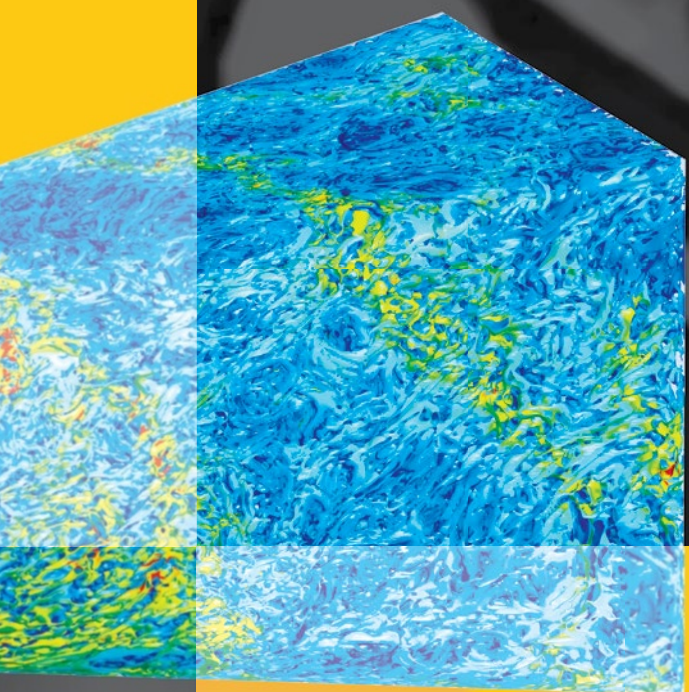
Wolfgang E. Nagel
Dietmar H. Kröner
Michael M. Resch *Editors*

High Performance Computing in Science and Engineering '17



H L R S

Springer

High Performance Computing in Science and Engineering ’17

Wolfgang E. Nagel • Dietmar H. Kröner •
Michael M. Resch
Editors

High Performance Computing in Science and Engineering ’17

Transactions of the High Performance Computing Center, Stuttgart (HLRS) 2017

Springer

Editors
Wolfgang E. Nagel
Zentrum für Informationsdienste
und Hochleistungsrechnen (ZIH)
Technische Universität Dresden
Dresden, Germany

Dietmar H. Kröner
Abteilung für Angewandte Mathematik
Universität Freiburg
Freiburg, Germany

Michael M. Resch
Höchstleistungsrechenzentrum
Stuttgart (HLRS)
Universität Stuttgart
Stuttgart, Germany

Front cover figure: Turbulent structures in the draft tube of a hydraulic propeller turbine operated in part load condition. The turbulent flow structures are colored with the eddy viscosity ratio. A high viscosity ratio (red) indicates regions with high turbulence. Details can be found in "Numerical Analysis of a Propeller Turbine Operated in Part Load Conditions" by Bernd Junginger and Stefan Riedelbauch, Institute of Fluid Mechanics and Hydraulic Machinery, University of Stuttgart, Germany on page 355ff.

ISBN 978-3-319-68393-5 ISBN 978-3-319-68394-2 (eBook)
https://doi.org/10.1007/978-3-319-68394-2

Library of Congress Control Number: 2017964382

Mathematics Subject Classification (2010): 65Cxx, 65C99, 68U20

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Part I
Physics

Peter Nielaba

In this section, six physics projects are described, which achieved important scientific results in 2016/2017 by using Hazel Hen at the HLRS and ForHLR of the Steinbuch Center.

Fascinating new results are being presented in the following pages on soft matter systems (simulations of fibrinogen properties and of the behaviour of blood proteins at nanoparticles), astrophysical systems (simulations of galaxy formation and of photoionization and radiative charge transfer), many body quantum systems (simulations of trapped ultracold quantum systems) and high energy physics systems (simulations of partial pressures of mesons and baryons with different strangeness quantum numbers and of quantum anomalies, topologies and particle production in strongly correlated gauge theories).

Studies of the soft matter systems have focused on the properties of fibrinogen and on the behaviour of blood proteins at nanoparticles.

T. Schäfer, J. Zhou, F. Schmid and G. Settanni from the University of Mainz investigated in their project *Flexadfg* properties of fibrinogen as well as the behaviour of blood proteins in the presence of the polymer poly(ethylene-glycol) by atomistic Molecular Dynamics simulations. Their simulations reveal important interaction details and permit the construction of simplified models.

The studies of the astrophysical systems have focused on galaxy formation and on photoionization and radiative charge transfer of several particles.

V. Springel, A. Pillepich, R. Weinberger, R. Pakmor, L. Hernquist, D. Nelson, S. Genel, M. Vogelsberger, F. Marinacci, J. Naiman, and P. Torrey from Heidelberg (V.S., A.P., R.W., R.P.), Cambridge, USA (L.H., J.N., M.V., F.M., P.T.), New York

P. Nielaba (✉)
Fachbereich Physik, Universität Konstanz, 78457 Konstanz, Germany
e-mail: peter.nielaba@uni-konstanz.de

(S.G.) and Garching (D.N.), present in their project *GCS-ILLU* results from a new generation of hydrodynamical simulations ("IllustrisTNG" project, AREPO code), including new black hole physics, and using more accurate techniques, an enlarged dynamical range, and system sizes with side lengths of up to 300 megaparsec. The authors report on early results on matter and galaxy clustering and find very good agreement with observation data for the two-point galaxy correlation function. According to their investigations, the effect of baryonic physics has to be taken into account in cosmological precision studies.

B.M. McLaughlin, C.P. Ballance, M.S. Pindzola, P.C. Stancil, J.F. Babb, S. Schippers and A. Müller from the Universities of Belfast (B.M.M., C.P.B.), Auburn (M.S.P.), Georgia (P.C.S.), Cambridge USA (J.F.B.), and Giessen (S.S., A.M.) investigated in their project *PAMOP* atomic, molecular and optical collisions on petaflop machines, relevant for astrophysical applications, for magnetically-confined fusion and plasma modeling, and as theoretical support for experiments and satellite observations. The Schrödinger and Dirac equations have been solved with the R-matrix or R-matrix with pseudo-states approach, and the time dependent close-coupling method has been used. Various systems and phenomena have been investigated, ranging from the photoionization of chlorine ions, of cobalt ions and of tungsten ions, and of He and H_2, to the collision scenario of SiO and H_2, and to the radiative charge transfer in systems of carbon atoms and He ions, and the radiative association in systems of carbon atoms and H ions.

In addition to the projects discussed above, interesting general relativistic properties of neutron stars and mergers have been investigated in project *HypeBBH* by using Hazel Hen.

In the last granting period, quantum mechanical properties of high energy physics systems have been investigated as well as the quantum many body dynamics of trapped bosonic systems.

P. Alba, R. Bellwied, S. Borsanyi, Z. Fodor, J. Günther, S.D. Katz , V. Mantovani Sarti, J. Noronha-Hostler, P. Parotto, A. Pasztor, I. Portillo Vazquez, and C. Ratti from the Universities of Frankfurt (P.A.), Houston (R.B., J.N.-H., P.P., I.P.V., C.R.), Wuppertal (S.B., Z.F., J.G., A.P.), Budapest (Z.F., S.D.K.), Torino (V.M.S.), and the FZ Jülich (Z.F.) computed in their project *GCS-POSR* the partial pressures of mesons and baryons with different strangeness quantum numbers by lattice simulations in the confined phase of quantum chromodynamics. A comparison of the effect of different hadronic spectra on thermodynamic observables in the Hadron Resonance Gas model with results from lattice quantum chromodynamics simulations reveals that additional states, e.g. from the quark model or the particle data group booklet, are are required in the Hadron Resonance Gas model for a better agreement.

