Journal of Computational Science 17 (2016) 446–456

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Two-phase flow analogy as an effective boundary condition for modelling liquids at atomistic resolution

Ivan Korotkin[a,e], Dmitry Nerukh[b], Elvira Tarasova[c], Vladimir Farafonov[d], Sergey Karabasov[e,*]

[a] Nuclear Safety Institute of Russian Academy of Sciences, Bolshaya Tulskaya 52, Moscow 115191, Russian Federation
[b] Systems Analytics Research Institute, Aston University, Birmingham B4 7ET, UK
[c] Immanuel Kant Baltic Federal University, A. Nevskogo str. 14, Kaliningrad 236041, Russian Federation
[d] Department of Physical Chemistry, V.N. Karazin Kharkiv National University, Svobody Square 4, Kharkiv 61022, Ukraine
[e] School of Engineering and Materials Science, Queen Mary University of London, Mile End, London E1 4NS, UK

ARTICLE INFO

Article history:
Received 29 November 2015
Received in revised form 17 March 2016
Accepted 21 March 2016
Available online 24 March 2016

Keywords:
Hybrid atomistic/continuum
Multiscale
Molecular dynamics
Fluctuating hydrodynamics
Dialanine
All-atom whole virus simulation
Capsid
PCV2

ABSTRACT

A hybrid Molecular Dynamics/Fluctuating Hydrodynamics framework based on the analogy with two-phase hydrodynamics has been extended to dynamically tracking the feature of interest at all-atom resolution. In the model, the hydrodynamics description is used as an effective boundary condition to close the molecular dynamics solution without resorting to standard periodic boundary conditions. The approach is implemented in a popular Molecular Dynamics package GROMACS and results for two biomolecular systems are reported. A small peptide dialanine and a complete capsid of a virus porcine circovirus 2 in water are considered and shown to reproduce the structural and dynamic properties compared to those obtained in theory, purely atomistic simulations, and experiment.

1. Introduction

Recent advances in hydrodynamics theory make it applicable to problems of space and time scales approaching molecular level. For example, it has been demonstrated [1,2] that the statistics of the thermal fluctuations of hydrodynamic fields can be quantitatively reproduced at the scale when the elementary volume contains only a few dozens of atoms. Such results motivated a wide spectrum of research where classical molecular dynamics is consistently connected to continuum hydrodynamics. These methods are particularly promising from the technical point of view since they have a potential of significant savings in computing resources without substantial loss of accuracy in critical regions. Biomolecular systems benefit especially strongly here because biomolecules necessarily have critical regions where atomistic representation is needed, while the system overall usually functions in close connection to a large medium of water and other molecules. The latter is critically important, but impossible to model at atomistic level because of the sheer size of the system.

In most approaches connecting particle and continuum dynamics either local hydrodynamic properties are computed with the help of molecular dynamics [3,4]; or hydrodynamic fields serve as boundary conditions to molecular dynamics part of the system [5]. This 'mechanistic' partitioning of the system often creates problems and even artefacts in cases where there is no clear scale separation between different parts of the multiscale system in space and time. A straightforward example is the problem of particles leaving and entering the molecular dynamics zone. Several computational tricks are suggested, for example periodic boundary conditions in the molecular dynamics zone with prescribing a required meanflow gradient, however, more consistent, physically justified approaches are needed in the case of a strong coupling between the atomistic and the fluid dynamics regions.

We have recently suggested and implemented a new approach to the problem of computing liquids at multiple scales in space and time based on the two-phase hydrodynamics analogy [6]. Instead of separating the system into parts using boundaries, we describe the liquid as a nominally two-phase system consisting of both particle and continuum at the same time. The contribution of each prop-

* Corresponding author.
E-mail address: s.karabasov@qmul.ac.uk (S. Karabasov).

http://dx.doi.org/10.1016/j.jocs.2016.03.012
1877-7503/

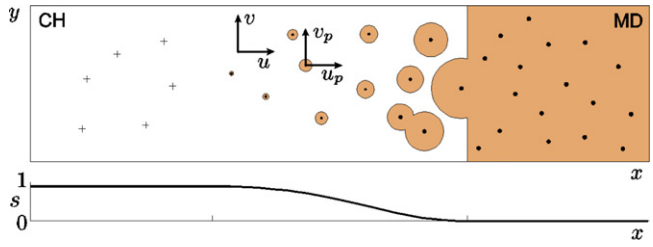


Fig. 1. A schematic representation of the hybrid simulation framework; because of the shape of the s function (bottom) purely MD particles on the right are gradually transformed into passive traces that follow the hydrodynamics flows on the left.

erty (particles or continuum), for example their 'partial volume', can vary depending on location and time, such as in the limiting cases only one of the descriptions remains. The governing equations are formulated as a system of the conservation laws for mass and momentum of the particle/continuum 'mixture'. This allows to solve several problems, for example the above mentioned problem of leaving/entering particles that is solved naturally as the particles are allowed to move freely in all parts of the system while having impact on the solution only in the regions where their 'partial volume' is nonzero, which happens without any violation of the governing conservation laws.

In Ref. [7] a one-way coupling implementation of the same hybrid framework, which accounts for the effect of molecular dynamics on continuum hydrodynamics without the feedback, was considered and tested as an alternative open boundary treatment in the popular Molecular Dynamics (MD) package GROMACS [8]. In the current paper, the former implementation in GROMACS has been extended to dynamically tracking the atomistic-resolution features of interest. To illustrate the accuracy and efficiency of the new extended method, two simulations of two biomolecular systems from the opposite ends of the spectrum of possibilities are considered: diffusion of a small peptide, dialanine, molecule in water for long simulation times and the interaction of the capsid of a complete virus, porcine circovirus 2, with water at equilibrium conditions.

2. The underlying two-way coupling approach and its reduction to a one-way coupling model

Two representations of liquid are combined in the same framework: 1) standard classical MD description where the matter is the collection of point masses interacting with each other through an empirically defined potential and 2) Landau-Lifshitz Fluctuating Hydrodynamics (LL-FH) continuum, that is the generalisation of Navier-Stokes hydrodynamics with stochastic sources added to statistically mimic thermal fluctuations at small scales. A nominally two-phase liquid model is considered as a representation of the same chemical substance (Fig. 1). The 'phases' are immersed into each other as 'fine grains', the surface tension effects are irrelevant, and both 'phases' simultaneously occupy the same control volume. The partial concentrations of the MD 'phase' and the LL-FH 'phase' are equal to s and $1-s$, respectively, where s is a parameter of the model $0 \le s \le 1$. In general s is a user-defined function of space and time which controls how much atomistic information is required in a particular region of the simulation domain, Fig. 1.

The most general form of the governing equations which are suitable for the continuous and discontinuous (particle) part of the hybrid model is integral conservation laws. The integral conservation laws are formulated through a state variable coupling method so that the mass of the mixture containing the two 'phases' is strictly preserved as well as satisfying the second law of Newton so that the change of total momentum of the system is equal to the sum of all forces. Notably, an alternative approach would be to formulate the same model in the framework of generalised functions where discrete particles are represented as Heaviside functions (e.g. by representing atoms as particles of a finite size corresponding to the characteristic width of the interaction potential well), their derivatives become Dirac delta functions, and so on, the model derivation is similar to obtaining the Navier-Stokes solutions with discontinuous variables [9].

