniel, J.L. Finney and J.C. Smith. Direct Determination of vibrational density of states change on ligand binding to a protein. *Physical Review Letters* 93 2 028103 (2004).
3) A. Tournier, J. Xu and J.C. Smith. Translational hydration water dynamics drives the protein glass transition. *Biophysical Journal*. **85**(3):1871-5. (2003).
4) A.L. Tournier and J.C. Smith. Principal Components of the Protein Dynamical Transition. *Physical Review Letters*. **91**, 208106 (2003).
5) A.C. Vaiana, H. Neuweiler, A. Schulz, J. Wolfrum, M. Sauer and J.C. Smith. Fluorescence Quenching of Dyes by Tryptophan: Interactions at Atomic Detail from Combination of Experiment and Computer Simulation. *Journal of the American Chemical Society* **125**(47):14564-72 (2003)
6) F. Merzel and J.C. Smith. Is the first hydration shell of lysozyme of higher density than bulk water? *Proceedings of the National Academy of Sciences (U.S.A.)* **99** 8 5378-5383 (2002).
7) H. Neuweiler, A. Schulz, A.C. Vaiana, J.C. Smith, S. Kaul, J. Wolfrum and M. Sauer. Detection of Individual p53-Autoantibodies Using Quenched Peptide-Based Molecular Probes. *Angewandte Chemie (Intl. Ed.)*, **41**(24):4769-73 (2002).
8) V. Reat, R. Dunn, M. Ferrand, J.L. Finney, R.M. Daniel and J.C. Smith. Solvent dependence of dynamic transitions in protein solutions. *Proceedings of the National Academy of Sciences (U.S.A.)* **97** 18 9961-9966 (2000).

Participant 7

University of Leipzig
Lehrstuhl für Bioinformatik, Institut für Informatik, Universität Leipzig, Kreuzstr. 7b, 04103 Leipzig, Germany
Principal Investigator: Prof. Peter F Stadler

Qualifications, expertise and intellectual resources

The research in Peter Stadler's group in Leipzig focusses on the mathematical description of landscapes, the development of algorithms for RNA structure prediction and comparison (in a close collaboration with the Vienna group), phylogenetic footprinting, and molecular phylogenetics. Martin Middendorf's group work primarily on various aspects of evolutionary optimization, complex systems, and phylogenetics.. The researchers both in Leipzig are associated with recently established Bioinformatics Centers funded by the DFG.

Research on complexity measures of landscapes will strengthen and focus the research interests within the department of computer science in Leipzig by joining together the efforts of the Bioinformatics and the Complex Systems groups. The collaboration between the Leipzig and Jena will enhance joint regional research efforts to the benefit of both institutions. The proposed research will provide the opportunity for the staff members to strengthen their profile in interdisciplinary research, thereby enhancing their career opportunities both inside and outside academia. In addition, the project will provide the possibility for contributing young scientists to achieve a PhD degree.

Role and main tasks

The role of Leipzig will be in investigating the energy landscapes. This work will be done in close collaboration with the Vienna group (Participant 5) in application to RNA models potential surfaces. The participant will provide the necessary data on the parameters of molecular energy surface for the research on the dynamical characteristics of biopolymer folding/unfolding, RNA kinetics, and dynamic hierarchies (Participants 3, 5, 4 respectively).

Personnel

Prof. Peter Stadler is the Chair of Bioinformatics at the University of Leipzig, Department of Computer Science and is member of the External Faculty of the Santa Fe Institute, Santa Fe, USA. He received his PhD from the University of Vienna in 1990 and his habilitation in Theoretical Chemistry from the University of Vienna in 1994. He has written more than 100 papers in international journals and was awarded the Novartis price for Chemistry in 2002. He is Editor-in-Chief of Advances in Complex Systems and member of the editorial board of the Journal of Experimental Zoology Part B and the Theory in Biosciences.

