Large Scale Simulation of Protein Mechanics and Function

Emad Tajkhorshid, Aleksij Aksimentiev, Ilya Balabin, Mu Gao, Barry Isralewitz, James C. Phillips, Fangqiang Zhu, and Klaus Schulten*

March 19, 2003

Theoretical and Computational Biophysics Group
Beckman Institute for Advanced Science and Technology
University of Illinois

*Corresponding author

Abstract

The biomedical sciences are enjoying a wealth of data from innovative new techniques in structure determination, DNA sequencing at the level of entire genomes, and direct manipulation and observation of single molecules. Computational biologists seek to extend and refine these experimental advances through atomic level modeling, often in direct collaboration with experimental researchers. Systems that capture the interest of experimental and computational biologists today are membrane channels, bioenergetic proteins like ATPase, or the machinery of the expression system like the ribosome, all of which should be simulated in realistic environments. Combining expertise from the physical, computational and life sciences, computational biologists have developed for this purpose modeling tools for biomolecular systems of 100,000 atoms and more. They have also advanced mathematical modeling using stochastic and non-equilibrium statistical mechanical methods to bridge the still huge gap between simulation (10 ns) and physiological times (1 ms). This article demonstrates some of the progress achieved so far and the discoveries made.

Introduction

During the past five years, life science research has been almost completely transformed. With the sequencing of the genomes of many organisms, in particular that of man, there has emerged a new view of organisms as a networked system of metabolic and signaling pathways. Today, organisms and cells are approached as a healthy functioning whole; studies of disease and therapy focus on failure and repair of pathways. The computer has become an indispensable instrument of all researchers confronted with the vast new data and their integration.

However, the same data remind biomedical scientists that for most proteins structures, functions, mechanisms, and pathways are still elusive, and that Science needs to make a strenuous, concerted effort to gain this missing knowledge. Biomolecular modeling is a key methodology that can bridge the knowledge gap and provide atomic level pictures of biomolecular processes. It is adopted today increasingly by computational as well as experimental biologists.

Proteins, the products of genomes, can only be completely understood in the context of their cellular environment. One must not lose sight of the fact that the genome only partly defines a cell; Virchow's famous 19th century statement,"It takes a cell to make a cell," is still true today [48]. This suggests that modelers embed proteins in their proper cellular environment, e.g., membranes and water, to study function and resolve mechanisms. Including the cellular environment is always computationally costly, but the dramatic increase of computing power makes this feasible today. In fact, the authors have already completed numerous molecular dynamics studies of proteins in membrane and aqueous environments, the most recent one reported here including 327,000 atoms and covering over 10 ns. A list of large scale simulations performed in the authors' group during the past decade is presented in Table 1. The list illustrates the progress made: ten years ago, to realize a then-heroic 27,000 atom 200 ps simulation required building a special purpose computer and over a year of run time; today, simulation for the same period of a much larger system of 270,000 atoms can be accomplished on 128 processors of a widely available machine within a day.

The increase in computing power stems from commodity workstation clusters with 10-100 processors and from very large parallel machines with thousands of processors. To harness these resources effectively required great effort. The authors' group has developed for this purpose the molecular dynamics program NAMD [74]. For the graphical analysis of the ensuing Gigabytes to Terabytes of data the group provides the program VMD [55]. These programs are widely used today since the increasing availability of protein structures has lead to most biomedical researchers using structural information for the design and analysis of their experiments. The next section will

Table 1: Large scale simulations performed by the authors. The size refers to the number of atoms simulated. The following abbreviations are used: SMD, Steered Molecular Dynamics; FMA, Fast Multipole Algorithm; NC, no cut-off; PME, Particle–Mesh Ewald. TCS and Platinum are computer clusters at the Pittsburgh (PSC) and Urbana (NCSA) supercomputing centers, respectively.

System	**Simulation**	**Size**	**Platform (# proc.)**	**Year, Ref.**
bilayer of 200 lipids	NVE, NC	27K	60 node transputer	1993, [49]
membrane-water interface	FMA	32K	SGI Crimson	1995, [150]
binding of estrogen receptor to DNA	FMA, NVT	36K	HP cluster (8)	1997, [78]
apolipoprotein A-I	NVE	46K	HP cluster (4)	1997, [107]
calmodulin	NVE	33K	T3D (64)	1998, [143]
Rieske subunit motion in cytochrome bc_1 complex	SMD	91K	T3E (64)	1999, [60]
bacteriorhodopsin and purple membrane	PME, NVE, NPT	24K	Alpha cluster (8)	2001, [7]
MscL	PME, NPT	55K	T3E (64)	2001, [47]
aquaporin-1	PME, NPT	60K	T3E (64)	2001, [152]
BamH1 endonuclease binding to DNA	PME, NPT	65K	T3E (64)	2001, [88]
photosynthetic light harvesting system	PME, NPT	87K	SGI Origin 2000 (4)	2001, [23]
fibronectin type III_{10}	PME, SMD	126K	Linux cluster (32)	2002, [39], 2003, here
aquaglyceroporin	PME, NPT	106K	TCS (128)	2002, [130] 2003, here
rhodopsin	PME, NPT	40K	T3E (128)	2002, [120]
CD2-CD58 complex	PME, SMD	91K, 104K	Linux Cluster (32)	2003, [8]
F_0-ATPase	PME, NPT, SMD	112K	T3E (128), TCS (256), Linux cluster (32)	2003, here
F_1-ATPase	PME, NPT, SMD	327K	Platinum (448), TCS (512)	2003, here

describe briefly NAMD and VMD.

The subsequent sections will illustrate three exemplary simulations of large scale systems that are presently the subject of intense research. The wide interest in these systems and the exciting research challenges they pose are the main reason why one feels compelled to model such large, cumbersome systems. The first example describes simulation of a membrane channel, aquaglyceroporin, that conducts water and small carbohydrates across the cell membrane. Placement of the protein into a membrane environment leads in this case to a 106,000 atom simulation. The second example is a study of ATP synthase, the renowned large protein complex that links an electro-mechanical motor, the F_0 subunit, with a chemo-mechanical motor, the F_1 subunit, through a rotatory elastic axle, the stalk. In this case the two subunits embedded in their appropriate membrane and aqueous environments comprise 112,000 and 327,000 atoms, respectively. The third example investigates a module of the extracellular matrix protein fibronectin, FN-III_{10}, that is involved in cell adhesion and motion. The protein has optimal mechanical elasticity and signals to cell surface receptors, integrins, the tension exerted on it. Atomic force microscopy studies suggest that the FN-III_{10} module, under the influence of mechanical stretching, unfold completely to a length of about 300 Å. Simulation of such unfolding event in which water plays a key role requires an appropriately long (370 Å) water box leading to a system of 130,000 atoms. The examples illustrate that the simulation size is mainly determined by the choice of environment, not by the actual protein simulated.

The functions of all three proteins investigated are mainly mechanical. To relate the observed mechanics to the architecture of proteins is the domain of so-called steered molecular dynamics simulations, contributing to the founding of the new field of mechanobiology, which studies the role of forces in cellular processes. Here forces appear as key ingredients of processes in cells: as "substrates" that drive reactions, as "products" of molecular motors, and as "signals". All modeling examples presented below employ steered molecular dynamics to accelerate the underlying processes.

Technology for Simulation and Visualization of Large Biomolecular Systems

With continuing increases in high performance computing technology, the domain of biomolecular simulation has rapidly expanded from isolated proteins in solvent to biomolecular complexes in their

native environments. Figure 1 compares simulation sizes of well-known biomolecular systems that have been studied by MD: bovine pancreatic trypsin inhibitor, the first protein simulated [93], a pair of DNA binding domains of the estrogen receptor complexed with DNA [10] and the F_1 fraction of ATP synthase (this paper). Today, simulations of biomolecular systems with over 100,000 atoms lasting 10 ns have become routine [26, 130] and discoveries of the structures of large biomolecular complexes like the RNA polymerase II [44] or the 70S ribosome complex made from the 30S [21] and 50S [5] subunits forecast simulations involving a few million atoms. Such simulations become feasible only through the use of large parallel computers and the development of molecular dynamics programs that run efficiently on such machines. 100,000 atom systems can be simulated with such programs very cost effectively on clusters of commodity workstations; for example, a cluster of 24 PCs running Linux and connected by commodity switched gigabit ethernet can be constructed for under $1000 per processor and utilized by the new generation of molecular dynamics programs at over 70% parallel efficiency, achieving a nanosecond or more of simulation time per week.

One such simulation program is NAMD [74], which has been developed specifically to simulate large biomolecular aggregates on both commodity clusters and massively parallel supercomputers. NAMD is primarily designed to work with CHARMM force field parameters [90]. Input files for NAMD can be generated using CHARMM [17], X-PLOR [18], or VMD [55]. NAMD has also been extended to read AMBER [141] and GROMACS [83] input file formats. NAMD running on 768 processors has enabled a record-breaking two million atom 5 ns simulation of the ribosome using an AMBER force field (K. Sanbonmatsu, private communication). The program NAMD has been used for the simulations described below.

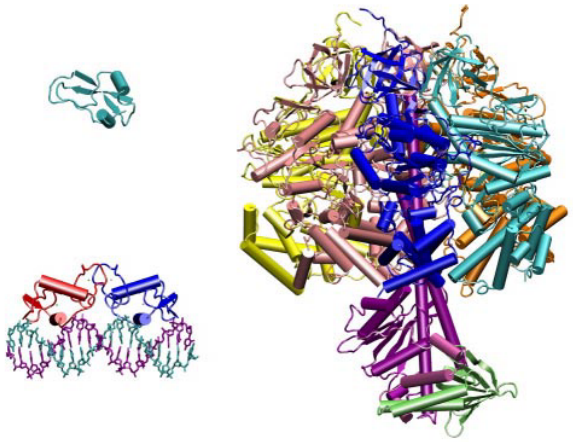

Figure 1: Simulations have increased in size, from bovine pancreatic trypsin inhibitor (upper left, about 3K atoms), through the estrogen receptor (lower left, 36K atoms), to F_1-ATPase (right, 327K atoms,). (Atom counts include solvent.)

NAMD provides the complete functionality needed to carry out standard molecular dynamics simulations. Efficient conjugate gradient minimization, fixed atoms and harmonic restraints are among the methods available for the initial assembly of aggregates. The resulting system

may be equilibrated and simulated in constant temperature and pressure ensembles, with periodic boundary conditions and full long-range electrostatics. NAMD supports steered molecular dynamics (SMD)[57, 61], a computational technique that applies external forces and accelerates reaction events to the nanosecond timescale, as illustrated in the examples below.

NAMD employs several methods to reduce the amount of calculations required by the MD algorithm. First, the complexity of the long-range electrostatic force evaluation is reduced from $O(N^2)$ to $O(N \log N)$ via the particle mesh Ewald (PME)[24] algorithm, which combines a cutoff direct calculation with an FFT-based mesh calculation. Second, nonbonded interactions are split into a smooth long-range component and a complementary short-range component; the period of long-range (electrostatic) evaluation is then increased from every 1 fs to every 4 fs via an impulse-based symplectic multiple timestepping method. Finally, either the period of short-range nonbonded force evaluation may be increased to 2 fs via multiple timestepping or the timestep itself increased to 2 fs by constraining the lengths of the highest frequency bonds. Timestep limits are imposed by the requirement of energy conservation (larger values result in energy drift); the use of a symplectic integrator allows us to interpret energy conservation as an indicator of accuracy.

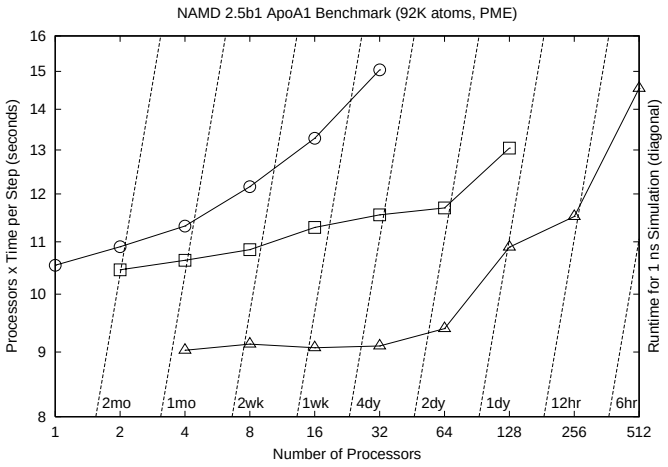

Figure 2: Computer resources consumed per timestep by NAMD for a 92,000 atom benchmark simulation using 1 fs timesteps and a 12 Å cutoff with PME, long range Coulomb forces being updated every 4 steps. Smaller numbers indicate better performance, perfect linear scaling with processor number is a horizontal line. Dotted diagonal lines indicate the runtime required for 1 ns of simulation. Machines tested are: circles: our group's single-processor 1333 MHz Athlon Linux cluster with switched 100 Mbit ethernet network; squares: National Center for Supercomputing Applications "Titan" dual-processor 800 MHz Itanium Linux cluster with Myrinet network; triangles: Pittsburgh Supercomputing Center "Lemieux" quad-processor 1000 MHz Alpha cluster with Quadrics network. One can recognize that NAMD running on 32 processors of our Linux cluster requires under 1 week run time for 1 ns.

The efficiency of NAMD on today's parallel computers is illustrated in Fig. 2. This efficiency is achieved through use of the Charm++/Converse parallel runtime system [75]. NAMD employs three basic techniques for this purpose: spatial domain decomposition, measurement-based load

balancing, and message-driven execution.

In NAMD, the region of space occupied by the atoms in a simulation is divided into regularly sized regions, and these regions are assigned to processors so as to distribute the data evenly. Processors may be assigned multiple regions, or there may be more processors than regions to distribute. All required force calculations in the simulation, other than the FFT-based PME reciprocal sum, are between atoms in neighboring regions. This serves to minimize communication, which often limits scalability in other programs as the number of atoms in the simulation grows.

In order to distribute the time-consuming force calculations evenly among the available processors, the calculations are divided into sets, such as the set of interactions between atoms in a particular pair of regions. These sets are initially assigned to processors so as to minimize the number of copies of each region's positions and forces that must be transmitted. During the initial stages of the simulation, the runtime of each set is measured and used to redistribute the calculations more evenly. This process is repeated at regular intervals to maintain performance during the entire simulation.

The order of force calculation in NAMD is not fixed, but rather each set is added to a queue as the required data from the regions it depends on becomes available. The data may arrive from another processor, or result from the integration of atomic coordinates and velocities locally. Calculations that generate results for atoms on other processors are assigned a higher priority. This mechanism allows NAMD to keep all processors doing useful work despite network latency or other minor delays.

Molecular modeling relies heavily on visualization technology. The simulations described below relied on the molecular visualization and analysis program VMD [55]. VMD was employed as a tool for building complete 3D structures of biomolecular aggregates. For example, using VMD, researchers embedded the structure of GlpF (see below) into the membrane and covered it with water. VMD permitted researchers to visualize the trajectories resulting from the simulations and to analyze the data in terms of root-mean square deviation (RMSD), the thickness of the membrane, and the flux of water molecules through channels. VMD permits the visualization of very large structures, for example, of a two million atom ribosome simulation (K. Sanbonmatsu, private communication). Systems of 100,000 atoms can be readily viewed with VMD on a modern desktop or laptop computer.

Aquaporins – Membrane Water Channels

In mammalian cells, water, driven by osmotic forces, can pass indiscriminately through the cell lipid membrane, but the conduction is too slow for an active regulation of water homeostasis. Membrane water channels that permit a rapid flux had been previously postulated on these grounds and were identified by Agre and coworkers [112], who termed them *aquaporins* (AQPs). This discovery promised insight into the fundamental physiology of water balance and the pathophysiology of water balance disorders. Soon after, it became evident that AQPs were abundantly present in all kingdoms of life, including mammals, amphibia, insects, plants, and bacteria [2, 12, 52]. In the human genome, eleven different genes have been identified which encode AQPs in organs that are either directly involved in handling large volumes of water, such as the kidneys, or need to continuously maintain a precise level of water, such as the brain. Impaired function of AQPs has been related to many pathophysiological situations such as nephrogenic diabetes insipidus and congenital cataract [2, 12, 28, 82].

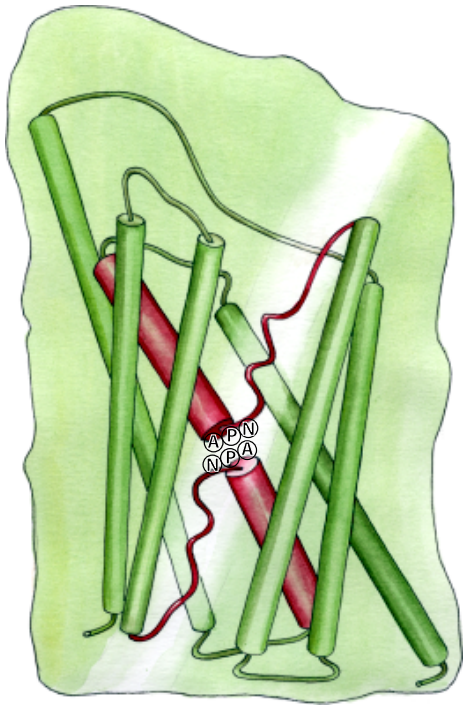

Figure 3: Architecture of an AQP channel. Six transmembrane α-helices (green) form a bundle inside the membrane. Two reentrant loops (red) then fold into the channel and complete the interior of the channel. The loops enter the channel in helical forms, meet each other at the center of the channel, and turn back in an energetically unfavorable extended form. The two conserved NPA motifs, which are located between the loops and short helices, provide multiple hydrogen bonds that keeps the loops together. The pathway of substrate transport is along the nonhelical halves of the reentrant loops (light area).

