Statement of teaching Philosophy for Dr. Florence Tama

My first teaching experience was as a teaching assistant for introductory chemistry at the University Paul Sabatier (France). I was fully responsible for a group of 30 students. I had great interactions with the students. The size of the class was perfect to closely monitor each student and provide alternative explanations. I was also in charge of two lab sessions in chemistry and spectroscopy. These classes were rewarding because not only I was able to directly interact with students and guide them with experiments, but also it gave me the opportunity to help them with the understanding of the theoretical background.

I have also experience in teaching advance courses. During my post-doctoral training, I have been invited several times to give lectures for faculties, post-doctoral researchers and graduate students in workshops held in La Jolla (Situs Workshop, TSRI, February 2003; A Practical Course in Molecular Microscopy, TSRI, November 2003; Theory and Computation in Molecular Biological Physics, CTPB and TSRI, August 2004). For these workshops I was fully responsible for the preparation of the lecture and the hands-on sessions.

Teaching is about transfer of knowledge and is an essential part of learning. My philosophy of teaching aims to stimulate students to develop abilities to solve problems utilizing their own skills. Most importantly, I believe that the role of the education is to guide the students to think on their own and develop critical thinking. To achieve this goal, I am convinced that the ideal learning environment is a small group where the interaction between the professor and the student is direct. The professor should create an environment that gives opportunity to the students to be enthusiastic about learning and being active. The active involvement of students should result in small projects in which students are required to work in groups to study a practical problem, design and implement a solution, document the results, and present those results to the rest of the class.

Effective teaching depends in part on the quality of the prepared material. It requires presentation methods that are stimulating, interactive and accessible to a diverse population of student. I believe that using simple examples that are easily understandable such as analogies and similarities with other disciplines makes abstract concepts more concrete. It is also always important to remind the students of the overall picture before going into details to stimulate their interest. In addition, I think it is essential to connect the current course with what students have learned in the past and are going to learn in the future to provide continuity in the learning curve.

With my undergraduate background in chemistry and physics and graduate background in biophysics, I believe that I am well qualified to teach most introductory courses in chemistry, physical chemistry and biochemistry. At the graduate level, I would be really interested in bringing computational biology and biophysics experiments together by designing a course that would approach dynamics of biomolecules from computational and experimental aspects and demonstrate how such techniques can complement each others to answer fundamental questions about the relation between the structure-function-dynamics of biological molecules.

Research accomplishments

PhD Thesis: Theoretical studies of large amplitude motions of proteins

Dynamics is essential for a protein to function. Recent advancements in computers and techniques enable us to simulate protein dynamics. One of the methods commonly used is the normal mode analysis (NMA).[1] In NMA, dynamics of a molecule are described as a set of normal mode, and low-frequency modes are closely related to large conformational change of proteins. However, the application of NMA critically depends on diagonalization of the Hessian, i.e. the matrix of the second derivatives of the potential energy, which has been a limiting factor in applying NMA to interesting large biological macromolecular assemblies due to its large memory requirement. We have introduced the RTB method to solve this issue by reducing the size of the Hessian to be diagonalized. The RTB method uses a simple physical idea: a protein or nucleic chain can be viewed as being comprised of rigid components linked together, such as residues/bases or group of residues/bases or more extensive segments of structure forming secondary structural elements.[2,3] Our work is an important advance in the field of normal mode analysis since now we can study any large biological systems with short computational time.

At the time I started my Ph.D, a very exciting paper by M. Tirion showed that elastic network models could be used to obtain normal modes of biological molecules.[4] In this approach, a simplified potential is used to represent the molecule as a set of pseudo-particles. These particles are coupled *via* harmonic springs to describe the system as an elastic network. This model does not necessitate preliminary energy minimization, which has been another limiting step in NMA, thereby permitting direct analysis of crystal and NMR coordinates. Reduced representation of the molecule, whereby a single coordinate is used to represent several atoms[5] provides a multi-scale description that can significantly reduce the computational expense. We investigated conformational changes of a large number of proteins and shown conformational change could be describe by a single low-frequency mode.[6] Therefore, theoretical tools enable us to discuss the kind of conformational change a protein can undergo, subsequently the function, from a single conformation.

I also became interested in the compressibility of proteins. Studies of this property lead to some useful insights into the mechanical properties of the proteins. We have presented a new approach, based on the concept of principal component analysis, to study volume fluctuation.[7] We have shown that low-frequency dynamics dominate the total volume fluctuation. Moreover, we pointed out that intrinsic compressibility of proteins may be larger than those usually deduced from experiment.

After I completed my PhD, I became extremely interested in very large macromolecular assemblies and their machineries to achieve biological functions. I joined Charles L. Brooks III's group to work within the Multiscale Modeling Tools Structural Biology NIH research resource. One objective of this resource is to develop multi-scale models to study dynamical properties of large macromolecular assemblies. My previous works on NMA methodology and reduced representation of proteins put me in an ideal position to tackle such problems. During the course of my post-doctoral research, I have been applying NMA to several biological systems to discuss their functional motions and developing new methodologies to assist the interpretations of experimental data.

