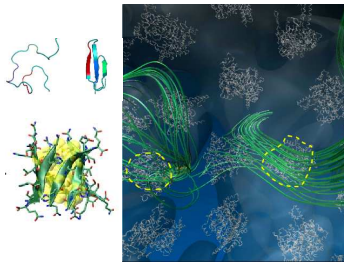
Chem Soc Rev

THE OPEP COARSE-GRAINED PROTEIN MODEL: FROM SINGLE MOLECULES, AMYLOID FORMATION, ROLE OF MACROMOLECULAR CROWDING AND HYDRODYNAMICS TO RNA/DNA COMPLEXES

Journal:	*Chemical Society Reviews*
Manuscript ID:	CS-REV-01-2014-000048.R1
Article Type:	Review Article
Date Submitted by the Author:	29-Mar-2014

Complete List of Authors:	Sterpone, Fabio; Laboratoire de Biochimie Théorique, UPR9080, CNRS, IBPC Melchionna, Simone; CNR, IPCF Tuffery, Pierre; INSERM U273, Université Paris Diderot Pasquali, Samuela; Laboratoire de Biochimie Theorique, UPR 9080 CNRS, IBPC Mousseau, Normand; Université de Montreal, Departement de Physique Cragnolini, Tristan; Laboratoire de Biochimie Theorique, UPR 9080 CNRS, IBPC Chebarro, Yasmine; Laboratoire de Biochimie Théorique, UPR9080 CNRS, IBPC; University of Cambridge, Departement of Chemistry Saint-Pierre, Jean-Francois; University of Montreal, Departement de Physique Kalimeri, Maria; Laboratoire de Biochimie Theorique, UPR 9080 CNRS, IBPC Barducci, Alessandro; EPFL, Laboratory of Statistical Biophysics Laurin, Yohan; Laboratoire de Biochimie Théorique, UPR9080 CNRS, IBPC Tek, Alex; Laboratoire de Biochimie Théorique, UPR9080 CNRS, IBPC; Upsalla University, Department of Cell and Molecular Biology Baaden, Marc; Laboratoire de Biochimie Theorique, UPR 9080 CNRS, IBPC Nguyen, Phuong; Laboratoire de Biochimie Theorique, UPR 9080 CNRS, IBPC Derreumaux, Philippe; Laboratoire de Biochimie Theorique, UPR 9080 CNRS, IBPC

THE OPEP COARSE-GRAINED MODEL: BIOLOGICAL PHYSICS FROM THE MOLECULAR LEVEL TO THE CELL



Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

THE OPEP COARSE-GRAINED PROTEIN MODEL: FROM SINGLE MOLECULES, AMYLOID FORMATION, ROLE OF MACROMOLECULAR CROWDING AND HYDRODYNAMICS TO RNA/DNA COMPLEXES.

Fabio Sterpone,[a] Simone Melchionna,[b] Pierre Tuffery,[c] Samuela Pasquali,[a] Normand Mousseau,[d] Tristan Cragnolini,[a] Yassmine Chebaro,[a,e] Jean-Francois Saint-Pierre,[d] Maria Kalimeri,[a] Alessandro Barducci,[f] Yohan Laurin,[a] Alex Tek,[a,g] Marc Baaden,[a] Phuong Hoang Nguyen,[a] and Philippe Derreumaux,[a,h,*]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

The OPEP coarse-grained protein model has been applied to a wide range of applications since its first release 15 years ago. The model, which combines energetic and structural accuracy and chemical specificity, allows studying single protein properties, DNA/RNA complexes, amyloid fibril formation and protein suspensions in a crowded environment. Here we first review the current state of the model and the most exciting applications using advanced conformational sampling methods. We then present the current limitations and a perspective on the on-going developments.

INTRODUCTION

Proteins, DNA and RNA carry out a variety of biochemical and biological tasks. These systems are very challenging experimentally and numerically due to their number of degrees of freedom, and the wide range of relevant time scales from nanoseconds to days associated with fluctuations about the native states, diffusion, folding and formation of harmful aggregates.

Classical atomistic molecular dynamics (MD) with explicit solvent and ions can complement experiments.[1,2] With the specially MD-designed Anton computer, performing MD 100-500 times faster than the standard computer, it has been possible to break the millisecond barrier and gain insights on the mechanisms, thermodynamics and kinetics of the folding of diverse proteins with 10-80 amino acids.[3] Anton has also proven useful in the development of allosteric inhibitors that target previously unknown binding sites.[4] The dynamic processes of life at the molecular level require, however, knowledge of the structure, dynamics and thermodynamics of biomolecules in a crowded environment. Similarly, we cannot wait for faster computers to design engineered proteins with specific properties or new molecules able to interfere with protein-protein or protein-DNA/RNA complexes associated with disease functions as a result of sporadic mutations or genetic risk factors. For instance, neurodegenerative diseases such as Alzheimer, Parkinson, Huntington challenge our society.[5] The V600E mutation in the BRAF protein is known to be responsible for 50% of melanoma cases,[6] and women with mutations in the BRCA1 and BRCA2 genes have markedly elevated risks of breast and ovarian cancer.[7]

For all these reasons, it is necessary to design multiscale approaches, coarse-grained models and advanced sampling methods to converge rapidly to equilibrium and explore dynamics in the time scale of microseconds and beyond for large systems.

The development of coarse graining (CG) and multiscale modeling is not new. It was an exciting day for computational chemistry and biology when the 2013 Nobel Prize in Chemistry was awarded to Martin Karplus, Michael Levitt, and Arieh Warshel for the "development of multiscale models for complex chemical systems." Among their contributions, Levitt and Warshel pioneered CG protein simulation, with atoms grouped into larger units or beads and normal modes to move the system on the energy landscape.[8] Despite extensive efforts, coarse graining still remains a challenge and poses the problem of how to derive potentials for the selected number of beads that maintain the all-atom physical behavior in test tubes and cellular environments. Recently, two reviews summarized the state of coarse-graining of biomolecular systems so as to couple information from different scales.[9,10] Marrink and Tieleman also presented the strengths and limitations of their CG Martini model, used for lipid membrane characterization, lipid polymorphism, membrane protein - lipid interplay, self-assembly of soluble proteins, and membrane protein oligomerization.[11]

The OPEP (Optimized Potential for Efficient protein structure Prediction) coarse-grained model with an implicit solvent model for soluble proteins in aqueous solution has evolved since its first version 15 years ago.[12-24] This model which retains structural accuracy and chemical specificity is free of any biases in contrast to Martini that imposes secondary structure constraints.[11,25,26] While CG models have been developed by Sansom[27,28] and Schulten[29] for membrane proteins, Klein,[30] Scheraga (UNRES),[31-33] Baker (Rosetta),[34,35] Voth,[36] Dokhloyan,[37] Hall (PRIME),[38] Lavery (PaLaCe),[39] Zacharias (ATTRACT),[40,41] Feig (PRIMO),[42] Shea,[43] Papoian (AWSEM)[44] and other scientists[45-48] for soluble proteins, to the best of our knowledge, none of these models

except Martini has been applied to a wide range of problems. Here, we present the design principles of the model and the state-of-the-art sampling techniques used. Furthermore this review provides an in-depth understanding on four timely topics in the chemical sciences.

The first topic is the self-assembly of amyloid proteins associated with neurodegenerative diseases. Alzheimer's disease affects 24 million people and drug after drug has failed to slow its progression.[49] The second topic is computer-assisted *de novo* design and structure prediction of peptides up to 52 amino acids, as they represent a source of novel antibiotics and therapeutics.[50-52] To this end, we need a fast and accurate method able to play with he amino acid sequence. The third topic is RNA structure prediction which is still in its infancy compared to protein structure prediction.[53] The small non-coding microRNA which regulate gene expression at a post-transcriptional level,[54] or more generally all non-coding RNAs are increasingly attracting the attention of cancer investigators.[55,56] The last topic concerns the effects of hydrodynamics and crowding which are mostly ignored in computer simulations and create a gap between the simulated and the real physics in the cell. We report an OPEP simulation of a system of unprecedented size and fully inclusive of hydrodynamic interactions, namely 18,000 flexible proteins and 70 million particles,[57] a breakthrough compared to the largest simulation of 1000 rigid proteins ignoring hydrodynamics.[58]

Finally, we present the current OPEP limitations and we sketch some on-going developments or applications. These include for instance understanding the physics behind the difference in the stability of thermophilic and mesophilic proteins, determining the effect of external conditions such as shear flows on the dynamics, kinetics and thermodynamics of non-amyloid and amyloid proteins, and interacting directly on the CG model system using virtual reality to probe the mechanical properties of molecular structures.

Overall, the review gives a state-of-the-art account of the various subjects treated and a well-balanced assessment of the current literature, comparing with previous computational results and experimental data when available.

Design principles of the OPEP model

Granularity

Various levels of granularity for amino acids have been developed ranging from two to six beads, and beyond. The OPEP CG model represents each amino acid by six centers of force: the side-chain is represented by a unique bead located at the center of mass of nonhydrogen atoms in the all-atom side chains of 2250 protein structures with sequence identity < 30%, while atomic resolution is used for the backbone that includes N, HN, Cα, C and O atoms. Proline is an exception represented by all its heavy atoms (Fig. 1A).[20] The disulfide (S-S) bonds can be treated as two non-bonded beads or described at an atomic level using local terms. This OPEP CG strategy was chosen to represent a good compromise between energetic and structural accuracy and chemical specificity, even if this limits the use of large time steps compared to other CG models. Note that the Rosetta fragment assembly Monte Carlo program uses the same level of description in the first step of its hierarchical procedure.[34,35]

Our level of granularity varies from the Martini model with coarse-grained solvent in which the main chain atoms of each residue are represented by a unique bead and, on average, four heavy side-chain atoms are represented by a single interaction center, with the exception of ring-like molecules.[25,26] Our representation varies also from eight other CG models in implicit solvent: (i) Klein's model where three to four heavy atoms are represented by a single CG bead. Most side-chains use one CG bead, except lysine and arginine with a hydrophobic and a hydrophilic site and the tyrosine, phenylalanine, and tryptophan residues represented by two, two, and three beads respectively,[30] (ii) UNRES two-bead model with one unified bead for side-chain and the peptide center, p, located in the centers of Cα-Cα bonds,[31-33] (iii) PRIMO using the Cα, Cβ and a combined CO particle for the backbone and one to several heavy side-chain atoms into CG sites,[42] (iv) PRIME with the N, Cα, and carbonyl C backbone atoms and up to three side-chain beads,[22] (v) Voth's model with a Cα for the backbone and as many as four beads for the side-chains,[36] (vi) ATTRACT with the N and O for the backbone and one or two beads for the side chains,[40] though its most recent version is very similar to OPEP,[41] (vii) AWSEM three-bead model with Cα, Cβ and O,[44] and (viii) PaLaCe with one to three beads for the main non-bonded interactions, combined with atomistic peptide groups and some side-chain atoms.[39] Scientists are also developing CG models for soluble proteins in simplified explicit solvent,[59] or atomistic soluble proteins with CG water models.[60,61]

Optimization Procedure and Analytical expression

There are multiple approaches to derive the bonded and non-bonded potentials.[9-11] The first approach, followed by PaLaCe and PRIMO, uses Boltzmann inversion of conformational probability distributions derived from a static or dynamic protein structure data set. The second one is to derive a CG potential from forces generated by atomistic simulations, referred to as force matching.[36] The third thermodynamics-based approach consists of fitting and predicting free energies such as water/oil partitioning coefficients of the amino acid side-chain analogues (Martini),[11] or density, surface tension (Klein)[30] for the non-bonded interactions and by using the distributions of bonds, bending angles and dihedral angles from the Protein Data Bank (PDB,) to optimize the bonded interactions. Another approach is the factor expansion method where the pairwise potentials of mean force (PMFs) between side-chains are obtained from atomistic simulations, and the torsional, double-torsional, backbone-electrostatic and correlation terms are fitted on quantum-mechanical *ab initio* calculations.[31-33] Other methods for force field derivation and optimization include minimization of relative entropy[38] and simulations to test whether hexapeptides form non-amyloid or amyloid fibrils (PRIME)[22] or proteins fold into their native states.[62]

The latest OPEP version uses a structure/ thermodynamic/PMF approach, since the parameters and analytical forms are trained on bonded and non-bonded distances and angle distributions of native and non-native protein structures, are fitted to reproduce the experimental lowest free-energy conformations and the melting temperatures (T_M) of a small set of peptides, and are derived from all-atom PMF simulations for the interactions between charged side-chains.[23,24]

The OPEP energy function is defined as a sum of local, non-bonded, and hydrogen bonding (H-bond) terms.[15,21,24] All analytical expressions are given in Supplementary Material. The local interactions include bond length, angle bending, and improper and proper dihedral angles. The improper torsions maintain the desired chirality of amino acids, and control the out-of plane motion of the C=O and N-H bonds about the peptide bond. All these terms were modeled on the analytical form of the AMBER[63] force field with an additional term for the Φ,Ψ dihedral angles to render realistic Ramachandran plots.[21,24] H-bonds of backbone atoms are accounted for by two- and four-body potentials, rather than Coulomb interactions. The two-body term for each H-bond is the product of a 10-12 term dependent on the O-H distance by the square of the cosine of the N-H..O angle.[15,21] The four-body term takes the form of the product of two Gaussian functions each monitoring the existence of one H-bond on the basis of distance criteria, and represents a cooperative energy if tight conditions on sequence-separation, Δ, between four residues are verified. If (*i*, *j*) and (*k*, *l*) are the residues involved in the two H-bonds, Δ(*ijkl*) =1 if (*k, l*) = (*i+1, j+1*) (α-helix), or Δ(*ijkl*) =1 if (*k, l*) = (*i+2 ,j-2*) or (*i+2, j+2*) (β-sheets), otherwise Δ is set to 0. These conditions stabilize secondary structures, independently of the Φ, Ψ angles, but also any segment satisfying the conditions on *ijkl*. [15,21,24]

It is essentially the van der Waals potential that has evolved from OPEPv1 to OPEPv5 by distinguishing its form as a function of the center of forces.[15,24] Each OPEP version does not have its own advantages, rather a new version is developed to solve unexpected failures of the previous version. In all versions, we use the 6-12 potential between the backbone atoms and between the backbone and side-chain atoms. In OPEPv3, the van der Waals energy between two side-chains was 6-12 if the interaction is hydrophobic or resulted from oppositely charged amino acids; otherwise an r^{-6} term was used.[20] In OPEPv4, following our work on RNA, the r^{-6} term was replaced by r^{-8} for purely repulsive interactions; otherwise the 6-12 term was replaced by an analytical formulation to limit the energy values of the side-chains at longer distances. We also distinguished 11 side-chain – side-chain interactions depending on their sequence-separation to stabilize α-helices.[23] From OPEPv4 to OPEPv5, we only changed the ion pair interactions from all-atom PMF potentials, characterized by one minimum for the pairs Lys-Asp and Lys-Glu and two minima for Arg-Asp and Arg-Glu.[24]

An overview of the optimization procedure is shown in the flowchart in Figure 2. The OPEPv1 and v2 parameters were adjusted by maximizing the energy gap between the native and misfolded states of six proteins, enabling the folding of 40 peptides of 12-46 amino acids consistent with NMR data in most cases.[12-18] OPEPv3, which used a training and validating set of 13 and 16 proteins to optimize the parameters,[20] was tested on a total of 11 proteins of 12-56 amino acids by MD,[21] REMD[21-22] or metadynamics[64] starting from random or NMR structures. OPEPv4 passed two tests: 17 proteins of 37−152 amino acids remained within 3.1 Å root-mean-square deviation (RMSD) from their native states after 30-100 ns MD at 300 K, and REMD of five peptides with β-hairpin, α-helix or a WW domain, and REMD of the ccβ 51-residue peptide delivered structures consistent with experiment starting from random states.[23]

Finally, the OPEPv5 parameters were tested on structurally diverse proteins differing in the number of charged residues by REMD.[24] These include two 13-residue α-helix and 16-residue β-hairpin peptides and the ccβ-p2 peptide switching from a coiled-coil structure at low T to amyloid fibrils at higher T and concentration. We also verified that MD preserved the structures of proteins with 37-75 residues at 300 K. The final test involved an 85-residue protein with 19 charged amino acids. Running REMD of 24 replicas, each of 300 ns, the predicted T_M is 360 K vs. 336 K experimentally. Overall, the OPEPv5 parameters, by refining the packing of the charged amino acids, impact the stability of secondary structure motifs and the population of intermediate states during temperature folding/unfolding; they also improve the aggregation propensity of peptides.[24]

In OPEPv5, the ε^0 value in kcal.mol^{-1} or well depth at the minimum is 3.89 for the Ile-Ile contact and 4.05 for a Lys-Glu salt-bridge. The ε^0 value, at the minimum, of an intramolecular H-bond is 3.3 and 2.7 kcal.mol^{-1} for (*i, i+4*) and (*i, j ≥ i+5*) interactions vs. 2.7 kcal.mol^{-1} for an intermolecular H-bond. The ε^0 values of the 4-body H-bond terms are 1.4 and 3.6 kcal.mol^{-1} for α-helices and β-sheets, and any segments satisfying the conditions on *ijkl*. The two-body H-bond terms are cut off at an O-H distance of 0.3 nm and an angle N-H-O < 90° and the energy is modulated by a switching function of CHARMM-type[65] from 0.25 to 0.3 nm. All other non-bonded interactions are cut off at 1.6 nm with a switching function starting at 1.3 nm.

Simulation Techniques

In what follows, we review the methods coupled to OPEP. These include the diffusion-controlled Monte Carlo (DCMC),[12,16] the Activation-Relaxation Technique (ART-*nouveau*),[66,67] molecular[21] and Langevin dynamics,[68] replica exchange (REMD)[22] or Hamiltonian (H-REMD) MD,[69] metadynamics,[64,70] simulated tempering (ST),[71] a greedy approach,[72,73] MUPHY,[57] and interactive MD simulations.