AQPs are open channels by design; no gating mechanism has been reported so far for water, however the water pore in most AQPs can be blocked by mercurial compounds [134]. AQPs are highly selective channels. Despite their high permeability to water, they completely exclude protons, a puzzling peculiarity since protons can be readily transported through a hydrogen bonded chain of water molecules according to the Grotthuss mechanism. Under physiological conditions, AQPs are water pores impermeable to ions and other charged species [3, 121, 147]. The exclusion of ions, in particular protons, is central to the biological function of AQPs, enabling them to transport high volumes of

water without dissipating the electrochemical potential of cell membranes.

A subfamily of AQPs, known as *aquaglyceroporins*, permit a highly stereoselective passage of small sugar molecules, such as glycerol [13, 45, 50]. The *E. coli* glycerol uptake facilitator, GlpF, is a prominent member of the subfamily, partly because it is the first AQP for which a high resolution structure was solved by x-ray crystallography [37]. GlpF provides the cell with efficient access to extracellular carbon sources and is particularly important under low sugar concentration conditions [118].

The architecture of AQPs was elucidated first by electron microscopy, revealing that AQPs form tetramers in the membrane [27, 137]. Further studies showed that the protein is composed of four functionally independent aqueous pores [70, 126]. The folded monomer is a right-handed helix bundle of six transmembrane α-helices, and includes two reentrant loops that meet each other at the center of the channel [134, 136]. A schematic presentation of an AQP monomer is provided in Fig. 3. The core of the channel is formed by the reentrant loops comprised of two α-helices (cylinders in Fig. 3) that protrude into the protein and return in the form of an inverted helix. The latter exposes the backbone atoms to the interior of the channel that guide single file water and glycerol through the channel [66]. The two half helices are held together by multiple hydrogen bonds between the two highly conserved NPA motifs (Asn-Pro-Ala; see Fig. 3) [117, 142].

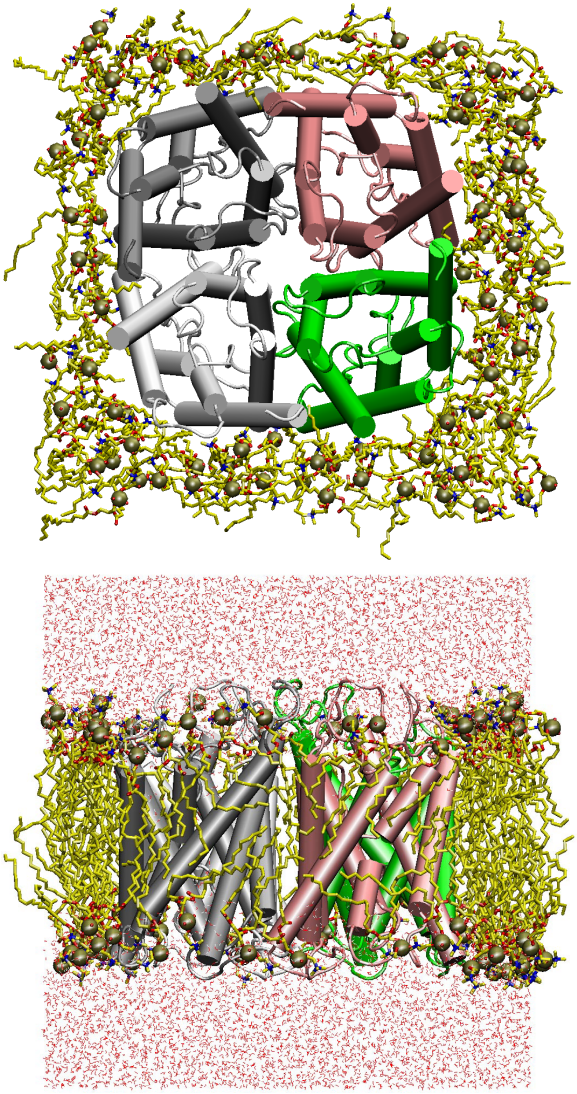

Figure 4: Top and side views of the tetrameric model of an AQP modeled in a fully hydrated patch of lipid bilayer (water molecules are not shown in the top view) [152]. The monomers are shown using cartoon representations and in different colors. In the side view (bottom), the position of phosphate groups of lipids are specified using vdW representation.

At the present time, crystallographically solved high resolution atomic structures are available only for two members of the AQP family, aquaporin-1 (AQP1) [128] and GlpF [37]. As described

below, MD simulations of these structures have significantly contributed to our understanding of the molecular basis of function and selectivity in this increasingly important family of membrane channels.

The simulations described in this section involve a membrane-embedded AQP tetramer solvated by two layers of water molecules, representing the functional form of the channel. This resulted in system sizes of more than 100,000 atoms (Fig. 4). Due to the biological function of the channel, water transport, water molecules needed to be explicitly treated in the model. All calculations described in this section lasted a few nanoseconds with an integration time step of 1 fs. Since water permeation occurs on a nanosecond time scale in AQPs, multinanosecond simulations were needed for sufficient sampling of this event. Using the program NAMD described above, it took about 10 days to complete 1 ns on a 32 (1.1 GHz Athlon) processor cluster. The same calculation took 1.2 days on 128 (1 GHz Alpha) processors of the Lemieux cluster at the Pittsburgh Supercomputing Center.

In order to best describe aquaporins under physiological conditions, the following methodologies were applied. First of all, although the system contains over 100,000 atoms, it is still much smaller than macroscopic systems in the cell and, hence, may suffer from boundary effects, such as artificial surface tension. The solution to this problem is to adopt periodic boundary conditions (PBC), where the basic system is a unit cell, but the actual system is infinite and periodic, obtained by copying and translating the unit cell in three dimensions. Applying PBCs is especially crucial for systems in which the membrane is essentially an infinite plane separating water molecules on its two sides. Langevin Dynamics was used to keep each part of the system at a desired constant temperature (310 K). The Langevin piston method was employed to maintain the pressure of the system at 1 atm. Such so-called NPT ensemble conditions represent a physiological environment. An efficient algorithm, PME, was used to calculate electrostatic forces without cut-off.

The first attempts of simulating AQPs [25, 152] were based on the medium-resolution electron microscopy structures of AQP1 [97, 115]. AQP1 was modeled by us in its tetrameric form embedded in a solvated palmitoyl-oleoyl-phosphatidyl-choline (POPC) lipid bilayer [152]. Nanosecond simulation [152] of the system for the first time revealed the spontaneous formation of a single file of water inside the channel, a feature which is very important for the selectivity of AQPs, as will be described later [130]. Examination of critical regions of the channel, particularly the highly conserved NPA motifs at the center, however showed that the structures were not stable enough for full atomic simulations. The instability was quite evident from the disrup-

tion of the hydrogen bonds between the two NPA motifs (Fig. 5) during the simulation [152]. Similar conclusions about the structural details inside the channel were made by other researchers [25] who reported a refined model of AQP1 on the basis of the high-resolution structure of GlpF [37], and were confirmed by comparison with a high resolution structure of AQP1 [128], which was solved later.

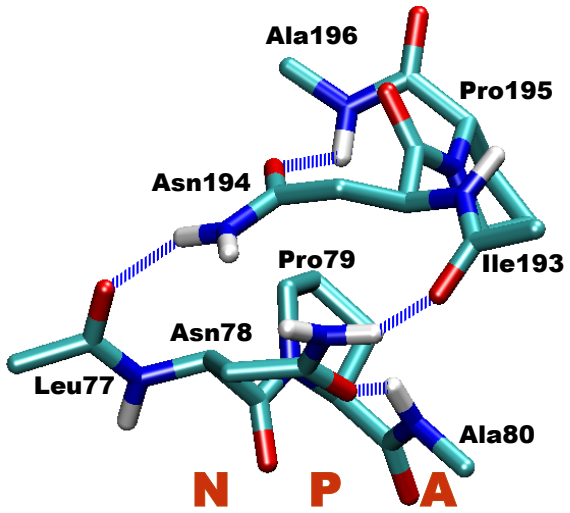

Figure 5: Hydrogen bonds between the two NPA motifs are essential for the stability of the structure and for the function of AQPs. Due to two stable hydrogen bonds between the amido group of the asparagine side chain of each NPA motif with neighboring side chains, one of the amido hydrogens of the asparagine is restrained to be fully exposed toward the interior of the channel, where it forms hydrogen bonds with the permeating substrate. Hydrogen bonds between the two NPA motifs are also important for the stability of the two reentrant loops.

Due to the above mentioned problems in the simulation of AQP1 models, the focus of our MD simulations was changed from AQP1 to GlpF, as soon as the high resolution (2.2 Å) crystallographic structure [37] became available. GlpF, which is an aquaglyceroporin, was crystallized in the presence of a high concentration of glycerol and, therefore, included three glycerol molecules in the pore region of each monomer [37]. We simulated [66] the tetrameric model of GlpF embedded in a palmitoyl-oleoyl-phosphatidyl-ethanolamine (POPE) bilayer, which is the best model for an *E. coli* cell membrane. The simulation included more than 106,000 atoms and extended over several nanoseconds [66].

During the simulations, thermal fluctuations promoted significant movements of individual glycerol molecules inside the four channels and allowed us to describe the complete conduction pathway through the channel, a finding that for the first time explained the presence of peculiar secondary structure elements in AQPs [66]. The discovered conduction pathway is mainly formed by the backbone carbonyl groups of the two non-helical parts of the reentrant loops. Together with a highly conserved pair of asparagines at the center of the channel (the NPA motifs) and an arginine located at the extracellular half channel, the carbonyl groups provide hydrogen binding sites that guide the substrate through an otherwise hydrophobic channel in a stepwise fashion [66]. In a helical conformation,

these carbonyl groups would be mainly involved in intra-helical hydrogen bonds stabilizing the helical arrangement and would not be readily accessible by the substrate permeating the channel. Therefore, the protein has to adopt a non-helical form, although energetically unfavorable, to optimize its functionality. It is interesting to note that the non-helical arrangement of these loops is stabilized through hydrogen bonds between the backbone amino groups of each loop and one conserved glutamate (Fig. 6). The glutamate residues are the only negatively charged residues in the transmembrane region of the protein, a fact that clearly reflects their important role in the structure and function of the channel. Water molecules accompanying glycerol in the channel proved to be very important for the conduction process, lubricating the movement of glycerol through competition for the hydrogen binding sites [66].

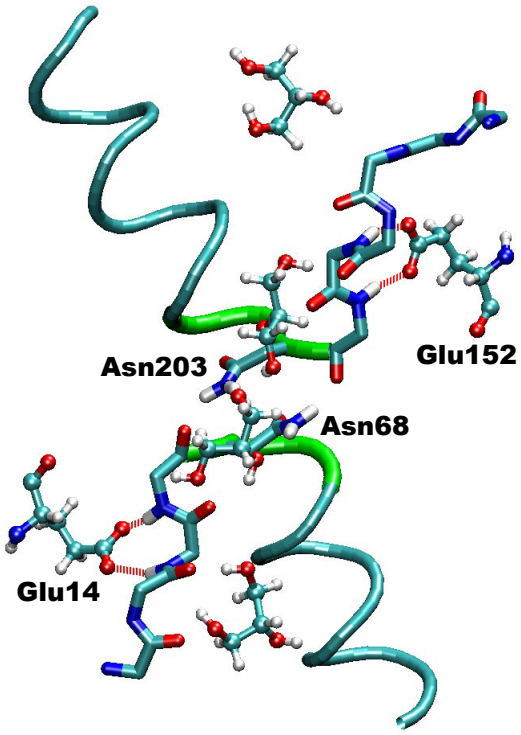

Figure 6: The curve-linear pathway of substrate conduction in AQPs formed by backbone oxygen atoms of reentrant loops. Carbonyl oxygens are shown in red. This peculiar secondary structure of the non-helical halves of the loops is stabilized by hydrogen bonds with Glu14:O_ϵ and Glu152:O_ϵ atoms. Also shown are the location of the NPA motifs (green), the α-helical parts of the reentrant loops (blue), and the side-chains of Asn68 and Asn203 at the center of the channel. Snapshots of four glycerol positions (shown in ball and stick representation) from the simulations are included to illustrate the conduction pathway.

In order to describe the energetics of glycerol conduction and to study the stereoselectivity of the channel, we performed a series of steered MD (SMD) simulations on the tetrameric GlpF model described [65]. In each run, a steering force was applied to individual glycerol molecules (one per monomer) to accelerate the trans-channel conduction in either direction and for two orientations of glycerol, i.e., head first and tail first. Conduction of glycerol occurs on time scales which are not accessible to equilibrium MD simulations, and therefore one needs to accelerate the event via external forces. By employing a simulation scheme that kept the system close to equilibrium and utilizing Jarzynski's

identity [62, 63] linking free energy and irreversible work, we reconstructed the potential of mean force (PMF) from an ensemble of trajectories [65], as shown in Fig. 7. The resulting free energy profile, one of the first ones determined for a biological system, captures major features of the glycerol-channel interaction. The positions of crystallographically observed glycerol molecules [37] closely correspond to minima in the energy profile. Several additional minima, due to multiple hydrogen bonds between glycerol and the channel, were also discovered. The largest barrier against the substrate conduction [65] was found to be located in a region that was suggested to function as the selectivity filter [37].

The constructed PMF (see Fig. 7) exhibits a clear asymmetry. Comparing the cytoplasmic and periplasmic vestibules of the channel, the periplasmic mouth of the channel is characterized by a deep potential well. Although GlpF is a passive channel and in principle, nutrient molecules can be transported either way across the membrane, the asymmetry reflected by PMF might be of importance for the dominant physiological role of the channel, namely uptake of glycerol from the periplasmic space. After entering the cytoplasmic region, glycerol becomes phosphorylated and cannot exit the cell. Interestingly, the asymmetric PMF corresponds to the asymmetric shape of the channel, when one compares the periplasmic and cytosolic segments of the channel. The significantly bigger protrusion of GlpF to the periplasmic region can provide an attractive site for nutrient molecules, which are usually not abundant. Kinetic rate models built up on the basis of the calculated PMF indicate that the attractive well in the periplasmic vestibule of GlpF may increase the efficiency of the channel for glycerol uptake, specially at low concentrations of the substrate [85]. In agreement with this suggestion, it is noteworthy that the asymmetry of the structure is much less pronounced in AQP1, a pure water channel, where substrate (water) is always present in the environment at a high concentration.

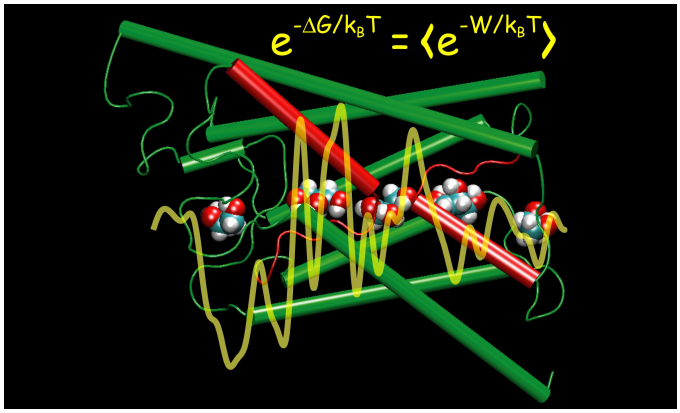

Figure 7: Potential of mean force (PMF) along the pathway of glycerol conduction in GlpF, constructed from SMD simulations [65]. The PMF is superimposed on a GlpF channel, in which several positions adopted by glycerol along its transport are highlighted.

In another set of simulations [130], we removed all glycerol molecules from GlpF and investigated

water permeation through the channel. An extensive set of MD simulations (a total of 15 ns) was performed on the wild-type GlpF and on a designed mutant (see below). During the simulations, single files of water were formed (in 100-200 ps) and maintained in the pore regions of the four channels. Several full permeation events through the 20 Å constriction region of the channels were observed in a 4 ns time frame [130]. The calculated conduction rate of native GlpF (1.1 H_2O/ns/channel), which is very close to the value reported by other researchers [26], compares satisfactorily with the experimentally deduced flux for GlpF, namely, 0.5×10^9 H_2O/s [12, 14, 52].

We also designed and prepared a GlpF mutant in which both the size and polarity of the narrowest part of the channel (the selectivity filter) were increased by mutation of two amino acids. The resultant mutant (W48F/F200T) was found to have an increased permeation rate both in measurements (25%) and in simulations (38%) [130], reflecting a surprisingly close agreement between experiment and theory. In both native and mutant species, water occupancy along the channel axis inferred from the simulations was closely matching the electron density of the crystal structures [130].

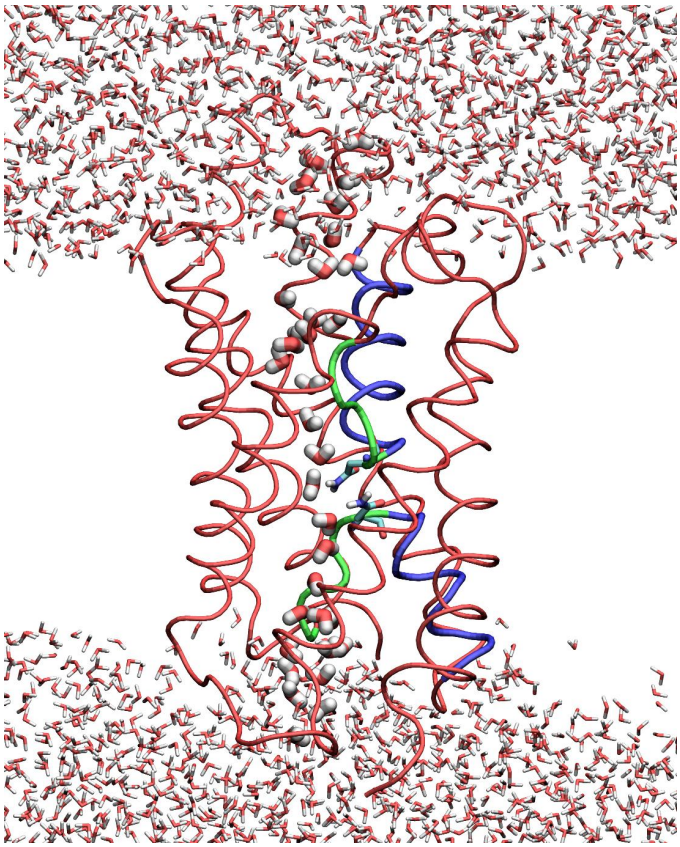

Figure 8: A snapshot from MD simulation of GlpF. Only one monomer is shown. A single file of water forms in the channel during the simulation. The orientation of water molecules in the single file is reversed in the two halves of the channel due to the electric field of the protein. This bipolar configuration of water (wing configuration) prevents proton conduction [130].

