Published in final edited form as:

J Chem Theory Comput. 2015 April 14; 11(4): 1843–1853. doi:10.1021/ct501015h.

Protein simulations in fluids: coupling the OPEP coarse-grained force field with hydrodynamics

Fabio Sterpone^{*,†}, Philippe Derreumaux[†], and Simone Melchionna[‡]

[†]Laboratoire de Biochimie Théorique, IBPC, CNRS UPR9080, Univ. Paris Diderot, Sorbonne Paris Cité, 13 rue Pierre et Marie Curie, 75005, Paris, France

[‡]IUF, Institut Universitaire de France, Boulevard Saint Michel, 75005 Paris, France

[‡]CNR-IPCF, Consiglio Nazionale delle Ricerche, Rome, Italy

Abstract

A novel simulation framework that integrates the OPEP coarse-grained (CG) model for proteins with the Lattice Boltzmann (LB) methodology to account for the fluid solvent at mesoscale level, is presented. OPEP is a very efficient, water-free and electrostatic-free force field that reproduces at quasi-atomistic detail processes like peptide folding, structural rearrangements and aggregation dynamics. The LB method is based on the kinetic description of the solvent in order to solve the fluid mechanics under a wide range of conditions, with the further advantage of being highly scalable on parallel architectures. The capabilities of the approach are presented and it is shown that the strategy is effective in exploring the role of hydrodynamics on protein relaxation and peptide aggregation. The end result is a strategy for modelling systems made up to thousands of proteins, such as in the case of dense protein suspensions. The future perspectives of the multi-scale approach are also discussed.

1 Introduction

The study of proteins in cell-like environments is a challenging frontier for modern biophysics and molecular biology^{1,2}. In fact, protein mobility^{3,4}, stability⁵ and function in the extremely heterogenous and crowded space as the interior of cells differ from those in dilute solution conditions. Just to say, in a eukariotic cell, depending on the organelle or subcellular location, 5 to 40% of the total volume is occupied by proteins, nucleic acids and other large molecules. Because of this molecular crowding, excluded volume is expected to affect protein relaxation, mobility and stability². For example, protein motion restrained in a crowded disordered medium is predicted to be slower and sub-diffusive⁴. Retardation has been confirmed by several investigations based on single particle tracking⁶, fluorescence correlation spectroscopy⁷, fluorescence-recovery-after-photobleaching⁸, and quasi elastic neutron scattering⁹. At occupied volume fraction approaching 30-40%, protein diffusion reduces by about one order of magnitude or more with respect to the dilute solution^{9–11}. On the other end, excluded volume stabilises proteins by disfavours highly elongated entropic states². However, the close packing and the action of specific or unspecific interactions, e.g.

^{*}To whom correspondence should be addressed fabio.sterpone@ibpc.fr.

of electrostatic origin, could favour misfolding and aggregation. Experiments have reported both stabilisation and destabilisation depending on the nature of the protein and of the crowders^{5,12,13}.

Despite the growing interest and the increasing experimental investigations, the net effect of crowding is still not completely understood. In fact, experimental studies are particularly challenging since the intracellular environment is difficult to manipulate and control.^{7,13–15} Therefore, to overcome some of the experimental limitations and obtain complementary information at high spatial and temporal resolution, a computational approach is necessary¹⁶. Following seminal work^{17,18}, important efforts are currently under way^{19–24}. Unfortunately, a fully atomistic molecular resolution for describing a cell-like environment is still prohibitive and is currently limited to simulate only small systems or short timescales. Therefore, a coarse-grained (CG) molecular model removing solvent degrees of freedom is a preferred strategy but requires including the computationally expensive hydrodynamic effects in order to account for solvent mediated interactions^{21,25–27}.

Ando and Skolnick²¹ have clearly shown that hydrodynamic interactions (HI) are essential to describe protein mobility under crowding and therefore to reproduce the experimental behaviour. They used Stokesian dynamics²⁸, Brownian dynamics²⁹ with near and far-field HI, applied to a model of *E. coli* cellular environment where proteins were treated as spheres of different sizes. Their model was used to show that only by including HI the mobility of a target protein (GFP) slowed down by a factor 10 as compared to the HI-free case, being then in agreement with experiments¹⁰. Very recently Bandyopahyay and coworkers²⁴ also presented a grained model of the *E. Coli* cytoplasm with proteins described by one or few connected beads, thus including a simplified intra-protein flexibility. The model was used to investigate protein mobility by applying Brownian Dynamics including approximated HI (mean-field). An attempt to provide a more realistic description of proteins in cell was done by McGuffee and Elcock¹⁹, their *E.coli* environment was made up of about 1000 proteins at atomistic resolution but moving as rigid bodies. Unfortunately for systems of such resolution, HI could not be included in the Brownian Dynamics scheme. In fact, HI comes at a high computational cost ($\mathcal{O}(N^3)$), and for methodologies based on Green's function, the non-local character of HI makes necessary to use drastic approximations when the number of particles becomes too high, see discussion in Ref.³⁰.

Simulations based on CG models and including HI find wide application in several contexts. HI crucially affects the protein mobility in dilute solution^{22,26}, the protein folding/unfolding kinetics^{26,31,32}, or the kinetics of lipid aggregation³³ and protein association³⁴. In this work we present and detail out an original multi-scale method that couples the CG model for proteins OPEP^{35–38} with an explicit representation of the solvent dynamics. At variance with previous schemes, the explicit and on-the-fly solution of the fluid kinetics is done via the Lattice Boltzmann (LB) methodology^{39,40} and allows including the HI in a natural way⁴¹. One of the appealing features of the approach is its amenability to parallel computing that makes it particularly efficient in simulating systems composed by many proteins. During the Gordon Bell Prize 2013 contest, the scheme was used to simulate about 18,000 fully flexible proteins for tenths of nanoseconds on about 18,000 GPUs. This

unprecedented simulation established the capability of the framework and an overview of the results was presented in a recent review paper on the OPEP force field³⁸.

In the following we first describe the computational method, namely the coarse-grained model for proteins OPEP and its coupling with the Lattice Boltzmann methodology. We then apply the technique to test the stability of isolated globular proteins in the hydrodynamic fluid, the effect of HI on the aggregation process of amyloid peptides, and the intrinsic multi-scale potentialities of the method for treating systems such as protein suspensions.

2 Methods

2.1 The force field OPEP

The CG model OPEP^{38,42} represents each amino acid by reserving an atomistic description to the backbone (5 centers of force for N, HN, C α , C, and O atoms respectively) while one bead is used for the protein side-chain. An exception is made by proline that is defined by all its heavy atoms. The OPEP force field is defined as a sum of local, non-bonded, and hydrogen bonding terms. The local terms describe bond stretching, angle bending, proper and improper dihedral rotations using functional form as in common to all-atoms force fields. The non-bonded terms describe attractive and repulsive interactions between the center of forces. OPEP is water and electrostatic free, however specific potentials mimic the backbone hydrogen-bonds propensity along with cooperativity and salt-bridge between ionic side-chains. A complete and detailed description of the whole Hamiltonian and the main changes introduced in the latest versions of the model^{36,37} is given in the supplemental material of Ref.³⁸.

