Chapter 9
Protein Dynamics: From Structure to Function

Marcus B. Kubitzki, Bert L. de Groot, and Daniel Seeliger

Abstract Understanding protein function requires detailed knowledge about protein dynamics, i.e. the different conformational states the system can adopt. Despite substantial experimental progress, simulation techniques such as molecular dynamics (MD) currently provide the only routine means to obtain dynamical information at an atomic level on timescales of nano- to microseconds. Even with the current development of computational power, sampling techniques beyond MD are necessary to enhance conformational sampling of large proteins and assemblies thereof. The use of collective coordinates has proven to be a promising means in this respect, either as a tool for analysis or as part of new sampling algorithms. Starting from MD simulations, several enhanced sampling algorithms for biomolecular simulations are reviewed in this chapter. Examples are given throughout illustrating how consideration of the dynamic properties of a protein sheds light on its function.

9.1 Molecular Dynamics Simulations

Over the last decades, experimental techniques have made substantial progress in revealing the three-dimensional structure of proteins, in particular X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy. Going beyond the static picture of single protein structures has proven to be more challenging, although, a number of techniques such as NMR relaxation, fluorescence spectroscopy or time-resolved X-ray crystallography have emerged (Kempf and Loria 2003; Weiss 1999; Moffat 2003; Schotte et al. 2003), yielding information about the inherent conformational flexibility of proteins. Despite this enormous variety, experimental techniques having spatio-temporal resolution in the

M.B. Kubitzki*, B.L. de Groot, and D. Seeliger
Computational Biomolecular Dynamics Group, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077, Goettingen, Germany
*Corresponding author: e-mail: mkubitz@gwdg.de

nano- to microsecond as well as the nanometre regime are not routinely available, and thus information on the conformational space accessible to proteins *in vivo* often remains obscure. In particular, details on the pathways between different known conformations, frequently essential for protein function, are usually unknown. Here, computer simulation techniques provide an attractive possibility to obtain dynamic information on proteins at atomic resolution in the nanosecond to microsecond time range. Of all ways to simulate protein motions (Adcock and McCammon 2006), molecular dynamics (MD) techniques are among the most popular.

Since the first report of MD simulations of a protein some 30 years ago (McCammon et al. 1977), MD has become an established tool in the study of bio-molecules. Like all computational branches of science, the MD field benefits from the ever increasing improvements in computational power. This progression also allowed for advancements in simulation methodology that have led to a large number of algorithms for such diverse problems as cellular transport, signal trans-duction, allostery, cellular recognition, ligand-docking, the simulation of atomic force microscopy and enzymatic catalysis.

9.1.1 Principles and Approximations

Despite substantial algorithmic advances, the basic theory behind MD simulations is fairly simple. For biomolecular systems having N particles, the numerical solu-tion of the time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \Psi(r,t) = H\Psi(r,t)$$

for the N-particle wave function $\psi(r,t)$ of the system is prohibitive. Several approxi-mations are therefore required to allow the simulation of solvated biomolecules at timescales on the order of nanoseconds. The first of these relates to positions of nuclei and electrons: due to the much lower mass and consequently much higher velocity of the electrons compared to the nuclei, electrons can often be assumed to instantaneously follow the motion of the nuclei. Thus, within the Born-Oppenheimer approximation, only the nuclear motion has to be considered, with the electronic degrees of freedom influencing the dynamics of the nuclei in the form of a potential energy surface V(r).

The second essential approximation used in MD is to describe nuclear motion classically by Newton's equations of motion

$$m_i \frac{d^2 r_i}{dt^2} = -\nabla_i V(r_1,\ldots,r_N),$$

where m_i and r_i are the mass and the position of the i-th nucleus. With the nuclear motion described classically, the Schrödinger equation for the electronic degrees of freedom has to be solved to obtain the potential energy V(r). However, due to

the large number of electrons involved, a further simplification is necessary. A semi-empirical force field is introduced which approximates V(r) by a large number of functionally simple energy terms for bonded and non-bonded interactions. In its general form

$$V(r) = V_{bonds} + V_{angles} + V_{dihedrals} + V_{improper} + V_{Coul} + V_{LJ}$$

$$= \sum_{bonds} \frac{1}{2} k_i^l (l_i - l_{i,0})^2 + \sum_{angles} \frac{1}{2} k_i^\theta (\theta_i - \theta_{i,0})^2$$

$$+ \sum_{dihedrals} \frac{V_n}{2} (1 + \cos(n\varphi - \delta)) + \sum_{improper} \frac{1}{2} k_\xi (\xi_{ijkl} - \xi_0)^2$$

$$+ \sum_{i,j;i \neq j} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} + \sum_{i,j;i \neq j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right].$$

The simple terms are often harmonic (e.g. $V_{bonds}, V_{angles}, V_{improper}$) or motivated by physical laws (e.g. Coulomb V_{Coul}, Lennard-Jones V_{LJ}). They are defined by their functional form and a small number of parameters, e.g. an atomic radius for van der Waals interactions. All parameters are determined using either *ab initio* quantum chemical calculations or comparisons of structural or thermodynamical data with suitable averages of small molecule MD ensembles. Between different force fields (Brooks et al. 1983; Weiner et al. 1986; Van Gunsteren and Berendsen 1987; Jorgensen et al. 1996) the number of energy terms, their functional form and their individual parameters can vary considerably.

Given the above description of proteins as point masses (positions r_i, velocities v_i) moving in a classical potential under external forces F_i, a standard MD simulation integrates Newton's equations of motion in discrete timesteps Δt on the femtosecond timescale by some numerical scheme, e.g. the leap-frog algorithm (Hockney et al. 1973):

$$v_i(t + \frac{\Delta t}{2}) = v_i(t - \frac{\Delta t}{2}) + \frac{F_i(t)}{m_i} \Delta t$$

$$r_i(t + \Delta t) = r_i(t) + v_i(t + \frac{\Delta t}{2})\Delta t.$$

Besides interactions with membranes and other macromolecules, water is the principal natural environment for proteins. For a simulation of a model system that matches the *in vivo* system as close as possible, water molecules and ions in physiological concentration are added to the system in order to solvate the protein. Having a simulation box filled with solvent and solute, artefacts due to the boundaries of the system may arise, such as evaporation, high pressure due to surface tension and preferred orientations of solvent molecules on the surface. To avoid such artefacts,

periodic boundary conditions are often applied. In this way, the simulation system does not have any surface. This, however, may lead to new artefacts if the molecules artificially interact with their periodic images due to e.g. long-range electrostatic interactions. These periodicity artefacts are minimized by increasing the size of the simulation box. Different choices of unit cells, e.g., cubic, dodecahedral or truncated octahedral allow an optimal fit to the shape of the protein, and, therefore, permit a suitable compromise between the number of solvent molecules while simultaneously keeping the crucial protein-protein distance high.

As the solvent environment strongly affects the structure and dynamics of proteins, water must be described accurately. Besides the introduction of implicit solvent models, where water molecules are represented as a continuous medium instead of individual "explicit" solvent molecules (Still et al. 1990; Gosh et al. 1998; Jean-Charles et al. 1991; Luo et al. 2002), a variety of explicit solvent models are used these days (e.g. Jorgensen et al. 1983). These models differ in the number of particles used to represent a water molecule and the assigned static partial charges, reflecting the polarity and, effectively, in most force fields, polarization. Because these charges are kept constant during the simulation, explicit polarization effects are thereby excluded. Nowadays, several polarizable water models (and force fields) exist, see Warshel et al. (2007) for a recent review.

In solving Newton's equations of motion, the total energy of the system is conserved, resulting in a microcanonical NVE ensemble having constant particle number N, volume V and energy E. However, real biological subsystems of the size studied in simulations constantly exchange energy with their surrounding. Furthermore, a constant pressure P of usually 1 bar is present. To account for these features, algorithms are introduced which couple the system to a temperature and pressure bath (Anderson 1980; Nose 1984; Berendsen et al. 1984), leading to a canonical NPT ensemble.

9.1.2 Applications

Molecular Dynamics simulations have become a standard technique in protein science and are routinely applied to a wide range of problems. Conformational dynamics of proteins, however, is still a demanding task for MD simulations since functional conformational transitions often occur at timescales of microseconds to seconds which are not routinely accessible with current algorithms and computer power.

9.1.2.1 Nuclear Transport Receptors

Despite their computational demands, MD simulations have been successfully applied to study functional modes of proteins. As an illustration, we will discuss in some detail a recent study (Zachariae and Grubmüller 2006) that revealed a strikingly fast conformational transition of the exportin CAS (Cse1p in yeast) from the

open to the closed state. CAS/Cse1p is a nuclear transport receptor consisting of 960 amino acids that binds importin-α and RanGTP in the nucleus. The heterotrimeric complex (Fig. 9.1) can cross nuclear pores and dissociates by catalyzed GTP hydrolysis in the cytoplasm and, thus, represents an important part of the nucleocytoplasmic transport cycle in cells.

For the function of the importin-α/CAS system it is essential that, after dissociation of the complex in the cytoplasm, CAS/Cse1p undergoes a large conformational change that prevents reassociation of the complex. X-ray structures of Cse1p show that the cargo bound conformation adopts a superhelical structure with curls around the bound RanGTP (Fig. 9.2 left), whereas the cytoplasmic form exhibits a closed ring conformation that leads to occlusion of the RanGTP binding site (Fig. 9.2 right). In order to understand the mechanism of this conformational switch, Zachariae and Grubmüller carried out MD simulations of Cse1p starting from the cargo bound conformation. They found that, mainly driven by electrostatic interactions, the structure of Cse1p spontaneously collapses and adopts a conformation close to the experimentally determined cytoplasmic form within a relatively short timescale of 10 ns. Simulations of mutants with different electrostatic surface potentials did not reveal a significant conformational change but remained in an open conformation which is in good agreement with experimental findings (Cook et al. 2005). This example shows that functionally relevant conformational changes that occur on short time scales can be studied by MD simulations. However, in this particular case the simulation has – due to the removal of importin-α and RanGTP – not been started from an equilibrium conformation
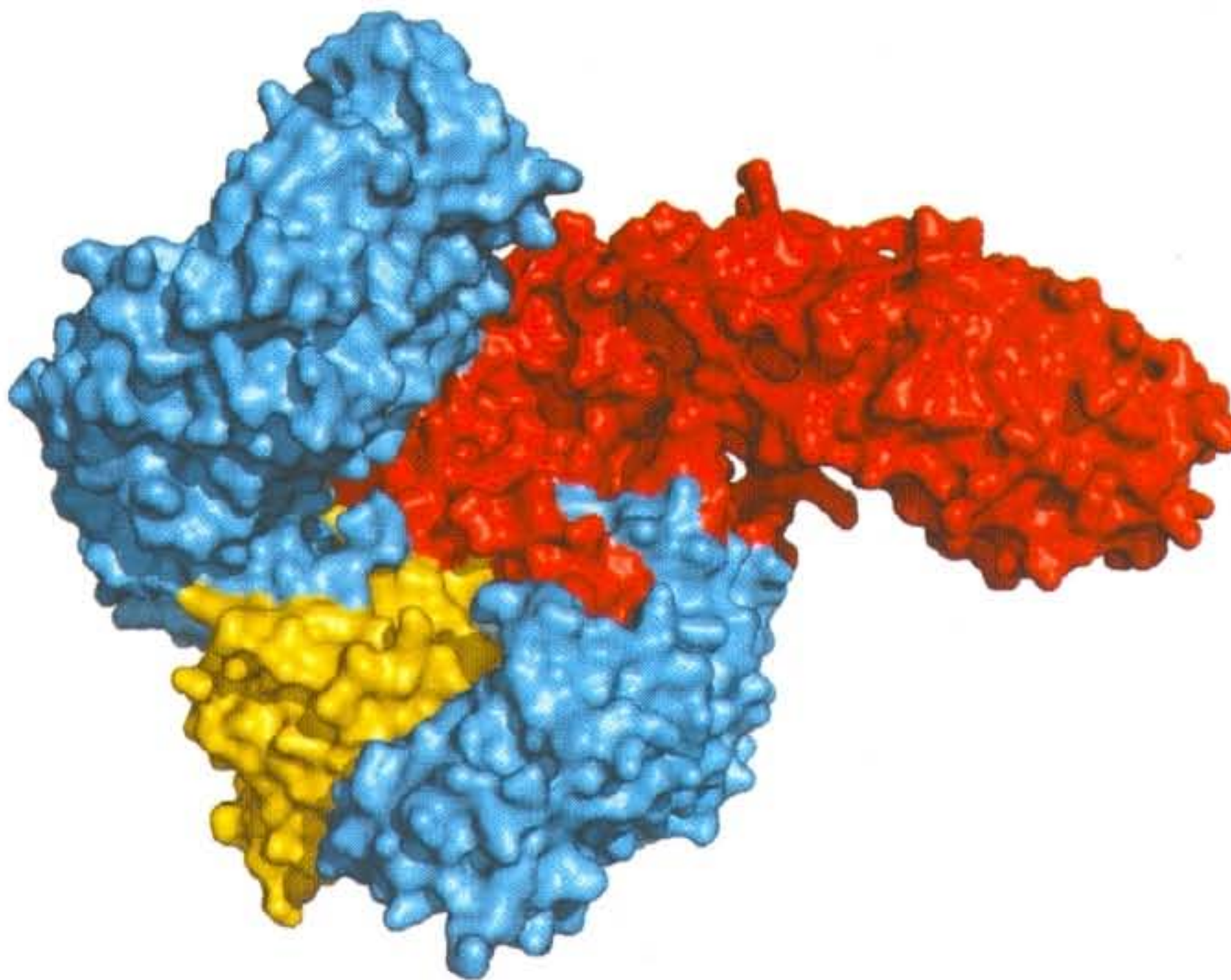