Following Ref. [7], let us consider a solution domain of volume V_0 which is broken down into elementary Eulerian cubical cells of volume V. Each cell has 6 faces $\gamma = 1, .., 6$ and it is filled with the continuum part of the liquid and, at the same time, with the MD particles which correspond to the discrete representation of the same chemical substance. It is assumed that the continuum part of the nominally two-phase fluid has the same transport velocity as that of the mixture. At isothermal condition this nominally two-phase liquid in addition to the macroscopic equation of state satisfies the following macroscopic conservation laws. For mass:

$$\delta_t(sm) + \sum_{\gamma=1,6} (s\rho\mathbf{u})\, d\boldsymbol{n}^\gamma \delta t = \delta_t J^{(\rho)}, \quad \text{for the LL} - \text{FH phase}, \tag{1}$$

$$\delta_t \left((1-s) \sum_{p=1,N(t)} m_p \right) + \sum_{\gamma=1,6} \left((1-s) \sum_{p=1,N_\gamma(t)} \rho_p \boldsymbol{u}_p \right) d\boldsymbol{n}^\gamma \delta t = -\delta_t J^{(\rho)}, \text{ for the MD phase}, \tag{2}$$

where m and $\rho = m/V$ are the mass and the density of the continuum 'phase' of the elementary volume V, m_p is the particle mass, $\boldsymbol{u}_p$ is the MD velocity, $\mathbf{u}$ is the average velocity of the 'mixture' $\tilde{u}_i = \left[s\rho u_i + (1-s) \sum_{p=1,N(t)} \rho_p u_{ip} \right] / \tilde{\rho}$, u_i is the velocity of the continuum LL-FH 'phase', $\tilde{\rho} = s\rho + (1-s) \sum_{p=1,N(t)} \rho_p$, $N(t)$ is the number of particles in the volume V. $N_\gamma(t)$ is the number of particles crossing the γ^{th} cell face with the normal $d\boldsymbol{n}^\gamma$, $\rho_p = m_p/V$ is the effective density of an MD particle p which occupies the volume V, and $\delta_t J^{(\rho)}$ is the mass source/sink term which describes the transformation of mass between the 'phases', δ_t describes the change of a quantity over time δt, e.g. the counters of particle mass and momentum in cell V accumulated over time δt.

For momentum this is:

$$\delta_t(smu_i) + \sum_{\gamma=1,6} (s\rho u_i \mathbf{u})\, d\boldsymbol{n}^\gamma \delta t = s \sum_{j=1,3} \sum_{\gamma=1,6} \left(\Pi_{ij} + \tilde{\Pi}_{ij} \right) dn_j^\gamma \delta t + \delta_t J_i^{(\boldsymbol{u})}, \tag{3}$$

$$\delta_t \left((1-s) \sum_{p=1,N(t)} m_p u_{ip} \right) + \sum_{\lambda=1,6} \left((1-s) \sum_{p=1,N_\gamma(t)} \rho_p u_{ip} \boldsymbol{u}_p \right) d\boldsymbol{n}^\gamma \delta t = (1-s) \sum_{p=1,N(t)} F_{ip} \delta t - \delta_t J_i^{(\boldsymbol{u})}, \tag{4}$$

where $\boldsymbol{\Pi}$ and $\tilde{\Pi}$ are the deterministic and stochastic parts of the Reynolds stress tensor in the LL-FH model, F_{ip} is the MD force exerted on particle p due to the pair potential interactions, and $\delta_t J_i^{(\boldsymbol{u})}$ is the LL-FH/MD exchange term corresponding to the i^{th} momentum component.

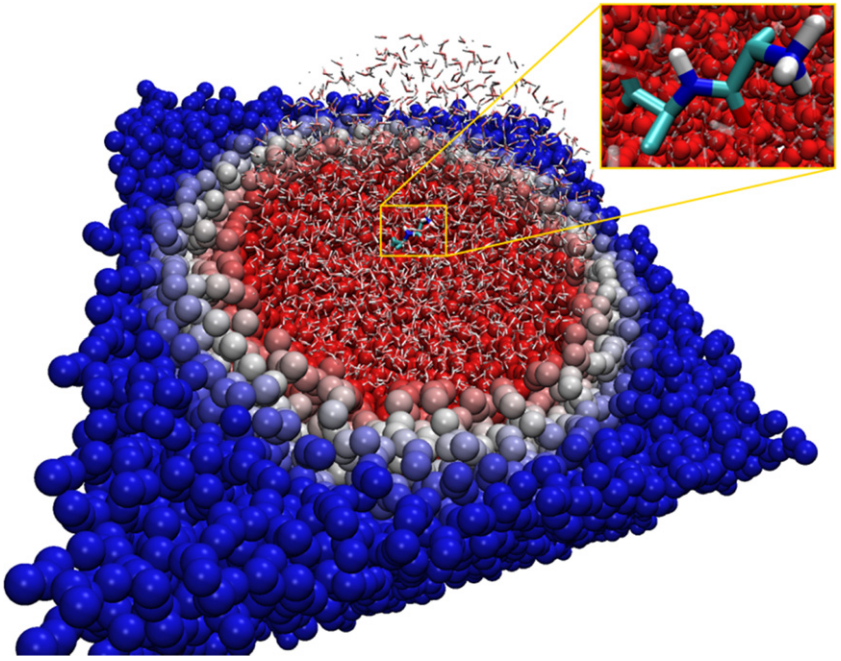


Fig. 2. Dialanine in hybrid MD/LL-FH water; the color of tracers gradually changes from blue (LL-FH region) through white (mixed MD/LL-FH region) to red (pure MD region). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The sums of fluxes $\sum_{\gamma=1,6}\left((1-s)\sum_{p=1,N_\gamma(t)}\rho_p\boldsymbol{u}_p\right)d\boldsymbol{n}^\gamma\delta t$ and $\sum_{\lambda=1,6}\left((1-s)\sum_{p=1,N_\gamma(t)}\rho_p u_{ip}\boldsymbol{u}_p\right)d\boldsymbol{n}^\gamma\delta t$ are the corresponding counters of particle mass and momentum crossing the cell's boundaries $\gamma = 1, .., 6$.

The flux terms can be calculated from the particle distributions at each point of the cell boundary either directly or approximately. In Refs. [6,7], for computing the cell-boundary values an interpolation method is used that is based on the particle distributions specified at the centres of adjacent volumes V in a finite-volume framework.

By summing up the mass Eqs. (1) and (2), and assuming the conservation fluxes vanish at the domain boundaries, it follows from the divergence theorem that the mass conservation law for the mixture is exactly satisfied, $\bar{m}(t+\delta t)=\bar{m}(t)$, $\bar{m}=\bar{\rho}V$. In a similar way, by combining the momentum Eqs. (3) and (4), it can be seen that Newton's second law, which equates the change of the total momentum $\bar{m}\times\mathbf{u}$ to the force applied to each control volume, $\bar{F}_i = s\sum_{j=1,3}\sum_{\alpha=1,6}\left(\Pi_{ij}+\tilde{\Pi}_{ij}\right)dn_j^\alpha\delta t+(1-s)\sum_{p=1,N(t)}F_{ip}$, is satisfied. Note that the forces include both conservative and non-conservative forces because of the continuum fluid dynamics involved in the hybrid coupling scheme.

In Eqs. (1)–(4), $\partial_t J^{(\rho)}$ and $\partial_t J_i^{(\boldsymbol{u})}$ are the user defined functions which need to be specified to close the model. Rather than specifying them explicitly, for example as constants, these functions are obtained from specifying how fast the mixture averaged values $\bar{\rho}$ and $\bar{u}_i\bar{\rho}$ should equilibrate to the cell averaged parameters from the MD 'phase' of the simulation, $\sum_{p=1,N(t)}m_p$ and $\sum_{p=1,N(t)}u_{ip}m_p$. This is achieved by setting the corresponding forcing terms in the mass and momentum equations for mixture averages so that the differences from the target cell-averaged particle values are gradually diffused. This is similar to the sponge zone boundary conditions for open domain problems in unsteady computational fluid dynamics (CFD). The gradual elimination of discrepancy between the cell-averaged properties of the continuum and atomistic parts of the solution is essential for excluding abrupt changes which could lead to artificial 'phase separation', i.e. to smoothly enforce the same density and momentum on both 'phases' which represent the same chemical substance.

Further details of the full two-way coupling model are available in Refs. [6,7] and the current work focuses on a simpler, one-way coupling model where the equation for mixture averages are computed separately from the particles. Indeed, for macroscopically stationary liquids in the absence of major hydrodynamic gradients and away from solid boundaries, thermal fluctuations are the only source of macroscopic fluctuations in liquids described by the Landau-Lifshitz Fluctuating Hydrodynamics (LL-FH) model.

