Eur Biophys J (2012) 41:647–661
DOI 10.1007/s00249-012-0837-1

ORIGINAL PAPER

EBSA
Biophysics in Europe

Solvating atomic level fine-grained proteins in supra-molecular level coarse-grained water for molecular dynamics simulations

Sereina Riniker · Andreas P. Eichenberger · Wilfred F. van Gunsteren

Received: 9 May 2012 / Revised: 18 June 2012 / Accepted: 25 June 2012 / Published online: 14 July 2012

Abstract Simulation of the dynamics of a protein in aqueous solution using an atomic model for both the protein and the many water molecules is still computationally extremely demanding considering the time scale of protein motions. The use of supra-atomic or supra-molecular coarse-grained (CG) models may enhance the computational efficiency, but inevitably at the cost of reduced accuracy. Coarse-graining solvent degrees of freedom is likely to yield a favourable balance between reduced accuracy and enhanced computational speed. Here, the use of a supra-molecular coarse-grained water model that largely preserves the thermodynamic and dielectric properties of atomic level fine-grained (FG) water in molecular dynamics simulations of an atomic model for four proteins is investigated. The results of using an FG, a CG, an implicit, or a vacuum solvent environment of the four proteins are compared, and for hen egg-white lysozyme a comparison to NMR data is made. The mixed-grained simulations do not show large differences compared to the FG atomic level simulations, apart from an increased tendency to form hydrogen bonds between long side chains, which is due to the reduced ability of the supra-molecular CG beads that represent five FG water molecules to make solvent-protein hydrogen bonds. But, the mixed-grained simulations are at least an order of magnitude faster than the atomic level ones.

Keywords Coarse-graining · Molecular dynamics · Protein · Water · Multigraining

S. Riniker · A. P. Eichenberger · W. F. van Gunsteren (✉)
Laboratory of Physical Chemistry, ETH Zurich,
8093 Zurich, Switzerland
e-mail: wfvgn@igc.phys.chem.ethz.ch

Introduction

To describe the dynamic processes of biomolecules such as proteins and peptides, atomic level molecular dynamics (MD) simulations have been proven to be a useful method. Although the increase of computer power has given access to larger and larger system sizes and time scales, MD simulations are generally still limited to the multi-nanometer and multi-nanosecond range nowadays. As most biological processes occur on larger time scales, the development of models at a more coarse-grained (CG) level of resolution to reduce the computational effort is of considerable interest. In the process of coarse graining, a chosen number of atoms is subsumed into a single CG bead, thus reducing the number of particle-particle interactions. In addition, the generally smoother interaction function between CG beads facilitates the sampling of configurations. However, the choice of the degrees of freedom to be eliminated in the process of coarse-graining is not trivial and depends largely on the questions that should be answered using the CG model, as coarse graining involves per se a loss of information. Generally, it is advantageous to select degrees of freedom that account for a large part of the computational cost. Coarse graining of these provides considerable speed-up of the simulation, while these may be of less interest for the property or biomolecular process to be investigated. To this category we count the solvent degrees of freedom and, in the case of membrane proteins, also intra-molecular degrees of freedom of lipids, whereas proteins usually present only a small part of the system and a detailed, i.e. atomistic, description is often essential for a correct representation of their structural and dynamic properties. Thus, combining an atomistic, fine-grained (FG) protein with a CG solvent may be a promising approach to simultaneously reduce the

computational cost of a simulation and retain the details for the region of interest, i.e. the protein. A variety of multi-graining schemes has been proposed recently [see Ref. Riniker and van Gunsteren (2012) for an overview]. Here, we focus on so-called hybrid systems of FG and CG particles where the FG and CG levels of modelling are combined analogous to hybrid quantum/classical (QM/MM) models (Neri et al. 2005; Shi et al. 2006; Michel et al. 2008; Masella et al. 2008, 2011; Orsi et al. 2009; Rzepiela et al. 2011). Thereby, molecules of different resolution, i.e. FG or CG, coexist in the system and the resolution of a molecule is fixed. The proposed hybrid approaches differ in the central question of how to define the interaction between the FG and CG particles.

Shi et al. (2006) chose to completely reparametrise all FG-CG interactions using the force-matching methodology. They applied their model to study the polypetide gramicidin A forming an α-helix in a CG DMPC membrane solvated in CG water. Thereby, a single water molecule is represented by a CG water bead and in the lipids groups of 3–6 atoms are subsumed into CG beads. The backbone atom-positional RMSD of the polypetide from the NMR solution model structure in the 10 ns mixed-grained simulation agrees with the RMSD of the purely atomistic simulation.

Michel et al. (2008) employ a standard 12-6 Lennard-Jones potential energy function for the CG beads and thus the FG-CG Lennard-Jones interaction parameters can be derived using standard combination rules. Two additional parameters were introduced to calibrate the FG-CG Lennard-Jones and electrostatic interactions to reproduce experimental partition coefficients of amino-acid side-chain analogues. As CG solvent the soft sticky dipole model (Liu and Ichiye 1996) was used, which represents a single water molecule. They also applied their model to study the transmembrane permeability of a series of small organic molecules through a CG DMPC membrane.

Another approach has been proposed by Masella et al. (2008, 2011) where the polarisable pseudo-particle water model (Basdevant et al. 2006) is mixed with a polarisable atomistic solute model. They found that the standard Lennard-Jones potential energy term was not hard enough to prevent a polarisation catastrophe and had to be altered accordingly. Their solvent model represents a single water molecule, where the solvent dipoles only interact with the solute dipoles and not with each other. For the mixed simulations, a model-dependent renormalisation factor had to be introduced for the dipoles of the solute. The model has been applied to study the structural stability of bovine pancreatic trypsin inhibitor (BPTI) (Masella et al. 2008, 2011), as well as G-protein and a ubiquitin-like domain (Masella et al. 2011) by analysing the atom-positional RMSD of the backbone from the X-ray structure and the potential energy terms.

A fourth method avoids the definition of FG-CG interactions altogether by introducing virtual CG sites for groups of FG atoms, where the CG solvent only interacts with these virtual CG sites (Rzepiela et al. 2011). The forces of the CG virtual sites are then mapped back on the underlying FG atoms. The model has been tested on FG butane molecules in CG butane and on FG dialanine peptides in CG water or CG butane. Thereby, four water molecules are represented by a Lennard-Jones bead without electrostatic interactions.

Most of the hybrid models discussed above employ a CG solvent model that represents only a single water molecule, thus limiting the achievable speed-up. Here, we use a recently introduced CG water model (Riniker and van Gunsteren 2011) where five water molecules are subsumed into a CG bead with explicit electrostatic interactions, thus reproducing the dielectric screening properties of water. The application of this model in mixed-grained simulations has been investigated for FG/CG mixtures of liquid water (Riniker and van Gunsteren 2012). Only a minor parametrisation effort was required to reproduce the solvation free energy of water and the solvation free energies of small alkanes in water. Here, we test the generality of this mixed-grained model on four proteins, i.e. hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM), solvated in CG water at room temperature and pressure. The four proteins represent different secondary structure motifs, i.e. all α-helical, all β-strand, and combinations thereof. The structural and dynamic properties are analysed and compared to those obtained from purely atomic level simulations, from simulations using an implicit solvent or in vacuo. In addition, the configurational ensemble of hen egg-white lysozyme is used for a comparison of simulated averages to experimental data for a number of observable quantities.

Methods

Fine-grained and coarse-grained models

The atomistic GROMOS force field (Schmid et al. 2011) 54A7 was used for the proteins, and the simple-point-charge model (Berendsen et al. 1981) (SPC) as atomistic fine-grained solvent. In the coarse-grained solvent model, five SPC water molecules are subsumed into a spherical CG bead, which consists of two interaction sites, named "center water" (CW) and "dipole particle" (DP), representing a polarisable dipole. The two sites are connected by an unconstrained bond with an attractive quartic potential energy function, making the model polarisable. This CG model is described in detail in Ref. Riniker and van Gunsteren (2011).

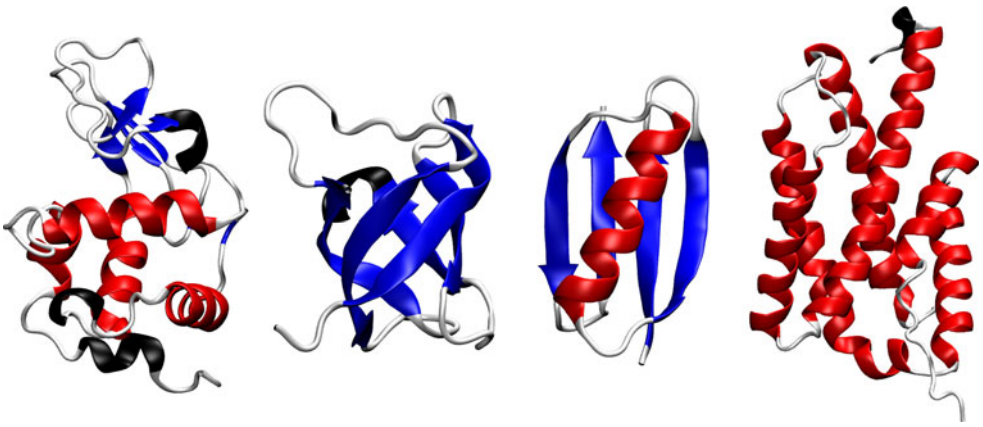

Fig. 1 From *left* to *right*, crystal structures of hen egg-white lysozyme [HEWL, PDB entry 1AKI (Artymiuk et al. 1982)], major cold shock protein [CspA, PDB entry 1MJC (Schindelin et al. 1994)], protein G [GP, PDB entry 1PGB (Gallagher et al. 1994)], and chorismate mutase [CM, PDB entry 2FP2 (Okvist et al. 2006)]. α-helices are shown in *red*, 3_{10}-helices in *black*, and β-strands in *blue*

The Lennard-Jones parameters for the FG-CG interactions are determined using the standard GROMOS combination rules (van Gunsteren et al. 1996). The DP site of the CG bead, which interacts purely electrostatically with other CG beads, has an additional repulsive C_{12} term for the interaction with FG atoms to prevent a polarisation catastrophe. The dielectric permittivity within the cutoff sphere is $\varepsilon_{cs}^{FG} = 1$ for FG-FG electrostatic interactions, $\varepsilon_{cs}^{mix} = 2.3$ for FG-CG interactions, and $\varepsilon_{cs}^{CG} = 2.5$ for CG-CG interactions. The parametrisation of the FG-CG interactions is described in more detail in Ref. Riniker and van Gunsteren (2012).

