Multilevel Lattice Boltzmann-Particle Dynamics simulations at the Physics-Biology interface

Massimo Bernaschi¹, Simone Melchionna², Philippe Derreumaux³, Fabio Sterpone³, Sauro Succi¹,⁴

¹ IAC-CNR, via dei Taurini 19, 00185, Rome, Italy
² ISC-CNR, Dipartimento di Fisica, Università Sapienza, P.le A. Moro, 5, 00185, Rome, Italy
³ LBT, IBPC, CNRS UPR9080, Univ. Paris Diderot, Sorbonne Paris Cité, 13 rue Pierre et Marie Curie, 75005, Paris, France
⁴ IACS-SEAS, Harvard University, Oxford Street 33, 02432, Cambridge, USA

E-mail: succi@iac.cnr.it

Abstract. We discuss current multilevel Lattice Boltzmann-Particle Dynamics (LBPD) simulations of complex states of flowing matter at the interface between physics and biology. We also offer a few considerations on prospective Exascale simulations in the direction of computational physiology.

1. Introduction

The quantitative description of flowing matter traditionally relies upon two basic pillars: continuum fluid mechanics and atomistic/molecular dynamics. The former describes matter as a collection of space-filling (continuum) fields, say density, pressure and flow, whereas the latter explicitly acknowledges the discrete nature of the microscopic constituents of matter, i.e., atoms and molecules. By definition, they deal with very different ranges of space and time scales, which we denote generically as macro and micro, respectively. As computing power and algorithms grow faster and stronger, the distinction between the macro and microlevels becomes increasingly fuzzier, as the corresponding range of operation start to develop a sizeable overlap. More precisely, current upper-end molecular dynamics simulations can simulate micrometric regions over milliseconds [1, 2, 3, 4], which is sufficient to generate a sizeable overlap with continuum fluid dynamics; a cubic micron of water contains about thirty billion molecules, more than enough for the continuum description to apply. Under such conditions, it is only natural to approach the quantitative simulation of flowing matter by multiscale strategies capable of accommodating both micro and microlevels within a unified computational harness.

Such strategy, multiscale modelling for short, experienced a tremendous boost over the last two decades across virtually all walks of science and engineering. Particularly noteworthy, though, is the case of computational biology, which offers a plethora of instances whereby the concurrent interaction of several scales is key to a proper understanding of the phenomena in point. Perhaps the most spectacular case being the molecular mechanics/quantum-mechanics simulation of proteins, which made it to the Nobel recognition in 2013 [5, 6].

In this paper we discuss another instance of multiscale modelling that we label LBPD, for Lattice Boltzmann Particle Dynamics.
The LB part takes care of the continuum sector, typically the flow solvent, while the PD is in charge of molecular motion. The distinctive trait of LBPD is that both micro and macro descriptions are inherently based on mesoscale (coarse-grained) representation of both solvent and molecular components. In other words, LBPD is strongly rooted within the two pillars of kinetic theory: the Boltzmann equation and stochastic particle dynamics, as we are going to briefly detail in the sequel.

1.1. Lattice Boltzmann

The basic object of Boltzmann kinetic theory is the distribution function \( f(\vec{r}, \vec{v}; t) \) measuring the probability density of finding a molecule at position \( \vec{r} \) and time \( t \) with molecular velocity \( \vec{v} \).

This makes a pretty demanding computational object, which obeys a complicated non-linear integro-differential equation in 6 dimensional space, plus time. Therefore, notwithstanding impressive advances, solving the Boltzmann equation remains a very challenging task to this day.

The basic idea of the Lattice Boltzmann method is to constrain the velocity degrees of freedom to a discrete lattice with sufficient symmetry to protect the conservation laws which secure the emergence of standard hydrodynamic behavior in the macroscopic limit.

The classic LB scheme in single-relaxation time (BGK) form reads as follows [7, 8]:

\[
\begin{align*}
  f_i(\vec{r} + \vec{c}_i; t + \Delta t) - f_i(\vec{r}; t) &= -\omega \Delta t (f_i(\vec{r}; t) - f_{eq}^i(\vec{r}; t)) + S_i(\vec{r}; t) \Delta t, \quad i = 0, b
\end{align*}
\]

where \( f_i \) is the discrete Boltzmann distribution associated with the discrete velocity \( \vec{c}_i, i = 0, b \) running over the discrete lattice, to be detailed shortly. In the above, \( \omega \) is a relaxation parameter controlling the fluid viscosity and \( f_{eq}^i \) is the lattice local equilibrium, basically the local Maxwell-Boltzmann distribution truncated to the second order in the Mach number. The truncation is not a luxury, but a necessity dictated by Galilean invariance. The point is that the mass-momentum-momentum flux conservations in discrete velocity space (latin subscripts run over spatial dimensions \( x, y, z \) and \( p \) denotes the fluid pressure),

\[
\sum_i f_i^{eq}\{1, c_{ia}, c_{ia}c_{ib}\} = \{\rho, \rho u_a, \rho u_a u_b + P\delta_{ab}\}
\]

can not be fulfilled for an arbitrary flow field \( u_a \), if the local Maxwell-Boltzmann distribution is retained with the plain replacement \( \vec{v} = \vec{c}_i \).