Under this approximation, the 'two-phase mixture' mass and momentum are completely decoupled from the MD 'phase', and the corresponding conservation variables, $\bar{\rho}$ and $\bar{u}_i$, are treated as a solution of the LL-FH equations:

$$\begin{aligned}&\frac{\partial\bar{\rho}}{\partial t}+div(\bar{\rho}\times\mathbf{u})=0,\\&\frac{\partial(\bar{\rho}\times\bar{u}_i)}{\partial t}+div(\bar{\rho}\times\bar{u}_i\times\mathbf{u})=\sum_{j=1,3}\nabla_j\left(\Pi_{ij}+\tilde{\Pi}_{ij}\right),\quad i=1,2,3\end{aligned}\tag{5}$$

where Equation Of State (EOS), $\bar{p}=\bar{p}(\bar{\rho})$, and the shear and bulk viscosity coefficients, η and ζ, of the Reynolds stress $\boldsymbol{\Pi}$ and its fluctuating component $\tilde{\Pi}$,

$$\begin{aligned}&\Pi_{i,j}=-(\bar{p}-\zeta div\,\mathbf{u})\delta_{i,j}+\eta\left(\partial_i\bar{u}_j+\partial_i\bar{u}_j-2D^{-1}div\;\mathbf{u}\delta_{i,j}\right),\\&\tilde{\Pi}_{i,j}=\zeta div\;\tilde{\mathbf{u}}\delta_{i,j}+\eta\left(\partial_i\tilde{u}_j+\partial_i\tilde{u}_j-2D^{-1}div\;\tilde{\mathbf{u}}\delta_{i,j}\right),\quad i,j=1,2,3\end{aligned}\tag{6}$$

are defined in accordance with the MD model, and where the stochastic stress tensor $\tilde{\Pi}$ is described as a random Gaussian matrix with zero mean and covariance, given by the formula

$$\begin{aligned}&\langle\tilde{\Pi}_{i,j}(\boldsymbol{r}_1,t_1)\tilde{\Pi}_{k,l}(\boldsymbol{r}_2,t_2)\rangle=2k_BT\left[\eta\left(\delta_{i,k}\delta_{j,l}+\delta_{i,l}\delta_{j,k}\right)\right.\\&\left.+\left(\zeta-2D^{-1}\eta\right)\delta_{i,j}\delta_{k,l}\right]\delta(t_1-t_2)\delta(\boldsymbol{r}_1-\boldsymbol{r}_2).\end{aligned}\tag{7}$$

The LL-FH equations are solved with a central finite-volume method [10] to provide the 'target' cell-centre hydrodynamic variables for the modified MD particle velocity and acceleration, $\frac{d\boldsymbol{x}_p}{dt}$ and $\frac{du_{ip}}{dt}$, equations:

$$\frac{d\mathbf{x}_p}{dt} = \boldsymbol{u}_p + s(\mathbf{u} - \boldsymbol{u}_p) + s(1-s) \times \alpha \times \frac{\sum_{\gamma=1,6} \left(\bar{\rho} - \sum_{q=1,N(t)} \rho_q \right) d\boldsymbol{n}^{\gamma}}{\sum_{q=1,N(t)} m_q},$$

$$\frac{du_{ip}}{dt} = (1-s)F_{ip}/m_{ip} +$$

$$+ \sum_{k=1,3} \sum_{\gamma=1,6} \left(s(1-s) \times \alpha \times \sum_{q=1,N(t)} \rho_q u_{iq} \times \left(\frac{\sum_{\lambda=1,6} \left(\bar{\rho} - \sum_{q=1,N(t)} \rho_q \right) dn_k^{\lambda}}{\sum_{q=1,N(t)} m_q} \right)_{\gamma} \right) dn_k^{\gamma} / \sum_{q=1,N(t)} m_q$$

$$+ \sum_{k=1,3} \sum_{\gamma=1,6} \left(s(1-s) \times \beta \times \frac{1}{V} \left(\sum_{\lambda=1,6} \left(\bar{\rho} \times \bar{\boldsymbol{u}}_i - \sum_{q=1,N(t)} \rho_q u_{iq} \right) dn_k^{\lambda} \right)_{\gamma} \right) dn_k^{\gamma} / \sum_{q=1,N(t)} m_q, \; i = 1, 3, \tag{8}$$

where $\alpha, \beta > 0$ are numerical adjustable parameters, which characterize the 'diffusion rate' of the difference between the two 'phases'. The latter parameters define how fast the two 'phases' equilibrate to the same macroscopic condition, i.e. converge to the same liquid they represent. The characteristic relaxation time associated with these parameters

$$\tau^{diff} \sim \Delta x^2/\alpha \sim \Delta x^2/\beta, \tag{9}$$

where $\Delta x = \sqrt[3]{V}$ is the length scale associated with the control volume V, should be comparable to the time step of the particles $\tau^{diff} \sim \tau^{MD}$ so that the relaxation process affects the particle trajectories over their characteristic molecular dynamics time scale.

Detailed derivation of the above equations can be found in Refs. [6,7] but the following should be highlighted.

1. The cell-average fields $\bar{\rho}$, $\mathbf{u}$, $\sum_{q=1,N(t)} \rho_q$ and $\sum_{p=1,N(t)} \rho_q \boldsymbol{u}_q$, which appear on the right-hand-side of the above equations, correspond to the centres of the cell-volumes and need to be evaluated for each location $\boldsymbol{x}$ of MD particle p, for example, using a bilinear interpolation in space for the particle location between the cell centre points, and an interpolation in time in case the hydrodynamics solution is advanced in time with a bigger time step in comparison with the MD solution.
2. The summations on the right-hand side correspond to the volume average field gradients (first derivatives and second derivatives in $\frac{d\boldsymbol{x}_p}{dt}$ and $\frac{du_{ip}}{dt}$ equations, respectively), e.g. using the divergence theorem to eliminate one derivative and assuming the implementation within a finite-volume framework.

3. Implementation

In Ref. [7], the one-way coupling model (5–8) was implemented for a spatially variable resolution with the following 3D shape function $s = s(x, y, z)$ that was centred at the origin of the solution domain:

$$s(x,y,z) = \begin{cases} S_{\min}, & r \le R_{MD}; \\ \dfrac{r - R_{MD}}{R_{FH} - R_{MD}} (S_{\max} - S_{\min}) + S_{\min}, & R_{MD} < r < R_{FH}; \\ S_{\max}, & r \ge R_{FH}. \end{cases} \tag{10}$$

where $r = \left(x - L/2\right)^2 + \left(y - L/2\right)^2 + \left(z - L/2\right)^2$, L is the computation box size, $x, y, z \in [0, L]$, $S_{\min} = 0$, $S_{\max} = 1$ centred at the origin of the molecular dynamics zone ($r = 0$).

The former implementation was based on the following algorithm:

- Define s-function, which specifies the (stationary) space decomposition between the MD and the hydrodynamics driven parts of the computational domain,
- For each particle location, find the corresponding control volume it belongs to, interpolate the hydrodynamic fields obtained from solving Eqs. (5)–(7), compute s-function value at the particle location, and integrate Eq. (8) over one time step,
- Update positions of all particles and repeat.

However, the assumption about the stationarity of s-function is not necessary. Indeed, the governing 'two-phase like' Eqs. (1)–(4) have been formulated with the space-time variable s-function in mind, which plays the role of a phase concentration in the two-phase hydrodynamics analogy model. The derivation of Eq. (8) neither requires the stationarity of s-function. Moreover, in principle, the s-function may not even be singly connected: one can specify and dynamically track more than one MD zone of interest.

Hence, the following modification of the original algorithm is suggested to adaptively track the atomistic region of interest:

- Specify the feature of interest, e.g. the centre of mass of a complex molecule that needs to be treated at all-atom resolution.
- Lock the centre of s-function to this feature so that the coordinates of the origin are updated at each time step.
- For each particle location corresponding to a certain time moment, find the control volume it belongs to, interpolate the hydrodynamic fields from Eqs. (5)–(7) to its location, calculate the space-time variable s-function at that location, and integrate Eq. (8) over one time step, then update positions of all particles, move the origin of S-function in accordance with the kinematics of the feature of interest, and repeat.

The current one-way coupling model with locking the origin of s-function to the centre of the MD region of interest has been implemented in the open-source package GROMACS. GROMACS is a versatile C/C++ code allowing one to conduct high performance molecular dynamics simulations, that is to solve the Newtonian equations of motion for systems of up to several millions of particles. Our coupled MD/LL-FH model implementation consists of two general parts: the MD solver (provided by GROMACS) and the LL-FH solver (standalone C function). The Fluctuating Hydrodynamics equations are solved on a Eulerian uniform orthogonal grid using a conservative low-dissipative finite-difference scheme developed in Ref. [10]. The MD solver is based on the velocity Verlet scheme that allows one to easily include additional forcing terms in the equations of motion for the coordinate and velocity of each particle at each time step in accordance with Eq. (8). The additional forcing terms correspond to the fields obtained from the LL-FH solution defined on the Eulerian grid with spacing defined in accordance with the control volume size. In order to obtain the LL-FH field values for particle locations between the grid points a linear interpolation is used. The implementation supports GROMACS intrinsic parallelism meaning that the simulation with the hybrid MD/LL-FH model can be performed using CUDA, MPI or GROMACS Thread MPI communication protocols.