Proteins

The four proteins were chosen to cover the diverse secondary and tertiary structure elements commonly found, i.e. purely α-helical, purely β-strand, and combinations thereof. The proteins were hen egg-white lysozyme [HEWL, RCSB protein databank (PDB) entry 1AKI (Artymiuk et al. 1982)], major cold shock protein [CspA, PDB entry 1MJC (Schindelin et al. 1994)], protein G [GP, PDB entry 1PGB (Gallagher et al. 1994)], and chorismate mutase [CM, PDB entry 2FP2 (Okvist et al. 2006)] (Fig. 1). The protonation states were chosen to match a pH of 7 and hydrogen atoms were added to the X-ray structures according to standard geometric criteria. The histidine side chains were protonated at N_δ or N_ε depending on their hydrogen-bonding environment. In CM, the first three residues were not considered.

Simulation details

All simulations were performed for 20 ns under NPT conditions using the GROMOS package of programs (Schmid et al. 2012; Kunz et al. 2012).[1] The bond lengths of the protein were constrained to the ideal values applying the SHAKE algorithm (Ryckaert et al. 1977). The temperature was maintained close to its reference value $T = 298$ K by weak coupling to a temperature bath with a relaxation time of 0.1 ps (Berendsen et al. 1984). The pressure was maintained close to its reference value $P = 1.013$ bar (1 atm) by weak coupling to a pressure bath with a relaxation time of 0.5 ps and using the isothermal compressibility $\kappa_T = 4.575 \times 10^{-4}(\text{kJ mol}^{-1}\text{nm}^{-3})^{-1}$. Newton's equations of motion were integrated using the leap-frog scheme (Hockney 1970) with a time step of 2 fs. A reaction field force (Tironi et al. 1995) was applied using the experimental relative dielectric permittivity $\varepsilon_{rf} = 78.5$ (Lide 2007) in the case of an explicit solvent model, or $\varepsilon_{rf} = 1$ for the simulations using an implicit solvent SASA model or vacuum boundary conditions. The contributions to the reaction field by the covalently bound nearest-neighbour atoms that are excluded from the non-bonded van der Waals and Coulomb interaction were considered in all simulations.

Simulations in atomistic water

The four energy-minimized proteins were solvated in a cubic box with a minimum distance to the wall of 1.4 nm, i.e. HEWL in 14,332 water molecules, CspA in 8,128, GP in 7,444, and CM in 20,643, and simulated with periodic boundary conditions. The solvent was energy-minimized with the protein kept fixed by restraining all solute atoms to their positions in the starting structure through a harmonic potential energy term with a force constant of 2.5×10^4 kJ mol^{-1}nm^{-2}. Sodium or chlorine counter ions were added to neutralize the total charge of the protein. Initial velocities were randomly generated from a Maxwell-Boltzmann distribution at 60 K. During the equilibration, the temperature was increased stepwise from 60 to 300 K in steps of 60 K, with a step size of 20 ps. The protein position restraints were simultaneously relaxed by lowering the force constant by a factor 10 at each step. The final configurations

[1] The GROMOS package of programs can be obtained from http://www.gromos.net.

were used as starting structures for the 20-ns production runs. For the non-bonded interactions, a twin cutoff method was used with a short-range cutoff radius of 0.8 nm, an intermediate-range cutoff radius of 1.4 nm, and an update frequency of five time steps for the short-range pairlist and intermediate-range interactions. The O–H and H–H bond lengths in the solvent were constrained to the ideal SPC values applying the SHAKE algorithm (Ryckaert et al. 1977).

Simulations in coarse-grained water

The four proteins were solvated in a cubic box with a minimum distance to the wall of 1.4–1.6 nm, i.e. HEWL in 3,217 CG beads, CspA in 1,659, GP in 1,669, and CM in 3,939, and simulated with periodic boundary conditions. The larger minimum distance to the wall is required by the larger non-bonded cutoff used in the mixed-grained simulations (see below). As starting structures, the equilibrated structures of the proteins from the simulations in FG water were used. The CG water was energy-minimized and subsequently equilibrated in three steps. In the first 100 ps, the protein atoms were positionally restrained with a force constant of 2.5×10^4 kJ mol^{-1} nm^{-2} and the volume was held constant. In the second step, a NPT simulation of 250 ps length was carried out with a position-restraining force constant of 2.5×10^2 kJ mol^{-1} nm^{-2}, and subsequently a 250-ps simulation was performed without any position restraints. For the non-bonded interactions, a twin cutoff method was used with a short-range cutoff radius of 1.4 nm, an intermediate-range cutoff radius of 2.0 nm, and an update frequency of five time steps for the short-range pairlist and intermediate-range interactions.

Simulations in implicit solvent and vacuum

The four proteins were simulated using vacuum boundary conditions and the vacuum GROMOS force field 54B7 (Schmid et al. 2011). As starting structures, energy-minimised X-ray structures were used. The implicit solvent model (Fraternali and van Gunsteren 1996), which describes the force of the solvent on the protein atoms as a function of the solvent accessible surface area (SASA) of the protein, was used without volume correction (Allison et al. 2011). The model is described in detail in Refs. Fraternali and van Gunsteren (1996), Allison et al. (2011). For the non-bonded interactions, a twin cutoff method was used with a short-range cutoff radius of 0.8 nm, an intermediate-range cutoff radius of 1.4 nm, and an update frequency of five time steps for the short-range pairlist and intermediate-range interactions.

Analysis

The GROMOS analysis programs to calculate energies *ene_ana*, atom-positional root-mean-square deviations *rmsd*, atom-positional root-mean-square fluctuations *rmsf*, secondary structure elements *dssp*, hydrogen bonds *hbond*, radii of gyration *rgyr*, torsional-angle transitions *ditrans*, NOE distance-bound violations *prep_noe*, *noe* and *post_noe*, ^3J-coupling values *jval*, residual dipolar couplings *svd_fit*, and ^{1}H–^{15}N order parameters *nhoparam* are described in more detail in Ref. Eichenberger et al. (2011).

RMSD and RMSF

The backbone atom-positional root-mean-square deviation (RMSD) between a configuration and a reference configuration has been calculated based on all backbone non-hydrogen atoms (N, C_α, C, O) using the following formula,

$$\mathrm{RMSD\ (bb)} = \sqrt{\frac{1}{N_{\mathrm{bb}}}\sum_{i=1}^{N_{\mathrm{bb}}}\left(\mathbf{r}_i - \mathbf{r}_{i,\mathrm{ref}}\right)^2}, \tag{1}$$

where N_{bb} is the number of backbone atoms considered, $\mathbf{r}_i$ the position of atom i, and $\mathbf{r}_{i,\mathrm{ref}}$ the position of atom i in the reference configuration. As reference configurations, the X-ray structures of the proteins studied were used (Artymiuk et al. 1982; Schindelin et al. 1994; Gallagher et al. 1994; Okvist et al. 2006).

The atom-positional root-mean-square fluctuations (RMSF) were calculated for the C_α atoms of the backbone and for the last non-hydrogen atom of each side chain [C_β (ALA), $N_{\eta 1}$ (ARG), $N_{\delta 2}$ (ASN), $O_{\delta 1}$ (ASP), S_γ (CYS), $N_{\varepsilon 2}$ (GLN), $O_{\varepsilon 1}$ (GLU), C_α (GLY), $C_{\varepsilon 1}$ (HIS), C_δ (ILE), $C_{\delta 1}$ (LEU), N_ζ (LYS), C_ε (MET), C_ζ (PHE), C_γ (PRO), O_γ (SER), $O_{\gamma 1}$ (THR), $C_{\zeta 2}$ (TRP), O_η (TYR), $C_{\gamma 1}$ (VAL)] using the following formula,

$$\mathrm{RMSF}_i = \sqrt{\frac{1}{N_t}\sum_{t=1}^{N_t}\left(\mathbf{r}_i(t) - \langle \mathbf{r}_i \rangle\right)^2}, \tag{2}$$

where i indicates the C_α or side-chain atom of residue i, $\langle \mathbf{r}_i \rangle$ its average position, and N_t is the number of configurations in the simulation. The RMSF was further averaged over all residues using a mean-square sum,

$$\langle \mathrm{RMSF} \rangle_x = \sqrt{\frac{1}{N_i}\sum_{i=1}^{N_i}\left(\mathrm{RMSF}_{x,i}\right)^2}, \tag{3}$$

where x indicates the backbone (C_α) or side chain atoms (sc), and N_i is the number of residues.