DCMC and ART. The basic idea of DCMC is to limit the search to conformations that are thermodynamically accessible from a given conformation in a reasonable time.[12] In principle; one has to determine the nearest saddle point, the energy barrier and the contribution of entropy. Here, we assume that the motion results from the diffusion in (Φ,Ψ) space and the transition time scales as $\Theta^2/D'\eta$, where Θ is the angular deviation of the residue from one state to another, D' is a diffusion parameter and η is related the ruggedness of the energy landscape. DCMC was used to fold 40 structurally diverse proteins. [13-18,74-75]

ART-*nouveau* goes one step beyond by generating non-biased pathways connecting adjacent local minima via exact first-order saddle points and was first developed for hard spheres.[66] Coupled to OPEP, the procedure works as follows (Fig. 1B). First, the system is deformed from its current minimum in a random direction (all atoms for a peptide up to 15 amino acids, and a subset of atoms for larger proteins or oligomers) until the lowest eigenvalue of the Hessian becomes negative, and the system is pushed along this direction while the energy is minimized in the orthogonal directions. Once the saddle point is reached, the system is relaxed to the other side of the barrier and minimized, and finally the move is accepted depending on the Metropolis criterion.[67,76-85]

The advantage of ART, as any activated methods,[86-88] is that it is not sensitive to the energy barriers allowing the system to move rapidly on the energy landscape. Even if ART lacks a proper thermodynamics basis, in contrast to a discrete sampling method,[89] ART-OPEP simulations revealed frustration in the energy landscape of the 60-residue protein A with multiple funnels,[85] consistent with global optimization of a 69-residue protein using basin-hopping and genetic algorithms with Gō energy models.[90] ART-OPEP also predicted conformations of the Aβ21-30 peptide consistent with NMR,[83] and located new minima and mechanisms for amyloid oligomers that were validated experimentally.

Molecular and Langevin Dynamics. Newton's equations of motion are integrated using the velocity-Verlet method. Each main atom has its standard mass while the side-chain beads have a mass equal to the total mass of their atomic constituents.[21] MD runs are performed with a time step of 1 fs or 2 fs using the SHAKE algorithm.[91] The system is first minimized and then heated to the desired temperature. Production runs in the NVT ensemble are performed either with Berendsen thermostat[92] and a coupling parameter $\tau = 0.5$ ps or the Langevin thermostat[68] with a collision frequency $\gamma = 1$ ps^{-1}. Note we found little variation in the equilibrium structures and heat capacity curves of two model monomer and trimer peptides using the two thermostats.[68] Simulations can be performed either in a sphere with reflecting boundary conditions or in a box with periodic boundary conditions. Presently, the non-bonded interactions are updated at each time step, but the code can be easily improved in terms of CPU efficacy by using a multiple time step framework.

REMD and H-REMD. REMD simulations are carried out with a number of replica running in parallel and a temperature range dependent on the system size.[22,93] For instance, we found that 8 replicas for 50 ns, 20 replicas for 600 ns and 22 replicas for 1200 ns are sufficient for the dimers of Aβ16-22[94] and Aβ16-35,[95] and the trimer of Aβ16-35,[96] respectively to reach equilibrium. An exponential temperature distribution is used and exchanges between two consecutive replicas are attempted every 5.0 to 7.5 ps, leading on average to an acceptance ratio of 30-40%.[22] To enhance sampling, it is useful to combine REMD with a Hamiltonian exchange procedure,[97] where we use at the highest temperature several replicas with reduced non-bonded energies.[69]

We assess convergence of the simulations near the physiological temperature by using different time intervals and metrics. These metrics are the distributions of the radius of gyration and end-to-end distances, the secondary structure along the amino acid sequence and the total number of clusters.[95,98] Convergence is also verified by the curves of the heat capacity and conformational entropy using different time windows.[95,99]

ST. In simulated tempering, temperature is a dynamical variable taking discrete values T*n*. Standard ST requires the determination of *a priori* unknown weight parameters to ensure a random walk in T space, the Helmholtz free energies at T*n*.[100,101] Recently, we developed an ST algorithm with on-the-fly weight determination. The weights are self-updated via a trapezoid rule during the run,[71] eliminating the need of trial simulations,[102] or complicated update schemes.[103,104] The advantage of our ST method over REMD was demonstrated using OPEP on Ala_{20} and the Aβ16-22 trimer.[101] The same efficiency is observed in explicit solvent for Ala_{10}, the 20-residue Trp-cage and the 37-residue WW-domain starting from random states and deviating by less than 0.2 nm RMSD from the NMR structure after less than 700 ns (in preparation).

Greedy Algorithm. This method differs from MC and genetic algorithms by growing a chain one fragment after another.[105-107] Our procedure for structure prediction performs a rigid assembly of fragments of 4-residue length by superimposing the first three α-carbons of the new fragment onto the last three of the previously built structure. Our early version used forward (from N- to C-terminal) and backward (from C- to N-) operators to grow the chain.[108,109] Our new version uses a zip operator to start the building process at any randomly chosen position, alternatively adding one residue at each side of the growing structure.[110] At each position, the algorithm keeps 3000 states, the 1000 energetically best OPEP states and 2000 randomly selected ones in the pool of the remaining generated conformations.[110]

Metadynamics. Metadynamics is an advanced technique for enhancing sampling in MD simulations,[70] with widespread applications in material science and chemical reactivity,[111] protein-drug recognition,[112] protein aggregation,[113] and allosteric pathways.[114] Enhanced sampling is achieved by introducing an external, history-dependent bias potential affecting few selected degrees of freedom, usually referred to as collective variables (CVs). The bias is adaptively constructed as a sum of Gaussians deposited along the system trajectory in CV space to discourage the system from revisiting regions that have already been explored (Fig. 1C). If the CVs capture all the slow, relevant degrees of freedom of the system, metadynamics provides a correct estimate of the system free-energy surface. An appropriate application of metadynamics requires the identification of a limited yet effective set of CVs. This may represent an intimidating task when dealing with extremely complex molecular processes. This limitation can be circumvented by combining metadynamics with replica exchange methods.[115] This scheme, where several metadynamics runs are performed using the same CV set at different temperatures and are swapped following a Metropolis criterion, has been applied to assess the quality of the OPEPv3 potential with respect to the all-atom OPLS and AMBER99SB force fields in explicit solvent.[64] Using two β-hairpin and α-helix peptides and an intrinsically disordered peptide, and by comparing the free energy surfaces (FES), the free energy differences between the folded and unfolded states and between the folded state and the transition state, we found remarkable agreement between the OPEP FES at 345-360 K and those using all-atom calculations in explicit solvent at 300 K.[64] This information was used to improve the model during the refinement of OPEPv4 and OPEPv5.

MUPHY. This name refers to a way to embed molecules of generic complexity in a hydrodynamic solvent. The interaction of proteins with the surrounding solvent implies accounting for the hydrodynamic interactions exerted between particles of the macromolecules. As one particle moves in space, it creates a velocity field in the environment that acts on other particles,

therefore generating effective, solvent-mediated interactions. One way to include hydrodynamics is the Brownian Dynamics technique developed by McCammon.[116] However, because hydrodynamic interactions are in principle long-range, the basic technique has a computational cost proportional to cube of the number of particles, so that it is computationally extremely difficult even to handle a small set of proteins in suspension.

The MUPHY software has been developed to study generic biofluidics systems[117] and, in order to specialize to protein suspensions, has implemented OPEP.[57] MUPHY handles the dynamical evolution of generic fluids and particles treated via the dual mechanism, a mesh-based Lattice Boltzmann method for fluids, and specialized versions of MD for particles.[117] Indeed, a powerful alternative to McCammon's method is provided by explicitly solving for the evolution of the solvent, as encoded by the Navier-Stokes equations, by using the Lattice Boltzmann (LB) numerical method.[118] In this approach, the coupling between fluid and particles takes place via specifically designed kernels based on kinetic modeling, significantly distinct from methods based on macroscopic hydrodynamics. Such a methodology is genuinely multiscale as it entails different levels of physical description (such as field-based for the fluid and particle-based for the proteins) within a single unifying framework.[119]

In LB, the solvent is described via the "populations" $f_p(\vec{x}, t)$ representing the probabilities of finding solvent molecules at a given position $\vec{x}$ and time t and moving along a discrete direction p. The populations, represented over a mesh, evolve as: $f_p(\vec{x} + \vec{c}_p, t+1) = \omega f_p^{eq}(\vec{x}, t) + (1-\omega) f_p(\vec{x}, t) + \Delta f_p(\vec{x}, t)$, where $f_p^{eq} = w_p n\left[1 + 3\vec{u}\cdot\vec{c}_p + \left(9(\vec{u}\cdot\vec{c}_p)^2 - 3u^2\right)/2\right]$ is the discrete Maxwell-Boltzmann equilibrium, associated to the weight w_p and discrete speed $\vec{c}_p$, with $n = \sum_p f_p$ and $\vec{u} = \sum_p \vec{c}_p f_p / n$ being the fluid density and velocity, respectively. The term Δf_p is proportional to the drag force exerted by the fluid on the particle and vice versa, being a bidirectional of fluid and particles according to the action-reaction principle. Thus the drag force, together with a stochastic force, is included in the particles' evolution besides the mechanical forces stemming from the OPEP force field.

The fluid-proteins concurrent evolution can be specialized to hybrid situations, such that the hydrodynamic interactions are decomposed into intra- and intermolecular components, the first ones evaluated analytically and the second ones handled via an under-resolved version of the LB fluid. Such a decomposition proves further advantageous in terms of CPU efficiency as it ideally balances the cost of computing intra- and intermolecular hydrodynamic interactions.[120] Finally, MUPHY is fully parallel and distributes the computational load on multiple cores or multiple Graphical Processing Units (GPUs).

Interactive Molecular Dynamics (IMD). Using our previously developed MDDriver software library,[121] we rendered the CG simulation engine interactive by implementing the IMD network protocol and interactive steering modules. Using a TCP (transmission control protocol) network socket, any IMD-protocol-aware frontend is able to connect to the running OPEP/Hire-RNA simulation engine and inject additional user-forces to drive the experiment. For initial validation, we used both our custom UnityMol[122] and the more widely distributed VMD[123] frontends. Simulation and frontend may be run on the same machine or remotely to optimize performance and fluidity.

OPEP applications

Though OPEP was used to explore large-scale motions, such as the pathways from the *holo* to the *apo* states of two EF-hand proteins,[124] or the conformations of 8-20 amino acid loops,[125] we focus here on the following timely topics: self-assembly of amyloid proteins, structure prediction of linear and disulfide bonded cyclic peptides, thermodynamic properties of RNA, and protein dynamics in a crowded environment with hydrodynamics. Table 1 gives a summary of the different systems simulated with OPEP indicating the methodology adopted and the total time lengths.

Understanding the self-assembly of amyloid proteins

Alzheimer's disease (AD) is marked by atrophy of cerebral cortex and loss of cortical and subcortical neurons. Autopsy reveals accumulation of amyloid plaques and numerous neurofibrillary tangles made of filaments of the phosphorylated tau proteins. The major constituents of plaques are made of the amyloid-β (Aβ) peptides of 40 and then 42 amino acids formed from the amyloid precursor protein via the actions of the β- and γ-secretases.[126] But many truncated variants, such as Aβ1-30 and Aβ1-26, and Aβ with proteolytic removal of D1 and A2 and subsequent cyclization of E3 to a pyroglutamate, have been detected by mass spectrometry in human AD brains.[127,128] The human Aβ1-42 sequence, designated Aβ42, is DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGV VIA with a charged N-terminus (A1-K16) and two hydrophobic patches L17-A21 (central hydrophobic core, CHC) and A30-A42 (C-terminus) separated by a hydrophilic patch E22-G29. Despite many clinical trials, drug after drug has failed to slow the progression of AD for three main reasons.[49]

While the experimental sigmoidal kinetics of amyloid formation with a lag-phase can be accounted for by means of primary classical nucleation theory (CNT) and/or secondary (fragmentation or lateral) nucleation processes,[129,130] we lack information on the topology, structure and size of the primary nucleus (N*).

Secondly, though the low molecular weight (LMW) Aβ40/42 aggregates are the most critical players in the pathology, we have little information on their structure, rate and extent of formation. Due to their high aggregation propensity, the LMW oligomers are not amenable to solution nuclear magnetic resonance (NMR) and X-ray crystallography. As a result, only low-resolution structural data from circular dichroism (CD),[131] Fourier transform infrared spectroscopy (FTIR),[132] ion-mobility mass spectrometry (IM-MS),[133] solid-state NMR,[134] pulsed hydrogen-deuterium exchange coupled to MS,[135] transmission electron (TEM) and atomic force microscopies (AFM) are available.[136] The final Aβ40/42 products are insoluble and only solid-state NMR models are available. While fibrils of synthetic Aβ40/42 peptides display perfect U-shaped forms with β-strands spanning the CHC and the C-terminus, and the N-terminus disordered, fibrils of AD-brain derived Aβ40 peptides show deformed U-shaped states and, remarkably, the structure varies from one patient to another.[137] A common feature of all fibrils is the inter-digitation of the side-

chains, the so-called steric zipper.[138]

Thirdly, though the general consensus is that drugs are given too late,[49] we lack the structures of Aβ40/42 peptides with known inhibitors of aggregation and toxicity, paving the way for the design of specific drugs with the highest affinities for Aβ40/42 oligomers. Overall, OPEP simulations have played a significant role to complement experiment on five aspects.

(1) Independently of the force field and the sampling method, self-assembly starts by a hydrophobic collapse and the formation of molten oligomers, which is modulated by the degree of hydrophobicity of the peptide. Then, the H-bonds drive the system to highly flexible and transient β-rich oligomers.[139-141] These β-rich oligomers have various topologies, and we were the first using OPEP simulations to (i) observe assemblies with various sheet-to-sheet pairing angles[79,81,94] that were confirmed by structures of macrocyclic β-sheet mimics,[142] and other force field calculations,[43,140,141] (ii) evidence β-barrels (Fig. 3A)[79,143] that were validated by the X-ray structure of a toxic hexamer of a 11-residue amyloid peptide (Fig. 3B),[144] (iii) predict formation of antiparallel double stranded poly-L-glutamine nanotubes with 22 residues per turn (Fig. 3C),[145] reminiscent of the water-filled model proposed by Perutz,[146] and (iv) identify reptation moves of the β-strands in the late steps of aggregation[78] that were validated by FTIR[147] and atomistic simulations.[140,148]

(2) The aggregates of 7- to 20-mers of GNNQQNY, NNQQ, Aβ16-22, KFFE and NHVTLSQ using OPEP are mostly amorphous and consist of a heterogeneous ensemble of β-rich states.[99,149-153] In all systems, the transition at the melting temperature involves a change in the distributions of oligomer and β-sheet sizes, but mixed parallel/antiparallel (P/AP) β-strands dominate.[99] This β-strand mismatch, observed with various force fields,[38,139,154] provides strong evidence that one limiting-factor for fibril formation is the transition from mixed P/AP to fully P or AP strands. This was confirmed by bias-exchange metadynamics of 18Val8 and 18Aβ35-40 peptides in explicit solvent, where the crossing of the highest free energy involves the transition from mixed P/AP to P β-strands that can be accompanied by the formation of the steric zipper.[113,155]

(3) A fundamental question pertains to the size of the primary nucleus, N*. Recently, two atomistic simulations in explicit solvent and one OPEP simulation have provided insights into N*. In the first study, an effective nucleus size on the order of 14 was proposed for 18 Val8 peptides by metadynamics.[113] In the second atomistic study, the aggregation of 16Aβ37-42 peptides was investigated by REMD,[156] and the population of 4-5 fully P β-strands, consistent with the fibril structure, was 1-2% at 300 K. Whether N* is around 15 for Aβ37-42 as for the Val8 system cannot be determined, due to finite-size effects and the fact that fibril formation is under kinetic and not thermodynamic control as evidenced experimentally[157] and by Langevin dynamics of a mesoscopic model.[48] Using unbiased MD-OPEP, we investigated the onset of aggregation in a 20-mer of GNNQQNY.[158] Running 16.9 μs at 280 K and 300 K, we showed that aggregation follows the CNT and N* is 4-5 at 280 K and 5-6 at 300 K. The kinetics of growth cannot be fully described by the CNT, however, because there are important rearrangements after the nucleus is formed, as the aggregates attempt to optimize their organization.[158]

OPEP simulations do not systematically show fibril formation. One reason found for the peptide spanning the residues 144−153 of the prion protein is that oligomerization is not thermodynamically favorable, in agreement with a turbidimetric experiment.[23] Another reason is the presence of a proline which can either destabilize the β-strand conformation of the monomer and totally prevent aggregation, or reduce the packing of β-sheets rendering fibril formation a slow process,[80,150] consistent with experiments.[159] Overall, many factors, in addition to pH and T, modulate amyloid formation (N* size, fibril topology and lag-phase) ranging from the energy landscape of the monomer,[43,48,160] the entropy of the loops[161] or the intrinsic disorder of the whole peptide,[162,163] to the supersaturation of the protein solution.[164,165]

(4) The solution NMR structure of the Aβ42 monomer reveals weak β-strand propensities at the CHC and the residues I31–V36 and V39–I41, and turns at D7–E11 and F20–S26.[166] NMR relaxation data reveal that Aβ42 is more rigid at the C-terminus than Aβ40. IM-MS reports a collision cross-section of 1256 Å^2 for Aβ42 dimers.[133] Using different preparation methods, CD leads to a β-strand between 12% and 25% and an α-helix between 3% to 9% at 295 K, pH 7 and day 0, i.e. for an heterogeneous ensemble of oligomers.[131,133] Remarkably, the Aβ40-D23N peptide forms fibrils with in-register antiparallel and parallel β-sheets under quiescent and strong agitations, respectively.[157]

To get insights into Aβ flexibility, we determined the free energy landscapes of the monomers[167] and dimers[168] of Aβ40, Aβ42, and with D23N using H-REMD-OPEP. We found that if the three monomeric alloforms are mostly disordered, in agreement with experimental data[166] and confirmed by all-atom simulations,[169-171] they display distinct morphologies. Aβ42 and Aβ40-D23N have higher β-strand propensities at residues 30–42 than Aβ40. D23N changes the Aβ40 structures; the residues 1–16 becoming more independent of the rest of the protein,[167] which may explain in part why the kinetics and the final products vary between Aβ40 and Aβ40-D23N under quiescent agitation. Our results on the dimers showed that Aβ42 has a higher propensity than Aβ40 to form β-strands at the CHC and residues 30-42, explaining the higher Aβ42 aggregation kinetics.[168] In none of the systems we observed any parallel β-sheet structure between the two CHC's. D23N impacts the free energy landscape by increasing the population of states with higher β-strand propensities at the C-terminal and antiparallel β-sheet between the two C-termini, and this motif could be important in the nucleation of Aβ40-D23N toward parallel β-sheets. Our results also revealed many configurations stabilized by N-terminal interactions[168] that were observed by single-molecule atomic force spectroscopy[172] and all-atom REMD simulations.[173]

(5) Based on the microcrystal structure of Aβ16-21 fibrils with the dye Orange G, Eisenberg designed compounds that reduce toxicity by preventing fragmentation of the Aβ42 fibrils without binding to the oligomers.[174] Despite many experimental attempts, scientists have not succeeded to provide the structures of Aβ40/42 monomers or Aβ40/42 oligomers with inhibitors. Using a shorter fragment, Segal solved the NMR structure of NQTrp bound to the Aβ12-28 monomer, revealing three dominant binding sites between NQTrp and the Aβ18-21 region.[175] As a first step toward understanding the interaction of Aβ oligomers with NQTrp we focused on the Aβ17-42 peptide also found in AD plaques and used a multiscale procedure.[96] Our extensive

OPEP-REMD simulation of the Aβ17-42 trimer, followed by all-atom docking of five molecules on the most populated Aβ structures, showed that NQTrp is a more favorable inhibitor than EGCG, 2002-H20 and resveratrol. In agreement with the NMR structure of NQTrp/Aβ12-28,[175] NQTrp binds to Aβ through the side chains of F19 and F20 and the main chain atoms of F19-E22. Our simulations reveal, however, many transient binding sites (Fig. 3D),[96] consistent with all-atom REMD of the Aβ1-42 dimer with 2NQTrp,[176] indicating that the design of more efficient drugs targeting the Aβ42 dimer is not an easy task.