Unveiling molecular mechanisms of biological functions in large macromolecular assemblies

Viruses

Important aspects in virology regard the assembly and maturation of virus particles which involve large-scale rearrangements.[8] We studied Cowpea Chlorotic Mottle Virus (CCMV), which is made up of 180 proteins that form a 286 Å diameter shell. Native CCMV is stable around pH 5.0, at pH 7.0 the particles undergo a transition to a swollen form (size increased by ~10 %[9]). We performed NMA using the 28620 Cα atoms of

CCMV. Starting from the native CCMV, intermediate structures were created along low frequency modes, for which we calculated the agreement (R-factor) with the experimental electron density of swollen CCMV. From pKa calculations on these structures, we determined that a single acidic residue (GLU81), with higher pKa, might trigger the swelling process

We also analyzed the mechanical properties of a variety of viruses comprising different sizes and quasi-equivalence symmetries. We found that pentamers generally have higher flexibility and the ability to move freely from the others capsomers, which facilitates shape adaptation and could be important in the viral life cycle.[10]

70S ribosome

We have been collaborating with Prof. J. Frank (HHMI, Wadsworth Center, Albany) to examine atomic level models to describe the mechanism of translocation of the ribosome. One of the key mechanical steps of protein synthesis is the translocation process, in which the tRNAs move from the A-aminoacyl and P-peptidyl site into the P and E-exit site respectively. This process is promoted by binding of the elongation factor G protein (EF-G) and GTP hydrolysis, and is accompanied by large conformational rearrangements of the ribosome.[11] We studied the rearrangements of the ribosome using the theoretical methods of elastic network NMA to provide a near atomic level description of these structural rearrangements linking motions obtained from the NMA to experimental observations.

One of the resulting normal mode shows a rotation of the 30S relative to the 50S subunit[12] similar to the one described by our collaborator Joachim Frank by cryo-EM experiment.[11] The analysis of this mode reveals that the rotation leads to relatively large conformational rearrangements in the stalk base in the 50S subunit and the head and the shoulder in the 30S subunit that interact with EF-G (see Fig. 1) which suggests a mechanical coupling between the binding sites of EF-G and the ratchet motion.

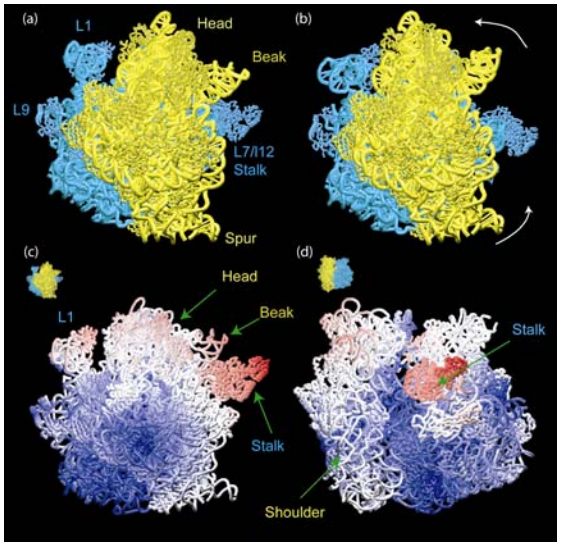

Figure 1: Elastic network NMA of the 70S ribosome. (a) The X-ray structure (the 30S subunit in yellow and the 50S in blue), (b) Rearrangements of after displacement along the ratchet-like mode. (c-d) Two different views, as indicated by thumbnails, of the atomic displacements along the ratchet-like mode. The atoms are colored according to their amplitude along the mode. The scale range from red: largest rearrangements to blue: no motions

The underlying mechanism of the ratchet like motion could be examined in structural detail using this calculated mode.[12] First, helix 27 of the 16S RNA emerges as a part of the axis of the rotation of the 30S subunit. Secondly, two types of bridges emerge from our study: some facilitate the conformational switching by being flexible, while others are very rigid and serve to maintain the integrity of the architecture during the ratchet-related conformational change. Thirdly, the change in the interactions between the ribosome and the tRNAs suggests that the ratchet-like motion may facilitate the movement of the tRNAs through the inter-

subunit space, by optimizing their positions for the translocation process. We complemented this study with some electrostatics calculations on the structures before and after the ratchet like motion.[13]

Myosin

We are modeling, in collaboration with Prof. K. Taylor (FSU), rearrangements associated with transitions between functional and inhibited states of myosin. Dephosphorylated myosin (10S) is soluble and adopts a folded rod domain, while phosphorylation leads to a conformational change to an extended form (6S) that is competent for filament assembly. The mechanism of formation of the inhibited state is unexplained, in particular how the two myosin heads (S1 in Fig. 2) fold backwards onto the rod domain and how the arrangement of the two heads can destabilize the filaments.

To address these fundamental questions about the functioning of myosin, we model the conformational transition from the activated state of smooth muscle heavy meromyosin (HMM) to its inhibited state using NMA.[14] An atomic model for the inhibited structure has been described[15] (see Fig. 2d). It is known that the active state is symmetrical and S2 adopts a coiled-coil structure, however the number of residues that should be uncoiled at the S1-S2 junction is not known. Therefore, we have constructed several hypothetical models for the active state with 0, 14, 21 or 28 residues uncoiled at the S1-S2 junctions (see Fig. 2a,2f) and examined using NMA which model of uncoiled junction best conforms with other experimental data.