Radius of gyration

The radius of gyration of a protein is a measure of the compactness of a structure and can be related to the light-scattering intensity,

$$R_{\mathrm{gyr}} = \sqrt{\frac{1}{N_a}\sum_{i=1}^{N_a}(\mathbf{r}_i - \mathbf{R}_{\mathrm{com}})^2} \tag{4}$$

with

$$\mathbf{R}_{\mathrm{com}} = \frac{1}{M}\sum_{i=1}^{N_a} m_i \mathbf{r}_i \tag{5}$$

where N_a is the number of protein atoms, $\mathbf{r}_i$ the position of atom i, m_i the mass of atom i, and M the total mass of the protein atoms.

Secondary structure elements

The secondary structure elements of the four proteins were detected and monitored based on the rules of Kabsch and Sander (1983). In order to avoid ambiguous assignment if a residue is part of two different secondary structure elements, the following priority rules were applied: β-strand/β-bridge > α-helix > π-helix > 3_{10}-helix > hydrogen bonded turn > bend.

Hydrogen bonds

The geometric criterion for a hydrogen bond was a minimum donor-hydrogen-acceptor angle of 135° and a maximum hydrogen-acceptor distance of 0.25 nm. The hydrogen bonds were split for each protein into three groups: backbone-backbone (bb-bb), backbone-side chain (bb-sc), and side chain-side chain (sc-sc). During a simulation, the number of configurations in which a hydrogen bond is present, N_{hb}, is counted and the occurrence is given in percentage,

$$\mathrm{occ} = 100 \times \frac{1}{N_f}\sum_{t=1}^{N_f} N_{\mathrm{hb}}(t), \tag{6}$$

where N_f is the total number of configurations in a simulation trajectory. For each of the three groups bb-bb, bb-sc and sc-sc, the occurrence of the corresponding hydrogen bonds was summed up.

Torsional-angle transitions

The 60° and 120°-transitions of the torsional angles $\phi, \psi, \chi_1, \chi_2$, and χ_3 were counted during the 20-ns simulation time. The root-mean-square fluctuations (RMSF) of the torsional angles were calculated and averaged analogously to Eqs. (2) and (3).

^1H–^{15}N order parameters

The ^{1}H–^{15}N order parameters (S^2) of hen egg-white lysozyme (HEWL) were calculated using the following equation (Henry and Szabo 1985; Chandrasekhar et al. 1992),

$$S^2 = \frac{1}{2}\left[3\sum_{\alpha=1}^{3}\sum_{\beta=1}^{3}\langle\mu_\alpha\mu_\beta\rangle^2 - 1\right], \tag{7}$$

where μ_α are x- , y-, and z-components of the normalised inter-atomic N-H vector. The order parameters were compared to experimental data of HEWL (Buck et al. 1995).

NMR NOE proton-proton distances and ^3J-coupling constants

A total of 1,630 experimental NMR proton-proton upper distance bounds of HEWL were taken from Ref. Schwalbe et al. (2001) and pseudo-atom distance corrections as given by Wüthrich et al. (1983) were added. The proton-proton distances in the simulation were calculated using $1/r^6$ averaging, i.e. $\langle r^{-6}\rangle^{-\frac{1}{6}}$. Experimental ^3J-coupling constants of HEWL reported by Smith et al. (1991) were used for comparison to calculated coupling constants.

Backbone ^{15}N–^1H, $^{13}C_\alpha$–^{13}C, and ^{13}C–^{15}N residual dipolar couplings

The residual dipolar coupling (RDC) D_{ij} between two spins i and j is given by

$$D_{ij} = -\frac{\gamma_i\gamma_j\mu_0 h}{8\pi^3}\left\langle\frac{P_2(cos(\theta_{ij}))}{r_{ij}^3}\right\rangle, \tag{8}$$

where γ_i is the gyromagnetic ratio of spin i, μ_0 the magnetic permittivity of vacuum, h Planck's constant, P_2 the second-order Legendre polynomial, and θ_{ij} the angle between the inter-nuclear vector $\mathbf{r}_{ij}$ and the static magnetic field. If the protein is more or less rigid, Eq. (8) can be reformulated such that the averaging over different orientations of the protein with respect to the magnetic field is represented by an alignment tensor $\underline{\mathbf{A}}$,

$$D_{ij} = -\frac{\gamma_i\gamma_j\mu_0 h}{8\pi^3}\sum_{k\in x,y,z}\sum_{l\in x,y,z}\underline{A}_{kl}\cos(\zeta_k)\cos(\zeta_l), \tag{9}$$

where ζ_x, ζ_y and ζ_z are the angles between the inter-nuclear vector and the x-, y-, and z-axes, respectively. The alignment tensor $\underline{\mathbf{A}}$ is determined for every configuration in the trajectory by a singular-value decomposition (SVD), solving the equation

$$\underline{\mathbf{C}}\,\mathbf{a} = \underline{\mathbf{R}}, \tag{10}$$

where the matrix $\mathbf{a}$ contains the five independent elements of the 3 × 3 alignment tensor, and the matrix $\underline{\mathbf{R}}$ contains the N_{RDC} experimental RDCs used for the fit. The resulting Q-value reflects the quality of each SVD fit,

$$Q = \frac{\sqrt{\frac{1}{N_{\mathrm{RDC}}}\sum_{i=1}^{N_{\mathrm{RDC}}}(D_i^{\mathrm{exp}} - D_i^{\mathrm{calc}})^2}}{\sqrt{\frac{1}{N_{\mathrm{RDC}}}\sum_{i=1}^{N_{\mathrm{RDC}}}(D_i^{\mathrm{exp}})^2}}, \quad (11)$$

and can be displayed as a distribution for all configurations in the trajectory.

Results and discussion

First, the influence of the process of coarse graining on various non-measurable properties is analysed by a comparison of the configurational ensembles generated in the four different environments. The focus is on the FG and CG simulations, while the results of the SASA and vacuum simulations were added for completeness. Second, a comparison of simulated with measured values for quantities that can be measured or derived from experimental data is presented for HEWL, for which ample NMR data are available.

Comparison of environments

The four proteins hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM) were simulated for 20 ns in four different environments, i.e. in FG water (SPC), in CG water, in implicit solvent (a SASA model), and in vacuo, and their structural and dynamic properties were analysed and compared. As reference thereby the simulation in FG water is used, although an atomic level simulation also has finite accuracy in regard to the reproduction of experimental data.

Atom-positional RMSD, RMSF, and radius of gyration

The atom-positional root-mean-square deviations (RMSD) of the backbone atoms with respect to the crystal structure are shown in Fig. 2. No clear trend can be observed. For HEWL and CM, the RMSD in CG water is closest to the RMSD in FG water. The protein in the implicit solvent and in vacuo deviates more from the crystal structure in the case of HEWL and less in the case of CM. For CspA, the RMSDs of the protein in CG water and in vacuo are closest to the RMSD of the protein in FG water. In the case of GP, all other environments have a larger RMSD compared to the protein in FG water, with the protein in CG water and in implicit solvent deviating the most.

The average atom-positional root-mean-square fluctuations (RMSF) over 20 ns of simulation of the backbone C_α atoms and of the last non-hydrogen atom of each side chain are given in Table 1. Again, no clear trend is observable. Both the average backbone RMSF and the average side-chain RMSF of the protein in CG water underestimate the ones in FG water in the case of HEWL, CspA, and CM, but overestimate them in the case of GP. The average RMSFs in implicit solvent and in vacuo are sometimes closer to the value in FG water than the value in CG water is, and sometimes further away.

The radius of gyration (RGYR) of the protein in CG water is similar to the one of the protein in FG water in the case of HEWL, CspA, and CM, and slightly increased in the case of GP (Fig. 3). In the latter case, the simulation in vacuo results in the most similar RGYR.