Finally, \( S_i \) is a generic source of mass/momentum/energy, describing the coupling of the fluid with the surrounding environment. This term is crucial for the extension of LB beyond Navier-Stokes hydrodynamics.

The above constraints, though, can be matched perturbatively by expanding the local Maxwellian to second order in the Mach number, which is sufficient to recover the (isothermal) Navier-Stokes equations, since these equations are quadratic in the fluid velocity. Full Galilean invariance for an arbitrary flow field \( u_a \) implies instead an infinite series in the Mach number, corresponding to the full expansion of the local Maxwellian in Hermite polynomials.

The actual expression of the discrete local equilibria reads as follows:

\[
\begin{align*}
  f_i^{eq} &= w_i \rho (1 + u_i + \frac{1}{2} q_i)
\end{align*}
\]

where \( u_i = c_{ia}u_a / c_s^2 \) and \( q_i = (c_{ia}c_{ib} - c_s^2\delta_{ab})u_a u_b / c_s^4 \) represent the dipole and quadrupole contributions, respectively. Hereafter, repeated indices are summed upon. The lattice equilibria are polynomial truncations of their continuum Maxwell-Boltzmann counterparts, which means that they are non-negative definite only in a finite range of fluid velocities, typically of the
order of $u/c_s \sim 0.3$. This configures the LB method as an appropriate description of quasi-incompressible, low Mach-number flows.

In the above, $w_i$ is a set of lattice-dependent weights, normalized to unity, which represent the lattice analogue of the global Maxwell-Boltzmann distribution. Finally $c_s^2 = \sum_i w_i c_{ia}^2$ is the lattice sound speed.

The corresponding discrete velocities are subsets of the D3Q27 mother lattice with 27 discrete speeds in three spatial dimensions. D2Q9 and D3Q27 are the direct tensor product of the elementary one-dimensional D1Q3 stencil, $c_{ix} = \{-1, 0, +1\}$, in two and three spatial dimensions, respectively (see Fig. 1).

The expansion (3) implies that hydrodynamic LB flows are bound to be quasi-incompressible, i.e., Mach number well below unity. Likewise, third order kinetic moments, describing energy and heat flux, are not correctly reproduced, since these terms require sixth-order isotropic lattice tensors. Such constraints can not be met by lattices confined to the first Brillouin region described by D3Q27: to this purpose, higher order lattices extending beyond the first Brillouin cell are necessary.

Based on (2) and (3), it can be shown that in order for the LB to recover the NS equations in the zero-Knudsen hydrodynamic limit, isotropic tensors only up to order four need to be reproduced exactly in the lattice.

The main computational merits of LB as a fluid solver have been amply documented in a vast literature, hence here we shall simply remind its main highlight: regardless of the spacetime complexity of the emerging macrocell physics, the LB information always move along straight lines, the molecular streamlines defined by the condition

$$\Delta \vec{r}_i - \vec{c}_i \Delta t = 0$$

This is the source of major computational benefits, including an outstanding amenability to parallel computing, a key aspect for leading edge LBPD simulations at the physics/biology interface.

It is also implies limitations towards complex geometries, as typically handled by body-fitted coordinates or finite volumes or elements, an interesting topic which is beyond the scope of the present paper.

1.2. Particle Dynamics

Most biological phenomena depend crucially on the interaction of biological bodies with the surrounding solvent, i.e. water. Biological bodies can range from small molecules, such as organic molecules or peptides, to large ones such as proteins, nucleic acids, up to various type of cells. Such wide spectrum is echoed by a corresponding variety of computational representations, a topic which makes a vigorous branch of modern research on mesoscale methods [9]. For brevity,
here we shall refer to the biological body as to a collection of $p = 1,N$ “particles” following a suitable (stochastic) dynamics. i.e.

$$\frac{d\vec{r}_p}{dt} = \vec{v}_p$$

(4)

$$\frac{d\vec{v}_p}{dt} = -\vec{F}_{pp} + \vec{F}_{fp}$$

(5)

$$\frac{d\vec{F}_{pf}}{dt} = m_p$$

(6)

The key, of course, is in the specific expression of the force term, describing both particle-particle ($pp$) and fluid-particle interactions ($fp$). Consistently with the mesoscale nature of LBPD, both are the result of some form of coarse-graining. The $pp$-part is in charge of reflecting the effect of suitable force fields, while the $fp$-part typically consists of a deterministic term (systematic drag) and a stochastic one (thermal fluctuations), linked via the Fluctuation Dissipation theorem. From the practical standpoint, a key feature of the LBPD procedure is the cross-talking between the LB fluid and the PD dynamics. Indeed, it is clear that the particles react on the fluid through a particle-fluid force $\vec{F}_{pf}$ which is equal and opposite to $\vec{F}_{fp}$.