Fig. 9.1 Heterotrimeric complex of Cse1p (blue), RanGTP (yellow) and importin-α (red). Cse1p adopts a superhelical structure and binds RanGTP and importin–α. The complex can cross nuclear pores and dissociates by catalyzed GTP hydrolysis in the cytoplasm
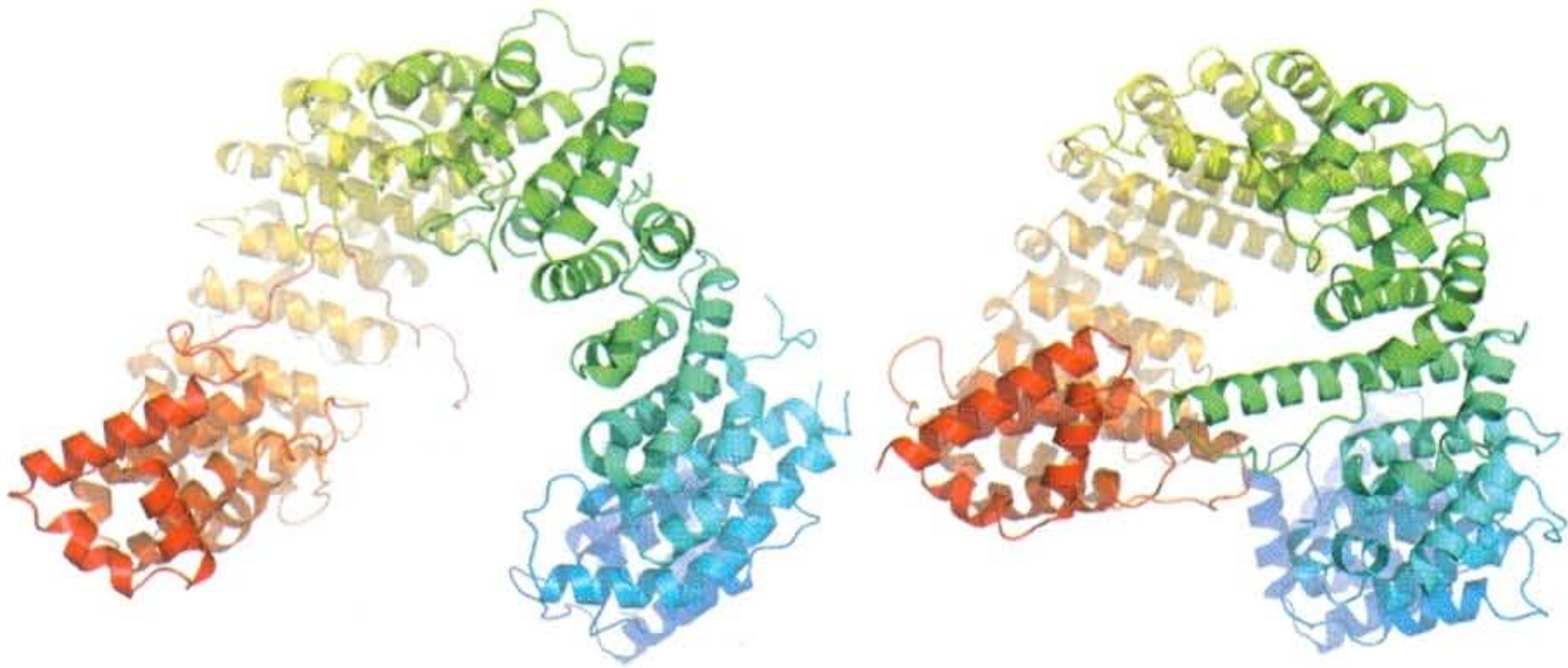
Fig. 9.2 Nucleoplasmic (left) and cytoplasmic (right) form of Cse1p. In the nucleoplasmic form, Cse1p is bound to RanGTP and importin-α (both not shown) and adopts a superhelical structure. After dissociation in the cytoplasm, Cse1p undergoes a large conformational change and forms a ring conformation that occludes the RanGTP binding site and prevents reassociation of the complex. The structures are coloured in a spectrum from blue (N-terminus) to red (C-terminus)

and thus, presumably, no significant energy barrier had to be overcome to reach the closed conformation. When simulations are started from a free energy minimum, which is usually the case, the accessible time scales are often too short to overcome higher energy barriers and, thus, to observe functionally relevant conformational transitions. This is known as the "sampling problem" and is a general problem for MD simulations.

9.1.2.2 Lysozyme

MD simulations of bacteriophage T4-lysozyme (T4L), an enzyme which is six times smaller than Cse1p, impressively illustrate this sampling problem for relatively long MD trajectories. T4L has been extensively studied with X-ray crystallography (Faber and Matthews 1990; Kuroki et al. 1993) and, since it has been crystallized in many different conformations, represents one of the rare cases where information about functionally relevant modes can be directly obtained at atomic resolution from experimental data (Zhang et al. 1995; de Groot et al. 1998). The domain character of this enzyme is very pronounced (Matthews and Remington 1974) and from the differences between crystallographic structures of various mutants of T4L it has been suggested that a hinge-bending mode of T4L (Fig. 9.3) is an intrinsic property of the molecule (Dixon et al. 1992). Moreover, the domain fluctuations are predicted to be essential for the function of the enzyme, allowing the substrate to enter and the products to leave the active site in the open configuration, with the closed state presumably required for catalysis.
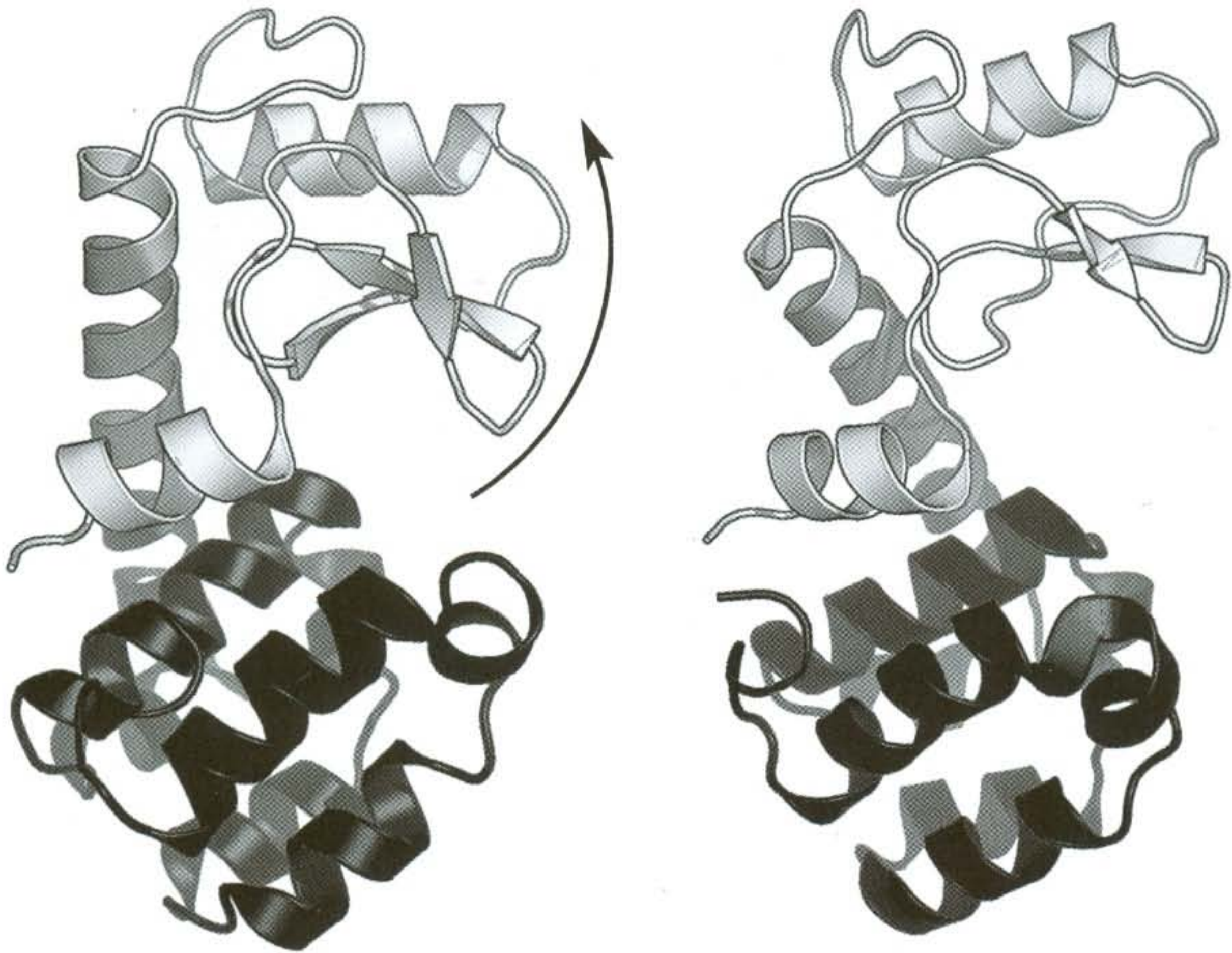
Fig. 9.3 Hinge-bending motion in bacteriophage T4-lysozyme. Domain fluctuations (domains are coloured differently) are essential for enzyme function, allowing the substrate to enter and the products to leave the active site

The wealth of experimental data also provides the opportunity to assess the reliability and sampling performance of simulation methods. Two MD simulations have been carried out using a closed (simulation 1) and an open conformation (simulation 2) as starting points, respectively. In order to assess the sampling efficiency a principal components analysis (PCA, see Section 9.2 below) has been carried out on the ensemble of experimentally determined structures and the X-ray ensemble and the two MD trajectories have been projected onto the first two eigenvectors. The first eigenvector represents the hinge-bending motion, whereas the second eigenvector represents a twist of the two domains of T4L. The projections are shown in Fig. 9.4. The X-ray ensemble is represented by dots, each dot representing a single conformation. Movement along the first eigenvector (x-axis) describes a collective motion from the closed to the open state. It can be seen that neither of the individual the MD trajectories, represented by lines, fully samples the entire conformational space covered by the X-ray ensemble, although the simulation times (184 ns for simulation 1 and 117 ns for simulation 2) are one order of magnitude larger than in the previously discussed Cse1p simulation. From the phase space density one can assume that an energy barrier exists between the closed and the open state and neither simulation achieves a full transition, from the closed to the open state, or *vice versa*.
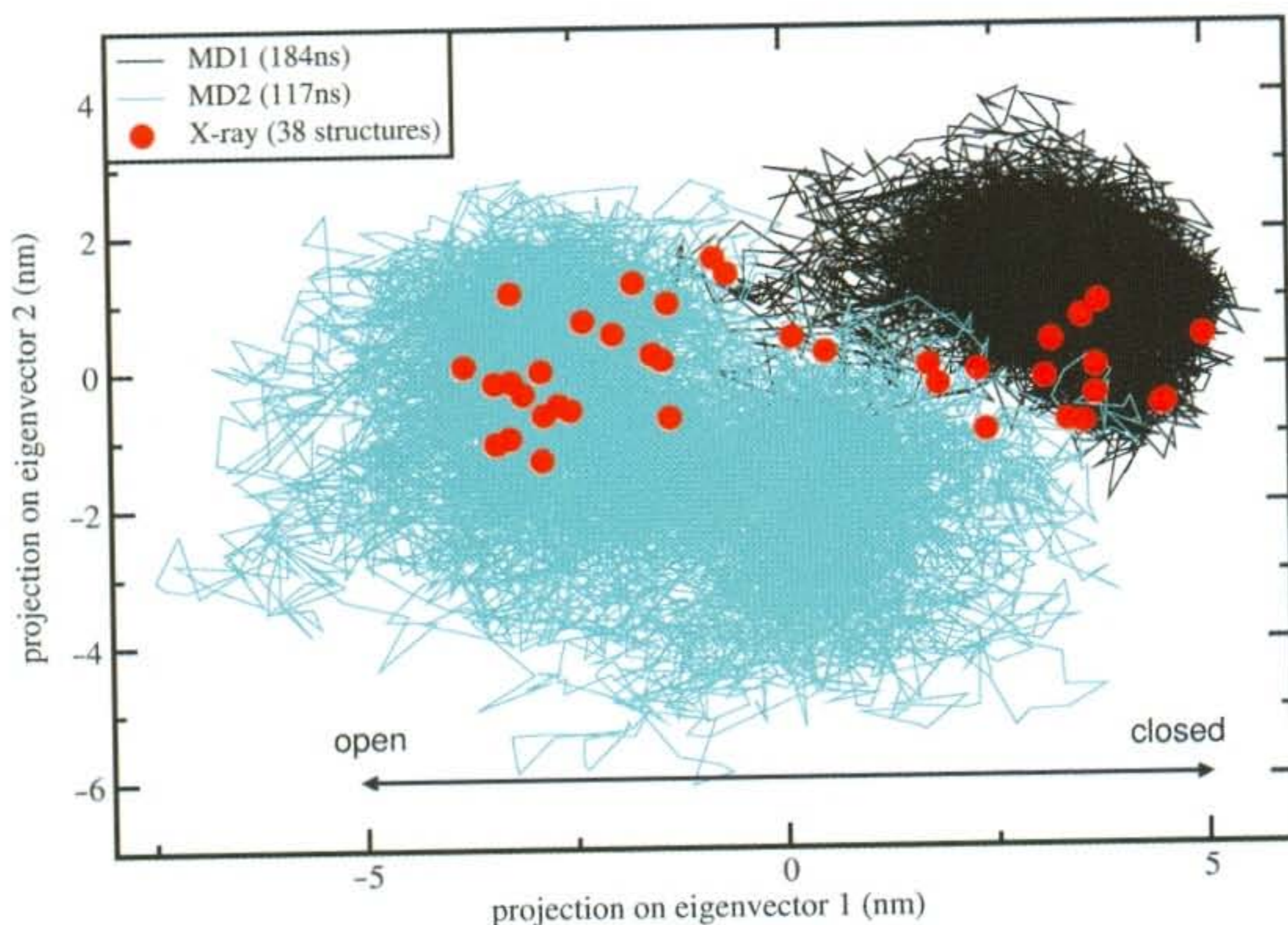
Fig. 9.4 Principal components analysis of bacteriophage T4-lysozyme. The X-ray ensemble is represented by dots, MD trajectories by lines. A movement along the first eigenvector (x-axis) represents a collective motion from the open to the closed state. Neither simulation 1 – started from a closed conformation – nor simulation 2 – started from an open conformation – show a full transition due to an energy barrier that separates the conformational states

9.1.2.3 Aquaporins

Aquaporins present a prime example of how MD simulations have contributed to the understanding of protein function both in terms of dynamics and energetics. Aquaporins facilitate efficient and selective permeation of water across biological membranes. Related aquaglyceroporins in addition also permeate small neutral solutes like glycerol. Available high-resolution structures provided invaluable insights in the molecular mechanisms acting in aquaporins (Fu et al. 2000; Murata et al. 2000; de Groot et al. 2001; Sui et al. 2001). However, mostly static information is available from such structures and we can therefore not directly observe aquaporins "at work". So far, there is no experimental method that offers sufficient spatial and time resolution to monitor permeation through aquaporins on a molecular level. MD simulations therefore complement experiments by providing the progression of the biomolecular system at atomic resolution. As permeation is known to take place on the nanosecond timescale, spontaneous permeation can be expected to take place in multi-nanosecond simulations, allowing a direct observation of the functional dynamics. Hence, such simulations have been termed "real-time simulations" (de Groot and Grubmüller 2001).