Secondary structure elements, intra-protein hydrogen bonds and energies

The time evolution of the secondary structure elements is shown in Fig. 4. Overall, the secondary structure of all four proteins is stable in CG water. In the case of HEWL, seven of the eight helices (residues 5–14, 20–22, 25–36, 80–84, 89–101, 109–114, and 120–123) and both β-strands (residues 43–45 and 51–53) reported in the crystal structure (Artymiuk et al. 1982) are maintained. Only the 3_{10}-helix

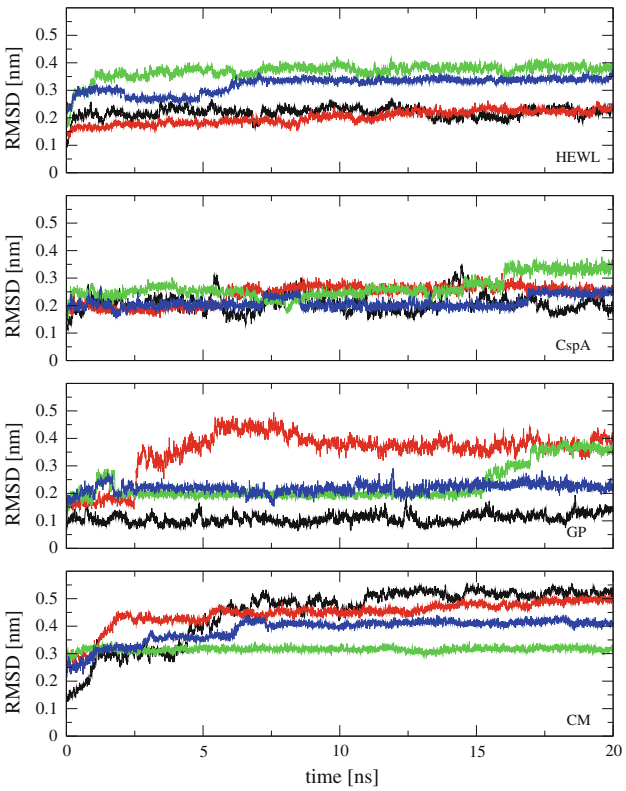

Fig. 2 Atom-positional root-mean-square deviations (RMSD) of the backbone with respect to the crystal structure of hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM). Simulations in FG water are shown in *black*, those in CG water in *red*, those using the SASA model in *green*, and those in vacuum in *blue*

Table 1 Average atom-positional root-mean-square fluctuation ⟨RMSF⟩ of the backbone (C_α) and the last atom of the side chains (sc) using Eqs. (2) and (3) of hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM). All values in nm and calculated over 20 ns

Solvent	HEWL		CspA		GP		CM	
	C_α	sc	C_α	sc	C_α	sc	C_α	sc
FG	0.108	0.186	0.133	0.225	0.079	0.170	0.205	0.258
CG	0.096	0.153	0.091	0.141	0.148	0.205	0.122	0.180
SASA	0.105	0.179	0.132	0.131	0.221	0.217	0.093	0.167
Vacuo	0.084	0.160	0.085	0.161	0.122	0.219	0.126	0.198

(residues 104–107) is lost during the simulation. This helix is also lost in the simulation of the protein in vacuo together with the last helix (residues 120–123), which is lost in the implicit solvent simulation. The crystal structure (Schindelin et al. 1994) of CspA reports one helix (residues 35–37) and five β-strands (residues 5–13, 18–23, 30–34, 50–56, and 63–69), which are all maintained in the protein in CG water as well as in the other environments, although with more fluctuations. In the case of GP, one helix (residues 23–36) and four β-strands (residues 1–9, 12–20, 42–46 and 50–56) are reported in the crystal structure (Gallagher et al. 1994). One end of the helix unfolds partly during the simulation in CG water, and the third β-strand (residues 42–46) becomes thinner, while these secondary elements are completely stable in the protein in FG water. However, the same β-strand is also only partly stable in implicit solvent where in addition the second β-strand (residues 12–20) partly disappears. The secondary structure of CM consists of nine α-helices (residues 3–16, 16–27, 33–51, 54–83, 84–87, 94–115, 115–120, 123–140, 142–154) as reported in the crystal structure (Okvist et al. 2006). These helices are maintained throughout the 20-ns simulation time in all environments.

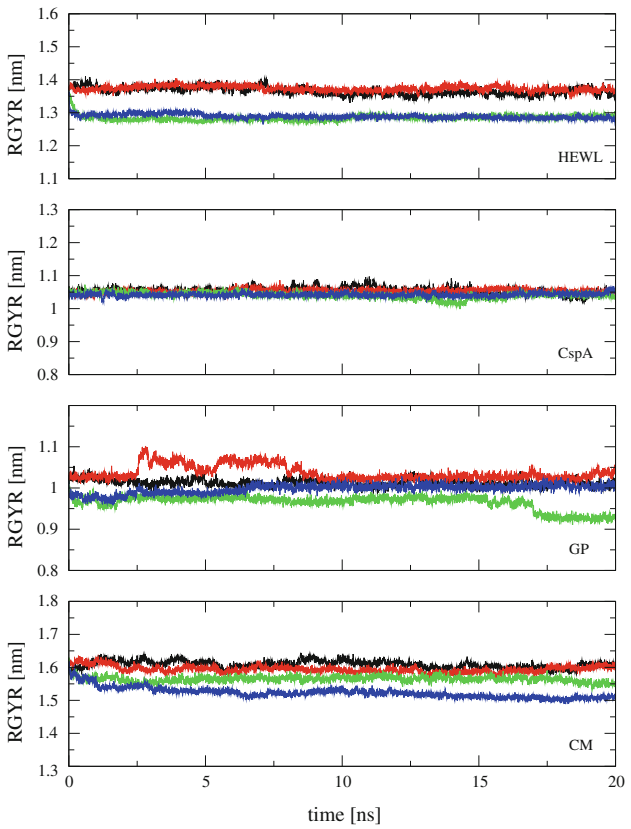

Fig. 3 Radius of gyration (RGYR) of hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM). Simulations in FG water are shown in *black*, those in CG water in *red*, those using the SASA model in *green*, and those in vacuum in *blue*

The intra-protein hydrogen bonds (H-bonds) were split into backbone-backbone (bb-bb), backbone-side chain (bb-sc), and side chain-side chain (sc-sc) H-bonds, and are listed in Table 2. Both the total number of H-bonds observed and the sum of the occurrence in percentage are given. In the case of bb-bb H-bonds, the total number observed in CG water is generally closer to the number in FG water compared to the other environments. In the simulations in vacuo, both the total number of H-bonds and the summed occurrence are increased compared to FG water, indicating that the vacuum surrounding the protein favours these interactions. The same is found for the simulations in implicit solvent for HEWL and GP. In the case of CspA and CM, the total number of H-bonds is increased in the implicit solvent but not the summed occurrence. The analysis of the bb-sc H-bonds shows that more and more stable H-bonds occur in all other environments compared to FG water, with the largest numbers of H-bonds found in CG water. In the case of sc-sc H-bonds, on the other hand, significantly more and more stable H-bonds are observed in the protein in CG water compared to the three other environments. This is not entirely surprising given that the polar side chains at the surface of the protein can no longer form H-bonds with the solvent and there is no additional term driving the polar side chains away from the protein as in the SASA model. In addition, the vacuum force field 54B7 used for the simulations in vacuum and in implicit solvent does not have full charges on polar side chains such

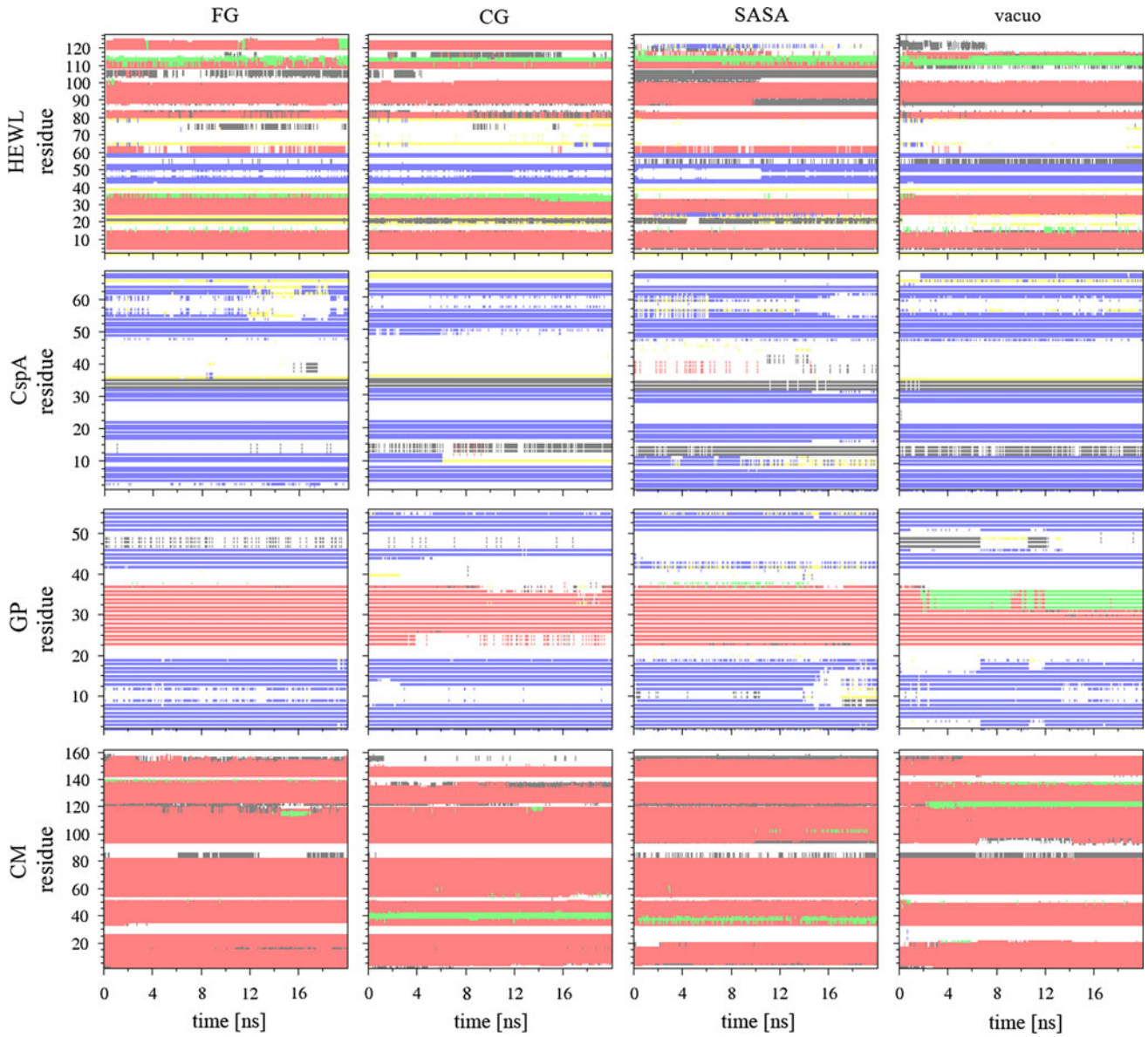

Fig. 4 From *top* to *bottom*, secondary structure elements of hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM). α-Helices are shown in red, 3_{10}-helices in *black*, π-helices in *green*, β-bridges in *yellow*, and β-strands in *blue*

as lysine, arginine, etc., which also results in a smaller driving force to form sc-sc hydrogen bonds.