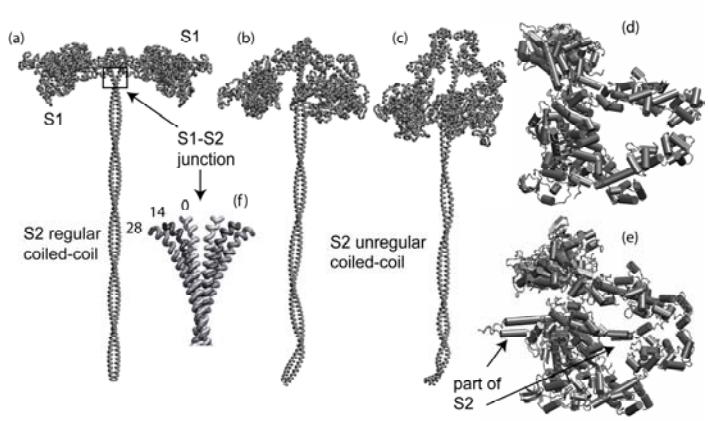

Figure 2: NMA of HMM. (a) the model of the active state, (b) intermediate structure for the transition (c) the final structure obtained from NMA, (d) S1 domains of the target structure, (e) S1 domains of the final structure obtained from NMA, (f) Modeled S1-S2 junction with 0, 14 and 28 uncoiled residues.

The analysis of the final structures obtained from a normal mode iterative procedure indicates at least 14 but not more than 21 residues at the S1-S2 junction must be uncoiled. We observed that the transition between the symmetrical active state to the asymmetrical inhibited state induces distortions throughout the S2 coiled-coil, which explains for the first time a mechanism for the changes in myosin solubility due to dephosphorylation. Such distortions produced in a sufficient number of myosin molecules would destabilize the filament as the tight filament packing of S2 domains of assembling myosins, which possible with a regular coiled-coil is disrupted. As the myosin is released from the filaments, further folding into the 10S conformation occurs until all the myosin is dissolved.

Exploration of global dynamics and interpretation of low-resolution structural information

The recent developments in electron microscopy (EM) have been producing a growing number of low-resolution structures of large assemblies. In particular, large conformational changes of macromolecular complexes have been characterized by cryo-EM.[16] Nevertheless, few theoretical methods have been developed to assist the interpretation of their dynamical properties. In this section, I present the first application of NMA to study global deformations of biological systems from low-resolution structure and a new procedure for the flexible fitting of high-resolution structures into low-resolution structures.

Dynamical properties

Highly reduced representations are sufficient to provide the dynamical information on X-ray and NMR protein structures. Thus, in principle it should be possible to study dynamical properties of low-resolution structures such as electron density maps if one develops a discrete representation of the density. In our approach, the shape of the biological object is encoded in so-called codebook vectors[17] that identify the structural features of the system. In combination with the elastic network model, we were able to obtain the dynamical properties of low-resolution structures.[18] Applications of our methodology to experimental data revealed functionally important rearrangements of several macromolecular assemblies as illustrated in Fig 3 with the 70S ribosome.[19]

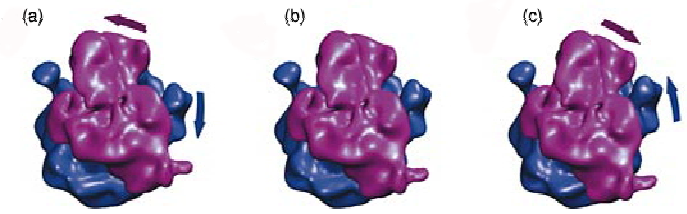

Figure 3: Ratcheting motion of the 25 Å EM map of the ribosome captured by the lowest-frequency mode. The 30 S subunit is shown in purple, the 50 S subunit in blue. Panel (a) correspond to the maximum positive deformation, panel (b) to the original map, and panel (c) to the maximum negative deformation.

NMFF: Normal mode flexible fitting

To complement emerging information from EM, it is tempting to use existing X-ray data to construct atomic models. As large assemblies often undergo large functional rearrangements, the conformations of the available X-ray structures may not correspond to the conformation apparent in the cryo-EM data. In such cases, the fitting can be rather complicated, since not only the correct orientation needs to be found but conformational rearrangements must also be considered. However, current approaches deform the molecule in an *ad-hoc* way.

We have introduced a novel method for the quantitative flexible docking of a high-resolution structure into low-resolution maps of macromolecular complexes from EM that takes into account the conformational flexibility of biological systems.[20] This method uses a linear combination of low frequency normal modes obtained from elastic network model in an iterative manner to deform the structure optimally to conform to the low-resolution electron density map. Using only the low frequency normal modes, the proteins are deformed in a mechanically realistic way as demonstrated with our applications on experimental data.[21] In collaboration with Mark Fisher, we applied this technique to model the conformational change of GroEL upon substrate binding[22] and the lethal anthrax complex.

Research plans

I intend to focus my future research on the emerging area of structural proteomics by use of computational methods to assist the building of atomic structural models of macromolecular complexes, which are critically important for the machinery of the cell. In addition, I would like to approach biophysical problems of large macromolecular assemblies from a more theoretical perspective to understand the physical principles of their assembly processes and mechanical properties. Physicochemical understandings of these processes will surely deepen our knowledge of the biological machinery of these macromolecules. In the following, I first describe some projects aimed as a base foundation to tackle the challenging questions arising from structural proteomics and then discuss my long-term research plans.

Introduction

It is now well established that large-scale rearrangements in proteins are important for a variety of protein functions including catalysis and regulation of activity. The recent developments in experimental methods, especially cryo-electron microscopy (EM), have revealed that large-molecular assemblies are also highly dynamic. While experiment provides a tremendous source of information on these dynamical properties, computational methods must be employed to complement experimental observations. Indeed, by using theory to explore functionally important rearrangements observed in experiments at low-resolution, it is possible to gain new insights into the mechanism of these transformations that are presently inaccessible to experiments.