This is where technical details take the lead; a suitable description of these technical matters goes beyond the scope of this paper. Here, we just mention that the LBPD coupling typically involves G2P (Grid-to-Particle) interpolations to transfer the flow field into the actual particle position, and the conjugate P2G (Particle-to-Grid) interpolation to transfer the particle forces into the grid locations. The choice of the interpolation kernels can range from point-like Dirac deltas to smoother versions based on higher order finite-support shape functions, pretty much in the spirit of smoothed particle hydrodynamics [10]. An accurate and efficient implementation of these “handshaking” interfaces is key to the success of the LBPD scheme.

1.3. Remarks on LBPD

The LBPD procedure falls within the time-honoured tradition of hybrid particle-in-cell methods, in use since long for particle-field simulations [11, 12]. Nevertheless, LBPD presents some specific features, which we briefly discuss in the following. The mesoscopic representation of macromolecules and surrounding solvent allows to account not only for nanometric-scale hydrodynamics, but in principle, also for accurate macromolecular interactions. The key question in this respect is what degree of molecular realism can be regarded as sufficient for the purpose of a realistic biological representation. The answer is problem-specific: representing a lipid chain, a protein, or a DNA chain may require different degrees of chemical specificity, depending on the problem at hand. It is also true, however, that one can employ such coarse-grained potentials in a rather flexible way, as long as the mesoscopic properties, let’s say at the nanometer / nanosecond scale, are correctly reproduced. Such scale, fortunately, is precisely where kinetic modelling, particularly for the liquid state of the biological solutions, and the force fields match in accuracy. At larger scales, say microns and above, such level of detail may become irrelevant, therefore a fair representation of thermodynamics, possibly via an equation of state and an accurate representation of fluid mechanics, fills most needs. The latter marks a major point for LBPD, specially in connection with massively parallel implementations in complex geometries. This blue-sky scenario, however, should cope also with long-range forces, such as the electrostatics ones. Fortunately, under typical cellular conditions electrostatic interactions are strongly screened above nanometric separations.

As discussed above, Boltzmann kinetic theory describes the aqueous solvent as a continuum in a probabilistic sense. On the other hand, the LB mesh spacing $\Delta x$ is defined as a representation (coarse-graining) of the collective kinetic behaviour of a group of solvent molecules. In order to observe hydrodynamic behavior down to the mesh spacing distances, the mean-free-path
generally falls below $\Delta x$. In the liquid state, the molecular mean free path extends over a few angstroms, and a sub-nanometric lattice spacing can be supported in LB, allowing for the hydrodynamic behaviour to emerge at a distance larger than $\Delta x$. Therefore, the lattice grid element must not be viewed as a volumetric entity that contains a fixed number of particles, but instead a numerical support for the probabilistic description of the averaged single particle trajectories. Horbach and Succi [13] have shown that this strategy is effective for the simulation of nano fluids and the obtained results agree very closely with molecular dynamics simulations of simple fluids.

In order to improve physico-chemical realism, with the incorporation of the relevant force-fields, LB models should also be enriched with water-like features, such as directional interactions, so as to make the LB-water solvent more realistic. Preliminary efforts in this direction have been developed in the past, but their thorough validation remains largely open [14].

Many applications naturally suggest themselves; thermal stability of proteins, the onset of neurodegenerative diseases due to peptidic aggregation, diffusion of proteins and trafficking in cellular crowding, being just some examples in point in which a tight coupling between realistic hydrodynamics and chemical specificity is essential.

In passing, we note that LBPD resonates well with the modern focus of in-cell and computational biology. While the past success of the fully atomistic approaches for simulating proteins and DNA was strongly linked to the development of the experimental structural biology [5], and flourished around the central paradigm of the structure/function relationship, the future boost of multi-scale modelling will represent the novel microscope to support system biology where biological function is declined in term of protein/protein multiple interactions, cascade effects and long range transport of information.