3.1. Dialanine simulations

Dialanine (L-alanyl-L-alanine) is the smallest protein consisting of only two aminoacid residues. It is a workhorse for biomolecular research because of its simplicity and ease of analysis. In its zwitterionic form the opposite ends of the peptide are charged that makes them strongly interact with the surrounding water. The interaction is realised through hydrogen bonding to water molecules, therefore, an explicit atom representation of water in the vicinity of the peptide molecule is critical. At the same time, many water layers around the peptide molecule are involved in this interaction, therefore, a correct coupling to the bulk of the liquid is necessary. The dynamics of the biomolecule cannot be modelled correctly in isolation from the relatively large volume of water.

The peptide diffusion in water was simulated in accordance with the hybrid model following Ref. [7]. The GROMOS96 53a6 force field [11] was used for the peptide and the SPC/E model was used for water. The conditions for molecular dynamics calculation were the following: $T = 298$ K (maintained by the Nose-Hoover thermostat with time constant 0.5 ps), $\rho = 993.3$ g/cm^3, the time step was 1 fs, the cut-off length of reaction field electrostatics was 1.0 nm, the dielectric constant was 78, and the van der Waals cut-off length was 1.0 nm. The entire computational domain including the hydrodynamic region corresponded to a cube with the edge length of 7.2 nm (~12 thousands water molecules) with periodic boundary conditions for solving Eqs. (5)–(8). The location of the pure MD and LL-FH zones are shown in Fig. 2.

In order to investigate the influence of the hydrodynamic region ($s > 0$) on the pure atomistic region ($s = 0$), the translational diffusion coefficient D was calculated for the peptide molecule. The result was compared with the reference solution obtained from the all-atom simulation that conducted in the same domain with the standard periodic boundary conditions. For calculating D, the Einstein relation was used $MSD(t) = \langle \Delta r(t)^2 \rangle = A + 6Dt$, where A is an arbitrary constant and MSD is mean-square deviation. The formula is valid in the limit of large statistical sampling that corresponds to long times t.

In our previous work [7] the method with a stationary s-function considered with the origin fixed to the centre of computational domain, which corresponded to the initial location of the peptide molecule ($r = 0$). For the stationary function (10), as the peptide drifts away from the domain centre, only a relatively short part of its trajectory lies inside the 'useful' pure atomistic region, which resulted in relatively short time of the signal available for post-processing. To compensate for the short simulations times of the hybrid model, sophisticated post-processing of molecular trajectories was necessary, which inevitably reduced the accuracy of the final result in Ref. [7]. In the current work, the centre of the shape function $s = s(x, y, z, t)$ was linked to the peptide's centre of mass so that the peptide molecule remained inside the atomistic region at all times. This has allowed us to increase the simulation times by an order of magnitude and, thus, directly calculate the trajectories of the peptide without any intermediate assumptions. In turn, this resulted in obtaining an estimate of the diffusion coefficient of peptide in water much less contaminated by the statistical post-processing.

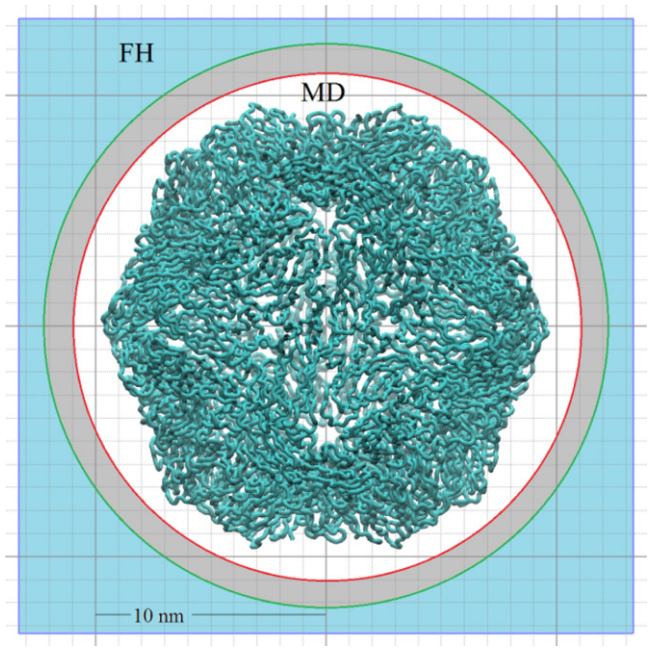


Fig. 3. PCV2 capsid in hybrid LL-FH/MD water; the capsid is inside the pure MD region (white) that gradually changes through the intermediate LL-FH/MD region (grey) to the FH region (blue); the small cells have the size of 1 nm that approximately corresponds to the FH cell size. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Whole virus simulation

Porcine circovirus type 2 (PCV2) has been linked to various diseases in pigs including postweaning multisystemic wasting syndrome (PMWS) and porcine dermatitis and nephropathy syndrome (PDNS), collectively named as porcine circovirus-associated disease (PCVADs), causing significant economic problems in swine industry worldwide [12–14]. As a member of the family Circoviridae, genus Circovirus, PCV2 has small nonenveloped icosahedral virus shell with a single-stranded circular DNA genome inside. The genome contains two major open-reading frames (ORFs), where ORF1 encodes the replicase proteins (Rep, Rep') and ORF2 encodes the capsid protein (Cap) [12,15,16]. The capsid protein can spontaneously assemble into empty PCV2-like particles. Empty capsids with morphology similar to PCV2 virus were detected experimentally [16,17].

The authors are not familiar with any previous simulation of PVC2 virus at all-atom resolution, which will be considered in the present work, but the topic of all-atom simulation of viruses has attracted a notable attention in the literature. One of the key questions addressed in these simulations is the stability of the virus

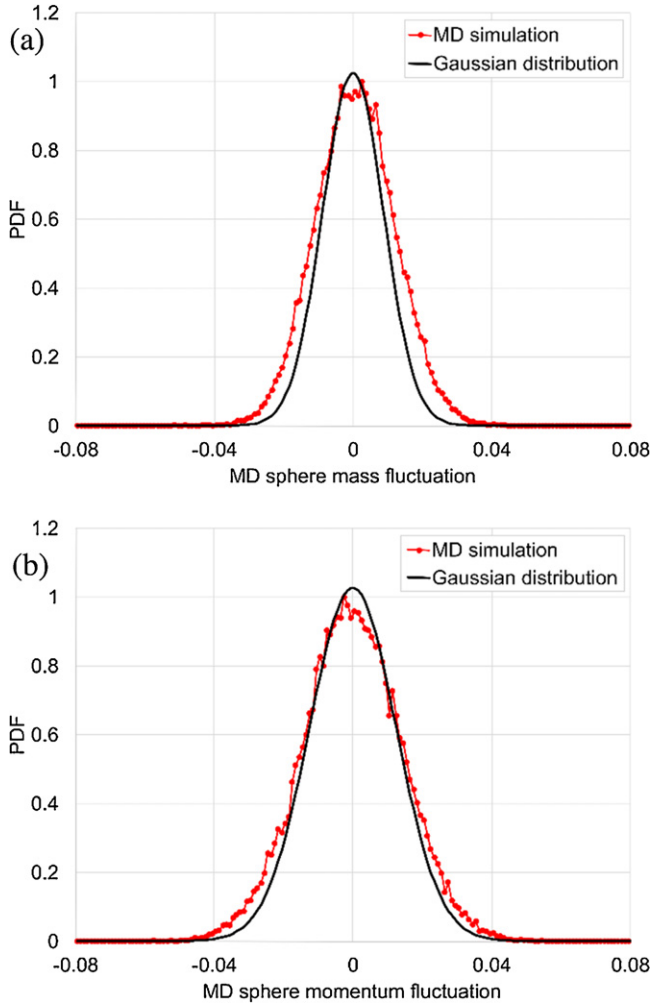


Fig. 4. Probability Density Function (PDF) distribution for mass (a) and x-component of the momentum (b); theoretical distributions in each case are shown as solid black line.

capsid in solution. For example, previous publications on all-atom simulation of the satellite tobacco mosaic virus [18] reported that after 10 ns of simulation the empty capsid was deformed. The simulated system had 932 500 atoms which included an empty capsid of the virus (without RNA) immersed in water, where the capsid of the virus alone was around 132 000 atoms. The authors argued that the collapse of the virus could be caused by the absence of RNA. All-atom simulation was also used for studying the capsid of the satellite tobacco necrosis virus [19], where similar changes in the capsid were connected to the lack of the calcium ions. In this case the system consisted of roughly 1 200 000 atoms.