Detailed analysis of the dynamics of water in the channel lumen in the simulations answered a long standing puzzle in the function of AQPs. The simulations showed how AQPs employ a global tuning mechanism which permits the fast transport of water, but prevents water-mediated proton transfer. The electric field of the protein dictates a peculiar configuration of water molecules inside the channel that was not reported in any other system before [130]. Starting from the NPA center, water molecules are oriented in opposite directions in the two halves of the channel, with

their hydrogen atoms pointing toward the exits as shown in Fig. 8. This arrangement of water molecules, the wing orientation, in the channel is in contrast to the proton wires [16, 110] formed in other water channels, such as gramicidin A [109, 111], synthetic channels [113], and carbon nanotubes [54], and completely blocks proton conduction [130]. This feature has been demonstrated most clearly through model channels of carbon nanotubes using MD simulations and network thermodynamics [151]

As described above, conventional MD simulations can be used to describe water molecules passing through the channel via diffusion, where the net transmembrane water flux through the channel is zero. While random diffusion of water is always happening, a physiologically more important event, which can also be measured experimentally, is the net flow of water induced by an osmotic or hydrostatic pressure gradient across the membrane. In fact, the osmotic permeability, p_f, an important quantity for a water channel, defined as the ratio of net water flux to the osmotic pressure difference, cannot be directly obtained by observing water diffusion in equilibrium MD simulations. In order to study such events, MD simulations are needed, in which bulk water located on the two sides of the membrane experience different osmotic or hydrostatic pressures, a method that seems to be difficult to implement since under PBC conditions, the bulk water on the two sides of the membrane are actually connected.

We have developed a novel method to overcome this difficulty, which can be used to quantify the osmotic permeability p_f of water channels by MD simulations [153]. In this technique, through application of external forces to every bulk water molecule, one generates a hydrostatic pressure gradient across the membrane that promotes water transport. Through adjustment of the force applied on individual water molecules, the pressure difference can be easily controlled. Since the membrane is experiencing different hydrostatic pressures, there is a net force on it from water, which will cause it to move. In order to keep the membrane in position, constraints can be used, or alternatively, counter forces can be applied to the membrane to make the net force of the whole system zero. This method can be used to generate any magnitude of pressure difference across the membrane, which enabled us to observe a net water flux through the channel and measure the channel's osmotic permeability from the simulation. The water flux can be measured by counting the water molecules passing through the channel during a certain time. From the ratio of water flux and pressure difference we could determine the osmotic permeability p_f. Simulations, performed on GlpF, revealed a linear relationship between net flux and applied pressure gradient [153].

Water translocation in AQPs occurs via a concerted displacement of a single file of water

constituted on average by eight water molecules inside the channel. In their bipolar configuration, water molecules are electrostatically stabilized, through specific interactions with the protein and with neighboring water molecules, by 15-30 kcal/mol per water molecule [64]. Electrostatic analysis of the interaction between a monovalent anion and the protein at various points along the channel axis indicates a very strong repulsion at the NPA motifs; this suggests that the center of the channel, which provides the main proton blocking mechanism, also plays an essential role in the selectivity of the channel against other ions [64].

The molecular mechanisms by which membrane channels facilitate a highly selective transport of materials across biological membranes is of great importance in biology of all cells. Such mechanisms are certainly due to specific architecture of the protein(s) that forms the channel and, therefore, detailed understanding of structure and dynamics of membrane channels is essential for understanding the physical basis of their selectivity. From this point of view, AQPs provide an excellent example, since several selectivity mechanisms at different functional levels have been successfully explored by means of modeling and MD simulations. Future simulations would investigate the possible involvement of the central tetrameric pore of AQP in ion transport. Three dimensional structures of other AQPs could be modeled based on their high degree of similarity with structurally known AQPs, AQP1 and GlpF. Such models will be used for studying the effect of mutations with known clinical implications and/or for investigation of transport of other small moelcules across biological membranes.

Energy conversion in ATP synthase

Living cells depend on an efficient transformation of the energy derived from light and foodstuff into the chemical energy of adenosine triphosphate (ATP), the universal energy carrier. The enzyme that facilitates this type of energy conversion in bacteria, chloroplast and mitochondria is ATP synthase, a complex of two molecular motors mechanically coupled by a common central stalk, as shown in Fig. 9. The membrane unit F_o converts the transmembrane electrochemical potential into mechanical energy that drives the central stalk rotation inside the solvent exposed F_1 unit. The rotation causes cyclic conformational changes in the F_1 unit catalytic sites, which drive ATP synthesis through the so-called binding change mechanism [15]. ATP synthase can also operate in the reverse direction, utilizing the energy released in the process of ATP hydrolysis to pump protons across the membrane.

The F_o unit consists of three types of subunits: subunit a, believed to mediate proton translocation; a dimer of b subunits, which extends from the membrane to the solvent, mechanically connecting F_o and F_1; and a ring-like oligomer of c subunits [9, 43, 69, 132, 133]. The number of c subunits in the oligomer is believed to vary between 9 and 14 for different species; for *E. coli*, which has the simplest F_o structure, the number was found to be 10 [67]. A single c subunit folds in the membrane in a hairpin-like structure of two transmembrane α-helices, the inner one and the outer one, connected by a short polar loop [149]; the loops form a mechanical contact with the central stalk (the γ, ϵ, and, in the bovine mitochondrial species, δ subunits of the F_1 unit). Residue Asp61 of the c subunits, located in the outer transmembrane helix, is believed to be a proton binding site. Its protonation and deprotonation has been proposed to drive rotation of the c_{10} oligomer relative to the ab_2 subunit complex, the motion being coupled

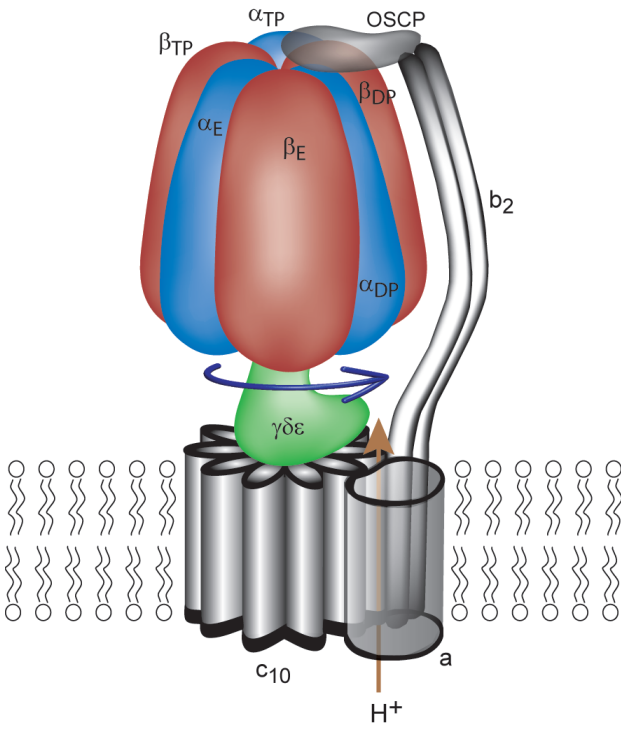

Figure 9: Major subunits of F_1-F_o ATP synthase. The curved arrow indicates the direction of central stalk rotation during synthesis. Subunits are labeled using the bovine mitochondrial ATP synthase naming scheme.

to the central stalk rotation in the F_1 motor. A single 360° rotation of the c_{10} complex and the stalk involves ten protons translocating across the membrane, releasing free energy sufficient to produce three ATP molecules. This process is considered further below.

Crystallographic structures of F_1 from bovine mitochondria [1, 42, 94] show three catalytic sites pesudo-symmetrically arranged around the assymetric central stalk, each site in a different configuration. The central stalk of F_1 comprises an α-helical coiled-coil running along most of its length, with a wide base at one end where it contacts F_o (Fig. 9). The central stalk is surrounded by three α/β subunit pairs in a pseudo-symmetric configuration, the pairs forming the three catalytic sites. F_o-driven rotation of the F_1 central stalk relative to the α/β subunits is thought to drive the subunits, and thus the catalytic sites, through at least three conformational states per rotation, with each site synthesizing one ATP molecule per rotation [15], at a position 100 Å away from where the torque is applied by F_o. The central stalk can spin like a rotor within the $(\alpha\beta)_3$ stator since a secondary, non-rotating stalk fixes $(\alpha\beta)_3$ by connecting it to the F_o stator. F_1 has been directly visualized performing the reverse of synthesis, e.g., when ATP is hydrolyzed by F_1, the central stalk is observed to spin [92, 99, 146], with changes in $(\alpha\beta)_3$ connected with cyclic ATP hydrolysis presumably producing the required torque. Conformation changes in the α/β pairs appear to be cooperative, the coupling being transmitted both between neighboring pairs and through the central stalk [15, 38, 73, 116]. Mutation experiments have identified residues important to α/β/stalk interactions [38, 73] and several schemes for F_1 synthase subunit interactions have been proposed [20, 103, 116, 140]. The rotation of the F_1 central stalk and its coupling to changes in the $(\alpha\beta)_3$ binding sites is considered further below.

Torque generation in F_o.

While the rotary catalysis mechanism of the F_1 unit operation has been recently demonstrated in a series of spectacular single molecule experiments [53, 76, 92, 98, 99, 146], much less is known about atomic scale events involved in the F_o motor function. How is proton translocation across the membrane coupled with the mechanical rotation of the c ring against the load imposed by the F_1 unit? What protein groups, and possibly water atoms, mediate the proton path to and from the essential Asp61 residues? What domain motions occur in the protein, and how are they coupled to the protonation and deprotonation of the key protein residues? How many protons are needed for a synthesis of one ATP molecule, and what makes the F_o operation almost 100% efficient? How does the F_o unit operation depend on the environment, for example, pH? In this section we present

an overview of the recent results obtained through MD simulations and mathematical modeling. The results are described in more details in ref. [4].

The principal problem that limits understanding of F_o operation on the atomic scale is lack of a sufficiently complete protein structure. The only available crystallographic structure obtained for the mitochondrial F_1-c_{10} complex at 3.9Å resolution [127] does not include subunits a and b, which are critical for the protein function. Several structural models have been developed on the basis of NMR experiments with individual subunits of F_o in polar solvents and detergents, disulfide cross-linking, scanning mutagenesis, and analysis of suppressor mutations [31, 46, 69, 72, 114]. It is not clear, however, in how far the protein structure in detergent is similar to that in membranes [46]. In this regard, all-atom MD simulations provide a unique opportunity to computationally investigate the proposed models on the atomic scale, estimate their stability, identify functionally relevant domain motions, and provide parameters necessary for modeling the protein motor operation on the physiological (millisecond) time scale.

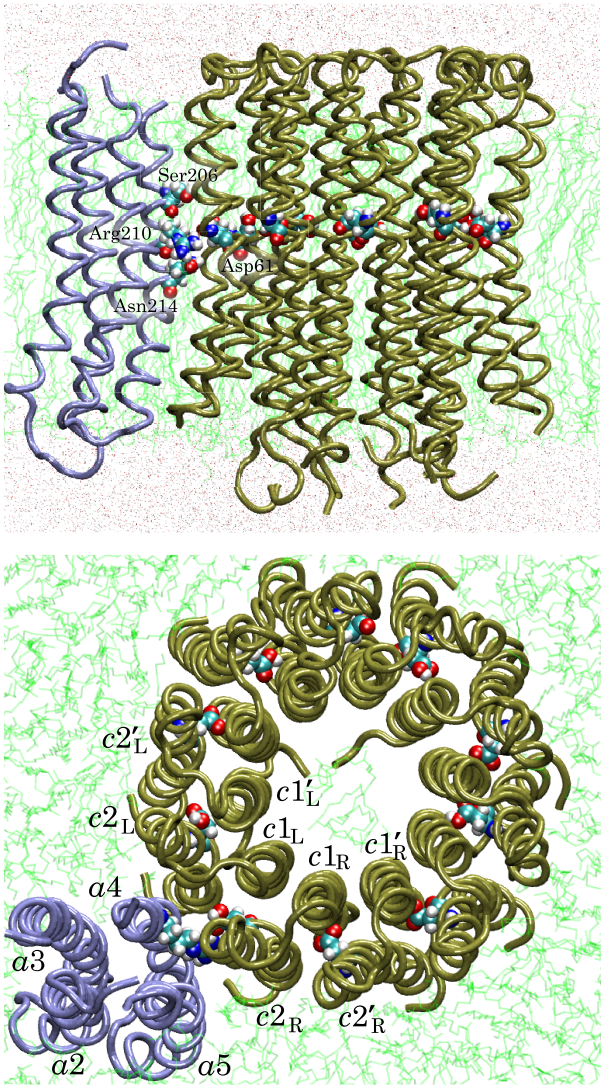

Figure 10: Structural model of F_o ATP synthase. Top: side view (cytoplasmic side up, periplasmic side down); bottom: front view from the cytoplasmic side. The F_1 unit is not shown for clarity.

To perform an MD study, a minimal structural model of the F_o motor was built using the available experimental data on subunits from *E. coli* ATP synthase, which is known to have the simplest structure [33, 68]. The model included subunit a and a 10-mer of subunits c, about 14,300 atoms total. Although most functionally relevant events are assumed to occur at the interface between subunit a and two nearest subunits c, all 10 subunits c were included to assure the oligomer stability in the membrane/solvent environment. Since it is still unclear whether subunits b are involved in proton translocation[1] [30], and their

[1]Reported experimental data remain controversial: Fillingame and co-workers [51] found that proton translocation

exact position in F_o is not known, they were not included into the model. The structure of the c_{10} complex was provided by Fillingame [29]. The only available structure of subunit *a*, which had been obtained from an ac_{12} complex [114], was docked to the c_{10} structure and then modeled. The outer transmembrane helices (TMH) of the four subunits *c* of the ac_{12} complex, closest to subunit *a*, were aligned to the outer TMH's of four sequential subunits *c* in the c_{10} complex; then subunit *a* was merged to the latter, forming an ac_{10} structure. Since the structure of subunits *a* was determined in a polar environment, whereas the *c* oligomer structure was obtained in a detergent mimicking the hydrophobic environment inside biological membranes, obtaining the correct structure of the interface between subunits *a* and *c* required additional modeling. The system underwent 10,000 minimization steps followed by 130 ps equilibration *in vacuo* at 4K with all subunits *c* constrained. While the temperature was too low for dihedral angle rotation, preventing substantial re-orientation of any relevant structural group, the equilibration allowed the helices in subunit *a* to assume the conformation that fits the structure of the *c* subunits complex.

To correctly model the operation of F_o, the protein was embedded in a membrane/solvent environment, which stabilized the protein structure. The membrane was modeled as a 130×130 Å^2 patch of a pre-equilibrated POPE bilayer[2]. After the protein had been embedded in the bilayer, lipid molecules that overlapped with the protein were removed. The protein-lipid complex was then solvated with TIP3 water using the *solvate* plugin of program VMD. Although an implicit solvent model could provide a better performance of the MD simulations, using TIP3 water was necessary to observe events in which individual water molecules may participate, for example, interacting with solvent exposed protein residues or possible spontaneous formation of the proton inlet or outlet half-channels. The water volume was large enough to provide the minimum distance between the protein and its periodic images of over 30Å, which was sufficient to screen out the electrostatic interactions. To electrically neutralize the system, which was required for simulations with periodic boundary conditions, chlorine and sodium ions were added using VMD extended through `Tcl` scripts. The resulting system, which included about 112,000 atoms, underwent minimization and equilibration for about 3 nanoseconds in the Charmm27 forcefield [91].

The principal goal of the work was to understand microscopic events that couple proton translocation across the membrane with the mechanical rotation of the *c* subunit oligomer. A single cycle

(but not active transport) is possible if subunits *b* (then named Psi) are missing, whereas Takeyama [131] and Monticello [96] concluded that they are required for any proton transport in F_o. Since all these studies had been done before the first atomistic structure [1] was available, the role of subunits *b* in proton conduction is not clear yet.

[2] `http://www.ks.uiuc.edu/~ilya/Membranes`

of the F_o operation involves processes that take place on dramatically different time scales: protonation/deprotonation of the key protein residues occurs about a billion times as fast as a revolution of the c oligomer. Neither the first nor the second process can be addressed directly by state of the art MD simulations: it is not feasible to simulate proton transfer over distances as large as about 15Å (half of the membrane hydrophobic layer thickness), and the millisecond time scale of the mechanical rotation is far beyond reach of conventional MD methods. Given the limitation, can simulations help us learn anything about the microscopic mechanisms of the F_o operation? We believe that this is possible, even though it is difficult. To bridge the time scale gap, we employed a multiscale approach, assuming that F_o operates as a molecular ratchet in which the rotor shows rotational diffusion in a potential biased by the proton-motive force [71]. Given that ratchets are known to be well described by stochastic models [95], we developed such a model to extend the investigation of F_o to the physiological (millisecond) time scale. In addition to the model, a series of MD simulations was performed to obtain quantitative parameters and test different scenarios of key events in F_o. Unlike the earlier idealized one-dimensional ratchet model suggested in [31], which essentially describes a generic protein motor, our approach directly relates the atomistic structure of F_o to the stochastic model. As we show below, using the atomistic structure not only leads to a more realistic dynamical description, but it also provides insight into the relationship between the protein structural features and its function.