N. Mueller, J. Berges, O. Garcia, and N. Tanji from the University of Heidelberg investigate in the project *RTGT* quantum anomalies, topologies and particle production in strongly correlated gauge theories by real-time lattice simulations in the Glasma model, using computer time at the ForHLR. The project has achieved many interesting results for various high energy physics phenomena, ranging from out-

of-equilibrium transient anomalous charge production from coherent color fields, to the anomalous transport in ultra-relativistic heavy ion collisions, and to the anomalous dynamical refringence in quantum electrodynamics experiments beyond the Schwinger limits.

O.E. Alon, R. Beinke, C. Bruder, L.S. Cederbaum, S. Klaiman, A.U.J. Lode, K. Sakman, M. Theisen, M.C. Tsatsos, S.E. Weiner, and A.I. Streltsov from the Universities of Haifa (O.E.A.), Heidelberg (R.B., L.S.C., S.K., M.T., A.I.S.), Basel (C.B., A.U.J.L.), Wien (K.S.), Sao Paulo (M.C.T.), and Berkeley (A.S.) studied in their project *MCTDHB* trapped ultracold atomic systems by their method termed multiconfigurational time-dependent Hartree for bosons (MCTDHB). The principal investigators investigated the interplay of light and ultra-cold matter on fragmented superradiance, the effect of angular momentum on fragmented vortices, the phenomena accompanying a splitting of a two-dimensional Bose-Einstein-Condensate, many-body excitations in finite condensates in double-well and triple-well traps, and the differences of the Bose-Einstein-Condensate wave function in mean-field and many-body computations.

Blood Proteins and Their Interactions with Nanoparticles Investigated Using Molecular Dynamics Simulations

Timo Schäfer, Jiajia Zhou, Friederike Schmid, and Giovanni Settanni

Abstract Blood proteins play a fundamental role in determining the response of the organism to the injection of drugs or, more in general, of therapeutic preparations in the blood stream. Some of these proteins are responsible for mediating immune response and coagulation. Nanoparticles, which are being intensely investigated as possible drug nanocarriers, heavily interact with blood proteins and their ultimate fate is determined by these interactions. Here we report the results of molecular dynamics simulations of several blood proteins aimed to determining their possible behavior at the nanoparticle surface. On one hand we investigated the behavior of fibrinogen, a glycoprotein, which polymerizes into fibrin during coagulation. On the other hand we investigated the behavior of several blood proteins in the presence of the polymer poly (ethylene-glycol), often used as nanoparticle coating to reduce unspecific interactions with the surrounding environment.

1 Introduction

Our research activity, centered on studying the behavior of blood proteins at the interface with (nano)materials has followed two main directions: on one hand, after revealing the flexibility of Fibrinogen [1], a central protein of the coagulation cascade, and its adsorption properties on model surfaces [2], we have studied with enhanced sampling techniques the full extent of its flexibility, paving the way to effective simulations of its adsorption process on realistic surface; on the other hand we have studied the interactions of proteins with poly(ethylene-glycol), a polymer which is very frequently used to improve the properties of therapeutic preparations including nanoparticles. The report is thus subdivided in two sections.

T. Schäfer • F. Schmid • G. Settanni (✉)
Institut für Physik, Johannes Gutenberg University, Mainz, Germany
e-mail: settanni@uni-mainz.de

J. Zhou
Institute of Physics, Johannes-Gutenberg University, Mainz, Germany

W.E. Nagel et al. (eds.), *High Performance Computing in Science and Engineering '17*, https://doi.org/10.1007/978-3-319-68394-2_1

2 Metadynamics of the Fibrinogen Protomer

Fibrinogen (Fg) is a multi-chain serum glycoprotein consisting of a dimer of identical protomers, each made up of three different peptide chains Aα, Bβ and γ and two carbohydrates (Fig. 1). Fg has a molecular weight of 340 kDa and an elongated structure. Each protomer consists of the globular D- and E-domains, connected via a coiled-coil region [3]. At the dimerization interface in the E-domain, the three chains of either protomer are connected covalently via disulphide bridges. Each protomer is diglycosylated at the D-domain and the coiled-coil region. Parts of ends of the three peptide chains are unstructured and flexible.

Fg is a major element in blood clot formation [4]. In the coagulation cascade, thrombin transforms Fg into fibrin, one of the main components of blood clots. Fibrinolysis, the cleavage of the fibrin mesh by the enzyme plasmin, controls the dissolution of the blood clot. In this process, plasmin cleaves fibrin at specific plasmin cleavage sites, some of which are located at the coiled-coil region between the D- and E-domains.

In our previous investigations of Fg dynamics [1] and adsorption behavior [2], carried out using classical atomistic molecular dynamics simulations, the presence of consistent bending motions at a hinge centered on the coiled-coil region of each protomer was observed. It was shown that these motions could represent a conserved mechanism for the exposure of early plasmin cleavage sites.

While Fg underwent significant bending motions in our molecular dynamics simulations under physiological conditions, considerable simulation time would be necessary to fully sample them due to their large timescale and variability. To speed up their sampling and reduce the computational burden, we have employed metadynamics [5], an enhanced sampling method which biases the system towards unexplored regions of the conformational space. In these simulations, the projections of the motion along its principal components (obtained via principal component analysis (PCA) [6] of our previous simulations of Fg [1]) were adopted as collective variables (CVs) that capture the main characteristics of Fg motion. In a metadynamics simulation, an external biasing potential acting on the CVs is steadily built up to force the system and explore new regions of CV space.

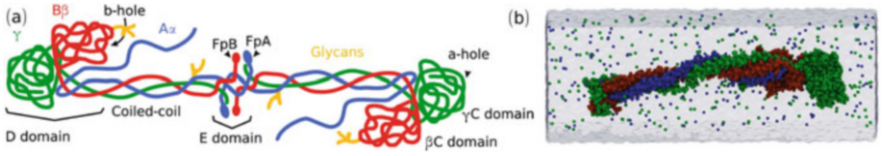


Fig. 1 Fibrinogen Structure and protomer system: (**a**) Sketch of the Fibrinogen Dimer. (**b**) Van der Waals representation of the protomer in solution, water molecules omitted for clarity, sodium and chloride ions in blue and green, respectively

2.1 Methods

The crystal structure of human Fg (PDB ID: 3GHG) [3] was used for the initial atomic coordinates of the protein. Due to the flexibility of the αC region and the N-terminal regions of all three chains, they do not reach a consistent stable state in the crystal, leaving their coordinates unresolved. Only the resolved carbohydrate group attached at residue β364 was included. One protomer (with the addition of only a small segment of the second protomer to complete the E-domain) was used as the minimal system that captures all relevant aspects of Fg structure.

The protomer was solvated in a rectangular simulation box with explicit TIP3P water [7] and a physiological concentration of ions (0.15 Mol [NaCl]) using VMD [8]. The initial size of 10.14 nm × 30.33 nm × 10.14 nm of the periodic unit cell of the system ensured a minimal padding of 1.5 nm, with a total number of atoms of 316,862. Preparation of the system was performed with GROMACS 4.6.5 [9], using an energy minimization followed by an equilibration in the Isothermal-isobaric ensemble at a temperature of 310 K and a pressure of 1 atm for 150 ps. A stochastic velocity rescaling thermostat [10] with a time constant of 0.1 ps and the Parrinello-Rahman barostat [11] with a time constant of 1 ps were used. The CHARMM36 force field with CMAP correction [12] was employed. Cutoffs of 1.2 nm for van der Waals and short-range electrostatic interactions were used, with treatment of long-range electrostatic interactions via Particle-Mesh-Ewald with a grid spacing of 0.1 nm. The first 50 ps of equilibration were performed using a time step of 1 fs, for everything else, a time step of 2 fs was used. All covalent bonds involving hydrogen were constrained in their length.