1.4. The OPEP coarse-grained model
We pause here to describe shortly the implicit solvent coarse-grained (CG) model for protein OPEP (Optimized Potential for Efficient peptide structure Prediction) [15] that has been integrated in the LBPD scheme [16] and that allows to treat biomolecules at a reasonable chemical accuracy but at a cheaper cost than traditional all-atoms molecular mechanics models. The OPEP force field reserves an atomistic description to the amino-acid backbone including hydrogens, but reduces all side-chain atoms to a single CG bead. The Hamiltonian of the force field is given by bonded and non bonded terms. In the former class we find standard potentials, e.g. harmonic, for bonds, angles and dihedrals. Some extra special terms allow also to map correctly the Ramachandra plots for secondary structures. While the model does not assign explicit charges to the particles, some interactions of electrostatic origins are modelled via ad-hoc potentials. This is the case of the Hydrogen Bond propensity and its cooperative effects via two-body and four-body terms, a critical ingredient to reproduce the sampling of $\alpha$-helices and $\beta$-strands secondary structures. Also ion-pairing between charged amino-acids is modelled with effective potentials of mean force. The non bonded interactions are optimised using standard Lennard-Jones or other short-range type of potentials. As an effect of its resolution and the parametrized potentials, OPEP, at variance with mainstream CG models as Martini, folds peptides and proteins, is a very good tool to sample protein fluctuations including long range rearrangements, and protein aggregation. Because of the atomistic resolution of the backbone, a relatively short time-step as compared to other CG model is needed to correctly integrate the equation of motion, 1.5fs that can be raised to 4fs if backbone mass rescaling is adopted to reduce the high frequency of hydrogen motions. For many purposes the details of intramolecular flexibility is not necessary, OPEP intramolecular interactions can therefore be replaced by an effective elastic network thus making possible to access longer time-scale in the simulation because of the reduced number of degrees of freedom and a larger time-step (10fs).
2. LBPD SIMULATIONS AT THE PHYSICS/BIOLOGY INTERFACE
Several different examples have already shown the potential of LB simulations in multiple areas across Physics and Biology. A selected set of examples will be briefly reviewed in this Section, to provide a taste for the breadth of applications that have been tackled in the recent past. The Physics-Biology interface is enormously rich and varied, ranging from nanometric macromolecular phenomena, to peptidic aggregation and biopolymer translocation, to cellular motion and active matter organization. However, the topic goes way beyond the scope of the present work. Nevertheless, it is important to appreciate the great flexibility of the method to cope with multiple flow conditions and, more importantly, to accommodate the proper degree of chemical specificity required by biological applications.

2.1. Biopolymer Translocation
The translocation of biopolymers, and in particular DNA strands in nanometric pores, provides a showcase of the synergistic hydrodynamic effects assisting or interfering with the translocation process. Such process mimics a genuinely biologic one, whereby viral penetration takes place via the injection of viral DNA into the host cell’s cytoplasm. Computer simulations can provide an exhaustive access to the translocation process, both for technological innovations and for a better understanding of the migration of small biopolymers. At technological level, analysis of nanopores could provide a new path to DNA fast-sequencing, provided the physics of translocation is better understood [19]. The key to nanopore-based sequencing is the ability to translocate DNA chains through a nanoconfined environment, where the genetic information can be decoded by optical mapping, ionic or electronic detection.

Translocating a long DNA chain into extremely narrow pores results in large entropy loss caused by the confinement and the need to stretch the DNA macromolecule creating a free energy barrier, which reduces DNA capture rates and causes clogging at the nanochannel/pore entrance. Most of translocation is modulated by the competing effects taking place in the chambers and in the pore, giving rise to a genuine multiscale scenario [20, 21, 22, 23]. When facing such complex set-up, all-atoms Molecular Dynamics methods, or even coarse-grained representations of the translocating biopolymer, necessarily have to give up the explicit representation of the solvent, thus imposing severe limitations to the simulation accuracy. On the other hand, resorting to other simulation methods, based on a direct solution of the Navier-Stokes equations, or by using other mesoscopic numerical methods (Lagrangian or Eulerian based), face several challenges including the generation of consistent noise under confinement, the stability of the numerical method and the request for time averaging of a process involving a handful of molecules/atoms flowing in a narrow space.

One more level of complication arises from the inclusion of electrokinetics, that is, by describing the multi-component saline solution that flows together with DNA from chamber to chamber. To handle such scenario ionic currents and solvent flow are coupled via the exchange of momentum, that gives rise to electro-osmotic and electrophoretic phenomena. The self-consistent equations can still be casted in the LBPD framework in a natural way [24]. In Figure 2 we provide two examples of polymer translocation studied via LBPD.

