Chem Soc Rev

A theoretical view of protein dynamics

Journal:	*Chemical Society Reviews*
Manuscript ID:	CS-TRV-12-2013-060474.R2
Article Type:	Review Article
Date Submitted by the Author:	18-Mar-2014
Complete List of Authors:	Orozco, Modesto; Institute for Research in Biomedicine and Barcelona Supercomputer Center, Life Sciences

Orozco, February 19th 2014

A THEORETICAL VIEW OF PROTEIN DYNAMICS

Modesto Orozco[1,2,3]

[1] Institute for Research in Biomedicine (IRB Barcelona), Baldiri i Reixac 8, Barcelona 08028, Spain

[2] Joint BSC-GRG-IRB Program in Computational Biology, Barcelona, Spain

[3] Department of Biochemistry and Molecular Biology, Biology Faculty. University of Barcelona, Barcelona, Spain

To whom correspondence should be addressed:

modesto.orozco@irbbarcelona.org

SUMMARY

Proteins are fascinating supramolecular structures, which are able to recognize ligands transforming binding information into chemical signals. They can transfer information across the cell, can catalyse complex chemical reactions, and are able to transform energy into work with much more efficiency than any human engine. The unique abilities of proteins are tightly coupled with their dynamics properties, which are coded in a complex way in the sequence and that have been carefully refined by evolution. Despite its importance, our experimental knowledge on protein dynamics is still rather limited, and mostly derived from theoretical calculations. I will review here, in a systematic way, the current state-of-the-art of theoretical approaches to protein dynamics, emphasizing the most recent advances, examples of use and the expected lines of development in the near future.

PROTEINS: THE MACHINES OF LIFE

Proteins are large macromolecules responsible of the most complex processes in the cell. They are fascinating molecular machines able to accommodate structural and dynamical changes in response to external stimuli, such as alterations in the molecular environment (i.e. changes in membrane composition, temperature or pH), chemical modifications (i.e., phosphorylation), or the binding of other molecules. Conformational plasticity is crucial for protein function (1-7), and preservation of flexibility through specific deformation modes is instrumental for the functional role of proteins (1-8). It seems that evolution has not only refined the flexibility pattern of proteins, but it has also exploited it to create new proteins following a conservative low-risk paradigm (9-14).

A comprehensive analysis of protein-protein complexes revealed that the movements required for effective dimerization are often just an extension of the spontaneous deformation modes of the unbound structures (15). Furthermore, large conformational transitions (in some cases more than 20 Å in RMSd) correlate very well with protein deformation patterns (16-17).

The impact of dynamics in the properties of proteins is of particular relevance in the case of “intrinsic disordered proteins (IDPs)”. These proteins, which were ignored for structural biology for decades, account for nearly 40% of human proteome, and are especially involved in regulatory functions (18-21). Under native conditions, IDPs lack a defined 3D structure, and should be described as a dynamic ensemble (18-21). Extreme structural diversity is also a characteristic of some “moonlighting” proteins, which under physiological conditions can adopt a variety of distinct folds, each of them adapted to a given function (22).

In summary, there is an overwhelming amount of evidence demonstrating that dynamics is essential for protein function, and we should move to a new paradigm where proteins should not be defined anymore as single structures, but as ensembles of conformations. The challenge of structural biology for the next years is how to capture such dynamics, and how to derive reliable conformational ensembles.

HIGH RESOLUTION STRUCTURAL APPROACHES TO THE STUDY OF PROTEIN DYNAMICS

Structural studies have traditionally frozen protein dynamics, since it facilitates the resolution of the structure. However, even X-ray structures provide indirect information of protein flexibility. For example, parts of the protein that cannot be fitted in the density maps are often flexibly stretches. Likewise, residues that can adopt different conformations often appear with partial occupancy in the X-ray structures. Finally, all X-ray structures in the Protein Data Bank report a rough measure of flexibility: the X-ray B-factors (Figure 1). They are determined from the mean fluctuation of the residue i around its average position at a given temperature (Eq. 1), and are useful to distinguish between rigid and flexible regions: However, the information coded in all these X-ray derived descriptors should not be overestimated. For example, the widely used B-factors account only for harmonic movements, and systematically underestimate the magnitude of protein flexibility (23).

$$B(i) = 8\pi^2\langle(r_i \quad - r_i^0)^2\rangle \qquad (1)$$

where index r stands for the position of a given particle and the index 0 refers to the equilibrium position.

→ Insert Figure 1

Recent works have illustrated the exciting possibility to mix X-Ray structures deposited in structural databases with coevolutionary signals derived from sequence analysis to derive the magnitude of the dynamic conformational space accessible to proteins in large time scale (24,25). The idea has been successfully coupled to simple simulation engines to determine alternative conformations for proteins (24,25).

NMR spectroscopy is much richer than X-ray crystallography as a source of information on protein dynamics (26-30). Current NMR experiments provide information on Nuclear Overhauser Effects (NOE), paramagnetic relaxation enhacement (PRE), three bonds scalar couplings (^{3}J), transhydrogen bond scalar couplings (^{3H}J), chemical shifts (CS) and residual dipolar coupling (RDC), which in all cases are determined as “ensemble properties” (30-32). These parameters can be used then to evaluate the structural diversity of proteins, and even to refine atomistic force fields (33-35). Furthermore, accurate ensemble-based observables can be combined with simulation tools to explore the structural diversity of partially folded, or fully unfolded proteins (36-39).

NMR experiments also provide information of protein dynamics at different time scales, ranging from the second-minutes regime in amide proton exchange saturation experiments to the micro-millisecond scale in relaxation-dispersion measures and to the pico-nanosecond scale from spin relaxation type experiments. Nevertheless, transforming these observables into 3D images of protein dynamics is not trivial and demands a very careful integration of theoretical methods (for a comprehensive review, see ref. 26).

A last source of indirect, but useful information on protein flexibility emerges from the analysis of structural diversity of proteins deposited in the Protein Data Bank (PDB; 40). For example, NMR-derived entries contain several structures compatible with NMR-derived restraints (typically NOEs and ^{3}Js). Even though the structural diversity in NMR-PDB entries may reflect, *a priori,* only technical noise, *in practice,* the dynamic patterns that emerge from NMR-based ensembles correlate quite well with flexibility descriptors derived from other sources (16). In addition, knowledge of structural diversity can also be gained from static X-ray structures in those cases where the protein has been solved several time under different conditions (i.e., apo and holo states, different pH or mutated forms). Again, caution is required in order to infer flexibility patterns from alternative X-ray structures, since structural diversity might emerge from crystallization artefacts or a number of spurious reasons. However, several studies (12-14) have shown that, by aligning the structures of a given family, one can build up a pseudo-trajectory that represents well the intrinsic dynamics of the protein and that the flexibility pattern agrees well with the physical deformation that can be obtained by other methods such as molecular dynamics (MD; Figure 2).

→ Insert Figure 2

THEORETICAL APPROACHES TO PROTEIN DYNAMICS

The intrinsic problems of experimental techniques to provide direct information on protein flexibility have fuelled the development of theoretical approaches to describe protein dynamics. All these techniques differ in the level of resolution used to describe the residues in the protein and their interactions, as well as in the algorithms used to sample the conformational landscape.

THE LEVEL OF RESOLUTION: ATOMISTIC MODELS

Within the atomistic level approach proteins and their environment (mostly solvent) are treated with atomic detail, which means that flexibility is explicitly captured by analysing the movement of every single atom of the structure. The different atomistic models arise from the level of complexity used to describe the energy associated to molecular interactions. We can in principle distinguish three levels of complexity: i) pure quantum mechanical (QM) description, ii) classical (MM) representation, and iii) hybrid quantum/classical (QM/MM) models.

Quantum description. In principle the study of the dynamics of any molecule is the study of the movements along time of its nuclei and electrons. In principle, this can be recovered by solving time-dependent Schrödinger equation (eq.2), something that, in practice, is impossible to do exactly for any molecule of interest. Approximated methods have been then developed for decades to obtain fast estimates of the electronic energy of a given nuclei configuration. For example, a common approximation is to work with the time-independent Schrödinger equation, and to follow the molecular orbital representation of the wavefunction, where such orbitals are expressed as linear combination of atomic orbitals. Methods based on this approach are the traditional *ab initio* or semiempirical (depending on the level of approximations used to solve numerically Schrödinger equation) QM methods.

$$\mathrm{H}\psi = \mathrm{E}\psi \qquad (2)$$

where H is the Hamilton operator, E is the energy and ψ is the molecular wavefunction.

An alternative to methods based in the resolution of the Schrödinger equation is to solve the simpler Kohn-Sham equation (for a review see 41), which expresses molecular energy in terms only of electron density. Approaches based on this equation are named “density functional theory” (DFT) methods (41).

Due to the extreme cost of the calculations, QM methods have been mostly used in the context of isolated molecules in the gas phase, or molecules embedded in continuum solvents (42-43). Application in the study of large molecules, especially in the context of protein dynamics, is more limited and typically implies: i) use of very simple QM descriptions (semiempirical Hamiltonians or low level DFT calculations combined with plane waves (42) or atomic orbitals (43)), and/or ii) the use of “divided and conquer” type of approaches (44-47), were the large system is represented as a series of small interacting subsystems. Some of these models have been coupled to Monte Carlo, or more often to molecular dynamics algorithms (see below) allowing the user to introduce explicitly quantum effects in the calculations.

Classical (MM) methods. Within this approach protein, and often solvent are treated at atomic resolution using classical force fields, which have been parametrized against high quality QM calculations and/or experimental observables obtained for model systems. The force field energy is expressed by simple classical potentials, which at the expense of neglecting explicitly quantum effects, guarantees a large computational efficiency. The first generation of force-fields were created in the late sixties (for an excellent historical overview see 48). Since then, they have not evolved much in their basic formalisms, but have been improved dramatically in terms of parametrization. Current force-fields are extremely refined and if used properly, provide results of a surprisingly good quality. Future developments are expected to arrive from two different directions: i) integration of dynamic information derived from NMR experiments in the parametrization (see above), and ii) extension of the formalism to include "quantum" terms like polarization in a more realistic manner (for a review on classical approaches to introduce polarization see 49). Recent attempts to use polarized force-fields from CHARMM-community are especially remarkable, as well as the efforts to facilitate parametrization of these force-field for new molecules (50-52).

$$E = \sum_{all\ bonds} K_{str}(l - l_0)^2 + \sum_{all\ angles} K_{bend}(\alpha - \alpha_0)^2 + \sum_{all\ torsion\ angles} 0.5\, V_{tor}[1 - cos(n\varphi + \delta)] + \sum_{all\ charges} \frac{Q_i Q_j}{R_{ij}} + \sum_{all\ nonbonded\ pairs} \left[\left(\frac{A_{ij}}{R_{ij}}\right)^{12} - 2\left(\frac{C_{ij}}{R_{ij}}\right)^{6}\right]$$

(3)

where K stands for the stiffness of stretching (str) or bending (bend), l and α stand for bond lengths and angles (the 0 subindex stands for equilibrium values), V_{tor}

represents torsional barrier, φ is the torsion angle, n,and δ are for the periodicity and phase angle of the Fourier term used to represent torsion, Q stands for charges, R represents a non-bonded interatomic distance, and finally A and C are van der Waals parameters characterizing Lennard-Jones interactions between particles.

Hybrid QM/MM methods were created to deal with systems where QM effects are important (for example reactivity happens), but that are too large to be treated entirely at the QM level. QM/MM methods divide the system in two parts: i) a small one that is described at a QM level of theory, and a large one that is represented by means of classical potentials. The two most fruitful developments in the QM/MM field follow two different approaches: i) the empirical valence bond theory (EVB; 53-56), and ii) the molecular orbital self-consistent field (MO-SCF) approach (57-63). QM/MM coupled to molecular dynamics algorithms they have been widely used to study enzymatic reactivity, helping to decipher the molecular mechanisms that allow protein to accelerate so efficiently chemical reactivity, illustrating for example, the fine coupling between conformational dynamics and catalytic function (61,62).

The elegant Warshel's EVB theory (53-56) follows the valence bond framework (and alternative not fully explored to the prevalent molecular orbital theory) adapting it to deal with large macromolecular systems. In EVB the Hamiltonian of the reacting system is described by a limited series of resonance states (defining for example the reactants and products of a reaction). The diagonal elements of the Hamiltonian (H_{ii}) correspond to the energies of the entire system in each resonance state (i), while the off diagonal terms (H_{ij}) are represented by empirically-fitted exponential functions of the distances between reacting atoms. In practice, for the study of reactivity in complex environments (like enzymes) H_{ii} is determined by fractioning the system in a small part ("the reacting system"), whose energy is computed at the QM level or by means of empirical functions fitted to reproduce QM profiles (or experimental data if available), plus classical terms accounting for the internal energy of the environment (for example a protein) and its interaction with the reacting system.