The observation of an increased number of sc-sc H-bonds is reflected in the intra-protein potential energy and its components are listed in Table 3. H-bonds are electrostatically highly favourable, but the close vicinity of the atoms involved leads to a repulsive Lennard-Jones energy contribution. For all four proteins, the electrostatic energy term V_{CRF} is thus lower and the Lennard-Jones energy term V_{LJ} higher in CG water, resulting in an overall lower potential energy compared to FG water. Interestingly, both the implicit solvent and the vacuum environment lead to a lower V_{LJ} and a higher V_{CRF} compared to FG water.

Torsional-angle analysis

The Ramachandran plots of the backbone torsional angles ϕ and ψ are shown in Fig. 5. As was already observed in the secondary structure element analysis, the secondary structure is well preserved in the mixed-grained simulations, and thus the size and position of the maxima in the Ramachandran plots are maintained. In the implicit solvent and in vacuo, additional configurations with torsional

Table 2 Total number (#) and total occurrence (occ) over 20 ns of hydrogen bonds in the protein of hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM), split into backbone-backbone (bb-bb), backbone-side chains (bb-sc), and side chain-side chain (sc-sc) hydrogen bonds. The total occurrence is given as the sum of the occurrences of the individual hydrogen bonds obtained using Eq. (6)

Protein	Solvent	bb-bb		bb-sc		sc-sc	
		#	occ	#	occ	#	occ
HEWL	FG	89	5,518	67	2,023	36	1,312
	CG	88	5,854	116	4,288	75	2,766
	SASA	102	6,050	73	2,805	34	1,629
	Vacuo	101	6,634	78	3,299	33	1,477
CspA	FG	43	2,967	13	468	22	573
	CG	46	2,953	59	1,640	51	1,663
	SASA	48	2,765	25	1,137	8	385
	Vacuo	53	3,782	26	1,080	13	603
GP	FG	36	2,938	20	570	16	557
	CG	40	2,381	44	1,393	64	1,689
	SASA	62	3,616	26	1,140	14	548
	Vacuo	55	2,918	34	1,015	17	548
CM	FG	124	9,404	43	1,206	47	1,477
	CG	143	9,246	92	3,438	153	5,217
	SASA	141	9,359	80	3,213	42	1,415
	Vacuo	158	9,968	82	2,899	50	1,775

Table 3 Internal potential energy of the protein V_{pot} and its components, the Lennard-Jones energy V_{LJ}, and the electrostatic energy V_{CRF} of hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM)

Protein	Solvent	V_{pot} (kJ mol^{-1})	V_{LJ} (kJ mol^{-1})	V_{CRF} (kJ mol^{-1})
HEWL	FG	−14,385	−3,417	−10,967
	CG	−16,884	−3,146	−13,738
	SASA	−13,434	−3,539	−9,895
	Vacuo	−13,643	−3,603	−10,040
CspA	FG	−6,154	−1,621	−4,533
	CG	−8,406	−1,403	−7,003
	SASA	−4,160	−1,664	−2,496
	Vacuo	−4,211	−1,728	−2,483
GP	FG	−6,275	−1,375	−4,899
	CG	−8,253	−1,099	−7,154
	SASA	−3,672	−1,373	−2,299
	Vacuo	−3,762	−1,446	−2,316
CM	FG	−17,701	−4,256	−13,446
	CG	−22,253	−3,902	−18,352
	SASA	−15,270	−4,610	−10,659
	Vacuo	−15,730	−4,687	−11,043

angles in the upper right quarter were sampled in all four proteins compared to the protein in FG water.

The total number of transitions of the backbone torsional angles ϕ and ψ and of the side-chain torsional angles χ_1, χ_2 and χ_3 are listed together with the respective average root-mean-square fluctuation in Table 4. For the backbone angles, the total number of transitions observed in the protein in CG water is closest to the value of the protein in FG water (except for ϕ in CM and ψ in CspA). There is no clear trend in the average RMSF values but they are generally close between all environments with a few exceptions, e.g. the RMSF of ϕ in CspA and of ψ in GP in the vacuum environment. The same general trend holds for the first two side-chain torsional angles χ_1 and χ_2, although the average RMSF of χ_2 is generally lower in CG water than in the other environments. This is even more pronounced for χ_3, where the average RMSF in CG water is significantly lower and also the total number of transitions is smaller for all four proteins. This finding is in agreement with the increased number of hydrogen bonds involving side chains. There are only five amino acids with a χ_3 torsional angle and four of these five amino acids are polar ones forming hydrogen bonds, thus dominating the analysis for this torsional angle.

Comparison to experiment

The simulations of hen egg-white lysozyme (HEWL) in FG water, in CG water, and in vacuo were compared to experimental data, i.e. NOE distance bounds, ^{1}H–^{15}N order parameters, $^{3}J_{H_N H_\alpha}$-coupling constants, and residual dipolar couplings (RDC).

The distributions of the NOE distance bound violations observed in the three simulations are shown in Fig. 6. In general, the number and size of the violations are slightly increased in the mixed-grained simulation but smaller compared to the vacuum environment. There were six violations larger than 0.3 nm found in the protein in FG water, which are, however, not crucial as all of them involve the carboxy terminal residue (LEU129), which is very mobile. In the protein in CG water, also six violations larger than 0.3 nm were found and they also involve LEU129. In the vacuum simulation, 10 violations larger than 0.3 nm were observed. One of them involves LEU129 and three involve TRP28. The other six violations are between residues in different parts of the protein.

The ^{1}H–^{15}N order parameters of the protein in the three environments are compared to order parameters derived from experiment in Fig. 7. The order parameters from the simulations are generally smaller than those derived from the experiment. The largest order parameters are found for residues in helices. This pattern is reproduced roughly by

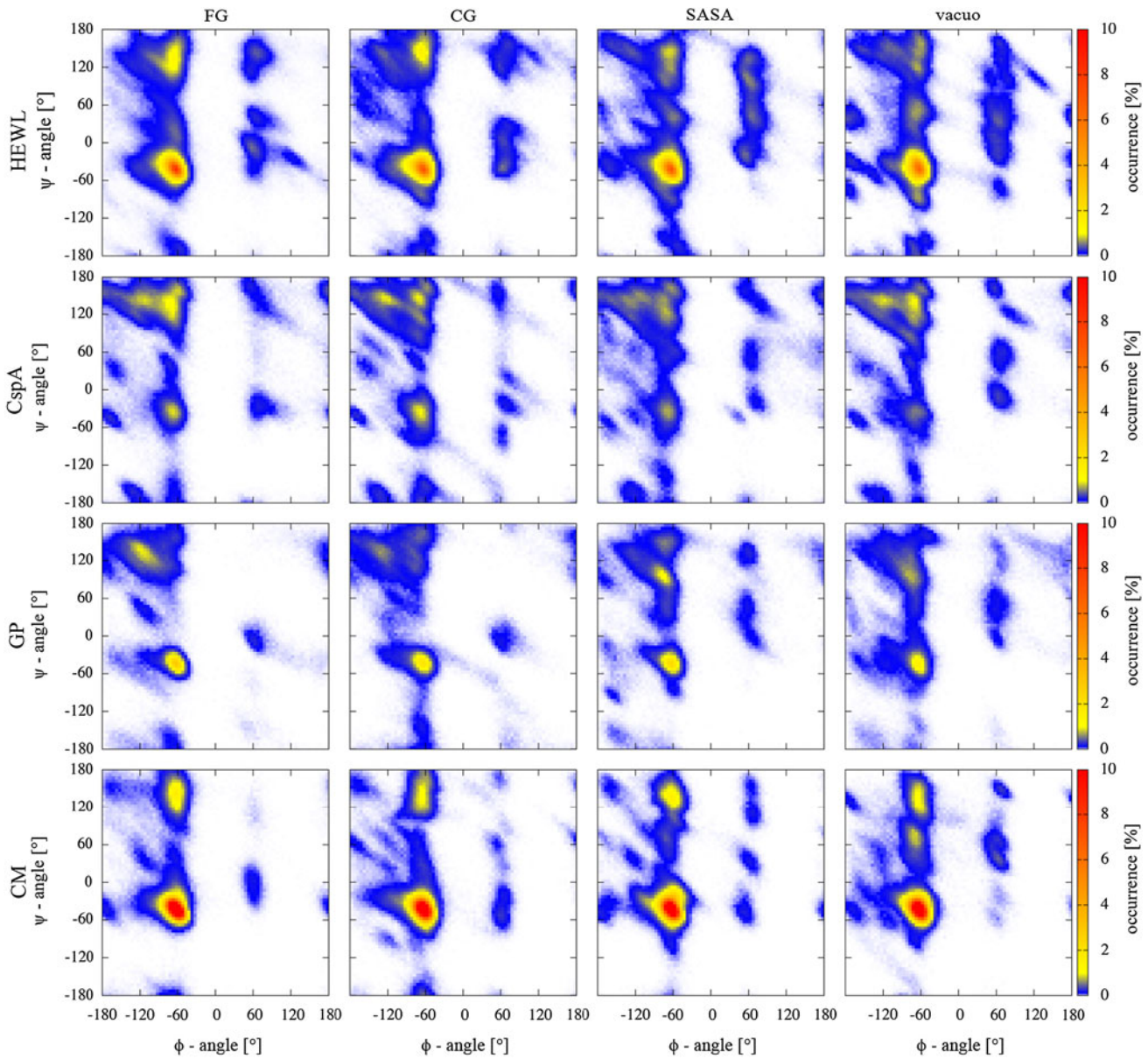

Fig. 5 From *top* to *bottom*, Ramachandran plots of hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM)