To study motions of macromolecules, normal mode analysis (NMA) provides an alternative to molecular dynamics. NMA consists in the decomposition of the motion into vibrational modes. This approach extends the time scale accessible to theoretical work and has proven to be extremely useful for studying collective motions of biological systems.[23,24] Exploration of the normal modes of a molecular system can yield insights, at the atomic level, on the mechanism of large-scale rearrangements of proteins/proteins complexes that occur upon ligand/protein binding.[25,26] Studies employing NMA generally focus on a few large-amplitude/low-frequency normal modes, which are expected to be relevant to function.

Studies from my Ph.D to postdoctoral research have enabled the extension of applications of NMA to large macromolecular assemblies.[3] The primary motivation behind my research is to complement experimental observations by exploring, at near-atomic levels of resolution, functionally important rearrangements observed in experiments at low-resolution, and to obtain new insights into the mechanism of these transformations. My application of these approaches has focused on several macromolecular assemblies and aimed to provide a better understanding of their functional motions, i.e., the swelling transition of viruses,[10,18] the ratchet-like motion of the ribosome[12] and the conformational transition between active and inhibited myosin.[14] In addition, I have developed methods to apply NMA to data arising from cryo-EM experiments[19,27] and to assist the interpretation of molecular rearrangements arising in these systems.[20,21]

Research Interests: New directions

Electrostatic properties/Free energy profile of conformational changes of biological molecules

My previous work on the ribosome, myosin, viruses and dynamics of low-resolution data has provided a new perspective on how macromolecular assemblies work: they evolve to have a shape that is amenable for large scale, machine-like motions. While this principle would explain the motions that the molecules can undergo, in order to fully understand the mechanism of biological molecules, physical causes triggering these

transitions need to be understood. For this, detailed studies on the energetics of the transition processes related to the triggering factors are necessary.

An intriguing example is seen in the ribosome. During protein synthesis, the ribosome interacts with several factors. The ternary complex[28] brings tRNAs to the ribosome, the elongation factor[29] is necessary to add residues to the chain and finally the termination factor stops the transcription process.[30] Each of these factors binds the ribosome in the same region, however they steer the machinery of the ribosome in different ways. Experimentally, each of them is known to undergo large conformational changes upon associating to the ribosome. Presumably, strain accumulated by these factors during the conformational change induces conformational transitions of the ribosome. Thus studying the energetics of the conformational changes of these factors would certainly help our understanding of the machinery of the ribosome.

Therefore, to deepen the physicochemical understanding of biomolecular functions of these factors, I intend to pursue the quantitative analyses of their conformational changes. Using normal mode analysis in an iterative manner,[14] which includes the non-harmonic character of the energy landscape, one can obtain an accurate description of the conformational change pathway between two conformations. With energy calculations on the intermediate structures along the path, the energy surface of the conformational change can be studied. Furthermore, we can perform simulations from these intermediate structures with suitable biased sampling techniques to determine the free energy profile of the transition, which would provide an atomic level detailed picture of the conformational changes.

The description of the conformational change using NMA also opens the way for further studies on large macromolecular assemblies. At the level of super-complexes of proteins and/or RNAs, its machine-like motion is often achieved by using molecules as a mechanical parts, for example, the ratchet-like motion of the ribosome is the relative rotation of 50S and 30S subunits.[11] Thus, the physicochemical properties of the interface between protein/RNA are a key for understanding the mechanism of important macromolecules. From structures along the conformational change path, it would be interesting to look at the change in the association energy between each component of the complex to determine the role of the interfaces in the conformational transitions. Indeed, some interfaces may maintain the core of the complex during the conformational change while others may facilitate it by being more flexible. Such differences should appear in the evolution of the association energies along the conformational transition path.

A candidate system for these studies would be the RNA polymerase, a macromolecular assembly composed of several proteins that synthesizes RNA. Large conformational rearrangements of this assembly have been related to its function. Several structures have been solved[31,32] and we have already demonstrated that a description of its conformational change can be obtained using NMA.[19] This preliminary result indicates that I will be able to construct intermediates structures along the conformational change path between the two states. Detailed calculations of the association energies on each intermediate structure would certainly shed the light on some aspects of the machinery of this important macromolecular machine.

Dynamics events related to the cell motility: Actin

Cells contain elaborate arrays of protein fibers, the cytoskeleton. There are three types of cytoskeletal filaments: actin, microtubule and intermediate filaments. Dynamic assembly and disassembly of these filaments and their interactions with several proteins or complexes of proteins are required to achieve a number of critical biological processes such as establishing cell shape, providing mechanical strength, locomotion and chromosome separation in mitosis and meiosis. While several interesting questions and problems arise for each of these systems, I am particularly eager to study the actin filament. Indeed, for filamentous actin (F-actin), it has been shown that the binding of some of these proteins are critical to maintain cellular viability.[33] Therefore, using the wealth of available experimental data (structures of actin

and actin binding proteins, mutational studies), I propose to study using computational techniques the binding of these proteins and their effects on F-actin conformation, dynamics and stability.