The established overall view of the F_o operation assumes that the c_{10} complex, being in contact with the central stalk of the F_1 motor, rotates relative to the ab_2 complex. Each subunit c has a proton binding site, Asp61, which is located in the middle of the membrane hydrophobic layer. Given that the membrane-spanning part of the c subunits is formed almost entirely by hydrophobic residues, which can not mediate proton pathways, the binding site can only change its protonation when it is located against TMH4 of subunit a. The latter includes several polar groups, which are thought to mediate two proton half-channels. The inlet channel leads from the proton-rich periplasm to the center of the membrane, and the outlet channel connects the membrane center to the cytoplasm. The binding site in subunits c can assume the protonated (neutral) or deprotonated (charged) form depending on the environment, which may affect the effective pK_a value. The subunit c located at the interface with subunit a, which is formed by polar protein residues that include positively charged Arg210, assumes the deprotonated state. However, for the c_{10} complex to rotate, the binding site needs to be protonated before it leaves the interface: the energy penalty for exposing an electrically charged group to the hydrophobic membrane environment is too high.

Therefore, a proton from the periplasm travels via the inlet half-channel and protonates the binding site. By a mechanism not yet understood, the rotation of the c_{10} complex brings, after an almost 360° turn, that binding site close to the a subunit again, causing it to release a proton, which travels via the outlet half-channel to the cytoplasm. Each step of the proposed scenario needs to be investigated in detail; here, we focus on two events: rotation of the entire c_{10} oligomer relative to the ab_2 complex, and rotation of individual membrane-spanning TMH's in the protein subunits. A more complete description of our simulations and model is given elsewhere [4].

A series of steered MD simulations was performed to examine the feasibility of rotation of the c_{10} complex relative to subunit a in the lipid bilayer. To avoid distortion of the protein structure in the process of rotation on the nanosecond time scale (about 10^5 times faster than the physiological rotation rate), forces were applied not only to the polar loops of subunits c, which form contact with the central stalk, but to all 3170 backbone atoms of all c subunits. The backbone atoms of TMH's 2, 3, and 5 of subunit a were restrained to prevent subunit a from being dragged along with lipid molecules surrounding the protein. TMH4 of subunit a, which forms the interface with subunits c, was not restrained.

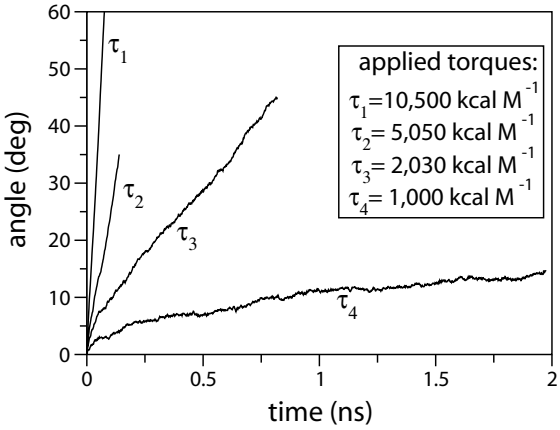

Figure 11: Forced rotation of the c_{10} oligomer in the ac_{10} protein complex in membrane/solvent. The system included about 112,000 atoms. The simulations were performed using the program NAMD2 with full electrostatics and Langevin thermostat damping of 5ps^{-1}.

The rotation axis was normal to the membrane plane, and the center of rotation was at the center of mass of the c subunit oligomer. The magnitude of the force acting on each atom was proportional to the distance between the atom and the rotation axis. The simulations were performed in the NVT ensemble. To control the system temperature, dissipate the heat produced by the applied forces, and additionally reduce distortion of the protein structure, NAMD Langevin forces were applied to all heavy atoms in the system. To allow a full revolution of the c_{10} complex, Asp61 residues in all subunits c were protonated.

The angle of the c_{10} oligomer rotation as a function of simulation time t is shown in Fig. 11 for several values of applied torque. The angle was calculated by averaging the rotation angles for each c subunit, which were computed using the subunit center of mass positions at time t and in the

beginning of the MD simulations. For all torque values, the protein structure remained stable, and no significant dependence of the angular velocity on the Langevin damping strength was observed. The angular velocity increased with the torque. Interestingly, the rotation substantially slowed down for torques of less than 2,000 kcal/M, indicating a transition to a high friction regime at lower angular velocities.

Another series of MD simulations addressed the feasibility of rotation of individual TMH's in subunit c, that had been proposed by Fillingame and co-workers [34–36]. NMR and cross-linking data showed that the position of Asp61 in the outer TMH of the c subunits depends on its protonation: in the protonated state, Asp61 is hidden in the hydrophobic core of subunit c, whereas in the deprotonated state, Asp61 extends from the outer TMH towards Arg210 of subunit a [69, 72, 114]. In order to release (or bind) a proton, Asp61 needs to approach the terminus of one of the proton half-channels, which are believed to be residues Ser206 and Asn214 in subunit a; both residues are located at the interface with subunits c next to the positive Arg210 residue of subunit a [34]. It was, therefore, suggested that a rotation of the outer TMH of subunit c ($c2_{\mathrm{L}}$) around its axis by about 120° (or 240° in the other direction) is necessary to bring Asp61 to the interface before it can release a proton (note that, at the interface, the proximity of Arg210 also facilitates deprotonation of Asp61) [34]. The scenario proposed in [34] also included rotations of TMH4 of subunit a and the outer TMH of the other nearest subunit c ($c2_{\mathrm{R}}$), which were thought to bring Asp61 close to the terminus of the other half-channel, allowing its re-protonation necessary for the c_{10} complex to rotate further. While the idea of individual TMH rotation successfully explains the observed conformational changes in subunit c upon protonation or deprotonation of Asp61, the particular scenario raises a number of questions. Most importantly, deprotonated Asp61 is likely to form a salt bridge with Arg210, which could effectively prevent rotation of either TMH4 of subunit a or the outer TMH of the c subunit.

To find out if the rotation of the c_{10} complex can occur, a 1 ns steered rotation of the c_{10} oligomer with one or two deprotonated Asp61 residues was performed. We observed formation of the salt bridge between Arg210 and Asp61, which dragged TMH4 of subunit a along with the c_{10} complex, breaking apart the four-helix bundle structure of the subunit. To preserve the structure, in another 1 ns simulation all TMH's of subunit a were restrained; that, however, resulted in unwinding the outer TMH of the c subunit. Therefore, simple rotation of the c_{10} complex appears unlikely to break the salt bridge; instead, the rotation leads to significant distortion of the subunit's structure. These results suggest a concerted rotation of the c_{10} complex and the $c2_{\mathrm{R}}$ TMH. To investigate this

possibility, a special algorithm for rotating the TMH was used. The $c2_{\mathrm{R}}$ TMH has a kink caused by the Pro64 residue located right below Asp61; the helix symmetry axis changes its direction at the kink. Therefore, a rotation of the entire helix around a single axis requires application of an artificially high torque because of the steric hindrance. To minimize the resistance, the rotation axes were assigned individually to every residue of the helix. As in the earlier simulations, the forces were applied to every heavy atom, and the force magnitude was proportional to the distance from the atom to the corresponding rotation axis. Driven by these forces, the helix rotated entirely within its reptation tube formed by the surrounding proteins and lipids. To the best of our knowledge, this type of protein domain motion has not been investigated before.

In the simulation of the combined rotation of the c_{10} oligomer and the $c2_{\mathrm{R}}$ helix, the two rotations were performed in several steps for technical reasons. First, the c_{10} oligomer was rotated clockwise by about 37 degrees in a 1 ns simulation. After that, the $c2_{\mathrm{R}}$ helix was rotated counterclockwise by 220°. The salt bridge stayed intact, and no significant distortions of the structure were observed. Then, the nearest binding site clockwise from $c2_{\mathrm{R}}$ (which belongs to the $c2'_{\mathrm{R}}$ helix, see Figs 10 and 12) was deprotonated. The simulation of the clockwise rotation of the c_{10} complex was continued. After rotating c_{10} by about 10° more, a complex of three charged residues was formed, as shown in Fig. 12. The dissociation energy of the salt bridge was dramatically reduced by the presence of another charged residue, making it possible to transfer the salt bridge from one subunit c to the other. Indeed, when we continued the simulation of the clockwise rotation, the salt bridge transfer was observed. In this simulation, we demonstrated a mechanism, by which the rotation of the c_{10} oligomer can proceed when the deprotonation of the binding sites take place.

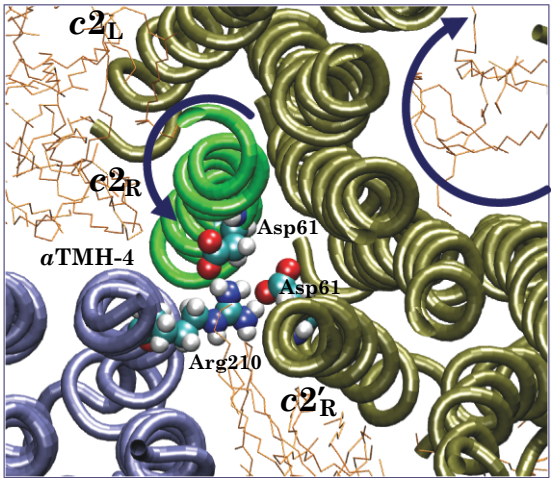

Figure 12: Concerted rotation of the c subunit outer helix and the c_{10} complex in a lipid bilayer. The highlighted helix has been forced to rotate counterclockwise by 220°. The moment of the salt bridge transfer between the two neighboring c subunits is shown. The initial structure is shown in Fig. 10.

While MD proved to be very successful in describing separate microscopic events that are involved in the process of the torque generation by $\mathrm{F_o}$, it cannot provide an overal account of the

F_o function on the physiological time scale. Instead, we constructed a simplified, mathematical model of the F_o motor, based on the stochastic equations of motion.

The model was derived from the microscopic description by reducing the number of degrees of freedom of the system. It was assumed that all torque generating events take place at the interface of the a and c subunits (see Fig. 13). The only motions allowed in the model are the rotations of the transmembrane helices and the c_{10} complex, respectively, which were already investigated by MD. Only one residue from each helix is considered explicitly in the model. Those residues are the binding sites (Asp61s) of the ten c subunits and the Arg210 residue of subunit a. The conformations of the residues are restricted geometrically, mimicking their covalent attachment to the backbone of the protein. It is assumed that the rotation of the residues is equivalent to the rotation of their parent helices and *vice versa*.

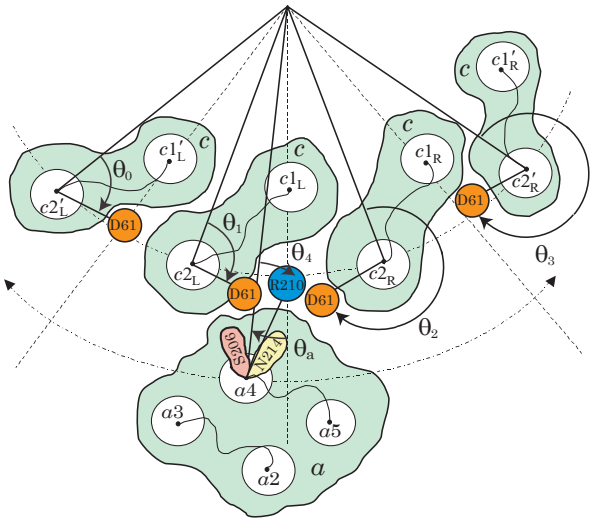

Figure 13: Mechanical model for F_o(view from the cytoplasm). The c_{10} complex is fixed (only four c-subunits are shown) while the a subunit can move in either direction. The second trans-membrane helix ("$c2$") of each c subunit can rotate independently, moving the key D61 residues, which are the proton binding sites. Similarly, the fourth helix of the a subunit ("$a4$") can turn, moving the Arg210 residue. The proton transfer is assumed to occur between the terminal residue of the periplasm channel (N214) and the binding site on the $c2_R$ helix, and between the terminal residue of the cytoplasm channel (S206) and the binding site on the $c2_L$ helix. All mechanical motions are confined to the plane of the figure. The system is fully described by the six angles and the protonation state of the two aspartates on helices $c2_L$ and $c2_R$.

Intuitively, the motor operates as follows. When a proton binds or unbinds from the binding site, it alters the potential that governs the motion of all helices. A repeated, asymmetric alternation of the potential transforms the random rotary fluctuations of the helices and c_{10} oligomer into a unidirectional motion. The direction of motion is dictated by the counteraction of the proton electro-chemical gradient and the load torque from F_1.

A typical result from the mathematical model is reproduced in Fig. 14, where the angular coordinates of the key residues are plotted versus time. The conformations of these residues are constrained by the Potential of Mean Force (PMF) acting on their parent helices. In this particular

simulation, the PMF acting on the parent helices of the binding sites, has an asymmetric double-well shape. Therefore, the binding sites spend different amount of time at the minima of the potential (black and red lines). The barrier between the minima is low enough to permit the binding sites to change frequently their orientations. The Arg210 residue (green) is directed mostly towards the center of the c_{10} oligomer, as it is dictated by the PMF acting on the $a4$ helix. The latter potential has only one minimum at zero, i.e., when Arg210 is oriented towards the center of c_{10} . The blue line in Fig. 14 indicates the relative position of the a subunit with respect to the c_{10} oligomer. In this simulation, the oligomer rotates against the physiological load of 41pN·nm imposed by the F_1unit. The rotation proceeds in steps, with the average velocity of 0.1 revolution per ms.

Thus, we have demonstrated how a microscopic event, discovered in a large scale MD simulation, can be linked to the physiological time scale through mathematical modelling. The present study is only the beginning of our computational effort directed to unveil the great puzzle of the operating mechanism of the F_o ATP synthase. Our mathematical model is open to accommodate new structural information, specifically anticipating the atom resolution crystal structure of F_o.

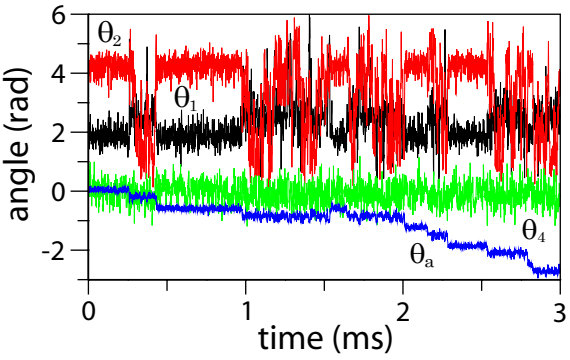

Figure 14: Stochastic simulations of the events involved in the F_o function: time evolution of rotation angles θ_1 (black), θ_2 (red), θ_4 (green) and θ_a (blue). The angles are defined in Fig. 13. The a subunit rotation takes place in steps (blue line).

Torque-driven catalysis in F_1.

ATP synthase is a double motor. In the previous section we have focused on the F_o component of this motor turning a proton motive force into a torque. This torque is coupled through the stalk, mainly its γ subunit which includes the long coiled-coil α-helix structure, to catalysis in the F_1 ATP binding sites. The mechanical properties of the stalk are obviously an important part of nature's design of ATPase. A key question arises: Is the stalk rigid? Does turning of its base move all parts as a rigid body at the same angular velocity all way up to the tip, or is it highly elastic such that it develops first a twist of its bottom segment before its main segment embedded in $(\alpha\beta)_3$ begins rotation? Possibly, if the latter scenario applies, the partial twist of the stalk induces a change of shape in the stalk, e.g., alters the coil-coil helix arrangement. It is even possible that the action of

the rotating stalk on F_1 catalysis can only be understood on the basis of the correct shape change of the stalk. Indeed, this change and, hence, the mechanism of coupling between stalk torque and catalysis might be different for the forward (synthesis) and backward (hydrolysis) reaction of ATP synthase.

The rotation of the stalk had been investigated already in the ground breaking simulation of Böckmann and Grubmüller [11]. However, these authors did not rotate the base of the stalk – which in fact was not even included in the simulation – but only rotated the stalk center, closer to the catalytic binding sites, enforcing the rotation equally along the entire stalk section. As a result, geometry change of the stalk induced by its base segment, elastic energy storage in this part, and results of the interaction of the base with surrounding structures might have been missed in this simulation.

Once rotation of the stalk is properly understood one can focus on the torque–catalysis coupling in F_1. Ma et al. have carried out so-called targeted MD simulations (TMD) [122] that addressed this coupling. The TMD simulations apply external forces to the system that pulled F_1 ATPase from its observed structure S0 to another structure S1 that results from S0 through 120° rotation. The forces pull all S0 atoms to their new positions in S1 along the shortest paths. Such an interpolation-like method makes it difficult to tell apart behavior resulting from properties of the studied system from behavior imposed by the simulation procedure. To explain the shortcoming of TMD we consider in Fig. 15 its application to study the toppling in a line of of dominoes leading to the well-known domino effect. One can discern that TMD misses entirely the causal chain of the domino effect, instead predicting falsely that all dominoes without collision rotate in concert from an upright to a horizontal position. One wishes that a simulation applies forces as they actually occur in the normal function of ATPase, namely, forces that are consistent with a torque at the base of the stalk. Only such simulations can reveal the actual mechanism of ATPase.