The metadynamics bias potential V is built up over the course of a simulation by periodically adding (i.e., at regular intervals τ) to the potential energy of the system a Gaussian function in CV space centered at the current position of the system s_t . For the purpose of performing metadynamics of Fg, version 5.0.4 of GROMACS [13] patched with PLUMED 2.2 [14] was used.

Results of principal component analysis (PCA) [6] performed on the Cα atoms of equilibrium trajectories of the Fg protomer in solution [1] were used to characterize relevant degrees of freedom. The first two PCs, describing bending motions at the hinge region, were used as CVs for the metadynamics. A multiple walker implementation of metadynamics with 10 parallel simulations sharing the same bias potential was used to speed up the sampling of PC space. Gaussian hills with a fixed height of $w = 0.7437 \frac{\text{kJ}}{\text{mol}}$, corresponding to a thermal energy of 0.3 $k_\text{B}T$ were deposited by each walker every $\tau = 4$ ps. The Gaussian width of the hills was set to $\delta = 100\,\text{nm}^2$ for both PCs.

Due to the significant computational cost of calculating the PCs, a multiple time stepping (MTS) integrator of order 10 was applied to the bias potential [15]. To ensure sufficient accuracy of the simulation, the effective energy drift [15] introduced by the MTS integrator was monitored. The effective energy of the bias potential describes the generalized detailed balance violation of the system, taking into account the inherent non-equilibrium character of metadynamics.

GROMACS is a molecular dynamics package optimized for performance and scaling. PLUMED provides the implementation of the metadynamics scheme and interfaces with the GROMACS code. In this way, PLUMED extracts coordinates and forces from GROMACS at runtime and adds the forces caused by the metadynamics bias as well as providing the communication between the individual simulations in the multiple walker scheme. For the production run of the simulations performed on Hazelhen at the High performance computing center Stuttgart, GROMACS and PLUMED were compiled specifically for the CRAY XC40 architecture with MPI parallelization. Good scaling up for up to thousands of cores was achieved. Each production job used 100 nodes, lasted at most 3.5 h and gathered 5 ns of trajectory distributed equally onto the ten walkers.

2.2 *Results*

The length of the individual trajectories of the walkers reached 100 ns, resulting in 1 μs of total simulation time. At one part of the simulation, jumps in the PCs appeared as artifacts, due to the ambiguity of the optimal alignment of instantaneous and reference coordinates. The decision was therefore made to remove this part of the simulations from all further analysis, with a total of 980 ns of trajectories remaining. For the remaining simulations the average effective energy drift of $4.8\,\frac{\text{kJ}}{\text{mol·ns}}$ was comparatively low, indicating multiple time stepping without significant losses in accuracy.

At convergence, the negative of the accumulated bias potential of metadynamics simulations provides an estimate of the underlying free energy surface of the system in terms of the CVs. The free energy landscape of Fg as identified by our simulations showed a single minimum centered roughly around the crystallographic conformation (Fig. 2a). The minimum was relatively deep, differing by 158 $\frac{\text{kJ}}{\text{mol}}$ from

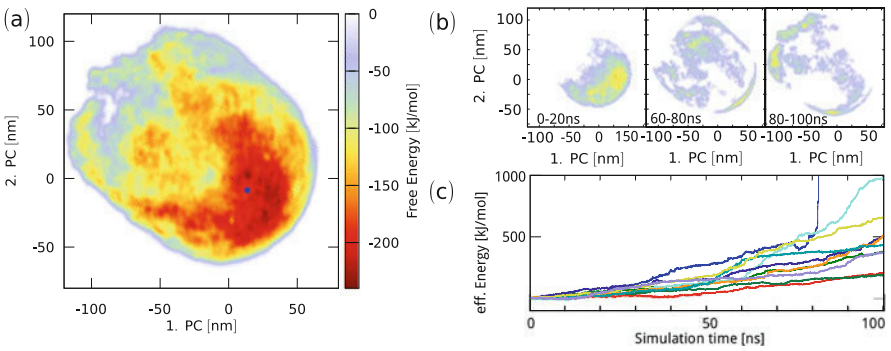


Fig. 2 Free energy surface: (**a**) Estimated free energy from full simulation. The blue dot marks the position of the crystallographic structure. (**b**) Free energy contribution of each walker from 0 to 20 ns, 60 to 80 ns and 80 to 100 ns of simulation time. (**c**) Effective energy of each walker, third walker (blue) reaches 35,000 kJ/mol after 100 ns

the average value of free energy measured across the sampled regions. In the initial part of the simulations the various walkers sampled mostly the minimum region. Later, the accumulated bias potential filled up this minimum allowing all the walkers to explore a wide area of CV space.

One pair of collective variables that has shown its use in the description of the protomer is the bending and torsion angles γ and ϕ. γ is the angle between the center of mass of E-domain, hinge and D-domain, while two different points in the E-domain together with the centers of hinge and D-domain were used to define the torsion angle ϕ. The correlation between bending and torsion at the hinge and the PCs that was noted before [1] was also present in the metadynamics simulations: the Pearson correlation coefficient r between γ and a linear combination of the first two PCs (1st–2nd–3rd) was 0.98, while r was 0.45 between ϕ and a combination of the first four PCs ($-$2nd/5+3rd+4th).

Figure 3a shows the overall distribution of the bending angle γ and the torsion angle ϕ as observed in the metadynamics simulations presented here. The angle distribution showed that Fg reaches heavily bent conformations with a bending angle below 30°, and that such a bending is only possible at certain torsion angles. In general, torsion angles around $\pm 180°$ or $-50°$ allowed strong bending, while values of ϕ around 70° heavily restrict it.

Analysis reveals two distinct bending modes (B1 and B2) characterized by different values of the torsion angle (Fig. 3) as well as an area where bending is heavily restricted (R). B1 and B2 conformations were clearly separated in PC-space, occupying roughly half of the accessible PC-space each. Figure 3 shows representative snapshots of the R, B1 and B2 conformations.

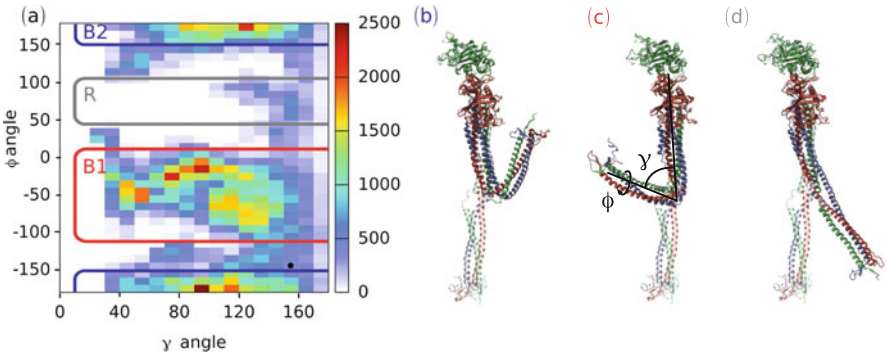


Fig. 3 Bending modes of the fibrinogen protomer: (**a**) Distribution of bending and torsion angles γ and ϕ. γ is the angle between the center of mass of the Cα atoms of the D-domain (β200–458, γ140–394), the hinge region (α99–110, β130–55, γ70–100) and the E-domain(α50–58, β82–90, γ23–31). For ϕ, the centers of mass of the Aα and γ chain or the Bβ and γ chain of the E-domain are used separately to create four reference points together with the centers of mass of the hinge region and the D-domain. The black dot marks the position of the crystallographic structure in angle space. (**b–d**) Representative snapshots showing bending in B1, B2 and R conformations