I intend first to focus on conformational changes of F-actin that have been observed experimentally. F-actin can be viewed as a right-handed long pitch helix of actin monomers (globular-actin or G-actin). F-actin and G-actin bind a large number of proteins,[34] some of which can affect their conformations. Among such proteins are the actin-depolymerization factor (ADF)/cofilin, which are responsible for rapid rearrangements of actin structures in the cell. The binding of cofilin along the length of the actin filament forces the filament to twist a little more tightly (see Fig 4a, Fig 4b) which could lead to the depolimerization of F-actin.[33,35] The ability of ADF/cofilin to assemble or disassemble F-actin is pH dependent in vitro. It was also recently shown that actin monomers, within a normal F-actin filament, can exist in a state of 'tilt' that involves very different contacts between subunits.[36]

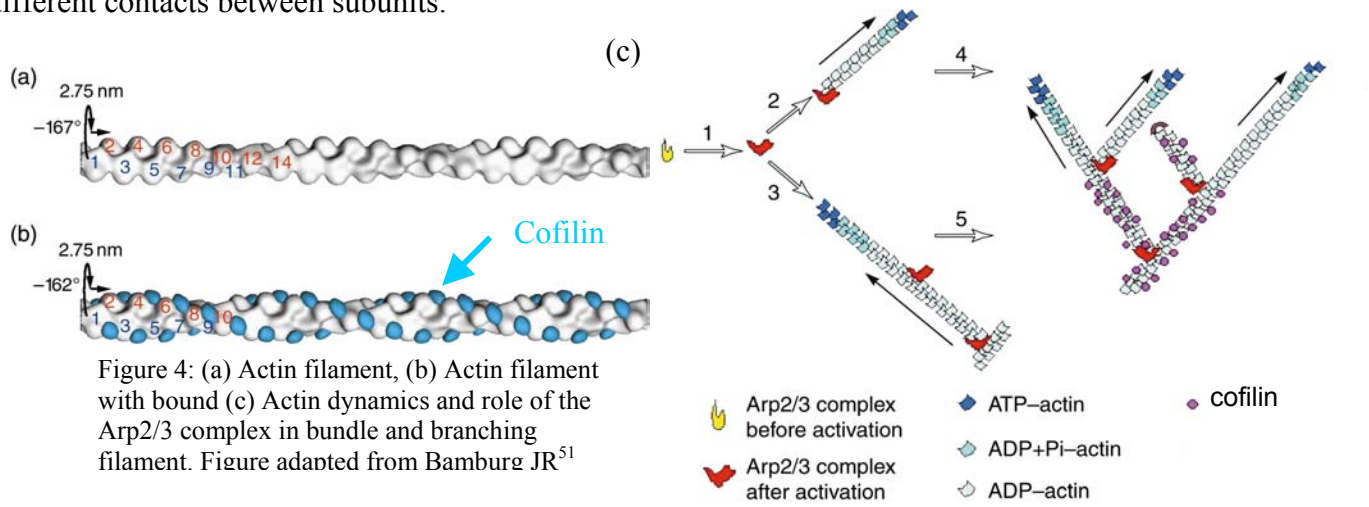

Figure 4: (a) Actin filament, (b) Actin filament with bound (c) Actin dynamics and role of the Arp2/3 complex in bundle and branching filament. Figure adapted from Bamburg JR[51]

Such conformational changes have been observed by cryo-EM and, while atomic models have been constructed, computational studies to address the effect of actin binding proteins on actin filament dynamics have yet to be performed. From my previous normal mode studies, we have seen that an elastic network model is able to capture the mechanical properties of many biological molecules. Moving toward filamentous systems such as F-actin, a helical polymer, the elastic model should remain efficient to investigate its dynamical properties. Indeed, its dynamical properties must be more dominated by elastic properties than a single protein. Therefore, to tackle questions related to the different conformations of F-actin, I will first use the elastic network normal mode analysis with a coarse-grained description to investigate, at the atomic level, the effect of cofilin on the flexibility of F-actin. I will examine the mechanical stress on the filament is induced by cofilin and moreover how the stress imposed on the filament may weaken the contact between the actin monomers and cause depolymerization. To complement these studies, electrostatic calculations using all-atoms models will be performed to elucidate the role of the pH on the assembly/disassembly of F-actin by ADF/cofilin. Such studies would certainly help our understanding of the depolymerization process of the actin filament. These studies will be extended to other actin binding proteins.

Another interesting system that binds F-actin is the Arp2/3 complex. This complex plays a major role in the assembly of actin filaments. It contain at least two binding sites for F-actin, thus facilitating bundles, branching filaments and a three dimensional network of actin (see Fig 4c).[37] Its inactive conformation was only recently described.[38,39] The complex is a large macromolecular assembly of two-actin related proteins and five novel proteins. There is evidence from cryo-EM that the active complex has a conformation more compact than the inactive one.[40] Using the X-ray structure and information from cryo-EM, I intend to model this conformational change using normal mode analysis in order to obtain a better understanding of the

initiation process of the branched actin filament. Once the active complex is determined, electrostatic calculations could be performed to determine which residues are important to stabilize the compact form. Further simulations of F-actin capped with the Arp2/3 complex could then be performed to measure the effect of Arp2/3 on the stability of the F-actin filament.

These studies aim to understand the dynamics of the actin filaments. The methods used and developed for these studies will certainly be useful to address similar questions related to the dynamics of the others cytoskeletal filaments i.e. microtubules and intermediate filaments. Therefore, I intend in the future to extend these studies to microtubules and intermediate filaments.

Long term research goals: Computational approaches to structural proteomics

Structural proteomics of biological complexes is emerging as a new endeavor at the confluence of several scientific areas. An ultimate goal of structural proteomics is the characterization and determination of macromolecular complexes and their functional states. The structures of these complexes are of great interest since many important cellular functions are accomplished by multi-component complexes of molecular machines. Indeed, these machines are involved in replication, transcription, protein synthesis, regulation of cellular transport and other core biological functions.