Fast and Accurate 3D Peptide Structure Prediction

Peptides have regained considerable interest as they represent alternative ways to design therapeutics, vaccines or molecular probes. However, fast and accurate peptide structure determination remains a long-standing goal in structural biology and peptide engineering.[34] Pep-Fold is an innovative approach aimed at *de novo* structure prediction of linear and disulfide bonded cyclic peptides with 9-52 amino acids (aa).[177-178] Pep-Fold relies on a Hidden Markov Model derived structural alphabet (SA) of 27 letters to describe proteins as series of overlapping fragments of four aa.[179] The SA letters can be assimilated to a generalized secondary structure, extending the number of states from 4 (α-helix, coil, turn or bend, and β-strand) to 27, but not all transitions are possible between two consecutive letters. The Pep-fold procedure consists in three steps. First, Pep-fold predicts a limited set of SA letters at each position from the sequence, and then performs a progressive assembly of the prototype fragments associated with each selected SA letter using our greedy algorithm[108-110] driven by OPEP. As Pep-Fold uses a rigid assembly, we found necessary to smooth the OPEP side chain - side chain potential.[177] The third step refines the CG models by Monte-Carlo before generating all-atom models and performing a clustering of all models returned by the simulations.

Pep-Fold1 efficiency was shown on 24 linear peptides of 9-25-aa in aqueous solution and neutral pH by predicting lowest-energy states with a mean 2.5 Å RMSD from the NMR rigid cores (RC, excluding the flexible parts).[110] Pep-fold2, which revisited the prediction of the SA letters from the sequence and considers several filters to generate a variety of SA trajectories, was tested on peptide lengths up to 36-aa.[178] The server allows the biologists or chemists to define S-S bonds or any residue-residue contact. Using 34 peptides with one to three S-S bonds, the best Pep-Fold2 models had a RMSD of 2.7 Å from the full NMR structures. Using 37 linear peptides, Pep-Fold2 located lowest-energy states with a 3 Å RMSD from the NMR RCs. We also showed the gain in the identification of the native state by filtering the Pep-Fold2 models using the backbone proton chemical shifts easily available from 2D NMR.[180]

Finally, Pep-Fold2 was compared to the state-of-the-art Rosetta program on 56 peptides with 25-52-aa.[180] Rosetta starts sampling with a CG model and fragment assembly MC, and then through successive steps, selects models for all-atom refinement.[18-19] By using a total of 200 Rosetta and Pep-fold runs for each peptide, and a new Binet-Cauchy (BC) score,[181] the mean BC score of the best models (lowest RMSD with respect to NMR) generated by Rosetta and Pep-fold are 0.83 and 0.87 (in preparation). While Rosetta generates high quality models (BC score > 0.9) for 34 targets vs. 29 for Pep-fold, suggesting that Pep-fold could benefit from an all-atom sampling refinement, Pep-fold generates near-native or native states for 53 peptides vs. 49 for Rosetta (BC score > 0.6). Fig. 4 shows the predicted structures of four peptides.

Pep-Fold is freely available as a web server[110,178] and has proven to be very useful by many scientists for different applications that can be broadly classified into six categories. The first application is predicting the conformations of protein fragments.[182] Structural characterization of the C-terminal 27-aa tail of HIV gp41 remained relatively limited and contradictory. Pep-Fold tail models showed conserved α-helix structures despite significant sequence variations among diverse clades, and this is supported by CD.[183] 3D models of the N-terminal 20-aa of human cytochrome c and several cytochrome c_2 variants from *R. capsulatus* were also generated by Pep-Fold and helped understand why the insertion of an alanine residue between Phe11 and Cys15 and substitution of residues Glu8 and Glu10 are critical for heme attachment by the mitochondrial protein holocytochrome c synthase.[184] Pep-Fold was also used to generate protein N- and C-terminal conformations.[185] Certain immune-driven mutations in HIV-1, such as those arising in $p24^{Gag}$, decrease viral replicative capacity. In HIV-1 subtype B, the $p24^{Gag}$ M250I mutation is a rare variant, while in subtype C, it is a relatively common minor polymorphic variant (10 to 15%). The structural implications of M250I were predicted by Pep-Fold to be greater in subtype B versus C, providing a potential explanation for its lower frequency and enhanced replicative defects in subtype B.[186] In addition, Pep-fold was used to model protein loops[187] or protein linkers.[188,189] A study on the linkers in a new class of modular alpha-amylases showed that the Pep-fold conformations are diverse, but match the data obtained from small-angle X-ray scattering.[189]

The second application is related to protein–peptide interaction in general, fundamentally important for signal transduction, transcription regulation and protein degradation.[190-194] Wu used Pep-Fold to generate the structures of 13 peptides of 20-aa as initial structures for short MD and showed a very good correlation between the experimental and the calculated MM-PB/SA binding free energies for the peptides interacting with the vascular endothelial growth factor A.[191] Kumar used the Pep-fold conformations of several 20-aa peptides to explore their binding mechanisms to calmodulin,[192] Chopra used Pep-Fold for the design of a peptide able to bind to Bacillus anthraxis toxin-antitoxin module,[193] while Stegman used Pep-fold prior to docking onto peptidyl-prolyl cis/trans isomerase PPIL1, a component of the human spliceosome.[194]

The third domain of application is the design of immunogenic peptides. Peptides play many roles in immunology, yet none are more important than their role as immunogenic epitopes driving the adaptive immune response against infectious disease.[195] Peptide epitopes are mediated primarily by their interaction with major histocompatibility complexes (T-cell epitopes) and antibodies (B-cell epitopes). In this context, Wingren reported the first detailed analysis of antibody–peptide interaction characteristics, by combining large-scale experimental peptide binding data with the structural analysis of eight human recombinant antibodies and numerous peptides using Pep-fold, targeting tryptic mammalian and eukaryote proteomes.[196]

Another application concerns antiviral peptides (AVP) and vaccines. Pep-Fold contributed to the design of peptides inhibiting *in vitro* the Influenza A virus[197] or other viruses[198,199] and is now defined in AVPdb, a server allowing the design of AVP.[200] Pep-fold was also used to design a DNA vaccine against human papillomavirus causing cervical cancer.[201]

Although OPEP has been optimized for aqueous solution, Pep-fold has been used on peptides in an apolar milieu,[202-204] and in particular on antimicrobial peptides (AMP) regarded as one of the most promising alternatives to antibiotics affected by resistance mechanisms. Using *in silico* predictions including Pep-Fold and *in vitro* assays led to the discovery of potential AMPs with high activity and low toxicity from the entire human genome.[204]

Finally, Pep-fold has been found useful in understanding the solvent-dependent CD spectrum of a 24-aa peptide corresponding to the tubulin-binding site of the neurofilament light subunit[205] and the effect of gold nanoparticle conjugation on peptide structure and dynamics.[206] Pep-fold has been used in various design situations: new molecules for induction of bone formation,[207] peptides binding lipids,[208] peptides coating carbon nanotubes,[209] and a peptide-based Hsp90 inhibitor leading to a novel anticancer agent[210] that will enter preclinical trials conducted on patients with breast cancer, prostate cancer and skin cancer. To date, there is only one case of conflict between Pep-Fold and *in vitro* results. While Gautam designed 15-aa peptides with coil-turn CD, Pep-Fold predicts β-hairpins, but the experimental conditions (pH and ionic strength) are not reported.[211]

A framework for RNA and DNA coarse-grained models

In many vital cellular processes, especially in regulatory functions related to transcription and translation, proteins interact with nucleic acids. We recently developed a nucleic acid CG model, called Hire-RNA/DNA, by following the physical principles used for OPEP, in order to better understand the thermodynamics and dynamics of RNA/DNA.

The most widely used all-atom force field for nucleic acids is undoubtedly AMBER with ff99 achieving a good agreement with experiment for DNA double helices.[63] The parameters are, however, constantly adjusted to better represent non-canonical structures in loops and bulges,[212] and a new parameterization obtained by reproducing known thermodynamic and kinetic measurements of RNA monomers and dimers was just reported allowing *de novo* folding of three hyperstable RNA tetraloops to 1–3 Å RMSD from their experimental structures.[213] Folding a single stranded RNA free of any biases remains, however, a computer challenge for nucleotide (nt) lengths > 20.[61]

Different strategies are applied to go beyond all-atom simulations and they can be organized into three categories:[214] homology modeling, hybrid and *ab initio* methods. Homology modeling works well, if one can find a good template in the NDB, but this is typically not the case for single stranded RNAs.[215,216] Hybrid methods based on knowledge-based energy functions vary from fragment reconstruction (MC-Fold and MC-Sym),[217] fragment assembly (FARNA)[218] to multiscale approaches relying on 2D structure predictions, CG 3D models based on the fragments selected from the NDB followed by all-atom minimization,[219] or junction topology prediction and graph modeling followed by all-atom refinement.[220] However, the best 2D structure prediction algorithms reach only 60% accuracy.[221] Another bottleneck is the low population of non-canonical Watson-Crick (WC) base pairs in the NDB, and the prediction of pseudo-knots and junctions. This problem is also faced by *ab initio* CG force fields, built from atomistic simulations[222,223] and electronic structure calculations[223] or by using experimental data to assign parameters,[224-226] e.g., iFold,[225] or derive statistical potentials,[227-229] e.g., NAST[227]. Most methods were recently evaluated in RNA-Puzzles and predicted a dimer of 46-nt with a RMSD from 0.34 to 0.69 nm, a 100-nt square of double-stranded RNA with a RMSD from 0.23 to 0.36 nm, and a 86-nt riboswitch domain with a RMSD from 0.72 to 2.3 nm.[35]

Following OPEP, our CG model has an energy function derived from physical intuition and parameters based on known structures. Among all CG models, our CG representation has the highest resolution, with an explicit representation of the heavy atoms of the sugar-phosphate backbone (P, O5', C5', C4' and C1'), one bead for pyrimidine bases (C and U) and two beads for the purine bases (G and A), see Fig. 5. For comparison, the NAST,[227] iFold[225] and Xia[228] models have 1, 3 and 5 beads, respectively. Note OH is not treated explicitly and therefore the distinction between RNA and DNA in our models is done solely on different equilibrium angles and torsions, with some angles and torsions allowed for one molecule but inaccessible to the other, and different base pairs, with DNA making only a subset of all RNA base pairs.[230]

The particles in Hire-RNAv1[231] interact via standard local terms for covalent bonds, bond angles and dihedral rotations, an electrostatic repulsion between phosphate groups, a modified Lennard-Jones potential for long range van der Waals interactions as used in OPEPv4-v5 and base pair terms. Base pairing is the most crucial interaction and we treat it with more detail than all other top-down *ab initio* models. A two-body term, as in OPEP, depends on the relative distance and angles formed by two base beads interacting through their WC sides, see Westhof's classification.[232] All the bases can form, and not just A-U and G-C, with different strengths. A three-body repulsive term prevents different bases to simultaneously interact, even though transient multiple pairs can form, and a four-body term, as in OPEP, helps stabilize pairs of consecutive bases.

With Hire-RNAv1, we folded two RNA of 26- and 40-nt into hairpins from fully extended states by MD (Fig. 5). Running REMD, we showed that the NMR configuration is the most populated structure at low T.[231] In a second study, we slightly modified the form of the two-body and four-body terms for base pairs and with Hire-RNAv2, we examined the assembly of DNA and RNA duplexes by REMD.[233] For the two RNA and the DNA consisting of 36- and 24-nt, we calculated the heat capacity curves and found one transition from an assembled state (RMSD of 0.18-0.26 nm with respect to the crystal structure) to disassembled states (Fig. 5). In principle, RNA could fold on itself and form a hairpin, but this is not a favorable free energy state for our nt sequences. The melting temperatures we find for the three systems deviate at most by 17 K from the values obtained by the HyTher algorithm, a reference in the field.[234] Overall, the same energetic parameters perform well for single- and double-stranded systems of 40-nt, and based on our algorithm generating an all-atom model from a CG state (Fig. 5), we

showed the equivalence of MD results using AMBER ff99 with explicit ions and water starting from our REMD-predicted and the experimental structures for both RNA and DNA.[233]

Simulations of protein suspensions with hydrodynamics

In the last few years macromolecular crowding has been the subject of several investigations since it has crucial implications on cell functioning.[235,236] There is increasing evidence that macromolecular crowding exerts large effects on the protein mobility, association and stability.[237,238] It is generally thought that crowding serves as a means of confining proteins in space, where enzymatic activity is undertaken. Also, data suggest that at high concentrations proteins non-specifically enhance association rates, with *in vivo* and *in vitro* rates and equilibria differing by orders of magnitude. Macromolecular crowding also affects hydration structure and dynamics,[239] and protein conformational change.[240] Evolution has fined-tune microtubule-based motor proteins to deliver cargoes rapidly and reliably throughout the cytoplasm by having molecular properties that prevent them from forming traffic jams.[241]

Hydrodynamic and excluded volume interactions are likely the two main factors that account for the large reduction of protein diffusivity in crowding conditions.[242] While hydrodynamics interactions do not alter the equilibrium distribution of states of a system, they potentially affect the local dynamics as well as the escape from metastable states characterizing the spatially and energetically heterogeneous crowded system. For example, hydrodynamics has a primary role on the transport properties, as for the translational and rotational diffusivities, and in general the importance on the dynamics of suspensions is well known. Along with simulations, diffusional data can be accessed experimentally via quasi-elastic neutron scattering, single molecule tracking, fluorescence correlation spectroscopy and fluorescence-recovery-after-photobleaching,[243] so the multiscale approach can be directly compared with *in vitro* and *in vivo* data.

One large-scale MUPHY/OPEP application is offered here to illustrate the potentiality of the coupling of the CG force field with hydrodynamics interactions. A large system composed of 17576 Rat1 yeast proteins in solution is simulated at 300 K for 30 ns. Rat1 is a 666-aa protein that functions primarily in the nucleus and plays an important role in transcription.[244] Altogether we consider a system of 70 million particles, each Rat1 having 4013 particles. To account for the solvent, a hybrid LB/Brownian Dynamics scheme with a time step of 1 fs was used on the Titan supercomputer, exploiting 17576 GPUs in parallel.[120] The highest volume fraction considered (40%) emulates the crowding conditions found in the cytoplasm, typically with 20-30% of the cytoplasmic volume occupied by proteins, nucleic acids and other macromolecules. As a result, the distance between proteins is comparable to the size of the proteins.[235]

During the evolution of the Rat1 suspension, proteins move and tumble together. Fig. 6 shows the typical protein configuration in the suspension, in particular by highlighting the hydrodynamic "bubble" that each protein carries along, representing the isosurface of constant velocity surrounding proteins. Each bubble is further distorted and connected with those generated by neighboring proteins. Visual inspection of the flow streamlines reveals that, as proteins move, they generate a substantial accompanying drain on the solvent. Even at physiological concentrations, the streamlines travel mostly undisturbed over several protein sizes, that is, distant proteins effectively experience solvent mediated mutual interactions.

Crowding is generally thought to induce sub-diffusive and slow dynamics on the short timescale and diffusive dynamics at longer times.[245] In principle, the coherent long-ranged organization of the solvent flow field can act on the suspension as a lubricant, in order to facilitating the protein motion. On the other hand, the hydrodynamic field can interfere with the protein motion, since viscous dissipation can drain momentum away from the suspension.