In the past, OPEP has been used in combination with several computational techniques, such as Monte Carlo⁴³, molecular dynamics⁴⁴, replica exchange molecular dynamics (REMD)⁴⁵, simulated tempering⁴⁶ and successfully applied to study peptide folding⁴⁷ and amyloid aggregation^{48–50}. At variance with other simplified force fields, OPEP is able to maintain a protein stable at long time scale without recurring to external constraints that “freeze” the secondary structures. Moreover, the force field is not biased toward the native configuration of proteins, so folding just emerges from the basic elementary forces of the model. Finally, we mention that even if – as for any other coarse graining scheme – the energy scale of the force field is somehow arbitrary, the effective tuning of the relative weights of the potential energy terms results in protein melting temperatures ranging from 280–400 K, with deviation from experimental ones by about 20 K for the system tested so far³⁸. As a consequence of this quality, recently the model was successfully applied to study the different thermal stability of a pair of homologues proteins of relatively large size, ~ 200 amino acids⁵¹. When compared to other CG models, a limitation of OPEP is that the atomistic description of the protein backbone requires a short integration timestep to account for the high frequency intramolecular dynamics, a value that does not exceed 1.5 – 2 fs. In the simulations presented in this work we used a timestep of 1.5 fs.

2.2 Coupling OPEP to Lattice Boltzmann

2.2.1 The Lattice Boltzmann Method—The motion of protein atoms has a direct effect on the surrounding aqueous environment. Hydrodynamic interactions arise from the motion of atoms that generate a velocity field in the surrounding space. The resulting flow acts on other protein atoms and therefore produces effective, solvent-mediated interactions. Our purpose is to represent explicitly the dynamics of the solvent rather than employing an effective, analytical expression for hydrodynamic interactions.

The solvent is described by considering the fluid as a continuum in a probabilistic sense. This is the realm of kinetic theory, as envisioned by Boltzmann, where the fluid molecules are described by the single particle distribution function, $f(\mathbf{r}, \mathbf{v}, t)$, that encodes the probability of finding at time t a group of molecules at position $\mathbf{r}$ and carrying a velocity $\mathbf{v}$. While the Boltzmann description holds down to the molecular scale, collisional events taking place among the molecules complete the description. However, it is particularly difficult to determine the precise form of the collisions in the liquid phase and such events can be accurately modelled only in special cases, as for the hard spheres fluid⁵².

The fluid is intrinsically characterized by the Knudsen number, $Kn = \lambda/L_H$, that regulates the deviation of the distribution from the local equilibrium state, where λ is the water mean free path ($\approx 5 \text{ \AA}$ ⁵³) and L_H is a characteristic hydrodynamic scale, ranging from a few intermolecular distances to the protein length scale ($10 - 100 \text{ \AA}$), so that $0.05 < Kn < 0.5$. Such range implies that in the liquid state water typically features small departures from local equilibrium. In the macroscopic limit, that is, after several molecular collisions and over a few molecular diameters, kinetic theory recovers the Navier-Stokes equations of fluid mechanics.

It is crucial to note that the kinetic description is richer than the macroscopic one, as it contains informations regarding the full probabilistic character of matter via a given statistical distribution, while fluid mechanics only deals with the first two moments of such distribution, namely the solvent density field, $\rho(\mathbf{r}, t)$, and the velocity field, $\mathbf{u}(\mathbf{r}, t)$. These are

related by the continuity equation $\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$, and the Navier-Stokes equation describes their macroscopic time evolution, namely

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla P + \nu \nabla^2 \mathbf{u} + \frac{\mathbf{G}}{m} \quad (1)$$

Here, P is the pressure, accounting for the thermodynamic forces mutually exerted between fluid elements, ν is the kinematic viscosity, accounting for dissipation arising from gradients in the velocity field, and $\mathbf{G}$ is the external force exerted by the protein atoms at given locations, exerted over the fluid molecules of mass m . In the following, for the sake of simplicity we initially neglect from the treatment the presence of spontaneous microscopic fluctuations in water, and it will be explicitly introduced in section 2.2.3.

The kinetic evolution of the fluid solvent is formulated in numerical terms by the Lattice Boltzmann method. In essence, the LB method is intrinsically based on the assumption of small departures from local equilibrium, a condition that is obeyed by liquid water under standard conditions. Given a spatial discretization of the distribution, the $f(\mathbf{r}, \mathbf{v}, t)$ is represented over a cartesian mesh with mesh points collectively indicated as $\{\mathbf{x}\}$ and mesh spacing Δx .

In the LB method, the velocity space is represented by a small set of discrete speeds and the dynamics evolves in timesteps of duration Δt . In essence, $f(\mathbf{r}, \mathbf{v}, t)$ is replaced by the simpler quantity, $f_p(\mathbf{x}, t)$, where the index p indicates a discrete velocity vector. The kinetic moments of the populations $f_p(\mathbf{x}, t)$ provide the local mass density $\rho(\mathbf{x}, t) = \sum_p f_p(\mathbf{x}, t)$ and velocity

$\mathbf{u}(\mathbf{x}, t) = \frac{1}{\rho} \sum_p \mathbf{c}_p f_p(\mathbf{x}, t)$. The stress tensor $\overleftrightarrow{P}(\mathbf{x}, t) = \frac{1}{\rho} \sum_p \mathbf{c}_p \mathbf{c}_p f_p(\mathbf{x}, t)$ has diagonal components related to the fluid pressure while the off-diagonal terms relate to the shear stress. At thermodynamic level, the solvent pressure obeys the ideal gas equation of state, a simple, yet effective description such that only hydrodynamic interactions and internal viscous forces are accounted for.

The fluid populations then evolve via the following updating scheme

$$f_p(\mathbf{x} + \mathbf{c}_p \Delta t, t + \Delta t) = f_p(\mathbf{x}, t) - \omega \Delta t (f_p - f_p^{eq})(\mathbf{x}, t) + g_p(\mathbf{x}, t) \quad (2)$$

The right hand side of Eq. (2) contains the effect of fluid-fluid molecular collisions that relax the populations towards a Maxwellian equilibrium, f_p^{eq} . This is the Bhatnagar-Gross-Krook collisional operator, local in space and time, and being proportional to the non-equilibrium component of the distribution, $(f_p - f_p^{eq})$. The Maxwellian equilibrium is expressed by a second-order expansion in the macroscopic velocity $\mathbf{u}$,

$$f_p^{eq} = w_p \rho \left[1 + \frac{\mathbf{u} \cdot \mathbf{c}_p}{c_s^2} + \frac{\mathbf{u} \mathbf{u} : (\mathbf{c}_p \mathbf{c}_p - c_s^2 \mathbf{I})}{2c_s^4} \right].$$

In addition, $c_s = 1/\sqrt{3}$ is the speed of sound, w_p is a set of weights normalized to unity, and $\mathbf{I}$ is the unit tensor in Cartesian space. The Chapman-Enskog multiscale analysis shows that the relaxation frequency ω is related to the

fluid kinematic viscosity via the expression $\nu = c_s^2 \left(\frac{1}{\omega} - \frac{\Delta t}{2} \right)$ 40.

The $\{w_p\}$ form a set of normalized weights, and we follow the implementation provided by the D3Q19 lattice with $p = 0, \dots, 18$. The discrete speeds span the crystal structure of a face

centered cubic lattice, so that $\sum_p w_p \frac{\mathbf{c}_p \mathbf{c}_p}{c^2} = \overleftrightarrow{\mathbf{1}}$, where the double arrow denotes a rank-two tensor $\overleftrightarrow{A} = A_{ab}$, a dyadic product $\mathbf{u} \mathbf{v} = u_a v_b$ and $\overleftrightarrow{\mathbf{1}} = \delta_{ab}$. This choice preserves isotropy with

respect to dyadic products, since $\frac{1}{c^4} \sum_p w_p \sum_{bc} \overline{(c_p c_p)_{ab} (c_p c_p)_{bc}} u_c u_d = u_a u_d$, with the overbar denoting the traceless tensor. For the D3Q19 scheme, the weights w_p are equal to 1/3 for the

population corresponding to the null discrete speed, 1/18 for the ones connecting first mesh neighbors and 1/36 for the ones connecting second mesh neighbors.

Finally, the last term in eq. (2) represents the coupling between the solvent and the atoms.