all three simulations. The vacuum simulation, however, largely overestimates the order parameter of residues 85 and 102/103, while the CG simulation only overestimates the order parameter of residue 85. The largest differences between the order parameters in the CG simulation compared to the FG simulation and the experimental data are observed for residues 74 and 102–105, for which the CG simulation yields much less order. Yet, the C_α positional RMS fluctuations for these residues display similar values, i.e. mobility (data not shown). The total number of bb-bb hydrogen bonds involving residues in the range 70–80 and 100–110 is 25 for the FG simulation and 24 for the CG simulation. For the bb-sc hydrogen bonds, these numbers are 8 (FG) and 25 (CG), and for the sc-sc hydrogen bonds these numbers are 6 (FG) and 18 (CG). A detailed comparison of these hydrogen bonds with high occurrence, i.e. $\geq$50 %, is given in Table 5. The bb-bb hydrogen bond patterns in the range 70–80 for the FG and CG simulations are similar. In the range of 100–110, the patterns are different, e.g. the hydrogen bond between residue 105 and 108 is very stable in the FG simulation but barely present in the CG simulation while for the hydrogen bond between residue 105 and 107 the picture is inverted. The hydrogen bonds in the CG simulation are generally more stable, which would indicate more order instead of less. Taken this together, there is no clear explanation for the difference in

Table 4 Total number (#) and average root-mean-square fluctuation ⟨RMSF⟩ (in degrees) of the torsional angles ϕ, ψ, χ_1, χ_2, and χ_3 using Eqs. (2) and (3) of hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM)

Angle	Solvent	HEWL		CspA		GP		CM	
		#	RMSF	#	RMSF	#	RMSF	#	RMSF
ϕ	FG	25,022	65	26,818	109	30,459	76	10,335	26
	CG	35,464	77	22,973	79	23,624	76	15,389	31
	SASA	17,366	72	32,426	95	12,989	76	13,055	44
	Vacuo	19,588	81	29,703	140	18,676	77	11,864	42
ψ	FG	8,800	103	5,324	108	7,600	112	3,845	69
	CG	10,232	92	9,381	101	11,167	108	8,670	81
	SASA	19,571	94	11,452	124	19,444	96	10,191	70
	Vacuo	15,267	104	11,767	105	16,237	282	10,865	73
χ_1	FG	8,404	173	5,669	196	4,324	131	13,658	204
	CG	9,433	158	4,496	147	4,759	146	23,057	175
	SASA	13,136	152	13,051	203	10,844	205	20,497	169
	Vacuo	10,908	144	8,422	160	9,001	165	19,383	152
χ_2	FG	44,809	420	15,636	1,105	16,696	661	37,981	804
	CG	47,733	304	16,640	322	16,661	347	43,480	360
	SASA	46,484	719	26,517	2,396	23,550	1,953	57,734	1,165
	Vacuo	42,776	848	21,773	1,078	21,779	817	50,065	974
χ_3	FG	12,265	721	9,944	1,071	13,275	1,245	31,937	759
	CG	10,075	414	6,916	230	8,384	360	28,440	452
	SASA	13,664	1,354	12,607	1,442	18,499	2,805	37,196	1,782
	Vacuo	12,332	541	11,857	1,668	16,120	1,087	37,387	1,368

the order parameters. However, the different number and occurrence of the side chain hydrogen bonds may explain the loss of the 3_{10} helix (residues 104–107) observed in the CG simulation. Especially the high occurrence of a hydrogen bond between the side chains of residues 106 and 116 as well as between the side chain of residue 103 and the backbone of residue 21 leads to an unfolding of the helical structure at this location.

Experimental proton–proton $^3J_{H_N H_\alpha}$-coupling constants of HEWL (Smith et al. 1991) were compared to calculated values from the simulations (Fig. 8). The root-mean-square deviations (RMSD) of the calculated coupling constants from the experimental values are 1.6, 1.8, and 1.8 Hz for HEWL in FG water, in CG water and in vacuo, respectively. However, the quality of the calculated ^3J-couplings is low because of the inaccuracy due to the empirical nature of the parameters of the Karplus relation, which is used to obtain a ^3J-coupling from a configuration (Allison and van Gunsteren 2009; Steiner et al. 2012). Additionally, the sampling in a simulation is limited compared to the experimental time scale.

The quality of fitting the backbone ^{15}N–^{1}H, $^{13}C_\alpha$–^{13}C, and ^{13}C–^{15}N residual dipolar couplings (RDC) calculated from simulation to the experimental RDCs of HEWL (Higman 2004) is assessed by a so-called Q-value (Eq. 11). The closer a Q-value is to zero, the better is the quality of the fit. The distribution of the Q-values observed in the three simulations is shown in Fig. 9. The values of the simulations are all much higher than the value calculated from the crystal structure (Artymiuk et al. 1982). The protein in FG water results in Q-values closest to zero compared to the other two environments. Interestingly, the mixed-grained simulation performs similarly to the purely atomistic simulation for the $^{13}C_\alpha$–^{13}C RDCs and worse for the ^{13}C–^{15}N RDCs, while the performance for the ^{15}N–^{1}H RDCs is worse than in the vacuum simulation.

Computational efficiency

A simulation of 500 steps of HEWL in FG water, in CG water, and in implicit solvent was performed with a single CPU and with four CPUs using MPI or OpenMP parallelisation. The number of CG beads in the mixed-grained simulation represented the same number of water molecules as present in the purely atomistic simulation to allow a fair comparison. The simulation in CG water is an order of magnitude faster than the purely atomistic simulation independent of the number of CPUs used, although much larger cutoffs for all non-bonded interactions are used in the mixed-grained simulations, i.e. 1.4 and 2.0 nm versus

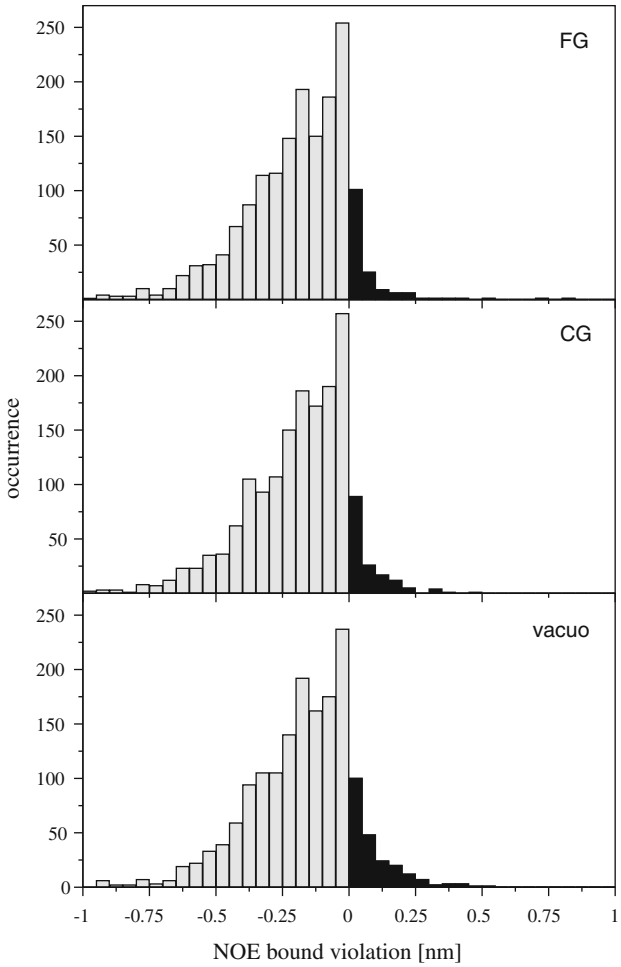

Fig. 6 Distribution of 1,630 experimental NOE-bound violations for the 20-ns simulation of hen egg-white lysozyme (HEWL) in FG water (*top*), CG water (*middle*), and in vacuo (*bottom*). NOE-bound violations are coloured *black*, while the satisfied NOE bounds appear in *grey*