In the following, we illustrate simulation data for the Rat1 suspension for the translational diffusion coefficients that pertain to the short-time dynamics (10 ns). The diffusion coefficient is shown in Fig. 7 and is evaluated via the integration of the protein center of mass velocity autocorrelation function. For the sake of comparison, experimental data on the translational coefficient obtained by quasi-elastic neutron scattering for the bovine *serum albumin* protein are also shown.

On the considered timescale, diffusion shows anomalous behavior, in the sense that the effective mean square displacement does not scale linearly with time but is rather subdiffusive (data not shown). As Fig. 7 shows, the translational diffusion coefficients provide similar, although systematically larger values than the experimental ones, probably related to the larger temporal scale accessed by the simulation as compared to the one pertinent to the scattering spectra ($3.5\text{ns} < \tau_{expt} < \square 5$ ns).[246] This temporal window exceeds the hydrodynamic one, which arises from the propagation of vorticity over the protein linear size (~ 100 ps) and slightly slows down the protein self-diffusion. The drop of the translational diffusion coefficient for volume fractions comprised between 10% and 30% signals the onset of caging effects on account of steric interactions. At larger volume fractions the diffusivity drops to one order of magnitude smaller than that at the high-dilution value, with proteins possessing some residual mobility. Analysis of the trajectories of the macromolecules, in particular by focusing on the intermolecular contacts, highlights that during their erratic encounters proteins display structural heterogeneity, with several non-specific and specific interactions. Heterogeneity reflects the presence of small clusters made of two or three proteins, together with the presence of isolated (singlet) proteins. At the same time, the simulations show a certain dispersion of the diffusion coefficient increasing with the level of volume fraction.

On the structural side, crowding conditions are usually considered to stabilize protein structures, due to the concomitant presence of specific and non-specific intermolecular interactions. At the same time, entropic effects due to the suppression of available space can destabilize the macromolecular scaffolding. Simulations of the Trp cage protein in cavity-like environments highlighted the thermodynamic shifts induced by polar (destabilizing) vs. non-polar (stabilizing) interactions between the protein and the confining surface.[237b] It was also reported that, when confined in a reverse micelle, atomistic fluctuations are reduced.[237c] Fig. 7 reports the Rat1 RMS fluctuations from their initial states. As crowding induces a larger number of intermolecular contacts, a mild destabilization of the proteins takes place, in agreement with recent experimental[247] and

computational studies[237d] on other proteins. However, the Rat1 system shows that the enhanced fluctuations do not induce substantial departures from the initial structures and are in the conventional range of values. Analysis of the whole ensemble of proteins (R_g inset, Fig. 7) shows that a small heterogeneity of structures is detected at the highest volume fraction considered, indicating that the presence of small clusters in the suspension does not trigger partial unfolding of the molecules.

This simulation of unprecedented size made possible by a multiscale methodology, bringing together the OPEP CG and a consistent treatment of the hydrodynamic interactions, is a first step towards simulating the real physics in the cell. The coupling of proteins and solvent reveals the interplay between specific and non-specific intermolecular interactions, and the role of hydrodynamic forces on the structural and diffusional properties of proteins in crowded environments (in preparation).

OPEP Limitations

The OPEP force field has several limitations, as is the case for any other CG or all-atom force field. Some are easy to alleviate and are the subject of on-going projects. Other issues are more delicate. It is important to be aware of the strengths, weaknesses, and limitations so as to use OPEP for the right questions.

Buffer Conditions and pH. We can block the N-terminus by an acetyl group (CH_3-CO) and the C-terminus by an NH_2 group, or block one end while the other is in its zwitterion form. Alternatively, the proteins can be in their zwitterion forms. The charged residues are parameterized for neutral pH. This means the N-terminus is NH_3^+, the C-terminus is CO_2^-, the Arg and Lys residues are treated as positively charged (NH_3^+), the Glu and Asp residues are treated as negatively charged (CO_2^-), and the His residues are neutral. So, OPEP can be safely used only in the pH range of 6-7. The pH effect can be illustrated on the Aβ12-24 peptide forming amyloid fibrils very rapidly at pH ≤ 5 and very slowly at pH 8.4 using the same *in vitro* conditions.[248] Another aspect to be known is that the non-bonded parameters have been parameterized in "normal" aqueous solution. So we expect deviations with experiments at high ionic strengths or buffers made of H_2KPO_4 and adjusted with H_2SO_4, or with DMSO.

Non-natural Amino Acids and Small Molecules. OPEP has been extensively tested for the 20 standard natural L-amino acids, but D-amino acids can be used as well. The three-proteinogenic amino acids occurring in all kingdoms of life, selenocysteine, pyrrolysine and N-formylmethione, cannot be treated. S-S bonds can be treated at a bead level using a 6-12 potential when folding peptides with Pep-Fold, or described at an atomic level using standard local terms and Amber parameters for the bond angles and dihedral angles. While the N-methylated amino acids have been parameterized using quantum mechanics calculations,[129] many non-canonical amino acids cannot be used. These include peptoids,[249] β-amino acids for designing antibiotics where the amino group is bonded to the β carbon rather than the α carbon,[250] γ-amino acids where the amino group at the third carbon atom is after the carboxyl carbon atom, such as γ-aminobutyric acid the most important neurotransmitter in the central nervous system,[251] side chains with cyclo-hexyl groups to design inhibitors of Aβ40/42 aggregation,[252] and Aβ40/42 with a pyroglutamate.[127] For all these systems, it is now straightforward within the framework of the OPEPv5 code to derive effective potentials and forces from all-atom simulations.[24] Generating OPEP parameters for small drugs is out of reach since the explicit representation of H-bond donors or acceptors is an essential requirement of the rule of "five",[253] but we can imagine a multi-resolution method on the fly with an all-atom representation of the protein and the drug in the regions of interest.

Effective Time Scale and Long-time Dynamics. In OPEP, the solvent contributions are treated through effective non-bonded interactions and a single bead replaces most side chains. So why does OPEP use 2 fs for integrating the equations of motion? For comparison, Deserno uses 100 fs with a resolution model of four beads,[46] Klein 25 fs, Martini simulations 20-40 fs,[11] Shea 10 fs with a three-bead model,[43] PaLaCe and UNRES 5 and 4.9 fs,[39,33] PRIMO 4 fs,[42] and Voth uses 2 fs.[36] The reason is that OPEP explicitly represents the N-H bond and its vibrational mode[254,255] at 3600 cm^{-1} which limits the time step for conserving the total energy in the NVE ensemble. The second reason is that augmenting the time step to 3-4 fs by changing the mass of the hydrogen atom would introduce dynamics perturbations compared to all-atom simulations.

Using an all-atom force field in implicit solvent, Rao showed that folding of three peptides is accelerated by two orders of magnitude.[256] The relationship between the OPEP-MD simulation time and the experimental time varies with the system. Poly-L-alanine and poly-L-proline have the same number of degrees of freedom in OPEP and in an all-atom model, while poly-L-valine has not. The implicit solvent and CG side chains do not affect the motions uniformly and even if dynamics were investigated by Langevin simulations, we would miss important dynamical contributions as a result of the momentum transfer that would occur through the solvent. Overall, our experience suggests a 5- to 10-fold speed-up compared to all-atom MD in explicit solvent, and the OPEP-generated dynamics cannot totally reflect the dynamics in explicit solvent. The OPEP-MD time is therefore smaller than the CG-DMD time[37,38] and the Martini-MD[11] time, preventing the self-assembly of large oligomers of Aβ42[257] or di-phenylalanine[258] peptides using reasonable computer time.

Short-time Dynamics. Using multiple MD trajectories of 30-100 ns at 300 K, the RMS deviations of all proteins are 0.15 nm higher than in all-atom MD simulations in explicit solvent,[23-24] though all-atom force fields do not describe similarly the folded[259] or unfolded[260] proteins. While the secondary structures are well preserved and display RMS fluctuations consistent with NMR, the loops display higher mobility. This results from their intrinsic flexibility and the simplified side chains, but more importantly from the absence of interactions between the loop residues and the solvent. As is the case for all-atom and CG force fields, OPEP has limitations in describing the vibrational modes.[255,261,262]

Thermodynamic Properties. The heat capacity and the melting temperature play a major role in relating microscopic and macroscopic properties of proteins. Their accurate predictions by

simulations remain a significant challenge due to the complex and dynamic nature of protein structures, their solvent environment, and conformation averaging. Constructing the heat capacity curves, C_V, as a function of T from REMD, ST or metadynamics simulations is an easy task using PTWHAM[263] or MBAR.[264] OPEPv5 was optimized to fit the experimental T_M of a β-hairpin (297 K).[24] Using the same parameters, the monomer of the ccβ-p2 monomer has a calculated T_M of 275 K and a α-helix content of 70% fully consistent with the Agadir program. Experimentally, the peptide displays a α-helical CD signal at 277 K, suggesting a T_M within 290-300 K. Finally, we predicted a T_M of 360 K vs. 336 K experimentally for the 85-residue HPr protein. Although a larger test set of proteins is needed, there is a systematic deviation of ± 25 K between the OPEPv5 calculated and experimental melting temperatures.

Four points are worth noting: (i) few CG models report T_M values. While earlier UNRES simulations found T_M of 1000 K,[31] the last UNRES version reports 297 and 317 K for the trpzip1 and trpzip2 peptides vs. 323 and 345 K experimentally.[33] No other systems are, however, available for validation. Note that by using OPEPv4 parameters, we found a T_M of 360 K for trpzip2.[23] In contrast, Voth's model finds T_M values 120 K lower than experiments for trpzip2 and Trp-cage,[36] (ii) OPEPv3[22] and other calculations[265] showed that an overestimation of T_M can result from the absence of a desolvation energy barrier; (iii) even all-atom force fields in explicit solvent overestimate T_M by 30−40 K,[3] and (iv) due to this T_M shift, we recommend to start with a minimal T of 260 K for ST- or REMD-OPEP simulations.

A second aspect to be aware of is that the OPEP heat capacities above T_M, are smaller than the experimental values. This is not surprising since three major terms account for the absolute heat capacity of a protein: one first term depending on the covalent structure and the contributions from all internal vibrational modes; a second term arising from non-covalent interactions of the 2D and 3D structures; and a third term from hydration. For a typical globular protein in solution the heat capacity at 25°C is given by the covalent structure term (85%) and the hydration term (15%). In contrast, the change in heat capacity upon unfolding results from the increase in the hydration term (95%) and then the loss of non-covalent interactions (5%).[266] Simplified side chains and the implicit solvent in OPEP make it difficult therefore to estimate the hydration contribution accurately above T_M. Although free energy differences may fit experimental data, a breakdown of free energies into enthalpies and entropies can be reliable for the backbone, but is not accurate for the side chains. We can however envision running a number of all-atom simulations in explicit solvent from a selected list of poses in the folded and unfolded states.

On-Going Projects and Developments

Physics behind thermophilic and mesophilic proteins

The capability of OPEP to simulate protein folding/unfolding and temperature melting makes the model a powerful tool to study the elementary stabilizing forces in biomolecules. In this regard, proteins from thermophilic organisms are ideal study-cases. These proteins are stable and functional up to 100°C.[267,268] While the general mechanisms that sustain such an extreme behavior remain to be determined, some molecular peculiarities have been singled out. A comparative structural analysis indicates that short loops are important motifs for stability[269] and *de novo* protein design based on the ROSETTA force field successfully predicted enhanced stability of proteins with minimal loops.[270] At the level of chemical composition, thermophiles have a systematic higher population of charged amino acids and salt-bridges, thus optimized electrostatics are thought to be a key ingredient for enhanced stability. Optimizing these interactions at the protein surface, based on the simplified Kirkwood-Tanford electrostatic model, allowed the design of proteins with increased stability.[271] However playing with electrostatics is not always an effective route to enhance stability, because mutations designed to introduce ion pairs can compromise stability due to the large desolvation penalty associated with buried ionic groups.[272-274]

The OPEPv4 model was used to explore the thermal stability of two homologues, the G-domains of EF-Tu and 1α proteins. These 200-aa domains were simulated by REMD using 24 replicas spanning 260-580 K, each for 300 ns. The specific heats of unfolding, reported in Fig. 8, show two main peaks. Though convergence is not reached, remarkably, the curve of the hyper-thermophilic protein is systematically shifted to higher temperatures (see horizontal arrows) mirroring its enhanced thermal stability. The calculated shift between the two homologues is 35 K, comparing favorably with the experimental difference of 40 K. Extended simulations and tests with OPEPv5 that includes improved potentials for salt bridges are in progress.

This preliminary result shows that OPEP with REMD can shed light on the intriguing problem of thermal stability. First, it is possible to obtain at a reasonable computational cost the melting temperature of homologues either by monitoring the peak of the heat capacity or by reconstructing the stability curve, $\Delta G^{f/u} = -kT \cdot \ln(p_u/p_f)$, where $p_{f(u)}$ is the probability to occupy the folded and unfolded states. This latter strategy could be crucial to understand the thermodynamic mechanism sustaining protein activity at high temperatures.[268, 275-277]

The extensive sampling of conformations in both the folded and unfolded states then would provide key information on the protein flexibility at ambient condition and the presence of motifs in the unfolded state. Moreover, the decomposition of the free energy gap into enthalpy and entropy, here clearly limited by the nature of the CG to the behavior of the backbone, could provide extra information on the stability mechanism.[278-280]

Effect of shear flow on protein folding and amyloid formation

It has been reported that hydrodynamic interactions accelerate collapse during polymer coil-to-globule transition[281] or protein folding,[282-283] and affect the kinetics of lipid membrane self-assembly.[284] Thanks to the MUPHY/OPEP coupling it is now possible to explore the behavior of proteins under shear flows. Assessing the effect of shear flow on the stability of proteins is of interest for biotechnological applications because proteins might be degraded due to filtering or injection processes.

Thus far, experimental studies have reached contradictory conclusions about the minimal shear-rate $\dot{\gamma}$ needed to perturb globular proteins.[285] Computational studies have also tackled this problem using simplified (generally Gō-like) models and showed that under strong uniform or elongation flow, proteins do unfold.[283] However, the minimal shear-rate generating unfolding

or the necessary time for cumulating shear stress remains an open issue.[285] We are applying OPEP to shed light on theses issues. In Fig. 9 we present preliminary results of the MD-OPEPv5 dynamics of a β-hairpin peptide in a strong laminar shear flow, $\dot{\gamma} = 10^{10}\ s^{-1}$. We see that after a few nanoseconds the peptide suddenly unfolds as marked by the RMSD increase, and explores several configurations that extend along the velocity gradient.

A systematic study of shear-induced unfolding is appealing also for probing mechanical stability as compared to atomic force spectroscopy experiments; in the former case the drag force is sensed in different locations due to the thermal motion of the protein while in the latter the external pulling applies only along the end-to-end distance axis. Moreover, as for the glycoprotein Ibα receptor, conformational change induced by shear flow can be essential for function, i.e. the binding to the cofactor.[286]

The effect of shear was also appreciated in the context of amyloid fibril formation.[287-290] For instance, it has been observed that in an uniform laminar flow generated in a Couette cell, Aβ1-40 sample forms fibrils within 15 hours at 37°C while in the absence of shear, the process requires at least 1 month.[288] This acceleration corresponds to a decrease of the activation barrier of 4.3 kT or the loss of one hydrogen bond per monomer in solution.[288] A possible mechanism for the effect of shear is that it may lead to the alignment of aggregates, which in turn facilitates their assembly into fibrils. It was further probed that changing the nature of the shear flow, i.e. a heterogeneous field generated by a magnetic stirrer bar, enhances the formation of protofibrils and the growth of fibrils[288] and affects the rate of fragmentation.[290] We are investigating the shear-induced effects on amyloid peptides using OPEP.

Buffer- and pH-dependent OPEP force field

We are currently using all-atom MD simulations and Boltzmann inversion to generate salt-bridge potentials between Lys, Arg and Asp and Glu as a function of pH. Similarly, we are deriving OPEP potentials for polypeptides at various high ionic strengths and in buffers made of H_2KPO_4, H_2SO_4 and DMSO so as to mimic as closely as possible the *in vitro* conditions used to form amyloids. These potentials will be useful for Pep-fold predictions.

Hire-RNA version 3

To predict complex RNA topologies, two features are critical. The first is electrostatics, RNAs being highly charged and ions playing important roles in the structures and thermodynamics. Many models consider ion screening at long distances between phosphate groups via the Debye-Hückel treatment of electrostatic interactions. Other models go more into details by considering either one layer of explicit ions surrounding RNA, referred to as ion condensation,[291] some implicit "structural ions",[228,292] or explicit ions and solvent.[293] In Hire-RNAv3, we added a Debye-Hückel term and the screening parameter is being calibrated against experimental melting temperatures of duplexes as a function of ionic concentrations. We are also exploring the presence of explicit ions.

The second issue is the treatment of non-canonical W-C interactions never taken into account by *ab initio* models. Our new force field allows two bases to interact on all sides, giving rise to about 30 recognition motifs, each one with its specific geometry and strength. Overall, Hire-RNAv3 consists of local terms, excluded volume, ionic screening electrostatics, a proper stacking interaction depending on base position and orientation, and terms accounting for both canonical and non-canonical interactions on the three base sides and also for the co-planarity of the interacting bases (in preparation).

With Hire-RNAv3, we are now able to predict complex RNA structures. As a first benchmark, we studied three systems starting from fully extended states: a 22-nt pseudo-knot, a 49-nt telomerase triple-helix pseudo-knot, and a 79-nt riboswitch with a kissing loop. For the 22-nt RNA, we recovered the native state by running 1.2 μs ST simulation with 15 discrete temperatures from 300 to 450 K (Fig. 10). With a REMD simulation of 64 replicas, each for 0.5 μs, we recovered the native topology of the 49-nt RNA with a small shift of the base pairs, however. For the 79-nt riboswitch, we implemented the possibility of including some restraints. Information about base pairing is easily obtained by preliminary NMR data and is not sufficient to assign the full 3D structure. Imposing 4 restraints in the three helices, we were able to fold this molecule to its NMR topology by a MD trajectory of 0.6 μs at 300 K and recovered the kissing hairpin configuration, although the guanine ligand was not considered (Fig. 10).