MO-SCF methods approaches the QM/MM problem (57-62) by simply solving time independent Schrödinger equation (eq. 2) for a global Hamiltonian defined as three blocks (eq. 4) : i) the QM part (H_{QM}) for the "reacting system", that is fully represented at the QM level (typically with a low-level Hamiltonian), a classical part (H_{MM}) that represent the intramolecular interaction of the environment (that for a given nuclei configuration does not affect directly the wavefunction of the reacting system), and finally a coupling term ($H_{QM/MM}$) that accounts for the interactions between the reacting system and the environment and that includes an electrostatic term acting as perturbational operator that modified the wavefunction of the reacting system. In the biochemical scenario MO-SCF QM/MM methods are typically coupled to MD algorithms to provide hybrid samplings (typically in the Born-Oppenheimer limit (i.e. electron distribution is SCF relaxed for each nuclei movement), see below).

$$H_{eff} = H_{QM} + H_{MM} + H_{QM/MM} \qquad (4)$$

THE LEVEL OF RESOLUTION: COARSE-GRAINED MODELS

Dynamic representation of large protein systems, especially of protein aggregates, can be extremely expensive when coupled to atomistic models (even when classical Hamiltonians are used), which has led to the development of lower resolution models, where often solvent is not explicitly considered (or it is represented by particles representing clusters of solvent molecules), and where several atoms of the proteins are grouped in a single bead to reduce even more the degrees of freedom of the system (63-68).

Coarse Grained (CG) methods are extremely efficient from a computational point of view, but since atomic detail is lost, they require the use of non-physical statistical potentials (in some cases with a general formalism that resembles physical potentials), which need to be carefully calibrated in order to reproduce the structural and flexibility properties of proteins (for recent ideas on how to build optimized models with low and very low resolution levels see reference 70 and text below).

Go-potential (69,70) is probably the oldest and the simplest of CG potentials. It assumes that the experimental structure of a protein corresponds to the force field energy minimum (i.e. it assumes that the experimental structure shows no frustration of protein-protein interactions), and that native residue-residue contacts are favourable, while non-native contacts are either irrelevant or unfavourable:

$$E = \sum_{i,j} \delta_{i,j}\, \varepsilon_{ij}$$

(5),

where *i* and *j* stands for two residues, $\varepsilon_{ij}$is a stability energy value (constant for all pairs in the uncoloured Go-model, or different in coloured Go-model depending on the nature of the interacting residues), and $\delta_{i,j}$ takes values of -1 if contact *i-j* is native and 0 or +1 otherwise.

Go-potential typically relies on a representation of the protein limited to the alpha carbon atoms, and despite its simplicity it has been very useful in the description of protein flexibility, specially related to folding and often coupled to Monte Carlo sampling algorithms. Recent analysis of very long Anton's atomistic molecular dynamics simulation for a series of fast folding proteins (see below) demonstrated (71) that native contacts define protein folding simulations, providing an unexpected support to simple Go and Go-like force-fields.

Go-potential has been a source of inspiration for other closely related CG potentials. One of them is *Onuchic's functional* (72,73), where a C$_\alpha$ representation of the protein is used, and the energy is determined by adding bonded terms computed for all contacts, as in a normal physical force field, and non-bonded interactions, which are computed separately for native and non-native contacts. The native contacts are represented by combining a repulsive r^{-12} term and an attractive r^{-10} component, whereas the non-native ones are represented only by a repulsive term. Onuchic's potential energy terms (see Eq. 6) were calibrated to reproduce protein structure and dynamics.

$$
\begin{aligned}
E = \sum_{bonds} K_s(r-r_o)^2 &+ \sum_{angles} K_\theta(\theta-\theta_o)^2 \\
&+ \sum_{dihedrals} \left\{K_\varphi^1[1-\cos(\varphi-\varphi_0)] + K_{\square}^3[1-\cos 3(\varphi-\varphi_0)]\right\} \\
&+ \sum_{i,j\,\in\, native} \varepsilon\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right] \\
&+ \sum_{i,j\,\in\, native} \varepsilon\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right] + \sum_{i,j\,\in\, non-native} \varepsilon\left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12}\right]
\end{aligned}
$$

(6),

where bond, angle and dihedral parameters have the same meaning that in physical force-fields (though limited to the description at the C_α level), ε is an energy value, σ_{ij} is the *i-j* distance (r_{ij}) in the native conformation and σ_o is set to 4 Å to avoid close non-native contacts. Onuchic's potential assumes the same principle of minimum frustration in protein structure than pure Go-pure potential, but is much more accurate and flexible. Furthermore, it has proven to be very efficient in simulations of protein folding, and on the description of the conformational landscape of complex proteins from Langevin-Brownian molecular dynamics simulations (72-74).

Other potentials not far from the concept of the Go model are the *elastic network models* (ENM; 75,76; see Figure 3). They are typically used in the context of C_α representations of the proteins, even though higher resolution models have been proposed. As Go-model ENM also assume that the experimental structure is a minimum in the conformational energy landscape, and that the protein reacts harmonically to the perturbation of inter-residue distances:

$$E = \sum_{i,j} \frac{1}{2} \kappa \Gamma_{ij} \left(r_{ij} - r_{ij}^{0} \right)^2$$

(7)

where κ is a force constant, typically the same for all pairs of residues, Γ_{ij} is the Kirchoff topology matrix with elements equal to 1 if the experimental inter-residue distance r_{ij}^{0} is smaller than a cut-off and zero otherwise.

The selection of the cut-off value in Eq. 7 is crucial for a correct description of flexibility. Typically values around 10 Å have been used, but the optimum value can change for proteins of different size. Kovacs et al. (77) proposed an alternative expression to Eq. 7, where the use of the Kirchoff matrix is avoided and instead of annihilating the springs among residues at the cut-off distance, the force constants are scaled down with the sixth power of the inter-residue distance, leading then to:

$$E = C\left(\frac{\alpha}{r_{ij}^{0}}\right)^{6}(r_{ij} \ - r_{ij}^{0})^{2}$$

(8)

where C is an effective force-constant and α is a constant taken as the shortest possible C_{α}-C_{α} distance (3.8 Å).

Recently Orellana et al. (78) refined Kovac's model by using three layers for the definition of the inter-residue force-constants, which were denoted i) very stiff sequence-mediated contacts, ii) space-dependent contacts with stiffness scaled down with a sixth power exponential term, and iii) a cut-off to completely remove irrelevant very distant contacts. The refined method proved to be superior to simpler ENM formalism in reproducing flexibility patterns retrieved from NMR ensembles, structural diversity in PDB, known conformational transitions, and atomistic MD simulations in explicit solvent (78,79; see Figure 3).

→ Insert Figure 3

ENM harmonic potentials can be used in conjunction of Langevin-Brownian dynamics (79), but the standard framework is within normal mode analysis (NMA; see below). Alternatively, ENM potentials can be discretized into harmonic wells to allow us the use of discrete molecular dynamics (dMD; 79,80) sampling algorithms (see below).

A large variety of *pseudo-physical CG potentials* have been suggested. They maintain a formalism that resemble those of atomistic force-fields, and they require only partial information on the three dimensional structure of the protein. Marrink's MARTINI force field (63,81,82) is one of the most used of this family of force-fields. MARTINI maps four heavy atoms with a bead, which is labelled according to its physical characteristics (charge, polarity, hydrogen bond donor/acceptor capabilities, etc). The interactions among beads adopt a quite standard "all atoms-like formalism", adding torsional restraints to maintain native secondary structure. MARTINI force-field (as physical force-fields) can be used in the context of either normal molecular dynamics, or Langevin-Brownian dynamics. The use of Marrink's force-field allows a significant computational efficiency, due not only to the decrease in the number of degrees of freedom, but also to the large masses of the beads, that allow the use of large integration steps. The force field has been largely used to study large conformational transitions, especially in membrane proteins (82-86); for a more detailed description of the force field and its application we address the reader to ref. 82.

Lavery and coworkers (87) have recently tried to alleviate some of the problems of MARTINI force field (mainly the need to force secondary structure, which limits its applicability in the study of conformational changes implying disruption of secondary structure) by developing a more elaborate model, were each side chain is reproduced by one or two beads (apart from the C_α atoms), using a quite detailed definition of the backbone that allows to capture backbone hydrogen bonding. The interactions between beads are represented by a standard bonded "all atoms-like" force field supplemented by a coupling torsion term used to refine the description of torsions that are strongly correlated (like Ψ and Φ). The non-bonded contribution contains four terms: a smooth r^{-8}-r^{-6} term for van der Waals, a standard Coulomb term, with a distance-dependent dielectric constant, a specific hydrogen-bond r^{-12}-r^{-6} term, and a very simple one-body solvation term that depends on the accessibility of every bead. The different parameters in the force field were fitted to reproduce neighbor distribution probabilities in a large set of proteins.

The idea that MD-based CG potentials should be able to reproduce atomistic MD trajectories was present in the development of many other CG force fields. For example, Kmiecik and coworkers used MoDEL database (88,89) to calibrate and validate the CABS force-field (90), which uses a simplified version of the protein (one bead at C_α, another in the middle of the peptide bond and one or more beads for the side chains). On similar lines, Voth and coworkers (91) used MD simulations at high temperature of small proteins and peptides to calibrate a CG potential with different levels of resolution for residues (from 1 to 5 beads), a sophisticated bonded term, and a non-bonded expansion that resembles the terms used in the MARTINI force field. Schulten's group has also worked in similar lines developing a very simple force field (two beads per residue), and later proposed an even lower resolution one, where beads were placed not in real residues, but in points selected to reproduce protein shape (92-94). The parameters required in these force fields were fitted case by case to reproduce atomistic CHARMM simulations for the same protein, which reduces universality and transferability in the force-field, but increases very significantly its accuracy for the system of interest.

It is impossible to discuss, or even cite here all the different CG models and force field developed in the last years. Hence, we limit ourselves to conclude this section by citing Scheraga's UNRES force field due to its impact in folding studies of proteins

(95,96). In this case each residue is represented with two beads, one located in the center of the peptide bond and the other representing the side chain. UNRES force field is formally complex and was calibrated using atomistic MD data and three-dimensional structures of folded protein deposited in PDB (95,96).

THE SAMPLING METHODS: MOLECULAR DYNAMICS

Irrespectively of the Hamiltonian used to represent molecular interactions, representation of flexibility requires a sampling method to generate ensembles of structures for a given environmental conditions, or in some cases, to simulate a transition path.

Quantum molecular dynamics. Molecular dynamics (MD) techniques derive ensembles by integrating equations of motion. If the Hamiltonian is classical, Netwon's physics is valid and the integration of equations of motions is simple. However, if electron degrees of freedom are considered, and the Hamiltonian is fully or totally quantum, integration of equations of motions is much more complex, since quantum physics needs to be considered. Quantum molecular dynamics (QMD) can be done within two major frameworks: i) Born-Oppenheimer molecular dynamics (BOMD) and ii) Carr-Parrinello MD (CPMD).

BOMD simplify the problem of integrating electron degrees of freedom by considering that electrons has reached equilibrium for each nuclei movement (i.e. the system follows the Born-Oppenheimer regime), that means that full SCF convergence of electronic wavefunction is performed for each nuclei movement. BOMD is extremely expensive, since for each femtosecond (the usual integration step for nuclei movements) one full SCF optimization has to be done. So, BOMD has been used mostly within the QM/MM framework (57-62). Recently divide-and-conquer type of approaches such as Gao's Xpol approach (44,45,47) have made possible to perform multi-picosecond full QM BOMD simulations of small proteins in realistic physiological conditions, using simple semiempirical Hamiltonians to represent protein and solvent. With new generation of computers, faster QM methods, and more efficient parallelization schemes it will be possible to extend trajectories to the multi-nanosecond time scale, opening interesting possibilities in the study of proteins where quantum effects are present across the entire structure.

Carr-Parrinello molecular dynamics (CPMD; 42) is probably the most popular approach to perform full QM MD calculations. This elegant technique uses DFT planewaves to describe electron distribution and consider explicitly electron degrees of freedom in integrating system's equations of motion (i.e. no Born-Openheimer approach is followed). This is made by considering an extended Lagrangian, which includes energy terms depending on time-dependent electron distributions. Integration is then made at sub-femtosecond scale to allow the relaxation of electron degrees of freedom. CPMD has had a dramatic impact in material sciences, and applications in proteins are increasing, especially when embedded in a general QM/MM treatment of the system (97,98).

Classical molecular dynamics is probably the better known and widespread method for the study of protein flexibility. Within this approach, both the protein and the solvent environment are treated at atomic resolution level using classical force fields, which have been parametrized against high quality quantum mechanical calculations and/or experimental observables obtained for model systems (see above). The force field energy (expressed by simple potentials) can be differentiated to obtain forces on individual atoms, and the corresponding accelerations are then numerically integrated (typically in the femtosecond range) to obtain new velocities and positions for the atoms. This process leads to a trajectory of the protein along time, which explicitly contains all information on the protein flexibility under the simulation conditions.

First MD simulations of proteins were performed in the late seventies (99,100) and the formalism was refined in the eighties (for a nice historical perspective see 101). Since then, the technique has continued evolved, gaining accuracy and predictive power, and becoming a widely accepted tool in hundreds of laboratories. Current protocols allow the representation of a variety of biological systems in realistic environments for simulation periods typically in the multi-nanosecond to microsecond time scale. Popularization of the technique has been possible thanks to the advances in MD algorithms implemented in codes such as GROMACS, AMBER, CHARMM and NAMD, among others, which thanks to the support of enthusiastic people offer the community in a free (or nearly-free) basis extraordinary pieces of computing engineering. As an example, GROMACS (http://www.gromacs.org/About_Gromacs) is the result of 482 person-year work with an estimate cost of more than € 26 million (by November 2013). Without this altruistic effort, the MD field would have been very different to what is at present.

Current progresses in atomistic MD simulations involve the extension of i) the size of the system and ii) the simulation length, iii) parallel-ensemble simulations, and iv) biased MD simulations.