2.2.2 Coupling molecules and LB—As stated above the evolution of the fluid is perturbed by the presence of solute particles via the action of the term $g_p(\mathbf{x}, t)$. An accurate expression is given by^{54,55}

$$g_p(\mathbf{x}, t) = -w_p \Delta t \left[\frac{\mathbf{G} \cdot \mathbf{c}_p}{c_s^2} + \frac{(\mathbf{G} \cdot \mathbf{c}_p)(\mathbf{u} \cdot \mathbf{c}_p) - c_s^2 \mathbf{G} \cdot \mathbf{u}}{c_s^4} \right] \quad (3)$$

The force $\mathbf{G}$ contains any external force and the exchange of momentum induced by N moving atoms. This term for example includes the drag force and extra contributions as the random noise encoded in the molecular dynamics. In order to have a consistent treatment of

the external forces, the fluid velocity is redefined as $\mathbf{u} = \frac{1}{\rho} \sum_p \mathbf{c}_p f_p + \frac{\Delta t}{2} \mathbf{G}$ ⁵⁴.

The drag force exerted between the i -th particle and the fluid is modelled as^{41,56}

$$\mathbf{F}_i^D(\mathbf{R}_i) = -\gamma (\mathbf{V}_i - \tilde{\mathbf{u}})(\mathbf{R}_i) \quad (4)$$

where $\{\mathbf{V}_i\}$ is the atom velocities and $\tilde{\mathbf{u}}$ indicates the fluid velocity field distributed over the region occupied by the atom. The constant γ represents the translational coupling coefficients that regulates the exchange of momentum between particles and fluid and is related to the particle diffusion coefficient at infinite dilution via the fluctuation-dissipation relation $D = kT/m\gamma$. At macroscopic level the coupling constant γ accounts for Stokes law, such that a sphere of radius a moving in a quiescent fluid experiences a drag force $\mathbf{F}^D = -6\pi\eta a\mathbf{V}$, with ρ the fluid density, and dynamic viscosity $\eta = \rho\nu$. In the simulations discussed later on, the parameter γ was set to 0.1 fs^{-1} , the density $\rho = 10^3 \text{ (kg/m}^3\text{)}$ and the kinematic viscosity $\nu = 0.166 \text{ (\AA}^2/\text{fs)}$, a value that mimics liquid water, whose kinematic viscosity varies between 0.18 and $0.08 \text{ \AA}^2/\text{fs}$ in the temperature interval 0-30° C. The case of different values of these parameters is explicitly stated in the text.

As we intend to treat proteins at microscopic level, Stokes law is not generally applicable and the finite extension of particles is modeled by the lumped form, eq. (4). Accounting for the particle extension sets the relative length scale between the fluid and the particle, being the fluid velocity distributed over the region centered on the particle position. A first choice is to take the fluid velocity at the nearest grid point to the particle position. However, such choice does not allow altering the relative length scale between fluid and particle and could cause small discontinuities in the drag force as the particle moves to different locations. A more flexible scheme is such that the particle velocity is distributed over multiple lattice points by a smooth version of the Dirac delta function, and the fluid velocity is obtained by

interpolation of velocity from its neighboring lattice points. The smooth delta function acts as an interpolation kernel and is given by the area-weighted average function, reading

$$\delta_h(\mathbf{x} - \mathbf{R}_i) = d_h(x - X_i)d_h(y - Y_i)d_h(z - Z_i) \quad (5)$$

where the hat function is defined as

$$d_h(r) = \begin{cases} (1 - |r|/h) & \text{for } |r| \leq h \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

At the smallest scale, h is chosen as the lattice spacing Δx with an effective range of $2\Delta x$, but a larger support given when multiples of Δx are adopted. The hat function consists in a trilinear interpolation scheme but one can also smear according to a higher order scheme, by taking

$$d_h(r) = \begin{cases} \frac{\Delta x}{2h} \left[1 + \cos\left(\frac{\pi|r|}{h}\right) \right] & \text{for } |r| \leq h \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Again, depending on the parameter h the lattice support varies. In the simulations presented in the next sections we used this latter scheme with a value $h = 2\Delta x$ and a spacing $\Delta x = 1\text{\AA}$. The effect of several resolution and interpolation schemes is discussed in the last part of the manuscript. It is a simple matter to verify that the chosen smearing function sums over the lattice to unity. In addition it guarantees the invariance over continuous translations.

For a given choice of the interpolation kernel, the fluid velocity at the particle position is taken to be the weighted velocity,

$$\tilde{\mathbf{u}}(\mathbf{R}_i) = \sum_{\mathbf{x}} \mathbf{u}(\mathbf{x}) \delta_h(\mathbf{x} - \mathbf{R}_i) \quad (8)$$

Clearly, the wider is the support the heavier is the computational effort, with the resulting discrete convolutions being proportional to 8, 64, ... points for $h = \Delta x, 2\Delta x, \dots$.

The particle dynamics is given by the following equations of motion:

$$\begin{aligned} \dot{\mathbf{R}}_i &= \mathbf{V}_i \\ \dot{\mathbf{V}}_i &= \frac{\mathbf{F}_i}{M_i} - \gamma(\mathbf{V}_i - \tilde{\mathbf{u}}_i) + \boldsymbol{\mu}_i \end{aligned} \quad (9)$$

$\boldsymbol{\mu}$ is a white noise with $\langle \boldsymbol{\mu}_i(t) \rangle = 0$ and $\langle \boldsymbol{\mu}_i(t) \boldsymbol{\mu}_j(t') \rangle = 2\gamma k_B T / M_i \delta(t - t') \mathbf{I}$ where k_B is the Boltzmann constant and T the temperature. Eqs. (9) are integrated in time over the timestep t_{MD} according to the Symplectic Position Verlet algorithm⁵⁷ and, if $t_{MD} = t$, the particle

and LB dynamics are updated in a synchronous way. It is worth noting that by setting $\tilde{\alpha} = 0$ the standard Langevin dynamics is recovered.

In order to observe hydrodynamic behavior down to the λ scale, it is usually considered that the fluid mean free path should not exceed $\lambda/8$. In our case, we lift this requirement by allowing for a sub-nanometric mesh spacing so that hydrodynamic behavior will spontaneously emerge over distances larger than λ . This strategy has proven to be accurate when the LB method is used to simulate nanofluids^{59,60}. At the same time, as shown in previous literature^{41,61}, given the smooth BGK collisional operator HI between pairs of particles is reproduced when λ is comparable to the interparticle distance. This occurrence should be considered as an artefact of the model, since the hydrodynamic scale should set in only for distances larger than a few (typically > 2) molecular diameters. Consistently, choosing the size of λ , as compared for instance to the atomic size, is a matter of tradeoff between including finer although somehow over-hydrodynamic, versus coarser interactions and the computational effort.

In the temporal domain, the LB timestep Δt can be chosen to be comparable (or synchronous) to Δt_{MD} . However, the speed of sound of the LB fluid cannot be modified and, by taking $\lambda = 1 \text{ \AA}$ and $\Delta t = 1 \text{ fs}$, it results to be about 100-fold faster than for real water, that is, the LB solvent is much less compressible than what water is in reality. Such difference would be minimized by considering a much larger LB timestep, but at the cost of reducing numerical robustness. For our needs, however, an artificially faster speed of sound does not spoil the accuracy of HI.

A schematic view of the coupling between LB methods and the molecular dynamics technique is given in figure (1). The coupling has been implemented in the in-house developed code for multi-physics simulation, MUPHY⁶². A release of the coupled engine is planned for the next future.