Virtual reality and interactive simulations

The use of haptic manipulations of molecular models has been well described.[294-295] The technical requirements are modest; and it is nowadays easy to setup interactive simulations.[296] Even quantum chemistry applications are within reach.[297] For the manipulation of complex biological assemblies, coarser methods are preferable and have been exploited notably for fitting models into experimentally determined envelopes.[298] Generally speaking such approaches build on the idea to render accurate molecular models more real and tangible to the scientists.[299]

We previously pointed out that CG models play a particular role in virtual interactive experiments.[121,300] CG descriptions represent an excellent compromise between simulation speed and biological fidelity. Furthermore our experience suggests that CG-level simulations are generally more robust with respect to user interactions than computations carried out at an all-atom level.

OPEP and Hire-RNA are of particular interest in this context and provide original features that we could not address previously, due to their relatively high resolution of the backbone representation and the presence of directional bonded terms. Both OPEP and Hire-RNA simulation engines were extended for interactive manipulation as described in the methods section. Here, we will mainly brush over the potential benefits of such an approach and restrict ourselves to present a very first, simple toy application, as the validation of this recent MDDriver/IMD implementation is still ongoing.

Generally speaking, the interactive approach opens up perspectives to guide simulations via user input, for example using a haptic device, within a dedicated graphical environment. Hence, the user feels an immediate force feedback by a straightforward combination of classical molecular modeling and virtual reality. An instant benefit is to gather an intuitive understanding of the causal relationship between the theoretical model and its chemically and biologically relevant properties.

These hands-on investigations echo recent experimental ventures into the mechanical properties of molecular structures, and can be associated with the term mechanochemistry.

Experiments, such as AFM pulling and allosteric spring probes, can be reproduced on the fly. Multiple forces may be applied simultaneously to reproduce complex deformations and assemble or disassemble several molecules. Such an interactive exploration can provide insight into the key interactions that govern the mechanical properties of molecular structures and is a unique tool to probe mechanochemistry at a molecular level.

We previously carried out such investigations at a CG level using elastic network models as in the studies of the SNARE complex,[121] RecA nucleofilament[301] and dystrophin fibril.[302] Such spring-based CG models do however preclude any significant changes in the underlying molecular structure that may occur upon tension, a limitation that we are able to lift using OPEP.

In order to illustrate these enhanced possibilities, we interactively manipulated a Hire-RNA model of an RNA hairpin (Fig. 11). By pulling on one or both ends of the structure, it is fairly easy to control the successive detachment of the base pairs. When the added external forces are released, the structure may either progressively return to the initial hairpin state or feature a base shift. This reversible process takes place on the order of a few hundred picoseconds, depending on how much the ends were torn apart. This numerical experiment provides insight on how the hairpin behaves under such stress, similarly to what can be probed experimentally with optical tweezers.[303]

One apparent caveat is the relatively short time step (0.1 to 2 fs) and/or low temperature (100 K) that have to be used to reduce vibrations in order to allow for accurate manipulation. This inconvenience could be lifted by adding the ability to interactively change these parameters during the virtual experiment: using low time-step/temperature values when selecting and applying the forces, then going back to standard values to observe the resulting effects. This kind of simple manipulation can be useful to quickly probe features of the force field or to generate non-trivial starting structures. When carried out more rigorously, the approach may be used to interpret experimental results.

Future extensions of the interactive approach similar to those previously reported[298,302] will enable the use of OPEP and Hire-RNA models to integrate low-resolution experimental data, from small-angle X-ray scattering (SAXS) or Cryo-EM, where both the force field and the user intuition will guide the refinement.

Conclusions

We have presented some of the good applications of OPEP and what OPEP-based simulations can tell us about the structures, dynamics, kinetics and thermodynamics of single proteins, amyloid fibril formation, proteins in a crowded environment with hydrodynamics and RNA/DNA complexes. Whether OPEP can reproduce the effects of a single mutation on protein energy landscapes remains to be determined.[163,304] Compared to the nine CG models described in this report, the OPEP CG strategy with inclusion of the amide hydrogen allows generating more accurate melting temperatures than Voth's model consistent with experimental values and all-atom simulations with growing accuracy of the force fields. OPEP is also free of any bioinformatics-based information (AWSEW), restraints on the backbone (ATTRACT, Klein's model and MARTINI) and has been extensively used on amyloid and non-amyloid systems in contrast to the PRIME, PRIMO, UNRES and Palace systems. The main OPEP disadvantage is that it can only use 2 fs time step. OPEP is coupled to many unbiased advanced conformational sampling methods and interactive virtual reality approaches. We have briefly sketched the main on-going applications and developments. Others, that have just started, include flexible protein/protein and protein/RNA-DNA docking with the necessity to couple the protein and nucleic acid force fields, and the use of the basin hopping method to locate the global energy minimum and calculate disconnectivity graphs to visualize the energy landscape.[305] All these studies will help improve the OPEP parameters and gain a better understanding of how living systems function, and how these functions can be perturbed by internal or external factors.

Acknowledgements

Fabio Sterpone and Maria Kalimeri thank funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013) Grant Agreement no.258748. Part of this work used HPC resources from GENCI [CINES and TGCC] (Grant 2012 c2012086818 and 2013 x201376818). Simone Melchionna thanks M. Bernaschi, M. Bisson, M. Fatica, C. Pierleoni and U. Marconi for discussions, and the Oak Ridge Leadership Computing Facility and CINECA supercomputing center (ISCRA grants KINPROT and FLEXPROT). Pierre Tuffery thanks support of ANR IA "BipBip" and IBiSA for funding the RPBS platform. Alessandro Barducci thanks Massimiliano Bonomi for discussion and the Swiss National Science Foundation for financial support under the Ambizione grant PZ00P2_136856. Marc Baaden thanks the French Agency for Research Grant "ExaViz" (ANR-11-MONU-003), and CNRS (Grant PEPS BMI 2012). Normand Mousseau and Philippe Derreumaux thank the Alzheimer Society of Canada and its 2005 postdoc program. Finally, Philippe Derreumaux thanks support of University of Aix-Marseille II (1999-2003), University of Paris Diderot from 2003, ANR SIMI7 GRAL 12-BS07-0017, ANR LABEX Grant "DYNAMO" (ANR-11-LABX-0011), 6th European PRCD (Immunoprion, FP6-Food023144), Institut Universitaire de France, French/Singapore Merlion PhD program (Grant 5.08.10), Pierre de Gilles de Gennes Foundation and its international PhD grant, Fudan University in China, CNRS - Académie Polonaise des Sciences (Grant 168836), and Institut de Chimie du CNRS over all these 15 years.

Notes and references

[a] *Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Université Paris Diderot, Sorbonne Paris Cité, IBPC, 13 rue Pierre et Marie Curie, 75005, Paris, France. E-mail: Fabio.sterpone@ibpc.fr, samuela.pasquali@ibpc.fr, Tristan.cragnolini@ibpc.fr, maria.kalimeri@ibpc.fr, yohan.laurin@ibpc.fr, marc.baaden@ibpc.fr, phuong.nguyen@ibpc.fr*

[b] *Instituto Processi Chimico-Fisici, CNR-IPCF, Consiglio Nazionale delle Ricerche, Roma, Italy. E-mail: simone.melchionna@roma1.infn.it*

[c] *INSERM U973, Université Paris Diderot, Sorbonne Paris Cité, 75013, France, E-mail: pierre.tuffery@paris7.jussieu.fr*

[d] *Département de Physique, Université de Montréal, C.P. 6128, succ. Centre Ville, Montréal, Québec, H3C 3J7, Canada. E-mail: normand.mousseau@umontreal.ca, jf.stpierre@calculquebec.ca*

[e] *Now at Department of Chemistry, University of Cambridge, UK. E-mail:* ycc34@cam.ac.uk

[f] *Laboratoire de Physique Statistique, Ecole Polytechnique de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, E-mail:* alessandro.barducci@epfl.ch.
[g] *Now at* Department of Cell and Molecular Biology, Uppsala University, Sweden, E-mail: alex.tek@icm.uu.se.
[h] *Institut Universitaire de France, 103 Boulevard Saint-Michel, 75005, Paris, France. Fax: 33 1 58 41 51 74; Tel: 33 1 58 41 51 72; E-mail: philippe.derreumaux@ibpc.fr*
** Corresponding author: philippe.derreumaux@ibpc.fr*

The full terms of the OPEP energy function and the accessibility of the parameters are available at the electronic supplementary information (ESI).

1 J.A. McCammon, B.R. Gelin and M. Karplus, *Nature*, 1977, **267**, 585.
2 Y. Duan and P. Kollman, *Science*, 1998, **282**, 740.
3 K. Lindorff-Larsen, S. Piana, R.O Dror and D.E. Shaw, *Science*, 2011, **334**, 517; S. Piana S, K. Lindorff-Larsen K and D.E. Shaw, *Proc Natl Acad Sci U S A.*, 2013, **110**, 5915.
4 R.O. Dror, H.F. Green, C. Valant, D.W. Borhani, J.R. Valcourt, A.C. Pan, D.H. Arlow, M. Canals, J.R. Lane, R. Rahmani, J.B. Baell, P.M. Sexton, A. Christopoulos and D.E. Shaw, *Nature*, 2013, **503**, 295.
5. M. Stefani and C.M. Dobson, *J Mol Med (Berl)* , 2003, **81**, 678.
6 P. Lito, N. Rosen and D.B. Solit, *Nat Med.*, 2013, **19**, 1401.
7 K.N. Maxwell and S.M. Domchek, *Nat Rev Clin Oncol*., 2012, **9**, 520.
8 M. Levitt and A. Warshel, *Nature*, 1975, **253**, 694.
9 M.G. Saunders and G.A. Voth, *Annu Rev Biophys*., 2013, **42**, 73.
10 W.G. Noid, *J Chem Phys*., 2013, **139**, 090901.
11 S. J. Marrink and D.P. Tieleman, *Chem. Soc. Rev*., 2013, **42**, 6801.
12 P. Derreumaux, *J. Chem. Phys*., 1997, **107**, 1941.
13 P. Derreumaux, *J. Chem. Phys.*, 1997, **106**, 5260.
14 P. Derreumaux, *J. Chem. Phys*., 1998, **109**, 1567.
15 P. Derreumaux, *J. Chem. Phys*., 1999, **111**, 2310.
16 P. Derreumaux, *Phys. Rev. Lett*., 2000, **85**, 206.
17 F. Forcellino and P. Derreumaux, *Proteins*, 2001, **45**, 159.
18 P. Derreumaux, *J. Chem. Phys*., 2002, **117**, 3499.
19 G. Wei, P. Derreumaux and N. Mousseau, *J. Chem. Phys*., 2003, **119**, 6403.
20 J. Maupetit, P. Tuffery and P. Derreumaux, *Proteins*, 2007, **69**, 394.
21 P. Derreumaux, and N. Mousseau, *J. Chem. Phys*., 2007, **126**, 025101.
22 Y. Chebaro, X. Dong, R. Laghaei, P. Derreumaux and N. Mousseau, *J. Phys. Chem. B*, 2009, **113**, 267.
23 Y. Chebaro, S. Pasquali and P. Derreumaux, *J. Phys. Chem. B*, 2012, **116**, 8741.
24 F. Sterpone, P.H. Nguyen, M. Kalimeri and P. Derreumaux, *J. Chem. Theory Comput.*, 2013, **9**, 4574.
25 L. Monticelli, S.K. Kandasamy, X. Periole, R.G. Larson, D.P. Tieleman and S.J. Marrink, *J. Chem. Theor. Comput*., 2008, **4**, 819.
26 D.H. de Jong, G. Singh, W.F.D. Bennett, C. Arnarez, T.A. Wassenaar, L.V. Schafer, X. Periole, D.P. Tieleman and S.J. Marrink, *J. Chem. Theor. Comput*., 2013, **9**, 687.
27 P.J. Bond and M.S.P. Sansom, *J. Am. Chem. Soc*., 2006, **128**, 2697.
28 D.L. Parton, A. Tek, M. Baaden and M.S. Sansom, *PLoS Comput Biol.*, 2013, **9**, e1003034.
29 A.Y. Shih, A. Arkhipov, P.L. Freddolino and K. Schulten, *J Phys Chem B*., 2006 , **110**, 3674.
30 R. Devane, W. Shinoda, P.B. Moore and M.L. Klein, *J. Chem. Theory Comput*., 2009, **5**, 2115.
31 M. Nanias, C. Czaplewski and H.A. Scheraga, *J Chem Theor Comput*., 2006, **2**, 513.
32 A. Liwo, Y. He and H.A. Scheraga, *Phys Chem Chem Phys*., 2011, **13**, 16890.
33 M. Makowski, A. Liwo and H.A. Scheraga, *J Phys Chem B*., 2011, **115**, 6130.
34 K. Simons, C. Kooperberg, E. Huang and D. Baker, *J Mol Biol*, 1997, **268**, 209.
35 A. Leaver-Fay, M.J. O'Meara, M. Tyka, R. Jacak, Y. Song, E.H. Kellogg, J. Thompson, I.W. Davis, R.A. Pache, S. Lyskov, J.J. Gray, T. Kortemme, J.S. Richardson, J.J. Havranek, J. Snoeyink, D. Baker D and B. Kuhlman, *Methods Enzymol*., 2013, **523**, 109.
36 RD. Hills Jr, L. Lu and G.A. Voth, *PLoS Comput Biol*., 2010, **6**, e1000827.
37 S. Peng, F. Ding, B. Urbanc, S.V. Buldyrev, L. Cruz, H.E. Stanley and N.V. Dokholyan, *Phys Rev E Stat Nonlin Soft Matter Phys*., 2004, **4**, 041908.
38 M. Cheon, I. Chang and CK. Hall, *Proteins*, 2010, **78**, 2950.
39 M. Pasi, R. Lavery and N. Ceres, *J Chem Theory Comput*., 2013, **9**, 785.
40 M. Zacharias, *Protein Sci*., 2003, **12**, 1271.
41 M. Zacharias, *Proteins*, 2013, **8**1, 81.
42 P. Kar, S.M. Gopal, Y.M. Cheng, A. Predeus and M. Feig, *J Chem Theory Comput.*, 2013, **9**, 3769.
43 G. Bellesia and J.E. Shea, *J. Chem. Phys*., 2009, **130**, 145103.
44 A. Davtyan, N.P. Schafer, W. Zheng, C. Clementi, P.G. Wolynes and G.A. Papoian, *J Phys Chem B*., 2012, **116**, 8494.
45 N. Basdevant, D. Borgis and T. Ha-Duong, *J Phys Chem B*., 2007, **111**, 9390; N. Basdevant, D. Borgis and T. Ha-Duong, *J Chem Theory Comput*., 2013, **9**, 803.
46 T. Bereau and M. Deserno, *J. Chem. Phys*., 2009, **130**, 235106.
47 P. Májek and R. Elber, *Proteins*, 2009, **76**, 822.
48 R. Pellarin, P. Schuetz, E. Guarnera and A. Caflisch, *J. Am. Chem. Soc.*, 2010, **132**, 14960?
49 P.S. Aisen, B. Vellas and H. Hampel, *Nat Rev Drug Discov*., 2013, **4**, 324.
50 P. Vlieghe, V. Lisowski, J. Martinez and M. Khrestchatisky, *Drug Discov Today*, 2010, **15**, 40.
51 S. Lohan and G.S. Bisht, *Mini Rev Med Chem*., 2013, **13**, 1073.
52 A.G. Jamieson, N. Boutard, D. Sabatino and W.D. Lubell, *Chem Biol Drug Des.*, 2013, **81**, 148.
53 J.A. Cruz, M.F. Blanchet, M. Boniecki, J.M. Bujnicki, S.J. Chen, S. Cao, R. Das, F. Ding, NV. Dokholyan, SC. Flores, L. Huang, CA. Lavender, V. Lisi, F. Major, K. Mikolajczak, DJ. Patel, A. Philips, T. Puton, J. Santalucia, F. Sijenyi, T. Hermann, K. Rother, M. Rother, A. Serganov, M. Skorupski, T. Soltysinski, P. Sripakdeevong, I. Tuszynska, KM. Weeks, C. Waldsich, M. Wildauer, N.B. Leontis and E. Westhof, *RNA*, 2012, **4**, 610.
54 J. Allmer, *Methods Mol Biol*., 2014, **1107**, 157.
55 F. Crea, P.L. Clermont, A. Parolia, Y. Wang and C.D Helgason, *Cancer Metastasis Rev*., 2013 in press PMID: 24346158.
56 Y. He, X.M. Meng, C. Huang, B.M. Wu, L. Zhang, X.W. Lv and J. Li, *Cancer Lett*., 2014, **344**, 20.
57 M. Bernaschi, M. Bisson, M. Fatica, S. Melchionna, S. Succi, *Comm. Comput. Phys.*, 2012, **184**, 329.
58 S.R. McGuffee and A.H. Elcock, *PLoS Comput Biol*., 2010, **6**, e1000694.
59 S. Sharma, S.K. Kumar, S.V. Buldyrev, P.G. Debenedetti, P.J. Rossky and H.E. Stanley, *Sci Rep.,* 2013, **3**, 1841.
60 H.C. Gonzalez, L. Darré and S.J. Pantan, J *Phys Chem B*., 2013, **117**, 14438.
61 W. Han W and K. Schulten, *J Chem Theory Comput.,* 2012, **8**, 4413.
62 A. Kapoor and A. Travesset, *Proteins*, 2014, **82**, 505.
63 T.E. Cheatham III and D.A. Case, *Biopolymers*, 2013, **99**, 969.
64 A. Barducci, M. Bonomi and P. Derreumaux, *J. Chem. Theory Comput*., 2011, **7**, 1928.
65 B. R. Brooks, C. L. Brooks III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, *J. Comp. Chem*., 2009, **30**, 1545.
66 R. Malek and N. Mousseau, *Phys Rev. E*, 2000, **62**, 7723.
67 G. Wei, N. Mousseau and P. Derreumaux, *J. Chem. Phys*., 2002, **117**, 113749.
68 Y. G. Spill, S. Pasquali and P. Derreumaux, *J. Chem. Theory Comput*., 2011, **7**, 1502.
69 R. Laghaei, N. Mousseau and G. Wei, *J. Phys. Chem. B*, 2010, **114**, 7071.
70 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A*., 2002, **99**, 12562.
71 P.H. Nguyen, Y. Okamoto and P. Derreumaux, *J. Chem. Phys*., 2013, **138**, 061102.