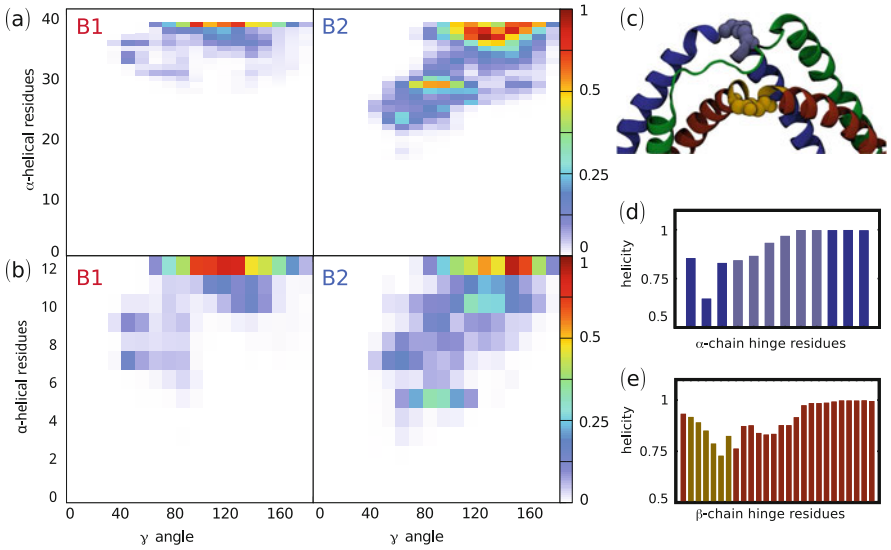


Fig. 4 α-helical content of the hinge region. (**a**) Distribution of the bending angle γ and the number of α-helical residues in the hinge region (α99–110, β130–155) of the Aα and Bβ chains for B1 and B2 conformations. (**b**) Distribution of the bending angle γ and the number of α-helical residues in the plasmin cleavage sites (α102–107, β131–136) for B1 and B2 conformations. (**c**) Bent B1 conformation, showing exposed plasmin cleavage sites. (**d–e**) Frequency of α-helical secondary structure by residue for the Aα and Bβ chain, respectively. Plasmin cleavage sites highlighted

Figure 4a shows the relation between the number of α-helical residues in the hinge and the bending angle. In bent B1 conformations (bending angle below 120°) the Aα and Bβ chains lost 2.4 of their α-helical residues in the hinge region, while it was 7.6 residues in B2 conformations. In R conformations, bending was heavily suppressed especially, never reaching low bending angles. Even at bending angles between 100° and 120° 10 residues lost their α-helical structure in the Bβ chain alone. Generally, the non α-helical residues formed hydrogen-bonded turns, bends or loops.

Overall, the simulations showed that bending occurs along two preferential directions B1 and B2, while it is heavily restricted otherwise. This behavior is related to the stability of the helical segments of the Aα and Bβ chain in the coiled-coil region. Bending in B1 conformations only required small non-helical turns in the hinge region of the Aα and Bβ chains. Bending in B2 conformations on the other hand relied on significantly larger regions of flexible, non-helical residues in the Aα and Bβ chains. In R conformations the Bβ chain heavily restricted bending, even slight bending required large losses of α-helical structure.

Apart from an impact on overall flexibility of Fg and fibrin fibers, bending in the hinge region was also relevant for the exposure of plasmin cleavage sites, as previously hypothesized [16]. The two cleavage sites at the residues α103–104 and

β133–134 are part of the hinge region of the Aα and Bβ chain, respectively. Previous simulations of Fg showed the correlation between strong bending and non-helical secondary structure around the cleavage sites [1].

This result could be confirmed here for both the B1 and B2 bending modes. While the B1 bending generally required only a few non-helical residues in the Aα and Bβ chains (Fig. 4a), they were predominantly located at the plasmin cleavage sites. The frequency of α-helical secondary structure for the residues in the hinge region presented in Fig. 4d–e showed that the residues around the plasmin cleavage sites lost helicity more frequently than other parts of the hinge region. This means that bending of Fg in either B1 or B2 conformations resulted in an increased exposure of the plasmin cleavage sites to the surrounding environment.

3 Interactions Between Proteins and Poly(Ethylene-Glycol) Investigated Using Molecular Dynamics Simulations

Poly(ethylene-glycol) (PEG) is a non-toxic polymer highly soluble both in water and non-polar solvents and finds application in a broad variety of contexts, including for example, as dispersant in toothpastes, lubricant in skin creams and anti-foaming agent in the food industry. In biomedical applications, PEG, when attached to a therapeutic molecule or nanoparticle (PEGylation), confers them extra resistance to degradation, extended blood clearance time and, ultimately, improved therapeutic efficacy. These improvements are obtained by a combination of factors like increased overall volume, which prevents renal clearance, but also a reduction of unspecific interactions with the surrounding environment. This last phenomenon goes under the name of stealth effect, and results in the ability of PEG to screen the PEGylated species from the immune response of the host organism.

Although PEG helps reduce unspecific interactions with the surrounding biological milieu, it does not remove them completely. It has been shown that, also due the constant exposure to PEG-containing products, the organism can produce antibodies against PEG. In addition, many different proteins adsorb on PEG-coated nanoparticles, as soon as these enter in contact with blood [17]. A molecular understanding of these protein adsorption phenomena on PEG-coated surfaces and more in general of the interactions between PEG and proteins is still missing.

Atomistic molecular dynamics simulations can help to fill this gap. They have been successfully used to address protein adsorption on a large variety of surfaces [18, 19], as well as the behavior of specific blood proteins on model surfaces [1, 2], While accurate atomistic models and force field already exist for many metallic, metal-oxides and some polymeric surfaces [19], the surface exposed by PEGylated species poses extra challenges due to its brushy nature. These challenges have been recently addressed by simulating the proteins of interest in a mixture of water and PEG molecules [20, 21]. With this approach, which is similar in spirit to what done in computational drug screening exercises [22, 23], the possible binding modes of

PEG with the protein surface are sampled accurately and their relative importance is determined from the simulations.

Here, we report the results obtained with these simulations, which were performed mostly on Hazelhen and have been published in ref. [20] and in ref. [24] where we have also determined preferential binding coefficients of several blood proteins for PEG.