Extension of simulated systems. Last generation MD software (see for example 102,103) achieves a good degree of parallelization, allowing a quite efficient use of last generation supercomputers. Thus, the possibility to distribute a single trajectory among hundreds or thousands of processors has enabled to perform huge simulations (104-106), such as the impressive study of the HIV virus capsid, a system with many millions of atoms recently simulated by Schulten's group (104). Unfortunately, even the best parallelized codes have generally problems to use efficiently more than a few thousands cores, whereas the top supercomputers have millions of cores (for example, Tianhe-2, the first supercomputer in November 2013 Top500 list, has more than 3 million cores; using current technology around 10^8 cores would be needed for an Exaflop computer). It is difficult to believe that MD codes could be further improved as to allow us an efficient use of millions of cores in a single biologically relevant MD trajectory, but massively parallel computers significantly enhance our capabilities to study bigger, more realistic systems.

Extension of the trajectory length. Current state-of-the art simulations are still too short to allow a direct comparison with experimental data, and it is always unclear whether or not a trajectory is long enough as to achieve a proper sampling of the conformational space. Modest parallelization (scaling to 128-512 cores) helps to increase the accessible time-scale of most simulations, but unfortunately, for most systems of interest (solvated proteins containing 10^4-10^5 atoms), no dramatic improvements in simulation speed is obtained when more cores are used. This has encouraged different groups toward: *i) the design MD-specific hardware and ii) the development of GPU –adapted MD codes*.

Shaw and co-workers have built the Anton computer specifically for MD simulations (107), leading to an enormous gain in the simulation length. Thanks to this specific purpose-designed computer, they published the first millisecond atomistic MD simulation of a folded protein (108), and successfully folded a series of small proteins in the sub-millisecond time scale (109), thus providing a basis for significant advances in the understanding of the kinetics and thermodynamics of protein folding (109-112). More recently, Anton computers have also been used to study processes such as ligand binding, allosteric transitions in membrane receptors and the mechanism of ion channels (113-117). The possibility of running these extremely long simulations is opening new fields, which will be fully explored when computer architectures, such as Anton, will become available to the entire community.

GPUs are highly specialized architectures, which offer a cheap alternative to accelerate MD simulations in cases where poor parallelism does not justify adding more processors to the calculation. In the last years GPU-specific codes such as AceMD (118) have been developed, and GPU-optimized versions of popular codes such as GROMACS (102), AMBER (119,120) or NAMD (103) have been created. The acceleration of calculations by GPUs is not as impressive as that achieved with Anton, but GPU-based cluster are accessible to the entire community, allowing many groups the access to the multi-microsecond regime for small and medium-sized proteins (120,121). Furthermore, GPU-based hardware allows an easy integration of ensemble-based simulations, including those performed in a distributed manner (120-123). As the increase in GPU performance and the parallelization in CPU/GPU clusters continue, we can expect an even larger impact of GPUs in the molecular dynamics field.

Parallel ensemble MD simulations. As discussed above, extension of the simulation time of individual trajectories is limited by the difficulty for an efficient parallelization of systems comprising around 10^5 atoms (i.e., a the typical size of a solvated protein). Fortunately, in the *ergodic* limit, there is no need to have an infinitely long trajectory, since the same information could be derived from an infinite number of short simulations. This has raised the concept of "ensemble simulations", where a very large number of small trajectories is collected and processed to reconstruct long time scale processes. The ensemble approach allows a very efficient use of computer resources, since individual simulations are computed with a reduced number of processors, which guarantees full parallelism, adapting well to current high performance computers.

Pande and co-workers popularized the ensemble MD procedure showing that quantitative information on the folding of small proteins may be derived by integrating the information of thousands of trajectories, each of them shorter than the expected length of the folding process (124,125). The same group developed a worldwide initiative named "folding_at_home" (www.folding.standford.edu/home), which allows them to use a gigantic number of (otherwise dormant) CPUs provided by volunteers around the world to advance in the knowledge of mechanistic aspects of protein folding, intrinsic disordered proteins or drug-binding (124-129). As a recent impressive example, Pande's group has made use of Google's Exacyle cloud-computing platform to simulate an aggregated time of two milliseconds of β2AR, which were combined by

Markov's models to obtain a clear picture of activation pathways of this important pharmaceutical target (127).

How individual trajectories are combined to obtain a Boltzmann's sampling of the protein conformational space is still a major issue, and research efforts have led to a variety of methods. The replica exchange simulation (RExMD; also named parallel tempering MD (130,131) uses multiple trajectories at different temperatures that are interchanged every number of integration steps based on a Metropolis Monte Carlo acceptance algorithm (Figure 4) with the acceptance probability being defined as:

$$p = min\left\{1, e^{\Delta E_{ij}\left(\frac{1}{kT_i}-\frac{1}{kT_j}\right)}\right\} \qquad (9)$$

where i and j stand for two protein configurations obtained at temperatures T_i and T_j. and ΔE_{ij} denotes the potential energy difference between configurations i and j.

It can be proved that for an infinite range of closely spaced temperatures, RExMD provides a Boltzmann's ensemble at a given reference temperature. As the replicas run independently until Metropolis Monte-Carlo attempts to interchange them, RExMD simulations are well suited for massive parallel computers. Since two layers of parallelization can be performed: one for individual replicas and another one for the temperatures, RExMD parallizes well in big clusters and supercomputers. However, RExMD is, in principle, not so efficient in the context of distributed computing, where asynchrony in the replicas can diminish severely the overall performance. However, recent efforts by Pande and others (see for example ref. 132, 133) have yielded modified algorithms that reduce synchrony problems, allowing the use of distributed infrastructures for these simulations. Note that the RExMD formalism also enables the use of Hamiltonian perturbations to improve sampling in a similar manner than the temperature perturbations (134), which extend the range of applicability of the general replica exchange idea.

→ Insert Figure 4

Other alternatives to RExMD have been proposed to process and integrate MD ensembles. For example, different groups have developed algorithms to integrate massive number of trajectories collected for the same system at the same temperature, but starting from different conformations and velocities. The ensembles are then processed through rigorous statistical mechanics algorithms, such as Markov State Models (MSM; 135-138), to define the entire conformational space. The basic idea of MSM is to capture the equilibrium population of two states A and B by looking at the equilibrated transition probability between them. Implementation of this idea is difficult if both states are too distant, since individual MD simulations are unable to detect enough transitions as to guarantee convergence of the transition probability ($P_{A|B}$). To solve this problem MSMs define a path of transitions through other intermediate states $P_{A|A'}$, $P_{A'|A''}$,…, $P_{An'|B}$, linking A and B. If the total collected ensemble is good enough, not only the relative population of the two conformational states is retrieved, but also the major transition pathways and their associated kinetics are obtained. In practice (139,140) MSM/MD calculations require thousands of individual independent trajectories, with after equilibration are processed by clustering methods to define populated microstates, which are later re-clustered in larger states from which transition probabilities are determined. The method is sensitive to limited sampling and to the definition of microstates, and often the transition path between states A and B is either interrupted, a given transition is not converged, or a state becomes isolated from the rest. In these cases iterative algorithms are used to reinforce sampling in poorly explored regions of the conformational space. MSM/MD is gaining relevance in the field and has been largely and successfully used to study slow processes (126-129,135-140) including protein folding and ligand binding and fits very well in supercomputer architecture, which suggest that its use will be extended in the future.

Parallel MD simulations have been used in conjunction with experimental ensemble information in order to characterize the structural transitions in complex processes. Thus, global ensembles are obtained by combining a slow time scale, which is covered by the experiment, and a fast one, which is by MD simulations For example, Grübmuller and coworkers (141) have combined in an elegant way electron microscopy and X-Ray data with atomistic MD simulations to examine the ribosomal elongation mechanism.. Following similar ideas Candotti et al. (40) characterized the unfolded state of Ubiquitin by combining MD trajectories (short time-scale movements) with NMR-derived ensembles (large time-scale movements), de Groot and coworkers mixed electron crystallographic data with MD simulations in packed environments to refine structure and dynamics of lipid-protein complexes (142), and again Grubmüller has recently published how to integrate FRET experiments in atomistic MD simulations (143-144).

Biased MD simulations. A series of techniques are available to bias a MD simulation to favour a given transition. *Umbrella sampling* (US) is probably the oldest, and still one of the most popular ones to bias a trajectory along a potential transition pathwar.

To this end, US defines a “transition coordinate” (γ) that drives the transition from state A to B ($\gamma_A \rightarrow \gamma_B$), forcing the system to move in small steps (windows) along it ($\gamma_A \rightarrow \gamma_{A+d\lambda},\ldots, \gamma_{B-d\lambda} \rightarrow \gamma_B$, $d\lambda$ being a small increment in the reaction coordinate from state A to B). A biasing potential (the umbrella potential, typically an harmonic function) is added to guarantee sampling around the desired window. The free energy associated to the transition is then determined from the probability function P(γ)* obtained from the MD simulation performed with a Hamiltonian defined by the standard force field supplemented with the umbrella potential (U(γ)):

$$W(\gamma) = -\mathrm{k}_b\mathrm{T}\ln\mathrm{P}(\gamma)^* - U(\gamma) - \mathrm{k}_b\mathrm{T}\ln\langle e^{-\frac{U(\gamma)}{k_bT}}\rangle_* \qquad (10)$$

where the index * refers to the MD ensemble obtained with the biasing potential (for a more detailed analysis see refs. 144,145).

Different variants of the US technique have been successfully used to study a number of conformational movements in proteins. However, a common problem to US-based methods is that finding a privileged “transition coordinate” to be modified smoothly and reversibly between states A to B might not trivial for collective motions in proteins. This has fuelled the development of many alternative strategies. We will limit ourselves to mention a few of these methods.

Targeted MD enforces the transition A→B by introducing restraints that force the system to reduce slowly the RMSd to a target structure (146). Alternatively, in *steered MD* (sMD) a steering force is added to the Hamiltonian to pull the system towards a defined reaction coordinate (Eq. 11). sMD has been largely used to simulate protein unfolding, ligand unbinding or to mimic *in silico* atomic force-microscopy experiments (147-150). In a very recent work, Hummer’s group (151) has shown how by combining steered MD and metadynamics (see below) is possible to analyse in detail the mechanism of interconvertion of chemical energy and work by F1-ATPase (one of the most important motors in the cell)

$$F = K(\gamma_0 + vt - \gamma_0) \qquad (11)$$

where K is the stiffness of the biasing potential $U = K(\gamma - \gamma_0)^2$, v is the velocity at which the system is pulled along the reaction coordinate γ, and t is the time.

Integration of the pulling force provides the associated "irreversible" work (W), from which the free energy profile can be derived by using Jarzynski's equality (Eq. 12).

$$e^{-\Delta G/k_bT} = \langle e^{-W/k_bT} \rangle \qquad (12)$$

where the brackets in the right-hand part of the equation mean that the "pulling" experiment needs to be repeated many times to achieve a converged free energy (for details see ref. 150).

A variety of alternative methods denoted as *Maxwell-Demon MD*, *dynamic importance sampling* (DIM), or *soft-racketing* (152,153) use a different approach to bias the trajectory. Instead of adding a biasing potential (or force), they bias the trajectory by selecting, and eventually cloning those snapshots that spontaneously approach the target structure. In practice the method implies parallel MD runs followed by Metropolis Monte Carlo selection steps, which select (and clone) snapshots based on how well they approach the target conformation.

Metadynamics is another popular alternative to facilitate the sampling of conformational transitions. The technique, developed by Parrinello and coworkers (154) from previous models by van Gunsteren's group (155), was originally intended to help the trajectory to escape from a local minimum. This is achieved by adding every certain numbers of steps a biasing term to the Hamiltonian that penalizes the system to re-visit regions previously sampled (Eq. 13).

$$V_{\beta}(\gamma,t)=\sum_{t=0,n\tau} we^{-\frac{\gamma-\gamma_t}{2\sigma^2}}$$

(13)

where γ is here a collective variable defining a given state, *w* and σ are the height and width of the Gaussians which are added at time τ.

As shown by Parrinello and coworkers (154-157), the free energy associated to the scape from the minimum can be determined as the added potential required to free the system from the original basin ($V_{\beta}(\gamma)=-F(\gamma)$). This definition, even in principle exact, suffers from convergence problems, since it is not trivial to decide when to stop the simulation, and if this is not done more and more useless Gaussian are added, increasing in an artifactual manner the free energy. To alleviate this problem, in the well-tempered variant of metadynamics the Gaussian height (*w*) is also history-dependent, forcing a decrease in Gaussian deposition as simulation time advances, improving then the convergence in the free energy estimates (156). The generalization of *metadynamics* to study any given transition is simple, provided collective variables (γ in eq. 8) describing a reasonable path are defined. A large number of different collective variables are accessible (157), allowing the method to be used for allosteric motions in proteins, ligand binding and many other slow processes (151,157-161).

Activated MD (162,163) is a biasing technique that is applied after there is a first approximation to the reaction pathway, from which a first guess of the placement of the free energy barrier (this can be obtained by standard umbrella, steered MD or any other similar method). Once the maximum energy point in the conformational pathway is known, many trajectories starting for this conformation with random velocities are lunched and followed to rebuild a refined conformational pathway and the associated kinetics, now without the bias imposed by the distinguished coordinate. Conceptually close to activate MD are a myriad of other techniques that refine a first guess of a transition pathway by moving in the potential energy (or free energy) hyperspace. Two examples of these techniques are the Karplus's conjugate peak refinement method (164,165) and the dynamic string method derived by Hummer and others (166). Transition pathway sampling technique (TPST; 167,168) can be seen also as member of this family of methods. In TPST a first guess of a potential pathway is used as seed for a Monte Carlo procedure where new pathways are generated and selected by Metropolis test based on how efficiently they link reactants and products (defined by some order variable). The final output of TPST is an ensemble of trajectories linking as efficiently as possible the two states of interest.