72 P. Tuffery and P. Derreumaux, *Proteins*, 2005, **61**, 732.
73 P. Tuffery, F. Guyon and P. Derreumaux, *J. Comput. Chem.,* 2005, **26**, 506.
74 P. Derreumaux, *Biophys J*., 2001, **81**, 1657.
75 P. Derreumaux, *J. Chem. Phys*., 2003, **119**, 4940.
76 G. Wei, N. Mousseau and P. Derreumaux, *Proteins*, 2004, **3**, 464.
77 S. Santini, G. Wei, N. Mousseau and P. Derreumaux, *Structure*, 2004, **12**, 1245.
78 S. Santini, N. Mousseau and P. Derreumaux, *J. Am. Chem. Soc.*, 2004, **126**, 11509.
79 G. Wei, N. Mousseau and P. Derreumaux, *Biophys. J.*, 2004, **87**, 3648.
80 A. Melquiond, G. Boucher, N. Mousseau and P. Derreumaux, *J. Chem. Phys*., 2005, **122**, 174904.
81 N. Mousseau and P. Derreumaux, *Acc Chem Res*, 2005, **38**, 885.
82 A. Melquiond, N. Mousseau and P. Derreumaux, *Proteins*, 2006, **65**, 180.
83 W. Chen, N. Mousseau and P. Derreumaux, *J. Chem. Phys*., 2006, **125**, 084911.
84 N. Mousseau and P. Derreumaux, *Front. Biosci*., 2008, **13**, 4495.
85 J.F St-Pierre, N. Mousseau and P. Derreumaux, *J. Chem. Phys*., 2008, **128**, 045101.
86 J. P. K. Doye and D. J. Wales, *Z. Phys. D*, 1997, **40**, 194.
87 R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson and H. Jónsson, *J. Chem. Phys*., 2004, **121**, 9776.
88 M.R. Yun, N. Mousseau and P. Derreumaux**,** *J Chem Phys*., 2007, **126**, 105101.
89. D. A. Evans and D. J. Wales, *J. Chem. Phys*., 2003, **119**, 9947.
90. M.T. Oakley, D.J. Wales and R.L. Johnston, *J Phys Chem B*., 2011 **115**, 11525-9.
91 H.C. Andersen, *J. Comput. Phys*., 1983, **52**, 24.
92 H.J.C. Berendsen, J.P.M. Postma, W.F. van Gusnteren, A. DiNola and J.R. Haak, *J. Chem. Phys*., 1984, **81**, 3668.
93 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett*., 1999, **314**, 141.
94. G.H. Wei, N. Mousseau and P. Derreumaux, *Prion*, 2007, **1**, 3.
95 Y. Chebaro, N. Mousseau and P. Derreumaux, *J. Phys. Chem. B*, 2009, **113**, 7668.
96 Y. Chebaro, P. Jiang, T. Zhang, Y. Mu, P. H. Nguyen, N. Mousseau and P. Derreumaux, *J. Phys. Chem. B*, 2012, **116**, 8412.
97 H. Fukunishi, O. Watanabe and S. Takada, *J. Chem. Phys*., 2002, **116**, 9058.
98 Y. Mo, Y. Lu, G. Wei and P. Derreumaux, *J Chem Phys*., 2009, **130**, 125101.
99 Y. Lu, G. Wei and P. Derreumaux, *J. Chem. Phys.*, 2012, **137**, 025101.
100 U.H. Hansmann and Y. Okamoto, *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*., 1996, **5**, 5863.
101 E. Rosta E and G. Hummer, *J Chem Phys*., 2010, **132**, 034102.
102 S. Park and V. Pande, *Phys Rev E*, 2007, **E76**, 016703.
103 C. Zhang and J. Ma, *J. Chem. Phys.*, 2008, **129**, 134112.
104 Y. Liu, J. Strümpfer, P.L. Freddolino, M. Gruebele and K. Schulten, *J Phys Chem Lett*., 2012, **3**, 1117.
105 B.H. Park and M. Levitt, *J Mol Biol*, 1995, **249**, 493.
106 R. Kolodny, P. Koehl, L. Guibas and M. Levitt, *J Mol Biol*, 2002, **323**, 297.
107 M. Vendruscolo, E. Kussell E and E. Domany, *Fold Des*, 1997, **2**, 295.
108 P. Tuffery and P. Derreumaux, *Proteins*, 2005, **61**, 732.
109 P. Tuffery, F. Guyon and P. Derreumaux, *J. Comput. Chem.*, 2005, **26**, 506.
110 J. Maupetit J, P. Derreumaux and P. Tufféry, *Nucleic Acids Res.*, 2009, **37**, W498.
111 A. Barducci, M. Bonomi and M. Parrinello, *Wiley Interdiscip Rev Comput Mol Sci*, 2011, **1**, 826.
112 V. Limongelli, L. Marinelli, S. Cosconati, C. La Motta, S. Sartini, L. Mugnaini, F. Da Settimo, E. Novellino and M. Parrinello, *Proc Natl Acad Sci USA*, 2012,**109**,1467.
113 F. Baftizadeh, X. Biarnés, F. Pietrucci, F. Affinito and A. Laio, *J. Am. Chem. Soc.* , 2012, **134**, 3886.
114 F. Palazzesi, A. Barducci, M. Tollinger and M. Parrinello, *Proc Natl Acad Sci USA*, 2013, **110** ,14237.
115 G. Bussi, F. L. Gervasio, A. Laio and M. Parrinello, *J Am Chem Soc* , 2006, **128**, 13435.
116 E.L. Ermak and J.A. McCammon, *J. Chem. Phys.*, 1978, **69**, 1352.
117 M. Bernaschi, S. Melchionna, S. Succi, M. Fyta, E. Kaxiras, J.K. Sircar, *Comput. Phys. Comm.*, 2009, **180**, 1495.
118 R. Benzi, S. Succi, and M. Vergassola, *Phys. Rep.*, 1992, **222**, 145.
119 M. Fyta, S. Melchionna, E. Kaxiras and S. Succi, *Comput. Sci. & Eng.*, 2008, **10**, 10.
120 M. Bernaschi, M. Bisson, M. Fatica and S. Melchionna, Proc. ACM/IEEE Intl. Conf. for High Performance Computing, SC13 (2013).
121 O. Delalande, N. Férey, G. Grasseau and M. Baaden, *J. Comput. Chem.*, 2009, **30**, 2375.
122 Z. Lv, A. Tek, F. Da Silva, C. Empereur-mot, M. Chavent and M. Baaden, 2013, *PLoS ONE*, **8**, e57990.
123 W. Humphrey, A. Dalke and K. Schulten, 1996, *J. Mol. Graph.*, **14**, 33.
124 L. Dupuis and N. Mousseau**,** *J Chem Phys*., 2012, **136**, 035101.
125 J.F. St-Pierre and N. Mousseau**,** *Proteins*, 2012, **80**, 1883
126 D.J. Selkoe, *Nature,* 2003, **426**, 900.
127 C.L. Masters and D.J. Selkoe, *Cold Spring Harb. Perspect. Med.* 2012, **2**, a006262.
128 B.D. Moore, P. Chakrabarty, Y. Levites, T.L. Kukar, A.M. Baine, T. Moroni, T.B. Ladd, P. Das, D.W. Dickson and T.E. Golde, *Alzheimers Res Ther.* 2012, **4**, 18.
129 T.P. Knowles, C.A. Waudby, G.L. Devlin, S.I. Cohen, A. Aguzzi, M. Vendruscolo, E.M. Terentjev, M.E. Welland and C.M. Dobson CM, *Science,* 2009, **326**, 1533.
130 S.I. Cohen, S. Linse, L.M. Luheshi, E. Hellstrand, D.A. White, L. Rajah, D.E. Otzen, M. Vendruscolo, C.M. Dobson and T.P. Knowles, *Proc Natl Acad Sci U S A*., 2013, **110**, 975.
131 M.D. Kirkitadze, M. M. Condron and D.B. Teplow, *J. Mol. Biol*., 2001, **312**, 1103.
132 R. Sarroukh, E. Goormaghtigh, J.M. Ruysschaert and V. Raussens, *Biochim Biophys Acta*., 2013, **1828**, 2328.
133 S.L. Bernstein, N.F. Dupuis, N.D. Lazo, T. Wyttenbach, M.M. Condron, G. Bitan, D.B. Teplow, J.E. Shea, B.T. Ruotolo, C.V. Robinson and M.T. Bowers, *Nat Chem,* 2009, **1**, 326.
134 W.M. Tay, D. Huang, T.L. Rosenberry TL and A.K. Paravastu, *J Mol Biol*., 2013, **425**, 2494.
135 Y. Zhang, D.L. Rempel, J. Zhang, A.K. Sharma, L.M. Mirica and M.L. Gross, *Proc Natl Acad Sci U S A*., 2013, **110**, 14604.
136 S. García, C. Cuscó, R.F. Brissos, E. Torrents, A. Caubet and P. Gamez, *J Inorg Biochem*., 2012, **116**, 26.
137 J-X. Lu, W. Qiang, W.M. Yau, C.D. Schwieters, S.C. Meredith and R. Tycko, *Cell*, 2013, **154**, 1257.
138 M.R. Sawaya, S. Sambashivan, R. Nelson, M.I. Ivanova, S.A. Sievers, M.I. Apostol, M.J. Thompson, M. Balbirnie, J.J. Wiltzius, H.T. McFarlane, A. Madsen, C. Riekel and D. Eisenberg**,** *Nature*, 2007, **447**, 453.
139 D. Matthes, V. Gapsys and B.L. de Groot, *J. Mol. Biol*., 2012, **421**, 390.
140 D. W. Li, S. Mohanty, A. Irback and S. Huo, *PLoS Comput. Biol*., 2008, **4**, e1000238.
141 A. De Simone and P. Derreumaux, *J. Chem. Phys*., 2010, **132**, 165103.
142 C. Liu, M.R. Sawaya, P.N. Cheng, J. Zheng, J.S. Nowick and D. Eisenberg, *J. Am. Chem. Soc.,* 2011, **133**, 6736.
143 W. Song, G. Wei, N. Mousseau and P. Derreumaux, *J. Phys. Chem. B*, 2008, **112**, 4410.
144 A. Laganowsky, C. Liu, M. R. Sawaya, J. P. Whitelegge, J. Park, M. Zhao, A. Pensalfini, A. B. Soriaga, M. Landau, P.K. Teng, D. Cascio, C. Glabe and D. Eisenberg, *Science*, 2012, **335**, 1228.
145 R. Laghaei and N. Mousseau, J. Chem. Phys., 2010, **132**, 165102.
146 M.F. Perutz, J.T. Finch, J. Berrilan and A. Lesk, *Proc Natl Acad Sci U S A*., 2002, ***99***, 5591.
147 S.A. Petty and S.M. Decatur, *Proc Natl Acad Sci U S A*., 2005, ***102***, 14272.
148 M. Kittner and V. Knecht, *J. Phys. Chem. B*, 2010, **114**, 15288; C.S. Whittleston and D.J. Wales, *J Am Chem Soc*., 2007, **129**, 16005.
149 Y. Chebaro and P. Derreumaux, *Proteins*, 2009, **75**, 442.
150 G. Wei, W. Song, P. Derreumaux and N. Mousseau, *Front Biosci.* 2008, **13**, 5681.