Selected publications:
1) P. F. Stadler, C. R. Stephens: Landscapes and Effective Fitness. *Comments Theor. Biol*., **8**, 339-356 (2003).
2) P. F. Stadler, W. Hordijk, J. F. Fontanari: Phase transition and landscape statistics of the number partitioning problem. *Phys.Rev.E*, **67**, 056701 (2003).
3) P. F. Stadler, C. Flamm: Barrier Trees on Poset-Valued Landscapes. *Genet. Prog. Evolv. Mach.*, **4**, 7-20 (2003).
4) B.M.R. Stadler, P.F. Stadler: Generalized Topological Spaces in Evolutionary Theory and Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci*., **42**, 577-585 (2002).
5) C. M. Reidys P. F. Stadler: Combinatorial Landscapes. *SIAM Review*, **44**, 3-54 (2002).

Prof. Middendorf is professor for parallel computing and complex systems at the University of Leipzig, Department of Computer Science. He received his Ph.D. degree from the University of Hannover in 1992 and his habilitation from the University of Karlsruhe in 1998. He has been working on each of the fields swarm intelligence, optimization, and parallel computing for more than 5 years. He has published in these areas more than 30 scientific articles in international journals, more than 40 scientific articles in international conferences, and edited 4 books and journal special issues. He has been in the programme committee of more than 35 international conferences in the last 5 years.

Selected publications:
1) M. Middendorf, D. Manlove. Combined Super-Substring and Super-/Subsequence problems. *Theoretical Computer Science*, **320**(2-3): 247-267, (2004).
2) S. Janson, M. Middendorf. A Hierarchical Particle Swarm Optimizer for Dynamic Optimization Problems. Proc. Evoworkshops 2004, Springer, LNCS 3005, 513- 524 (2004).
3) M. Middendorf, F. Reischle, and H. Schmeck. Multi Colony Ant Algorithms. *Journal of Heuristics*, **8**, 305-320, (2002).
4) M. Middendorf and V. Timkovsky. On scheduling cycle shops: classification, complexity and approximation. *Journal of Scheduling*, **5**, 135-169 (2002).

Dr. Merkle has written his doctoral thesis about Ant Colony Optimization and is an active researcher in the field of swarm intelligence. He completed his doctoral thesis in 2002 at the University of Karlsruhe and is now assistant professor (Wiss. Assistent) at the University of Leipzig. He has co-authored 6 articles in international journals and 16 scientific articles in international conference proceedings or peer reviewed books.

Selected publications:
1) S. Janson, D. Merkle, M. Middendorf, H. ElGindy, H. Schmeck. On Enforced Convergence of ACO and its Implementation on the Reconfigurable Mesh Architecture Using Size Reduction Tasks. *Journal of Supercomputing*, **26**, 221-238 (2003).
2) D. Merkle and M. Middendorf. Modelling the Dynamics of Ant Colony Optimization Algorithms. *Evolutionary Computation*, **10**, 235-262 (2002).
3) D. Merkle, M. Middendorf, and H. Schmeck. Ant Colony Optimization for Resource- Constrained Project Scheduling. *IEEE Transactions on Evolutionary Computation*, **6**, 333-346 (2002).
4) D. Merkle and M. Middendorf. Fast Ant Colony Optimization on Runtime Reconfigurable Processor Arrays. *Genetic Programming and Evolvable Machines*, **3**, 345-361 (2002).

Dr. Axel Mosig received his PhD on Efficient Algorithms for Shape and Pattern Matching from the University of Bonn, Germany, in 2004. His research has been presented at the European Workshop on Computational Geometry and the NATO Advanced Study Institute on Non-Commutative Algebra and its Applications in 2003. Further work is about to be published in Computational Geometry: Theory and Applications, which he also served as a referee for. Currently, Dr. Mosig is a research assistant at the University of Leipzig Bioinformatics group.

Selected publications:
1) A. Mosig. Matching Polygonal Curves using Group Theory, *NATO Advanced Study on Computational Noncommutative Algebra and Applications*, Tuscany, Italy (2003).
2) M. Clausen and A. Mosig. Approximately Matching Polygonal Curves with Respect to the Frechet Distance, *19th Europ. Worksh. Comp. Geom.* (2003).