Indeed, spontaneous permeation events were observed in MD simulations of aquaporin-1 and the aquaglyceroporin GlpF. These simulations identified that the

efficiency of water permeation is accomplished by providing a hydrogen bond complementarity inside the channel comparable to bulk water, thereby establishing a low permeation barrier (de Groot and Grubmüller 2001; Tajkhorshid et al. 2002). The simulations furthermore identified that the selectivity in these channels is accomplished by a two-stage filter. The first stage of the filter is located in the central part of the channel at the conserved asparagine/proline/alanine (NPA) region; the second stage is located on the extracellular face of the channel in the aromatic/arginine (ar/R) constriction region (Fig. 9.5). As water permeation takes place on the nanosecond timescale, permeation coefficients can be directly computed from the simulations, and compared to experiment. Quantitative agreement was found between permeation coefficients from experiment and simulation, thereby validating the simulations.

A long standing question in aquaporin research has been the mechanism by which protons are excluded from the aqueous pores. The MD simulations addressing water permeation revealed a pronounced water dipole orientation pattern across the channel, with the NPA region as its symmetry centre (de Groot and
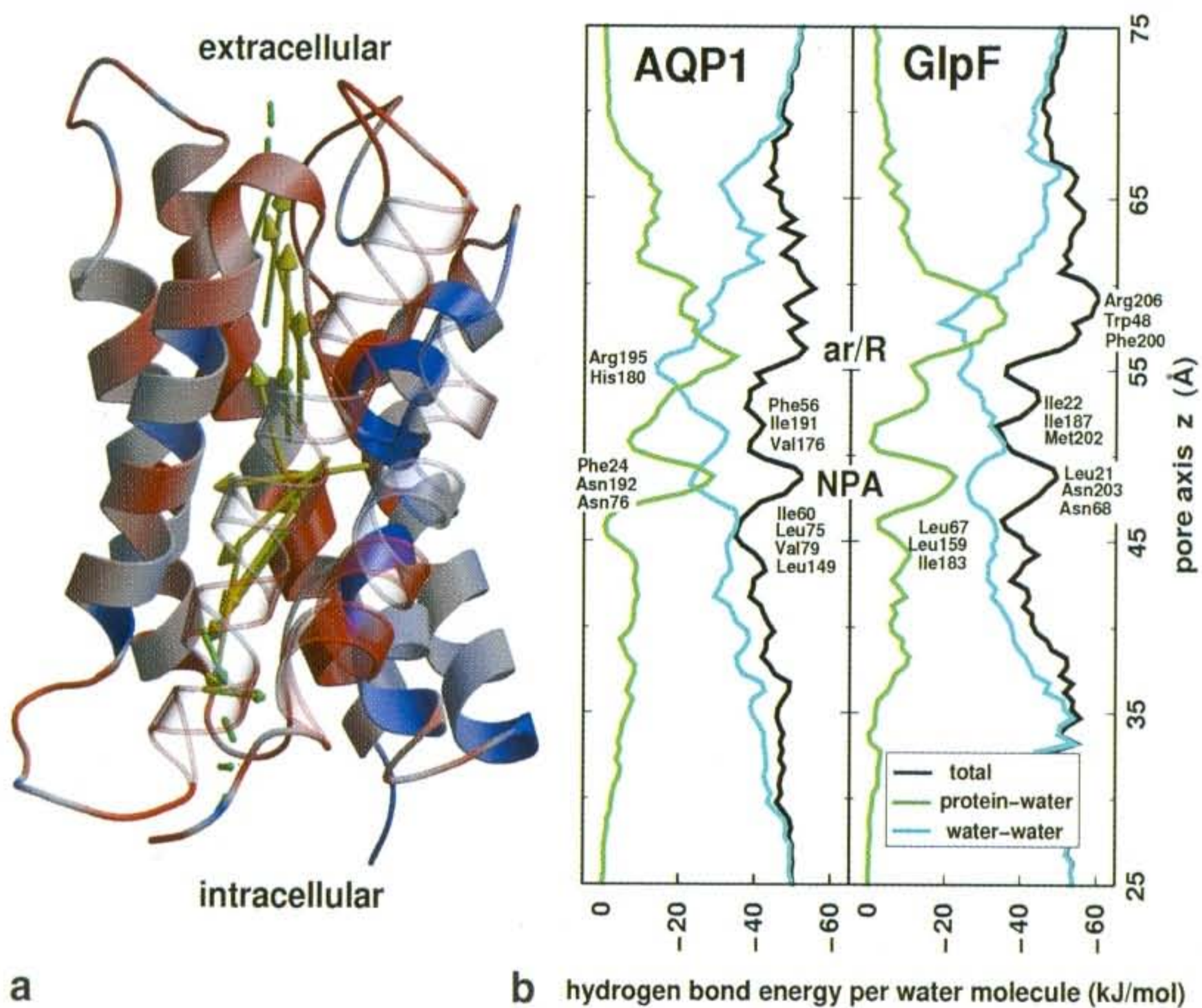

Fig. 9.5 (a) Water molecules are strongly aligned inside the aquaporin-1 channel, with their dipoles pointing away from the central NPA region (de Groot and Grubmüller 2001). The water dipoles (yellow arrows) rotate by approximately 180 degrees while permeating though the AQP1 pore. The red and blue colours indicate local electrostatic potential, negative and positive, respectively. (b) Hydrogen bond energies per water molecule (solid black lines) in AQP1 (left) and GlpF (right). Protein-water hydrogen bonds (green) compensate for the loss of water-water hydrogen bonds (cyan). The main protein-water interaction sites are the ar/R region and the NPA site

Grubmüller 2001). In the simulations, the water molecules were found to rotate by 180 degrees on their path through the pore (Fig. 9.5a). In a series of simulations addressing the mechanism of proton exclusion it was found that the pronounced water orientation is due to an electric field in the channel centred at the NPA region (de Groot et al. 2003; Chakrabarti et al. 2004; Ilan et al. 2004). Electrostatic effects therefore form the structural basis of proton exclusion. A debate continues about the origin of the electrostatic barrier, where both direct electrostatic effects caused by helix dipoles has been suggested (de Groot et al. 2003; Chakrabarti et al. 2004), as well as a specific desolvation effects (Burykin and Warshel 2003). The most recent results suggest that both effects contribute approximately equally (Chen et al. 2006).

Recently, MD simulations allowed for the elucidation of the mechanism of selectivity of neutral solutes in aquaporins and aquaglyceroporins. Aquaporins were found to be permeated solely by small polar molecules like water, and to some extent also ammonia, whereas aquaglyceroporins are also permeated by apolar molecules like CO_2 and larger molecules like glycerol, but not urea (Hub and de Groot 2008). For aquaporins, an inverse relation was observed between permeability and solute hydrophobicity – solutes competing with permeating water molecules for hydrogen bonds with the channel determine the permeation barrier. A combination of size exclusion and hydrophobicity therefore underlies the selectivity in aquaporins and aquaglyceroporins.

9.1.3 Limitations – Enhanced Sampling Algorithms

Although molecular dynamics simulations have become an integral part of structural biology and provided numerous invaluable insights into biological processes at the atomic level, limitations occur due to both methodological restrictions and limited computer power. Methodological limitations arise from the classical description of atoms and the approximation of interactions by simple energy terms instead of the Schrödinger equation. This means that chemical reactions (bond breaking and formation) can not be described. Also polarization effects and proton tunnelling lie out of the scope of classical MD simulations.

The second class of limitations arises from the computational demands of MD simulations. Although bonds are usually treated as constraints thereby eliminating the highest frequency motions, the timestep length in MD simulations usually cannot be chosen longer than 2 fs. Hence, a nanosecond simulation requires 500,000 force calculations and integration steps. Given current algorithm techniques and computer power, timescales of 100 ns are accessible within 3–4 weeks for a solvated 200 amino acid protein.

However, biologically relevant protein motions like large conformational transitions, folding and unfolding usually take place on the micro- to (milli)second timescale. Thus it becomes evident that, despite ever increasing computer power, which roughly grows by a factor of 100 per decade, MD simulations will not solve the "sampling problem" anytime soon by just waiting for faster computers.

Therefore, alternative methods – partly based on MD – have been developed to specifically address the problem of conformational sampling and to predict functionally relevant protein motions.

Reducing the number of particles is one approach. Since proteins are usually studied in solution most of the simulation system consists of water molecules. The development of implicit solvent models is therefore a promising means to reduce computational demands (Still et al. 1990; Gosh et al. 1998; Jean-Charles et al. 1991; Luo et al. 2002). Another possibility to reduce the number of particles is the use of so-called coarse-grained models (Bond et al. 2007). In these models, atoms are grouped together, for instance typically four water molecules are treated as one pseudo-particle (bead). These groupings have two effects. First, the number of particles is reduced and, second, the timestep, depending on the fastest motions in the system, can be increased. However, coarse-graining is not restricted to water molecules. Representations of several atoms up to complete amino acids by a single bead are nowadays used. This allows for a drastic reduction of computational demands, thereby enabling the simulation of large macromolecular aggregates on micro- to millisecond timescales. This gain in efficiency, however, comes with an inherent reduction of accuracy compared to all-atom descriptions of proteins, restricting current models to semi-quantitative statements. Essential for the success of coarse-grained simulations are the parametrizations of force fields that are both accurate and transferable, i.e. force fields capable of describing the general dynamics of systems having different compositions and configurations. As the graining becomes coarser, this process becomes increasingly difficult, since more specific interactions must effectively be included in fewer parameters and functional forms. This has led to a variety of models for proteins, lipids and water, representing different compromises between accuracy and transferability (see e.g. Marrink et al. 2004).

Other MD based enhanced sampling methods, which retain the atomistic description, include replica exchange molecular dynamics (REMD) and essential dynamics (ED) which are discussed in subsequent sections. Moreover, a number of non-MD based methods are discussed that aim towards the prediction of functional modes of proteins.

9.1.3.1 Replica Exchange

The aim of most computer simulations of biomolecular systems is to calculate macroscopic behaviour from microscopic interactions. Following equilibrium statistical mechanics, any observable that can be connected to macroscopic experiments is defined as an ensemble average over all possible realizations of the system. However, given current computer hardware, a fully converged sampling of all possible conformational states with the respective Boltzmann weight is only attainable for simple systems comprising a small number of amino acids (see e.g. Kubitzki and de Groot 2007). For proteins, consisting of hundreds to thousands of amino acids, conventional MD simulations often do not converge and reliable estimates of experimental quantities can not be calculated.

This inefficiency in sampling is a result of the ruggedness of the systems' free energy landscape, a concept put forward by Frauenfelder (Frauenfelder et al. 1991; Frauenfelder and Leeson 1998). The global shape is supposed to be funnel-like, with the native state populating the global free energy minimum (Anfinsen 1973). Looking in more detail, the complex high-dimensional free energy landscape is characterized by a multitude of almost iso-energetic minima, separated from each other by energy barriers of various heights. Each of these minima corresponds to one particular conformational substate, with neighboring minima corresponding to simi- lar conformations. Within this picture, structural transitions are barrier crossings, with the transition rate depending on the height of the barrier. For MD simulations at room temperature, only those barriers are easily overcome that are smaller than or comparable to the thermal energy $k_B T$ and the observed structural changes are small, e.g. side chain rearrangements. Therefore the system will spend most of its time in locally stable states (*kinetic trapping*) instead of exploring different conformational states. This wider exploration is of greater interest, due to its connection to biologi- cal function, but requires that the system be able to overcome large energy barriers. Unfortunately, since MD simulations are mostly restricted to the nanosecond times- cale, functionally relevant conformational transitions are rarely observed.

A plethora of enhanced sampling methods have been developed to tackle this multi-minima problem (see e.g. Van Gunsteren and Berendsen 1990; Tai 2004; Adcock and McCammon 2006 and references therein). Among them, generalized ensemble algorithms have been widely used in recent years (for a review, see e.g. Mitsutake et al. 2001; Iba 2001). Generalized ensemble algorithms sample an arti- ficial ensemble that is either constructed from combinations or alterations of the original ensemble. Algorithms of the second category (e.g. Berg and Neuhaus 1991) basically modify the original bell-shaped potential energy distribution p(V) of the system by introducing a so-called multicanonical weight factor w(V), such that the resulting distribution is uniform, p(V)w(V) = const. This flat distribution can then be sampled extensively by MD or Monte-Carlo techniques because poten- tial energy barriers are no longer present. Due to the modifications introduced, estimates for canonical ensemble averages of physical quantities need to be obtained by reweighting techniques (Kumar et al. 1992; Chodera et al. 2007). The main problem with these algorithms, however, is the non-trivial determination of the different multicanonical weight factors by an iterative process involving short trial simulations. For complex systems this procedure can be very tedious and attempts have been made to accelerate convergence of the iterative process (Berg and Celik 1992; Kumar et al. 1996; Smith and Bruce 1996; Hansmann 1997; Bartels and Karplus 1998).

The replica exchange (REX) algorithm, developed as an extension of simulated tempering (Marinari and Parisi 1992), removes the problem of finding correct weight factors. It belongs to the first category of algorithms where a generalized ensemble, built from several instances of the original ensemble, is sampled. Due to its simplicity and ease of implementation, it has been widely used in recent years. Most often, the standard temperature formulation of REX is employed (Sugita and Okamoto 1999), with the general Hamiltonian REX framework gaining increasing

attention (Fukunishi et al. 2002; Liu et al. 2005; Sugita et al. 2000; Affentranger et al. 2006; Christen and van Gunsteren 2006; Lyman and Zuckerman 2006).

In standard temperature REX MD (Sugita and Okamoto 1999), a generalized ensemble is constructed from M + 1 non-interacting copies, or "replicas", of the system at a range of temperatures $\{T_0,\ldots,T_M\}$ $(T_m \leq T_{m+1}; m = 0,\ldots,M)$, e.g. by distributing the simulation over M + 1 nodes of a parallel computer (Fig. 9.6 left). A state of this generalized ensemble is characterized by $S = \{\ldots,s_m,\ldots\}$, where s_m represents the state of replica m having temperature T_m. The algorithm now consists of two consecutive steps: (a), independent constant-temperature simulations of each replica, and (b), exchange of two replicas $S = \{\ldots,s_m,\ldots,s_n,\ldots\} \rightarrow S' = \{\ldots,s_n',\ldots,s_m',\ldots\}$ according to a Metropolis-like criterion. The exchange acceptance probability is thereby given by

$$P(S \rightarrow S') = \min \{1, \exp \{(\beta_m - \beta_n)[V_m - V_n]\}\} \tag{9.1}$$

with V_m being the potential energy and $\beta_m^{-1} = k_B T_m$. Iterating steps a and b, the trajectories of the generalized ensemble perform a random walk in temperature space, which in turn induces a random walk in energy space. This facilitates an efficient and statistically correct conformational sampling of the energy landscape of the system, even in the presence of multiple local minima.