151 W. Song, G. Wei, N. Mousseau and P. Derreumaux, *J. Phys. Chem. B*, 2008, **112**, 4410.
152 Y. Lu, P. Derreumaux, Z. Guo, N. Mousseau and G. Wei, *Proteins*, 2009, **75**, 954.
153 J. Nasica-Labouze, M. Meli, P. Derreumaux, G. Colombo and N. Mousseau, *PLoS Comput. Biol.*, 2011, **7**, e1002051.
154 B. Barz, D.J. Wales and B. Strodel, *J Phys Chem B*., 2014, **118**, 1003.
155 F. Baftizadeh, F. Pietrucci, X. Biarnés and A. Laio, *Phys Rev Lett.,* 2013, **110**, 168103.
156 P.H. Nguyen and P. Derreumaux, *J. Phys. Chem. B*, 2013, **117**, 5831.
157 R. Tycko and R.B. Wickner, *Acc. Chem. Res*., 2013, **46**, 1487.
158 J. Nasica-Labouze and N. Mousseau, *PLos Comp. Biol*., 2012, **8**, e1002782.
159 C.T. Middleton, P. Marek, P. Cao, C. Chiu, S. Singh, A.M. Woys, J.J. de Pablo, D.P. Raleigh and M.T. Zanni, *Nature Chem*, 2012, **4**, 355.
160 J.E. Straub and D. Thirumalai, *Annu Rev Phys Chem*., 2011, ***62***, 437.
161 A. De Simone, C. Kitchen, A.H. Kwan, M. Sunde, C.M. Dobson and D. Frenkel, *Proc Natl Acad Sci U S A*., 2012, **109**, 6951.
162 Q. Qiao, G.R. Bowman and X. Huang, *J. Am. Chem. Soc.*, 2013, **135**, 16092.
163 P.H. Nguyen, B. Tarus and P. Derreumaux, *J Phys Chem B*., 2014, **118**, 501.
164 P.H. Nguyen and P. Derreumaux, *Acc. Chem. Res*., 2014, **47**, 603.
165 P. Ciryam, G.G. Tartaglia, R.I. Morimoto, C.M. Dobson and M. Vendruscolo, *Cell Rep*., 2013, **5**, 781.
166 D.V. Laurents, P.M. Gorman, M. Guo, M. Rico, A. Chakrabartty and M. Bruix, *J. Biol. Chem*., 2005, **280**, 3675.
167 S. Côté, P. Derreumaux and N. Mousseau, *J. Chem. Theory Comput.,* 2011, **7**, 2584.
168 S. Cote, R. Laghaei, P. Derreumaux, and N. Mousseau, N. *J. Phys. Chem. B, 2012, **116***, 4043.
169 Y.S. Lin YS and V.S. Pande**,** *Biophys J.,* 2012, **103**, L47.
170 D.J. Rosenman, C.R. Connors, W. Chen, C. Wang and A.E. Garcia, *J. Mol. Biol.,* 2013, **425**, 3338.
171 K.A. Ball, A.H. Phillips, D.E. Wemmer, D. E. and T. Head-Gordon, *Biophys. J.*, 2013, **104**, 2714.
172 Z. Lv, R. Roychaudhuri, M.M. Condron, D.B. Teplow and Y.L. Lyubchenko, *Sci. Rep.*, 2013, **3**, 2880.
173 T. Zhang, J. Zhang, P. Derreumaux and Y. Mu, *J. Phys. Chem. B*, 2013, **117**, 3993.
174 L. Jiang, C. Liu, D. Leibly, M. Landau, M. Zhao, M.P. Hughes and D.S. Eisenberg, *Elife.*, 2013, **2**, e00857.
175 R. Scherzer-Attali, R. Pellarin, M. Convertino, A. Frydman-Marom, N. Egoz-Matia, S. Peled, M. Levy-Sakin, D.E. Shalev, A. Caflisch, E. Gazit and D. Segal, *PLoS One*, 2010, **5**, e11101.
176 T. Zhang, W. Xu, Y. Mu and P. Derreumaux, *ACS Chem. Neurosci.*, 2014, **5**, 148.
177 J. Maupetit J, P. Derreumaux and P. Tufféry, *J. Comput. Chem.,* 2010, **31**, 726.
178 P. Thévenet, Y. Shen, J. Maupetit, F. Guyon, P. Derreumaux and P. Tufféry, *Nucleic Acids Res*., 2012, **40**, W288.
179 A.C. Camproux, R. Gautier and P. Tuffery, *J. Mol. Biol.,* 2004, **339**, 591.
180 P. Thévenet, Y. Shen , J. Maupetit, F. Guyon F, A. Padilla , P. Derreumaux and P. Tuffery, *J. Peptide Sci.,* 2012, **18**, 60.
181 F. Guyon and P. Tuffery, *Bioinformatics*, 2014, **30**, 784.
182 Y. Shen, G. Picord, F. Guyon, and P. Tuffery**,** *PLoS One,* 2013, **8**, e80493.
183 J.D. Steckbeck, J.K. Craigo, C.O. Barnes and R.C. Montelaro, *J. Biol. Chem*., 2011, **286**, 27156.
184 B. San Francisco, E.C. Bretsnyder and R.G. Kranz, *Proc Natl Acad Sci U S A.*, 2013, **110**, E788.
185 S. Horjales, D. Schmidt-Arras, R.R. Limardo, O. Leclercq, G. Obal, E. Prina, A.G. Turjanski, G.F. Späth and A. Buschiazzo, *Structure*, 2012, **20**, 1649.
186 D.R. Chopera, L.A. Cotton, A. Zawaira, J.K. Mann, N.K. Ngandu, R. Ntale, J.M. Carlson, K. Mlisana, Z. Woodman, D. de Assis Rosa, E. Martin, T. Miura, F. Pereyra, B.D. Walker, C.M. Gray, D.P. Martin, T. Ndung'u, M.A. Brockman, S.A. Karim, Z.L. Brumme, and C. Williamson; C CAPRISA 002 Study Team. *J Virol,* 2012, **86**, 13423.
187 J.F. Mouscadet, R. Arora, J. André, J.C. Lambry, O. Delelis, I. Malet, A.G. Marcelin, V. Calvez and L. Tchertanov, *J. Mol. Recognit*, 2009**,** **22**, 480.
188 H. K. Malik-Chaudhry, A. Saavedra and J. Lia, *Biotechnology and Bioengineering*, 2014, **DOI:** 10.1002/bit.25183
189 G. Feller, D. Dehareng and J.L. Lage, *FEBS J.*, 2011, **278**, 2333.
190 M. Li, S. Chang, L. Yang, J. Shi, K. McFarland, X. Yang, A. Moller, C. Wang, X. Zou, C. Chi and J. Cui, *J. Biol. Chem*., 2014, **289**, 4735.
191 G. Wu, K. Han, F. Lv., *J. Theor. Biol*., 2013, **317**, 293.
192 M. Kumar, S. Ahmad, E. Ahmad, M.A. Saifi and R.H. Khan, *PLoS One*, 2012, **7**, e36770.
193 N. Chopra, S. Agarwal, S. Verma, S. Bhatnagar and R. Bhatnagar, *J. Comput. Aided Mol. Des.,* 2011, **25**, 275.
194 C.M. Stegmann, R. Lührmann, M.C. and Wahl, *PLoS One*, 2010, **5**, e10013.
195 D.R. Flower, *Nature Chem. Biol.,* 2013, **9**, 749.
196 N. Olsson, S. Wallin, P. James, C.A. Borrebaeck and C. Wingren, *Protein Sci.,* 2012, **21**, 1897.
197 R. López-Martínez, G.L. Ramírez-Salinas, J. Correa-Basurto and B.L. Barrón, *PLoS One*, 2013, **8**, e76876.
198 T. Jesús, L. Rogelio, C. Abraham, L. Uriel, G. J-Daniel G, M.T. Alfonso MT and B.B. Lilia, *Bioinformation*, 2012, **8**, 870.
199 W. Deng, C. Guan, K. Liu, X. Zhang, X. Feng, B. Zhou, X. Su and P. Chen, *Virus Research*, 2013, **13**, 00369.
200 A. Qureshi, N. Thakur, H. Tandon and M. Kumar, *Nucleic Acids Res.,* 2014, **42**, D1147.
201 S.K. Gupta, A. Singh, M. Srivastava, S.K. Gupta and B.A. Akhoon, *Vaccine*, 2009, **28**, 120.
202 ML. Teixeira, A. Dalla Rosa and A. Brandelli, *Microbiology*, 2013, **159**, 980.
203 J. Lin and A. Alexander-Katz, *ACS Nano*, 2013, **7**, 10799.
204 L. Yan, Y.Yan, H. Liu and Q.Lv, *BioSystems*, 2013, **113**, 1.
205 R. Berges, J. Balzeau, M. Takahashi, C. Prevost and J. Eyer, *PLoS One*, 2012, **7**, e49436.
206 K.H .Lee and F.M.Ytreberg, *Entropy*, 2012, **14**, 630.
207 J.M. Ramis, M. Rubert, J. Vondrasek, A. Gayà, S.P. Lyngstadaas and M. Monjo, *Tissue Eng. Part A*, 2012, **18**, 1253.
208 M. G. Friedrich, J. Lam and R.J.W. Truscott, *J. Biol. Chem.*, 2012, **287**, 39012.
209 X. Chen, X. Yu, Y. Liu and J. Zhang, *J. Mol. Graph. Model.,* 2013, **46**, 83.
210 U.K. Gupta, S. Mahanta and P. Subhankar, *Medical Hypotheses*, 2013, **81**, 853.
211 S. Gautam, K.C. Loh, *Separation and purification technology*, 2013, **102**, 173.
212 M. Zgarbova, M. Otyepka, J. Sponer, A. Mladek, P. Banas, T.E. Cheatham TE and P. Jurecka, *J Chem Theory Comput.,* 2011, **7**, 2886.
213 A.A. Chen and A.E. Garcia, *Proc Natl Acad Sci U S A*., 2013, **110**, 16820.
214 K. Rother, M. Rother, M. Boniecki, T. Puton and J.M. Bujnicki, *J Mol Model*, 2011, **17**, 2325.
215 K. Rother, M. Rother, T. Puton and J.M. Bujnicki, *Nucleic Acids Res.*, 2011, **39**, 4007.
216 S. Coulbourn, X. Flores and R.B. Altman, *RNA*, 2010, **16**, 1769.
217 M. Parisien and F. Major, *Nature*, 2008, **452**, 51.
218 R. Das and D. Baker, *Proc Natl Acad Sci U S A*., 2011, **104**,14664.
219 S. Cao and S.J. Chen, *J. Phys. Chem B*, 2011, **115**, 4216.
220 C. Laing, S. Jung, N. Kim, S. Elmetwaly, M. Zahran and T. Schlick, *PLoS One*, 2013, **8**, e71947.
221 E. Rivas, *RNA Biol.*, 2013, **10**, 1185.
222 A. Morris-Andrews, J. Rottler and S.S. Plotkin, *J. Chem. Phys.*, 2010, **132**, 35105.
223 Y. He, M. Maciejczyk, S. Oldziej, H.A. Scheraga HA and A. Liwo, *Phys Rev Lett*., 2013, **110**, 098101; C.W. Hsu, M. Fyta, G. Lakatos, S. Melchionna and E. Kaxiras, *J. Chem. Phys.*, 2012, **137**, 105102.
224 T.E. Ouldridge, A.A. Louis and JPK. Doye, *J. Chem. Phys.*, 2011, **134**, 085101.
225 F. Ding, S. Sharma, P. Chalasani, V.V. Demidov, N.E. Broude and N.V. Dokholyan, *RNA*, 2008, **14**, 1164.
226 C. Hyeon and D. Thirumalai, *Proc Natl Acad Sci U S A*., 2005, **102**, 6789.

227 M.A. Jonikas, R.J. Radmer, A. Laederach, R. Das, S. Pearlman, D. Herschlag and R.B. Altman, *RNA*, 2009, **15**, 189.
228 Z. Xia, D.P. Gardner, R.R. Gutell and P. Ren, *J. Phys Chem B*, 2010, **114**, 13497; Z. Xia, D.R. Bell, Y. Shi and P. Ren, *J Phys Chem B*., 2013, **117**, 3135.
229 J. Bernauer, X. Huang, A.Y. Sim, and M. Levitt M, *RNA*, 2011, **17**, 1066.
230 E.J. Denning and A.D. MacKerell Jr., *J. Am. Chem. Soc*., 2012, **134**, 2800.
231 S. Pasquali and P. Derreumaux, *J. Phys. Chem B*, 2010, **114**, 11957.
232 N.B. Leontis, J. Stombaugh and E. Westhof, *NAR*, 2002, **30**, 3497.
233 T. Cragnolini, P. Derreumaux, S. Pasquali, *J. Phys. Chem B*, 2013, **117**, 8047.
234 J. Santa Lucia, *Proc Natl Acad Sci USA*, 1998, **95**, 1460.
235 R.J. Ellis, *Trends Biochem. Sci*. 2001, **26**, 597; S.B. Zimmerman, A. P. Minton, *Ann. Rev. Biophys. & Biomol. Struct.*, 1993, **22**, 27; H-X. Zhou and S. Qin, *Biophys. Rev.,* 2013, **5**, 207.
236 T. Frembgen-Kesner and A.H. Elcock, *Biophys Rev.,* 2013, **5**, 109.
237 R. Harada, N. Tochio, T. Kigawa, Y. Sugita and M. Feig, *J Am Chem Soc.*, 2013, **135**, 3696; (b) J. Tian and A.E. Garcia, *J. Am. Chem. Soc.*, 2011, **133**, 15157; c) J. Tian and A.E. Garcia, *J. Chem. Phys.*, 2011, **134**, 225101; (d) S.R. McGuffee and A.H. Elcock, *J. Am. Chem. Soc.*, 2006, **128**, 12098;
238 N.A Kurniawan, S. Enemark and R. Rajagopalan, *J Am Chem Soc.*, 2012, **134**, 10200.
239 R. Harada, Y. Sugita and M. Feig, *J Am Chem Soc*., 2012, **134**, 4842.
240 H. Dong, S. Qin and H.X. Zhou, *PLoS Comput Biol*., 2010, **6**, e1000833.
241 C. Leduc, K. Padberg-Gehle, V. Varga, D. Helbing, S. Diez and J. Howard, *Proc Natl Acad Sci U S A*., 2012, **109**, 6100.
242 T. Ando and J. Skolnick, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 18457.
243 J.T. Mika and B. Poolman, *Curr. Opin. Biotech.*, 2011, **22**, 117.
244 L.-H. Qu, A. Henras, Y.-J. Lu, H. Zhou, W.-X. Zhou, Y.-Q. Zhu, J. Zhao, Y. Henry, M. Caizergues-Ferrer and J.-P. Bachellerie, *Mol. & Cell. Biol.*, 1999, **19,** 1144.
245 M. Weiss, M. Elsner, F. Kartberg and T. Nilsson, *Biophys. J.*, 2004, **87**, 3518; D.S. Banks and C. Fradin, *Biophys. J*., 2005, **89**, 296.
246 F. Roosen-Runge, M. Hennig, F. Zhang, R.M.J. Jacobs, M.Sztucki, H. Schober, T. Seydel and F. Schreiber, *Proc. Natl. Acad. Sci USA*, 2011, **108**, 11815.
247 L.A. Benton, A. E. Smith, G.B. Young and G.J. Pielak, *Biochem.*, 2012, **51**, 9773.
248 W. Xu, C. Zhang, P. Derreumaux, A. Gräslund, L. Morozova-Roche and Y. Mu, *PLoS One.*, 2011, **6**, e24329.
249 N.P. Chongsiriwatana, J.A. Patch, A.M. Czyzewski, M.T. Dohm, A. Ivankin et al., *Proc Natl Acad Sci USA*, 2008, **105**, 2794.
250 D.H. Appella, L.A. Christianson, I.L. Karle, D.R. Powell and S.H. Gellman, *J. Am. Chem. Soc.* 1996, **118**, 13071.
251 J. Tian, H. Dang, Z. Chen, A. Guan, Y. Jin, M.A. Atkinson and D.L. Kaufman, *Diabetes*, 2013, **62**, 3760.
252 H. Amijee, C. Bate, A. Williams, J. Virdee, R. Jeggo, D. Spanswick, D.I. Scopes, J.M. Treherne, S. Mazzitelli, R. Chawner, C.E. Eyers and A.J. Doig, *Biochemistry*, 2012, **51**, 8338.
253 J. Baell, M. Congreve, P. Leeson and C. Abad-Zapatero, *Future Med Chem.*, 2013, **5**, 745.
254 P. Derreumaux, G. Vergoten and P. Lagant, *J. Comput. Chem.,* 1990, **11**, 560.
255 P. Derreumaux and G. Vergoten, *J. Chem. Phys.,* 1995, **102**, 8586.
256 F. Rao, G. Settanni, E. Guarnera and A. Caflisch, *J. Chem. Phys*., 2005, **122**, 184901.
257 B. Urbanc, M. Betnel, L. Cruz, G. Bitan and D.B. Teplow, *J Am Chem Soc.*, 2010, **132**, 4266.
258 C. Guo, Y. Luo, R. Zhou and G. Wei, *ACS Nano*, 2012, **6**, 3907.
259 K. Lindorff-Larsen, P. Maragakis, S. Piana, M.P. Eastwood, R.O Dror RO and D.E. Shaw**,** *PLoS One*. 2012, **7**, e32131.
260 P.H. Nguyen, M.S. Li and P. Derreumaux, *Phys. Chem. Chem. Phys*., 2011, **13**, 9778.
261 P. Derreumaux, M. Dauchez and G. Vergoten, *J. Mol. Struct*, 1993, **295**, 203.
262 P. Derreumaux, K.J. Wilson, G. Vergoten and W.L. Peticolas, *J. Phys. Chem.*, 1989, **93**, 1338.
263 A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.*, 1988, **61**, 2635.
264 M. R. Shirts and J. D. Chodera, *J. Chem. Phys*., 2008, **129**, 124105.
265 Z. Liu Z and H.S. Chan, *J Mol Biol*, 2005, **349**, 872.
266 A. Cooper, D. Cameron, J. Jakus and G.W. Pettigrew, *Biochem Soc Trans*., 2007, **35**, 1547.
267 C. Vieille and G.J. Zeikus, *Microbiol.Mol.Biol.Rev.*, 2001, **65**, 1.
268 F. Sterpone and S. Melchionna, *Chem.Soc.Rev* ., 2012, **41**, 1665.
269 N. Balasco, L. Esposito, A. De Simone and L. Vitigliano, *Protein Sci*, 2013, **22**, 1016.
270 N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T.B. Acton, G.T. Montelione and D. Baker, *Nature*, 2012, **491**, 222.
271 A.V. Gribenko, M.M. Patel, J. Liu, S.A. McCallum, C. Wang and G.I. Makhatadze, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 2601.
272 L. Xiao and B. Honig, *J. Mol. Biol.*, 1999, **289**, 1435.
273 S. Kumar, B. Ma, C.J. Tsai and R. Nussinov, *Proteins*, 2000, **38**, 368.
274 S. Xiao, V. Patsalo, B. Shan, Y. Bi, D. Green and D.P. Raleigh, *Proc Natl Acad Sci USA*, 2013, **110**, 11337.
275 H. Nojima, A. Ikai, T. Oshima and H. Noda, *J. Mol .Biol.*, 1977, **116**, 429.
276 A. Razvi and J.M. Scholtz, *Protein Sci.*, 2006, **15**, 1569.
277 G. Feller, *J Phys : Condens Matter*, 2010, **22**, 323101.
278 C.C. Liu and V.J. Licata, *Proteins*, 2013, **DOI**:10.1002/prot.24458.
279 M. Kalimeri, O. Rahaman, S. Melchionna and F. Sterpone, *J Phys Chem B*, 2013, **117**, 13775.
280 S. Dagan, T. Hagai, Y. Gavrilov, R. Kapon, Y. Levy and Z. Reich, *Proc Natl Acad Sci USA*, 2013, **110**, 10628.
281 K. Kamata, T. Araki and H. Tanaka, *Phys Rev Lett*, 2009, **102**, 108303.
282 T. FrembgenKesner and A.H. Elcock, *J Chem Theory Comput*, 2009, **5**, 242.
283 P. Szymczak and M. Cieplak, *J Phys: Condens Matter*, 2011, **23**, 033102.
284 T. Ando and J. Skolnick, *Biophys. J*., 2013, **104**, 96.
285 J. Jaspe and S.J. Hagen, *Biophys J*., 2006, **91**, 3415.
286 X. Zou, Y. Liu, Z. Chen, G.I. Cárdenas-Jirón and K. Schulten, *Biophys J*., 2010, **99**, 1182.
287 P. Hamilton-Brown, I. Bekard, W.A. Ducker, D. Dunstan and E. Dave, *J Phys Chem B*, 2008, **112**, 16249.
288 D. Dunstan, P. Hamilton-Brown, P. Asimakis, W. Ducker and J. Bertolini, *Protein Eng Des Sel*, 2009, **22**, 741.
289 C. Lee, S. Bird, M. Shaw, L. Jean and D.J. Vaux, *J Biol Chem*, 2012, **287**, 38006.
290 W. Qiang, K. Kelley and R. Tycko, *J Am Chem Soc*., 2013, **135**, 6860.
291 N.A. Denesyuk and D. Thirumalai, *J Phys Chem B*, 2013, **117**, 4901.
292 D.E. Draper, *RNA*, 2004, **10**, 335.
293 R.C. DeMille, T.E. Cheatham III and V. Molinero, *J. Phys. Chem B*, 2011, **115**, 132.
294 A. Ricci, A. Anthopoulos, A. Massarotti, I. Grimstead and A. Brancale, *Future Medicinal Chemistry*, 2012, **4**, 1219.
295 A.M. Wollacott and K.M. Merz Jr., *J. Mol. Graphics Model.*, 2007, **25**, 801.
296 J.E. Stone, A. Kohlmeyer, K.L. Vandivort et al. *in ADVANCES IN VISUAL COMPUTING, PT II Book Series: Lecture Notes in Computer Science*, 2010, **6454**, 382.
297 M.P. Haag and M. Reiher, *Int. J. Quantum Chem.*, 2013, **113**, 8.
298 S. Birmanns, M. Rusu and W. Wriggers, *J. Struct. Biol.*, 2011, **173**, 428.
299 A. Gillet, M. Sanner, D. Stoffler and A. Olson, *Structure*, 2005, **13**, 483.
300 M. Baaden and S.J. Marrink, *Current Opin. Struct. Biol.*, 2013, **23**, 878.
301 A. Saladin, C. Amourda, P. Poulain, N. Férey, M. Baaden, M. Zacharias, O. Delalande and C. Prévost, *Nucleic Acids Res.*, 2010, **38**, 6313.
302 A.E. Molza, N. Ferey, M. Czjzek, E. Le Rumeur, J.F. Hubert, A. Tek, B. Laurent, M. Baaden and O. Delalande, *Faraday Discuss.*, 2014, **DOI:** 10.1039/C3FD00134B.
303 M. Manosas, J.D. Wen, P.T. Li, S.B. Smith, C. Bustamante, I. Tinoco

Jr and F. Ritort, *Biophys J.*, 2007, **92**, 3010.
304 M.H. Viet, P.H. Nguyen, S.T Ngo, M.S. Li and P. Derreumaux, *ACS Chem. Neurosci*, 2013, **4**, 1446; A. Huet and P. Derreumaux, Biophys. J., 2006, **91,** 3829.
304 J. Doye and D. Wales, *J. Phys. Chem. A*, 1997, **101**, 5111.