2.2.3 Thermal noise in the fluid—Finally, in order to be consistent with the microscopic treatment of the protein particles and the fluid, the effect of thermal noise needs to be taken into consideration also at the fluid mesoscopic level. The stochastic version of the LB is given by the following modification of eq. (2)⁶¹

$$f_p(\mathbf{x} + \mathbf{c}_p \Delta t, t + \Delta t) = f_p(\mathbf{x}, t) - \omega \Delta t (f_p - f_p^{eq})(\mathbf{x}, t) + g_p(\mathbf{x}, t) + \xi_p(\mathbf{x}, t) \quad (10)$$

where the fluctuating term $\xi_p(\mathbf{x}, t)$ acts at the level of the stress tensor and non-hydrodynamic modes and reads

$$\xi_p = \sqrt{\frac{\rho k_B T \omega \Delta t (2 - \omega \Delta t)}{\Delta x^3 c_s^2}} \sum_{k=4}^{18} w_p \chi_{kp} \mathcal{N}_k$$

$\{\chi_k\}_{k=0,18}$ is a set of 18 adimensional lattice eigenvectors and orthonormal according to the scalar product

$$\sum_{p=0}^{18} w_p \chi_{kp} \chi_{lp} = \delta_{kl} \quad (11)$$

$$\sum_{k=0}^{18} w_p \chi_{kp} \chi_{kp'} = \delta_{pp'} \quad (12)$$

and placed in a box of dimension L is a set of 15 random numbers extracted from a Gaussian distribution of zero mean and unit variance. The eigenvectors $\{\chi_k\}$ are constructed based on the LB scheme in use (as for D3Q19), the interested reader can find the technical details in ref.61. They correspond to hydrodynamic moments: $k=0$ is relative to mass density, $k=1-3$ to mass current, $k=4-9$ on the (symmetric) momentum flux tensor, while the remaining $k=10-18$ correspond to non-hydrodynamic modes. The kinetic eq. (10) reproduces the

fluctuating Navier-Stokes equation⁶³, $\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla P + \nu \nabla^2 \mathbf{u} + \frac{1}{\rho} \nabla \cdot \overleftrightarrow{\sigma} + \frac{\mathbf{G}}{m}$, where $\overleftrightarrow{\sigma}$ is the stochastic component of the stress tensor. The stochastic force acts at the level of the smallest length scale Δx and produces consistent fluctuations up to the largest wavelength of the simulation box⁶¹.

2.3 Multi-time and space resolution

The coupled Lattice Boltzmann and Molecular Dynamics (LBMD) techniques provide an intrinsically multi-physics multi-scale framework for molecular simulations. The further use of the OPEP force field specializes the technique to proteins in solution. The multi-scaling can be designed at different levels, concerning both the molecular or fluid resolution. First, the molecular motion can be evolved using a multiple timestep technique for integrating separately bonded and non bonded interactions of the OPEP Hamiltonian. Moreover, the molecular representation based on the OPEP force field can be further grained by simplifying the intramolecular interactions and reduce them to an effective elastic networks or to a G-like model. On the other hand, the lattice resolution underlying the fluid dynamics can be changed in order to reduce the cost of accounting for hydrodynamics. Finally, the coupling can be designed to be synchronous, with the fluid and the molecular motion propagated at equal times, or asynchronous, with the fluid evolved only every n molecular steps. A detailed discussion of the multi-scaling strategies is reserved to the final section of this work.

3 Results

In the following, we present a detailed testing of the methodology. At first, we have investigated the capabilities of our implementation and inquired the long-time stability of a protein in the hydrodynamic fluid. Then, we have explored the effect of HI on the kinetics of

amyloid aggregation, a relevant biological process. Finally, we have benchmarked the multi-scale features of the LBMD technique. For the sake of clarity, in table 1 we summarise the body of simulations performed here with the relevant parameters used. Details about the simulation performances are reported in SI, Table S.1.

3.1 Stability of model proteins

We have considered five small proteins previously used to assess the quality of the latest versions (v.436 and v.564) of the OPEP force field. The size of these proteins ranges from 37 to 75 amino acids, see table 2. For each protein we performed simulations of about 90/100 ns with and without hydrodynamics. Our simulations show that OPEP is capable of maintaining the native fold of the proteins stable within 100 ns, in absence and also in presence of the surrounding fluid. In table 2 we report the average values of a set of collective variables (CV) monitoring the stability of the proteins, namely the C_α root mean square deviation (rmsd) calculated with respect to the initial configuration (from NMR or X-ray), the fraction of native contact (Q) and the fraction of torsional angle (n_t). The formal definition of these variables⁶⁵ is given in SI; here we just mention that the rmsd was computed only for the rigid core of the protein, the well defined region by NMR as reported in table 2. For all the CVs, the average was computed using the all the trajectory length and is reported in table 2, in each column the first value refers to the solvent free MD and the second one to the simulations including the LB evolution (LBMD).

By visual inspection we first note that the protein relaxed configurations look very similar in both simulations. This can be also appreciated by looking at the final steady values of rmsd and n_t in Figure 2. This result is not surprising since hydrodynamics does not affect the thermodynamics of the system, hence the states at equilibrium (and their distribution) should be the same. Averaged over the full trajectories, the five proteins display an average rmsd varying between 3.0 (1EOL) and 4.3 (1SHG) in solvent free MD and between 2.7 (1CLB) and 4.0 (1SHG) including hydrodynamics. At the end of the simulations, three proteins over five exhibit a very low rmsd, about 3/4 Å, in two other cases (1FCL and 1SHG proteins) the deviation from the initial state is larger but the drift is mainly due to the rearrangement of the secondary structures with no systematic unfolding observed, see details in SI. However, when looking at the relaxation process in itself, figure 2, it appears that the protein evolution is somehow softened by HI, and we observe longer lived intermediate states. This is also mirrored by the different average values for each CVs when computed along all the trajectories, see table 2.

Since the coupling between protein dynamics and the fluid is encoded in the friction coefficient γ , see eq. (4), we tested its effect on protein structural relaxation. From the final equilibrated configuration of the protein 1EOL, we restarted a series of simulations with different value for the parameter γ and re-assigned the particle velocities. For the pure Langevin dynamics, γ relates to the memory-less motion, but when HI are included the coupling accounts for both the fluid viscosity and the effective hydrodynamic radius of the molecular center of force, i.e. according to Stokes law $\gamma = 6\pi\eta a/m$ being η , a and m respectively the dynamic viscosity of the fluid and the size and the mass of the particle. In figure 4 the time evolution of the fraction of native torsion n_t is reported for the two type of

simulations and for several values of γ . During the solvent free MD the protein tends to drift away from the initial structure, occurring for all values of γ considered. Diffusion of a single Langevin particle is inversely proportional to the friction coefficient, $D = kT/m\gamma$, so that in the weak coupling regime particle motion is faster. On the contrary, when HI are included, the system shows a long-lived memory and tightly fluctuates around the starting configuration. Only in the limit of vanishing coupling the relaxation drives the protein toward new configurational states. This finding shows how HI drain or pump energy in the protein by altering the relaxation process from one state to another. Interestingly, while HI have been reported to accelerate folding events in G-like model^{26,31}, by favouring the collapse towards the compact structure, no author has yet studied how HI affect the internal relaxation of a folded protein. Our preliminary results indicate that for the compact folded state the system may survive longer in conformational substates when HI are included.

The possibility to control the coupling between the fluid evolution and the protein internal dynamics, for example by tuning either the friction γ or the viscosity η , makes our framework a powerful tool to investigate how protein motion is influenced by the solvent viscosity without an explicit representation of the latter. A related, interesting debate concerns the effect of solvent viscosity on protein folding kinetics^{66–68}.

A weak dependence on solvent viscosity has been reported experimentally and rationalised as due to protein internal friction on dominating the folding kinetics^{67,68}.

Moreover, we also mention that the coupling with an effective solvent also opens the possibility to explore energy transfer processes from solvent to protein and vice versa⁶⁹ in the context of water-free CG models.