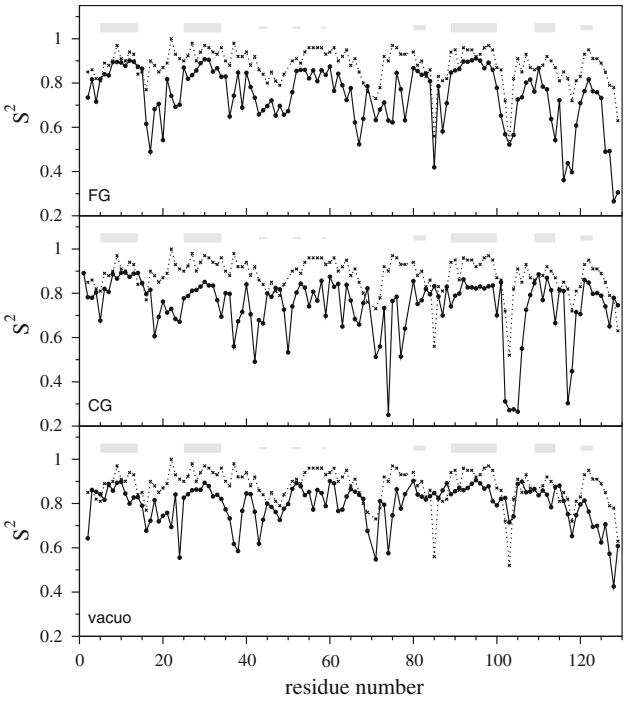

Fig. 7 $^{1}H–^{15}N$ order parameters of hen egg-white lysozyme (HEWL) derived from experiment (Buck et al. 1995) (*dotted line with crosses*) compared to calculated order parameters (*solid line with circles*) of HEWL in FG water (*top*), CG water (*middle*), and in vacuum (*bottom*). $^{1}H–^{15}N$ order parameters are calculated over the whole 20 ns of simulation time using a 1-ns time window for averaging. The *grey bars* on top of each graph indicate the secondary structure elements of HEWL, i.e. α-helix (*thick bars*), 3_{10}-helix (*medium bars*), and β-strand (*thin bars*)

0.8 and 1.4 nm. The use of four instead of one CPU leads to a speed-up of a factor three. On a single CPU, the simulations in CG water and in implicit solvent are equally fast. However, the SASA model does not allow parallelisation as well as the CG model, resulting in a factor two difference in speed on four CPUs.

Summary and conclusions

Molecular dynamics (MD) simulations of four proteins, hen egg-white lysozyme (HEWL), major cold shock protein (CspA), G-protein (GP), and chorismate mutase (CM), solvated in coarse-grained (CG) water were analysed in terms of structural and dynamic properties, and compared to simulations of the same proteins in atomistic, fine-grained (FG) water, in implicit solvent (SASA model), and in vacuo. In addition, the simulations of HEWL were compared to experimental NOE distance bounds, ^{3}J-coupling constants, N–H order parameters, and backbone residual dipolar couplings.

The comparison of the backbone atom-positional root-mean-square deviation (RMSD) with respect to the crystal structures, the atom-positional root-mean-square fluctuations (RMSF) of the backbone and the side chains, the radius of gyration, the evolution of the secondary structure elements, and the Ramachandran plots shows no significant difference between the purely atomistic simulations and mixed-grained simulations. The protein in CG water generally performs better compared to the simulations in implicit solvent or in vacuo. The major difference observed is an increased occurrence of side chain-side chain (sc-sc) hydrogen bonds due to the lack of hydrogen-bond partners in the solvent. For the simulations in implicit solvent and in vacuo, the corresponding vacuum force field was used, which does not have full charges on the ARG, LYS, ASP, and GLU side chain and the polypeptide chain termini, therefore preventing the formation of salt bridges. In addition, the SASA model contains a term to drive polar side chains on the surface away from the protein. The

Table 5 Details of the backbone—side chain (bb-sc) and side chain–side chain (sc-sc) hydrogen bonds involving residues 70–80 or 100–110 of hen egg-white lysozyme (HEWL) observed in the simulations with FG water and CG water with an occurrence of at least 50 % in at least one of the simulations

Donor residue	Acceptor residue	Donor atoms	Acceptor atom	Occurrence [%]	
				FG	CG
65 ASN	78 ILE	N–H	O	63	65
73 ARG	61 ARG	N–H	O	70	86
76 CYS	63 TRP	N–H	O	94	93
77 ASN	74 ASN	N–H	O	54	79
78 ILE	76 CYS	N–H	O	77	80
80 CYS	65 ASN	N–H	O	94	99
82 ALA	79 PRO	N–H	O	56	59
83 LEU	80 CYS	N–H	O	61	65
100 SER	96 LYS	N–H	O	81	38
101 ASP	97 LYS	N–H	O	37	96
102 GLY	98 ILE	N–H	O	1	61
106 ASN	104 GLY	N–H	O	37	70
107 ALA	105 MET	N–H	O	0	61
108 TRP	105 MET	N–H	O	90	2
112 TRP	108 TRP	N–H	O	65	96
113 ASN	109 VAL	N–H	O	56	87
115 CYS	110 ALA	N–H	O	83	99
67 GLY	72 SER	N–H	OG	3	71
72 SER	60 SER	OG–HG	O	41	58
100 SER	96 LYS	OG–HG	O	8	88
103 ASN	21 ARG	ND2–HD2	O	0	57
105 MET	23 TYR	N–H	OH	60	33
108 TRP	56 LEU	NE1–HE1	O	74	51
110 ALA	35 GLU	N–H	OE2	51	65
112 ARG	107 ALA	NE–HE	O	2	57
112 ARG	107 ALA	NH2–HH2	O	0.4	55
20 TYR	100 SER	OH–HH	OG	28	92
27 ASN	105 MET	ND2–HD2	SD	59	23
65 ASN	74 ASN	ND2–HD2	OD1	26	58
97 LYS	101 ASP	NZ–HZ	OD1	6	62
97 LYS	101 ASP	NZ–HZ	OD2	0.4	65
116 LYS	106 ASN	NZ–HZ	OD1	0.3	78

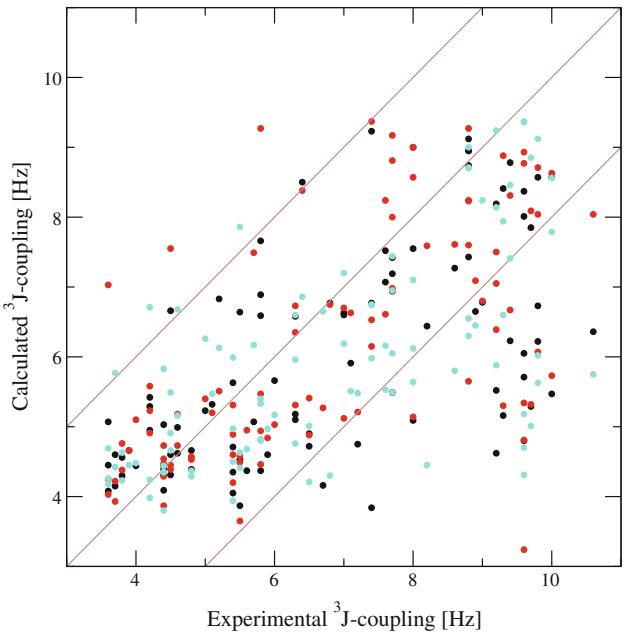

Fig. 8 Comparison of calculated $^3J_{H_N H_\alpha}$-coupling constants of hen egg-white lysozyme (HEWL) in FG water (*black*), CG water (*red*), and in vacuum (*cyan*) to the experimental coupling constants of HEWL (Smith et al. 1991). The coupling constants were calculated from 20-ns simulation time. A 2-Hz region of uncertainty is indicated by *brown lines* (Steiner et al. 2012)

increased number of sc-sc hydrogen bonds in the mixed-grained simulations is reflected in the intra-protein potential energy V_{pot} and its components. Hydrogen bonds are electrostatically favourable but due to the short distance between the involved atoms the Lennard-Jones energy is less negative, i.e. more repulsive. Thus, the Lennard-Jones energy V_{LJ} of the protein in CG water is less negative compared to that of the protein in FG water and the electrostatic energy V_{CRF} is more negative. Interestingly, the opposite trend is found for the protein in implicit solvent and in vacuo, where V_{LJ} is more negative and V_{CRF} is less negative compared to the protein in FG water. The effect of the larger occurrence of sc-sc hydrogen bonds can also be observed in the analysis of side-chain torsional angles χ_1, χ_2 and χ_3. While the total number of transitions and the RMSF of the first two angles, which are present both in side chains forming hydrogen bonds (mostly polar) and in such not forming any (apolar), show no clear difference between the proteins in CG or FG water, significantly fewer transitions occur for the last angle χ_3 and also the RMSF is much smaller in CG water compared to FG water. This angle occurs mainly in polar side chains. If these side chains form more and more stable hydrogen bonds, the number of transitions will decrease and thus the RMSF will be reduced.

Further analysis of the simulations of HEWL in FG water, in CG water, and in vacuo shows that the extent of agreement with the available experimental data is not much decreased in the mixed-grained simulation compared to the purely atomistic simulation. The number of NOE distance bound violations increases slightly and the RMSD of the calculated and measured $^3J_{H_N H_\alpha}$-coupling constants increases from 1.6 to 1.8 Hz. The distributions of the Q-values of the backbone RDCs are shifted but in general still smaller than in the vacuum simulation.