Structural characterization of such complexes provides us information for understanding their functional roles and mechanisms, which will play an important role in lead discovery, and optimization of new drugs to treat human disease. However, the number of solved structures of such systems is still small compared to the number of possible complexes indicated from genomics studies. Structure determination of these macromolecules is an extremely difficult task. Therefore, the demand for computational algorithms for macromolecular structure determination continues to increase.

I intend to orient my long-term research toward the computational structural determination of large protein assemblies, and subsequent studies of the molecular mechanisms of their biological function. Research during my PhD focused on the machinery of proteins, and through my post-doctoral experience, I have been studying large macromolecular assemblies of proteins and working in close collaboration with experimental groups. These experiences lay the foundation of my research on supramolecular machines.

One of my focuses will be the development of new computational algorithms to construct atomic-level models of large macromolecular complexes. Such macromolecular assemblies comprise a large number of proteins and/or of RNA molecules. Currently, the structure determination by X-ray crystallography of the overall complex is extremely difficult. Still, X-ray crystallography and NMR provide atomic resolution structural data of the independent protein/RNA components. Thus, it is beneficial if an atomic model of the complete assembly could be constructed by assembling the high-resolution structures of each component. Theoretically, one should be able to reconstruct a structural model of the overall macromolecular assembly if all of the components are known. However, considering the level of structural complexity of these macromolecular machines, one needs to integrate information from several experimental and theoretical sources in order to assemble proteins and/or RNA into a correct structural model.[41,42] Cryo electron microscopy (EM) plays a significant role in providing information about the structure of these assemblies, the ribosome being an example.[11] Electron tomography is also emerging as a new technique, which enables one to obtain low-resolution structural information of large macromolecular assemblies.[43]

A few groups have been developing computational algorithms to construct assemblies from their individual components. Russell and collaborators have taken advantage of bioinformatics and protein-protein interactions of known structures to construct a possible atomic level model for several assemblies.[44]

However, these methods are applicable only when a similar pattern of protein association has been already observed. In addition the flexibility of the proteins are not taken into account in this approach, which can lead to significant clashes between proteins. Models of the 80S and 70S ribosome have also been constructed using comparative homology modeling and fitting into cryo-EM maps.[45,46] These studies were made possible because the structure of each of the two subunits of the ribosome were previously solved at high-resolution. Therefore, an algorithm that does not depend on a known sub-complex is, arguably, more useful. An attempt to construct a model for the nuclear pore complex from a variety of experimental data has been performed by Sali and collaborators.[41,47] However, in this work, the structure of each protein is not taken into account. Because the structure of each protein component is unknown, each is represented by a simple sphere and therefore the result provides only the low-resolution shape of the macromolecular complex; the physical properties of each component are also ignored.

My goal will be the development of new algorithms to construct atomic-level models of large macromolecular complex assemblies from the known x-ray or NMR structure of the individual components, in which experimental input as well as the physical properties of each protein/RNA will to be considered. To obtain a model with a high level of confidence, the research plan proposed here will consist of several developmental stages, each of them being an attempt to validate the approach described below. The development could be first undertaken for systems for which the structure of the macromolecular assembly is already known, to evaluate the robustness of the algorithm. In particular the uniqueness of the obtained model needs to be ensured, therefore we will need to determine which minimal set of experimental constraints are necessary to achieve convergence toward a unique model. Some key aspects of the developments of the algorithm are summarized in the following and in figure 5.