Dr. Konstantin Klemm completed his PhD studies in Physics at the Niels Bohr Institute, Denmark, in 2002. He was employed at the physics departments of Kiel, Copenhagen, and Palma de Mallorca. He has published 10 peer-reviewed papers in the fields of Complex Networks, Neuroscience and non-equilibrium phase transitions. In his current position as a postdoctoral researcher at the Interdisciplinary Centre for Bioinformatics in Leipzig his work focuses on models of gene regulatory networks. He is a reviewer for Physical Review Letters, Physical Review E, Advances in Complex

Systems and European Physical Journal B. He has co-organized international conferences on Complex Networks, Econophysics and Weak Turbulence in Germany, Belgium and Denmark.

Selected publications:
1) V. M. Eguiluz and K. Klemm. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett*., **89**, 108701 (2002).
2) K. Klemm, V. M. Eguiluz, R. Toral, and M. San Miguel. Nonequilibrium transitions in complex networks: a model of social interaction. *Phys. Rev. E*, **67**, 026120 (2003).
3) K. Klemm, V. M. Eguiluz, R. Toral, and M. San Miguel. Global culture: A noise-induced transition in finite systems. *Phys. Rev. E*, **67**, 045101(R) (2003).
4) K. Klemm, V. M. Eguiluz, R. Toral, and M. San Miguel. Role of dimensionality in Axelrod's model for the dissemination of culture. *Physica A* ,**327**, 1 (2003).

Participant 8

Friedrich-Schiller-Universität Jena
Lehrstuhl für Bioinformatik, Institut für Informatik, Ernst-Abbe-Platz 2, 07743 Jena
Principal Investigator: Prof. Rolf Backofen

Qualifications, expertise and intellectual resources

The group in Jena focuses on structure prediction in simplified models, sequence/structure alignment of RNAs, the investigation of alternative splice forms of proteins, rules-based techniques for the description of regulatory sequences (promoters) as well as constraint programming. The researchers in Jena are associated with recently established Bioinformatics Centers funded by the German Ministery of Sciences.

Role and main tasks

The role of Jena will be in investigating the energy landscapes. We have recently solved the structure prediction problem for the HP-type lattice proteins, we we are able to find native conformations for sequences up to length 300. This work will be done in close collaboration with the Leipzig group (Participant 7) to explore the low energy regions of the conformation space, which will be used to determine the parameters of molecular energy surface for the research on the dynamical characteristics of biopolymer folding/unfolding.

Personnel

Prof. Rolf Backofen) is a Chair in Bioinformatics at the Friedrich-Schiller-University Jena, Institute of Computer Science. He received his Ph.D. degree in computer science from the University of Saarland in 1994, where he worked at the German Research Center for Artificial Intelligence (DFKI). He received his habilitation from the University Munich (LMU). He is co-author of the book "Computational Molecular Biology: An Introduction" (Wiley&Sons, Mathematical and Computational Biology Series, 2000). His current research interests are protein structure prediction, RNA-motif finding, alternative splicing, transcription factors and promoters, selenoproteins and constraint programming and its application in bioinformatics.

Selected publications:
1) R. Backofen. A Polynomial Time Upper Bound for the Number of Contacts in the HP-Model on the Face-Centered-Cubic Lattice (FCC). *Journal of Discrete Algorithms*, in press (2003).
2) R. Backofen, N. S. Narayanaswamy, F. Swidan. Protein similarity search under mRNA structural constraints: application to targeted selenocysteine insertion. *In Silico Biology*, **2**(3), 275-90 (2002).
3) R. Backofen, D. Gilbert. Bioinformatics and Constraints. *Constraints*, **6**, 141-156 (2001).
4) R. Backofen. The Protein Structure Prediction Problem: A Constraint Optimisation Approach using a New Lower Bound. *Constraints*, **6**, 223-255 (2001).

Sebastian Will is Ph.D. student at the Chair for Bioinformatics at the Department for Mathematics and Computer Science of the Friedrich-Schiller University Jena. He received his master degree from the Institute of Computer Science at the Ludwig-Maximillians University of Munich in 2000, where he won a price as one of the best students in this year. Before his change to Jena in early 2002, he held a scholarship of the PhD programme "Logic in Computer Science" in Munich.