In the current implementation, the coupling parameter γ is the same for all molecular centers of force but heterogenous, per-particle values depending on the amino-acid nature or local topology of the protein/solvent interface can be easily incorporated. In fact, recently all atoms simulations clearly probed the heterogeneity of the solvent viscosity in proximity of biological interfaces⁷⁰.

3.2 Transport properties

We now turn our attention to the transport properties of the proteins. From the performed simulations we extracted the translational and rotational diffusion constants. For the former we performed a linear fit of the mean square displacement (*MSD*) at long time scale (> 40 ns) and extracted the diffusion constant D_T using the Einstein relation, $\langle \delta R^2 \rangle = 6D_T t$. The rotational diffusion constant was calculated by considering the second order rotational relaxation of selected rigid vectors of the protein structures. Namely for each molecule we have individuated two vectors associated to the orientation of two well defined rigid secondary structure elements. The fragment of the proteins defining these vectors are reported in Table 3. The decay ($t < 30$ ns) of the time-correlation functions of the vector

orientation, $C_2(t) = \langle \frac{1}{2}(3\cos(\theta(t)) - 1) \rangle$ were then fitted with an exponential function, $C_2(t) = e^{-6D_R t}$. The estimated values of D_T and D_R are reported in table 3 for all the

simulations. Following the framework of²⁶, we have also estimated for our systems the experimental diffusion constants by using the software HYDROPRO⁷¹.

As already reported by other authors, i.e.^{26,29,33}, the HI systematically enhance the diffusivity of the molecules in the solution. For example when HI are included in the Brownian Dynamics of Go-like model of proteins the diffusion coefficient for folded and unfolded proteins²⁶ is a factor 20 larger as compared to the absence of HI, and are essential to correctly reproduce the experimental data. In our case, the speed up is about a factor two for the translational motion. For the LBMD simulations the D_T is comprised in the range 4-8 $\text{\AA}^2/\text{ns}$, and it results about 2-3 times smaller than the experimental data. The same behaviour is recovered for all but one system when the rotational diffusion is considered, D_R being 1.2 to 4 times larger when HI are included in the simulation than in the pure Langevin dynamics. The only exception is the 1LCB system for which the values for D_R are close to each other, but HI slightly reduces rotational motion. The values calculated from the LBMD trajectories are, as for translational motion, 1.2-4 times smaller than the experimental estimates. However, as already anticipated, the mobility of the protein depends on the parameters characterising our fluid. In particular we have probed for one protein (1E0L) that a fine tuning of the friction γ results in a very good agreement with the experimental HYDROPRO value. In fact, when using $\gamma = 0.05 \text{ fs}^{-1}$ we found that the diffusion constants are $D_T = 17 \text{ \AA}^2/\text{ns}$ and $D_R = 4.6 * 10^{-2} \text{ ns}^{-1}$.

As already reported in³⁸, the LBMD scheme based on the OPEP force field was delivered to investigate the effect of crowding on protein mobility, showing that for high occupied volume fraction, $\sim 30\%$, diffusion at the nanosecond time-scale slows down of about one order of magnitude with respect to the dilute solution limit, in good agreement with data from neutron scattering⁹.

3.3 Amyloid aggregation

When applying the computational scheme to the problem of amyloid aggregation, the inclusion of HI in CG simulations is expected to crucially influence the kinetics of molecular aggregation^{22,33,34}. We consider a system composed of 18 monomers of the peptide $A\beta_{16-22}$ capped by Ace and NH_2 groups and placed in a box of dimension $65 \times 65 \times 65 \text{ \AA}^3$, corresponding to a concentration of 100 mM. The system contains 864 particles and is simulated for about 100/150 ns depending on the simulation scheme. The multi-scale simulation is compared to the standard Langevin dynamics. The aggregation process is quantified by calculating the gyration radius of the system, R_g , the time evolution of which is reported in figure 5. Our results clearly show that when OPEP is coupled to the solvent the initial aggregation process speeds up by about 15%. In the presence of HI the $A\beta_{16-22}$ monomers collapse quickly and a first steady state is reached in about 20 ns, with the size of the aggregate being $R_g \simeq 15.7 \text{ \AA}$. A second minor structural reorganisation is observed at a longer time, 50 ns, making the aggregate more compact as a result of an optimized packing. This state persists for the rest of the simulation. A similar two-step mechanism is also observed for the pure MD but the size of the intermediate and final states are rather larger than the ones observed in the active fluid. The first aggregation takes place in about 20 ns, the metastable structure has a $R_g \simeq 18 \text{ \AA}$. The final collapse occurs at about 70 ns, is

characterised by slow kinetics and leads to an aggregate of $R_g \simeq 16 \text{ \AA}$. A pictorial representation of the aggregation is given in figure 7.

This finding clearly shows how solvent mediated interactions facilitate the encounter and structural shaping of $A\beta_{16-22}$ monomers, with the final collapse being accelerated by about 40%. Our result agrees with recent simulations reported by Ando and Skolnik³³ where, using a sort of specialised dumbbell model, it was probed that HI substantially increase the rate of lipid aggregation. The faster collapse of lipids was rationalised in terms of a reaction-diffusion model⁷² and by observing that for polymeric entities modelled as a series of N linked beads, the encounter rate is favoured by HI due to the different scaling of the diffusion constant with the polymer size as compared to the free draining case. With HI the diffusion coefficient scales as $D^{HI} \approx D_0 N^{-\nu}$ with $0 < \nu < 1$ whereas, in the free draining limit, $D^{FD} \approx D_0 N^{-1}$.

We therefore compute the diffusion constant for the monomers in our system as a function of time. This allows detecting changes in monomer diffusivity during the aggregation process. The trajectories are separated in time windows of 12 ns each. For each time window we calculate the *MSD* of the monomer center of mass and then average over all peptides. The average diffusion constant is extracted from the linear fit of the *MSD*(t) for each window. The obtained values are reported in figure 6.

We first notice that the mobility of the monomers is enhanced by HI by a factor 2.4 in the initial time window where the solution can be considered rather dispersed. As the aggregation process proceeds, the speed-up of monomers mobility reduces to a factor 1.5 with respect to the solvent free MD. An accurate investigation of the diffusion of amyloid proteins such as $A\beta$ in a hydrodynamic fluid along with its dependence on fluid friction and viscosity is reserved to a forthcoming work; here we just mention that with HI the initial diffusion coefficient is about a factor four smaller than the experimental value $D = 2.3 \cdot 10^{-10} \text{ m}^2/\text{s}$ ($23 \text{ \AA}^2/\text{ns}$)⁷³. This difference is caused by two factors. First, as reported in the previous paragraph, with the initial choice of our parameters protein diffusion results 2-3 times slower than in experiments. Secondly, because of the high concentration even in the initial time window, the monomers do not diffuse as in the very dilute solution.

From the single long simulations we just presented it is not possible to extract statistically meaningful information about the kinetics of the collapse. To add extra weight to these preliminary results we have generated five initial states, all sharing an initial value for the gyration radius of the system, $R_g \simeq 22 - 23 \text{ \AA}$. The time evolution of R_g was then average over the independent trajectories. The results plotted in the bottom panel of figure 5 confirms that HI speed up the initial encounter of the peptides and their collapse.

A systematic test of how the diffusivity of amyloid species depends on the essential parameters of our simulations, i.e. the friction γ and the fluid viscosity ν , and on system concentration is reserved to a further investigation. However, our result confirms that also for the correct modelling of molecular aggregation in aqueous solution it is essential to include HI³³.