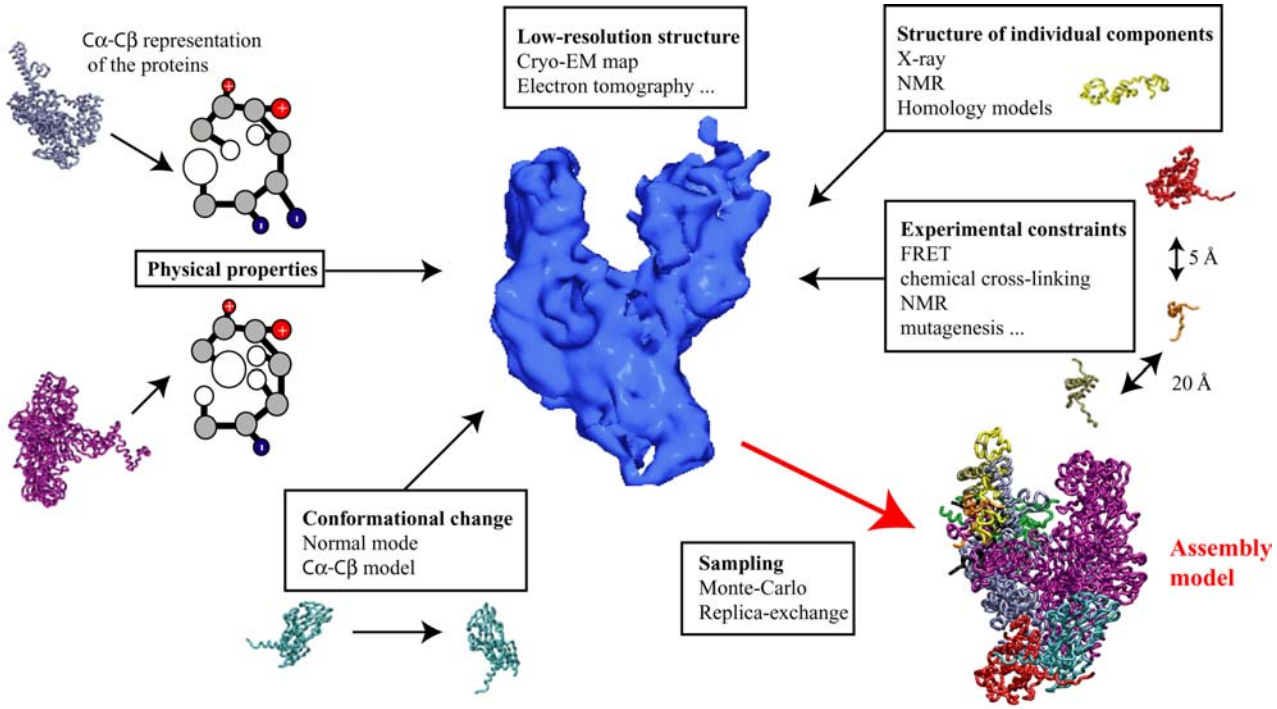

Figure 5: Integration of several experimental data and computational approaches for constructing a model of a complete macromolecular assembly.

- **Fitness to experimental data.** The primary constraint would be the fitness of the atomic structure into the low–resolution structure, i.e. shape of the molecule, as obtained from cryo-EM or electron tomography. Correlation between the experimental density maps and the one simulated from the modeled structure can be used to evaluate the fitness. Another type of constraint would be the

distances between particular parts of proteins. From FRET, cross-linking and mutagenesis experiments information of the inter-component contacts can be obtained. These data provide information on the proximity of each component, which is not explicitly available in the global EM density map.

- **Physical properties of each component.** Physical properties of the association interface between the proteins needs to be considered; in particular, surfaces and electrostatic complementarities between the proteins need to be ensured. Since these systems may have over 200,000 atoms, to optimize interactions between proteins, a simplified force field that takes into account the physical properties of the residues needs to be developed. The intra-protein force field that captures the physical properties of the components should be reproduced by models with differing levels of complexity, e.g. a Cα or Cα-Cβ model. The inter-protein force field will need to reproduce hydrophobic and hydrophilic interactions between proteins as seen with a more detailed force field.
- **Accounting for the flexibility of each component.** We also need to consider the fact that the conformation of a protein in the complex may be different from its known isolated state, which adds an additional degree of complexity. To model the possible conformational change of each protein/RNA component, normal mode analysis can be used. Since protein/RNA conformational change parallels lowest frequency modes, we can represent protein dynamics with a limited number of degree of freedom (~10) instead of the total number of degree of freedom in Cartesian coordinates. In this way, large conformational changes of each component can be sampled efficiently. On the other hand, to refine the contacts at the interface, standard molecular dynamics would be suitable.
- **Algorithms to sample possible assembly configurations.** All the properties presented above need to be considered to construct a possible model of the complex assembly. To achieve this goal, a large number of assembly configurations need to be sampled efficiently. To optimize the fitness of the model to the experimental data as well as the complementarity in protein interactions, a Monte-Carlo procedure can be employed using the conformation of the protein/RNA components and the rotation/translation of those as variables to change. In addition, the replica-exchange algorithm can be used to increase the efficiency of the sampling. Uniqueness of the resulting model would need to be examined.

I am convinced that progress on the study of such large macromolecular assemblies can only be achieved by combining several computational approaches and experimental techniques at different levels of resolution. Therefore, I am eager to establish active collaboration with experimental groups.

The projects presented above are devoted to the development of methodologies to utilize experimental data. In addition, I would like to approach the assembly process of large macromolecular assemblies from a more theoretical perspective. Theoretical structural predictions require a detailed understanding of the chemical and physical principles that dictate molecular interactions. These studies are necessary to comprehend how each piece of a macromolecular assembly comes together to form a functional complex. Addressing these questions will also provide insights that will be useful for the development of modeling algorithms discussed above.

Another interesting aspect is the dynamics of association and dissociation processes of the components of the complex during the expression of its function. As the allostery of kinases is controlled by ligands, large biological machines are often controlled by proteins. While the allosteric mechanisms of even small proteins are still not clear, control mechanisms of large biomolecular machines are even more challenging and could be another important theme in biophysics. I intend to address this question more generally in my future research.

The ultimate purpose of this line of research is to progress toward an understanding of function and expression of the function of supramolecular complexes. I am convinced that computational approaches have an important role to play in the field of structural proteomics and that new developments in computational tools will greatly help to improve our understanding of the machinery of the cell.