Selected publications:
1) R. Backofen, S. Will. Local Sequence-Structure Motifs in RNA, Journal of Bioinformatics and Computational Biology (JBCB), to appear (2004).

2) R. Backofen, S. Will. A Constraint-Based Approach to Structure Prediction for Simplified Protein Models that Outperforms Other Existing Methods. Proc. 19th Int. Conf. on Logic Programming (ICLP 2003), 49-71, (2003).
3) R. Backofen, S. Will. Excluding symmetries in constraint-based search. Constraints, 7(3), 333-349 (2002).

References

1. P.A.Y. Gunter, *Chaos, Solitons & Fractals*, **20**, 5 (2004).
2. I. Prigogine and I. Stengers, in *Order out of chaos*, **14**, 175 (1984)
3. J.P. Crutchfield, D.P. Feldman, and C.R. Shalizi, *Phys. Rev. E*, **62**, 2996 (2000)
4. P. Allegrini, M. Giuntoli, P. Grigolini, and B. J. West, From knowledge, knowability and the search for objective randomness to a new vision of complexity, *Chaos, Solitons & Fractals*, **20**(1), 11 (2004)
5. S. Nara, Can potentially useful dynamics to solve complex problems emerge from constrained chaos and/or chaotic itinerancy?, Chaos: An Interdisciplinary Journal of Nonlinear Science, 13(3), 1110 (2003)
6. G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani, Predictability: a way to characterize complexity, *Physics Reports*, **356**(6), 367 (2002)
7. R. Badii and A. Politi, *Complexity*, Cambridge University Press, Cambridge , 1997.
8. *Entropy, Complexity, and Physics of Information*, SFI Studies in the Sciences of Complexity, VIII, edited by W. Zurek, Addison-Wesley,Reading, Massachusetts (1990)
9. *Evolutionary dynamics. exploring the interplay of selection, accident, neutrality, and function*, edited by J.P. Crutchfield and P. Schuster, Oxford University Press, New York (2002)
10. S.A. Kauffman, *The Origins of Order. Self-organization and selection in evolution*, Oxford University Press, New York, Oxford (1993)
11. *Complexity in Chemistry Introduction and Fundamentals*, edited by: D. Bonchev, D. H Rouvray, Taylor & Francis (2003)
12. M. Braxenthaler, R. Unger, D. Auerbach, J.A. Given, and J. Moult, *Proteins: structure, function, and genetics*, 29, 417 (1997)
13. H.-b. Zhou. *J. Phys. Chem*., 100, 8101 (1996)
14. T. Rage, A. Neumaier, and C. Schlier, *Phys. Rev. E*, **50**, 2682 (1994)
15. R. Gaspard, M.E. Briggs, M.K. Francis, J.V. Sengers, R.W. Gammon, J.R. Dorfman, and R.V. Calabrese, *Nature*, **394**, 865 (1998)
16. D. Nerukh, G. Karvounis, and R. Glen, Complexity of classical dynamics of molecular systems. Part I: methodology and Part II: finite statistical complexity of a water-Na^+ system, *J. Chem. Phys.*, **117**(21), 9611-9617 (2002) and *J. Chem. Phys.*, **117**(21), 9618-9622 (2002)