The importance of having a large enough computational domain in the case of all-atom virus simulation was confirmed by Ref. [20] where capsids of the poliovirus in water solutions were found to be stable unlike in the previous examples. The simulation included 60 sphingosine molecules, 1 884 218 water molecules, 5310 sodium ions, 136 potassium ions, and 5206 chloride ions. The total number of atoms considered in that simulation was 6 480 236.

For the current publication, an isolated circovirus capsid placed in a large water bath under normal temperature and pressure conditions is considered. Under such conditions, in accordance with the experimental observations [16,17], the virus structure is known to remain stable. Detailed crystallographic structure of the protein subunit of the capsid has been reported [21] with atomistic resolution (PDB entry 3R0R). For the current modelling, the structure of the virus capsid reported in the latter experiment is used for

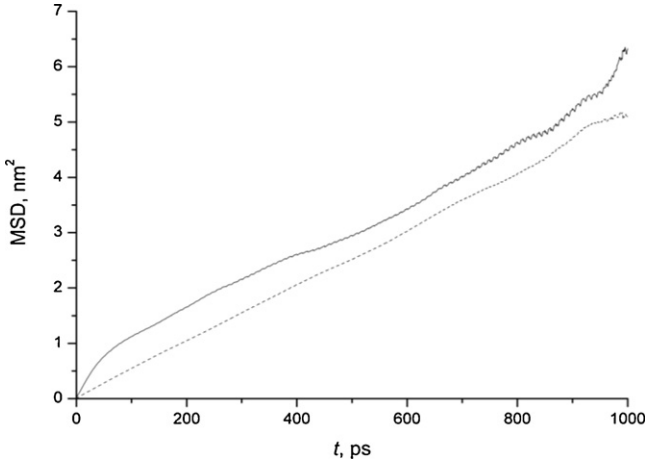


Fig. 5. The computed MSD(t) function for the peptide-in-water diffusion problem for the hybrid LL-FH/MD simulation (solid line) and in pure MD simulation (dashed line).

building the initial structure of the atomistic virus model without considering the DNA inside (Fig. 6). When immersed in water, the shape of the virus will self-adjust to correspond to the minimum free energy of the entire system and then the adjusted capsid shape should be further preserved in the simulation as it is observed in the experiment. Therefore, for atomistic modelling, any discrepancy of the solution from the initial state could be attributed to inaccuracies of the numerical boundary conditions which are imposed in the truncated computational domain to approximate the infinite water bath effects. For the same size of the computational domain, which contains the virus capsid surrounded by water, two boundary treatments were applied in the current study. One is the standard periodic boundary condition for the all-atom simulation and the other corresponds to the hybrid LL-FH/MD model (8).

The capsid has the diameter of ~18–20 nm and consists of 126 180 atoms which was immersed in the cubic cell with water with the edge length of 26.68 nm. For the LL-FH/MD model, the pure MD region was linked to the centre of mass of the capsid which could freely flow inside the computational domain, similar to the dialanine simulations. The internal MD region of 22.68 nm diameter also included a layer of water molecules that interacted with the capsid. The total number of atoms in the MD zone was approximately 590 000, which corresponded to the total of 1 837 905 particles in the entire computational domain including in the intermediate LL-FH/MD zone and the FH region (Fig. 3).

All MD calculations were performed with GROMOS96 53a6 force field [11], SPC water model. Hybrid LL-FH/MD simulations were performed with the Nose-Hoover thermostat (T = 300 K), reaction field electrostatics with cut-off length 1.0 nm and dielectric constant 78, van der Waals cut-off length 1.0 nm. The MD time step was 1 fs while the time step for the FH solver was 10 times larger (10 fs). The total simulation time was ~2 ns (220 000 MD time steps per day on 256 MPI cores).

For all-atom simulations, the same atomistic virus model inside the same cell was considered but with explicit atomistic modelling of all surrounding water and periodic boundary conditions applied at the cell boundaries. The total number of atoms in the all-atom model was equal to the number of particles in the LL-FH/MD model (1837905).

The parameters of the MD simulation are summarised in Table 1.

The molecular dynamics simulations were performed at constant volume using vdwtype equal to cut-off, Particle-Mesh Ewald electrostatics, v-rescale thermostat, and Verlet cutoff-scheme. For all-atom simulation, periodic boundary conditions in all 3 directions are used. It was made sure that the distance between the

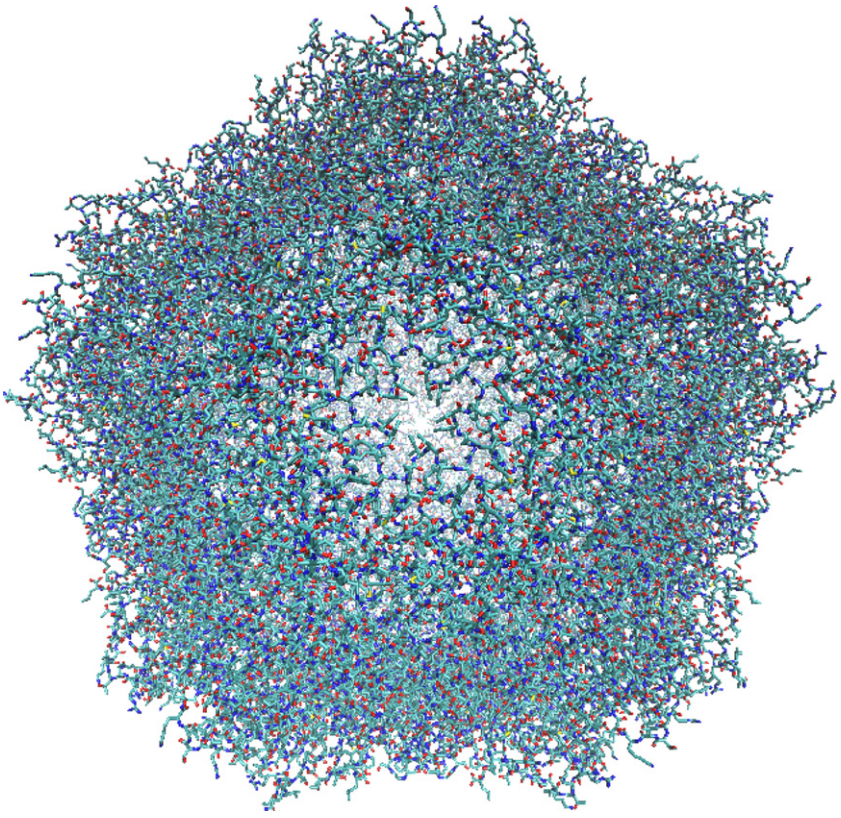


Fig. 6. PCV2 capsid initial configuration from the experimental X-ray data.

Table 1
Parameters of the MD simulations: L is the size of the cubic box, t is running time, Δt is the time step for integration, τ defines time constant for temperature coupling and T is the reference temperature for temperature coupling, R_{vdw} and R_{coul} are cut-off distances (Lennard-Jones and Particle-Mesh Ewald electrostatics for Coulomb respectively).

Parameter	Value	Units
L	26.68	nm
t	2000	ps
Δt	0.002	ps
τ	0.1	ps
T	300	K
R_{vdw}	1.0	nm
R_{coul}	1.0	nm

capsid and its periodic image is large enough in accordance with Ref. [22] who suggest to use at least 2 nm of water layers for making the effective protein–protein interaction negligible.

In preparing the initial conditions, after energy minimization, the MD calculation was carried out with position restraints of the capsid for the same MD parameters as used in the production run (Table 1). After 1 ns of simulation position restraints were removed and the equilibration run was conducted until the energy of the system reached a constant value.