3.1 Methods

All the simulations were carried out using the program NAMD [25], specifically compiled for the Cray XC40 architecture, and the charmm27 force field [26] with the extension for PEG [27]. Tip3p [7] was used as model for the explicit treatment of water. An integration time step of 1 fs was used across the simulations. Simulations were carried out using periodic boundary conditions. 1 atm Pressure and 300 K temperature were maintained constant during the simulations using the Langevin piston algorithm and Langevin thermostat [28, 29]. A cutoff of 1.2 nm was used for the non-bonded interactions with a switch function. Long range electrostatic interactions were treated using the smooth particle mesh Ewald (PME) method [30] with a grid spacing of about 0.1 nm. To prepare the PEG-water mixture, 64 PEG molecules ($H-[O-CH_2-CH_2]_n-OH$, with n either 4 or 7) were placed on a $4 \times 4 \times 4$ grid with 1.0 nm spacing between grid points. Then 1 ns high temperature (700 K) simulations in vacuum with damped electrostatics interactions (dielectric constant 200) were run to randomize the initial dihedral distribution of the PEG molecules. The PEG molecules were then immersed in a box of water molecules and sodium and chlorine ions were added to reach physiological concentration (0.15 M). Mixtures with different concentrations of PEG were obtained by changing the size of the water box surrounding the PEG molecules. The prepared mixtures were then equilibrated first at high temperature (373 K) for 1.0 ns and then at 300 K for 1.0 ns. Several plasma proteins have been considered: human serum albumin(HSA), bovine serum albumin(BSA), transferrin(TF) and apolipoprotein A1 (ApoA1). The initial coordinates of the proteins were taken from the Protein Data Bank(PDB) (see Table 1 for the list of PDBids). Each protein was immersed in a box filled by replicating the coordinates of the PEG-water mixtures obtained before in the three space directions and removing mixture atoms in close contact with protein atoms. The final PEG concentration in the simulation boxes is reported in Table 1. The size of the boxes was large enough to leave at least 1.0 /nm from each protein atom and the box boundary. In the case of HSA, larger box sizes with 1.5 and 2.0 nm distances between protein and box boundary were simulated to investigate the dependence of the simulation results upon box size. The total charge of the systems was further neutralized by changing an adequate number of water molecules into ions. The complete systems were then minimized using the steepest descent algorithm for 10,000 steps with harmonic restraints on the heavy atoms of the proteins. Then the systems were equilibrated at room temperature and pressure for 1.0 ns during which

Table 1 List of the simulations performed

System	Box size (Å)	N. Atoms	PEG length	[PEG] (g/ml)	Simulation time (ns)
HSA+lipids	103	114,000	4	0.13	4 × 105
HSA	98.8	100,881	4	0.08	4 × 100
	98.2	99,301	4	0.11	4 × 100
	108.6	134,134	4	0.12	4 × 125
	118.2	172,541	4	0.12	5 × 100
BSA	103.6	116,489	4	0.11	4 × 100
Transferrin	108	132,009	4	0.11	5 × 200
	108.7	134,663	4	0.07	5 × 200
	108.4	133,303	7	0.07	5 × 200
	108.8	134,372	7	0.04	5 × 200
ApoA1	192	739,653	4	0.13	4 × ≈ 100

the harmonic restraints were gradually removed and for 1.0 ns without restraints. Finally production runs were started with 4 or 5 replicas for each system. Length of the runs ranged from 100 to 200 ns (see Table 1). The simulations have been run using NAMD an MPI-based molecular dynamics simulation engine which has been specifically compiled for the Cray XC40 architecture of Hazelhen. The large number of atoms included in the simulations (10^5–10^6) assure good scaling up to thousands cores. Typical job sizes of 100 nodes have been used for the simulations. Typical job length of 30 min–3 h were submitted. Depending on system size this has allowed to collect from 1 to 8 ns/h/100nodes.

The direct PEG-protein interactions along the simulations were measured using the NAMD pair-interaction utility. The time series along the trajectories of the number of PEG and water heavy atoms found within a distance d of the protein or of each amino acid type was determined using the "pbwithin" selection command of VMD. The ratio of these numbers was compared to the ratio between all the PEG and water heavy atoms in the simulation box (bulk ratio). A PEG/water ratio larger than bulk for an amino acid implies the presence of an effective attractive interaction between PEG and the amino acid. On the other hand, amino acids with a PEG/Water ratio smaller than bulk exert and effective repulsion for PEG. For the residue specific PEG/Water ratio calculations we used $d = 5.0$ Å. The preferential binding coefficient measures the excess of PEG found in the local environment of the protein, i.e., in its vicinity (Fig. 5) and was defined as:

$$\Gamma = [PEG]_L - [H_2O]_L \frac{[PEG]_B}{[H_2O]_B} \tag{1}$$

where $[\cdot]_L$ is the concentration of PEG or water in the vicinity of the protein and $[\cdot]_B$ the concentration in the bulk, that is, away from the protein. The local environment of the protein is defined as a shell of thickness d around the protein. Γ converges for $d \geq 9.0$ Å, which is the value we used for these calculations.

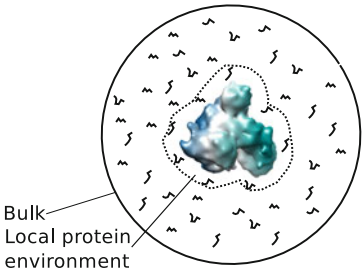


Fig. 5 The concentration of PEG in the local environment of the protein and in the bulk solution are used for the measure of the preferential binding coefficient Eq. (1)

3.2 Results and Discussion

Four different plasma proteins have been simulated immersed in a PEG/water mixture, as reported in Sect. 3.1. Three of them (HSA, BSA, TF) are relatively abundant in plasma but not on the surface of PEGylated nanoparticles and one (ApoA1) is abundant on the nanoparticle [17]. The simulations were performed at different PEG concentrations to verify a possible dependence of the results on this parameter. Additional simulations explored the role of PEG length and box size. Each simulation was repeated 4–5 times to improve the statistical accuracy of the results. Only short PEG molecules were considered (4 or 7 monomers), which diffuse fast and allow for a thorough exploration of the possible binding modes with the protein. This insures that the observed quantities converge in the accessible simulation times (100–200 ns). It is worth noting that, since the persistence length (3.8 Å) of PEG is comparable with the distance between monomers, longer PEG chains are expected to explore the same binding modes as the shorter chains simulated here. The simulations show that the protein structures are not affected by the presence of PEG. The average root mean square deviation with respect to the starting structure or, in the case of ApoA1 to a control run in pure water, is in agreement with native state fluctuations of proteins of similar size (3–5 Å for HSA, BSA and TF).

PEG molecules in the simulations reach and leave the surface of the protein continuously, in a rather dynamical equilibrium. However, some regions of the surface are more frequently populated by PEG (Fig. 6).

We, then, analyzed in more detail the possible origin of the observed uneven distribution of PEG on the protein surface. We hypothesized that each amino-acid type may have a different affinity for PEG molecules. To test the hypothesis, we measured the average ratio between the number of PEG and water heavy atoms (PEG/Water ratio) in a 5.0 Å shell surrounding each amino acid (Fig. 7). The data show that some amino acids are more likely to have a concentration of PEG higher than bulk, while others have a smaller concentration. In other words, some amino acids exert an effective attraction for PEG while other exert a repulsion. The PEG-amino acid interaction pattern look also very similar across the different simulations. Indeed, we verified that the residue-specific PEG/Water ratio, normalized to the bulk value in each simulation, are highly correlated across the simulations, at least for