17. P. Calmettes, B. Roux, D. Durand, M. Desmadril, and J.C. Smith, Configurational Distribution of Denatured Phosphoglycerate Kinase, *Journal of Molecular Biology*, **231**, 840-848 (1993)
18. P. Calmettes, D. Durand, P. Minard, M. Desmadril, V. Receveur, and J.C. Smith, How Random is a Highly Denatured Protein? *Biophysical Chemistry*, **53,** 105-114 (1994)
19. A.J. Petrescu, V. Receveur, P. Calmettes, D. Durand, M. Desmadril, B. Roux, and J.C. Smith, Small Angle Neutron Scattering by a Strongly Denatured Protein: Analysis using Random Polymer, *Theory. Biophysical Journal*, **72,** 335-342 (1997).
20. A.-J. Petrescu, P. Calmettes, D. Durand, V. Receveur, and J.C. Smith, Change in backbone torsion angle distribution on protein folding, *Protein Science*, **9,** 1129-1136 (2000)
21. A. Gruia, S. Fischer & J.C.Smith, Kinetics of breaking a salt-bridge critical in protein unfolding, *Chemical Physics Letters*, **385**, 337-340 (2004)
22. A.D. Gruia, S. Fischer, and J.C. Smith, Molecular dynamics simulation reveals a surface salt bridge forming a kinetic trap in the unfolding of truncated Staphylococcal nuclease. *Proteins: Structure, Function, and Genetics*, **50**, 507-515 (2003)
23. .J. Hayward and J.C. Smith, Temperature dependence of protein dynamics: computer simulation analysis of neutron scattering properties. *Biophysical Journal*, **82,** 1216-1225 (2002)
24. F. Merzel and J.C. Smith, Is the first hydration shell of lyzozyme of higher density than bulk water? *Proceedings of the National Academy of* Sciences (U.S.A.), **99**(8), 5378-5383 (2002)
25. J.A. Hayward, J.L. Finney, R. Daniel, and J.C. Smith, Molecular dynamics decomposition of temperature-dependent elastic neutron scattering by a protein solution. *Biophysical Journal*, **85**(2), 679-85 (2003)
26. A. Tournier, J. Xu, and J.C. Smith, Translational hydration water dynamics drives the protein glass transition. *Biophysical Journal*. **85**(3), 1871-5. (2003)
27. A. Tournier and J.C. Smith, Principal Components of the Protein Dynamical Transition. *Physical Review Letters*, **91**, 208106 (2003)
28. A. Torcini, R. Livi, and A. Politi, *J. Biol. Phys*. **27** 181 (2001)
29. L. Bongini et al., *Phys. Rev. E*, **68,** 061111 (2003)
30. J.D. Honeycutt and D. Thirumalai, *Proc. Natl. Acad. Sci*., **87,** 3526 (1990)
31. P. Bruscolini and L. Casetti, *J. Biol. Phys*. **27** (2001), 243; *Phys. Rev. E*, **64,** 051805 (2001)
32. L. Casetti, M. Pettini and E. G. D. Cohen, *Phys. Rep*., **337**,237 (2000)