THE SAMPLING METHODS: BROWNIAN/LANGEVIN MOLECULAR DYNAMICS

When part of the entire system is not simulated at the discrete level, but by means of a continuum that, for example, represents friction effects or general environmental effects, the basic MD algorithm needs to be modified by introducing Langevin and Brownian corrections, which allow the introduction of environment effects by coupling the particular system to an external bath that affects the trajectory in two different ways: i) by dissipating energy as heat (mimicking collision of the system with, for example, solvent molecules), and ii) by generating random forces that add energy into the system in a Brownian manner, as noted in Eq 14,

$$m_i\vec{a}_i = \vec{F}_i - \gamma\vec{v}_i + \vec{R}_i(t) \qquad (14)$$

where the first term in the right side of the equation stands for the forces (obtained from derivation of the force-field potential at the position of particle *i*, the second term is a Langevin term for a fluid with friction coefficient γ, and the last term accounts for a random force which is typically considered Gaussian with zero mean for large simulation times (for a detailed explanation of the technique we address the reader to refs. 169-171).

Brownian/Langevin dynamic algorithm is implemented in many MD codes, which allows a direct use of this technique with many continuum CG models. The method is especially powerful when combined with CG representations of the protein Hamiltonian and implicit representation of solvent, since the reduction in degrees of freedom, the simplicity of the CG Hamiltonian, and the heavy masses of the particles considered in CG simulations allow efficient samplings. Very recently, an alternative use for the technique has been proposed to deal flexibility effects linked to ligand binding (172,173). The new method uses a hybrid Hamiltonian like that described in Eq. 14, with the essential deformation modes (eigenvectors ν and eigenvalues λ) being determined from essential dynamics analysis of a previous MD simulation (174).

THE SAMPLING METHODS: NORMAL MODE ANALYSIS (NMA)

NMA is probably the simplest of all sampling methods currently used to describe protein dynamics. It can be used with both atomistic (with implicit solvent representations) or CG representations of the molecule, and with any kind of force-fields. However, in practice the vast majority of NMA studies with proteins are carried out considering a CG representation of the protein, typically coupled with ENM Hamiltonians (see above). NMA methods assume than the experimental (or in general the reference) structure of the protein corresponds to the free energy minimum and that changes in energy related to geometrical perturbations of such structure can be expressed as a Taylor series:

$$V(r) = V(r^0) + \sum_i \left(\frac{\delta V}{\delta r_i}\right)_0 (r_i - r_i^0) + \frac{1}{2}\sum_{i,j}\left(\frac{\delta^2 V}{\delta r_i \delta r_j}\right)_0 (r_i - r_i^0)(r_j - r_j^0) + \cdots$$

(15),

where both the (relative) energy and the first derivative are equal to zero at the reference point (r^0). If one ignores higher order contribution, the preceding expression can be rewritten as in Eq. 13, which indicates that the deformation energy can be easily computed from the Hessian.

$$V(r) = \frac{1}{2}\sum_{i\neq j}\left(\frac{\delta^2 V}{\delta r_i \delta r_j}\right)_0 (r_i - r_i^0)(r_j - r_j^0)$$

(16)

ENM-NMA methods (77-79,175-186) are simple and provide reliable estimates of X-Ray B-factors, the pattern of flexibility reported by NMR ensembles and even that derived from atomistic MD simulations (16,77-79,175-177). Furthermore, despite the simplicity of the model, ENM-NMA deformation modes capture well biologically relevant conformational transitions (9-10,15,16,179,180).

Recent advances in ENM-NMA have been focused on the refinement of the Hamiltonian (79), the use of internal coordinates instead of Cartesian ones to define protein movements (181-183), and the introduction of anharmonicity effects (184). Large effort has also been made in the development of friendly tools that facilitate the use of the technique to non-experts (76,79,185-187). Finally, attention has also been paid to the use of ENM-NMA information to determine potential transition pathways between alternative conformations of proteins (states A and B). Here the common idea of the different methods is to activate movements of the protein along the normal modes of state A in order to approach state B and *viceversa*. The procedure adopted for exchanging the two minima potentials (for A and B) has been subject to different strategies (188-191), and multi-reference approaches where normal modes are re-computed along a seed pathway (192), or where additional information is implemented (17,193) in the searching algorithm (see below) have been suggested.

THE SAMPLING METHODS: MONTE CARLO

Monte Carlo (MC) is an old lgorithm for generation of structural ensembles that still widely used. Within the traditional Metropolis implementation, the method generates potential random movements on a given structure accepting or rejecting them based on the relative energy of the previous (seed) structure. For an infinite number of attempts the method guarantees that a Boltzmann's ensemble is collected, but in practice the method is efficient only when a suitable set of sampling variables are used. For example, MC is quite inefficient when sampling protein movements in Cartesian space due to the high rejection rate, but it is very efficient to sample, for example to sample side chain movements (194,195) or ultra-simplified descriptions of proteins (196). The use of MC algorithms to refine binding modes in the context of flexible proteins has been crucial in the design of ultra-affinity drugs (197,198).

Recently (177), MC algorithms have been coupled with ENM-NMA methods to sample conformational changes along (in part) normal modes, using as potential energy function an effective Hamiltonian:

$$V(R) = \frac{1}{2}\sum_{i=1}^{M} \frac{k_B T}{\lambda_i} (\Delta v_i)^2 + V(X)$$

(17),

where *i* stands for one of the *M* important normal modes (typically those explained most of protein variance), λ_i is the eigenvalue associated to the mode i (in distance^{-2} units), $\Delta v_i$is the projection of the sampled movement on the eigenvalue associated to normal mode *i*. and finally V(X) represents changes in energy terms computed in the Cartesian space. A similar implementation in the context of Brownian/Langevin dynamics will be discussed below. Note also that the method can be used when instead of normal modes, deformation modes derived from essential dynamics are used (see above).

THE SAMPLING METHODS: DISCRETE MOLECULAR DYNAMICS

Discrete molecular dynamics (dMD) defines the Hamiltonian as a series of square potentials (see above) and accordingly, particles are expected to move in the ballistic regime, without changes in velocities until collision in the boundary of the square well happens, where new velocities are computed by imposing the maintenance of energy and momentum (see Figure 5):

$$m_i\overrightarrow{u_i} + m_i\overrightarrow{u_j} = m_i\overrightarrow{u_i'} + m_i\overrightarrow{u_j'} \qquad (18)$$

$$m_iu_i^2 + m_iu_j^2 = m_iu_i'^2 + m_iu_j'^2 + \Delta V \qquad (19)$$

where *i* and *j* stands for the two particles colliding, the index ‘ stands for the situation after the collision, *u* for the component of the velocity (*v*) on the direction of the collision (particles are assumed spherical), and *ΔV* is the height of the step in the inter-particle potential.

If the particle does not collide, its position is simply computed as:

$$\vec{r_i}(t + \Delta t) = \vec{r_i}(t) + \vec{v_i}(t) \cdot \Delta t \qquad (20)$$

with integration step $\Delta t < t_c$, which denotes the shortest collision time computed from the non-imaginary solution of:

$$t_c = \frac{-b_{ik} \pm \sqrt{b_{ik}^2 - v_{ik}^2(r_{ik}^2 - d^2)}}{v_{ik}^2} \qquad (21)$$

where k is the first particle colliding with particle i, $b_{ik} = \overrightarrow{r_{ik}} \cdot \overrightarrow{v_{ik}}$ and *d* is the distance corresponding to the wall of the square well.

→ Insert Figure 5

Within the dMD framework the particles move from collision to collision, without need to compute forces every few femtoseconds as in MD or Brownian/Langevin dynamics. This implies a large computational efficiency in sampling, especially in processes with slow dynamics, such as diffusion or folding (80,199-205). The method has been also very powerful to trace complex conformational transitions in proteins (17,193), especially when coupled to information-based biasing techniques (see above). Dokholyan and coworkers (199,200) have made large efforts to develop accurate force fields to study a variety of macromolecular systems and are pushing the technique for docking experiments, where binding and protein dynamics are explicitly coupled.

CONCLUSIONS

Proteins are flexible entities, and we should avoid looking to proteins as static macromolecules, carefully designed by evolution to show well defined structures, that under native conditions are kept rigid by a myriad of interactions. Protein dynamics is crucial for function and it should be explicitly captured. During the writing of this review we celebrate the Nobel Prize of Chemistry to the pioneering work of Karplus, Levitt and Warshel, which opened the possibility to study explicitly protein dynamics. Several decades after this pioneering work, we have a plethora of efficient theoretical methods that implemented in last generation computers allows a representation of unprecedented quality of protein dynamics.

Predicting the evolution of a mature field is always difficult. Surely, we are going to see simulation of larger systems, of multiple complexes, and of proteins in crowded environment approaching to real cellular conditions. Simulations are going to cover longer time scales, approaching to the real biological dominant time scales (milliseconds to seconds) making it possible to capture slow transition and diffusive processes linked to, for example, protein binding. The speed at which this evolution happens will be surely linked to improvement in hardware and in software, forcing a close relationship between computer scientists and computational chemists.

Force-fields will be improved, and surely multi-body effects, such as polarization or charge transfer (see discussion above) will be sooner or later incorporated in high-level calculations. New sampling algorithms will appear to facilitate the analysis of slow transitions. However, where I envision more disruptive changes is from a more agressive integration of simulation techniques with experimental data. Frontiers between simulations and experiments will become fuzzy, if not simply disappear.

ACKNOWLEDGMENTS

The author is indebted to Prof. F.J.Luque for many helpful discussions and for a careful reading of this manuscript. Thanks to Drs. Laura Orellana, Agustí Emperador and Marco d'Abramo for help in preparing several of the Figures, and Pedro Sfriso for pointing me to some exciting references. Our work has been supported by Spanish MINECO (BIO2012-32868), the European Scalalife Project, the Spanish National Institute of Bioinformatics (INB) and the European Research Council (ERC). The author is an ICREA-Academia fellow.

REFERENCES

1. Eisenmesser,E.Z.; Millet,O., Labeikovsky,W.; Korzhev,D.M.; Wolf-Watx,M.; Bosco,D.A.; Skalicky,J.J.; Kay,L.E.; Kern,D., *Nature* , 2005, **438**, 117.
2. Wolf-Warz,M.; Thai,V.; Henzler-Wildman,K.A.; Hadjipavlou,G.; Eisenmesser,E.Z. Kern,D. *Nature Struct. Mol. Biol.,* 2004, **11**, 945.
3. Henzler-Wildman,K.A.; Thai,V.; Lei,M.; Ott,M.; Wolf-Watz,M.; Fenn,T.; Pozharski,E.; Wilson,M.A.; Petsko,G.A.; Karplus,M.; Hubner,C.G.; Kern,D. *Nature*, 2007, **450**, 838.
4. Ma,J.; Karplus,M. *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 8502.
5. Karplus,M.; Kuriyan,J. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 6679.
6. KuhlanB.; Baker,D. *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 10363.
7. Cozzini,P.; Kellogg,G.E.; Spyrakis,F.; Abraham,D.J.; Costantino.G.; Emerson,A.; Fanelli,F.; Gohlke,H.; KuhnA.L.; Morris,G.M.; Orozco,M.; Pertinhez,T.A.; Rizzi,M.; Sotriffer,C.A. *J.Med.Chem.,* 2008, **51**, 6237.
8. Hensen,U.; Meyer,T.; Haas,J.; Rex,R.; Vriend,G.; Grubmüller,H. *PLOS One*, 2012, **7**, e33931.
9. Bahar,L.; Chennubhotla,C.; Tobi,D. *Curr.Opin.Struct.Biol.,* 2007, 17, 633.
10. Dobins,S.E.; Lesk,V.L.; Sternberg,M.J.E. *Proc.Natl.Acad.Sci.USA.,* 2008, **108**, 10390
11. Falke,J.J. *Science*, 2002, **295**, 1480.
12. Velazquez-Muriel,J.A.; Rueda,M.; Cuesta,L.; Pascual-Montano,A.; Orozco,M.; Carazo,J.M. *BMC Struct.Biol.,* 2009, **9**, 6.
13. Leo-Macias,A.; Lopez-Romeo,P.; Lupyan,D.; Zerbino,D.; Ortiz,A.R. *Biophys.J.,* 2005, **88**, 1291.