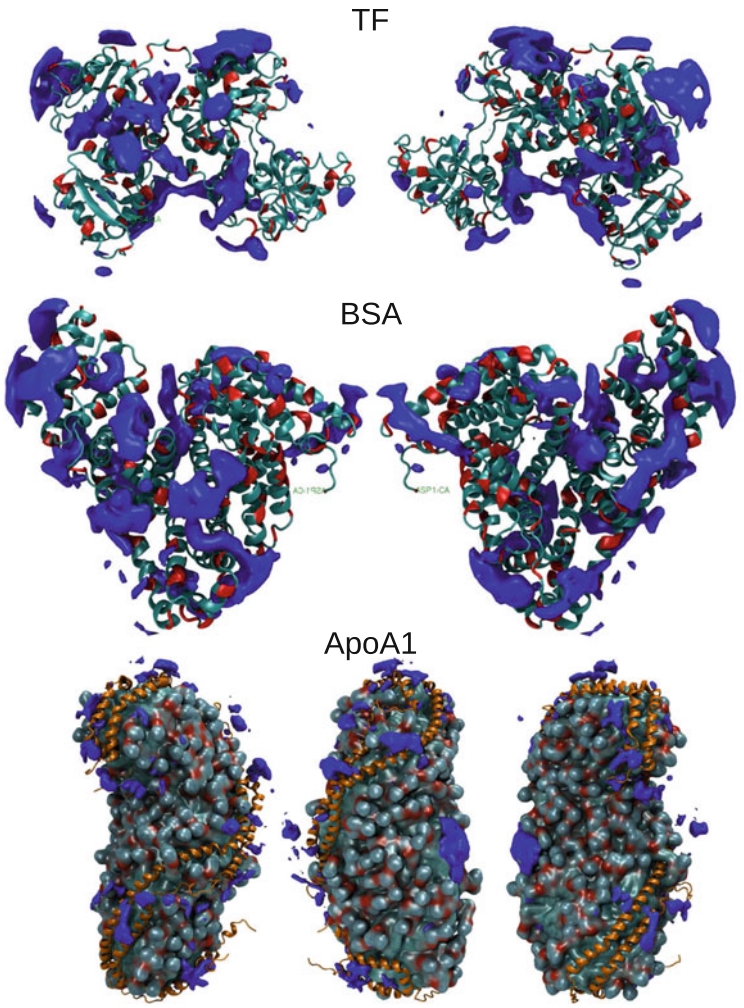


Fig. 6 Cartoons of three different simulated proteins (cyan) where regions with PEG density larger than twice as in the bulk have been highlighted (dark blue). Left and right and center pictures represent different views of the proteins. Negatively charged residues are highlighted in red. In the case of ApoA1, the lipids present in the apolipoprotein are also rendered as a continuous surface

the most exposed amino acids (Fig. 8). These information, obtained from similar simulations, allowed us to describe PEG-protein interactions with a simplified model based on the solvent accessibility of each amino acid on the protein surface [20].

A further analysis of the data consists in the measurements of the preferential binding coefficient of each protein for PEG. Preferential binding coefficients are derived from the Kirkwood-Buff integrals of the pair correlation functions in a

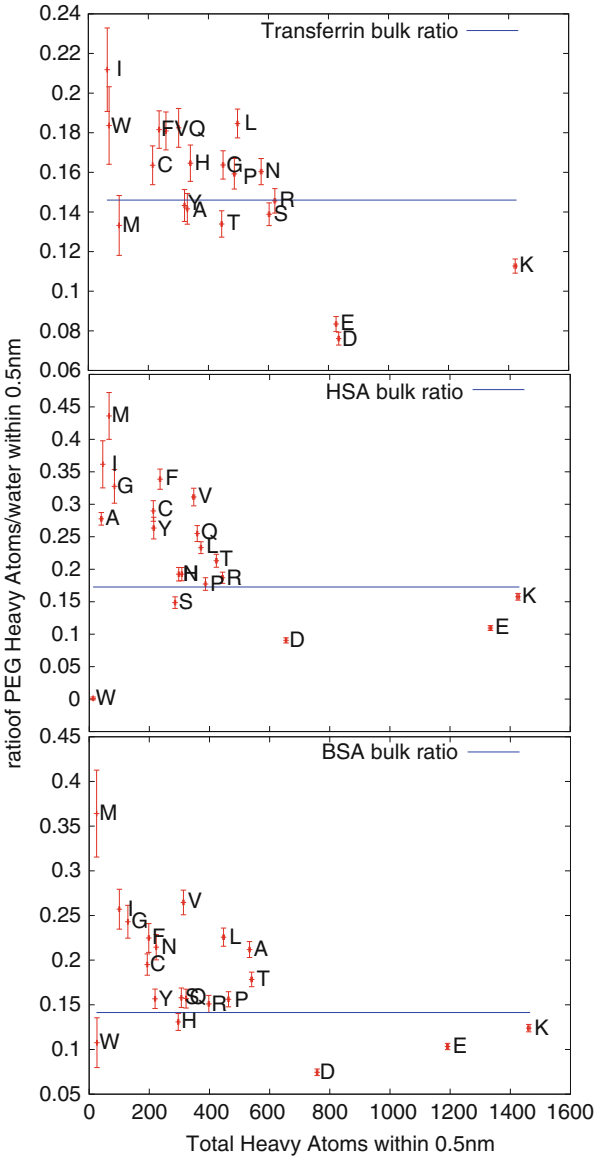


Fig. 7 Ratio between PEG heavy atoms and water molecules within 5 Å of each residue types vs sum of PEG+water (indicating the degree of exposure of each residues type). Negatively charged residues "repel" PEG

multi-component system [31]. In particular they measure the excess number of cosolvent molecules in the vicinity of each protein molecule, thus they are positive in case of a preferential binding of the cosolvent to the protein, and negative otherwise. Preferential binding coefficient are thermodynamic parameters that can

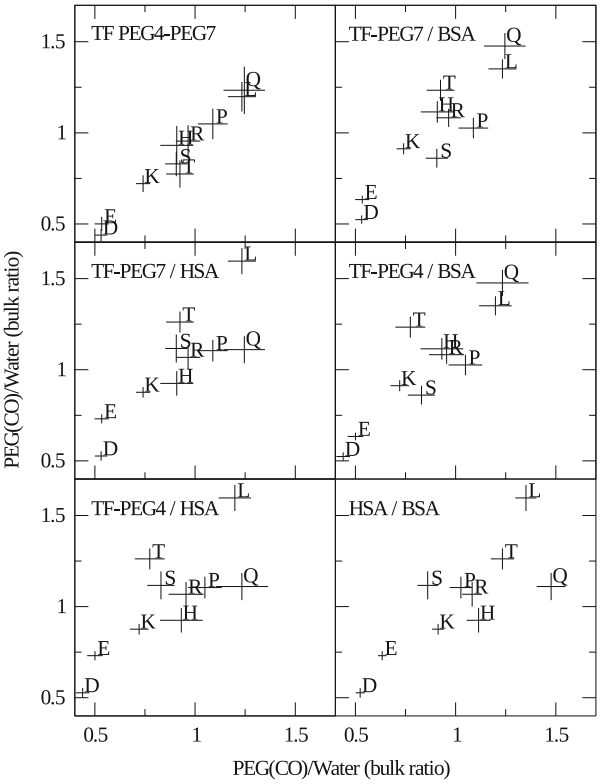


Fig. 8 Correlation plots of the observed residue specific PEG/water ratios between the various simulations. Correlation coefficients all above 0.8 indicate the substantial independence of the values from the parameters of the simulations

be accessed experimentally through high precision densitometry experiments [32], differential scanning calorimetry [33] and vapor pressure osmometry [34]. In the present case, they report the affinity of each protein for the cosolvent (PEG) *versus* the solvent (water) and can be approximated using Eq. (1) [35]. The measurements we have performed on the simulations (Table 2) show in most cases positive coefficients very close to zero, indicating a very weak preference of the analyzed proteins for PEG.