33. M. Carrion-Vasquez et al., *Proc. Natl. Acad. Sci*., **96**, 3694 (1999); A.F. Oberhauser et al., *Nature Structural Biology*, **6** 1025 (1999); A.F. Oberhauser et al., *Proc. Natl. Acad. Sci*., **98,** 468 (2001)
34. E. Evans and K. Ritchie, *Biophys. J.*, **72**, 1541 (1997); B. Heymann and Grubmueller, *Phys. Rev. Lett.*, **84,** 6126 (2000)
35. J. Liphardt et al., *Science,* **296,** 1833 (2002); J. Gore at al., *Proc. Natl. Acad. Sci.*, **100,** 12564 (2003)
36. O. Braun et al., Probing molecular free energy landscapes by periodic loading' (cond-mat/0402496)
37. F-Y li et al., *Phys. Rev. E,* **63** 021905 (2001)
38 R. Backofen, The protein structure prediction problem: A constraint optimisation approach using a new lower bound. *Constraints*, **6**,223-255 (2001)
39 R. Backofen, A polynomial time upper bound for the number of contacts in the HP-model on the face-centered-cubic lattice (fcc). *Journal of Discrete Algorithms*, in press (2003)
40 R. Backofen and S. Will, A constraint-based approach to structure prediction for simplified protein models that outperforms other existing methods. In: *Proceedings of the 19th International Conference on Logic Programming (ICLP 2003)* (2003).
41. R. Backofen, S. Will, and E. Bornberg-Bauer, Application of constraint programming techniques for structure prediction of lattice proteins with extended alphabets. *J. Bioinformatics*, **15**(3):234-242 (1999)
42 S. Janson and M. Middendorf, A Hierarchical Particle Swarm Optimizer. *Proc. 2003 Congress on Evolutionary Computation (CEC 2003)*, IEEE Press, 770-776 (2003)
43 S. Janson and M. Middendorf, A Hierarchical Particle Swarm Optimizer for Dynamic Optimization Problems. *Proc. EvoWorkshops 2004*, LNCS 3005, 513-524 (2004)
44 D. Merkle and M. Middendorf, Modelling the Dynamics of Ant Colony Optimization Algorithms. *Evolutionary Computation*, **10**(3), 235-262, (2002)
45 M. Middendorf, F. Reischle, and H. Schmeck, Multi Colony Ant Algorithms. *Journal of Heuristics*, **8**(3), 305-320 (2002)
46 C.M. Reidys and P.F. Stadler, Combinatorial Landscapes. *SIAM Review* **44**, 3-54 (2002)
47 C. Flamm, I.L. Hofacker, P.F. Stadler. and M.T. Wolfinger, Barrier Trees of Degenerate Landscapes. *Z. Phys. Chem*. **216**, 155-173 (2002)
48 J.F. Fontanari. and P.F. Stadler. Fractal geometry of spin-glass models, *J.Phys.A:Math.Gen*. **35**, 1509-1516 (2002)
49 D. Rockmore, P. Kostelec, W. Hordijk, and P.F. Stadler, Fast Fourier Transform for Fitness Landscapes. *Appl. Comput. Harmonic Anal*. **12**, 57-76 (2002)
50 C.M. Reidys and P.F. Stadler. and P. Schuster, Generic Properties of Combinatory Maps: Neutral Networks of RNA Secondary Structures. *Bull.Math.Biol*. **59**, 339-397 (1997)
51 C.M. Reidys and P.F. Stadler, Neutrality in Fitness Landscapes. *Appl. Math. & Comput*. **117**, 321-350 (2001)
52. K. Yue and K.A. Dill. Folding proteins with a simple energy function and extensive conformational searching. *Protein Science*, **5**(2), 254-261 (1996)
53. K. Yue, K. M. Fiebig, P. D. Thomas, H. S. Chan, E. I. Sha khnovich, and K. A. Dill, A test of lattice protein folding algorithms. *Proc. Natl. Acad. Sci. USA*, **92**(1),325-9 (1995).
54. Wuchty et al., *Biopolymers*, **49**, 145-165 (1999)
55. C. Flamm et al., *RNA*, **6**,325-338 (1999)
56. C. Flamm et al., *Z. Phys. Chem*., **216**, 155-173 (2002)
57. Wolfinger et al., *J. Phys. A*, in press (2004)
58. C. Flamm et al., *RNA*, **7**, 254-265 (2001)
59. P. Schuster et al., *Proc.Roy.Soc.B*, **255**, 279-284 (1994) and Fontana & Schuster, *Science* **280**, 1451-1455 (1998)
60. Schultes & Bartel, *Science* **289**, 448-452 (2000)
61. S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, (New York, NY: Springer-Verlag (1990), J. Guckenheimer and P.J. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, New York, NY, Springer-Verlag.(1983) J. Carr, *Applications of Center Manifold Theory*, New York, NY: Springer-Verlag (1981)
62. D. Ruelle, Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics, *Journal of Statistical Physics*, **95** (1/2), 393-468 (1999)
63. M. Nilsson, 2004, Hierarchical Structures in Smooth Dynamical Systems, submitted to Artificial Life.
64. Working paper: http://frt.fy.chalmers.se/cs/people/lindgren/flows_of_information.pdf accepted for presentation at *Artificial Life* **9** in Boston September (2004)
65. K.-E. Eriksson and K. Lindgren, Structural information in self-organizing systems, *Physica Scripta* **35**, 388-397 (1987)
66. K.-E. Eriksson, K. Lindgren, and B. Å. Månsson, *Structure, Context, Complexity, and Organization*, World Scientific, Singapore (1987)
67. K. Lindgren, Microscopic and macroscopic entropy, *Phys Rev A*, **38**, 4794-4798 (1988)