References

1.Goldstein, H., *Classical Mechanics*, 1950.
2.Durand, P., Trinquier, G. and Sanejouand, Y. H., *Biopolymers* 34, 759, 1994.
3.Tama, F., Gadea, F. X., Marques, O. and Sanejouand, Y. H., *Proteins* 41, 1, 2000.
4.Tirion, M. M., *Phys. Rev. Lett.* 77, 1905, 1996.
5.Bahar, I., Atilgan, A. R. and Erman, B., *Fold. Des.* 2, 173, 1997.
6.Tama, F. and Sanejouand, Y. H., *Protein Eng.* 14, 1, 2001.
7.Tama, F., Miyashita, O., Kitao, A. and Go, N., *Eur. Biophys. J. Biophys. Lett.* 29, 472, 2000.
8.Lee, K. K. and Johnson, J. E., *Curr. Opin. Struct. Biol.* 13, 558, 2003.
9.Bancroft, J. B. and Heibert, E., *Virology* 32, 354, 1967.
10.Tama, F. and Brooks III, C. L., *J. Mol. Biol.* in press, 2004.
11.Frank, J. and Agrawal, R. K., *Nature* 406, 318, 2000.
12.Tama, F., Valle, M., Frank, J. and Brooks, C. L., III, *Proc. Natl. Acad. Sci. USA,* 100, 9319, 2003.
13.Trylska, J., Konecny, R., Tama, F., Brooks III, C. L. and Mccammon, J. A., *Biopolymers* 74, 423, 2004.
14.Tama, F., Feig, M., Liu, J., Brooks III, C. L. and Taylor, K. A., *J. Mol. Biol.* in press, 2004.
15.Wendt, T., Taylor, D., Trybus, K. M. and Taylor, K., *Proc Natl Acad Sci USA* 98, 4361, 2001.
16.Saibil, H. R., *Nat. Struc. Biol.* 7, 711, 2000.
17.Wriggers, W. and Birmanns, S., *J. Struct. Biol.* 133, 193, 2001.
18.Tama, F. and Brooks, C. L., III, *J. Mol. Biol.* 318, 733, 2002.
19.Chacon, P., Tama, F. and Wriggers, W., *J. Mol. Biol.* 326, 485, 2003.
20.Tama, F., Miyashita, O. and Brooks III, C. L., *J. Mol. Biol.* 337, 985, 2004.
21.Tama, F., Miyashita, O. and Brooks III, C. L., *J. Struct. Biol.* 147, 315, 2004.
22.Falke, S., Tama, F., Brooks III, C. L., Gogol, E. P. and Fisher, M. T., *submitted*, 2004.
23.Go, N., Noguti, T. and Nishikawa, T., *Proc Natl Acad Sci USA* 80, 3696, 1983.
24.Brooks, B. R. and Karplus, M., *Proc Natl Acad Sci USA* 80, 6571, 1983.
25.Brooks, B. R. and Karplus, M., *Proc Natl Acad Sci USA* 82, 4995, 1985.
26.Seno, Y. and Go, N., *J. Mol. Biol.* 216, 95, 1990.
27.Tama, F., Wriggers, W. and Brooks, C. L., III, *J. Mol. Biol.* 321, 297, 2002.
28.Valle, M., Zavialov, A., Li, W., Stagg, S. M., Sengupta, J., Nielsen, R. C., Nissen, P., Harvey, S. C., Ehrenberg, M. and Frank, J., *Nat. Struc. Biol.* 10, 899, 2003.
29.Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M. and Frank, J., *Cell* 114, 123, 2003.
30.Rawat, U. B. S., Zavialov, A. V., Sengupta, J., Valle, M., Grassucci, R. A., Linde, J., Vestergaard, B., Ehrenberg, M. and Frank, J., *Nature* 421, 87, 2003.
31.Zhang, G. Y., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. and Darst, S. A., *Cell* 98, 811, 1999.
32.Murakami, K. S., Masuda, S. and Darst, S. A., *Science* 296, 1280, 2002.
33.Bamburg, J. R., McGough, A. and Ono, S., *Trends Cell Biol.* 9, 364, 1999.
34.Dos Remedios, C. G., Chhabra, D., Kekic, M., Dedova, I. V., Tsubakihara, M., Berry, D. A. and Nosworthy, N. J., *Physiol. Rev.* 83, 433, 2003.
35.McGough, A., Pope, B., Chiu, W. and Weeds, A., *J. Cell Biol.* 138, 771, 1997.
36.Galkin, V. E., Orlova, A., Lukoyanova, N., Wriggers, W. and Egelman, E. H., *J. Cell Biol.* 153, 75, 2001.
37.Pollard, T. D., Hanein, D. and Volkmann, N., *Nature* 418, 915, 2002.
38.Robinson, R. C., Turbedsky, K., Kaiser, D. A., Marchand, J. B., Higgs, H. N., Choe, S. and Pollard, T. D., *Science* 294, 1679, 2001.
39.Turbedsky, K., Robinson, R., Kaiser, D. A., Higgs, H. N., Marchand, J. B., Choe, S. and Pollard, T. D., *Mol. Biol. Cell* 12, 396A, 2001.
40.Volkmann, N., Amann, K. J., Stoilova-McPhie, S., Egile, C., Winter, D. C., Hazelwood, L., Heuser, J. E., Li, R., Pollard, T. D. and Hanein, D., *Science* 293, 2456, 2001.
41.Sali, A., Glaeser, R., Earnest, T. and Baumeister, W., *Nature* 422, 216, 2003.
42.Russell, R. B., Alber, F., Aloy, P., Davis, F. P., Korkin, D., Pichaud, M., Topf, M. and Sali, A., *Curr. Opin. Struct. Biol.* 14, 313, 2004.
43.Baumeister, W., *Curr. Opin. Struct. Biol.* 12, 679, 2002.
44.Aloy, P., Bottcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., Fischer, S., Gavin, A. C., Bork, P., Superti-Furga, G., Serrano, L. and Russell, R. B., *Science* 303, 2026, 2004.
45.Gao, H. X., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S. M., Van Roey, P., Agrawal, R. K., Harvey, S. C., Sali, A., Chapman, M. S. and Frank, J., *Cell* 113, 789, 2003.
46.Spahn, C. M. T., Beckmann, R., Eswar, N., Penczek, P. A., Sali, A., Blobel, G. and Frank, J., *Cell* 107, 373, 2001.
47.Alber, F., Eswar, N. and Sali, A., *Structure determination of macromolecular complexes by experiment and computation*, Vol 15, Springer-Verlag, Berlin Heidelberg, 2004.