2.2. Proteins in cell: motion, conformations and folding
Protein functionality often comes along with important conformational changes that mechanically transport biochemical signals. The extreme case is the folding process that allows a protein to acquire a well defined 3D structure ready to function. Reproducing protein folding during its complete temporal span is a major challenge in computational science, with plenty of implications for medicine and drug design. For isolated small and mid-size proteins placed
Figure 2. Panel a. Translocation of a neutral biopolymer in a narrow pore. Here, the translocating biopolymer is represented as a simple necklace of beads, by neglecting correlations stemming from the local molecular rigidity or backbone charge, but including thermal noise. Depending on the translocation conditions and pore size, the polymer threads as single-file (left-panel) or multi-file (right panel). Panel b. A charged biopolymer, translocating in a narrow pore. This model is representative of single-stranded DNA in which molecular rigidity is neglected. The charged macromolecule moves in an electrolyte solution, whereby a neutral solvent and counterions and coions migrate due to an externally applied electric field, giving rise to electro-osmotic flow that ultimately causes translocation [24].

in a water solution the problem can be considered computationally solved. All-atom Molecular Dynamics simulations performed at the millisecond timescale by the group of DE Shaw [25, 26] or calculations based on the enhanced sampling Replica Exchange strategy [27] demonstrated that molecular mechanics force fields developed along the years can fold successfully a protein in good agreement with experiments. Such efforts have allowed to explore the folding kinetics and paths, and test the reaction diffusion framework earlier introduced in conjunction with the funnel-theory. The matter changes drastically as soon as the focus is placed on a protein in its natural environment, that is the cell. Here, a protein experiences a strong crowded space where excluded volume effect, specific interactions and solvent mediated correlation significantly affect its dynamics and stability. The possibility to follow the "dance" of proteins in cellular conditions is considered the next challenge in molecular and computational biology, and is stimulating advances in spectroscopic single molecule techniques as well as ensemble methods. Computational biology is responding to the call too. To this purpose it is mandatory the use of simplified and cheap molecular force fields yet realistic enough to capture essential internal motions and intermolecular interactions. LB based approach can disclose the potentiality of a combined multi-scale approach including solvent mediated correlation in implicit solvent MD. In fact, in a cellular environment, along with the strong crowding condition, hydrodynamic interactions (HI) are expected to play a critical role by affecting the mobility of proteins, correlating at long range the cyclic conformational changes of molecular motors with the mobility of passive particles, selecting the paths of conformational changes – including folding events. In some special cases, the fluid environment plays a crucial role by triggering the activity of proteins, as in the case of catch-bonds that explicit their functionality under the presence of shear flow in cell adhesion or in blood coagulation. Several groups have started approaching the challenge, either by attempting to include HI in brownian dynamics[28], by using stochastic
rotation dynamics combined with simple coarse-grained model [29] and with the incorporation of the implicit solvent coarse-grained model for protein OEP within the LBPD harness [16, 17]. For instance, recent studies based on LBPD showed that crowding contributes to slow-down the dynamics of a target protein [15, 17] in very good agreement with Neutron Scattering and NMR experiments. Recent efforts were also directed to model catch-bonds proteins as the respond to Couette and Poiseuille flows [18].

2.3. Amyloid Aggregation
A special case where protein dynamics and conformational changes play a common role is the aggregation of misfolded soluble proteins into fibrils, the hallmark of several neurodegenerative diseases, such as Alzheimer, Parkinson, and Huntington. The aggregation of amyloid β proteins is a complex process, whereby amyloid fibrils extend up to hundred of nanometers, and the time scale of full growth exceeds hours in vitro. Understanding the mechanics of amyloid aggregation is crucial to the design of drugs able to prevent fibril formation and toxicity in the brain. Computer modelling is an essential tool to explore the aggregation process. Key thermodynamic insights have been obtained by using oversimplified molecular models [30], mimicking amyloid aggregation at microscopic level is still a challenge [31]. The all-atom approach is prohibitive and a bare implicit solvent model would lack the action of solvent mediated correlations. By deploying the LBPD methodology, it is shown that the effect of HIs is to decisively speed-up amyloid aggregation [32]. As a matter of fact, solvent mediated interactions enhance the peptides mobility easing their encounter and collapse in an aggregated structure. Moreover, as the size of the aggregates increases, the latter acts as an active body, attracting isolated monomers which fuse together according to a coagulation model. This is an important, possibly the first, real-life application where macromolecular realism and solvent-mediated interactions are shown to provide a more complete picture for amyloid aggregation.