The choice of temperatures is crucial for an optimal performance of the algorithm. Replica temperatures have to be chosen such that (a) the lowest temperature is small enough to sufficiently sample low-energy states, (b) the highest temperature is large enough to overcome energy barriers of the system of interest, and (c) the acceptance probability $P(S \rightarrow S')$ is sufficiently high, requiring adequate overlap of potential energy distributions for neighboring replicas. For larger systems simulated with explicit solvent the latter condition presents the main bottleneck. A simple estimate (Cheng et al. 2005; Fukunishi et al. 2002) shows that the potential energy difference $\Delta V \sim N_{df} \Delta T$ is dominated by the contribution from the solvent degrees of freedom N_{df}^{sol}, constituting the largest frac-
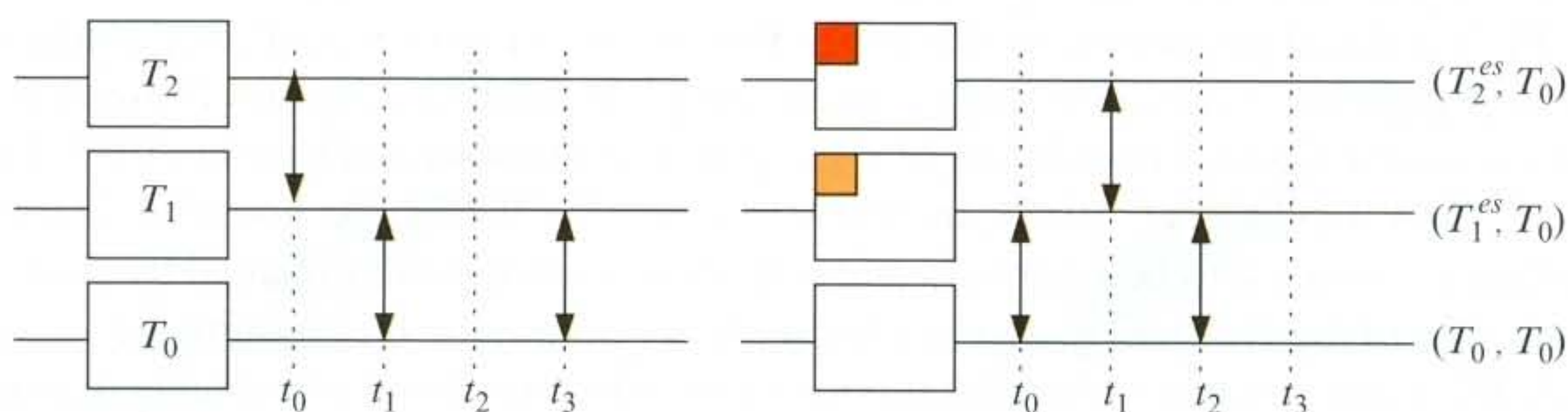

Fig. 9.6 Schematic comparison of standard temperature REX (left panel) and the TEE-REX algorithm (right panel) for a three-replica simulation. Temperatures are sorted in increasing order, $T_{i+1} > T_i$. Exchanges ($\leftrightarrow$) are attempted (…) with frequency ν_{ex}. Unlike REX, only an essential subspace {es} (red boxes) containing a few collective modes is excited within each TEE-REX replica. Reference replica (T_0, T_0), containing an approximate Boltzmann ensemble, is used for analysis

tion of the total number of degrees of freedom N_{df} of the system. Obtaining a reasonable acceptance probability therefore relies on keeping the temperature gaps $\Delta T = T_{m+1} - T_m$ small (typically only a few Kelvin) which drastically increases computational demands for systems having more than a few thousand particles. Despite this severe limitation, REX methods have become an established tool for the study of peptide folding/unfolding (Zhou et al. 2001; Rao and Caflisch 2003; García and Onuchic 2003; Pitera and Swope 2003; Seibert et al. 2005), structure prediction (Fukunishi et al. 2002; Kokubo and Okamoto 2004), phase transitions (Berg and Neuhaus 1991) and free energy calculations (Sugita et al. 2000; Lou and Cukier 2006).

Going beyond conventional MD, another class of enhanced sampling algorithms is successfully applied to the task of elucidating protein function. These algorithms make use of the fact that fluctuations in proteins are generally correlated. Extracting such collective modes of motion and their application in new sampling algorithms will be the focus of the following two sections.

9.2 Principal Component Analysis

Principal component analysis (PCA) is a well-established technique to obtain a low-dimensional description of high-dimensional data. Its applications include data compression, image processing, data visualization, exploratory data analysis, pattern recognition and time series prediction (Duda et al. 2001). In the context of biomolecular simulations PCA has become an important tool in the extraction and classification of relevant information about large conformational changes from an ensemble of protein structures, generated either experimentally or theoretically (García 1992; Gō et al. 1983; Amadei et al. 1993). Besides PCA, a number of similar techniques are nowadays used, most notably normal mode analysis (NMA) (Brooks and Karplus 1983; Gō et al. 1983; Levitt et al. 1983), quasi-harmonic analysis (Karplus and Kushick 1981; Levy et al. 1984a, b; Teeter and Case 1990) and singular-value decomposition (Romo et al. 1995; Bahar et al. 1997).

PCA is based on the notion that by far the largest fraction of positional fluctuations in proteins occurs along only a small subset of collective degrees of freedom. This was first realized from NMA of a small protein (Brooks and Karplus 1983; Gō et al. 1983; Levitt et al. 1983). In NMA (see Section 9.4.1), the potential energy surface is assumed to be harmonic and collective variables are obtained by diagonalization of the Hessian[1] matrix in a local energy minimum. Quasi-harmonic analysis, PCA and singular-value decomposition of MD trajectories of proteins that do not assume harmonicity of the dynamics, have shown that indeed protein dynamics

---

[1] Second derivative $(\partial^2 V)/(\partial x_i \partial x_j)$ of the potential energy

is dominated by a limited number of collective coordinates, even though the major modes are frequently found to be largely anharmonic. These methods identify those collective degrees of freedom that best approximate the total amount of fluctuation. The subset of largest-amplitude variables form a set of generalized internal coordinates that can be used to effectively describe the dynamics of a protein. Often, a small subset of 5–10% of the total number of degrees of freedom yields a remarkably accurate approximation. As opposed to torsion angles as internal coordinates, these collective internal coordinates are not known beforehand but must be defined either using experimental structures or an ensemble of simulated structures. Once an approximation of the collective degrees of freedom has been obtained, this information can be used for the analysis of simulations as well as in simulation protocols designed to enhance conformational sampling (Grubmüller 1995; Zhang et al. 2003; He et al. 2003; Amadei et al. 1996).

In essence, a principal component analysis is a multi-dimensional linear least squares fit procedure in configuration space. The structure ensemble of a molecule, having N particles, can be represented in 3N-dimensional configuration space as a distribution of points with each configuration represented by a single point. For this cloud, always one axis can be defined along which the maximal fluctuation takes place. As illustrated for a two-dimensional example (Fig. 9.7), if such a line fits the data well, all data points can be approximated by only the projection onto that axis, allowing a reasonable approximation of the position even when neglecting the position in all directions orthogonal to it. If this axis is chosen as coordinate axis, the position of a point can be represented by a single coordinate. The procedure in the general 3N-dimensional case works similarly. Given the first axis that best describes the data, successive directions orthogonal to the previous set are chosen such as to fit the data second-best, third-best, and so on (the *principal components*). Together, these directions span a 3N-dimensional space. Mathematically, these directions are given by the eigenvectors μ_i of the covariance matrix of atomic fluctuations
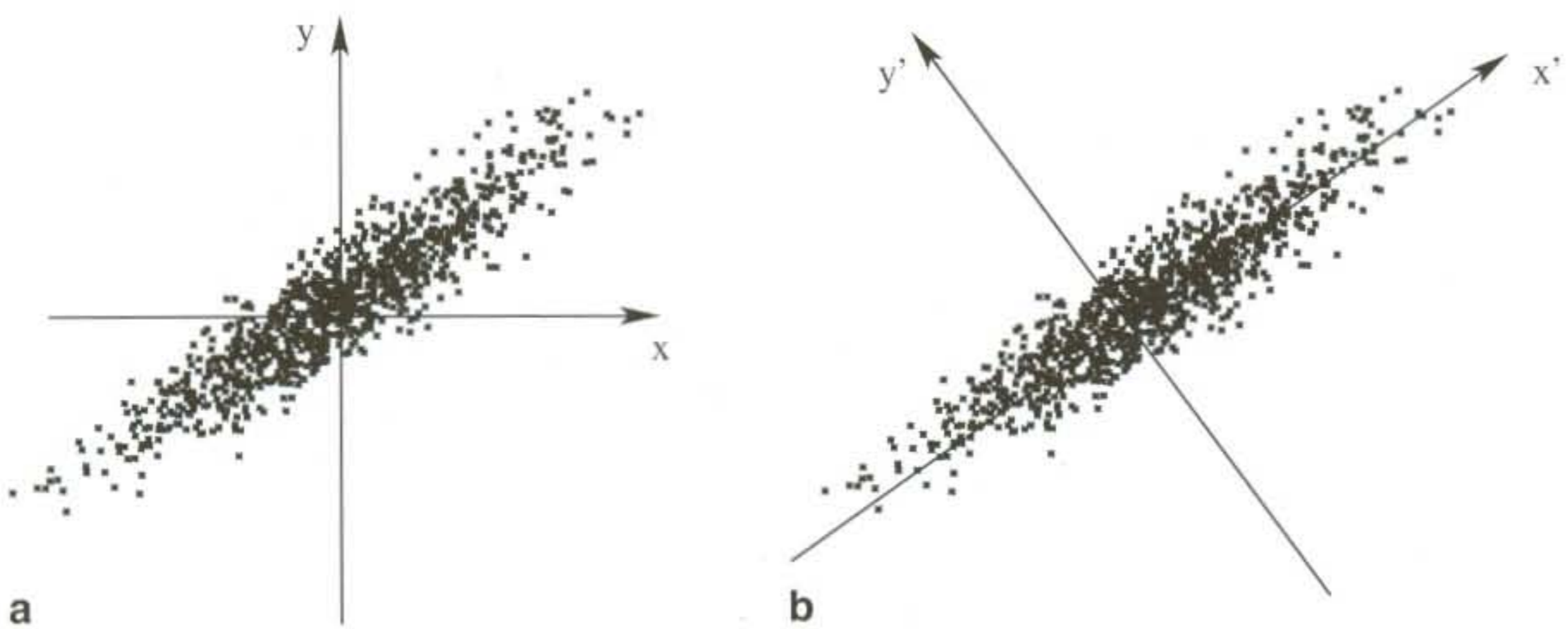

Fig. 9.7 Illustration of PCA in two dimensions. Two coordinates (x, y) are required to identify a point in the ensemble in panel (a), whereas one coordinate x′ approximately identifies a point in panel (b)

$$C = \left\langle \left(x(t) - \langle x \rangle \right) \left(x(t) - \langle x \rangle \right)^T \right\rangle,$$

with the angle brackets $\langle \cdot \rangle$ representing an ensemble average. The eigenvalues λ_i correspond to the mean square positional fluctuation along the respective eigenvector, and therefore contain the contribution of each principal component to the total fluctuation (Fig. 9.8).

Applications of such a multidimensional fit procedure on protein configurations from MD simulations of several proteins have proven that typically the first 10 to 20 principal components are responsible for 90% of the fluctuations of a protein (Kitao et al. 1991; García 1992; Amadei et al. 1993). These principal components correspond to collective coordinates, containing contributions from every atom of the protein. In a number of cases these principal modes were shown to be involved in the functional dynamics of the studied proteins (Amadei et al. 1993; Van Aalten et al. 1995a, b; de Groot et al. 1998). Hence, the subspace responsible for the majority of all fluctuations has been referred to as the *essential subspace* (Amadei et al. 1993).

The fact that a small subset of the total number of degrees of freedom (essential subspace) dominates the molecular dynamics of proteins originates from the presence of a large number of internal constraints and restrictions defined by the atomic interactions present in a biomolecule. These interactions range from strong covalent bonds to weak non-bonded interactions, whereas the restrictions are given by the dense packing of atoms in native-state structures.

Overall, protein dynamics at physiological temperatures has been described as diffusion among multiple minima (Kitao et al. 1998; Amadei et al. 1999; Kitao and Gō 1999). The dynamics on short timescales is dominated by fluctuations within a local minimum, corresponding to eigenvectors having low eigenvalues. On longer timescales large fluctuations are dominated by a largely anharmonic diffusion between multiple wells. These slow dynamical transitions are usually represented by the largest-amplitude modes of a PCA. In contrast to normal mode analysis, PCA of a MD simulation trajectory does not rest on the
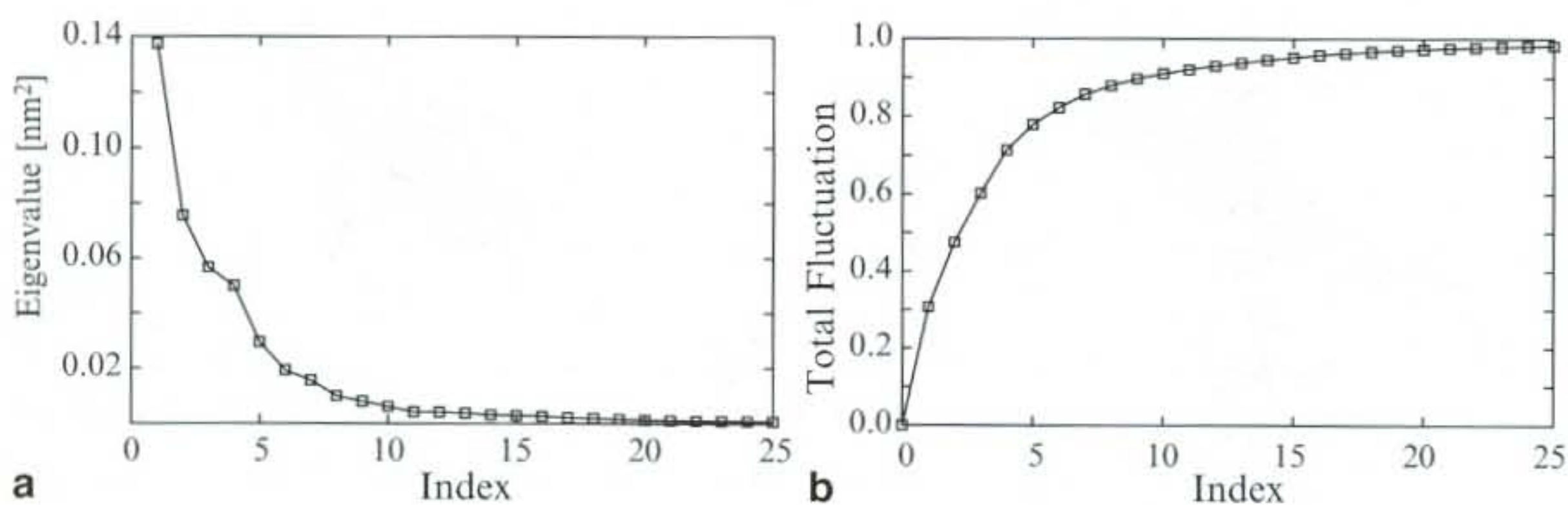
Fig. 9.8 Typical PCA eigenvalue spectrum (MD ensemble of guanylin backbone structures). The first five eigenvectors (panel a) cover 80% of all observed fluctuations (panel b)

assumption of a harmonic potential. In fact, PCA can be used to study the degree of anharmonicity in the molecular dynamics of the simulated system. For proteins, it was shown that at physiological temperatures, especially the major modes of collective fluctuation that are frequently functionally relevant, are dominated by anharmonic fluctuations (Amadei et al. 1993; Hayward et al. 1995).