Concluding, we show how atomistic molecular dynamics simulations can be used fruitfully to assess the interaction of plasma proteins with PEG. The simulations reveal that the interaction patterns with surface amino acids are similar across the different proteins and the different simulation conditions analyzed here. This fact has allowed for building simplified models of protein interactions with PEG [20]. The simulations provide also direct measurements of the preferential binding coefficients which could be assessed experimentally to test and help improve the available force fields. The approach used in the present work is rather general and

Table 2 Preferential binding coefficients measured along the simulations

System	Box size (Å)	PEG length	[PEG] (g/ml)	Γ (n.mol.)
HSA+lipids	103	4	0.13	3.4 (1.2)
HSA	98.8	4	0.08	5.0 (0.8)
	98.2	4	0.11	7.0 (1.3)
	108.6	4	0.12	2.9 (0.9)
	118.2	4	0.12	−0.1 (1.1)
BSA	103.6	4	0.11	0.0 (1.0)
Transferrin	108	4	0.11	4.8 (0.9)
	108.7	4	0.07	7.5 (1.0)
	108.4	7	0.07	4.7 (0.7)
	108.8	7	0.04	2.9 (0.7)
ApoA1	192	4	0.13	−1.2 (1.6)

could be extended to evaluate interactions of proteins with other polymers that, like PEG, are highly soluble in waters and do not induce large conformational changes to proteins. Indeed, large efforts are being made to identify polymers that could replace PEG in biomedical applications and solve the issues related to its possible long-term toxicity and nonbiodegradability. We expect the techniques developed here to help in that direction.

Acknowledgements TS gratefully acknowledges financial support from the Graduate School Materials Science in Mainz. GS gratefully acknowledges financial support from the Max-Planck Graduate Center with the University of Mainz. We gratefully acknowledge support with computing time from the HPC facility Hazelhen at the High performance computing center Stuttgart and the HPC facility Mogon at the university of Mainz. This work was supported by the German Science Foundation within SFB 1066 project Q1.

References

1. S. Köhler, F. Schmid, G. Settanni, PLoS Comput. Biol. **11**(9), 1 (2015)
2. S. Köhler, F. Schmid, G. Settanni, Langmuir **31**(48), 13180 (2015)
3. J. Kollman, L. Pandi, M. Sawaya, M. Riley, R. Doolittle, Biochem **48**(18), 3877 (2009)
4. G. Marguerie, E. Plow, T. Edgington, J. Biol. Chem. **254**(12), 5357 (1979)
5. A. Laio, M. Parrinello, Proc. Natl. Acad. Sci. USA **99**(20), 12562 (2002)
6. A.E. García, Phys. Rev. Lett. **68**, 2696 (1992)
7. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. **79**(2), 926 (1983)
8. W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. **14**, 33 (1996)
9. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, J. Chem. Theory Comput. **4**(3), 435 (2008)
10. G. Bussi, D. Donadio, M. Parrinello, J. Chem. Phys. **126**(1), 014101 (2007)
11. M. Parrinello, A. Rahman, J. Appl. Phys. **52**(12), 7182 (1981)
12. R.B. Best, X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, M. Feig, A.D. Mackerell Jr., J. Chem. Theory Comput. **8**(9), 3257 (2012)

13. M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, SoftwareX **1–2**, 19 (2015)
14. G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, Comput. Phys. Commun. **185**(2), 604 (2014)
15. M.J. Ferrarotti, S. Bottaro, A. Párez-Villa, G. Bussi, J. Chem. Theory Comput. **11**(1), 139 (2015)
16. R.F. Doolittle, D.M. Goldbaum, L.R. Doolittle, J. Mol. Biol. **120**(2), 311 (1978)
17. S. Schöttler, G. Becker, S. Winzen, T. Steinbach, K. Mohr, K. Landfester, V. Mailänder, F.R. Wurm, Nat. Nanotechnol. **11**, 372–377 (2016)
18. H. Heinz, H. Ramezani-Dakhel, Chem. Soc. Rev. **45**(2), 412 (2016)
19. M. Ozboyaci, D.B. Kokh, S. Corni, R.C. Wade, Q. Rev. Biophys. **49**, e4 (2016)
20. G. Settanni, J. Zhou, T. Suo, S. Schöttler, K. Landfester, F. Schmid, V. Mailänder, Nanoscale **9**(6), 2138 (2017)
21. Q. Shao, Y. He, A.D. White, S. Jiang, J. Chem. Phys. **136**(22), 225101 (2012)
22. N. Basse, J.L. Kaar, G. Settanni, A.C. Joerger, T.J. Rutherford, A.R. Fersht, Chem. Biol. **17**(1), 46 (2010)
23. J. Seco, F.J. Luque, X. Barril, J. Med. Chem. **52**(8), 2363 (2009)
24. G. Settanni, J. Zhou, F. Schmid, CSP2017 (accepted)
25. J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, J. Comput. Chem. **26**, 1781 (2005)
26. A.D. Mackerell, M. Feig, C.L. Brooks, J. Comput. Chem. **25**(11), 1400 (2004)
27. H. Lee, R.M. Venable, A.D. Mackerell, R.W. Pastor, Biophys. J. **95**(4), 1590 (2008)
28. G.J. Martyna, D.J. Tobias, M.L. Klein, J. Chem. Phys. **101**(5), 4177 (1994)
29. S.E. Feller, Y. Zhang, R.W. Pastor, B.R. Brooks, J. Chem. Phys. **103**(11), 4613 (1995)
30. U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, J. Chem. Phys. **103**(19), 8577 (1995)
31. J.G. Kirkwood, F.P. Buff, J. Chem. Phys. **19**(6), 774 (1951)
32. J.C. Lee, S.N. Timasheff, J. Biol. Chem. **256**(14), 7193 (1981)
33. N. Poklar, N. Petrovčič, M. Oblak, G. Vesnaver, Protein. Sci. **8**(4), 832 (1999)
34. E.S. Courtenay, M.W. Capp, C.F. Anderson, M.T. Record, Biochemistry **39**(15), 4455 (2000)
35. B.M. Baynes, B.L. Trout, J. Phys. Chem. B **107**(50), 14058 (2003)

Cosmic Large-Scale Structure in the IllustrisTNG Simulations

Volker Springel, Annalisa Pillepich, Rainer Weinberger, Rüdiger Pakmor, Lars Hernquist, Dylan Nelson, Shy Genel, Mark Vogelsberger, Federico Marinacci, Jill Naiman, and Paul Torrey

Abstract We have finished two new, extremely large hydrodynamical simulations of galaxy formation that significantly advance the state of the art in cosmology. Together with accompanying dark matter only runs, we call them 'IllustrisTNG', the next generation Illustris simulations. Our largest and most ambitious calculation follows a cosmological volume 300 megaparsecs on a side and self-consistently solves the equations of magnetohydrodynamics and self-gravity coupled to the fundamental physical processes driving galaxy formation. We have employed AREPO, a sophisticated moving-mesh code developed by our team over the past 7 years and

V. Springel (✉)
Zentrum für Astronomie der Universität Heidelberg, Astronomisches Recheninstitut, Mönchhofstr. 12-14, 69120 Heidelberg, Germany

Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany
e-mail: volker.springel@h-its.org

A. Pillepich
Max-Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
e-mail: pillepich@mpia-hd.mpg.de

R. Weinberger • R. Pakmor
Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany
e-mail: rainer.weinberger@h-its.org; ruediger.pakmor@h-its.org

L. Hernquist • J. Naiman
Center for Astrophysics, Harvard University, 60 Garden Street, Cambridge, MA 02138, USA
e-mail: lars.hernquist@cfa.harvard.edu; jill.naiman@cfa.harvard.edu

D. Nelson
Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, 85740 Garching, Germany
e-mail: dnelson@mpa-garching.mpg.de

S. Genel
Department of Astronomy, Columbia University, 550 W. 120th St., New York, NY 10027, USA
e-mail: shygenelastro@gmail.com

M. Vogelsberger • F. Marinacci • P. Torrey
Kavli Institute for Astrophysics and Space Research, MIT, Cambridge, MA 02139, USA
e-mail: mvogelsb@mit.edu; fmarinac@mit.edu; ptorrey@mit.edu

W.E. Nagel et al. (eds.), *High Performance Computing in Science and Engineering '17*, https://doi.org/10.1007/978-3-319-68394-2_2

equipped with an improved, multi-purpose galaxy formation physics model. The simulated universe contains tens of thousands of galaxies encompassing a variety of environments, mass scales and evolutionary stages. The groundbreaking volume of TNG enables us to sample statistically significant sets of rare astrophysical objects like rich galaxy clusters, and to study galaxy formation and the spatial clustering of matter over a very large range of spatial scales. Here we report some early results on the matter and galaxy clustering found in the simulations. The two-point galaxy correlation function of our largest simulation agrees extremely well with the best available observational constraints from the Sloan Digital Sky Survey, both as a function of galaxy stellar mass and color. The predicted impact of baryonic physics on the matter power spectrum is sizeable and needs to be taken into account in precision studies of cosmology. Interestingly, this impact appears to be fairly robust to the details of the modelling of supermassive black holes, provided this reproduces the scaling properties of the intracluster medium of galaxy clusters.