ATP synthase is an extremely complex machine that functions on a millisecond time scale. Many aspects of the mechanism of F_1 torque–catalysis coupling are not explained by existing experimental data [11, 20, 38, 89, 103]. A description of how F_1 couples applied torque to catalytic site changes requires answers to these questions: What are the changes to central stalk structure, and torque transmission along the stalk, during rotation? What is the sequence of deformations and interactions between/within F_1 subunits that lead to catalytic site conformation changes? How do the conformations of the catalytic sites change throughout a rotation?

Experimental studies of F_1 torque production [92, 105], the interactions among F_1 subunits,

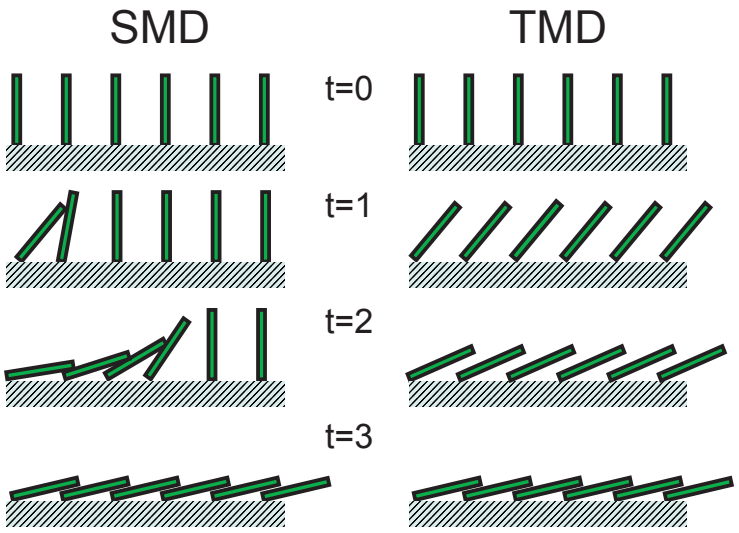

Figure 15: Loss of causal relationships in TMD simulations. A row of dominoes moves from the initial, upright, state to the final, collapsed, state. SMD applies a force to the leftmost domino and predicts the correct series of events. TMD predicts instead a uniform angular motion of all dominoes, failing to capture the "domino effect". Notice that the TMD dominoes do not even contact each other until the final state is reached.

and how these relate to catalytic function [73, 124] provide partial answers to these questions. However, mechanical details, properties, and intermediate states of the rotating system, required for an understanding of torque–catalysis coupling, can not be obtained experimentally. There is also the danger that model simulations on a nanosecond time scale don't capture all of the action of this machine and, if one accelerates motion to induce transformations over nanoseconds, that the machine may break, or at the minimum, distort and become artificially caught in a distortion. One must expect many such problems; simulation studies of ATP synthase likely have a long way to go in revealing the underlying physical mechanisms, yet on the other hand the currently feasible simulations may already provide a first glimpse into the mechanics of ATP synthase. In this section, we summarize results of full-atom MD simulations of F_1, which are presented in greater in Isralewitz et al. [59].

To simulate torqued rotation in the F_1 unit, we performed periodic boundary MD simulations on a 327,000 atom system including ATP synthase F_1, nucleotides, water, and ions, and applied a torque to the central stalk. The F_1 structure was built from coordinates of DCCD-inhibited bovine mitochondrial ATP synthase [42] (RCSB code `1e79`), which has the most complete structure of the

central stalk among available structures. To construct our model system, the inhibitor-modified β_{DP}-Glu-199 was changed back to its native structure, and two small γ chain gaps were modeled in place. We placed three ATP molecules in the α subunits, and an ATP and ADP molecule in the β_{TP} and β_{DP} respectively, even though the `1e79` structure shows ADP bound at both the β_{DP} and the β_{TP} catalytic sites. An Mg^{2+} was placed at each of the five occupied nucleotide binding sites. To fill in cavities with water molecules, water positions suggested by DOWSER [148] and additional proximity criteria were used. A system was then produced by solvating the assembly with a water shell four molecules thick, producing a system size of 100,000 atoms; the protein displayed large backbone conformation changes when equilibrated in this system. A larger system was then created by placing the complex in a rectangular prism of water and ions large enough to allow periodic boundary simulations, for a total 92,000 water molecules and 400 ions, producing a total system size of 327,000 atoms. This system was equilibrated for 1 ns as an NPT ensemble using (as for all F_1 MD trajectories described here) a 1 fs timestep with PME electrostatics, followed by 0.2 ns equilibration as an NVE ensemble.

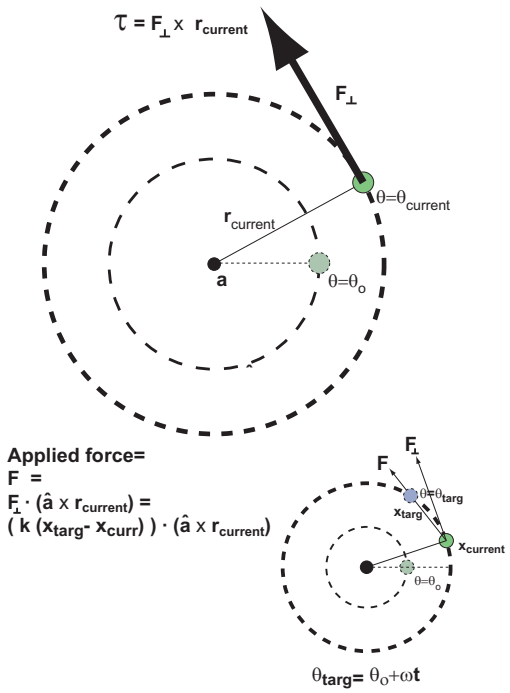

Figure 16: Torque is applied to rotate 16 C_α atoms at constant angular velocity $\omega = 24\,^\circ$/ns without restraining motion in the directions of **a** and **r**. Force is applied to the component tangent to $\mathbf{r}_{\mathrm{current}}$ of a harmonic restraint to the fictitious "target" atom.

We use a modified form of our SMD torque method [60] to simulate the torque that the F_o subunit normally applies to F_1 in fully assembled ATP synthase, applying torque to restrain a subset of atoms to constant angular velocity (see Fig. 16). Restraining only angular velocity makes a minimum of assumptions about the details of the as yet unknown F_1-F_o interface. Although the method can be shown to be unstable in any system if applied with a sufficiently high target angular velocity, testing showed this not to be a problem with this system at the angular velocities we employed. The axis of rotation for torque application was chosen as the pseudo-symmetry axis of F_1, determined as a fit through the averaged positions of backbone atoms in the N-terminal region of the three α subunits, a portion of F_1 with near-C3 symmetry. We applied torque to 16 atoms within 5 Å of the presumptive F_o interface, to reflect the physiological torque transmission (see Fig. 17a).

In our first round of simulations we applied torque to enforce 240 ° of 24°/ns stalk rotation, requiring 10 ns of simulation time, then subjected the system at different stages of rotation to a total of 8 ns of equilibration. The 18 ns of dynamics required 652,000 hours of processing time, utilizing at different times between 32 and 512 processors of NCSA and PSC supercomputer clusters (see Fig. 2).

The total external torque required to maintain 24°/ns constant angular velocity of the 16 SMD constrained atoms is shown in Fig. 17b, and ranges between roughly 50–100 times the experimentally observed total torque of 40 pN·nm [77]. The total torque, averaged with a 0.1 ps window, fluctuates around 2300 pN·nm for the first 2 ns of torque application, then displays three cycles of increase/decrease between 2.0 and 6.5 ns, with local maxima at 2.2, 3.3, and 5.1 ns, reaching a maximum of 4820 pN·nm. There is a 2000 pN·nm torque increase between 2.8 and 3.5 ns, a 1500 pN·nm increase from 4.0 to 4.7 ns, a sharp drop, then an increase of 2000 pN·nm from 4.8 to 5.1 ns. To measure the propagation of twist up the stalk, we graph the best rotation fit of slices of the stalk structure, proceeding along the rotation axis, in Fig. 17c. The enforced rotation propagates 60 Å up the stalk from the foot along the solvent-exposed region of the stalk, until about 6.4 ns (153 °), with the region 0–15 Å along the stalk acting nearly like a rigid body. In the region 60–90 Å along the stalk, where the stalk is buried in the $(\alpha\beta)_3$ complex, the overall propagation of rotation in the stalk sections along the rotation axis appears to halt at about 3.0 ns (72 °). The discontinuity in average rotation at 60 Å and 3.0 ns seems an unphysical stretching of the bonds connecting sections of the stalk, but actually represents the sum of forward rotation of one of the two helices making up the γ-subunit coiled-coil and local backwards rotation of the

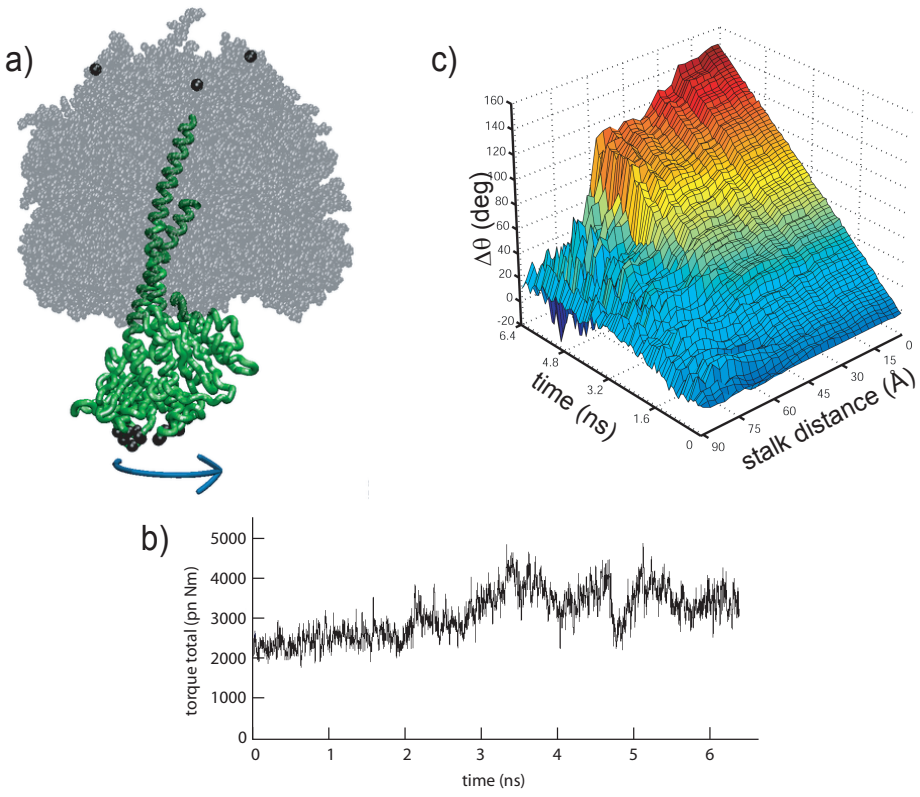

Figure 17: Events during F_1 central stalk rotation. (a) Torque application to F_1. The central stalk is shown as thick tubes, $(\alpha\beta)_3$ is shown as translucent spheres. Torque is applied to 16 C_α atoms (black spheres on stalk), the three C_α atoms at the N-termini of the three α chains are held fixed (black spheres in $(\alpha\beta)_3$). (b) Total applied torque during the first 6.5 ns of rotation. At each timestep the torque applied to each of the angular-velocity constrained atoms is summed; a 100 fs windowed average is shown here. (c) Twist propagation up the central stalk. $\Delta\theta$ is computed for C_α atoms, stalk height is measured along the rotation axis, parallel to the stalk major axis.

other. The behavior is not captured by the averaged rotation graph in Fig.17c since in this region the stalk is acting neither like a rigid body nor a single twisted rod. As shown in Fig. 18, the longer C-terminal helix (γ-192-272) continues to rotate around the symmetry axis, but the shorter N-terminal helix (γ-1–50) tilts across the symmetry axis, since it winds around the N-terminal helix, as shown in Fig. 18, making a backwards rotation contribution to the total average-fit rotation shown in Fig. 17b.

To characterize the interaction between the central stalk and the three γ subunits during the rotation, we examine the vdW interaction between the β subunits and the portion of the γ stalk (γ 12–19, 25–30, 234–242) which was observed to have close interactions with the C-terminal domains

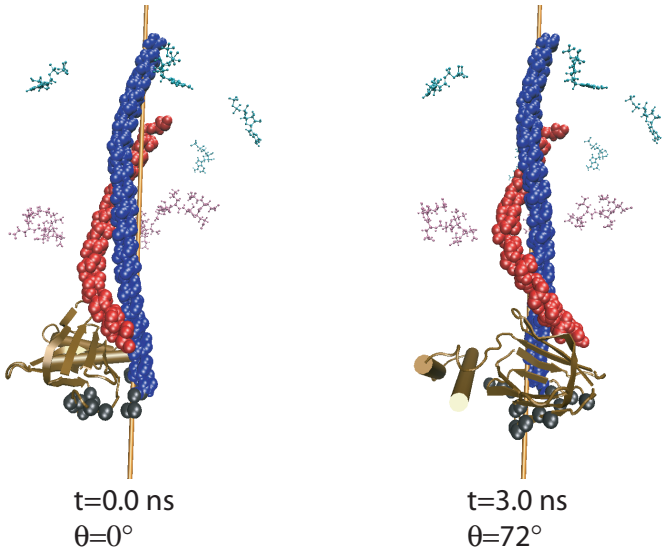

Figure 18: Winding of γ subunit coiled-coil. The longer helix (γ 197 – 272) of the coiled-coil is shown as blue vdW spheres, the shorter helix (γ 1 – 50) as red vdW spheres. The δ subunit is shown in cartoon representation (tan color), with secondary structure recalculated for each configuration. For spatial reference, the rotation axis, DELSEED sequences (pink CPK representation), bound nucleotides (green CPK), and torqued atoms (black spheres) are shown.

of β_{DP} and β_{E} in the AlF_4^- hydrolysis intermediate structure of F_1 [94]. As shown in Fig. 19a, the favorable decrease in vdW interaction energy with the γ section is observed for β_{E} around 1.8 ns, an unfavorable increase is observed for β_{TP} at 3.0 ns, and relatively little change is observed for β_{DP}.

If the system is obeying the binding change model, we expect the β_{TP} subunit to move to an open configuration as the central stalk rotates; the two parts of the ATP catalytic site in β_{TP}, the nucleotide binding pocket and the phosphate binding pocket (which binds P_i before synthesis and ATP-P_γ after synthesis), should unbind from ATP. In Fig. 19b,c, the charged groups of two phospate binding pocket residues, β_{TP}Thr-163 and β_{TP}Arg-189, move away from β_{TP}ATP-P_γ at roughly 2.8 ns (67 °) of rotation.

The SMD-enforced stalk rotation was carried out to 250°; but with nucleotide bound in a configuration compatible with at most 120° of stalk rotation, the extra rotation was performed in order to re-wind any untwisting of the gamma stalk that the DCCD-inhibition may have caused [19, 42]. A likely point to mark the end of the useful trajectory, after which the enforced stalk rotation is damaging the F_1 structure, is the movement of the β_{TP} DELSEED sequence at 134 ° (5.6 ns) of

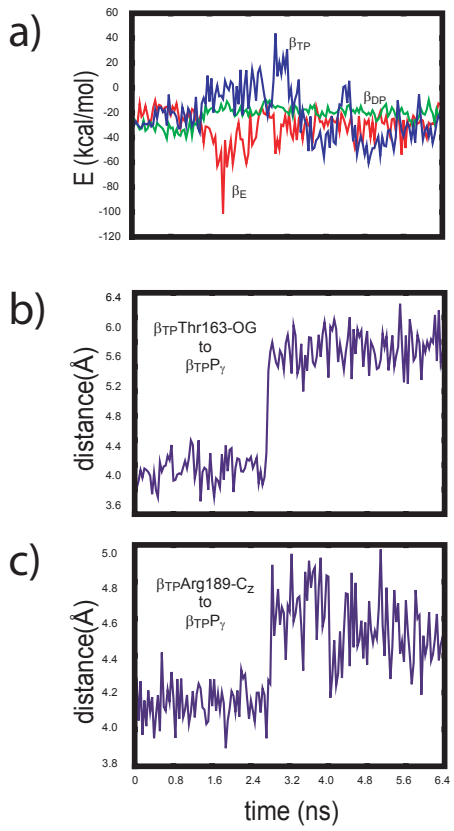

Figure 19: Subunit interaction and ATP unbinding during central stalk rotation. a) vdW interaction energy of a subset (see text) of central stalk groups with the three β subunits. b), c) distances between two β_{TP} phosphate pocket-forming residues and β_{TP}-ATP-P_{γ}.

stalk rotation, as shown in Fig. 20. The β_{TP} DELSEED is rotating towards α_{DP}, severely distorting F_1.

As the enforced central stalk rotation proceeds, two events consistent with a synthesis binding change mechanism take place in F_1, as far away as the 80 – 100 Å distance between the stalk foot and the the catalytic sites: central stalk – β subunit cooperative interactions (Fig. 19a), and ATP unbinding at the phospate pocket of the catalytic β_{TP} site (Fig. 19b). These two energy-requiring events take place during the first large increase in total torque application, between 2.8 and 3.5 ns (c.f. Fig. 17).