9.3 Collective Coordinate Sampling Algorithms

Analyzing MD simulations in terms of collective coordinates (obtained e.g. by PCA or NMA) reveals that only a small subset of the total number of degrees of freedom dominates the molecular dynamics of biomolecules. As protein function could in many cases been linked to these essential subspace modes (e.g. Brooks and Karplus 1983; Gō et al. 1983; Levitt et al. 1983), the dynamics within this low-dimensional space was termed "essential dynamics" (ED). This not only aids the analysis and interpretation of MD trajectories but also opens the way to enhanced sampling algorithms that search the essential subspace in either a systematic or exploratory fashion (Grubmüller 1995; Amadei et al. 1996).

9.3.1 Essential Dynamics

The first attempts in this direction were aimed at a simulation scheme in which the equations of motion were solely integrated along a selection of primary principal modes, thereby drastically reducing the number of degrees of freedom (Amadei et al. 1993). However, these attempts proved problematic because of non-trivial couplings between high- and low-amplitude modes, even though after diagonalization the modes are linearly independent (orthogonal). Therefore, instead, a series of techniques has prevailed that take into account the full-dimensional simulation system and enhance the motion along a selection of principal modes. The most common of these techniques are conformational flooding (Grubmüller 1995) and ED sampling (Amadei et al. 1996; de Groot et al. 1996a, b). In conformational flooding, an additional potential energy term that stimulates the simulated system to explore new regions of phase space is introduced on a selection of principal modes, whereas in ED sampling a similar goal is achieved by geometrical constraints along a selection of principal modes. With these techniques a sampling efficiency enhancement of up to an order of magnitude can be achieved, provided that a reasonable approximation of the principal modes has been obtained from a conventional simulation. However, due to the applied structural or energetic bias on the system, the ensemble generated by ED sampling and

conformational flooding is not canonical, restricting analysis to structural questions.

9.3.2 TEE-REX

Enhanced sampling methods such as ED (Amadei et al. 1996) achieve their sampling power (Amadei et al. 1996; de Groot et al. 1996a, b) primarily from the fact that a small number of internal collective degrees of freedom dominate the configurational dynamics of proteins. Yet, systems simulated with such methods are always in a non-equilibrium state, rendering it difficult to extract thermodynamic, i.e. equilibrium properties of the system from such simulations. On the other hand, generalized ensemble algorithms such as REX not only enhance sampling but yield correct statistical ensembles necessary for the calculation of equilibrium properties which can be subjected to experimental verification. However, REX quickly becomes computationally prohibitive for systems of more than a few thousand particles, limiting its current applicability to smaller peptides (Pitera and Swope 2003; Cecchini et al. 2004; Nguyen et al. 2005; Liu et al. 2005; Seibert et al. 2005). The newly developed Temperature Enhanced Essential dynamics Replica EXchange (TEE-REX) algorithm (Kubitzki and de Groot 2007) combines the favourable properties of REX with those resulting from a *specific* excitation of functionally relevant modes, while at the same time avoiding the drawbacks of both approaches.

Figure 9.6 shows a schematic comparison of standard temperature REX (left) and the TEE-REX algorithm (right). TEE-REX builds upon the REX framework, i.e. a number of replicas of the system are simulated independently in parallel with periodic exchange attempts between neighbouring replicas. In contrast to REX, in each but the reference replica, only those degrees of freedom are thermally stimulated that contribute significantly to the total fluctuations of the system (essential subspace {es}). This way, several benefits are combined and drawbacks avoided. In contrast to standard REX, the specific excitation of collective coordinates promotes sampling along these often functionally relevant modes of motion, i.e. the advantages of ED are used. To counterbalance the disadvantages associated with such a specific excitation, i.e. the construction of biased ensembles, the scheme is embedded within the REX protocol. Thereby ensembles are obtained having approximate Boltzmann statistics and the enhanced sampling properties of REX are utilized. The exchange probability (equation 9.1) between two replicas crucially depends on the excited number of degree of freedom of the system. Since the stimulated number of degrees of freedom makes up only a minute fraction of the total number of degrees of freedom of the system, the bottleneck of low exchange probabilities in all-atom REX simulations is bypassed. For given exchange probabilities, large temperature differences ΔT can thus be used, such that only a few replicas are required.

Figure 9.9 shows a two-dimensional projection of the free energy landscape of dialanine, calculated with MD (panel A) and TEE-REX (panel B). The thermodynamic behaviour of a system is completely known once a thermodynamic potential
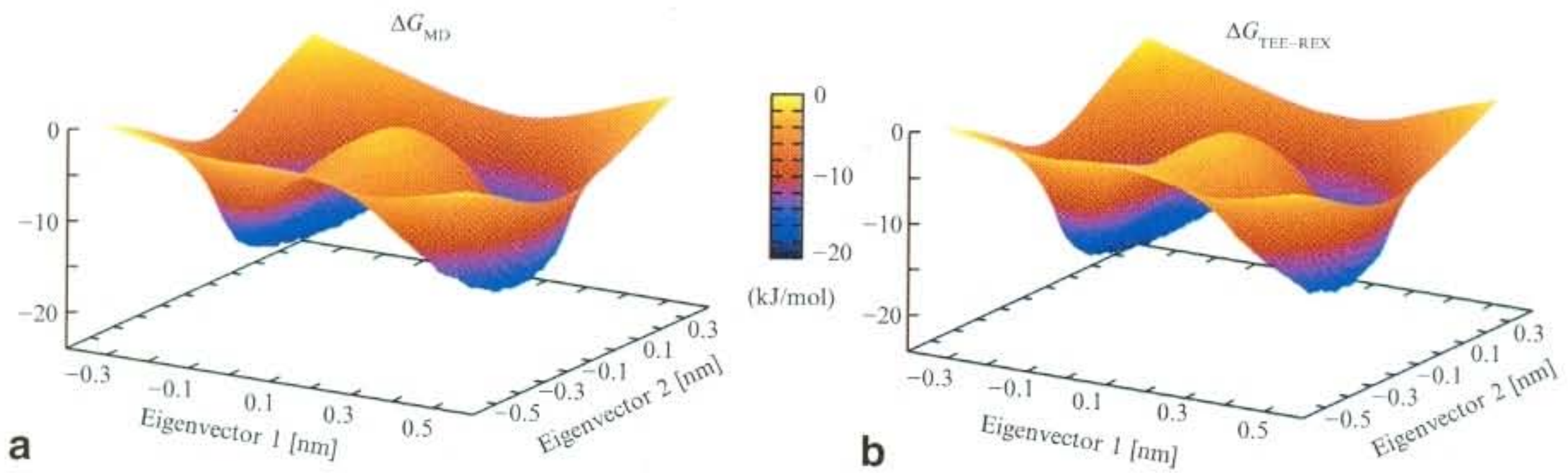
Fig. 9.9 Comparison of two-dimensional relative free energy surfaces (in units of kJ/mol) of dialanine generated by MD (panel a) and TEE-REX (panel b). Deviations $\Delta G_{TEE-REX} - G_{MD}$ are commensurate with the statistical errors of $\sim 0.1\,k_BT$

such as the relative Gibbs free energy ΔG is available. Comparing free energies thus enables us to decide to which degree ensembles created by different simulation methods coincide. In doing so, ensemble convergence is an absolute necessity. For the dialanine test case, this requirement is met. A detailed analysis of the shape of the free energy surfaces generated by MD and TEE-REX shows that the maximum absolute deviations of $1.5\,kJ/mol \cong 0.6\,k_BT$ from the ideal case $\Delta G_{TEE-REX} - G_{MD} = 0$, commensurate with the maximum statistical errors of $0.15\,k_BT$ found for each method. The small deviations found for the TEE-REX ensemble are presumably due to the exchange of non-equilibrium structures into the TEE-REX reference ensemble.

The sampling efficiency of the TEE-REX algorithm compared to MD was evaluated for guanylin, a small 13 amino-acid peptide hormone (Currie et al. 1992). Trajectories generated with both methods-using the same computational effort-were projected into (ϕ,ψ)-space as well as different two-dimensional subspaces spanned by PCA modes calculated from an MD ensemble of guanylin structures. From these projections, the time evolution of sampled configuration space volume was measured. Overall, the sampling performance of MD is quite limited compared to TEE-REX, the latter outperforming MD on average by a factor of 2.5, depending on the subspace used for projecting.

9.3.2.1 Applications: Finding Transition Pathways in Adenylate Kinase

Understanding the functional basis for many protein functions (Gerstein et al. 1994; Berg et al. 2002; Karplus and Gao 2004; Xu et al. 1997) requires detailed knowledge of transitions between functionally relevant conformations. Over the last years X-ray crystallography and NMR spectroscopy have provided mostly static pictures of different conformational states of proteins, leaving questions related to the underlying transition pathway unanswered. For atomistic MD simulations, elucidating the pathways and mechanisms of protein conformational dynamics poses a

challenge due to the long timescales involved. In this respect, *E. coli* adenylate kinase (ADK) is a prime example. ADK is a monomeric enzyme that plays a key role in energy maintenance within the cell, controlling cellular ATP levels by catalyzing the reaction $Mg^{2+}:ATP + AMP \leftrightarrow Mg^{2+}:ADP + ADP$. Structurally, the enzyme consists of three domains (Fig. 9.10): the large central "CORE" domain (light grey), an AMP binding domain referred to as "AMPbd" (black), and a lid-shaped ATP-binding domain termed "LID" (dark grey), which covers the phosphate groups at the active centre (Müller et al. 1996). In an unligated structure of ADK the LID and AMPbd adopt an open conformation, whereas they assume a closed conformation in a structure crystallized with the transition state inhibitor Ap_5A (Müller and Schulz 1992). Here, the ligands are contained in a highly specific environment required for catalysis. Recent ^{15}N nuclear magnetic resonance spin relaxation studies (Shapiro and Meirovitch 2006) have shown the existence of catalytic domain motions in the flexible AMPbd and LID domains on the nanosecond time scale, while the relaxation in the CORE domain is on the picosecond time scale (Tugarinov et al. 2002; Shapiro et al. 2002). For ADK, several computational studies have addressed its conformational flexibility (Temiz et al. 2004; Maragakis and Karplus 2005; Lou and Cukier 2006; Whitford et al. 2007; Snow et al. 2007). However, due to the magnitude and timescales involved, spontaneous transitions between the open and closed conformations have not been achieved until now by all-atom MD simulations. Using TEE-REX, spontaneous transitions between the open and closed structures of ADK are facilitated, and a fully atomistic description of the transition pathway and its underlying mechanics could be achieved (Kubitzki and de Groot 2008). To this end, different essential subspaces {es} were constructed from short MD simulations of either conformation as well as from a combined ensemble holding structures from both the open and closed conformation. In the latter case, {es} modes were excited containing the difference X-ray mode connecting the open and closed experimental structures.

The observed transition pathway can be characterized by two phases. Starting from the closed conformation (Fig. 9.10 left), the LID remains essentially closed
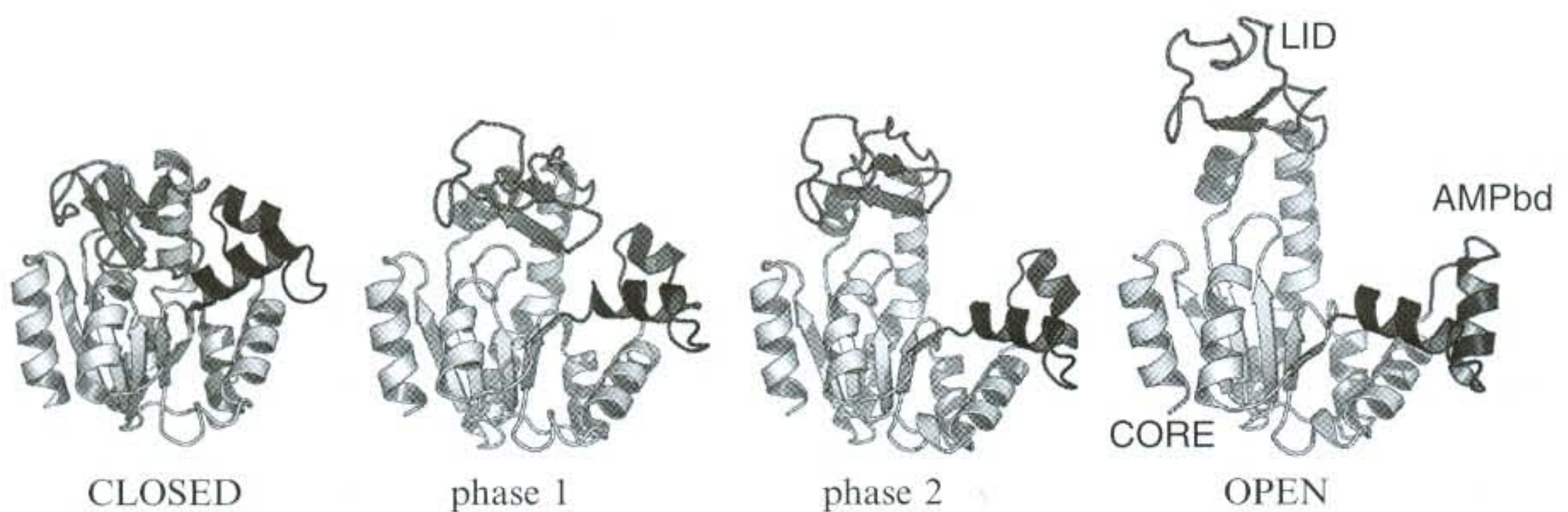

CLOSED   phase 1   phase 2   OPEN

Fig. 9.10 Closed (left) and open (right) crystal structures of *E. coli* adenylate kinase (ADK) together with intermediate structures characterizing the two phases of the closed-open transition. ADK has domains CORE (light grey), AMPbd (black) and LID (dark grey). The transition state inhibitor Ap_5A is removed in the closed crystal structure (left)

while the AMPbd, comprising helices α_2 and α_3, assumes a half-open conformation. In doing so, α_2 bends towards helix α_4 of the CORE by 15 degrees with respect to α_3. This opening of the AMP binding cleft could facilitate an efficient release of the formed product. For the second phase, a partially correlated opening of the LID domain together with the AMPbd is observed. Compared to coarse-grained approaches, all-atom TEE-REX simulations allow detailed analyses of inter-residue interactions. For ADK, a highly stable salt bridge between residues Asp118 and Lys136 forms during phase one, connecting the LID and CORE domains. Estimating the total non-bonded interaction between LID and CORE, it was found that this salt-bridge contributes substantially to the interaction of the two domains. Breaking this salt bridge via mutation, e.g. Asp118Ala, should thus decrease the stability of the open state. From a comparison of fourteen Protein Data Bank (PDB) structures from yeast, maize, human and bacterial ADK, eleven structures feature such a salt-bridge motif at the LID-CORE interface.