3.4 Towards crowding

In this final section, we unfold the potentiality of the LBMD technique to tackle the problem of protein crowding. The inclusion of HI in simulations of systems composed by tens, hundreds and ultimately thousands of proteins has a heavy computational cost. Indeed, in terms of the LB, the computational effort scales cubically with the number of linear lattice points. In the test cases presented so far, we used a high resolution scheme for the LB part with a lattice spacing $\Delta x = 1 \text{ \AA}$. In these conditions, depending on the size of the molecular system, the LB component and the LB-MD coupling contribute 30% – 60% of the total elapsed time, so that it is desirable to reduce the computational effort of handling the solvent.

Since OPEP, as other CG models, casts together multiple atoms to model the aminoacid side-chains (SC), thus it associates to each SC centre of force a characteristic radius of several Angstroms, the resolution of the HI lattice can be decreased without compromising significantly the accuracy of the overall method. We have tested this strategy on the amyloid system. In figure 8 we report the elapsed time for a parallel simulation of 1000 MD steps allocated on 8 cores (Intel xeon 5660) for several resolutions of the LB mesh, Δx . The red dots refer to a synchronous coupling between MD and LB, thus HI are evaluated at each MD step. As expected, the simulation time scales approximately as the cube of the mesh size and the benchmark points can be fitted with a function $t = A * (\Delta x^{-\alpha} + t_0)$. For synchronous time-stepping $\alpha = 3.12$, while for asynchronous time-stepping (HI evaluated each 2 MD steps), $\alpha = 1.95$. The value $A t_0 \sim 10 \text{ s}$ is the same for both calculations, signalling the irreducible time due to MD plus the particle / fluid coupling. In fact, the coupling weights about 60% ($\sim 6 \text{ s}$), and a standard 1000 MD steps costs about 4 s. For a selected protein, the 1E0L, we have probed that small changes in the lattice resolution have a very negligible impact on both protein stability and diffusion. For instance with a spacing $\Delta x = 1.33$ and $2.0 \text{ \AA}$, the average RMSD over a trajectory of 100 ns is 2.7 and $3.5 \text{ \AA}$, respectively. Moreover the effect on the diffusion constant is less than 10%. How lattice resolution influences the kinetic of complex processes like the amyloid aggregation will be discussed in a separated work.

As we increase the LB mesh spacing, the computational gain can be further improved by using a less expensive interpolation scheme to map the particle positions on the mesh points. The data presented in figure 8 are obtained by using the interpolation scheme according to eq. (6) and spanning 4 lattice spaces, as described in the section Method. Using a simpler scheme such the nearest point interpolation, the overall simulation cost further drops down by about 30%.

The performance of the simulations can be further improved by tuning the dynamics of the molecular systems. When the dynamics of the protein is governed by the complete OPEP Hamiltonian, the timestep for the simulations is indeed confined to very small values, practically the typical values used in standard all atom simulations, $1\text{--}2 \text{ fs}$. However, when dealing with a protein suspension composed of many molecules, a fast initial sampling of the molecular packing does not require the finest details of the molecular motion. In this respect, several strategies can be adopted for enlarging the integration step, such as to define an harmonic elastic network to restrain the protein motion or a simplified G-like model

based on the OPEP force field. The associated grained protein motion would be coupled to a grained fluid motion as discussed above. The finest resolution can be recovered to investigate in detail selected states of the suspension.

When molecular dynamics is evolved with large timesteps the coupling with HI must be carefully scaled in order to maintain a physical control of the parallel-time evolution. As discussed in the method, the intrinsic timescale of the fluid is controlled by the dynamic viscosity η (or ν) while the intrinsic timescale of the molecular system is encoded by the physical Hamiltonian and reflected by the integration timestep. In synchronous LBMD, the fluid relaxation is evaluated at each MD step, thus when the molecular system evolves with a larger timestep, the fluid relaxation occurs at a slower rate. In order to preserve the same physical viscosity of water and its internal dynamics, the numerical viscosity should be scaled in a such way that $\nu = \nu_0 * (\tau_0/\tau)$ where τ_0 and ν_0 are reference values. We have probed the scaling behaviour by considering the zero temperature velocity relaxation of an amyloid peptide $A\beta_{16-22}$ in the fluid. We have assigned initial velocities (10^{-4} Å/fs) to each particle of the peptide and followed the time relaxation of the velocity of the peptide centre of mass. When HI interactions are included, three different time scales are expected⁷⁴. As a reference we have considered the relaxation process governed by a MD integration timestep $\tau_0 = 1$ fs and a fluid viscosity $\nu_0 = 0.166$ (Å²/fs). The associated data are reported in figure 9. The relaxation was also monitored using a different timestep $\tau = 0.5$ and 1.5 fs, and using either a viscosity $\nu = \nu_0$ (solid red and orange lines) or a scaled one, $\nu = \nu_0 * (\tau_0/\tau)$ (red and orange dashed lines). The time relaxations obtained via the scaled approach accurately reproduce the target relaxation process.

4 Conclusions

The coupling of water-free, coarse-grained model with hydrodynamic interactions is an effective strategy to study protein suspensions and provides complementary informations to all atoms simulations, which necessarily accesses a smaller number of proteins and shorter timescales. A few research groups are active in this field and presented so far challenging applications^{16,19,21,22,24,26,33,34}. In the present work, we described a novel implementation that couples the OPEP force field for proteins with the Lattice Boltzmann methodology to resolve explicitly and on-the-fly the solvent dynamics, leading to the hydrodynamic interactions.

The framework combines the excellent performances of OPEP in reproducing the proteins motion at quasi-atomistic resolution, with the excellent scalability of the LB technique on parallel architectures, making the scheme a powerful tool for inquiring proteins structure and dynamics under crowding conditions³⁸. We tested the simulation technique and showed that it is effective in monitoring protein fluctuations and mobility in the folded state as well as in peptide aggregation.

In perspective, the proposed modeling of the solvent can be improved along the mesoscopic Lattice Boltzmann model for water-like fluids described in refs.^{75,76}. By water-like fluids, we imply that the fluid may include generic features of the hydrogen bonds, particularly the tendency to form ordered domains, with selected bond angles, against the disordering action

of thermal noise. The possibility to merge the OPEP force field with the water-like treatment of the solvent opens new ways to describe the fluid solvent with an accurate thermodynamics yet with a minimal set of degrees of freedom. As a result, the intrinsic multi-scale nature of the methodology would fully unfold and find application in those biological problems where such level of detail is required.

6 Supporting Information

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no.258748. Part of his work was performed using HPC resources from GENCI [CINES and TGCC] (Grant x201376818 and x2014076818). We acknowledge the financial support for infrastructures from ANR-11-LABX-0011-01. We thanks M. Chiricotto, M. Bernaschi e M. Bisson for useful discussions.