2.4. Towards Computational Physiology
Computational physiology deals with the dynamics of the main organs of the human body, such as the cardiovascular or the respiratory systems, which operate at macroscopic scales. LB hemodynamics has known a major burst of activity for the last decade, with several applications to coronary, carotid, and cerebral blood flow. These are macroscopic applications which fit within the continuum hydrodynamic picture. The challenge here, is to reach up to physiological scales (1-10 cm), without surrendering the essential microbiological features, the finite-size of red blood cells (8 micron) in the first place [33, 34, 35, 36, 37, 38] and possibly even their elasto-mechanical properties [39].

This is of major interest for many reasons; the granular nature of blood may have a significant impact on the recirculation patterns in the vicinity of natural geometrical irregularities, such as bifurcations, stenoses, aneurysms, or man-made ones, like stents and other medical devices. It is quite clear that a fully 4D (three-space dimensions plus time) real-time numerical and visual access at the blood dynamics from microns to the full-scale geometry, would disclose unprecedented opportunities for personalized precision-medicine. It is to be stressed here that statistical averaging is of little meaning in a physiological context, since each individual is a story of his own. As a result, the detailed access to the patient-specific 4D hemodynamic data across scales of motion, would offer a forward leap in the quality and accuracy of pre-emptive medicine [40, 41].

A direct extension is to simulate blood flows by accounting for the explicit presence of red blood cells (RBC) suspended in plasma. This is a typical case where the hydrodynamic medium hosts particles with finite-size, anisotropic shape, actually prolate ellipsoids that represent RBC cells to a first approximation. In hemodynamics, there are several situations, where the continuous nature of blood should be replaced by the two-component mixture of plasma
Figure 3. Panel a. Translational diffusion at the nanosecond timescale of the protein RAT I in solution at different concentrations and corresponding to a crowded occupied volume $\phi$ and computed via LBPD (blue circles) as compared to the experimental diffusion slowdown reported for a protein of similar size (BSA) and estimated by Neutron Scattering, details in Ref. [15]. Panel b. Modelling of a single domain of the von Willebrand factor protein. This elementary unit of the multi-domain chain is composed by the globular A1, A2 and A3 proteins linked by polypeditic chains. During blood coagulation, the vWF extends under the action of the shear and A2 partial unfolding contributes to further extend the chain. In the multiscale modelling of the vWF unit, A1 and A3 are represented by an elastic network, A2 is represented by the full flexible OPEP force field, while the linkers are modelled using a simple bead for each amino-acids and ad hoc ultra-coarse grained potentials (bond, angle, torsion and vdw). Panel c. Modelling of the FimH protein, representing the tip of E.coli pilum. This tip links E. coli to a target protein at the infected cell membrane and this binding is enhanced under shear flow. The behaviour under flow of the Fim proteins was recently investigated via LBPD and OPEP force field [18]. Panel d. Massive aggregation of amyloid peptide A$\beta$(16-22) followed in time for a system composed of 1000 peptides. After a first linear growth, the largest cluster increases size discontinuously by the fusion of preformed aggregates [32].

and cells. This is the case, for example, of stenotic vessels as caused by atherosclerosis, where common conditions are such that about 10 RBCs can pass in the lumen narrowing. In this condition, the granular nature of blood shows up, and non-trivial rheological properties emerge as a consequence of the concerted motion of plasma and RBCs. As we zoom into capillaries, with a lateral size of 100 microns and below, the motion of red blood cells reveals highly non-trivial effects and for capillaries with a diameter of a few microns, RBCs deform in order to squeeze in, in such a way that they manage to crawl into the micrometer-sized vessels. On the other hand, when inspecting larger-scale circulation, in the 100 - 500 microns range, it is generally sufficient to consider the cells as rigid bodies.

An all-embracing LBPD hemodynamic tool is still beyond reach, the various prominent codes being able to cover only three decades in space at best. Mesoparticle methods, such as Dissipative Particle Dynamics, can range from, say 0.1 to 1 microns [42], while extreme implementations of the MUPHY code [43] managed to cover about four decades, from centimetres to a few microns. Intermediate options are also available, like the multilevel code Hemocell, which is explicitly targeted to address the mechanical properties of RBCs so as to allow direct coupling between


fluid flow and biological processes such as cellular adhesion \[44\]. While a super-code based on
the modular coupling of the aforementioned approaches is in principle conceivable, it would be
highly labor-intensive in practice. Most likely, each of these tools will keep growing on their
own, by extending their range of operation and facilitating the mutual exchange of increasingly
accurate information.

3. SIMULATIONS AT THE PHYSICS/CHEMISTRY/BIOLOGY INTERFACE:
DREAMING AHEAD
Having briefly discussed the current LBPD applications, as a conclusive remark we next venture
into some visionary, yet hopefully realistic, considerations for the future, mostly in connection
with foreseeable progress in computer technology. From the conceptual viewpoint, these efforts
fully inscribe within the very elegant multiscale integrative approach to Biology advocated in
Denis Noble's books \[45, 46\].