Alternative transition pathways seem possible, however an analysis of all TEE-REX simulations suggests a high free energy barrier obstructing the full opening of the AMPbd after the LID has opened. Together with the observed larger fluctuations in secondary structure elements, indicating high internal strain energies, the enthalpic penalty along this route possibly renders it unfavourable as a transition pathway of ADK.

9.4   Methods for Functional Mode Prediction

As discussed in the previous section, functional modes in proteins are usually those with the lowest frequencies. Apart from molecular dynamics based techniques, there are several alternative methods that focus on the prediction of these essential degrees of freedom based on a single input structure.

9.4.1   Normal Mode Analysis

Normal mode analysis (NMA) is one of the major simulation techniques used to probe the large-scale, shape-changing motions in biological molecules (Gō et al. 1983; Brooks and Karplus 1983; Levitt et al. 1983). These motions are often coupled to function and a consequence of binding other molecules like substrates, drugs or other proteins. In NMA studies it is implicitly assumed that the normal modes with the largest fluctuation (lowest frequency modes) are the ones that are functionally relevant, because, like function they exist by evolutionary "design" rather than by chance.

Normal mode analysis is a harmonic analysis. The underlying assumption is that the conformational energy surface can be approximated by a parabola, despite the fact that functional modes at physiological temperatures are highly anharmonic

(Brooks and Karplus 1983; Austin et al. 1975). To perform a normal mode analysis one needs a set of coordinates, a force field describing the interactions between constituent atoms, and software to perform the required calculations. The performance of a normal mode analysis in Cartesian coordinate space requires three main calculation steps.

1. Minimization of the conformational potential energy as a function of the atomic coordinates.
2. The calculation of the so-called "Hessian" matrix

$$H = \frac{\partial^2 V}{\partial x_i \partial x_j}$$

which is the matrix of second derivatives of the potential energy with respect to the mass-weighted atomic coordinates.

3. The diagonalization of the Hessian matrix. This final steps yields eigenvalues and eigenvectors (the "normal modes").

Energy minimization can require quite a lot of CPU time. Furthermore, as the Hessian matrix is a $3N \times 3N$ matrix, where N is the number of atoms, the last step can be computationally demanding.

9.4.2 Elastic Network Models

Elastic or Gaussian network models (Tirion 1996) (ENM) are basically a simplification of normal mode analysis. Usually, instead of an all atom representation, only C_α atoms are taken into account. This means a ten-fold reduction of the number of particles which decreases the computational effort dramatically. Moreover, as the input coordinates are taken as representing the ground state, no energy minimization is required.

The potential energy is calculated according to

$$V = \frac{\gamma}{2} \sum_{|r_{ij}^0| < R_C} (r_{ij} - r_{ij}^0)^2$$

where γ denotes the spring constant and R_C the cut-off distance. Regarding the drastic assumptions inherent in normal mode analysis, these simplifications do not mean a severe loss of quality. This, together with the relatively low computational costs, explains the current popularity of elastic network models. ENM calculations are also offered on web servers such as ElNemo (Suhre and Sanejouand 2004a, b) (http://www.igs.cnrs-mrs.fr/elnemo/) and AD-ENM (Zheng and Doniach 2003; Zheng and Brooks 2005) (http://enm.lobos.nih.gov/).

9.4.3 CONCOORD

CONCOORD (de Groot et al. 1997) uses a geometry-based approach to predict protein flexibility. The three-dimensional structure of a protein is determined by various interactions such as covalent bonds, hydrogen bonds and non-polar interactions. Most of these interactions remain intact during functionally relevant conformational changes. This notion lies at the heart of the CONCOORD simulation method: based on an input structure, *alternative* structures are generated that share the large majority of interactions found in the original configuration. To this end, in the first step of a CONCOORD simulation (Fig. 9.11) interactions in a single input structure are analyzed and turned into geometrical constraints, mainly distance constraints with upper and lower bounds for atomic distances but also angle constraints and information about planar and chiral groups. This geometrical description of the structure can be compared to a construction plan of the protein. In the second step, starting from random atomic coordinates, the structure is iteratively rebuilt based on the predefined construction plan, commonly several hundreds of times. As each run starts from random coordinates, the method does not suffer from sampling problems like MD simulations and the resulting ensemble covers the whole conformational space that is available within the predefined constraints. However, the method does not provide information about the path between two conformational substates or about timescales and energies (Fig. 9.12).

9.4.3.1 Applications

CONCOORD and the newly developed extension tCONCOORD (t stands for transition) (Seeliger et al. 2007) have been applied to diverse proteins. Adenylate kinase displays a distinct domain-closing motion upon binding to its substrate (ATP/AMP) or an inhibitor (see Fig. 9.13 left) with a C_α-RMSD of 7.6 Å between the ligand-bound and the ligand-free conformation. Two tCONCOORD simulations were carried out using a closed conformation (PDB 1AKE) as input. In one simulation the ligand (Ap_5A) was removed. Fig. 9.13 (right) shows the result of a principal components analysis (PCA) applied to the experimental structures. The first eigenvector (x-axis) corresponds to the domain-opening motion indicated by the arrow in Fig. 9.13 (left). Every dot in the plot represents a single structure. Red dots represent the ensemble that has been generated using the closed conformation of adenylate kinase without ligand as input. Green dots represent the ensemble that has been generated using the ligand-bound structure as input. Whereas the simulation with inhibitor basically samples closed conformations around the ligand-bound state, the ligand-free simulation samples both, open and closed conformations, thereby reaching the experimentally determined open conformations with RMSD's of 2.4, 2.6, and 3.1 Å for 1DVR, 1AK2, and 4AKE, respectively. In structure-based drug design, often the reverse problem,
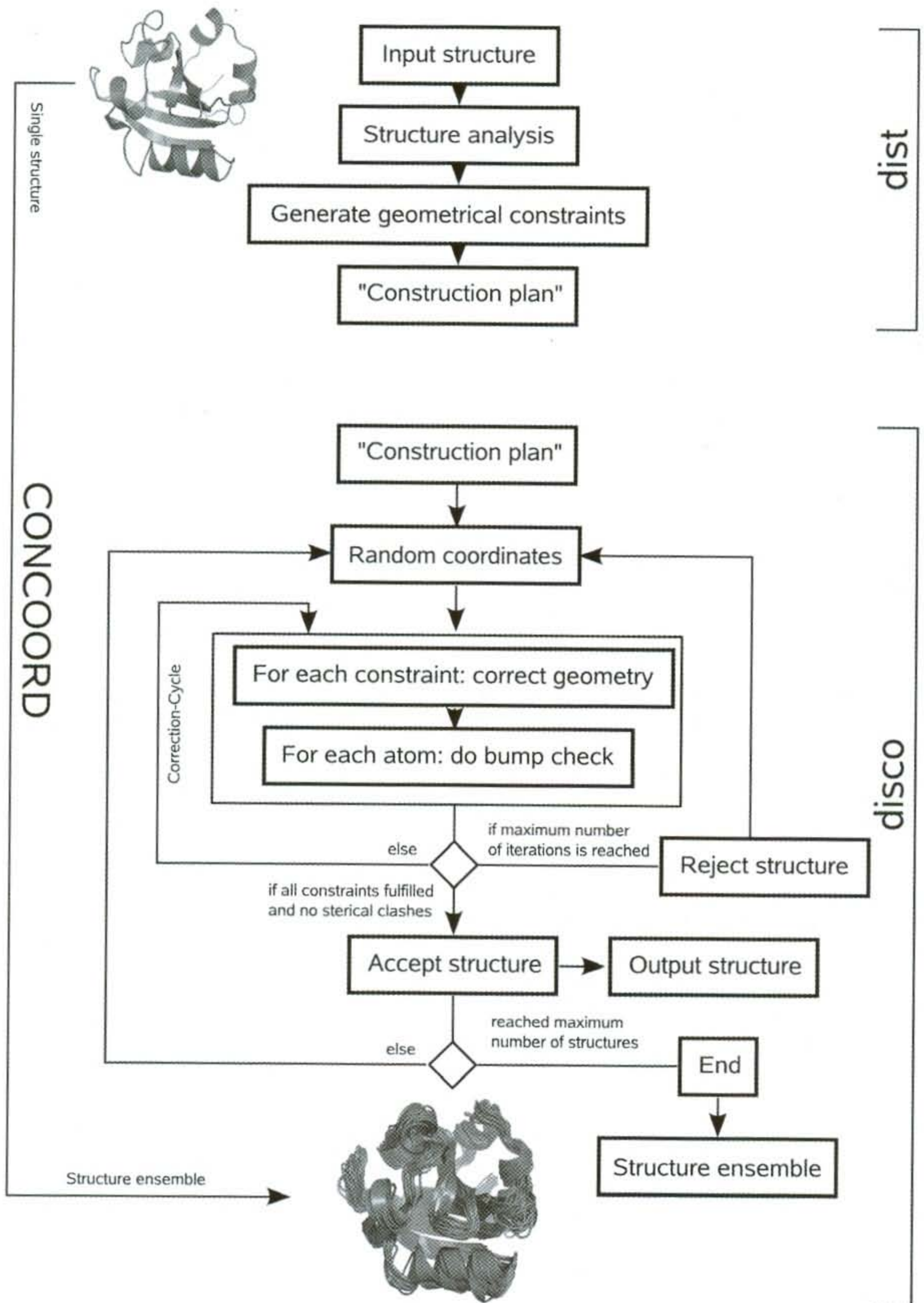
Fig. 9.11 Schematic representation of the CONCOORD method for generating structure ensembles from a single input structure. In a first step (program *dist*) a single input structure is analyzed and turned into a geometric description of the protein. In a second step (program *disco*) the structure is rebuilt based on the predefined constraints, starting from random coordinates
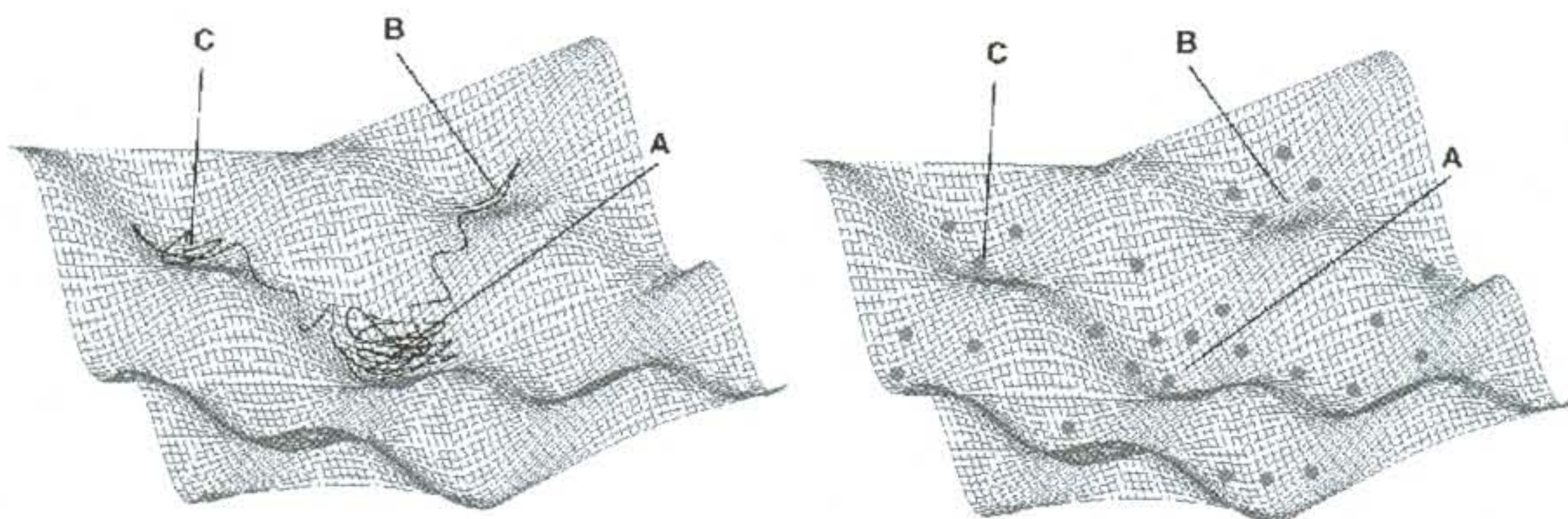
Fig. 9.12 Comparison of the sampling properties of Molecular Dynamics and CONCOORD on hypothetic energy landscapes. A MD-trajectory (left) "walks" on the energy landscape, thereby providing information about timescales and paths between conformational substates. The (non-deterministic) CONCOORD-ensemble (right) "jumps" on the energy landscape, thereby offering better sampling of the conformational space
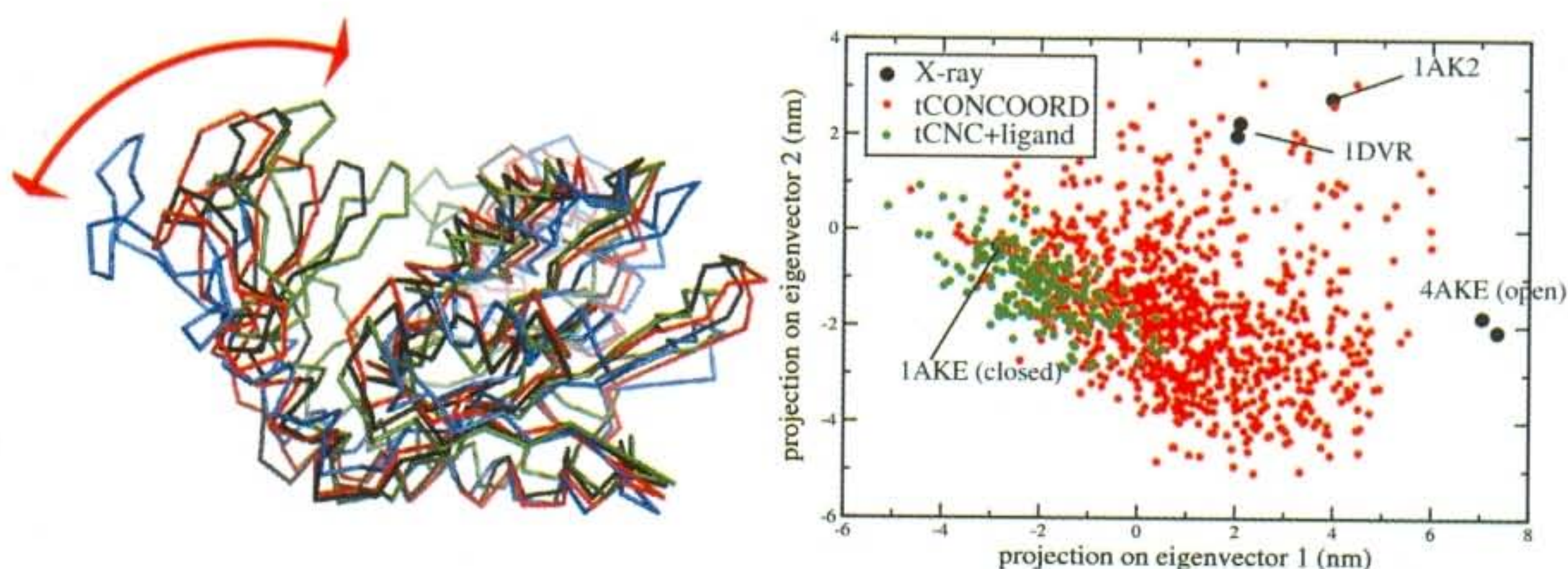