The helix winding shown in Fig. 18 may play an important role in energy storage and torque transmission in F_1 function, although the behavior may be largely an artifact of subjecting a system to unphysically high forces. The γ-subunit C-terminal helix, which rotates along with the stalk

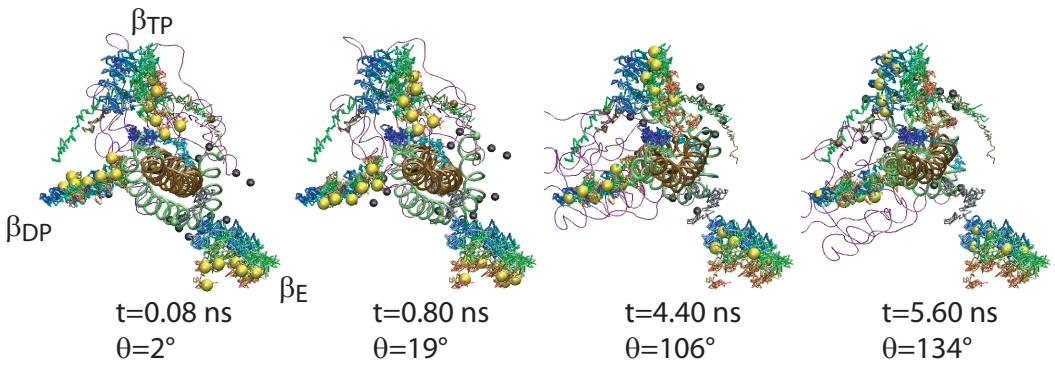

Figure 20: Motion of DELSEED regions (β 394 – 400) during torque application. The β_{TP} movement is a non-physiological distortion of the F_1 structure. Movements of all C_α atoms of the three DELSEED regions are traced as thin line segments, progressing in color from red to green to blue over the 5.6 ns of motion depicted, with the current positions shown as yellow spheres. Torqued atoms are shown as black spheres, the γ subunit as a purple line, the γ coiled-coil as green and brown tubes. The distortion of the β_{TP} DELSEED region is seen in the movement from the green trace area at 4.40 ns to the blue trace area at 5.60 ns.

base, has a direct connection to the putative F_o c_{10} ring interface, very near the pseudo-symmetry rotation axis. The γ-subunit N-terminal helix, which is observed to tilt as it wraps around the C-terminal helix, is connected to the F_o interface less directly, via the β-sandwich of the δ subunit, and at a greater distance from the rotation axis than the C-terminal helix. This would provide a c_{10} ring rotating around the symmetry axis with a greater moment arm with which to apply force to the N-terminal helix, causing the helices to store and release elastic energy independently, with additional possible elastic energy storage in deformation of the δ subunit β-sheet, as in the deformation of β-subunit β-sheets modeled by Sun et al. [129].

The torque increase and coincident increase in stalk-β_{TP} interaction energy depicted in Fig. 17b and Fig. 19a is consistent with what is required for a step forward in synthesis: an energy input to unbind ATP from β_{TP} to transform it's catalytic site to an open state. This energy input is connected with an overall hinge opening motion of of the β_{TP} and in the ATP unbinding from β_{TP}-P_γ. The attractive stalk-β_{E} interactions are consistent with the synthesis-direction transition to a tightly-closed binding site which β_{E} must undergo. The relatively small motions required to transform β_{DP} to a β_{TP}-state are reflected in the much smaller changes observed in the stalk-β_{DP} interaction energy observed.

The unbinding of two residues from β_{TP}-ATP-P_γ appears to open the β_{TP} site to release newly synthesized ATP, a clear step of synthesis-direction binding site change, nearly 100 Å away from

the region of external torque application. Since diffusion does not take place during the simulation time, and since the nucleotides remain bound throughout stalk rotation, the continued presence of ATP in the β_{TP} site will eventually cause the rotating stalk to distort the structure into non-physiological configurations. Future simulations should remove ATP from the β_{TP} site after a geometric or energetic unbinding criterion is met.

Considering the millisecond time scales of F_1 function, along with it's large size and multiple subunit interactions, a complimentary "top-down" approach to modeling is necessary; the needed kinetic model of F_1 might contain a small number of springs and rods, along with mechanical catches and levers linked to the states of the catalytic sites. However, as with all biomolecular systems, such a modeling approach requires deciding which degrees of freedom are important for function and which can be discarded; the problem is made especially difficult by the complexity and large size of F_1. There is still too little experimental data on mechanical changes within F_1 during catalysis to clearly guide construction of a top-down model, although several models have been proposed [72, 105, 106, 129, 138, 139] based on available structure and functional data.

The "bottom-up" approach of whole-system SMD simulations so far applied eliminates the problem of deciding which elements of the system to simulate, it seeks to identify important details and emergent properties of F_1 by preserving as much structural detail as possible while making a minimum of assumptions beyond what is experimentally known. We plan to combine both of these approaches in our further studies of F_1.

An ultimate goal of our work is to integrate our descriptions of torque generation in F_o and torque utilization in F_1 into a single model. An understanding of the elastic properties of both the central and secondary stalks will allow us to connect independent stochastic models of F_o and F_1.

Mechanical signaling in fibronectin

Biological cells are surrounded by the extracellular matrix (ECM), a flexible network containing several classes of proteins secreted by cells themselves. Fibronectin (FN), found in all vertebrates, is the first structurally well characterized ECM protein, yet its functional properties are not fully understood. FN is an important mechanical component of the ECM and acts as a specific adhesive, forming elastic FN fibrils that connect cells via transmembrane protein integrins and guide cell movement [56]. Crosslinked through a disulfide bond at the C-terminus, two identical FN subunits form a large dimeric FN molecule (450-500 kD) consisting of $\sim$20 different modules in each subunit. About 15 modules are type III (FN-III) modules, structurally exhibiting a so-called β-sandwich

motif that has been also identified in many force-bearing proteins, such as immunoglobulin-like (Ig) domains from the muscle protein titin [58].

Mechanical stress has been found to regulate the assembly of FN fibrils, a process termed fibrillogenesis, through integrin receptors that mechanically couple intracellular actin filaments to extracellular FN molecules (see reviews [41, 123]). It has been hypothesized that the stretching of FN fibrils unfolds individual FN-III modules [32], providing the necessary extension to the FN fibril that can be stretched four times as long as its relaxed length [102]. In addition to providing elasticity, the unfolding of FN-III modules is functionally relevant to mediating fibrillogenesis by exposing otherwise buried cryptic sites [41, 135]. Supporting this unfolding hypothesis, recent fluorescence studies found that cells not only integrate fibronectin in an extended conformation into fibrils, but also overstretch many fibrils so that FN-III modules may become unfolded [6].

The mechanical stability of individual FN-III modules has been probed in force-unfolding experiments using atomic force microscopy (AFM) [100, 101, 119]. The force-extension profiles obtained from AFM experiments display characteristic saw-tooth patterns, implying that these protein modules unfold in a one-by-one fashion. The unfolding is reversible as fully stretched modules refold in seconds after releasing the external force, which is a crucial property for elastic components. With genetically engineered identical repeats of FN-III modules, the mechanical stability of individual modules can be quantified in terms of rupture forces. Suprisingly, although FN-III modules possess quite similar tertiary structures, they demonstrate distinctly different mechanical stability with peak rupture forces ranging from 75 – 200 pN at a pulling velocity of 0.6 μm/s [100]. More interestingly, FN-III modules such as FN-III$_1$ may have multiple distinct stable conformations, indicating the existence of unfolding intermediates. The mechanical variability and the intermediate states of FN-III modules may be closely related to the function of FN.

The structural reseason why FN-III modules exhibit such mechanical properties is not clear. Although about half FN-III modules have been resolved at high resolution [81, 125], it is difficult to connect the static structures to the dynamic unfolding scenarios that one sees in AFM experiments. To link these static structures to their dynamically changing non-equilibrium states induced under external forces, steered molecular dynamics (SMD) simulations have been introduced to provide an atomic-level description of the unfolding processes for FN-III modules [22, 39, 79, 80]. The SMD method has also been successfully applied to study the mechanical unfolding of other mechanical proteins such as titin domains (reviewed in [40, 58]).

In early approaches the simulated FN-III modules were either solvated in a small water sphere

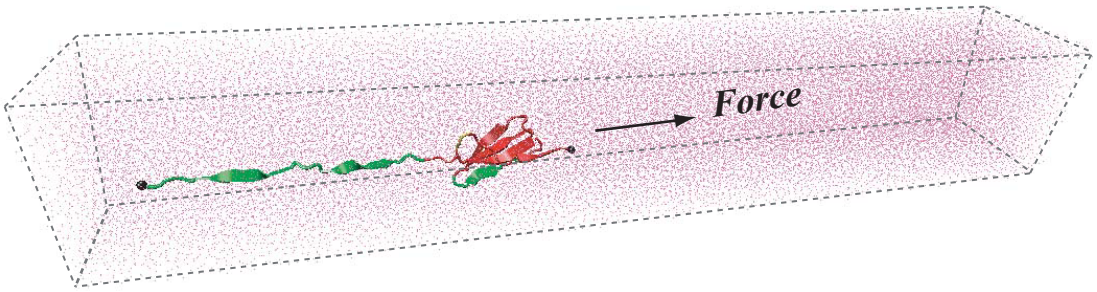

Figure 21: An FN-III$_{10}$ module was solvated and stretched in a 370 Å long water box (130,000 atoms). During the simulations the N-terminus was fixed, and the force was applied along the vector pointing from the N-terminus to the C-terminus.

(typically ∼70 Å in diameter) [22, 80], or treated with implicit solvent models [104]. While the spherical solvent models are sufficiently accurate for short stretching, i.e., stretching within 60 Å, they produce both surface tension and protein de-solvation artifacts, as a protein is stretched beyond the surface of the water sphere. On the other hand, the implicit models overlooked the interactions between the protein and explicit water molecules, which play a key role in disrupting the inter-strand hydrogen bonds that stabilize the whole protein structure [87]. To address the problem of insufficient solvation and obtain a more faithful description of the unfolding dynamics, FN-III modules have been solvated in long water boxes, resulting in large systems of more than 100,000 atoms [39, 79]. Figure 21 displays such a model with the water box long enough to allow a single FN-III$_{10}$ module to be completely extended over 300 Å without encountering a boundary, and accounting for full electrostatics [39]. The system contains 130,000 atoms in total, requiring for one nanosecond simulation with the program NAMD [74] 170 hours running on our group's cluster of 32 1.33GHz Athlon processors (see Fig. 2).

Two SMD protocols, constant velocity and constant force stretching, have been applied to our systems under NVE condition. Of course, under constant velocity stretching conditions energy is not conserved. SMD using restraints moving with constant velocity simulates the stretching of protein domains by an AFM cantilever. One terminus is fixed during the simulation, which corresponds to attaching the protein to a fixed substrate in the AFM experiment. The other terminus is restrained to a point in space (restraint point) by an external, e.g., harmonic, potential. The restraint point is then shifted in a chosen direction [86], typically along the vector pointing from the C-terminus to the N-terminus.

SMD simulations revealed that the main force-bearing elements of FN-III modules arise pri-

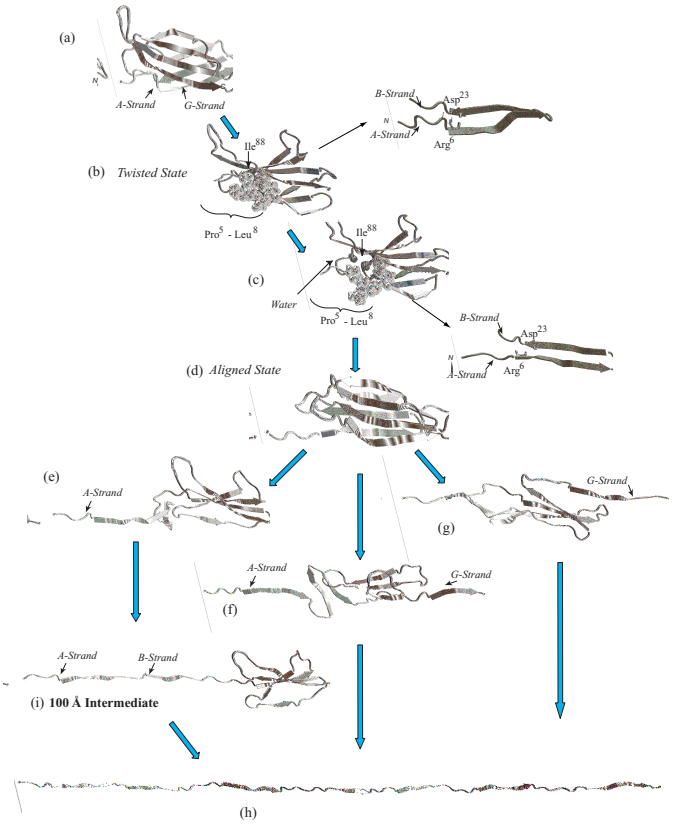

Figure 22: Unfolding pathways for FN-III$_{10}$ revealed by SMD simulations. (a) Equilibrated FN-III$_{10}$. (b) The randomly coiled termini are straightened slightly upon applying a constant force. (c) Rupture of two key inter-strand hydrogen bonds between A- and B-strands permits solvation of the hydrophobic core periphery. (d) Alignment of the two β-sheets before unraveling any β-strands. Separation of the β-strands begins by either separation of (e) A-strand first, (f) A- and G-strands simultaneously, or (g) G-strand first. (h) Fully unraveled structure. (i) In case that the A-strand separates first, a stable intermediate at 100 Å extension arises.

marily from the β-strands close to the termini, specifically, from hydrogen bonding between A- and B-strands and between F- and G-strands, rather than the whole β-sheets [39]. This is similar to titin domain I27, the mechanical stability of which is attributed to the β-strands interweaving two terminal portions. Despite the structure similarity with titin Ig domains, however, FN-III modules are typically weaker than Ig domains, and some modules exhibit stable intermediates with extension much longer than its fully folded length [100]. Unlike the single unfolding pathway that titin modules exhibit, FN-III modules may have multiple unfolding pathways and intermediate states as observed in previous simulations [22, 39, 104]. As shown in Fig. 22 during the unfolding of FN-III$_{10}$, the termini of the module were straightened first upon mechanical stress, entering an intermediate termed "twisted state". As the unfolding proceeds, the hydrophobic core was disrupted and a pair of backbone hydrogen bonds between Arg6 and Asp23 near the C-termini broke, permitting extension up to 25 Å. The protein subsequently entered an intermediate named "aligned state", in which the two β-sheets of FN-III$_{10}$ aligned along the direction of the external force. The two intermediates, twisted and aligned states, are universal to other FN-III modules [22], extending modules up to $\sim$70% of their original length without unraveling the tertiary structure. The aligned intermediate thus may expose partially buried cryptic binding sites. From the aligned state, FN-III$_{10}$ progresses along three alternative unfolding pathways: separating first the A-strand from the module, separating the A- and G-strands simultaneously, or separating first the G-strand. When unraveling of the β-strands began with the A-strand, a stable intermediate at 100 Å was observed. With four-fold extension compared to folded FN-III$_{10}$, the intermediate may serve as a basis for fibronectin self-assembly through a proposed β-strand swapping mechanism [84].

FN-III$_{10}$ mediates cell adhesion to the ECM via its integrin binding motif, Arg-Gly-Asp (RGD), that is located at the apex of the loop connecting β-strands F and G (Fig. 23). The binding of the RGD loop to certain integrins, such as $\alpha_5\beta_1$, is assisted by a synergy site on the neighboring domain FN-III$_9$. SMD simulations found that this interaction can be regulated either through directly detaching the RGD motif [39, 80] or through increasing the distance between the synergy site and the RGD loop [79]. By separating the G-strand from the remaining fold, the distance between the apex of the RGD-containing loop and the module surface was shortened. This conformational change reduces the affinity to integrin. The RGD loop thus constitutes a mechanosensitive switch for recognition by integrin receptors [80]. SMD simulations of the FN-III$_{9-10}$ dimer [79] identified an intermediate in which the length of the linker chain between the two modules is increased by 17 Å, a change that could also switch off the signaling between FN modules and transmembrane

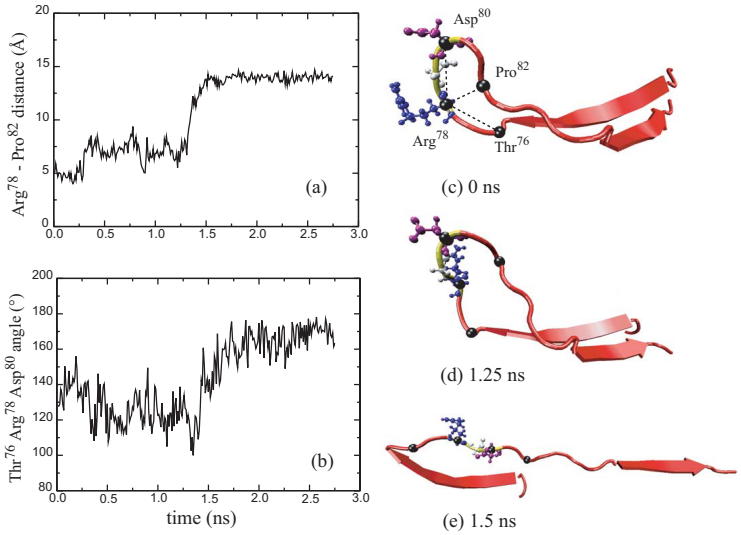

Figure 23: Conformational changes of the RGD-containing loop upon a constant external force. (a) The width of the RGD loop is measured by the distance between the C_α atoms of Arg78 and Pro82. (b) Angle formed by the C_α atoms of residues Thr76, Arg78, and Asp80 illustrates a change from a bent ($\sim 110°$) to a more planar ($\sim 170°$) conformation. (c) Snapshot of F- and G-strands and the connecting loop in the native structure. (d) The RGD loop remains intact in the intermediate state at 100 Å. (e) The RGD loop is straightened out after the protein passes the intermediate state.

protein integrins.