4. Results

4.1. Dialanine diffusion in water

Before calculating the diffusion coefficient, the hybrid model of the dialanine in water has been checked on the preservation of correct mass and momentum fluctuations inside the pure molecular dynamics zone s = 0. This test was essential to confirm that atoms corresponding to the volume locked to the moving centre of mass of the peptide has the correct statistics in accordance to the theory for pure water at equilibrium conditions (neglecting the effect of the peptide on the relatively large water volume inside the sphere). It can be recalled at this point that it is only the combined mixture fluctuations which are automatically preserved in the current one-way coupling model and the solution representing the MD particles needs to be checked a posteriori. Notably, the necessity of an a-posteriori performance analysis for the current LL-FH/MD technique is similar to what is typical of other numerical open-boundary treatments in the modelling literature.

Fig. 4 show histograms of the computed probability density functions of the fluctuations of mass, $\frac{\sum_{p=1,N} m_p u_p}{\rho_0 V_0 c_T}$ and momentum, $\frac{\sum_{p=1,N} m_p - \rho_0 V_0}{\rho_0 V_0}$ inside the moving MD sphere with the peptide, where ρ_0 is the mean water density, V_0 is the volume of the MD zone, c_T is the isothermal sound speed in water, m_p, u_p are the mass and x-velocity component of particle p.

Theoretical Gaussian distributions of the mass and momentum corresponding to the prescribed temperature, volume, and speed of sound are plotted in the same graphs for comparison.

The theoretical distributions correspond to the following functions [1]:

$$f(x) = A \times \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{11}$$

where A is a calibration constant, σ is dispersion which is related to standard deviation (STD) so that

$$\sigma_m = \frac{std(\rho)}{\rho_0} \text{ for mass and } \sigma_{\rho u} = \sigma_m\sqrt{2} \text{ for momentum,} \tag{12}$$

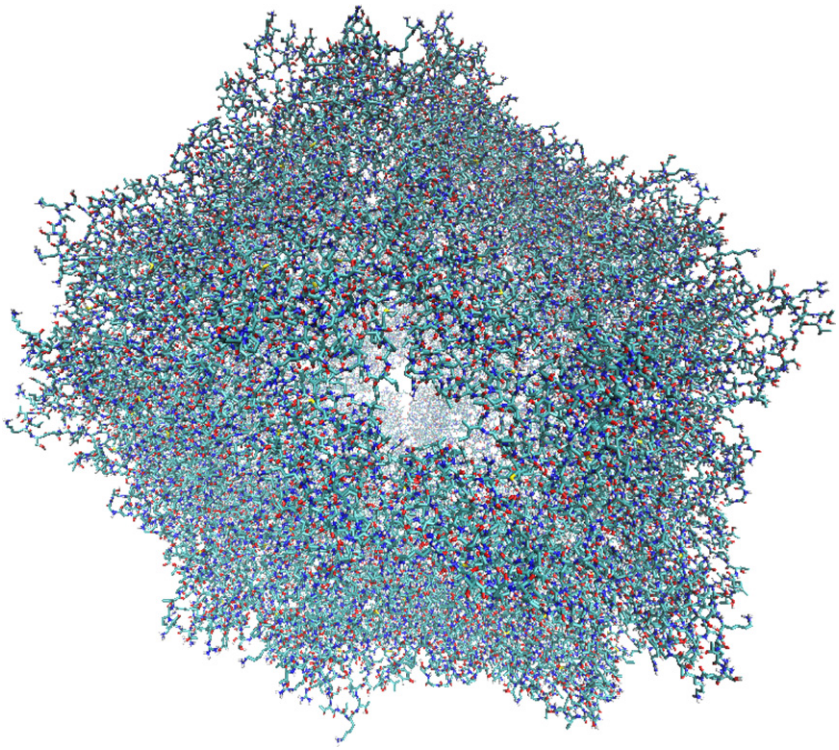


Fig. 7. PCV2 capsid is unstable in pure MD.

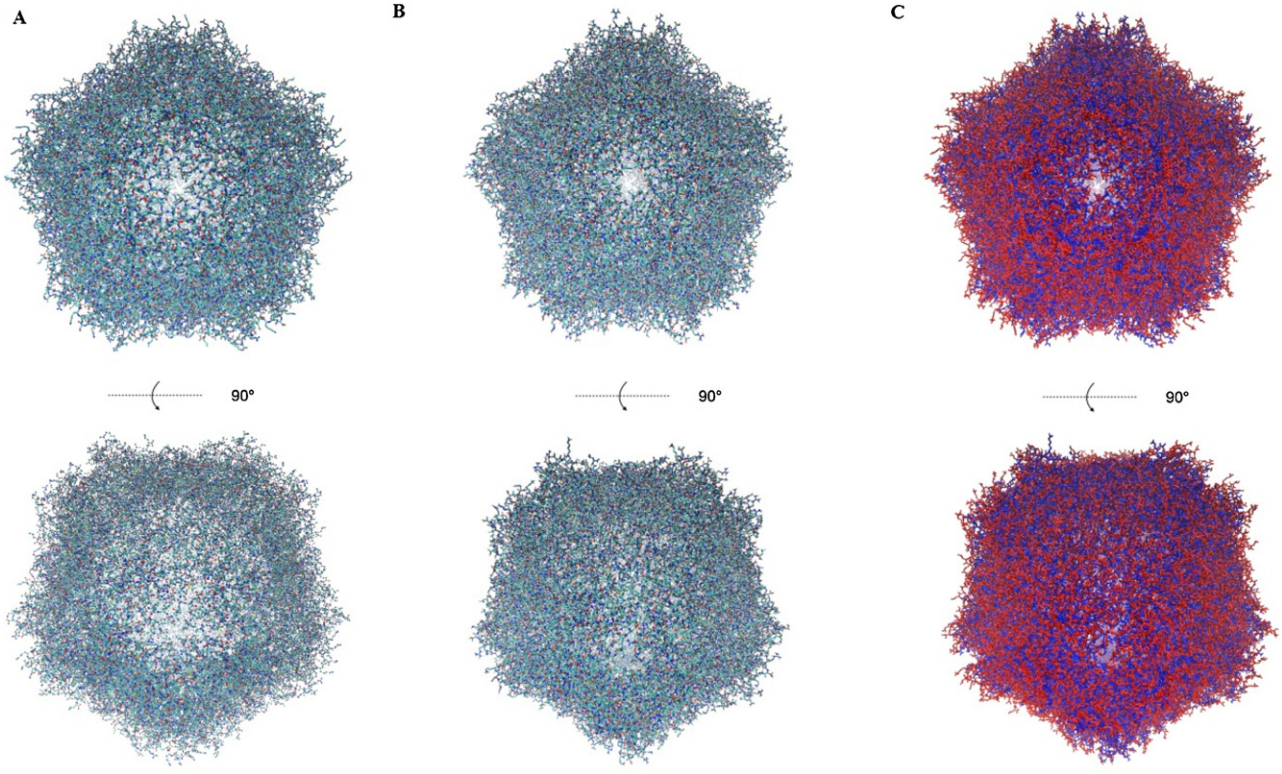


Fig. 8. PCV2 capsid is stable for the hybrid LL-FH/MD simulation: A is initial virus shape after the equilibration corresponding to the experimental structure; B is the simulated structure from hybrid LL-FH/MD; C is the comparison of A and B (blue—simulation, red—initial virus structure after the equilibration). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