1 Introduction

Observed galaxies range in mass from a few thousand to a few trillion times the mass of the sun, encompass physical sizes from a fraction to tens of kilo-parsecs, and span a variety of morphologies. Galaxies can reside in diverse environments—in isolation, or as members of rich groups and clusters. They are self-gravitating systems of stars and gas embedded in a halo of dark matter, and their distribution throughout space traces a 'cosmic web' defined by filaments, nodes, sheets, and voids of matter. The highly clustered large-scale structure of the Universe today, at mega-parsec and giga-parsec scales, arose from 13.8 billion years of evolution, starting from the nearly homogeneous distribution of matter in the early universe.

According to the current cosmological paradigm this large-scale structure emerges through the dominant presence of cold dark matter (CDM), which in turn fuels individual galaxies with cosmic gas, imparts gravitational torques and tides, and determines a bottom-up or 'hierarchical' growth, with smaller objects collapsing earlier by gravitational-instability and then later merging to form progressively more massive and rare systems. The cosmological environment and the resulting hierarchical growth govern the formation and transformation of galaxies through cosmic time, regulating their stellar content, star formation activity, gas and heavy-element composition, morphological structure, and so on. In order to gain a theoretical, ab-initio understanding of these processes and their role in shaping galaxies within the full cosmological context, simulations that account for the multi-scale physics involved are the tool of choice.

As a starting point, such calculations rely on initial conditions that are *known* and well constrained by observational data from the Cosmic Microwave Background radiation, now measured to exquisite precision by the Planck mission. The calculations then need to accurately compute the dominant physical force, namely gravity, acting upon all matter, within the accelerating expansion of the Universe.

Furthermore, (magneto)hydrodynamical processes for modelling the evolution of the gaseous component of the Universe need to be followed. Finally, one must account for all other relevant astrophysical processes: from the atomic level interactions that govern radiative cooling of a metal-enriched gas, to the formation of stars and supermassive black holes, with their subsequent expulsion of mass, metals, and 'feedback' energy, which can impact scales as large as entire galactic halos.

The mathematical equations governing these physical processes are non-linear, complex, and highly coupled. This makes their solution through numerical simulation the only practical approach, although the inherently multi-scale, multi-physics nature of the problem poses significant challenges. Nevertheless, rapid progress has been made. In the last 10 years, gravity-only simulations such as the Millennium Run [1], which model only the dark matter component, have reached an extraordinary level of size and sophistication—the largest such simulations now routinely include over one trillion resolution elements [2, 3], encompassing volumes 6–10 giga-parsecs on a side.

Simulations including gas have proven significantly more difficult. Yet, in the past few years projects such as Eagle [4], Horizon-AGN [5], and our Illustris simulation [6–8] have demonstrated that hydrodynamical simulations of structure formation at kilo-parsec spatial resolution can reasonably reproduce the basic properties and scaling relations of observed galaxies. Together with simulations of individual galaxies at even higher resolutions [9–13], these calculations have provided unparalleled insights into the processes underlying galaxy formation. Simultaneously, they have been calling for further improvements of the cosmological hydrodynamical simulations techniques.

In particular, the limited volume covered by simulations like Illustris—encompassing about 100 mega-parsecs on a side—have hindered the exploration of astrophysical processes which act on the largest scales of the Universe and precluded the study of very rare objects, like rich clusters of galaxies and the supermassive black holes sitting at their centers. Because of the enormous computational cost required to simulate the cosmological formation of very massive structures or very large cosmological volumes, at present just a handful of hydrodynamical simulations exist which are capable of accessing the realm of *statistical* samples of massive clusters [14–18]. Yet, none of them has the resolution needed to unveil the structural details of clusters, their intra-cluster plasma and the galaxies hosted therein. The latest generation of supercomputers and cosmological codes make it now possible to attain the simulated volumes necessary to study the regime of rare objects, yet simultaneously achieving mass and spatial resolutions suitable for galaxy formation applications.

Here we report on first results from our team's efforts in this direction, carried out with the computer time granted under our GCS allocation GCS-ILLU-44057 on the HazelHen supercomputer at the HLRS. We will first give a description of the large simulations we have carried out, and then briefly discuss in an exemplary fashion some first results on cosmic large-scale structure. We note, however, that the letter is only a small subset of a large number of new scientific results that we obtain from these simulations at the time of writing of this report.

2 TNG: The Next Generation Illustris Simulations

2.1 The Need to Go Beyond the State of the Art

A few years ago, our team has performed one of the first high-resolution, large-scale hydrodynamic simulations of galaxy formation [6–8]. This simulation suite—"Illustris"—used a novel and innovative numerical approach, employing a moving, unstructured mesh as implemented in the AREPO code [19]. One of the major achievements of Illustris was its ability to follow the small-scale evolution of gas and stars within a representative portion of the Universe, yielding a population of thousands of well-resolved elliptical and spiral galaxies. For the first time ever, it reproduced the observed morphological mix of galaxies and its dependence on stellar mass. It could also explain at the same time the characteristics of hydrogen on large scales, and the metal and neutral hydrogen content of galaxies on small scales.

In practice, all of the dominant galaxy formation processes needed to simulate galaxies over a broad mass range were identified, implemented, and tested in a full cosmological setting. These include primordial and metal line cooling of a gas with self-shielding from radiation; a time-varying, spatially uniform ultraviolet (UV) background; the evolution of stars and stellar populations, their energetic feedback, as well as mass and metal return from Supernova Type Ia, Type II and AGB stars; the formation, growth, and merging of supermassive black holes; models for distinct 'quasar', 'radio', and radiative feedback from active galactic nuclei (AGN). The Illustris stellar evolution and chemical enrichment network followed nine elements simultaneously (H, He, C, N, O, Ne, Mg, Si, Fe), while the stellar feedback was realised through stochastically driven, galactic-scale winds [20, 21].

While the Illustris model has been remarkably successful in reproducing a wide-range of galaxy properties, particularly for L^* galaxies, some of the results in Illustris were found to be in tension with a number of observable constraints [22]. In particular, the predicted central gas fractions of galaxy groups and clusters were clearly too low [8]. At the same time, the colours of massive galaxies were too blue, and their stellar masses too high when compared to observations. Similar tensions existed with the stellar mass content as well as the sizes of the smallest resolved galaxies [7, 23]. These important discrepancies could be traced to the Illustris feedback model from active galactic nuclei (AGN), pointing to deficits in the physics model for supermassive black hole accretion and its associated energy release.

2.2 New Modelling Techniques

In the last 3 years we have therefore undertaken a campaign of model development and improvement, in order to address this key deficiency and other lingering