Recently published crystal structures of integrin $\alpha_V\beta_3$ [144, 145] provide a further opportunity to study the signaling pathways between integrin and fibronectin modules. Simulations of integrin $\alpha_V\beta_3$ containing RGD peptide in the binding pocket with 136,000 atoms are in progress and may reveal how integrins mediate the adhesion with FN ligands. However, the ultimate goal of simulations is not only to reveal the dynamic processes of FN-III unfolding or dissociation from integrins, but also to understand, in combination with experiments, how the individual FN modules contribute to the overall mechanical properties of FN fibrils and how cell receptors and the ECM proteins respond and exchange mechanical signals.

Outlook

We have illustrated above the state of art in large scale biomolecular modeling. Such modeling has become feasible only through parallel computing utilizing hundreds and soon thousands of processors. Fortunately, the necessary computers are available to many researchers and the needed

new generation of molecular dynamics and molecular visualization programs have been developed and are widely shared. During the next years the new programs will mature and most likely merge with the familiar, feature rich first generation of modeling programs.

Interestingly, most large scale simulation projects in our and other groups have been carried out in close collaboration with experimentalists. The projects have not been accomplished for the record books, but were motivated by a desire to extend the resolution of observation to atomic level dynamics and to explain thereby the mechanisms underlying important cellular processes. As stated already above, biomolecular modeling is being adopted today by experimentalists to guide and explain observations; often these researchers simulate large systems.

Indeed, size is not really a key obstacle today and will be less so in the near future. Nevertheless, it seems frivolous to describe extremely large systems uniformly at the atomic level when one is only interested in a small part, e.g., an active site. Unfortunately, the old dream of computational approaches that permit one to dial an effective resolution for different parts of a system is apparently hard to realize; legions of researchers have attempted a solution of such multi-scale modeling, but successes have been rare. Computational biologists seek to replace an explicit simulation of water molecules in bulk water by a so-called implicit solvent model and much effort has been invested in this approach. Other approaches replace an atomic level simulation of DNA strands by elastic rod models to simplify the description of protein-DNA complexes. Attempts have also been made to systematically reduce the number of simulated atoms by placing a majority of atoms into a class of "slaved" atoms that move along with the other atoms according to a fixed mathematical formula [108]. None of the approaches described can yet be employed in an automated way with controlled accuracy. These approaches may also be ill-fated due to computer power permitting more and more easily full atom descriptions and outrunning the development of the necessarily complex multi-scale algorithms.

But universally accepted as a bottleneck is the presently short 10-100 ns time scale of simulations that is even for the fastest cellular processes five orders of magnitude too short. Parallelization is not a choice in this regard since long time dynamics is inherently sequential; a general parallel solution to the time scale problem has not been found and very well may not exist. The only remedy is an application of external forces in so-called steered molecular dynamics simulations that accelerate processes. The problem with this approach is not so much that the forces applied are artificial, but that one does not know *a priori* in which direction in the multi-dimensional space of biopolymer conformations forces should be applied. Such knowledge implies that the reaction

path and, hence, the mechanism for the cellular process studied is known, but often one neither knows this path nor does one even have a suitable hypothesis. The problem may be overcome with a method called interactive molecular dynamics [45] that applies forces like steered molecular dynamics, but does so interactively in a running simulation, permitting one to tinker with a system and, thereby, develop the needed hypotheses.

Simulations of 100,000 atoms and more accomplished so far for the purpose of explaining the physical mechanism of cellular processes focus as much as smaller scale simulations on atomic level detail in key moieties of biopolymers, e.g., on channel lining, force bearing secondary structure elements, active sites or even electronic degrees of freedom [23]. The main reason for such simulations is to capture mechanisms in the context of integral functions of biopolymeric machineries. Also, large scale simulations are necessary when one does not know from the outset the essential components of a cell's machinery. For example, presently the mechanistic essentials of ATP synthase that couple reversibly proton currents to stalk rotation, and stalk rotation to catalysis, are all but unknown forcing all encompassing, i.e., large scale, simulations to avoid missing key features.

A further argument for large simulations are the recently discovered structures of the most basic machineries of living cells, like redox chains in bioenergetic membranes, electrical channels of neural membranes, or the genome storage and expression machinery, which involve large scale systems that cannot be readily subdivided.

The ultimate argument for large scale simulations focuses on the most fundamental question of life sciences: Where in the hierarchies of assemblies found in cells does the step from inorganic matter to living systems occur? Living systems are all made of many molecular components that self-assemble, control and repair each other, and self-replicate. In the past, molecular biologists have taken the molecular machines in living cells apart and learned what they are made of. Now they seek to learn how the molecular constituents of cells assemble into machines, cellular structures, and complete living cells. Large scale simulations investigating assemblies of biopolymers will provide life scientists with a unique tool to answer the question " What is life?".

References

[1] J. Abrahams, A. Leslie, R. Lutter, and J. Walker. Structure at 2.8-Å resolution of F_1-ATPase from bovine heart mitochondria. *Nature*, 370:621–628, 1994.

[2] P. Agre, M. Bonhivers, and M. J. Borgnia. The aquaporins, blueprints for cellular plumbing systems. *Journal of Biological Chemistry*, 273:14659–14662, 1998.

[3] P. Agre, M. D. Lee, S. Devidas, and W. B. Guggino. Aquaporiuns and ion conductance. *Science*, 275:1490–1490, 1997.

[4] A. Aksimentiev, I. Balabin, R. H. Fillingame, and K. Schulten. Exploring protein motors on multiple time scales: F_o ATP synthase. *Submitted to PNAS*, 2003.

[5] N. Ban, P. Nissen, J. Hansen, M. Capel, P. B. Moore, and T. A. Steitz. Placement of protein and rna structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature*, 400:841–847, 1999.

[6] G. Baneyx, L. Baugh, and V. Vogel. Supramolecular chemistry and self-assembly special feature: fibronectin extension and unfolding within cell matrix fibrils controlled by cytoskeletal tension. *Proceedings of the National Academy of Sciences, USA*, 99:5139–43, 2002.

[7] J. Baudry, E. Tajkhorshid, F. Molnar, J. Phillips, and K. Schulten. Molecular dynamics study of bacteriorhodopsin and the purple membrane. *Journal of Physical Chemistry B*, 105:905–918, 2001.

[8] M. V. Bayas, K. Schulten, and D. Leckband. Simulations of the forced detachment of the CD2-CD58 complex. *Biophysical Journal*, 2003. In press.

[9] R. Birkenhager, M. Hoppert, G. Deckers-Hebestreit, F. Mayer, and K. Altendorf. The F_0 complex of the *Escherichia coli* ATP synthase. Investigation by electron spectroscopic imaging and immunoelectron microscopy. *European Journal of Biochemistry*, 230:58–67, 1995.

[10] T. C. Bishop, D. Kosztin, and K. Schulten. How hormone receptor–DNA binding affects nucleosomal DNA: The role of symmetry. *Biophysical Journal*, 72:2056–2067, 1997.

[11] R. A. Böckmann and H. Grubmüller. Nanoseconds molecular dynamics simulation of primary mechanical energy transfer steps in F_1-ATP synthase. *Nature Structural Biology*, 9:198–202, 2002.

[12] M. Borgnia, S. Nielsen, A. Engel, and P. Agre. Cellular and molecular biology of the aquaporin water channels. *Annual Review of Biochemistry*, 68:425–458, 1999.

[13] M. J. Borgnia and P. Agre. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA*, 98:2888–2893, 2001.

[14] M. J. Borgnia, D. Kozono, G. Calamita, P. C. Maloney, and P. Agre. Functional reconstitution and characterization of AqpZ, the E-coli water channel protein. *Journal of Molecular Biology*, 291(5):1169–1179, 1999.

[15] P. D. Boyer. Catalytic site forms and controls in ATP synthase catalysis. *Biochimica et Biophysica Acta – Bioenergetics*, 1458(2-3):252–262, 2000.

[16] M. L. Brewer, U. W. Schmitt, and G. A. Voth. The formation and dynamics of proton wires in channel environments. *Biophysical Journal*, 80(4):1691–1702, 1999.

[17] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4:187–217, 1983.

[18] A. T. Brünger. *X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR*. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, 1992.

[19] R. A. Capaldi and R. Aggeler. Mechanism of F_1F_0-type ATP synthase, a biological rotary motor. *Trends in Biochemical Sciences*, 27:154–160, 2002.

[20] D. A. Cherepanov, A. Y. Mulkidjaninan, and W. Junge. Transient accumulation of elastic energy in proton translocating ATP synthase. *FEBS Letters*, 449:1–6, 1999.

[21] W. M. Clemons Jr, J. L. C. May, B. T. Wimberly, J. P. McCutcheon, M. S. Capel, and V. Ramakrishnan. Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature*, 400:833–840, 1999.

[22] D. Craig, A. Krammer, K. Schulten, and V. Vogel. Comparison of the early stages of forced unfolding of fibronectin type III modules. *Proceedings of the National Academy of Sciences, USA*, 98:5590–5595, 2001.

[23] A. Damjanović, I. Kosztin, U. Kleinekathoefer, and K. Schulten. Excitons in a photosynthetic light-harvesting system: A combined molecular dynamics, quantum chemistry and polaron model study. *Physical Review E*, 65:031919, 2002. (24 pages).

[24] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald. An N·log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98:10089–10092, 1993.

[25] B. L. de Groot, A. Engel, and H. Grubmüller. A refined structure of human aquaporin-1. *FEBS Letters*, 504:206–211, 2001.

[26] B. L. de Groot and H. Grubmüller. Water permeation across biological membranes: Mechanism and dynamics of aquaporin-1 and GlpF. *Science*, 294:2353–2357, 2001.

[27] B. L. de Groot, J. B. Heymann, A. Engel, K. Mitsuoka, Y. Fujiyoshi, and H. Grubmüller. The fold of human aquaporin 1. *Journal of Molecular Biology*, 300:987–994, 2000.

[28] P. M. T. Deen and C. H. van Os. Epithelial aquaporins. *Current Opinion in Cell Biology*, 10(4):435–442, 1998.

[29] O. Dmitriev, P. C. Jones, and R. H. Fillingame. Structure of the subunit *c* oligomer in the F_1F_0 ATP synthase: Model derived from solution structure of the monomer and cross-linking in the native enzyme. *Proceedings of the National Academy of Sciences, USA*, 96:7785–7790, 1999.

[30] S. D. Dunn, M. Revington, D. J. Cipriano, and B. H. Shilton. The *b* subunit of *escherichia coli* ATP synthase. *Journal of Bioenergetics and Biomembranes*, 32(4):347–355, 2000.

[31] T. Elston, H. Wang, and G. Oster. Energy transduction in ATP synthase. *Nature*, 391:510–513, 1998.

[32] H. Erickson. Reversible unfolding of fibronectin type III and immunoglobulin domains provides the structural basis for stretch and elasticity of titin and fibronectin. *Proceedings of the National Academy of Sciences, USA*, 91:10114–10118, 1994.

[33] R. H. Fillingame. Coupling H^+ transport and ATP synthesis in F_1F_0-atp synthases: Glimpses of interacting parts in a dynamic molecular machine. *Journal of Experimental Biology*, 200:217–224, 1997.

[34] R. H. Fillingame, C. M. Angevine, and O. Y. Dmitriev. Coupling proton movements to *c*-ring rotation in F_1F_o ATP synthase: aqueous access channels and helix rotations at the *a*-*c* interface. *Biochimica et Biophysica Acta – Bioenergetics*, 1555(1–3):29–36, 2002.

[35] R. H. Fillingame, W. Jiang, and O. Y. Dmitriev. Coupling H^+ transport to rotary catalysis in F-type ATP synthases: Structure and organization of the transmembrane rotary motor. *Journal of Experimental Biology*, 203:9–17, 2000.

[36] R. H. Fillingame, W. Jiang, O. Y. Dmitriev, and P. C. Jones. Structural interpretation of F_0 rotary function in the *Escherichia coli* F_1F_0 ATP synthase. *Biochimica et Biophysica Acta*, 1458:387–403, 2000.

[37] D. Fu, A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski, and R. M. Stroud. Structure of a glycerol conducting channel and the basis for its selectivity. *Science*, 290:481–486, 2000.

[38] M. Futai, H. Omote, Y. Sambongi, and Y. Wada. Synthase (H^+ ATPase): Coupling between catalysis, mechanical work, and proton translocation. *Biochimica et Biophysica Acta – Bioenergetics*, 1458(2-3):276–288, 2000.

[39] M. Gao, D. Craig, V. Vogel, and K. Schulten. Identifying unfolding intermediates of FN-III_{10} by steered molecular dynamics. *Journal of Molecular Biology*, 323:939–950, 2002.

[40] M. Gao, H. Lu, and K. Schulten. Unfolding of titin domains studied by molecular dynamics simulations. *Journal of Muscle Research and Cell Motility*, 2002. In press.

[41] B. Geiger, A. Bershadsky, R. Pankov, and K. Yamada. Transmembrane crosstalk between the extracellular matrix and the cytoskeleton. *Nat. Rev. Mol. Cell. Biol.*, 2:793–805, 2001.

[42] C. Gibbons, M. G. Montgomery, A. G. W. Leslie, and J. E. Walker. The structure of the central stalk in bovine F_1-ATPase at 2.4 Å resolution. *Nature Structural Biology*, 7:1055–1061, 2000.

[43] M. E. Girvin, V. K. Rastogi, F. Abildgaard, J. L. Markley, and R. H. Fillingame. Solution structure of the transmembrane H^+-transporting subunit *c* of the F_1F_0 ATP synthase. *Biochemistry*, 37:8817–8824, 1998.

[44] A. L. Gnatt, P. Cramer, J. H. Fu, D. A. Bushnell, and R. D. Kornberg. Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science*, 292(5523):1876–1882, 2001.

[45] P. Grayson, E. Tajkhorshid, and K. Schulten. Mechanisms of selectivity in channels and enzymes studied with interactive molecular dynamics. *Biophysical Journal*, 2003. In press.

[46] G. Groth. Molecular models of structural arrangement of subunits and the mechanism of proton translocation in the membrane domain. *Biochim. Biophys. Acta*, 1458:417–427, 2000.

[47] J. Gullingsrud, D. Kosztin, and K. Schulten. Structural determinants of MscL gating studied by molecular dynamics simulations. *Biophysical Journal*, 80:2074–2081, 2001.

[48] F. M. Harold. *The Way of the Cell – Molecules, Organisms and the Order of Life*. Oxford University Press, New York, 2001.

[49] H. Heller, M. Schaefer, and K. Schulten. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal-phases. *Journal of Physical Chemistry*, 97:8343–8360, 1993.

[50] K. B. Heller, E. C. Lin, and T. H. Wilson. Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. *J. Bacteriol.*, 144:274–278, 1980.

[51] J. Hermolin, J. Gallant, and R. H. Fillingame. Topology, organization, and function of the psi subunit in the f_0 sector of the H^+-ATPase of *escherichia coli*. *Journal of Biological Chemistry*, 258(23):14550–14555, 1983.

[52] J. B. Heynmann and A. Engel. Aquaporins: Phylogeny, structure, and physiology of water channels. *News in Physiological Sciences*, 14:187–193, 1999.

[53] Y. Hirono-Hara, H. Noji, M. Nishiura, E. Muneyuki, K. Y. Hara, R. Yasuda, K. Kinosita, Jr., and M. Yoshida. Pause and rotation of F_1-ATPase during catalysis. *Proceedings of the National Academy of Sciences, USA*, 98:13649–13654, 2001.

[54] G. Hummer, J. C. Rasaiah, and J. P. Noworyta. Water conduction through the hydrophobic channel of a carbon nanotube. *Nature*, 414:188–190, 2001.

[55] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[56] R. O. Hynes. *Fibronectins*. Springer-Verlag, New York, 1990.

[57] B. Isralewitz, J. Baudry, J. Gullingsrud, D. Kosztin, and K. Schulten. Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics and Modeling*, 19:13–25, 2001. Also in *Protein Flexibility and Folding* , L. A. Kuhn and M. F. Thorpe, editors, *Biological Modeling Series* , (Elsevier).

[58] B. Isralewitz, M. Gao, and K. Schulten. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, 11:224–230, 2001.

[59] B. Isralewitz, E. Tajkhorshid, and K. Schulten. Simulated forced rotation of F_1-ATP synthase central stalk. 2003.

[60] S. Izrailev, A. R. Crofts, E. A. Berry, and K. Schulten. Steered molecular dynamics simulation of the Rieske subunit motion in the cytochrome bc_1 complex. *Biophysical Journal*, 77:1753–1768, 1999.

[61] S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten. Steered molecular dynamics. In P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel, editors, *Computational Molecular Dynamics: Challenges, Methods, Ideas*, volume 4 of *Lecture Notes in Computational Science and Engineering*, pages 39–65. Springer-Verlag, Berlin, 1998.

[62] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master equation approach. *Physical Review E*, 56:5018–5035, 1997.

[63] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78:2690–2693, 1997.

[64] M. Jensen, E. Tajkhorshid, and K. Schulten. Electrostatic control of permeation and selectivity in aquaporin water channels. 2003. submitted.

[65] M. Ø. Jensen, S. Park, E. Tajkhorshid, and K. Schulten. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proceedings of the National Academy of Sciences, USA*, 99:6731–6736, 2002.

[66] M. Ø. Jensen, E. Tajkhorshid, and K. Schulten. The mechanism of glycerol conduction in aquaglyceroporins. *Structure*, 9:1083–1093, 2001.

[67] W. Jiang, J. Hermolin, and R. H. Fillingame. The preferred stoichiometry of c subunits in the rotary motor sector of *Escherichia coli* ATP synthase is 10. *Proceedings of the National Academy of Sciences, USA*, 98:4966–4971, 2001.