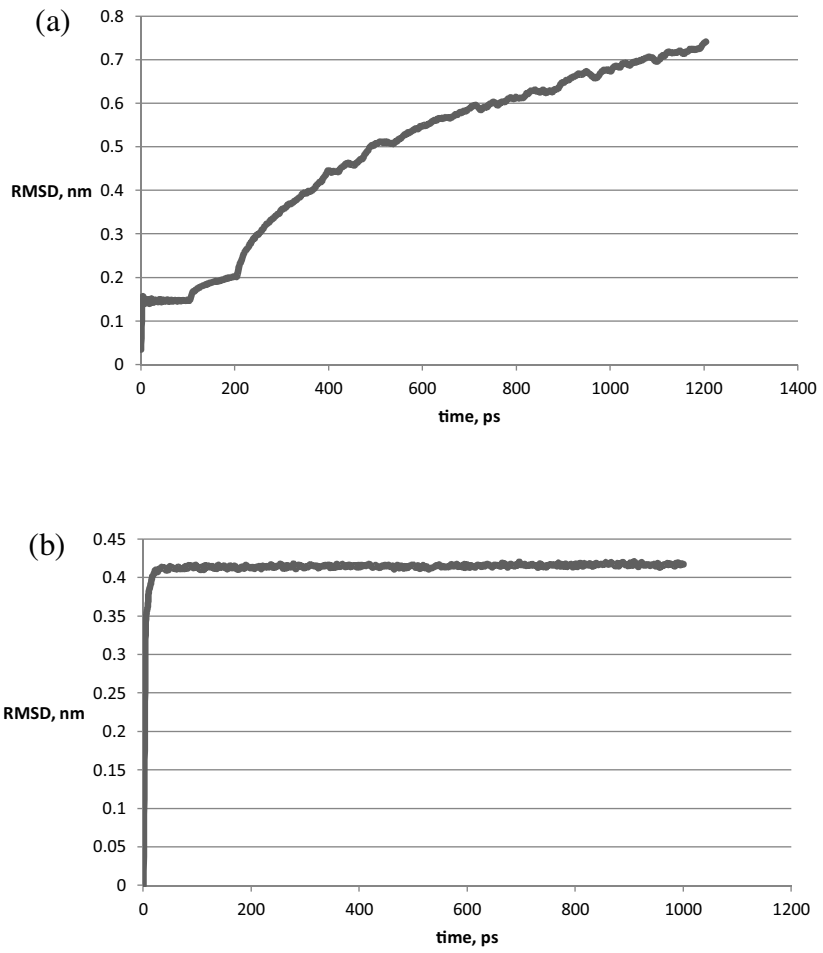


Fig. 9. Running averages of the root mean square deviation (RMSD) of backbone atoms from their initial structure for pure MD calculation (a) and the hybrid LL-FH/MD simulation (b).

and

$$\mathrm{std}(\rho) = \frac{1}{c_T}\sqrt{\rho_0 k_B \frac{T}{V_0}},\ \frac{std(u)}{c_T} = \frac{std(\rho)}{\rho_0}. \tag{13}$$

For the peptide diffusion in water, the length of the hybrid model simulation was 1 ns. In total, 50 runs were carried out starting from different initial configurations for better statistical averaging. For each run, the mean square deviation $MSD(t)$ was calculated and the final result was obtained by averaging all 50 trajectories. Fig. 5 shows the comparison of the resulting $MSD(t)$ trajectory obtained with the hybrid method where the reference MD result is also included.

To calculate the diffusion coefficient D, the approximately linear interval between 100 and 400 ps was used. To estimate the uncertainty of the value obtained, the original interval of 300 ps was broken down into several regions of 100 ps each with the 50 ps overlap. Each interval produced its own diffusion coefficient D_i in accordance with the local gradient and one half of the difference between the maximum and the minimum D_i was taken to be the uncertainty ΔD. The final value of the diffusion coefficient D is then found to be $(0.83 \pm 0.08) \times 10^{-5}\ \mathrm{cm}^2/\mathrm{s}$ which is in a very good agreement with the reference value from the pure MD simulation, $0.86 \times 10^{-5}\ \mathrm{cm}^2/\mathrm{s}$.

It can be noticed that unlike for the MD simulation, the hybrid solution experiences a step change of trajectory in the initial time moment. This step change is most likely associated with the stronger initial transients in the hybrid simulation case due to starting the calculation from the initial conditions which are more difficult to make entirely consistent for the hybrid model in comparison with the all-atom simulation.

4.2. Porcine circovirus in water

In comparison with the initial configuration and the LL-FH/MD model (Figs. 6 and 8), the all-atom computation in the domain of the same size with periodic boundary conditions gives a qualitatively different answer: the virus capsid collapses after approximately 1 ns of simulation time (Fig. 7). We attribute this to the artefact of the standard periodic boundary conditions in the all-atom simulation.

To assess the preservation of the virus capsid shape qualitatively, the root mean square deviation (RMSD) of the protein atoms in comparison with the initial structure is computed in each case: for the all-atom simulation and for the hybrid LL-FH/MD approach. Fig. 9a shows that the RMSD of the backbone atoms of the proteins in the case of pure MD simulations keeps increasing, which indicates that the system is not stabilised. In the case of LL-FH (Fig. 9b), calculations of the RMSD values asymptotes to a constant value after ~60 ps. This residual RMSD value is about 0.4 nm which small enough to suggest that the virus structure is stable in accordance with the literature [23] and [24] who find that the RMSD values of around 0.3 nm are acceptable for stability of large proteins.

Interestingly, the collapse of the virus capsid is observed in pure MD simulation even when the simulation box is increased to contain about 4 000 000 atoms, that is much more water molecules surround the virus. On the contrary, for the LL-FH/MD model, a twice larger computational domain, corresponding to a twice larger

size of the pure MD zone (approximately 1 180 000 atoms) and effectively bringing the 'hybrid boundary' away from the virus located in the centre of the pure MD zone, produced correct symmetric structure very similar to the experimental. The result of this verification test was that the shape of the virus remained virtually unchanged regardless of the size of computational domain used. Notably, RMSD of the capsid's backbone atoms were 0.4 nm for smaller computational domain and 0.5 nm for the larger one, which demonstrates essentially unchanged structure compared to the crystallographic experimental data.

The above results suggest that in order to preserve the virus in an all-atom simulation with the standard periodic boundary conditions the size of the computational domain, hence, the number of water atoms and the associated computational cost should be further increased. The introduction of the LL-FH/MD model, in turn, allows performing simulations of the problem in the same reasonably small computational box without notable numerical artefacts.

Moreover, since the calculation of MD forces is turned off inside the FH region of the LL-FH/MD model it speeds up the simulation by several times (depending on the MD/FH volume ratio) in comparison with the all-atom simulation for the same computational box size. For example, in the current simulation the FH region contains approximately 1/2 part of the total number of particles in the box. Assuming that the MD potential calculation is the most computationally expensive part of the numerical algorithm (MD force estimation takes up to 90% of all computational time per each time iteration), one obtains a factor of two computational speed-up when calculating the MD forces only for the particles inside the pure MD zone and the intermediate LL-FH/MD region. The benefits of hybrid approach should become even more pronounced in comparison with the all-atom simulation for larger hydrodynamic (FH) regions of the computational domain.

5. Conclusions

The hybrid LL-FH/MD method [6,7] has been extended to dynamically tracking the atomistic-resolution features of interest which can evolve in the flow. Simulations of two very different biomolecular systems have been performed: the peptide diffusion in water and interaction of a circovirus capsid with water. It has been shown that the hybrid model can preserve important properties of the molecular part of the system. For the problem of peptide diffusion in water, the fluctuations of the mass and momentum of the pure molecular dynamics zone surrounding the peptide were found in an excellent agreement with the theory. The resulting diffusion coefficient from the hybrid solution was found in a very good agreement with the reference all atom simulation. For the simulation of a capsid of virus porcine circovirus 2 in water, in comparison with the all-atom simulations, the current hybrid method preserves the structure of the virus and leads to the root mean square deviation (RMSD) of the backbone atom structure which asymptotes to a constant value. This constant value is small enough to suggest the structure stability and it remains reasonably insensitive to the size of the computational domain in case of the hybrid method. This is in contrast with the all-atom simulation with the standard periodic boundary conditions which can neither produce a constant RMSD, nor preserve stability of the virus capsid even for an increased size of the computational domain.

Further development of the current hybrid method will include its extensions to the two-way coupling algorithms in 3D to include the feedback from the molecular dynamics on hydrodynamics. Additional developments will involve extension of the current hybrid approach, which uses constant integration volumes for hydrodynamics and constant time steps for MD particles regardless of the value of s-parameter, which effectively plays the role of the 'coarse-graining' (MD to LL-FH) scale, to multi-space-time stepping algorithms such as TARDIS [25] in order to further increase the computational efficiency.

Acknowledgements

The work has been supported by Engineering and Physical Sciences Research Council (EP/J004308/1) in the framework of the G8 Research Councils Initiative on Multilateral Research Funding. SK is grateful to the Royal Society of London for their support. DN thanks the Royal Society of Chemistry for the JWT fellowship and Royal Academy of Engineering and Leverhulme Trust for Senior Research Fellowship. The supporting data of this study are stored at the University of Aston. Details of how to request access to these data are provided in the documentation available from the University of Aston research data repository at http://doi.org/10.17036/de4b474e-0497-4aa4-9676-439644b1eb48.