14. Micheletti,C. *Phys.Life.Rev.,* 2013, **10**, 1.
15. Stein,A.; Rueda,M.; Panjkovich,A.; Orozco,M.; Aloy,P. *Structure* 2011, **19**, 881.
16. Orellana,L.; Rueda,M.; Ferrer-Costa,C.; López-Blanco,J.R.; Chacón,P.; Orozco,M. *Journal of Chemical Theory and Computation*, 2010, **6**, 2910.
17. Sfriso,P.; Hospital,A.; Emperador,A.; Orozco,M. *Bioinformatics*, 2013, **16**, 1980.
18. Dunker,A.K.; Obradovic,Z.; Romero,P.; Garner,E.C.; Brown,C.J. *Genome Informatics,* 2000, **11**, 161.
19. Dunker,A.K.; Silman,I.; Unversky,V.N.; Sussman,J.L. Curr.Opin.Struct.Biol., 2008, 18, 756.
20. Iakoucheva,L.M.; Brown,C.J.; Lawson,J.D.; Obradovicc,Z.; Dunker,A.C. *J.Mol.Biol.,* 2002, **323**, 573.
21. Dyson,H.J.; Wright,PO.E. *Nat.Rev.Mol.Cell.Biol.,* 2005, **6**, 197.
22. Kjaergaard,M.; Teilum,K.; Poulsen,F.M. *Proc. Natl. Acad. Sci. USA,* 2010, **107**, 12535-1540.
23. Kuzmanic,A.; Pannu,N.S.; Zagrovic. *Nature Comm.,* 2014, Electronic Version available. DOI: 10.1038/ncomms4220.
24. Morcos,F.; Jana,B.; Hwa,T.; Onuchic,J.N. *Proc.Natl.Acad.Sci.USA*., 2013, **110**, 20533-38.
25. Hops,T.A.; Colwell,L.J.; Sheridan,R.; Rost,B.; Sander,C.; Marks,D.S. *Cell*, **149**, 1607-1621.
26. Esteban-Martín,S.; Fenwick,R.B.; Salvatella,X. *WIREs Comput. Mol. Sci.* 2012, **2**, 466-478.
27. Mittermaier,A.; Kay,L.E. *Science*, 2006, **312**, 224-228.
28. Mulder,F.A.; Mittermaier,A.; Hon,B.; Dahlquist,F.W.; Kay,L.F. *Nat. Struct. Biol.,* 2001, **8**, 932-935.
29. Markwick,P.R.L.; Malliavin,T.; Nilges,M. *PLoS Comput Biol.,* 2008, **4**, e1000168.
30. Lindorff-Larsen,K.; Best,R.B.; DePristo,M.A.; Dobson,C.M.; Vendruscolo,M., *Nature*, 2005, **433**, 128-132.
31. Showalter,S.A.; Brüschweiler,R. *J.Am.Chem. Soc.,* 2007, **129**, 4158-4159.
32. Fenwick,R.B.; Esteban-Martín,S.; Richter,B.; Lee,D.; Walter,K.F.A.; Milavanovic,D.; Becker,S.; Laomek,N.A.; Griesinger,C.; Salvatella,X. *J.Am.Chem.Soc.,* 2011, **133**, 10336-10339.
33. Lipari,L.; Szabo,A.; Levy,R.M. *Nature*, 1982, **300**, 197-198.
34. Wickstrom,L.; Okur,A.; Simmerling,C. *Biophys J.,* 2009, **97**, 853-856.
35. Li,D.W.; Brüschweiler,R. *Angew.Chem-Intl.Ed.Eng.,* 2010, **49**, 6778-6781.
36. Bernadó,P., Blanchard,L.; Timmins,P.; Marion,D.; Ruigrok,R.W.H.; Blackledge,M.A. *Proc. Natl. Acad. Sci. USA,* 2005, **102**, 17002-17007.
37. Bernado P, Blackledge M. *Biophys J.,* 2009, 97, 2839-2845.
38. Esteban-Martin S, Fenwick RB, Salvatella X. *J.Am.Chem.Soc.,* 2010, 132, 4626-4632.
39. Candotti,M.; Esteban-Martin,S.; Savaltella,X.; Orozco,M. *Proc. Natl. Acad.Sci. USA.,* 2013, 110, 5933-5938.
40. Berman,H.M.; Westbrook,J.; Feng,X.; Gilliland,G.; Bhat,T.N.; Weissig,H., et al., *Nucleic Acids Res.,* 2000, **28**, 235-242.

41. Parr,R.; Yang,W. Density-Functional Theory of Atoms and Molecules. , Springer. Berlin ISBN 3-540-51993-51999.
42. Car,R.; Parrinello,M. Phys.Rev.Lett., 1985, **55**, 2471-74
43. Soler,J.M.; Artacho,E.; Gale,J.D.; Garcia,A.; Junquera,J.; Ordejon,P.; Sanchez-Portal,D. J.Phys.Condens.Matt., 2002, **14**, 2745-2779
44. Song.L.; Hahn,J.; Lin,Y.L.; Xie,W.; Gao,J. , *J.Phys.Chem.,* A, 2009, **113**, 11656-64.
45. Xie,W.; Orozco,M.; Truhlar,D.; Gao,J. *J.Chem.Theor.Comput.*, 2009, **5**, 459-467.
46. Kitaura,K.; Sawai,T.; Asada,T.; Nakano,T.; Uebayasi,M. *Chem.Phys.Lett*., 1999, **312**, 319-324
47. Wang,Y.; Sosa,C.P.; Cembran,A.; Truhlar,D.G.; Gao,J. *J.Phys.Chem.B.,* 2012, **116**, 6781-6788.
48. Levitt,M. *Nature Structural Biology,* 2001, **8**, 392-393.
49. Luque,F.J.; Dehez,F.; Chipot,C.; Orozco,M. *WIRES Comput.Mol.Sci.,* (2011), **1**, 844-854.
50. Lopes,P.E.M.; Roux,B.; MacKerell,A.D. *Theor.Chem.Acc.,* 2009, **124**, 11-28.
51. Anisimov,V.M.; Vorobyov,I.V.; Roux,B.; MacKerell,A.D. *J.Chem.Theor. Comput.,* 2007, **3**, 1927-1946.
52. Huang,L.; Roux,B. *J.Chem.Theor.Comput.,* 2013, **9**, 3543-3556.
53. Warshel,A.; Levitt,M. *J.Mol.Biol.,* 1976, **103**, 227-249.
54. Warshel,A.; Weiss,R.M. *J.Am.Chem.Soc.,* 1980, **102**, 6218-6226.
55. Kamerlin,S.C.; Warshel,A. *WIRES Comput.Mol.Sci.,* 2011, **1**, 30-45.
56. Aaqvist,J.; Warshel,A. *Chem.Rev.,* 1993, **93**, 2523-2544.
57. Gao,J.; *Reviews in Comp.Chem.,* 1996, **7**, 119-185.
58. Gao,J.; Xia,X. *Science,* 1992, **258**, 631-635.
59. Gao,J. *Acc.Chem.Res.,* 1994, **29**, 298-307.
60. Bash,P.A.; Field,M.J.; Karplus,M. *J.Comput.Chem.,* 1990, **11**, 700-733.
61. Senn,H.M.; Thiel,W. *Angew.Chem.Int.Ed.Eng*., 2009, **48**, 1198-1229
62. Fan,Y.; Cembran,A.; Ma,S.; Gao,J. *Biochemistry*, 2012, **52**, 2036-2049.
63. Baaden,M.; Marrink,S.J. *Curr. Opin. Struct. Biol.,* 2013, **23**, 878-886.
64. Rinker,S.; Allison,J.R.; van Gunsteren,W.F., *Phys.Chem.Chem.Phys.,* 2012, **14**, 12423-12430.
65. Clementi,C. *Curr.Opin.Struct.Biol.,*2008, **18**, 10-15.
66. Tozzini,V. *Curr.Opin.Struct.Biol.,*2005, **15**, 144-150.
67. Kamerlin,S.C.L.; Vicaros,S.; Dryga,A.; Warshel,A. *Ann.Rev.Phys.Chem.,* 2011, **62**, 41-64.
68. Dama,J.F.; Sinitskiy,A.V.; McCullagh,M.; Weare,J.; Roux,B.; Dinner,A.R.; Voth,G.A. *J.Chem.Theor.Comput.,* 2013, **9**, 2466-2480.
69. Taketomi,H.; Ueda,Y.; Go,N. *Int.J.Pept.Prot.Res.,* 1975,**7**,445-448.
70. Go,N. *J.Stat.Phys.,* 1983, **30**, 413-423.
71. Best,R.B.; Hummer,G.; Eaton,W.A. *Proc. Natl. Acad. Sci. USA,* 2013, **110**, 17874-17879.
72. Clementi,C.; Nymeyer,H.; Onuchic,J.N. *J.Mol.Biol.,* 2000, **298**, 937-953.
73. Clementi,C.; Garcia,A.E.; Onuchic,J.N. *J.Mol.Biol.,* 2003, **326**, 933-954.

74. Naganathan,A.; Orozco,M. *J.Am.Chem.Soc.,* 2011, **133**, 12154-12161.
75. Tirion,M.M. *Phys.Rev.Lett.,* 1996, **77**, 1905-1908.
76. Atilgan,A.R.; Durell,S.R.; Jernigan,R.L.; Demirel,M.C.; Keskin,O.; Bahar,I. *Biophys.J.,* 2001, **80**, 505-515.
77. Kovacs,J.A.; Chacon,P.; Abagyan,R. *Proteins* 2004, **56**, 661-668.
78. Orellana,L. ; Rueda,M. ; Ferrer-Costa,C ; López-Blanco,J.R.; Chacón,P.; Orozco,M. *J.Chem.Theor.Comput.,* 2010, **6**, 2910-2923.
79. Camps,J.; Carrillo,O.; Emperador,A.; Orellana,L.; Hospital,A.; Rueda,M.; Cicin-Sain,D.; D'Abramo,M.; Gelpi,J.L.; Orozco,M. *Bioinformatics* 2009, **25**, 1709-1710.
80. Emperador,A.; Carrillo,O.; Rueda,M.; Orozco,M. *Biophys.J.,* 2008, **95**, 2127-2138.
81. Monticelli,L.; Kandasamy,S.K.; Periole,X.; Larson,R.G.; Tieleman,D.T.; Marrink,S.J. *J.Chem.Theor.Comput.,* 2008, **4**, 819-834.
82. Marrink,S.J.; Tieleman,D.P. *Chem.Soc.Rev.,* 2013, **42**, 6801-6822.
83. Kali,A.C.; Campbell,I.D.; Sansom,M.S.P. *Proc. Natl. Acad. Sci. USA,* 2011, **108**, 11890-11895.
84. Vostrikov,V.V.; Hall,B.A.; Greathouse,D.V.; Koeppe,R.E.; Sansom,M.S.P. *J.Am.Chem.Soc.,* 2010, **132**, 5803-5811.
85. Arnaez,C.; Mazat,J.P.; Elezgaray,J.; Marrink,S.J. Periole,X. *J.Am.Chem.Soc.,* 2013, **135**, 3112-3120.
86. Rollauer,S.E.; Tarry,M.J.; Graham,J.E.; Jääskeläinen,M.; Jäger,F.; Johnson,S.; Krebenbrink,M.; Liu,S.M.; Lukey,M.J.; Marcoux,J.; McDowell,M.A.; Rodriguez,F.; Roversi,P.; Stansfeld,P.J.; Robinson,C.V.; Sansom,M.S.P.; Palmer,T.; Hogbom,M. Berks,B.C.; Lea,S.M. *Nature*, 2012, **492**, 210-214.
87. Pasi,M.; Lavery,R.; Ceres,N. *J.Chem.Theor.Comput.,* 2013, **9**, 785-793.
88. Rueda,M.; Ferrer,C.; Meyer,T.; Perez,A.; Hospital,A.; Gelpí,J.L.; Orozo.M. *Proc. Natl. Acad. Sci. USA,* 2007, **104**, 796-801.
89. Meyer,T.; D'Abramo,M.; Hospita,A.; Rueda,M.; Ferrer-Costa,C.; Pérez,A.; Carrillo,O.; Fenollosa,C.; Rechevsky,D.; Gelpí,J.L.; Orozco,M. *Structure*, 2010, **18**, 1399-1409.
90. Jamroz,M.; Orozco,M.; Kolinski,A.; Kmiecik,S. *J.Chem.Theor.Comput.,* 2013, **9**, 119-125.
91. Hills,R.D.; Lu,L.; Voth,G.A. *PLOS Comput.Biol.,* 2010, **6**, e1000827
92. Shi,A.Y.; Arkhipov,A.; Freddolino,P.L.; Schulten,K. *J.Phys.Chem.B.,* 2006, **110**, 3674-3684.
93. Arkhipov,A.; Freddolino,P.L.; Imada,K.; Namba,K.; Schulten,K. *Biophys. J.,* 2006, **91**, 4589-4597.
94. Arkhipov,A.; Freddolino,P.L., Schulten,K. Structure, 2006, **14**, 1767-1777.
95. Liwo,A.; Khalili,M.; Scheraga,H.A. *Proteins*, 2005, **102**, 2362-2367.
96. Khalili,M.; Liwo,A.; Scheraga,H.A. *J.Mol.Biol.,* 2006, **355**, 536-547.
97. Piana,S.; Bucher,D.; Carloni,P.; Rothlisberger,U. *J.Phys.Chem.,* 2004, 108, 11139-11149.
98. Carloni,P.; Rothlisberger,U.; Parrinello,M. *Acc.Chem.Res.,* 2002, **35**, 455-464.
99. McCammon.J.A.; Gelon,B.R.; Karplus,M. *Nature*, 1977, **267**, 585-590.
100. Warshel,A. *Nature* 1976, **260**, 679-683.
101. Karplus,M.; McCammon,J.A. *Nat.Struct.Biol.,* 2002, **9**, 646-652.

102. Pronk,S.; Páll,S.; Schultz,R.; Larsson,P.; Bjelkmar,P.; Apostolov,R.; Shirts,M.R.; Smith,J.C.; Kasson,P.M.; van der Spoel,D.; Hess,B.; Linhdahl,E. *Bioinformatics*, 2013, **29**, 845-854.

103. Phillips,J.C.; Braun,R.; Wang,W.; Gumbart,J.; Tajkhorshid,E.; Villa,E.; Chipot,C.; Skeel,R.D.; Kále,L.; Schulten,K. *J.Comput.Chem.,* 2005, **26**, 1781-1802.