[68] P. C. Jones and R. H. Fillingame. Genetic fusion of subunit c in the F_o sector of H^+-transporting ATP synthase. *J. Biol. Chem.*, 273(45):29701–29705, 1998.

[69] P. C. Jones, W. Jiang, and R. H. Fillingame. Arrangement of the multicopy H^+-translocating subunit c in the membrane sector of the *Escherichia coli* F1F0 ATP synthase. *Journal of Biological Chemistry*, 273:17178–17185, 1998.

[70] J. S. Jung, G. M. Preston, B. L. Smith, W. B. Guggino, and P. Agre. Molecular structure of the water channel through aquaporin CHIP – the hourglass model. *Journal of Biological Chemistry*, 269:14648–14654, 1994.

[71] W. Junge, H. Lill, and S. Engelbrecht. Atp synthase: An electrochemical transducer with rotary mechanics. *Trends in Biochemical Sciences*, 22:420–423, 1997.

[72] W. Junge, O. Pänke, D. A. Cherepanov, K. Gumbiowski, M. Müller, and S. Englebrecht. Inter-subunit rotation and elastic power transmission in F_0F_1-ATPase. *FEBS Letters*, 504(3):152–160, 2001.

[73] Y. Kagawa, T. Hamamoto, and H. Endo. The α/β interfaces of $\alpha_1\beta_1$, $\alpha_3\beta_3$, and F_1: domain motions and elastic energy stored during γ rotation. *Journal of Bioenergetics and Biomembranes*, 32:471–484, 2000.

[74] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, 151:283–312, 1999.

[75] L. V. Kalé and S. Krishnan. Charm++: Parallel programming with message-driven objects. In G. V. Wilson and P. Lu, editors, *Parallel Programming using C++*, pages 175–213. MIT Press, 1996.

[76] Y. Kato-Yamada, H. Noji, K. K. J. Ryohei Yasuda, and M. Yoshida. Direct observation of the rotation of ϵ subunit in f_1-atpase. *Journal of Biological Chemistry*, 273:19375–19377, 1998.

[77] K. Kinosita, Jr., R. Yasuda, and K. Adachi. A rotary molecular motor that can work at near 100% efficiency. *Phil. Trans. Roy. Soc. B*, 355:473–489, 2000.

[78] D. Kosztin, T. C. Bishop, and K. Schulten. Binding of the estrogen receptor to DNA: The role of waters. *Biophysical Journal*, 73:557–570, 1997.

[79] A. Krammer, D. Craig, W. E. Thomas, K. Schulten, and V. Vogel. A structural model for force regulated integrin binding to fibronectin's RGD-synergy site. *Matrix Biology*, 21:139–147, 2002.

[80] A. Krammer, H. Lu, B. Isralewitz, K. Schulten, and V. Vogel. Forced unfolding of the fibronectin type III module reveals a tensile molecular recognition switch. *Proceedings of the National Academy of Sciences, USA*, 96:1351–1356, 1999.

[81] D. J. Leahy, I. Aukhil, and H. P. Erickson. 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell*, 84:155–164, 1996.

[82] J. Li and A. S. Verkman. Impaired hearing in mice lacking aquaporin-4 water channels. *Journal of Biological Chemistry*, 276:31233–31237, 2001.

[83] E. Lindahl, B. Hess, and D. van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 2001.

[84] S. V. Litvinovich, S. Brew, S. Aota, S. Akiyama, C. Haudenschild, and K. Ingham. Formation of amyloid-like fibrils by self-association of a partially unfolded fibronectin type III module. *Journal of Molecular Biology*, 280:245–58, 1998.

[85] D. LU, P. Grayson, and K. Schulten. Conductance and physical asymmetry of the *escherichia coli* glycerol facilitator GlpF. 2003. submitted.

[86] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical Journal*, 75:662–671, 1998.

[87] H. Lu and K. Schulten. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophysical Journal*, 79:51–65, 2000.

[88] T. W. Lynch, D. Kosztin, M. A. McLean, K. Schulten, and S. G. Sligar. Dissecting the molecular origins of protein-nucleic acid recognition: Hydrostatic pressure and molecular dynamics. *Biophysical Journal*, 82:93–98, 2002.

[89] J. Ma, T. C. Flynn, Q. Cui, A. G. W. Leslie, J. E. Walker, and M. Karplus. A dynamics analysis of the rotation mechanism for conformational change in F_1-ATPase. *Structure*, 10:921–931, 2002.

[90] A. D. MacKerell Jr., D. Bashford, M. Bellott, et al. Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. *FASEB Journal*, 6(1):A143–A143, 1992.

[91] A. D. MacKerell Jr., B. Brooks, I. C. L. Brooks, L. Nilsson, B. Roux, Y. Won, and M. Karplus. Charmm: The energy function and its parameterization with an overview of the program. pages 271–277. John Wiley & Sons, Chichester, 1998.

[92] T. Masaike, N. Mitome, H. Noji, E. Muneyuki, R. Yasuda, K. Kinosita, Jr., and M. Yoshida. Rotation of F_1-ATPase and the hinge structures of the β subunit. *Journal of Experimental Biology*, 203:1–8, 2000.

[93] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, June 1977.

[94] R. I. Menz, J. E. Walker, and A. G. W. Leslie. Structure of bovine mitochondrial F_1-ATPase with nucleotide bound to all three catalytic sites: Implications for the mechanism of rotary catalysis. *Cell*, 106:331–341, 2001.

[95] A. Mogilner, H. Wang, T. Elston, and G. Oster. Molecular motors: Theory & experiment. In C. Fall, E. Marland, J. Wagner, and J. Tyson, editors, *Computational Cell Biology*, chapter 12, pages 321–380. Springer, New York, 2002.

[96] R. A. Monticello and W. S. Brusilow. Role of the delta subunit in enhancing proton conduction through the F0 of the Escherichia coli F1F0 ATPase. *J. Bacteriol.*, 176(5):1383–1389, 1994.

[97] K. Murata, K. Mitsuoka, T. Hirai, T. Walz, P. Agre, J. B. Heymann, A. Engel, and Y. Fujiyoshi. Structural determinants of water permeation through aquaporin-1. *Nature*, 407:599–605, 2000.

[98] H. Noji, K. Häsler, W. Junge, K. Kinosita Jr, M. Yoshida, and S. Engelbrecht. Rotation of *escheria coli* F_1-ATPase. *Biochemical and Biophysical Research Communications*, 260:597–599, 1999.

[99] H. Noji, T. Yasuda, M. Yoshida, and K. Kinosita Jr. Direct observation of the rotation of F1-ATPase. *Nature*, 386:299–302, 1997.

[100] A. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vazquez, and J. Fernandez. The mechanical hierarchies of fibronectin observed with single molecule AFM. *Journal of Molecular Biology*, 319:433–447, 2002.

[101] A. F. Oberhauser, P. E. Marszalek, H. Erickson, and J. Fernandez. The molecular elasticity of tenascin, an extracellular matrix protein. *Nature*, 393:181–185, 1998.

[102] T. Ohashi, D. P. Kiehart, and H. P. Erickson. Dynamics and elasticity of the fibronectin matrix in living cell culture visualized by fibronectin-green fluorescent protein. *Proceedings of the National Academy of Sciences, USA*, 96:2153–2158, 1999.

[103] G. Oster and H. Wang. Reverse engineering a protein: the mechanochemistry of ATP synthase. *Biochimica et Biophysica Acta*, 1458:482–510, 2000.

[104] E. Paci and M. Karplus. Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. *Journal of Molecular Biology*, 288:441–459, 1999.

[105] O. Pänke, D. A. Cherepanov, K. Gumbowski, S. Engelbrecht, and W. Junge. Viscoelastic dyanamics of actin filaments coupled to rotary F-ATPase: angular torque profile of the enzyme. *Biophysical Journal*, 81:1220–1233, 2001.

[106] O. Pänke and B. Rumberg. Kinetic modeling of rotary CF_oF_1-ATP synthase: storage of elastic energy during energy transduction. *Biochimica et Biophysica Acta*, 1412:118–128, 1999.

[107] J. C. Phillips, W. Wriggers, Z. Li, A. Jonas, and K. Schulten. Predicting the structure of apolipoprotein A-I in reconstituted high density lipoprotein disks. *Biophysical Journal*, 73:2337–2346, 1997.

[108] R. Phillips, M. Dittrich, and K. Schulten. Quasicontinuum representations of atomic-scale mechanics: From proteins to dislocations. *Annual Review of Materials Research*, 32:219–233, 2002.

[109] R. Pomès and B. Roux. Structure and dynamics of a proton wire: A theoretical study of H^+ translocation along the single-file water chain in the gramicidin A channel. *Biophysical Journal*, 71(1):19–39, 1996.

[110] R. Pomès and B. Roux. Theoretical study of H^+ translocation along a model proton wire. *Journal of Physical Chemistry*, 100(7):2519–2527, 1996.

[111] R. Pomès and B. Roux. Molecular mechanism of H^+ conduction in the single-file water chain of the gramicidin channel. *Biophysical Journal*, 82:2304–2316, 2002.

[112] G. M. Preston and P. Agre. Isolation of the cDNA for erythrocyte integral membrane-protein of 28-kD – member of an ancient channel family. *Proceedings of the National Academy of Sciences, USA*, 88:11110–11114, 1991.

[113] H. S. Randa, L. R. Forrest, G. A. Voth, and M. S. P. Sansom. Molecular dynamics of synthetic leucine-serine ion channels in a phospholipid membrane. *Biophysical Journal*, 77(5):2400–2410, 1999.

[114] V. K. Rastogi and M. E. Girvin. Structural changes linked to proton translocation by subunit *c* of the ATP synthase. *Nature*, 402:263–268, 1999.

[115] G. Ren, V. S. Reddy, A. Cheng, P. Melnyk, and A. K. Mitra. Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proceedings of the National Academy of Sciences, USA*, 98:1398–1403, 2001.

[116] H. Ren and W. S. Allison. On what makes the γ subunit spin during ATP hydrolysis by F_1. *Biochimica et Biophysica Acta – Bioenergetics*, 1458:221–233, 2000.

[117] L. Reuss. Focus on "effect of expressing the water channel aquaporin-1 on the CO2 permeability of *Xenopus* oocytes". *American Journal of Physiology*, 274:C297–C298, 1998.

[118] D. P. Rickey and E. E. C. Lin. Importance of facilitated diffusion for effective utilization of glycerol by Escherichia coli. *Journal of Bacteriology*, 112(2):784–790, 1972.

[119] M. Rief, M. Gautel, A. Schemmel, and H. Gaub. The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by AFM. *Biophysical Journal*, 75:3008–3014, 1998.

[120] J. Saam, E. Tajkhorshid, S. Hayashi, and K. Schulten. Molecular dynamics investigation of primary photoinduced events in the activation of rhodopsin. *Biophysical Journal*, 83:3097–3112, 2002.

[121] M. S. P. Sansom and R. J. Law. Membrane proteins: Aquaporins – channels without ions. *Current Biology*, 11:R71–R73, 2001.

[122] J. Schlitter, M. Engels, P. Krüger, E. Jacoby, and A. Wollmer. Targeted molecular dynamics simulation of conformational change — application to the T $\leftrightarrow$ R transition in insulin. *Molecular Simulation*, 10(2–6):291–308, 1993.

[123] J. E. Schwarzbauer and J. L. Sechler. Fibronectin fibrillogenesis: a paradigm for extracellular matrix assembly. *Current Opinion in Cell Biology*, 11:622–627, 1999.

[124] A. E. Senior, S. Nadanaciva, and J. Weber. The molecular mechanism of ATP synthesis by F_1F_o-ATP synthase. *Biochimica et Biophysica Acta*, 1553:188–211, 2002.

[125] A. Sharma, J. A. Askari, M. J. Humphries, E. Y. Jones, and D. I. Stuart. Crystal structure of a heparin- and integrin-binding segment of human fibronectin. *EMBO Journal*, 18:1468–1479, 1999.

[126] L. B. Shi, W. R. Skach, and A. S. Verkman. Functional independence of monomeric CHIP28 water channels revealed by expression of wild-type-mutant heterodimers. *Journal of Biological Chemistry*, 269:10417–10422, 1994.

[127] D. Stock, A. G. W. Leslie, and J. E. Walker. Molecular architecture of the rotary motor in ATP synthase. *Science*, 286:1700–1705, 1999.

[128] H. Sui, B.-G. Han, J. K. Lee, P. Walian, and B. K. Jap. Structural basis of water-specific transport through the AQP1 water channel. *Nature*, 414:872–878, 2001.

[129] S. Sun, D. Chandler, A. R. Dinner, and G. Oster. Elastic energy storage in F_1-ATPase. *Biophysical Journal*, 2003. In Press.

[130] E. Tajkhorshid, P. Nollert, M. Ø. Jensen, L. J. W. Miercke, J. O'Connell, R. M. Stroud, and K. Schulten. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science*, 296:525–530, 2002.

[131] M. Takeyama, T. Noumi, M. Maeda, and M. Futai. F_o portion of *escherichia coli* H^+-ATPase. *Journal of Biological Chemistry*, 31(5):16106–16112, 1988.

[132] K. Takeyasu, H. Omote, S. Nettikadan, F. Tokumasu, A. Iwamoto-Kihara, and M. Futai. Molecular imaging of *Escherichia coli* F0F1-ATPase in reconstituted membranes using atomic force microscopy. *FEBS Letters*, 392:110–113, 1996.

[133] F. I. Valiyaveetil and R. H. Fillingame. On the role of Arg-210 and Glu-219 of subunit *a* in proton translocation by the *Escherichia coli* F_1F_0 ATP synthase. *Journal of Biological Chemistry*, 272:32635–32641, 1997.

[134] A. S. Verkman and A. K. Mitra. Structure and function of aquaporin water channels. *American Journal of Physiology – Renal Physiology*, 278:F13–F28, 2000.

[135] V. Vogel, W. E. Thomas, D. W. Craig, A. Krammer, and G. Baneyx. Structural insights into the mechanical regulation of molecular recognition sites. *Trends in Biotechnology*, 19:416–423, 2001.

[136] T. Walz and R. Ghosh. Two-dimensional crystallization of the light-harvesting I reaction centre photounit from *Rhodospirillum rubrum*. *Journal of Molecular Biology*, 265:107–111, 1997.

[137] T. Walz, T. Hirai, K. Murata, J. B. Heynmann, K. Mitsuoka, Y. Fujiyoshi, B. L. Smith, P. Agre, and A. Engel. The three-dimensional structure of aquaporin-1. *Nature*, 387:624–627, 1997.

[138] H. Wang and G. Oster. Ratchets, power strokes, and molecular motors. *Applied Physics A*, 75:315–323, 2002.

[139] H. Y. Wang and G. Oster. Energy transduction in the F_1 motor of ATP synthase. *Nature*, 396:279–282, 1998.

[140] J. Weber and A. E. Senior. ATP synthase: what we know about ATP hydrolysis and what we do not know about ATP synthesis. *Biochimica et Biophysica Acta – Bioenergetics*, 1458:300–309, 2000.

[141] P. K. Weiner and P. A. Kollman. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *Journal of Computational Chemistry*, 2(3):287–303, 1981.

[142] G. J. Wistow, M. M. Pisano, and A. B. Chepelinsky. Tandem sequence repeats in transmembrane channel proteins. *Trends in Biochemical Sciences*, 16:170–171, 1991.

[143] W. Wriggers and K. Schulten. Investigating a back door mechanism of actin phosphate release by steered molecular dynamics. *PROTEINS: Structure, Function, and Genetics*, 35:262–273, 1999.

[144] J. P. Xiong, T. Stehle, B. Diefenbach, R. Zhang, R. Dunker, D. L. Scott, A. Joachimiak, S. L. Goodman, and M. A. Arnaout. Crystal structure of the extracellular segment of integrin αvβ3. *Science*, 294:339–345, 2001.

[145] J. P. Xiong, T. Stehle, R. Zhang, A. Joachimiak, M. Frech, S. L. Goodman, and M. A. Arnaout. Crystal structure of the extracellular segment of integrin αvβ3 in complex with an arg-gly-asp ligand. *Science*, 296:151–155, 2002.

[146] R. Yasuda, H. Noji, M. Yoshida, K. Kinosita Jr, and H. Itoh. Resolution of distinct rotational substeps by submillisecond kinetic analysis of F_1-ATP-ase. *Nature*, 410:898–904, 2001.

[147] M. L. Zeidel, S. V. Ambudkar, B. L. Smith, and P. Agre. Reconstitution of functional water channels in liposomes containing purified red cell CHIP28 protein. *Biochemistry*, 31:7436–7440, 1992.

[148] L. Zhang and J. Hermans. Hydrophilicity of cavities in proteins. *PROTEINS: Structure, Function, and Genetics*, 24:433–438, 1996.

[149] Y. Zhang and R. H. Fillingame. Essential aspartates in subunit c of F_1F_0 ATP synthase. *Journal of Biological Chemistry*, 269:5473–5479, 1994.

[150] F. Zhou and K. Schulten. Molecular dynamics study of a membrane–water interface. *Journal of Physical Chemistry*, 99:2194–2208, 1995.

[151] F. Zhu and K. Schulten. Water and proton conduction through carbon nanotubes as models for biological channels. *Biophysical Journal*, 2003. In press.

[152] F. Zhu, E. Tajkhorshid, and K. Schulten. Molecular dynamics study of aquaporin-1 water channel in a lipid bilayer. *FEBS Letters*, 504:212–218, 2001.

[153] F. Zhu, E. Tajkhorshid, and K. Schulten. Pressure-induced water transport in membrane channels studied by molecular dynamics. *Biophysical Journal*, 83:154–160, 2002.