Fig. 9.13 Left: Overlay of X-ray structures of adenylate kinase. Right: principal component analysis. Two tCONCOORD ensembles are projected onto the first two eigenvectors of a PCA carried out on an ensemble of X-ray structures. The ensemble represented by red dots has been started from a closed conformation (1AKE) with removed inhibitor. The generated ensemble samples both, closed and open conformations. The ensemble represented with green dots has also been started from a closed conformation (1AKE) but with inhibitor present. The generated ensemble only samples closed conformations around the ligand bound conformation

predicting ligand-bound structures from unbound conformations, needs to be addressed. A tCONCOORD simulation starting with an open conformation (4AKE) as input produced structures that approach the closed conformations with RMSD's of 2.5, 2.9, and 3.3 Å for 1DVR, 1AK2, and 1AKE, respectively. Thus, the functional domain-opening motion has been predicted in both cases, when using a closed, ligand-bound conformation as input and when using an open, ligand-free conformation as input.

Because of its computational efficiency, CONCOORD can be routinely applied to extract functionally relevant modes of flexibility for molecular systems that are beyond the size limitations of other atomistic simulation techniques like molecular dynamics simulations. An application to the chaperonin GroEL-GroES complex

that contains more than 8,000 amino acids revealed a novel form of coupling between intra-ring and inter-ring cooperativity (de Groot et al. 1999). Each GroEL ring displays two main modes of collective motion: the main conformational transition upon binding of the co-chaperonin GroES, and a secondary transition upon ATP binding (Fig. 9.14 upper right panel). CONCOORD simulations of a single GroEL ring did not show any coupling between these modes, whereas simulations of the double ring system showed a strict correlation between the two modes, thereby providing an explanation for how nucleotide binding is coupled to GroES affinity in the double ring, but not in a single ring.

9.5 Summary and Outlook

Computational methods gain growing recognition in structural biology and protein research. Protein function is usually a dynamic process involving structural rearrangements and conformational transitions between stable states. Since such dynamic processes are difficult to study experimentally, *in silico* methods can significantly contribute to the understanding of protein function at atomic resolution.
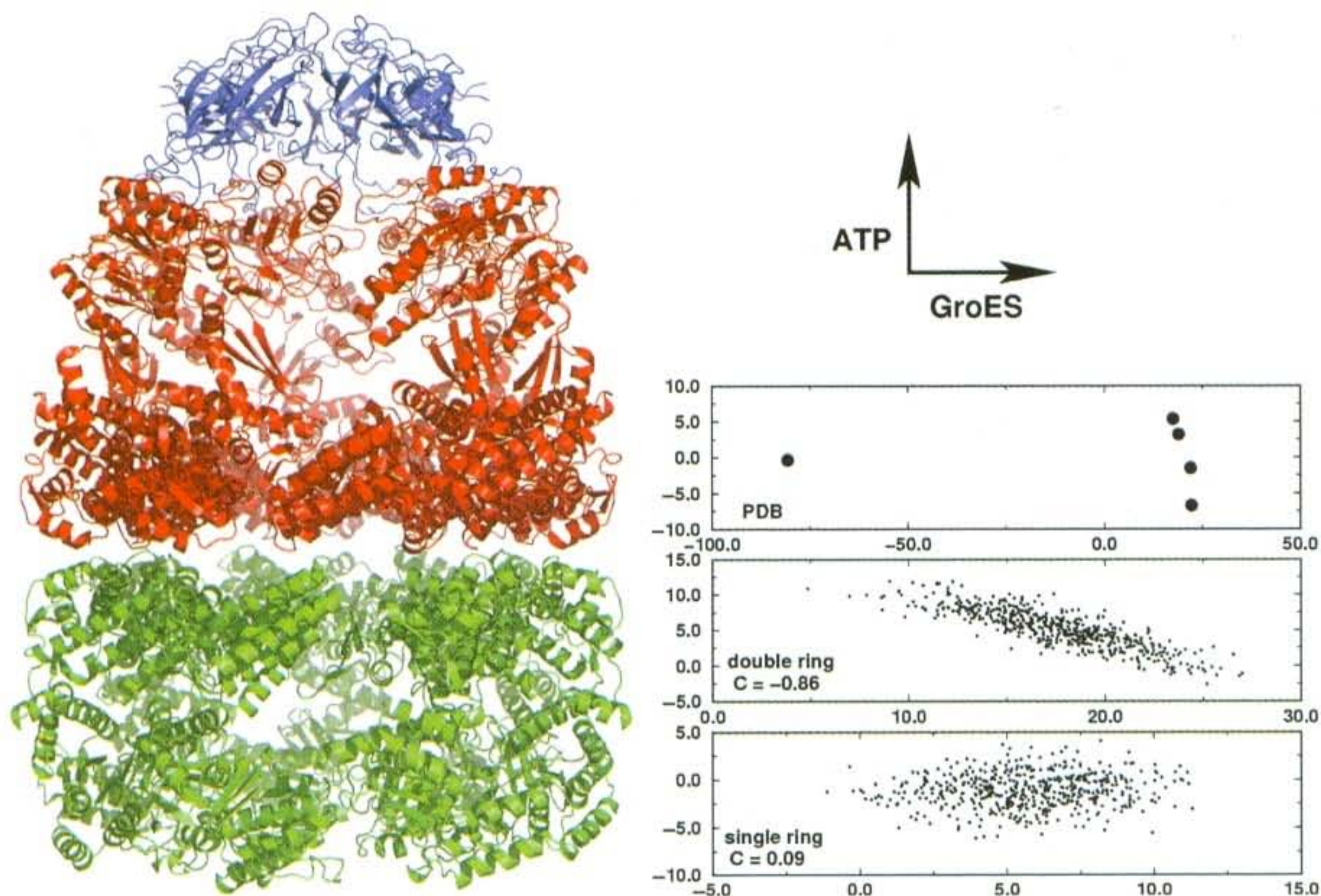

Fig. 9.14 Asymmetric GroEL-GroES complex (left), together with CONCOORD simulation results (right). The GroEL-GroES complex consists of the co-chaperonin GroES (blue), the trans-ring of GroEL, bound to GroES (red), and the cis-ring (green). A principal component analysis revealed two main structural transitions per GroEL ring, upon nucleotide binding (vertical axis in the right panels) and GroES binding (horizontal axis), respectively. In simulations of the double ring, but not in a single ring, these modes were found to be coupled, suggesting a coupling between intra-ring and inter-ring cooperativity

The most prominent method to study protein dynamics is molecular dynamics (MD), where atoms are treated as classical particles and their interactions are approximated by an empirical force field. Newton's equations of motion are solved at discrete time steps, leading to a trajectory that describes the dynamical behaviour of the system. Despite their growing popularity the scope of application for MD simulations is limited by computational demands. Within the next 10 years the accessible timescales for the simulation of average sized proteins will, in all likelihood, not exceed the low microsecond range for most biomolecular systems. However, since functionally relevant protein dynamics is usually represented by collective, low-frequency motions taking place on the micro- to millisecond timescale, standard MD simulations are ill-suited to be routinely applied to study conformational dynamics of large biomolecules.

Different methodologies have been developed to alleviate this sampling problem that standard MD suffers from. One approach is to reduce the number of particles, either by fusing groups of atoms into pseudo-atoms (coarse-graining), or by replacing explicit solvent molecules with an implicit solvent continuum model. In both cases the number of particles is significantly reduced, facilitating much longer time scales than in all-atom simulations using explicit solvent. However, the loss of "resolution" inherent to both methods may limit their accuracy and hence, their applicability. Other approaches retain the atomistic description and pursue different sampling strategies.

Generalized ensemble algorithms such as Replica Exchange (REX) make use of the fact that conformational transitions occur more frequently at higher temperatures. In standard temperature REX, several copies (replicas) of the system are simulated with MD at different temperatures, with frequent exchanges between replicas. Thereby, low-temperature replicas utilize the enhanced barrier-crossing capabilities of high-temperature replicas. Although dynamical information gets lost in this setup, each replica still represents a Boltzmann ensemble at its corresponding temperature, providing valuable information about thermodynamics and thus the stability of different conformational substates. Although often used in the context of protein folding, REX simulations at full atomic resolution quickly become computationally very demanding for systems comprising more than a few thousand atoms.

Whereas REX is an unbiased sampling method, several other methods exist that bias the system in order to enhance sampling predominantly along certain collective degrees of freedom. Functionally relevant protein motions often correspond to those eigenvectors of the covariance matrix of atomic fluctuations having the largest eigenvalues. If these eigenvectors are known from a principal component analysis (PCA), either using experimental data or previous simulations, they can be used in simulation protocols like Conformational Flooding or Essential Dynamics (ED). However, in both methods the enhancement in sampling is paid for by losing the canonical properties of the resulting trajectory.

The recently developed TEE-REX protocol combines the favourable properties of REX with those resulting from a specific excitation of functionally relevant modes (as e.g. in ED), while at the same time avoiding the aforementioned drawbacks of each

method. In particular, approximate canonical integrity of the reference ensemble is maintained and sampling along the main collective modes of motion is significantly enhanced. The resulting reference ensemble can thus be used to calculate equilibrium properties of the system which allows comparison with experimental data.

Although significant progress has been made in the development of enhanced sampling methods, computational demands of MD based methods are still large and simulations usually take several weeks to months of computation time on multiprocessor state-of-the-art computer clusters. For many questions in structural biology it is already beneficial to have an idea about possible protein conformations and functional modes without the need to get detailed information about energetics and timescales. In this respect, elastic network models offer a cheap way to get an estimate of possible functional protein motions. Although drastic assumptions are made and no atomistic picture is obtained the predicted collective motions are often in qualitatively good agreement with experimental results. Another computational efficient way which retains the atomistic description of protein structures is the CONCOORD method where a protein is described with geometrical constraints. Based on a construction plan derived from a single input structure, an ensemble of structures is generated which represents an exhaustive sampling of conformational space that is available within the predefined constraints. However, no information about timescales or energies is obtained.

Right now there is no single method that is routinely applicable to predict functionally relevant protein motions from a given three-dimensional structure. However, there are a large number of methods available, capturing different aspects of the problem and contributing to our understanding of protein function. Thus, combinations of existing methods will presumably be the most straightforward way of enhancing the predictive power of *in silico* methods.

References

Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev 106:1589–1615

Affentranger R, Tavernelli I, di Iorio E (2006) A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. J Chem Theory Comput 2:217–228

Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. Proteins 17:412–425

Amadei A, Linssen ABM, de Groot BL, et al. (1996) An efficient method for sampling the essential subspace of proteins. J Biom Str Dyn 13:615–626

Amadei A, de Groot BL, Ceruso M-A, et al. (1999) A kinetic model for the internal motions of proteins: diffusion between multiple harmonic wells. Proteins 35:283–292

Anderson HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 72:2384–2393

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230

Austin RH, Beeson KW, Eisenstein L, et al. (1975) Dynamics of ligand binding to myoglobin. Biochemistry 14(24):5355–5373

Bahar I, Erman B, Haliloglu T, et al. (1997) Efficient characterization of collective motions and inter-residue correlations in proteins by low-resolution simulations. Biochemistry 36:13512–13523

Bartels C, Karplus M (1998) Probability distributions for complex systems: Adaptive umbrella sampling of the potential energy. J Phys Chem B 102:865–880

Berendsen HJC, Postma JPM, di Nola A, et al. (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

Berg BA, Celik T (1992) New approach to spin-glass simulations. Phys Rev Lett 69:2292–2295

Berg BA, Neuhaus T (1991) Multicanonical algorithms for first-order phase transitions. Phys Lett 267:249–253

Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry, fifth edition. WH Freeman, New York

Bond PJ, Holyoake J, Ivetac A, et al. (2007) Coarse-grained molecular dynamics simulations of membrane proteins and peptides. J Struct Biol 157:593–605

Brooks B, Karplus M (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci USA 80:6571–6575

Brooks BR, Bruccoleri RE, Olafson BD, et al. (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculations. J Comp Chem 4:187–217

Burykin A, Warshel A (2003) What really prevents proton transport through aquaporin? Charge self-energy versus proton wire proposals. Biophys J 85:3696–3706

Cecchini M, Rao F, Seeber M, et al. (2004) Replica exchange molecular dynamics simulations of amyloid peptide aggregation. J Chem Phys 121:10748–10756

Chakrabarti N, Tajkhorshid E, Roux B, et al. (2004) Molecular basis of proton blockage in aquaporins. Structure 12:65–74

Chen H, Wu Y, Voth GA (2006) Origins of proton transport behavior from selectivity domain mutations of the aquaporin-1 channel. Biophys J 90:L73–L75