The LB method is often quoted for its amenability to parallel processing. The performance
of LB codes is best conveyed in terms of MLUPS, namely Million Lattice Updates Per Second; 1
MLUPS means updating one million sites by one step in time in one CPU second. Given that a
LB code for Navier-Stokes hydrodynamics requires about 200 Flops per site and step, 1 MLUPS corresponds to approximately 200 Mflops/s. On current high-end laptops, a LB code features
of the order of 5 MLUPS. On more complex applications, especially with LBPD interfaces, this
figure lowers to about 1 MLUPS or below, due to the increased computational density, typically
in the order of thousands Flops/site/step. At a 1 MLUPS rate, a 1Gflops/s computer updates
one million sites/second, meaning that an LBPD simulation spanning, say, four decades in time
(ten thousands time steps) would take just three hours.

We may realistically assume that the LBPD computational complexity scales linearly with
the volume of the simulated four-dimensional spacetime, namely:

\[
C \sim kV = kL^3T = kL^{3+\alpha}
\]

where \( k \sim 10^3 \) is the computational density in Flops/site/step and \( \alpha = 1 \) for the ballistic regime
and \( \alpha = 2 \) for the diffusive one. Even within the more demanding diffusive regime, a prospective
Exaflop computer would ideally permit to span over four decades in space and over eight in
time, \textit{in matter of a day}!

With these advances in mind, several exciting scenarios open-up in front of the LBPD
framework, such as the direct simulation of organelles, or highly complex physiological processes
like haemostasis, to be detailed shortly. Before such dreams come true, though, many technical
and conceptual issues have to be successfully addressed. To begin with, upsizing to such sizes
requires \textit{qualitatively} new parallel programming paradigms. In particular, as size increases, the
cost of accessing memory kicks heavily in; for instance, the streaming step, which entails literally
zero floating point operations, can currently account for a significant fraction (about 1/3) of the
execution time on current leading edge LB simulations. Consequently, much care must be
exercised to minimise the overheads of memory access, for instance by hierarchical organization
of the data arrays in energy shells, i.e. discrete velocities with the same magnitude \[47, 48\].

The efficient parallelisation of LB codes requires additional skilful techniques, aimed at i)
Overlapping communication and computation, ii) Distributing the workload evenly across the
various processors, iii) Accessing data efficiently. The above items are only accrued in the case
of multiscale LBPD, due to conflicting needs raised by grid-based data structures (LB) and grid-
free particles (PD). Such techniques have indeed led to impressive performances; for instance,
the MUPHY code has delivered up to 20 Pflops, with 90 percent parallel efficiency for the case
of protein dynamics in crowded solutions, see Figure 4.a.

With Exascale machines in mind, another major issue to be considered is \textit{fault-tolerance},
i.e. how to equip LBPD with feedback mechanisms capable of recovering from failures, as they
Figure 4. Panel a. The roadmap of the MUPHY code based on Gordon-Bell performance. Panel b. Multiscale representation of vesicles transporting proteins. Vesicle firing from a membrane where the vesicles are modelled as immiscible fluid (green) phase separating from the aqueous host (light blue background), the central panel zooms the vesicle bilayer with liquid water in the interior, the right panel depicts a molecular representation of the vesicle containers with an heterogenous protein suspension in the interior and membrane proteins embedded in the bilayer.

inevitably occur once millions and possibly billions of concurrent processes are in action. In this context, the new challenging applications would focus for instance on fundamental biological processes where the complex states of flowing matter across many scales of motion is essential, this includes for example the simulation of molecular transport in the cellular Golgi apparatus, the multi-scale investigation of haemostasis, neuronal firing and ions flow in neurones. Meeting the above challenges is all but a trivial task, but the pay-off might be immense.

4. Acknowledgements
The research leading to these results has received funding from the European Research Council under the Horizon 2020 Programme Grant Agreement n. 739964 ("COPMAT"). FS acknowledges funding from the ERC (FP7/2007-2013) Grant Agreement no.258748 and ANR-11-LABX-0011-01 (Dynamo).