Table 1. Summary of the different systems studied with the CG force field.

Applications	Methodology used	Total Time	References
$A\beta_{1-40}$, $A\beta_{1-42}$, WT(D23N) monomers and dimers	HT-REMD	70.0 μs 97.5 μs	167 168
Trimers of $A\beta_{17-42}$ and interactions with drugs	REMD	16.4 μs	96
Aggregation of 3- to 20-mers of amyloid fragments	MD ART REMD	30.0 μs no 120.0 μs	149, 150, 152 80,150 23, 68, 99, 149-153
Size of the primary nucleus for fibril formation	MD REMD	33.8 μs 90.0 μs	158 99, 153
Peptide Structure Prediction and Conformations of protein fragments	PEP-FOLD	no	177, 178, 180, 182-189
Protein-peptide interactions	PEP-FOLD	no	190-194
Design of immunogenic and antiviral peptides	PEP-FOLD	no	195-201
Impact of macromolecules and hydrodynamics	MUPHY/OPEP	no	this work
RNA and DNA folding	MD, REMD, ST	85.0 μs	231, 233, this work
Thermophilic and mesophilic proteins	REMD	72.0 μs	this work
Impact of shear flow	MUPHY/OPEP	30 ns	this work
Virtual reality	Interactive MD	0.2 ns	this work

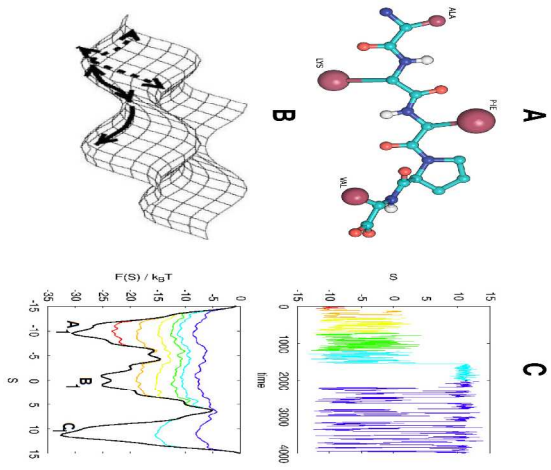
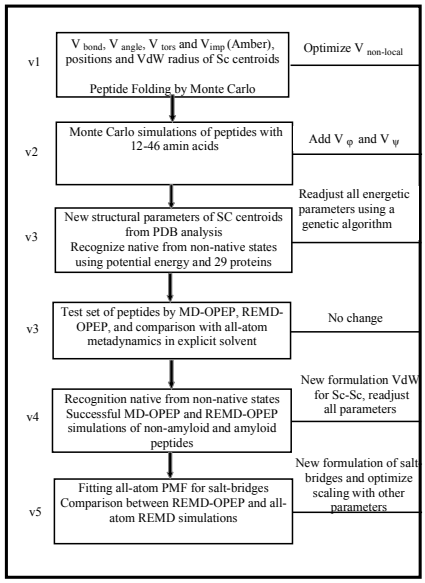


Fig.1 OPEP CG model and enhanced sampling methods. (A) We use the peptide Ala-Lys-Phe-Pro-Val in its zwitterion form to show the details of the backbone and the side-chains. (B) The Activation-Relaxation Technique connecting local minima by first-order saddle points. (C) Example of a metadynamics simulation in a one-dimensional landscape with multiple metastable minima separated by energy barriers. Top panel: System trajectory in CV space as a function of simulation time. Bottom panel: Progressive filling (colored lines) of the underlying potential (black line) by bias. In both the panels color code is used to measure the simulation time. The system starts in the basin A1 and it is pushed by the bias to overcome the free-energy barriers and to visit basin B1 (t~100) and basin C1 (t~1500). In the second half of the simulation, the system can easily sample the whole landscape and the bias can be used to estimate the underlying free-energy surface.

Fig.2 Flowchart depicting the OPEP force field parametrization scheme.



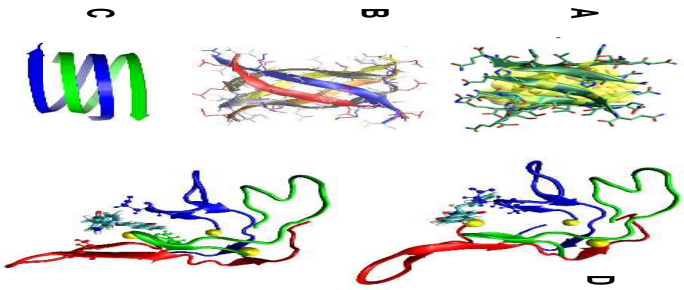


Fig.3 Amyloids. (A) The β-barrel of the β2m(82-87) peptide as predicted by OPEP.[79,143] (B) The hexamer of the KV11 peptide consisting of six antiparallel β strands forming a barrel as determined by X-ray crystallography.[144] (C) The predicted OPEP antiparallel double stranded poly-L-glutamine nanotube.[145] (D) Two binding modes of the NQTrp drug to the Aβ17-42 trimer as predicted by our multiscale simulation. The yellow balls indicate the 17th amino acids and the drug is shown with all-atoms.[96] Top view also shows all atoms of A21 and E22; bottom view shows all atoms of E22 (blue) and V39 (green).

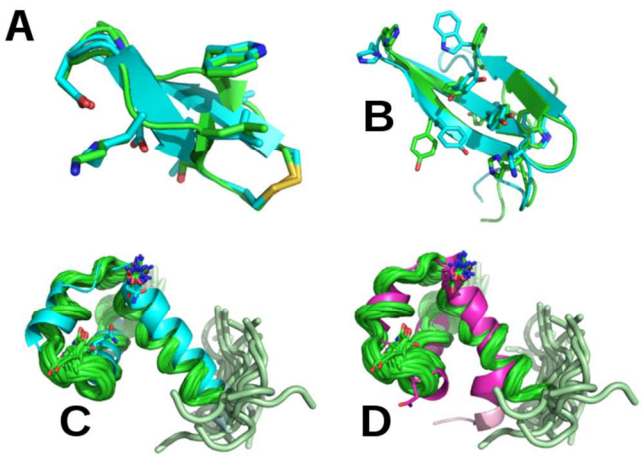


Fig.4 Structure predictions superposed on the experimental structures. **(**A): Best Pep-fold model of the peptide code PDB 1n0a (11-aa, BC score = 0.94) with one S-S bond; (B): Best Pep-fold model of the peptide 1e0m (37-aa, BC score = 0.88). In (A) and (B), we show the all-atom representation of some side chains; (C) and (D): Best Pep-fold and Rosetta models of the peptide 2j8p (49-aa, BC scores = 0.93 and 0.88) superposed on the 20 NMR structures and showing the flexibility of one extremity. Green: experimental conformations.

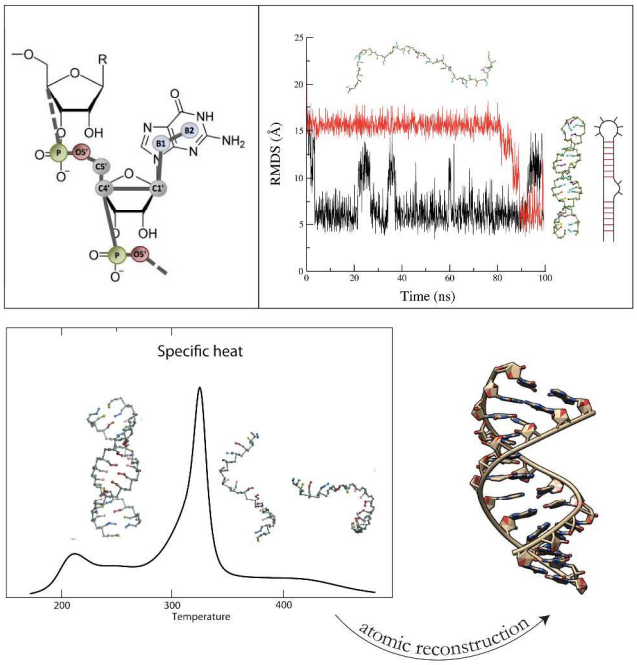


Fig.5 Hire-RNA model. Top, Left: Representation of a Guanine nucleotide with 7 beads. Top, Right: MD of the 36-nt 1N8X hairpin with Hire-RNAv1 recovering the native structure from a fully extended state. Secondary structure of the hairpin is shown on the right. Bottom, Left: Heat capacity plot of a 36-nt RNA duplex with HiRE-RNAv2. Bottom, Right: With our reconstruction algorithm, the predicted all-atom structure behaves similarly to the experimental structure using all-atom MD in explicit solvent.

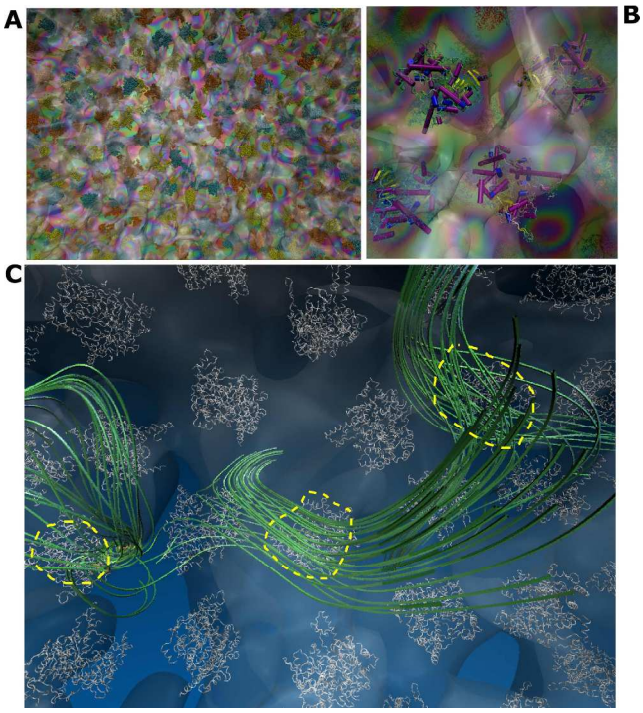


Fig.6 MUPHY/OPEP suspension. (A) Snapshot highlighting the Rat1 proteins and the solvent velocity field shown in transparency. (B) A zoom showing the Rat1 secondary structures and the local solvent velocity field. (C) Flow streamlines generated by three selected proteins in the suspension in a single timeframe.

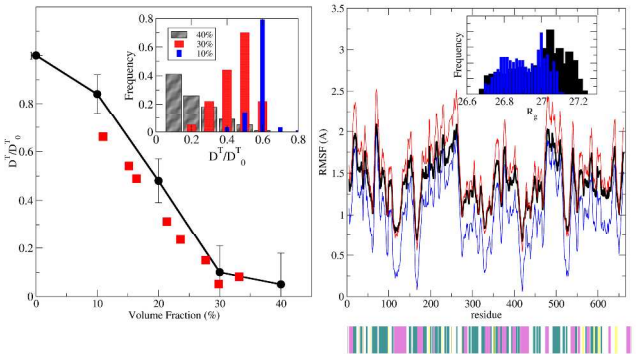


Fig.7 MUPHY/OPEP results. Left: Translational diffusion coefficients at various volume fractions (black circles) are compared to the experimental data of bovine *serum albumin*[246] (red squares). The diffusion coefficient is normalized by the value at virtually zero volume fraction, and the solid line is a guide to the eye. Inset: Histogram of diffusivity stemming from the ensemble of proteins. Right: RMS fluctuations from the crystallographic structure at 40% volume fraction (black curve). Structures with maximal (red) and minimal (blue) departures from the average value are shown. Secondary structure is indicated as a lower bar with colour green (turn), yellow (β-strand), magenta (α-helix) and white (coil). Inset: Histogram of Rg values.

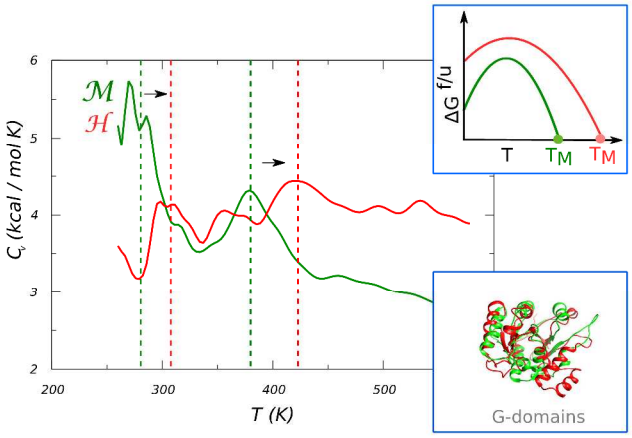
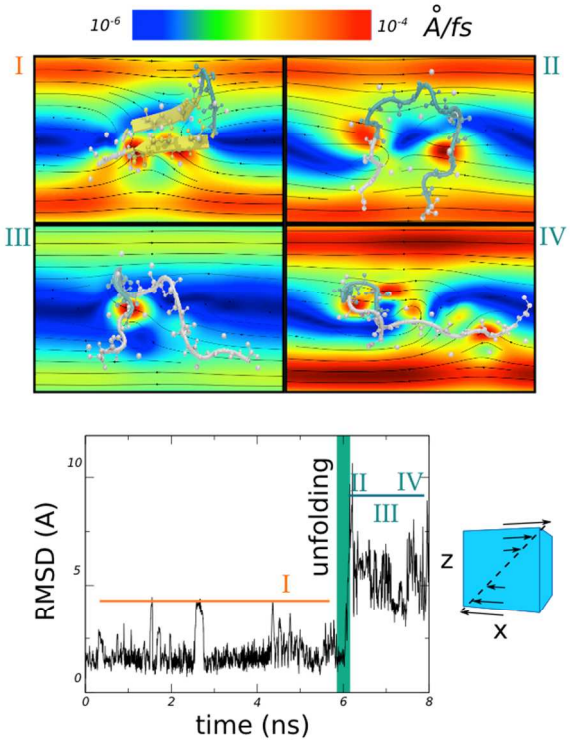


Fig.8 Specific heat of unfolding C_V for the mesophilic (green) and hyperthermophilic (red) domains of the EF-Tu and 1α proteins, respectively calculated from OPEP-REMD simulations. The structural homology of the two proteins is highlighted in the bottom right panel. The presence of two main peaks in the C_V profile is caused by the unfolding events of different secondary structure motifs. In a single two-state model, the C_V is expected to show a single peak at the melting temperature T_M at which the population of the folded (p_f) and unfolded (p_u) states are equal. The melting temperature indicates the zero of the stability curve, see the upper inset graph. Several mechanisms can cause the increase of the T_M of thermophiles, i.e. the upshift, the right shift or the broadening of the curve.[268,275-277.]

Fig.9 Time evolution of the RMSD and snapshots of a β-hairpin peptide (PDB code 1PGB, fragment 41-56) simulated in laminar shear flow. The velocity gradient is generated along the Z direction and corresponds in our simulation to a shear rate of $\dot{\gamma} = 10^{10}\, s^{-1}$. We show the detailed structures of the peptide prior to unfolding (I) and at various unfolding stages (II-IV), with the velocity field represented in background and a colour scale given in the top of the figure.



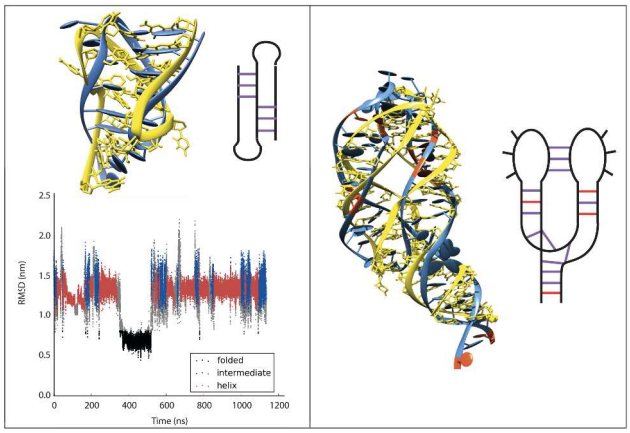


Fig.10 Hire-RNAv3 results. Left, Folding of the 22-nt pseudo-knot (2G1W) using ST simulation with the predicted state (blue) superposed on the NMR structure (yellow). The RMSD with respect to the NMR structure over time is shown. Right, MD-predicted structure of the 79-nt guanine riboswitch 1Y26 (blue) superposed on the experimental structure (yellow) using our four restraints (in red). In both panels the secondary structure of the system is shown on the right.

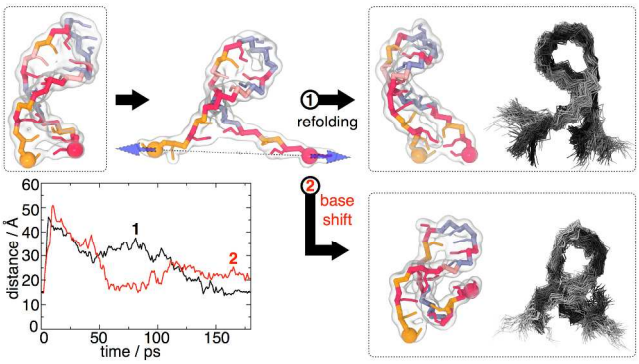


Fig.11 Interactive force unfolding of an RNA hairpin modelled by Hire-RNA. Visualization and interaction were performed within VMD. Starting from an initial folded conformation (top left), the user tears apart both ends by applying forces (blue arrows) on the C4' beads of the end bases. Two scenarios were observed after releasing the forces, either the structure refolds and returns to a full hairpin state (1) or a base shift occurs (2). A cumulated view of simulation snapshots is shown on the very right, coloured from white (start of the simulation) to black (final snapshot). The end base C4'-C4' distance curve over time is shown for both experiments.