Cheng X, Cui G, Hornak V, et al. (2005) Modified replica exchange simulation for local structure refinement. J Phys Chem B 109:8220–8230

Chodera JD, Swope WC, Pitera JW, et al. (2007) Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. J Chem Theory Comput 3:26–41

Christen M, van Gunsteren WF (2006) Multigraining: an algorithm for simultaneous fine-grained and coarse-grained simulation of molecular systems. J Chem Phys 124:154106

Cook A, Fernandez E, Lindner D, et al. (2005) The structure of the nuclear export receptor cse1 in its cytosolic state reveals a closed conformation incompatible with cargo binding. Mol Cell 18:355–357

Currie MG, Fok KF, Kato J, et al. (1992) Guanylin: an endogenous activator of intestinal guanylate cyclise. Proc Natl Acad Sci USA 89:947–951

de Groot BL, Grubmüller H (2001) Water permeation across biological membranes: Mechanism and dynamics of aquaporin-1 and GlpF. Science 294:2353–2357

de Groot BL, Amadei A, Scheek RM, et al. (1996a) An extended sampling of the configurational space of HPr from E coli. Proteins 26:314–322

de Groot BL, Amadei A, van Aalten DMF, et al. (1996b) Towards an exhaustive sampling of the configurational spaces of the two forms of the peptide hormone guanylin. J Biomol Str Dyn 13:741–751

de Groot BL, van Aalten DMF, Scheek RM, et al. (1997) Prediction of protein conformational freedom from distance constraints. Proteins 29:240–251

de Groot BL, Hayward S, van Aalten DMF, et al. (1998) Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. Proteins 31:116–127

de Groot BL, Vriend G, Berendsen HJC (1999) Conformational changes in the chaperonin GroEL: new insights into the allosteric mechanism. J Mol Biol 286:1241–1249

de Groot BL, Engel A, Grubmüller H (2001) A refined structure of human Aquaporin-1. FEBS Lett 504: 206–211

de Groot BL, Frigato T, Helms V, et al. (2003) The mechanism of proton exclusion in the aquaporin-1 water channel. J Mol Biol 333:279–293

Dixon MM, Nicholson H, Shewchuk L, et al. (1992) Structure of a hinge-bending bacteriophage T4 lysozyme mutant Ile3 → Pro. J Mol Biol 227:917–933

Duda RO, Hart PE, Stork DG (2001) Pattern Classification, second edition. Wiley, New York

Faber HR, Matthews BW (1990) A mutant T4 lysozyme displays five different crystal conformations. Nature 348:263–266

Frauenfelder H, Leeson DT (1998) The energy landscape in non-biological and biological molecules. Nat Struct Biol 5:757–759

Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254:1598–1603

Fu D, Libson A, Miercke LJ, et al. (2000) Structure of a glycerol-conducting channel and the basis for its selectivity. Science 290: 481–486

Fukunishi H, Watanabe O, Takada S (2002) On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. J Chem Phys 116:9058–9067

García AE (1992) Large-amplitude nonlinear motions in proteins. Phys Rev Lett 68:2696–2699

García AE, Onuchic JN (2003) Folding a protein in a computer: An atomic description of the folding/unfolding of protein A. Proc Natl Acad Sci USA 100:13898–13903

G N, Noguti T, Nishikawa T (1983) Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci USA 80:3696–3700

Gerstein M, Lesk AM, Chothia C (1994) Structural mechanisms for domain movements in proteins. Biochemistry 33:6739–6749

Gosh A, Rapp CS, Friesner RA (1998) Generalized Born model based on a surface integral formulation. J Phys Chem B 102:10983–10990

Grubmüller H (1995) Predicting slow structural transitions in macromolecular systems: Conformational flooding. Phys Rev E 52:2893–2906

Hansmann UHE (1997) Effective way for determination of multicanonical weights. Phys Rev E 56:6200–6203

Hayward S, Kitao A, Gō N (1995) Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. Proteins 23:177–186

He J, Zhang Z, Shi Y, et al. (2003) Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions. J Chem Phys 119:4005–4017

Hockney RW, Goel SP, Eastwood JW (1973) 10000 particle molecular dynamics model with long-range forces. Chem Phys Lett 21:589–591

Hub JS, de Groot BL (2008) Mechanism of selectivity in aquaporins and aquaglyceroporins. Proc Natl Acad Sci USA 105: 1198–1203

Iba Y (2001) Extended ensemble Monte Carlo. Int J Mod Phys C 12:623–656

Ilan B, Tajkhorshid E, Schulten K, et al. (2004) The mechanism of proton exclusion in aquaporin channels. Proteins 55:223–228

Jean-Charles A, Nicholls A, Sharp K, et al. (1991) Electrostatic contributions to solvation energies: comparison of free energy perturbation and continuum calculations. J Am Chem Soc 113:1454–1455

Jorgensen WL, Chandrasekhar J, Madura JD, et al. (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935

Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236

Karplus M, Gao YQ (2004) Biomolecular motors: the F1-ATPase paradigm. Curr Opin Struct Biol 14:250–259

Karplus M, Kushick JN (1981) Method for estimating the configurational entropy of macromolecules. Macromolecules 14:325–332

Kempf JG, Loria JP (2003) Protein dynamics from solution NMR theory and applications. Cell Biochem Biophys 37:187–211

Kitao A, Gō N (1999) Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol 9:143–281

Kitao A, Hirata F, Gō N (1991) The effects of solvent on the conformation and the collective motions of proteins - normal mode analysis and molecular-dynamics simulations of melittin in water and vacuum. Chem Phys 158:447–472

Kitao A, Hayward S, Gō N (1998) Energy landscape of a native protein: Jumping-among-minima model. Proteins 33:496–517

Kokubo H, Okamoto Y (2004) Prediction of membrane protein structures by replica-exchange Monte Carlo simulations: case of two helices. J Chem Phys 120:10837–10847

Kubitzki MB, de Groot BL (2007) Molecular dynamics simulations using temperature-enhanced essential dynamics replica exchange. Biophys J 92:4262–4270

Kubitzki MB, de Groot BL (2008) The atomistic mechanism of conformational transition in adenylate kinase: a TEE-REX molecular dynamics study. Structure 16:1175–1182

Kumar S, Bouzida D, Swendsen RH, et al. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method. J Comp Chem 13:1011–1021

Kumar S, Payne PW, Vásquez M (1996) Method for free-energy calculations using iterative techniques. J Comput Chem 17:1269–1275

Kuroki R, Weaver LH, Matthews BW (1993) A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. Science 262:2030–2033

Levitt M, Sander C, Stern PS (1983) Normal-mode dynamics of a protein: Bovine pancreatic trypsin inhibitor. Int J Quant Chem: Quant Biol Symp 10:181–199

Levy RM, Karplus M, Kushick J, et al. (1984a) Evaluation of the configurational entropy for proteins: application to molecular dynamics of an α-helix. Macromolecules 17:1370–1374

Levy RM, Srinivasan AR, Olsen WK, et al. (1984b) Quasi-harmonic method for studying very low frequency modes in proteins. Biopolymers 23:1099–1112

Liu P, Kim B, Friesner RA, et al. (2005) Replica exchange with solute tempering: A method for sampling biological systems in explicit water. Proc Natl Acad Sci USA 102:13749–13754

Lou H, Cukier RI (2006) Molecular dynamics of apo-adenylate kinase: a distance replica exchange method for the free energy of conformational fluctuations. J Phys Chem B 110:24121–24137

Luo R, David L, Gilson ML (2002) Accelerated Poisson-Boltzmann calculations for static and dynamic systems. J Comput Chem 23:1244–1253

Lyman E, Zuckerman DM (2006) Ensemble-based convergence analysis of biomolecular trajectories. Biophys J 91:164–172

Maragakis P, Karplus M (2005) Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. J Mol Biol 352:807–822

Marinari E, Parisi G (1992) Simulated tempering: a new Monte Carlo scheme. Europhys Lett 19:451–458

Marrink SJ, de Vries AH, Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. J Phys Chem B 108:750–760

Matthews BW, Remington SJ (1974) The three dimensional structure of the lysozyme from bacteriophage T4. Proc Natl Acad Sci USA 71:4178–4182

McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590

Mitsutake A, Sugita Y, Okamoto Y (2001) Generalized-ensemble algorithms for molecular simulations of biopolymers. Biopolymers 60:96–123

Moffat K (2003) The frontiers of time-resolved macromolecular crystallography: movies and chirped X-ray pulses. Faraday Discuss 122:65–77

Murata K, Mitsuoka K, Walz T, et al. (2000) Structural determinants of water permeation through Aquaporin-1. Nature 407: 599–605

Müller CW, Schulz GE (1992) Structure of the complex between adenylate kinase from *Eschericia coli* and the inhibitor Ap$_5$A refined at 19 Å resolution: a model for a catalytic transition state. J Mol Biol 224:159–177

Müller CW, Schlauderer G, Reinstein J, et al. (1996) Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. Structure 4:147–156

Nguyen PH, Mu Y, Stock G (2005) Structure and energy landscape of a photoswitchable peptide: a replica exchange molecular dynamics study. Proteins 60:485–494

Nose S (1984) A unified formulation of the constant temperature molecular dynamics method. J Chem Phys 81:511–519

Pitera JW, Swope W (2003) Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. Proc Natl Acad Sci USA 100:7587–7592

Rao F, Caflisch A (2003) Replica exchange molecular dynamics simulations of reversible folding. J Chem Phys 119:4035–4042

Romo TD, Clarage JB, Sorensen DC, et al. (1995) Automatic identification of discrete substates in proteins: singular value decomposition analysis of time-averaged crystallographic refinements. Proteins 22:311–321

Schotte F, Lim M, Jackson TA, et al. (2003) Watching a protein as it functions with 150 ps time-resolved X-ray crystallography. Science 300:1944–1947

Seeliger D, Haas J, de Groot BL (2007) Geometry-based sampling of conformational transitions in proteins. Structure 15:1482–1492

Seibert MM, Patriksson A, Hess B, et al. (2005) Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. J Mol Biol 354:173–183

Shapiro YE, Meirovitch E (2006) Activation energy of catalysis-related domain motion in E coli adenylate kinase. J Phys Chem B 110:11519–11524

Shapiro YE, Kahana E, Tugarinov V, et al. (2002) Domain flexibility in ligand-free and inhibitor bound Eschericia coli adenylate kinase based on a mode-coupling analysis of ^{15}N spin relaxation. Biochemistry 41:6271–6281

Smith GR, Bruce AD (1996) Multicanonical Monte Carlo study of solid-solid phase coexistence in a model colloid. Phys Rev E 53:6530–6543

Snow C, Qi G, Hayward S (2007) Essential dynamics sampling study of adenylate kinase: comparison to citrate synthase and implication for the hinge and shear mechanisms of domain motion. Proteins 67:325–337

Still WC, Tempczyk A, Hawley RC, et al. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 112:6127–6129

Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141–151

Sugita Y, Kitao A, Okamoto Y (2000) Multidimensional replica-exchange method for free-energy calculations. J Chem Phys 113:6042–6051

Suhre K, Sanejouand YH (2004a) ElNemo: a normal mode web-server for protein movement analysis and the generation of templates for molecular replacement. Nucl Acids Res 32:610–614

Suhre K, Sanejouand YH (2004b) On the potential of normal mode analysis for solving difficult molecular replacement problems. Act Cryst D 60:796–799

Sui H, Han B-G, Lee JK, et al. (2001) Structural basis of water-specific transport through the AQP1 water channel. Nature 414: 872–878

Tai K (2004) Conformational sampling for the impatient. Biophys Chem 107:213–220

Tajkhorshid E, Nollert P, Jensen MØ, et al. (2002) Control of the selectivity of the aquaporin water channel family by global orientational tuning. Science 296: 525–530

Teeter MM, Case DA (1990) Harmonic and quasi harmonic descriptions of crambin. J Phys Chem 94:8091–8097

Temiz NA, Meirovitch E, Bahar I (2004) Eschericia coli adenylate kinase dynamics: comparison of elastic network model modes with mode-coupling ^{15}N-NMR relaxation data. Proteins 57:468–480

Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter atomic analysis. Phys Rev Lett 77:186–195

Tugarinov V, Shapiro YE, Liang Z, et al. (2002) A novel view of domain flexibility in E coli adenylate kinase based on structural mode-coupling ^{15}N NMR spin relaxation. J Mol Biol 315:155–170

Van Aalten DMF, Amadei A, Vriend G, et al. (1995a) The essential dynamics of thermolysin - confirmation of hinge-bending motion and comparison of simulations in vacuum and water. Prot Eng 8:1129–1136

Van Aalten DMF, Findlay JBC, Amadei A, et al. (1995b) Essential dynamics of the cellular retinol binding protein - evidence for ligand induced conformational changes. Prot Eng 8:1129–1136

Van Gunsteren WF, Berendsen HJC (1987) Groningen Molecular Simulation (GROMOS) Library Manual. Biomos, Groningen

Van Gunsteren WF, Berendsen HJC (1990) Computer-simulation of molecular-dynamics – methodology, applications, and perspectives in chemistry. Angew Chem Int Edit Engl 29:992–1023

Warshel A, Kato M, Pisliakov AV (2007) Polarizable force fields: history test cases and prospects. J Chem Theory Comput 3:2034–2045

Weiner SJ, Kollman PA, Nguyen DT, et al. (1986) An all atom force field for simulations of proteins and nucleic acids. J Comp Chem 7:230–252

Weiss S (1999) Fluorescence spectroscopy of single biomolecules. Science 283:1676–1683

Whitford PC, Miyashita O, Levy Y, et al. (2007) Conformational transitions of adenylate kinase: switching by cracking. J Mol Biol 366:1661–1671

Xu Z, Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric Gro-EL-GroES-(ADP)$_7$ chaperonin complex. Nature 388:741–750

Zachariae U, Grubmüller H (2006) A highly strained nuclear conformation of the exportin Cse1p revealed by molecular dynamics simulations. Structure 14:1469–1478

Zhang X-J, Wozniak JA, Matthews BW (1995) Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. J Mol Biol 250:527–552

Zhang Z, Shi Y, Liu H (2003) Molecular dynamics simulations of peptides, and proteins with amplified collective motions. Biophys J 84:3583–3593

Zheng W, Brooks BR (2005) Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: myosin versus kinesin. Biophys J 89(1):167–178

Zheng W, Doniach S (2003) A comparative study of motor-protein motions by using a simple elastic-network model. Proc Natl Acad Sci USA 100(23):13253–13258

Zhou R Berne BJ, Germain R (2001) The free energy landscape for β-hairpin folding in explicit water. Proc Natl Acad Sci USA 98:14931–14936