104. Zhao,G.; Perilla,J.R.; Yufenyuy,E.K.; Meng,X.; Chen,B.; Ning,J.; Ahn,J.; Gronenborn,A.M.; Schulten,K.; Aiken,C.; Zhang,P. *Nature*, 2013, **497**, 643-646.

105. Klein,M.L.; Shinoda,W. *Science*, 2008, **321**, 798-800.

106. Larsson,D.S.D.; Liljas,L.; van der Spoel,D. *PLoS Comput. Biol.,* 2012, **8**, e1002502

107. Dror,R.O.; Dirks,R.M.; Grossman,J.P.; Xu,H.;M.; Shaw.D.E. *Annual Reviews in Biophysics,* 2012, **41**, 429-4452.

108. Shaw,D.E.; Margakis,P.; Lindorff-Larsen,K.; Piano,S.; Dror,R.O.; Eastwood,M.PO.; Bank,J.A.; Jumper,J.M.; Salmon,J.K.; Shan,Y.; Wriggers,W. *Science*, 2010, **330**, 341-346.

109. Lindorff-Larsen,K.; Piana,S.; Drod,R.O.; Shaw,D.E. *Science*, 2011, **334**, 517-520.

110. Piana,S.; Lindorff-Larsen,K.; Shaw,D.E. *Proc. Natl.Acad.Sci.USA.,* 2012, **109**, 17845-17850.

111. Dickson,A.; Brooks,C.L. *J.Am.Chem.Soc.,* 2013, **135**, 4729-4734.

112. Beauchamp,K.A.; McGibbon,R.; Lin,Y.-S.; Pande,V.S. *Proc.Natl.Acad.Sci. USA,* 2012, **109**, 17807-17813.

113. Piana,S.; Lindorff-Larsen,K.; Shaw,D.E. *Proc. Natl.Acad.Sci.USA.,* 2013, In Press

114. Jensen,M.; Jogini,V.; Borhani,D.W.; Leffler,A.E.; Dror,R.O.; Shaw,D.E. *Science*, 2012, **336**, 229-233.

115. Arkhipov,A.; Shan,Y.; Das,R.; Endres,N.F.; Eastwood,M.P.; Wenner,D.E.; Kuriyan,J.; Shaw,D.E. *Cell*, 2013, **152**, 557-569.

116. Ostmeyer,J.; Chakrapani,S.; Pan,A.C.; Perozo,E.; Roux,B. *Nature*, In Press 2013.

117. Dror,R.O.; GreenH.F.; Valant,C: Borhani,D.W.; Valcourt,J.R.; Pan,A.C.; Arlopw,D.H.; Canals,M.; Lane,R.; Rahmani,R.; Baell,J.B.; Sexton,P.M.; Cristopoulos,A.; Shaw,D.E. *Nature*, 2013, **503**, 295-299.

118. Harvey,M.; Giupponi,G.; De Fabritis,G. *J.Chem.Theor & Comput*., **2009**, *5*, 1632-1639.

119. Salomon-Ferrer;R.; Goetz.A.W.; Poole.D.; Le Grand,S.; Walker.R. *J. Chem. Theory Comput.*, 2013, in press.

120. Xu,D.; Williamson,M.J.; Walker,R.C. *Ann.Rep. in Comput Chem,* 2010, **6**, 2-19.

121. Buch,I.; Giorgino,T.; De Fabritis,G. *Proc. Natl. Acad. Sci. USA,* 2011, **108**, 10184-10189.

122. Buch,I.; Harvey,M.J.; Giorgino,T.; Anderson,D.P.; De Fabritis,G. *J.Chem.Inf & Mod.,* 2010, **50**, 397-405.

123. Sadiq,S.K.; Noe,F.; De Fabritis,G. *Proc. Natl. Acad. Sci. USA,* 2012, **109**, 20449-20454.

124. Pande,V.S.; Rokhsar,D.S. *Proc. Natl. Acad. Sci. USA,* 1999, **96**, 9062-9067.

125. Snow,C.D.; Nguyen,H.; Pande,V.S.; Gruebele,M. *Nature* 2002, **420**, 102-106.

126. Weber,J.K.; Jack,R.L.; Pande,V.S. *J.Am.Chem.Soc.,* 2013, **135**, 5501-5504.

127. Kolhoff,K.J.; Shukla,D.; Lawrenz,M.; Bowman,G.R.; Konerding,D.E.; Belov,D.; Altman,R.B.; Pande,V.S. *Nature Chemistry*, 2014, **6**, 15-21.

128. Tan,Y.S.; Sledz,P.; Lang,S.; Stubbs,C.J.; Spring,D.R.; Abell,C.; Best,R.B. Angew.Chem.Int.Ed.Eng. 2012, **51**, 10078-10081.

129. Knott,M.; Best,R.B. *PLOs Comput Biol.* 2012, **8**, e1002605

130. Sugita,Y.; Okamoto,Y.*; Chem.Phys.Lett.,* 1999, **314**, 141-151.

131. Earl,D.J.; Deem,M.W. *Phys.Chem.Chem.Phys.,* 2005, **7**, 3910-3916.

132. Rhee,Y.M.; Pande,V.S. *Biophys. J.*, 2003, 84, 775-786.

133. Rauscher,S.; Neale,C.; Pòmes,R. J.Chem.Theor.Comput., 2009, 5, 2640-2662.

134. Fukunishi,H.; Watanabe,O.; Takada,S. *J.Chem.Phys.,* 2002, **116**, 9058-9067

135. Bowman,G.R.; Pande,V.S.; Noé,F. “An introduction to Markov State Models and their application to long timescale molecular simulations”. Springer. 2014.

136. Bruchete.N.V.; Hummer,G. *J.Phys.Chem.B.;* 2008, **112**, 6057-6069.

137. Noe,F.; Horenko,I.; Schütte,C.; Smith,J.C. *J.Chem.Phys.,* 2007, **126**, 155102-155107.

138. Noe.F.; Schütte.; Vanden-Eijnden,E.; Lothat,R.; Weik,T.R. *Proc. Natl. Acad. USA.,* 2009, **106,** 19011-19016.

139. Pande,V.S.; Beauchamp,K.; Browman,G.R. *Methods*., 2010, **52**, 99-105.

140. Best,R.B. *Current Opin. Struct. Biol.,* 2012, **22**, 52-61.

141. Bock,L.V.; Blau,C.; Schröder,G.F.; Davydov,II.; Fisher,N.; Stark,H.; Rodnina,M.V.; Vaiana,A.C.; Grubmüller,H. *Nature Struct. Molec. Biol.,* In Press 2013.

142. Aponte-Santamaria,C.; Briones,R.; Schenk,A.D.; Walz,T.; de Groot,B.L. *Proc. Natl. Acad. Sci. USA.,* 2012, **109**, 44319-44325.

143. Hoefling,M.; Grubmüller,H. *Comp.Phys.Comm.,* 2013, **184**, 841-852.

144. Torrie,G.M.; Valleau,J.P. *J.Comput.Phys.,* 1977, 23, 187-199.

145. Kumar,S.; Rosenber,J.; Bouzina,D.; Swendsen,R.H.; Kollman,P.A. *J.Comput.Chem.,* 1992, **13**, 1011-1021.

146. Schlitter,J.; Engels,M.; Krüger,P.; Jacoby,E.; Willmer,A. Mol.Simul., 1993, **10**, 291-308.

147. Grubmüller,H.; Heymann,B.; Tavan.P. *Science*, 1996, **271**, 997-999.

148. Sotomayor,M.; Schulten,K. *Science*, 2007, **316**, 1144-1148.

149. Jensen,M.; Park,S.; Tajkhorshid,E.; Schulten,K. *Proc. Natl. Acad. Sci. USA.,* 2002, **99**, 6731-6736.

150. Izrailev,S.; Stepaniants,S.; Isralewitz,B.; Kosztin,D.; Lu,H.; Molnar,F.; Wriffers,W.; Schulten,K. In Algorithms for Macromolecular Modelling. Lecture Notes in Computational Science and Engineering. Vol 4.

P.Deaufhard,, J.Hermans, B.Leinkuhler, A.Mark, R.D.Skeel, and S.Reich eds. Springer-Verlag. Berlin, pp 39-65. 1998.

151. Okazaki,K.I.; Hummer,G. Proc. Natl. Acad. Sci. USA., 2013, 41, 16468-73.

152. Zuckerman,D.M.; Woolf,T.B. *J.Chem.Phys.,* 1999, **111**, 9475-9484.

153. Rueda,M.; Cubero,E.; Laughton,C.A. Orozco,M. *Biophys.J.,* 2004, **87**, 800-811.

154. Laio,A.; Parrinello,M. *Proc. Natl. Acad. Sci.USA.,* 2002, **99**, 12562-12566.

155. Huber,T.; Torda,A.E.; van Gunsteren,W.F. *J.Comput-Aided Mol. Des.,*1994, **8**, 695-708.

156. Barducci,A.; Bonomi,M.; Parrinello,M. *WIREs Comput. Mol. Sci.,* 2011, **1**, 826-843.

157. Sutto,L.; Marsili,S.; Gervasio,F.L. *WIREs Comput. Mol. Sci.,* 2012, **2**, 771-779.

158. Herbet,C.; Schieborr,U.; Saxena,K.; Juraszek,J.; De Smet,F.; Alcouffe,C.; Biacniotto,M.; Saladino,G.; Sibrac,D.; Kudlinzki,D.; Sreeramulu,S.; Bron,.A.; Rigon,P.; Herault,J.-P., Lassalle,G.; Blundell,T.; Rousseau,F.; Gils,A.; Schymkowitz,J.; Tompa,P.; Herbert,J.M.; Carmeliet,P.; Gervasio,F.L.; Schwalbe,H.; Bono,F. *Cancer Cell.,* 2013, **23**, 489-501.

159. Palazzesi,F.; Barducci,A.; Tollinger,M.; Parrinello,M. *Proc. Natl. Acad. Sci. USA*, 2003, **110**, 9201-9208

160. Gervasio,F.L.; Laio,A.; Parrinello,M. *J.Am.Chem.Soc.,* 2006, **128**, 13435-13441.

161. Barducci,A.; Chelli,R.; Procacci,P.; Schettino,F.L.; Gervasio,F.L.; Parrinello.M. *J.Am.Chem.Soc.,* 2006, **128**, 2705-2710.

162. McCammon,J.A.; Lee,C.Y.; Northrup,S.H. *J.Am.Chem.Soc*., 1993, **105**, 2352-23757.

163. Addock,S.A.; McCammon,J.A. *Chem.Rev*., 2006, **106**, 1589-1615.

164. Fischer,S.; Karplus,M. *Chem.Phys.Lett.,* 1992, **194**, 252-261.

165. Fisher,S.; Olsen,K.W.; Nam,K.; Karplus,M. *Proc.Natl.Acad.Sci.USA.,* 2011, **108**, 5608-5613

166. Johnson,M.E.; Hummer,G. *J.Phys.Chem.B.,* 2012, **116**, 8573-8583.

167. Bolhuis,P.; Chandler,D.; Dellago,C.; Geissler,P. *Annu.Rev.Phys.Chem.,* 2002, **59**, 291-318.

168. Dellago,C.; Bohuils,P.G.; Geissler,P.L. *Adv.Chem.Phys.,* 2002, **123**, 1-84.

169. Orozco,M.; Orellana,L.; Hospital,A.; Naganathan,A.; Emperador,A.; Carrillo,O.; Gelpi,J.L. Advances in Protein Chemistry and Structural Biology. Vol 85. C.Cristov (ed). Burlington: Academic Press., 2011, pp 183-215.

170. Ermak,D.L.; McCammon,J.A. *J.Chem.Phys.,* 1978, **69**, 1352-1360.

171. McCammon,J.A.; Harvey,S.C. Dynamics of proteins and nucleic acids. Cambridge University Press., New York. 1987

172. Carrillo,O.; Laughton,C.A.; Orozco,M. *J.Chem.Theor.Comput.,* 2012, 8, 792-799.

173. Chaudhuri,R.; Carrillo,O.; Laughton,C.A.; Orozco,M *J.Chem.Theor. Comput.,* 2012, **8**, 2204-2214.

174. Amadei,A.; Linssen,A.B.M.; Berendsen,H.J.C. *Proteins*., 1983, **17**, 412-425.

175. Hinsen,K.; Petrescu,A.; Dellerue,S.; Bellisent-Funel,M.; Kneller,G. *Chem.Phys.,* 2000, **261**, 25-30.

176. Sen,T.Z.; Jernigan,R.L. Optimizing the parameters for Gaussian network model for ATP-binding proteins. In: Normal Mode Analysis: Theory and applications. Cui,Q and Bahar,I (Eds). Pp 171-186. CRC Press. Boca Raton. CA. 2006

177. Rueda,M.; Chacon,P.; Orozco,M. *Structure*, 2007, **15**, 565-575.

178. Yang,L.; Eyal,F.; Chennubhoda,C.; Jee,J.G.; Gronenbon,A.; Bahar,I. *Structure*, 2007, **15**, 741-749.

179. Zheng,W.; Brooks,B.-R.; Thirumalai,D. *Curr. Protein Pept Sci.,* 2009, **10**, 128-132.

180. Zheng,W.-; Brooks,B.R.; Thirumalai,D. *Proc. Natl. Acad. Sci. USA.,* 2006, 103, 7664-7669.

181. Bray,J.K.; Weiss,D.R.; Levitt,M. *Biophys.J.,* 2011, **101**, 2966-2969.

182. Mendez,R.; Bastolla,U. *Phys.Rev.Lett.,* 2010, **104**, 228103-228107.

183. López-Blanco,J.R.; Garzón,J.I.; Chacon,P. *Bioinformatics*, 2011, **27**, 2843-2850.

184. Zheng,W. *Biophys.J.;* 2010, **98**, 3025-3034.

185. Keating,K.S.; Flores,S.C.; Gerstein,M.B.; Kuhn,L.A. *Protein Sci.* 2009, **18**, 359-371.

186. Suhre,K.; Sanejouand,Y.H. *Nucleic Acids Res*., 2004, **32**, W610-W614

187. Hinsen,K. *Proteins*, 1998, **33**, 417-429.

188. Maragakis,P.; Karplus,M. *J.Mol.Biol*., 2005, **352**, 807-822.

189. Seo,S.; Kim,M.K. *Nucleic Acids Res.,* 2012, **40**, W531-W536.

190. Chu,J.W. BVoth,G.A. *Biophys. J.,* 2007, **93**, 3860-3871.

191. Noy,A.; Pérez,A.; Laughton,C.; Orozco,M. *Nucleic Acids Res.,* 2007, **35**, 3330-3338.

192. Yang,Z.; Majeck,P.; Bahar,I. *PLOS Comput.Biol.*, 2009, **5**, e1000360.

193. Sfriso,P.; Emperador,A.; Orellana,L.; Hospital,A.; Gelpí,J.L.; Orozco,M. *J.Chem.Theor.Comput.,* 2012, **8**, 4707-4718.