Finally, it can be stated that the fold of the atomistic proteins in CG water is stable within the 20 ns simulation

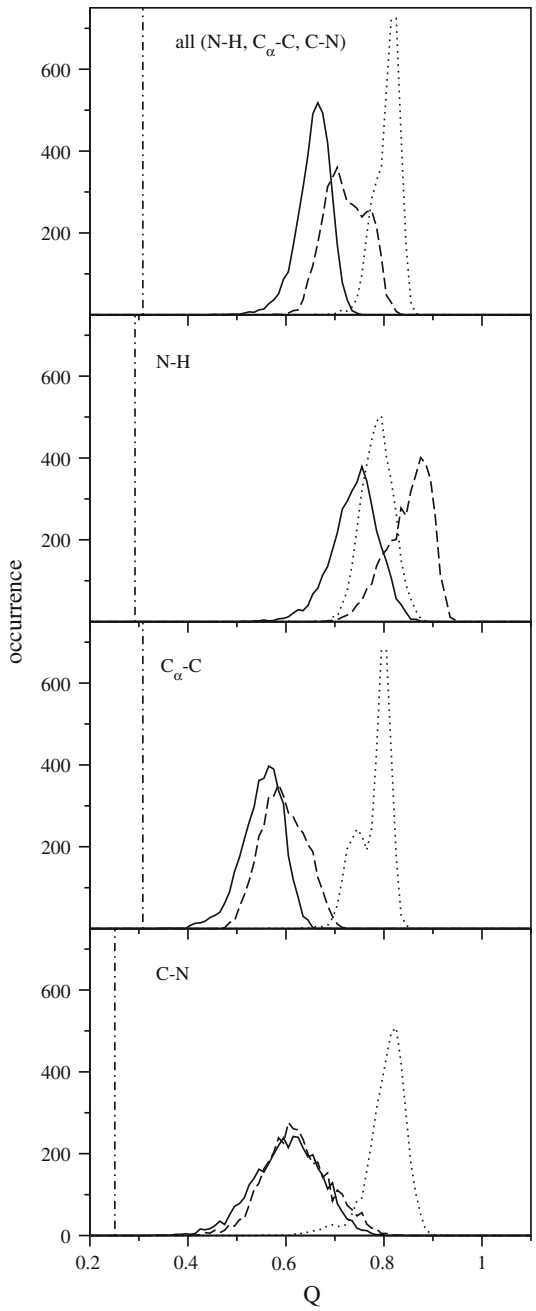

Fig. 9 Q value distribution of the calculated backbone N–H, C_α–C, and C–N RDCs of hen egg-white lysozyme (HEWL) in FG water (*solid line*), CG water (*dashed line*), and in vacuo (*dotted line*) when fitting the calculated residual dipolar couplings (RDC) of each configuration to the experimentally derived RDCs of HEWL (Higman 2004). The Q value of the fits from the X-ray crystal structure (Artymiuk et al. 1982) is shown by a *vertical dashed-dotted line*

length and no larger alteration in the structural and dynamic properties can be detected except an increased occurrence of sc-sc hydrogen bonds. This difference comes, however, with a speed-up of an order of magnitude compared to purely atomistic simulations. Therefore, a possible application of these mixed-grained simulations can be in free energy difference calculations between a series of ligands bound to a protein, where only the overall stability of the secondary structure and thus of the (buried) binding pocket is important. The speed-up gained through the use of CG solvent would allow the application of accurate methods such as thermodynamic integration on larger data sets.

Acknowledgments This work was financially supported by the National Center of Competence in Research (NCCR) in Structural Biology and by grant number 200020-137827 of the Swiss National Science Foundation and by grant no. 228076 of the European Research Council, which is gratefully acknowledged.

References

Allison J, van Gunsteren WF (2009) A method to explore protein side chain conformational variability using experimental data. ChemPhysChem 10:3213–3228

Allison JR, Boguslawski K, Fraternali F, van Gunsteren WF (2011) A refined, efficient mean solvation force model that includes the interior volume contribution. J Phys Chem B 115:4547–4557

Artymiuk PJ, Blake CCF, Rice DW, Wilson KS (1982) The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 Angstroms resolution. Acta Crystallogr Sec B 38:778–783

Basdevant N, Ha-Duong T, Borgis D (2006) Particle-based implicit solvent model for biosimulations: application to proteins and nucleic acids hydration. J Chem Theory Comput 2:1646–1656

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullmann B (ed) Intermolecular Forces. Reidel, Dordrecht, pp 331–342

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

Buck M, Boyd J, Redfield C, MacKenzie DA, Jeenes DJ, Archer DB, Dobson CM (1995) Structural determinants of protein dynamics: analysis of ^{15}N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. Biochemistry 34:4041–4055

Chandrasekhar I, Clore G, Szabo A, Gronenborn A, Brooks B (1992) A 500 ps molecular dynamics simulation study of interleukin-1β in water—correlation with nuclear magnetic resonance spectroscopy and crystallography. J Mol Biol 226:239–250

Eichenberger AP, Allison JR, Dolenc J, Geerke DP, Horta BAC, Meier K, Oostenbrink C, Schmid N, Steiner D, Wang D, van Gunsteren WF (2011) The GROMOS++ software for the analysis of biomolecular simulation trajectories. J Chem Theory Comp 7:3379–3390

Fraternali F, van Gunsteren WF (1996) An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. J Mol Biol 256:939–948

Gallagher T, Alexander P, Bryan P, Gilliland GL (1994) Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. Biochemistry 33:4721–4729

Henry E, Szabo A (1985) Influence of vibrational motion on solid state line shapes and NMR relaxation. J Chem Phys 82:4753–4761

Higman VA (2004) The use of bicelles and other ordered media to study protein structure and dynamics. Ph.D. thesis, University of Oxford

Hockney RW (1970) The potential calculation and some applications. Methods Comput Phys 9:136–211

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Kunz A-PE, Allison JR, Geerke DP, Horta BAC, Hünenberger PH, Riniker S, Schmid N, van Gunsteren WF (2012) New functionalities in the GROMOS biomolecular simulation software. J Comput Chem 33:340–353

Lide DR (2007–2008) Handbook of chemistry and physics, 88th edn. CRC Press/Taylor and Francis, Boca Raton

Liu Y, Ichiye T (1996) Soft sticky dipole potential for liquid water: a new model. J Phys Chem 100:2723–2730

Masella M, Borgis D, Cuniasse P (2008) Combining a polarizable force-field and a coarse-grained polarizable solvent model: application to long dynamics simulations of bovine pancreatic trypsin inhibitor. J Comput Chem 29:1707–1724

Masella M, Borgis D, Cuniasse P (2011) Combining a polarizable force-field and a coarse-grained polarizable solvent model. II. Accounting for hydrophobic effects. J Comput Chem 32:2664–2678

Michel J, Orsi M, Essex JW (2008) Prediction of partition coefficients by multiscale hybrid atom-level/coarse-grain simulations. J Phys Chem B 112:657–660

Neri M, Anselmi C, Cascella M, Maritan A, Carloni P (2005) Coarse-grained model of proteins incorporating atomistic detail of the active site. Phys Rev Lett 95:218102

Okvist M, Dey R, Sasso S, Grahn E, Kast P, Krengel U (2006) 1.6 Angstrom crystal structure of the secreted chorismate mutase from Mycobacterium tuberculosis: novel fold topology revealed. J Mol Biol 357:1483–1499

Orsi M, Sanderson WE, Essex JW (2009) Permeability of small molecules through a lipid bilayer: A multiscale simulation study. J Phys Chem B 113:12019–12029

Riniker S, van Gunsteren WF (2011) A simple, efficient polarizable coarse-grained water model for molecular dynamics simulation. J Chem Phys 134:084110

Riniker S, van Gunsteren WF (2012) Mixing coarse-grained and fine-grained water in molecular dynamics simulations of a single system. J Chem Phys (in press)

Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical-integration of Cartesian equations of motion of a system with contraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341

Rzepiela AJ, Louhivuori M, Peter C, Marrink SJ (2011) Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites. Phys Chem Chem Phys 13:10437–10448

Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF (2011) Definition and testing of the GROMOS force-field versions: 54A7 and 54B7. Eur Biophys J 40:843–856

Schmid N, Christ CD, Christen M, Eichenberger AP, van Gunsteren WF (2012) Architecture, implementation and parallelization of the GROMOS software for biomolecular simulation. Comput Phys Comm 183:890–903

Schindelin H, Jiang W, Inouye M, Heinemann U (1994) Crystal structure of CspA, the major cold shock protein of Escherichia coli. Proc Natl Acad Sci USA 91:5119–5123

Schwalbe H, Grimshaw S, Spencer A, Buch M, Boyd J, Dobson C, Redfield C, Smith L (2001) A refined solution structure of hen lysozyme determined using residual dipolar coupling data. Protein Sci 10:677–688

Shi Q, Izvekov S, Voth GA (2006) Mixed atomistic and coarse-grained molecular dynamics: simulation of membrane-bound ion channel. J Phys Chem B 110:15045–15048

Smith L, Sutcliffe M, Redfield C, Dobson C (1991) Analysis of F and χ_1 torsion angles for hen lysozyme in solution from ^{1}H NMR spin-spin coupling constants. Biochemistry 30:986–996

Steiner D, Allison JR, van Gunsteren WF (2012) On the calculation of $^3J_{\alpha\beta}$-coupling constants for side chains in proteins. J Biomol NMR. doi:10.1007/s10858-012-9634-5

Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. J Chem Phys 102:5451–5459

van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide; vdf Hochschulverlag AG an der ETH Zürich: Zürich, Groningen

Wüthrich K, Billeter M, Braun W (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J Mol Biol 169:949–961