References

- (1). Ellis RJ. Trends Biochem Sci. 2001; 26:597–604. [PubMed: 11590012]
- (2). Zhou HX, Rivas G, Minton AP. Annu Rev Biophys. 2008; 37:375–97. [PubMed: 18573087]
- (3). Dix JA, Verkman A. Annu Rev Biophys. 2008; 37:247–263. [PubMed: 18573081]
- (4). Hofling F, Franosch T. Rep Prog Phys. 2013; 76:046602. [PubMed: 23481518]
- (5). Wang Y, Mohona S, Smith AE, Krois AS, Pielak GJ. J Am Chem Soc. 2012; 134:16614–16618. [PubMed: 22954326]
- (6). Jin S, Verkman A. J Phys Chem B. 2007; 111:3625–3632. [PubMed: 17388520]
- (7). Malchus N, Weiss M. J of Fluoresc. 2010; 20:19–26. [PubMed: 19582558]
- (8). Verkman A. Meth Enzymol. 2003; 360:635–648. [PubMed: 12622172]
- (9). Roosen-Runge F, Hennig M, Zhang F, Jacobs RMJ, Sztucki M, Schober H, Seydel T, Schreiber F. Proc Natl Acad Sci USA. 2011; 108:11815–11820. [PubMed: 21730176]
- (10). Konopka MC, Shkel IA, Record SCM, Weisshaar J. J Bacteriol. 2006; 188:6115–6123. [PubMed: 16923878]
- (11). Wang Y, Li C, Pielak GJ. J Am Chem Soc. 2010; 132:9392–9397. [PubMed: 20560582]
- (12). Miklos AC, Sarkar M, Wang Y, Pielak GJ. J Am Chem Soc. 2011; 133:7116–7120. [PubMed: 21506571]
- (13). Monteith WB, Pielak GJ. Proc Natl Acad Sci USA. 2014; 31:11335–11340.
- (14). Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Walchli M, Smith BO, et al. Nature. 2009; 458:102–105. [PubMed: 19262674]
- (15). Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, Nakase I, Takeuchi T, Futaki S, Ito Y, Hiroaki H, Shirakawa M. Nature. 2009; 458:106–110. [PubMed: 19262675]
- (16). Frembgen-Kesner T, Elcock AH. Biophys Rev. 2013; 5:109–119. [PubMed: 23914257]
- (17). Bicout DJ, Field MJ. J Phys Chem. 1996; 100:2489–2497.
- (18). Ridgway D, Broderick G, Lopez-Campistrous A, Ruaini M, Winter P, Hamilton M, Boulanger P, Kovalenko A, Ellison MJ. Biophys J. 94:3748–3759. [PubMed: 18234819]
- (19). McGuffee SR, Elcock AH. PLoS Comput Biol. 2010; 6:e1000694. [PubMed: 20221255]
- (20). Harada R, Tochio N, Kigawa T, Sugita Y, Feig M. J Am Chem Soc. 2013; 135:3696–3701. [PubMed: 23402619]
- (21). Ando T, Skolnick J. Proc Natl Acad Sci USA. 2010; 107:18457–18462. [PubMed: 20937902]
- (22). Mereghetti P, Wade RC. J Phys Chem B. 2012; 116:8523–8533. [PubMed: 22594708]
- (23). Feig M, Sugita Y. J Mol Graph Model. 2013; 45:144–156. [PubMed: 24036504]

- (24). Hasnain S, McClendon C, Hsu M, Jacobson M, Bandyopadhyay P. PLoS ONE. 2014; 9:e106466. [PubMed: 25180859]
- (25). Gruebele M, Thirumalai D. J Chem Phys. 2013; 139:121701. [PubMed: 24089712]
- (26). Frembgen-Kesner T, Elcock AH. J Chem Theory Comput. 2009; 5:242–256. [PubMed: 26610102]
- (27). Szymczak P, Cieplak M. J Phys Condens Matter. 2011; 23:033102. [PubMed: 21406855]
- (28). Durlinsky L, Brady JF, Bossis G. J Fluid Mech. 1987; 180:21–49.
- (29). Ermak DL, McCammon JA. J Chem Phys. 1978; 69:1352–1360.
- (30). Ando T, Chow E, Skolnick J. J Chem Phys. 2013; 139:121922. [PubMed: 24089734]
- (31). Cieplak M, Niewiczzerza S. J Chem Phys. 2009; 130:124906. [PubMed: 19334888]
- (32). Kamata K, Araki T, Tanaka H. Phys Rev Lett. 2009; 102:108303. [PubMed: 19392167]
- (33). Ando T, Skolnick J. Biophys J. 2013; 104:96–105. [PubMed: 23332062]
- (34). Frembgen-Kesner T, Elcock AH. Biophys J. 2010; 99:L75–L77. [PubMed: 21044566]
- (35). Song W, Wei G, Mousseau N, Derreux P. J Phys Chem B. 2008; 112:4410–4418. [PubMed: 18341325]
- (36). Chebaro Y, Pasquali S, Derreux P. J Phys Chem B. 2012; 116:8741–8752. [PubMed: 22742737]
- (37). Sterpone F, Nguyen P, Kalimeri M, Derreux P. J Chem Theory Comput. 2013; 9:4574–4584. [PubMed: 25419192]
- (38). Sterpone F, Melchionna S, Tuffery P, Pasquali S, Mousseau N, Cragolini T, Chebaro Y, St-Pierre J-F, Kalimeri M, Barducci A, Laurin Y, et al. Chem Soc Rev. 2014; 43:4871–4893. [PubMed: 24759934]
- (39). Benzi R, Succi S, Vergassola M. Phys Rep. 1992; 222:145–197.
- (40). Chen S, Doolen GD. Ann Rev Fluid Mech. 1998; 30:329–364.
- (41). Ahlrichs P, Duenweg B. J Chem Phys. 1999; 111:8225.
- (42). Maupetit J, Tuffery P, Derreux P. Proteins: Structure, Function and Genetics. 2007; 69:394–408.
- (43). Derreux P. J Chem Phys. 1997; 106:5260–5270.
- (44). Derreux P, Mousseau N. J Chem Phys. 2007; 126:025101. [PubMed: 17228975]
- (45). Spill YG, Pasquali S, Derreux P. J Chem Theory Comput. 2011; 7:1502–1510. [PubMed: 26610141]
- (46). Nguyen PH, Okamoto Y, Derreux P. J Chem Phys. 2013; 138
- (47). Barducci A, Bonomi M, Derreux P. J Chem Theory Comput. 2011; 7:1928–1934. [PubMed: 26596453]
- (48). Chebaro Y, Derreux P. Proteins. 2009; 75:442–452. [PubMed: 18837034]
- (49). Chebaro Y, Jiang P, Zang T, Mu Y, Nguyen PH, Mousseau N, Derreux P. J Phys Chem B. 2012; 116:8412–8422. [PubMed: 22283547]
- (50). Nasica-Labouze J, Meli M, Derreux P, Colombo G, Mousseau N. PLoS Comput Biol. 2011; 7:e1002051. [PubMed: 21625573]
- (51). Kalimeri M, Derreux P, Sterpone F. J Non Crys Sol. 2015; 407:494–501.
- (52). Marconi UMB, Melchionna S. J Chem Phys. 2009; 131:014105. [PubMed: 19586094]
- (53). Miller AA. J Chem Phys. 1963; 38:1568–1571.
- (54). Guo Z, Zhen C, Shi B. Phys Rev E. 2002; 65:046308.
- (55). Shan X, Yuan X, Chen H. J Fluid Mech. 2006; 550:413.
- (56). Fyta M, Kaxiras E, Melchionna S, Succi S. Comput Sci Eng. 2008; 3:10.
- (57). Melchionna S. J Chem Phys. 2007; 127:044108. [PubMed: 17672682]
- (58). Succi S. The Lattice Boltzmann Equation for Fluid Dynamics and beyond. Clarendon Press; 2001.
- (59). Melchionna S, Marconi UMB. Europhys Lett. 2008; 81
- (60). Horbach J, Succi S. Phys Rev Lett. 2006; 96:224503. [PubMed: 16803312]
- (61). Duenweg B, Ladd A. Adv Polym Sci. 2008; 221:89.