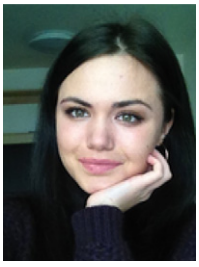
References

[1] N.K. Voulgarakis, J.-W. Chu, Bridging fluctuating hydrodynamics and molecular dynamics simulations of fluids, J. Chem. Phys 130 (2009) 134111.
[2] J.B. Bell, A.L. Garcia, S.A. Williams, Computational fluctuating fluid dynamics, ESAIM: Math. Model. Numer. Anal. 44 (2010) 1085–1105.
[3] D.M. Holland, M.K. Borg, D.A. Lockerby, J.M. Reese, Comput. Fluids 115 (2015) 46–53.
[4] A. Alexiadis, D.A. Lockerby, M.K. Borg, J.M. Reese, A particle-continuum hybrid framework for transport phenomena and chemical reactions in multicomponent systems at the micro and nanoscale, J. Heat Transfer 137 (9) (2015) 091010.
[5] G. De Fabritiis, R. Delgado-Buscalioni, P.V. Coveney, Multiscale modeling of liquids with molecular specificity, Phys. Rev. Lett. 97 (2006) 134501.
[6] A. Markesteijn, S. Karabasov, A. Scukins, D. Nerukh, V. Glotov, V. Goloviznin, Concurrent multiscale modelling of atomistic and hydrodynamic processes in liquids, Phil. Trans. R. Soc. A 372 (2014) 20130379.
[7] I. Korotkin, S. Karabasov, D. Nerukh, A. Markesteijn, A. Scukins, V. Farafonov, E. Pavlov, A hybrid molecular dynamics/fluctuating hydrodynamics method for modelling liquids at multiple scales in space and time, J. Chem. Phys. 143 (2015) 014110.
[8] H.J.C. Berendsen, et al., GROMACS: a message-passing parallel molecular dynamics implementation, Comp. Phys. Commun. 91 (1995) 43–56.
[9] D.G. Crighton, A.P. Dowling, J.E. Ffowcs Williams, M.A. Heckl, F.A. Leppington, Modern Methods in Analytical Acoustics, Lecture Notes (October) (2013).
[10] A.P. Markesteijn, S.A. Karabasov, Glotov VYu, V.M. Goloviznin, Comput. Methods Appl. Mech. Eng. 281 (2014) 29–53.
[11] C. Oostenbrink, A. Villa, A.E. Mark, W.F. Van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, J. Comput. Chem. 25 (2004) 1656.
[12] Xiang-Jin Meng, Circoviridae, in: D.M. Knipe, P.M. Howley (Eds.), Fields Virology, Lippincott Williams & Wilkins, Philadelphia, 2013, pp. 1792–1801.
[13] A.G. Cino-Ozuna, S. Henry, R. Hesse, J.C. Nietfeld, J. Bai, H.M. Scott, R.R.R. Rowland, Characterization of a new disease syndrome associated with porcine circovirus type 2 in previously vaccinated herds, J. Clin. Microbiol. 49 (2011) 2012–2016.
[14] J. Segalés, T. Kekarainen, M. Cortey, The natural history of porcine circovirus type 2: from an inoffensive virus to a devastating swine disease? Vet. Microbiol. 165 (2013) 13–20.
[15] S. Yin, S. Sun, S. Yang, Y. Shang, X. Cai, X. Liu, Self-assembly of virus-like particles of porcine circovirus type 2 capsid protein expressed from Escherichia coli, Virol. J. 7 (2010) 166.
[16] P. Nawagitgul, I. Morozov, S.R. Bolin, P.A. Harms, S.D. Sorden, P.S. Paul, Open reading frame 2 of porcine circovirus type 2 encodes a major capsid protein, J. Gen. Virol. 81 (2000) 2281–2287.
[17] H. Fan, C. Ju, T. Tong, H. Huang, J. Lv, H. Chen, Immunogenicity of empty capsids of porcine circovius type 2 produced in insect cells, Vet. Res. Commun. 31 (2007) 487–496.
[18] P.L. Freddolino, A.S. Arkhipov, S.B. Larson, A. McPherson, K. Schulten, Molecular dynamics simulations of the complete satellite tobacco mosaic virus, Structure 14 (2006) 437–449.
[19] D.S. Larsson, L. Liljas, D. van der Spoel, Virus capsid dissolution studied by microsecond molecular dynamics simulations, PLoS Comput. Biol. 8 (2012) e1002502.
[20] Y. Andoh, N. Yoshii, A. Yamada, K. Fujimoto, H. Kojima, K. Mizutani, A. Nakagawa, A. Nomoto, S. Okazaki, All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution, J. Chem. Phys. 141 (2014) 165101.
[21] R. Khayat, N. Brunn, J.A. Speir, J.M. Hardham, R.G. Ankenbauer, A. Schneemann, J.E. Johnson, The 2.3-Angstrom structure of porcine circovirus 2, J. Virology 85 (2011) 7856.

[22] J.W. Chu, G.A. Voth, Coarse-Grained modeling of the actin filament derived from atomistic-Scale, Simul. Biophys J. 90 (2006) 1572–1582.
[23] T. Baig, H. Sheikh, P.K. Tripathi, In-Silico studies of oncogene protein with anti-cancer drugs, Glob. J. Biotechnol. Biochem. 9 (2014) 65–75.
[24] P.L. Freddolino, A.S. Arkhipov, S.B. Larson, A. McPherson, K. Schulten, Molecular dynamics simulations of the complete satellite tobacco mosaic virus, Structure 14 (2006) 437–449.
[25] A.P. Markesteijn, S.A. Karabasov, Time asynchronous relative dimension in space method for multi-scale problems in fluid dynamics, J. Comput. Phys. 258 (2014) 137–164.

Dr Korotkin is a Senior Research Fellow at Moscow Institute for Nuclear Safety and also in the Laboratory of Industrial Mathematics at Lomonosov Moscow State University. He has been studying and working in top Russian institutions (Moscow Institute of Physics and Technology, the alma mater of the two recent Nobel Prize laureates in physics, and Moscow State University) as well as in Queen Mary University of London. His work has been recognised by several prestigious awards in Russia, e.g. the medal of Russian Academy of Sciences for the best scientific publication among students (2003), the Moscow Government award for scientific research (2005), the "New Generation" Contest award by RAO UES Russia (2007) and, internationally, by a CRDF Grant Assistance Program of Department of Energy, USA.

Dmitry Nerukh is a Senior Lecturer in the School of Engineering and Applied Science at Aston University. His research interests are in the field of complex non-linear dynamics in application to molecular systems. One of the main sources of data for his research is high performance simulations of large biomolecular systems. The simulations are the basis for his intense ongoing collaboration with RIKEN Advanced Institute for Computational Science, Japan where the world's fastest computer for protein simulation has been built. Before the appointment in Aston University he worked in the Unilever Centre for Molecular Informatics, a part of Chemistry Department, University of Cambridge where he has established and led for nine years the Centre's research in the field of complex dynamics of molecular systems (emergent behaviour of molecular systems resulting in spontaneous formation of bio-chemically important structures).



Elvira Tarasova is a PhD Candidate at Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation. Her area of research is molecular dynamics and she is a part of Dr Nerukh's group.

Vladimir Farafonov is a PhD Candidate in the Department of Physical Chemistry at V.N. Karazin Kharkiv National University, Kharkiv, Ukraine. His area of research is molecular dynamics and he is a part of Dr Nerukh's group.

Sergey Karabasov is a Reader in Computational Modelling in Queen Mary University of London, School of Engineering and Materials Science (SEMS). Prior to joining SEMS as a Senior Lecturer in 2012, SK was a Royal Society University Research Fellow in Cambridge University Engineering Department. His research is devoted to high-resolution and multiscale modelling in computational aero and hydrodynamics. Over the years, his research has been supported by the leading UK industry as well as UK Research Councils and the Royal Society of London. Internationally, his research into multiscale modelling methods for sound generated by jet turbulence in collaboration with Ohio Aerospace Institute was supported by Aero Acoustics Research Consortium (AARC). He was an organiser of the Royal Society Scientific Seminars in 2011, 2013 and a guest editor of the Royal Society PhilTrans A special issue "Multiscale systems in fluids and soft matter: approaches, numerics, and applications" in 2014.