194. Jorgensen,W.L.; Tirado-Rives,J. *J.Comput.Chem.,* 2005, **26**, 1689-1700.

195. Borrelli,K.W.; Vitalis,A.; Alcantara,R.; Guallar,V. *J.Chem.Theor.Comput.,* 2005, **1**, 1304-1311.

196. Dill,K.A.; Bromberg,S.; Yue,K.; Fiebig,K.M.; Yee,D.P.; Thomas,P.D.; Chan,H.S. *Protein Science*, 1995, **4**, 561-602.

197. Lee,W.-G.; Gallardo-Macias,R.; Frey,K.M.; Spasov,K.A.; Bollini,M.; Anderson,K.S.; Jorgensen,W.L. *J.Am.Chem.Soc.,* In Press 2014.

198. Bollini,M.; Domaoal,R.A.; Thakur,V.V.; Gallardo-Macias,R.; Spasov,K.A.; Anderson,K.R.; Jorgensen,W.L.; *J.Med.Chem.,*2011, **54**, 8582-8591.

199. Proctor,E.; Ding.F.; Doholyan,N.V. *WIRES Comput. Mol. Sci.,* 2011, **1**, 80-92.

200. Dokholyan,N.V. *Curr. Opin. Struct. Biol*., 2006, 16, 79-85.

201. Zhou,Y.Q.; Karplus,M. *Nature*, 1999, **401**, 400-403.

202. Ding,F.; Dokholyan,N.V. *Proc. Natl. Acad. Sci. USA,* 2008, **105**, 19696-19701

203. Ding,F.; Buldyrev,S.V.; Dokholyan,N.V. *Biophys.J.,* 2005, **88**,147-156.

204. Dokholyan,N.V.; Buldyrev,S.V.; Stanley,H.E.; Shakhnovich,E.L. *Fold Des.,* 1998, **3**, 577-587.

205. Emperador,A.; Meyer,T.; Orozco,M. *J.Chem.Theory Comput.,* 2008, **4**, 2001-2010.

FIGURE CAPTIONS

Figure 1. Representation of the B-factor profile of a large protein (Rhamnogalacturonan Lyase from Aspergillus Aculeatus; PDB entry 1NKG) computed from EN-NMA calculations and determined from X-Ray diffraction measures.

Figure 2. Structural variability sampled by evolution of the SH3 family (PDB entry 1ARK as the central structure for the family) and by MD simulation in our MoDEL database (101,102). Left panel show structure superposition and right panel B-factors obtained by considering either evolution or MD variance.

Figure 3. Example of coarse graining of a small a-helix for elastic network model calculations. Arrows indicate the interactions (springs) felt by the first residue, the widths of the arrows are representative of the stiffness of he springs, that are weighted according to either Cartesian distance, or a combination of sequence and Cartesian distance depending on the method (see text).

Figure 4. Basic algorithm of Replica Exchange. Many independent replicas are launched at closely spaced temperatures. Every certain number of replicas Metropolis test is applied to decide whether or not two replicas have to be interchanged according to its energy (see eq. 9; Metropolis test can be applied to all pairs of temperatures, not only to neighbor temperatures as shown for the shake of clarity in the Figure). At the end, Boltzmann's ensembles corresponding to each temperature are obtained. In the plot colour indentify the original replica.

Figure 5. General scheme of the discrete molecular dynamics algorithm for two particles subjected to a single one-dimensional square well potential. For all distances between $(1-\sigma)R_{eq}$ nd $(1+\sigma)R_{eq}$ the particles move at constant velocities. At the boundaries $(1-\sigma)R_{eq}$ and $(1+\sigma)R_{eq}$ they interchange momentum (assuming typically elastic collision model). The equilibrium value of the well (R_{eq}) and the width of the well is taken from oscillation values in MD simulations.

Figure 1

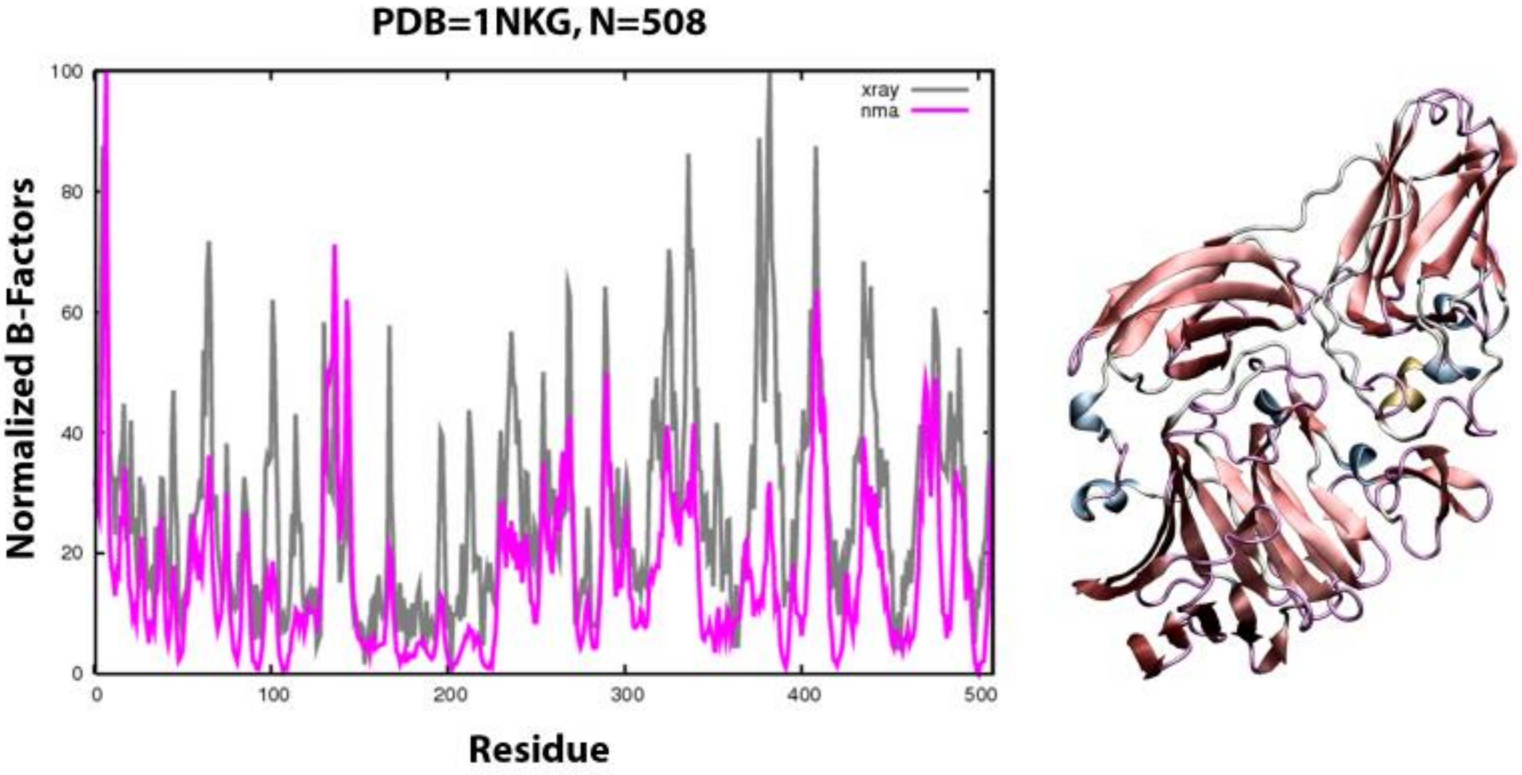
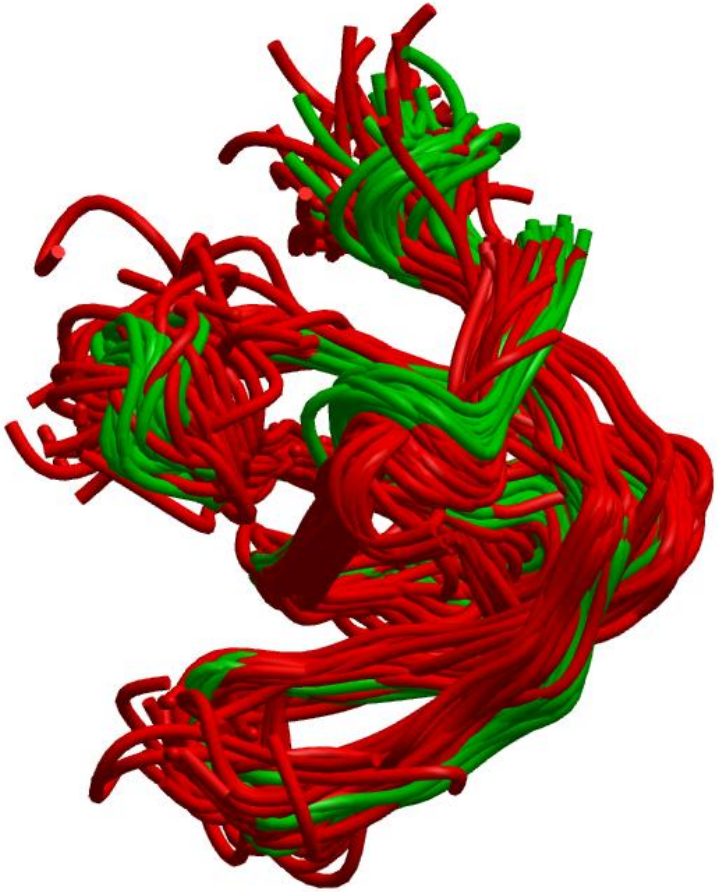
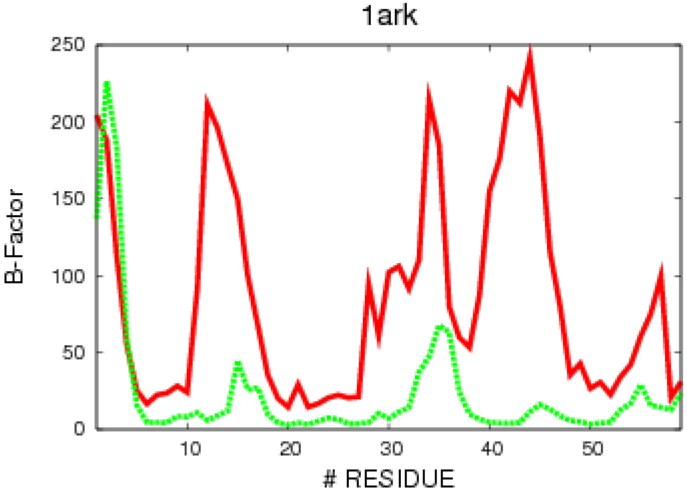

PDB=1NKG, N=508
Normalized B-Factors
Residue
xray
nma
0
20
40
60
80
100
0
100
200
300
400
500

Figure 2

SuperFamilies vs. MD



Domain 1ark (59 aa, 66 elements)