- (62). Bernaschi M, Melchionna S, Succi S, Fyta M, Kaxiras E, Sircar J. *Comput Phys Comm.* 2009; 180:1495–1502.
- (63). Landau, LD.; Lifshitz, EM. *Fluid Mechanics*. Pergamon Press; London: 1963.
- (64). Rahaman O, Melchionna S, Laage D, Sterpone F. *Phys Chem Chem Phys.* 2013; 15:3570–3576. [PubMed: 23381660]
- (65). Kalimeri M, Rahaman O, Melchionna S, Sterpone F. *J Phys Chem B.* 2013; 117:13775–13785. [PubMed: 24087838]
- (66). Schulz JCF, Schmidt L, Best RB, Dzubiella J, Netz RR. *J Am Chem Soc.* 2012; 134:6273–6279. [PubMed: 22414068]
- (67). Chung HS, Eaton WA. *Nature.* 2013; 502:685–688. [PubMed: 24153185]
- (68). de Sancho S, Sirur A, Best RB. *Nature Comm.* 2014; 5:4307.
- (69). Moritsugu K, Kidera A, Smith J. *J Phys Chem B.* 2014; 118:8559–8565. [PubMed: 24999844]
- (70). Pronk S, Lindahl E, Kasson PM. *Nature Comm.* 2014; 5:3034.
- (71). Torre GDL, Huertas JML, Carrasco B. *Biophys J.* 2000; 78:719–730. [PubMed: 10653785]
- (72). Northrup SH, Hynes JT. *J Chem Phys.* 1979; 71:871–883.
- (73). Danielsson J, Jarvet J, Damberg P, Gräslund A. *Magn Reson Chem.* 2002; 40:S89–S97.
- (74). Hinch EJ. *J Fluid Mech.* 1975; 72:499–511.
- (75). Mazzitelli I, Venturoli M, Melchionna S, Succi S. *J Chem Phys.* 2011; 135:124902. [PubMed: 21974559]
- (76). Moradi N, Greiner A, Melchionna S, Rao F, Succi S. *Phys Chem Chem Phys.* 2014; 16:15510–15518. [PubMed: 24953220]

Figure 1.

Schematic representation of the coupling between the molecular system and the fluid. The fluid "particles" are associated to a set of discrete probability distributions $f_p(x, t)$ where p indicates a discrete velocity along one of the lattice directions. The drag force F_i^D on the molecular particle i is obtained by interpolating the lattice velocity $\tilde{u}$ at the particle position.

Similarly, the drag force F_i^D and the random force μ_i act on the fluid and their values are extrapolated at the lattice point. Several schemes are available to map particles positions and lattice points, see text for details.

Figure 2.

Time evolution of the fraction of native torsion n_t and RMSD for the set of proteins. Data in green refer to simple Langevin dynamics, data in yellow refer to simulation with HI (LBMD).

Figure 3.

Time line of the fraction of native torsions for 1EOL. Restart from the equilibrated structure, and using different value of the friction γ for the Langevin dynamics (bottom) and LBMD (top).

Figure 4.
The protein 1E0L embedded in the hydrodynamic fluid: pictorial representation of the fluid streamlines pattern around the protein.

Figure 5.

Top. Effect of hydrodynamic interactions on amyloid aggregation as monitored by the time evolution of the radius of gyration, R_g . Orange and green curves refer to the simulations with and without hydrodynamics, respectively. The exponential fit of the initial collapse is reported in solid lines (red and light green). The characteristic time τ of the initial collapse extracted by the exponential fit, has value of $\tau = 7$ and $= 8$ ns, for the simulation including HI and the one without. Bottom. Initial aggregation of the peptides as it results from the

average of the time decay of R_g over independent trajectories. The error bars indicate the variability of the data computed as the (Max-Min)/2.

Figure 6. Average diffusion coefficient D calculated in several time window of 12 ns each by fitting the mean square displacement of the peptides centers of mass. Orange and green symbols refer to simulations including or not hydrodynamics interactions, respectively.

Figure 8.

Benchmark of the multi-resolution scheme. The time to produce 1000 MD steps on 8 cores for the system composed of 18 $A\beta$ monomers is reported as a function of the grid spacing. Red dots refer to synchronous coupling between the MD and the LB. Maroon dots refer to asynchronous coupling, with LB integration every two MD steps. The dashed lines indicate a fit, see text.

Figure 9.

Time-relaxation of the CM velocity at zero temperature in synchronous MD and LB dynamics. Black circles refer to the reference tandem, molecular timestep $\tau_0 = 1$ fs and fluid viscosity $\nu_0 = 0.166(\text{\AA}^2/\text{fs})$. Solid red and green lines refer to relaxation processes sampled by MD time step $\tau = 1.5$ fs and $\tau = 0.5$ fs and viscosity ν_0 ; dashed lines refer to the same process sampled using a rescaled fluid viscosity, $\nu = \nu_0 * (\tau/\tau_0)$.

Table 1

Summary of the simulations performed to test the LBMD methodology and values of the most important parameters used in the simulations. The asterisk indicates tests where parameters have been modified.

System	f.f.	Time [ns]	n particles	time step [fs]	L_x [Å]	x [Å]	ν [Å ² /fs]	γ [fs ⁻¹]
1CLB	v.4	90/100	447	1.5	50	1.0	0.1666	0.1
1EOL	v.4	90/100	221	1.5	40	1.0	0.1666	0.1
1FCL	v.4	90/100	332	1.5	40	1.0	0.1666	0.1
1SHG	v.4	90/100	342	1.5	40	1.0	0.1666	0.1
2DA1	v.4	90/100	418	1.5	50	1.0	0.1666	0.1
1EOL*	v.4	30	221	1.5	40	1.0	0.1666	0.1/0.001/0.001
A β_{16-22}	v.5	110/150	864	1.5	65	1.0	0.1666	0.1
A β_{16-22} *	v.5	50	864	1/1.5	65	1.3/1.44/1.857/2.6	0.1666	0.1

Table 2

Set of proteins investigated using OPEP v.4 coupled with hydrodynamic interactions. Simulations were performed at T=300 K. The residues forming the rigid core of each protein are reported in the third column. For the simulation time and the three collective variables, rmsd, fraction of native torsion (n_t) and fraction of native contacts (Q) we report data from simulations without and with hydrodynamic coupling, left and right data respectively in each column. The rigid core residues are indicated according to the PDB numbering.

system	size [a.a.]	rigid core	time [ns]	rmsd [$\text{\AA}$]	n_t	Q
1CLB	75	3-14,20-39,47-53,60-73	100/90	3.5/2.7	0.71/0.81	0.68/0.79
1EOL	37	7-33	100/90	3.0/2.8	0.67/0.70	0.65/0.67
1FCL	56	3-55	100/90	3.6/3.5	0.75/0.76	0.70/0.65
1SHG	57	8-11,29-61	100/90	4.3/4.0	0.67/0.72	0.51/0.64
2DA1	70	16-63	100/90	3.2/2.7	0.74/0.79	0.58/0.75

Table 3

Translational and rotational diffusion constants calculated from simulations including or excluding HI (left and right values in the columns, respectively). In the last two columns the "experimental" values estimated by the HYDROPRO software⁷¹ are reported for comparison. In the last row we report the values computed for a LBMD simulation using a friction $\gamma = 0.05 \text{ fs}^{-1}$.

system	vectors	$D_T [\text{\AA}^2 \text{ns}^{-1}]$	$D_R [\text{ns}^{-1}]$	HYDROPRO $D_T [\text{\AA}^2 \text{ns}^{-1}]$	HYDROPRO $D_R [\text{ns}^{-1}]$
ICLB	3-13,27-35	2.8/5.5	0.8/0.6*10 ⁻²	13.1	3.4*10 ⁻²
1EOL	9-12,16-21	4.5/8.4	1.2/1.4*10 ⁻²	15.6	5.6*10 ⁻²
1FCL	3-7,24-36	3.3/5.0	0.4/1.2*10 ⁻²	14.2	4.3*10 ⁻²
1SHG	24-29,44-48	4.5/6.5	0.5/1.0*10 ⁻²	14.0	4.4*10 ⁻²
2DA1	35-43,50-62	2.0/4.1	0.4/1.5*10 ⁻²	11.2	2.0*10 ⁻²
1EOL*	9-12,16-21	-/17.0	-/4.6*10 ⁻²	15.6	5.6*10 ⁻²