References
[1] Salomon-Ferrer R, Götz A W, Poole D, Le Grand S and Walker R C 2013 J. Chem. Theory Comput. 9 3878–3888
[2] Phillips J C, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R D, Kale L and Schulten K 2005 J. Comput. Chem. 26 1781–1802
[3] Abraham M J, Murtola T, Schulz R, Páll S, Smith J C, Hess B and Lindahl E 2015 SoftwareX 1 19–25
[4] Lindorff-Larsen K, Piana S, Dror R and Shaw D 2011 Science 334 517–520
[5] Levitt M 2001 Nature Struct. Bio. 8 392–393
[6] Karplus M 2014 Angew. Chem. Int. Ed. Engl. 53 9992
[7] Higuera F, Succi S and Benzi R 1989 Europhys. Lett. 9 345
[8] Qian Y, d’Humieres D and Lallemand P 1992 Europhys. Lett. 17 479
[9] Dünweg B and Ladd A J 2009 Advanced Computer Simulation Approaches for Soft Matter Sciences III (Springer) pp 89–166
[10] Monaghan J J 1992 Ann. Rev. Astron. and Astrophys. 30 543–574
[11] Eastwood J W 1991 Comp. Phys. Comm. 64 252–266
[12] Hockney R W and Eastwood J W 1988 Computer simulation using particles (CRC Press)
[13] Horbach J and Succi S 2006 Physical review letters 96 224503
[14] Succi S, Moradi N, Greiner A and Melchionna S 2014 Front. Phys. 2 22
[15] Sterpone F, Melchionna S, Tuffery P, Pasquali S, Mousseau N, Cragnolini T, Chebaro Y, St-Pierre J F, Kalimeri M, Barducci A, Laurin Y, Tek A, Baaden M, Nguyen P H and Derreumaux P 2014 Chem. Soc. Rev. 43 4871–4893
[16] Sterpone F, Derreumaux P and Melchionna S 2015 J. Chem. Theory Comput. 11 1843–1853
[17] Chricot F, Sterpone F, Derreumaux P and Melchionna S 2016 Philos. Trans. A 374 20160225
[18] Sterpone F, Melchionna S, Doutreligne S, Tran T, Nguyen P, Baaden M and Derreumaux P 2017 Bioch. Biophys. Res. Comm. 17 31739
[19] Fyta M, Melchionna S and Succi S 2011 J. Polym. Sci. B 49 985–1011
[20] Fyta M G, Melchionna S, Kaxiras E and Succi S 2006 Multiscale Mod. Sim. 5 1156–1173
[21] Fyta M, Melchionna S, Succi S and Kaxiras E 2008 Phys. Rev. E 78 036704
[22] Bernaschi M, Melchionna S, Succi S, Fyta M and Kaxiras E 2008 Nano lett. 8 1115–1119
[23] Fyta M, Melchionna S, Kaxiras E and Succi S 2008 Computing in Science & Engineering 10
[24] Datar A V, Fyta M, Marconi U M B and Melchionna S 2017 Science 350 341–346
[25] Lindorff-Larsen K, Trbovic N, Maragakis P, Succi S and Shaw D E 2012 J. Am. Chem. Soc. 134 3787–3791
[26] Paschek J, Dempela S and Garcia A E 2008 Proc. Natl. Acad. Sci. USA 105 17754–17759
[27] Ando T, Chow E and Skolnick J 2013 J. Chem. Phys. 139 121922
[28] Schoffield J, Inder P and Kapral R 2012 J. Chem. Phys. 136 205101
[29] Sarić A, Chebaro Y C, Knowles T P J and Frenkel D 2014 Proc. Natl. Acad. Sci. USA 111 17869–17874
[30] Nasica-Labouze J, Nguyen P H, Sterpone F, Berthoumieu O, Buchete N V, Coté S, Simone A D, Doig A J, Faller P, Garcia A, Laio A, Li M, Melchionna S, Mousseau N, Mu Y, Paravastu A, Pasquali S, Rosenman D J, Strodel B, Tarsus B, Viles J H, Wang C and Derreumaux P 2015 Langmuir
[31] Pontrelli G, Halliday I, Melchionna S, Spencer T J and Succi S 2014 Mathem. Comp. Mod. Dyn. Syst. 20 470–490
[32] Toschi F and Succi S 2005 Europhys. Lett. 69 549
[33] Melchionna S 2011 Macrom. Theory Sim. 20 548–561
[34] Artoli A M, Hoekstra A G and Sloot P M 2006 J. Biomech. 39 873–884
[35] Noble D 2008 The music of life: biology beyond genes (Oxford University Press)
[36] Noble D 2016 Dance to the Tune of Life: Biological Relativity (Cambridge University Press)
[37] Shet A G, Siddharth K, Sorathiya S H, Deshpande A M, Sherlekar S D, Kaul B and Ansumali S 2013 Inte. J. Mod. Phys. C 24 1340011
[38] Shet A G, Siddharth K, Sorathiya S H, Deshpande A M, Kaul B, Sherlekar S D and Ansumali S 2013 Phys. Rev. E 88 013314