🟥 **SF**

🟩 **MD**

Figure 3

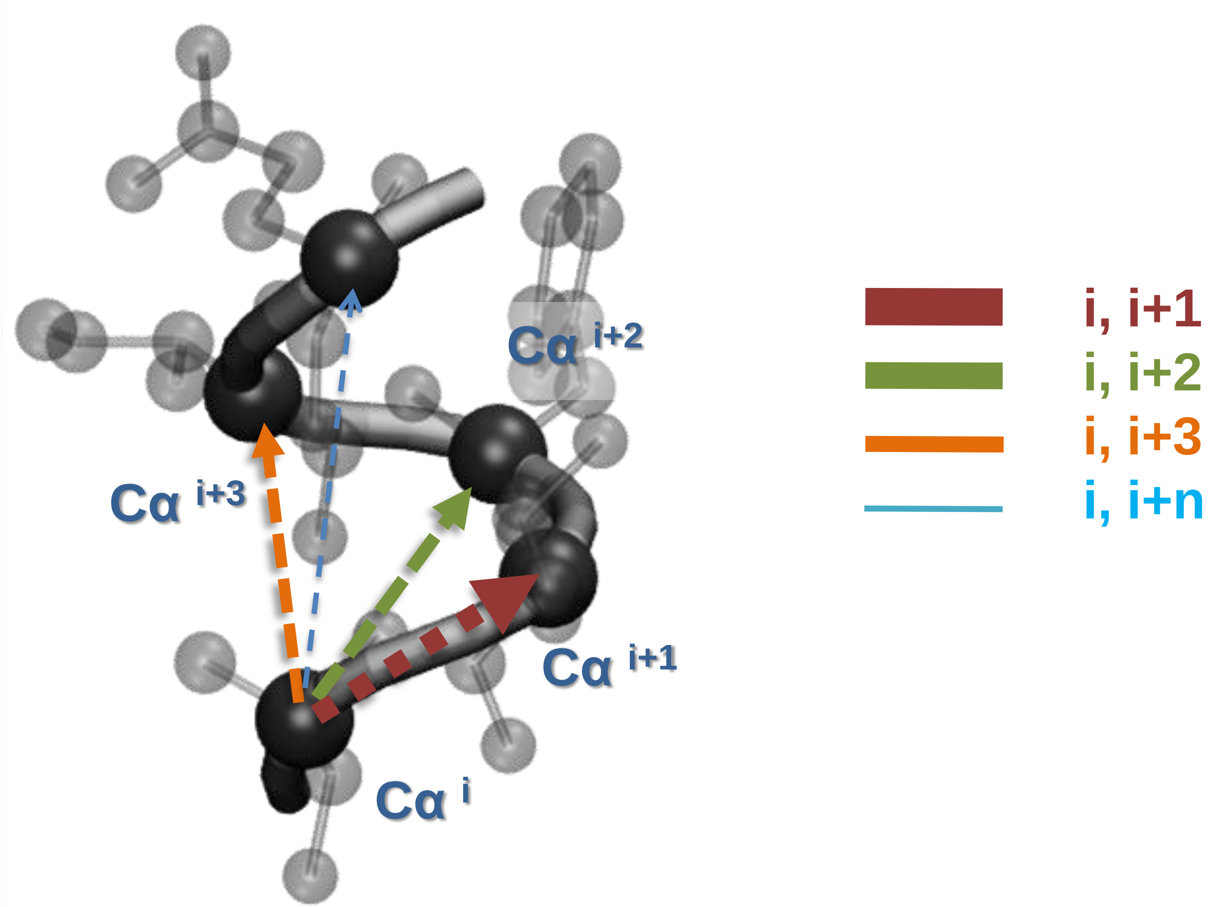


Figure 4

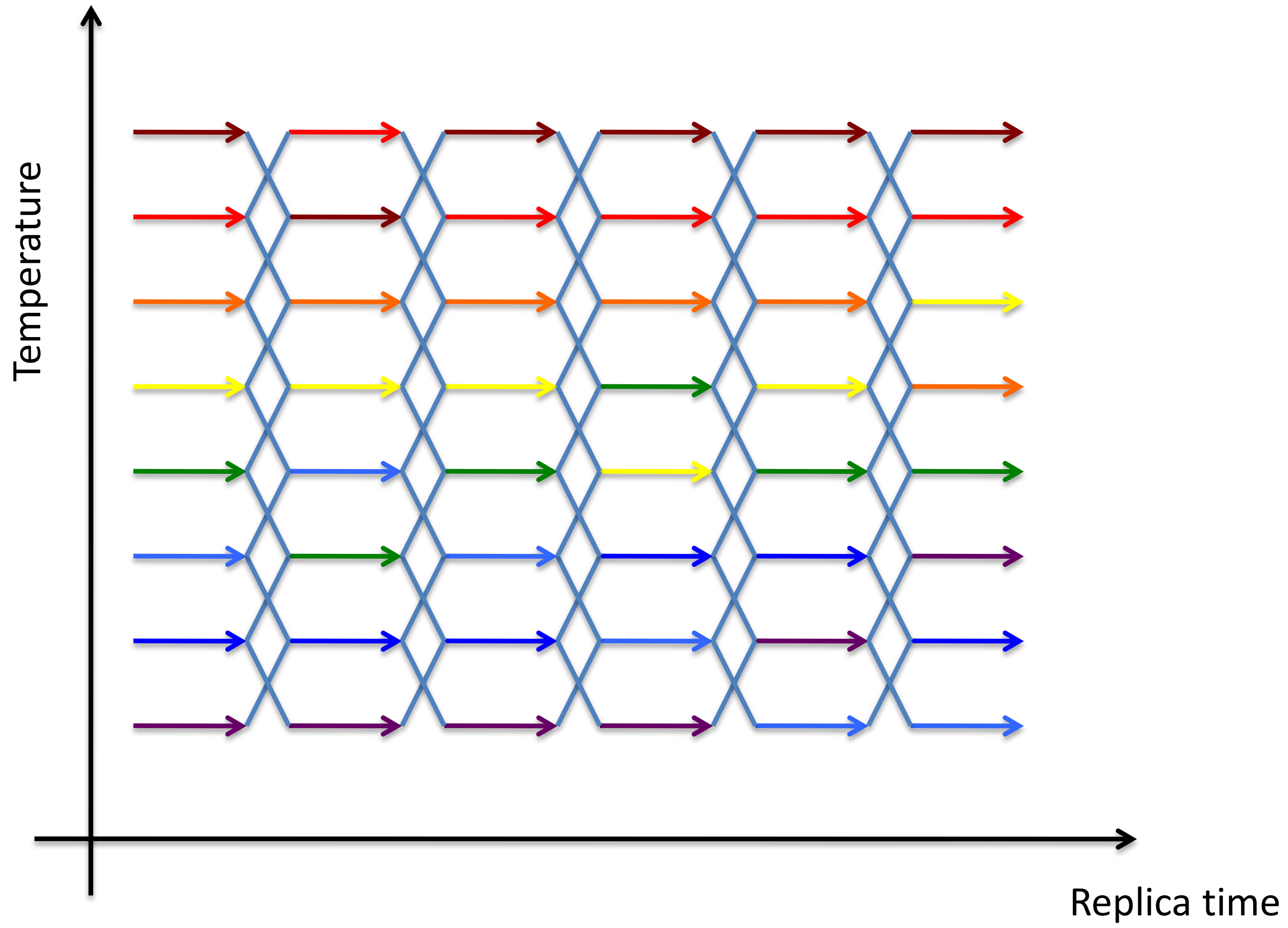
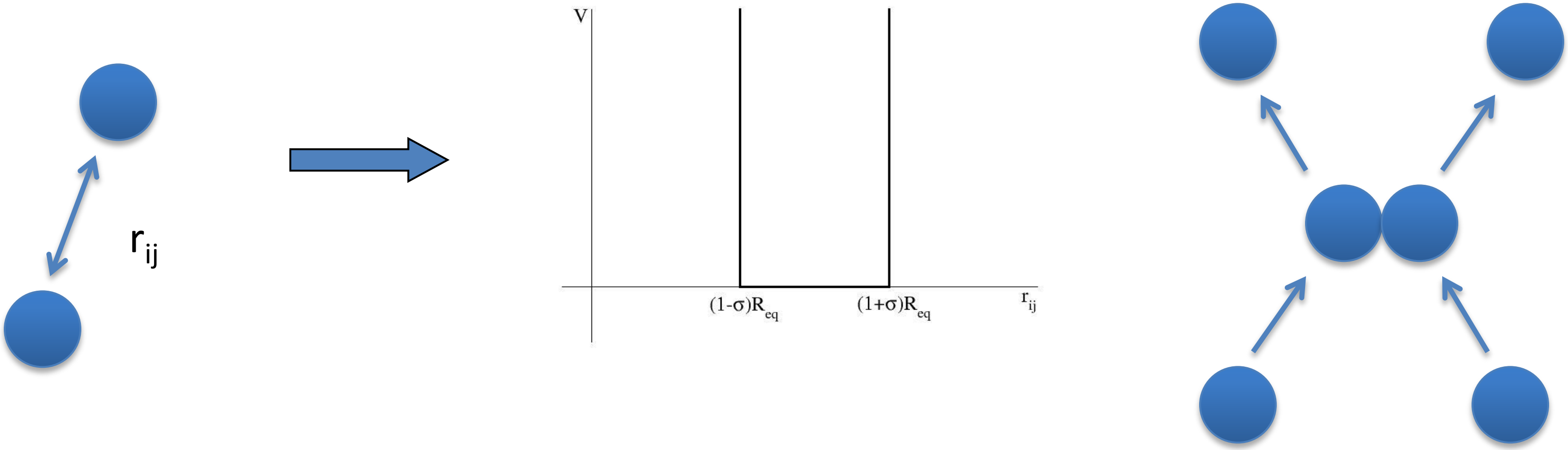


Figure 5



Out of well: forbidden

Within the well: $\vec{r}_i(t+t_c) = \vec{r}_i(t) + \vec{v}_i(t)t_c$

Collision: $m_j\vec{v}_j + \Delta\vec{p} = m_j\vec{v}_j{}'$ $\quad m_i\vec{v}_i = m_i\vec{v}_i{}' + \Delta\vec{p}$

with: $t_{ij} = \frac{-b_{ij} \pm \sqrt{b_{ij}{}^2 - v_{ij}{}^2(r_{ij}{}^2 - d^2)}}{v_{ij}{}^2}$ $\quad \Delta p = \frac{2m_i m_j}{m_i + m_j}\left(v_i - v_j\right)$

BIOGRAPHY

Modesto Orozco got his Ph.D. in chemistry in 1990 from the Universitat Autònoma de Barcelona. He is Full Professor of Biochemistry at the Universitat de Barcelona since 2004. He is principal researcher at the Institute for Research in Biomedicine (IRB Barcelona), director of the Life Science Department at the Barcelona Supercomputing Center (BSC), as well as the director of the Joint BSC-CRG-IRB Program in Computational Biology. MO main interests are in the field of computational biology and chemistry. He is author of more than 340 research papers and has collected over 14000 citations (2013).

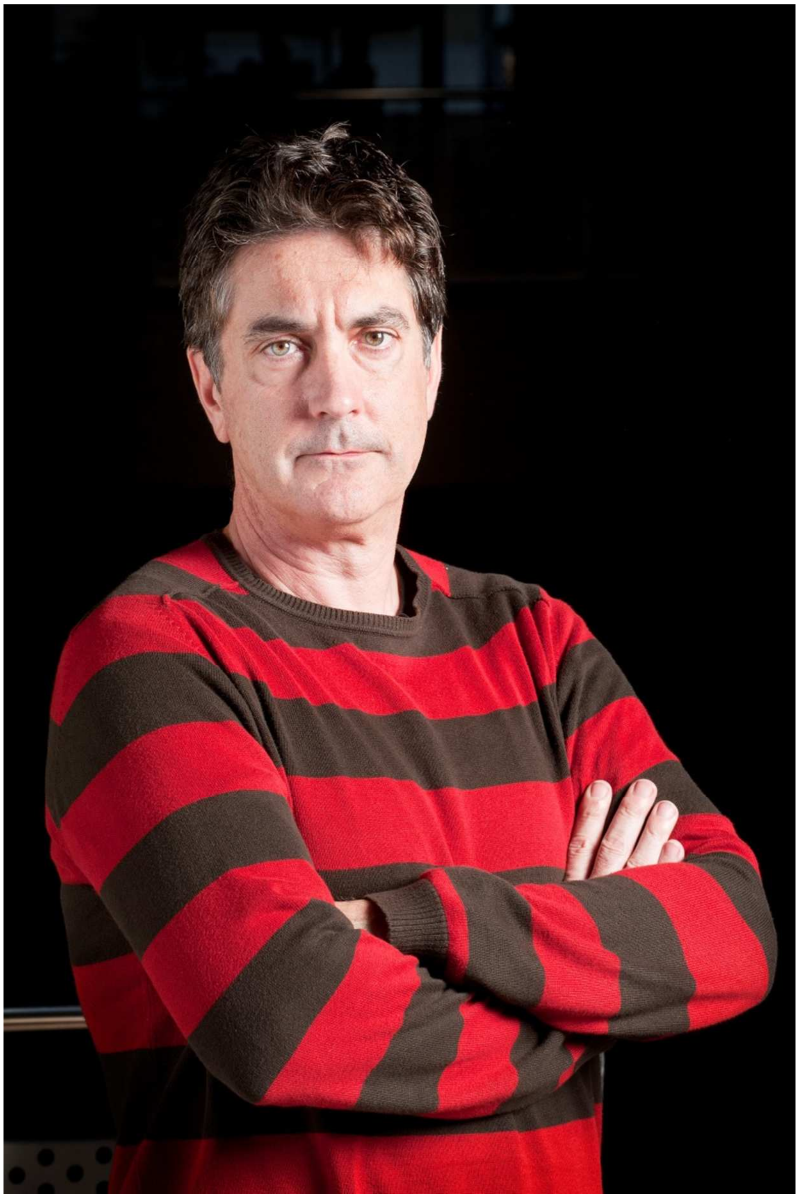


TABLE OF CONTEXT

Moving from a traditional static picture of proteins to an alternative dynamic paradigm is one of the biggest challenges of structural biology, and the point where modeling can contribute the most. I review here the current state of the art in theoretical methods for dynamics representations of proteins.

TOC_Figure