Some Notes on the Thermodynamic Accuracy of Coarse-Grained Models

Ewa Anna Oprzeska-Zingrebe and Jens Smiatek*

Institute for Computational Physics, Theoretical Chemical Physics, University of Stuttgart, Stuttgart, Germany

Over the last decades, multiscale molecular dynamics (MD) simulations including *ab initio*, atomistic as well as coarse-grained models have significantly expanded our understanding of biologically relevant macromolecules like DNA, RNA, or proteins and their properties in solution. Despite the broad applicability, we comment here on some general challenges for coarse-grained approaches, the most important being a reliable thermodynamic description at large time and length scales.

Due to a massive increase in computational power, classical atomistic MD simulations are nowadays the method of choice for the study of complex molecular mechanisms, thereby taking into consideration hundreds of thousands of atoms on time scales of several microseconds. Although classical atomistic models provide a higher level of detail when compared to coarse-grained approaches, it has to be noted that the simplification of electronic behavior in terms of potential functions, so called force fields, introduces some conceptual artifacts into the dynamic and structural properties of the simulated molecular species (Dommert et al., 2012). Furthermore, polarization and charge-transfer mechanisms are usually ignored, such that more sophisticated *ab initio* or empirical models have to be used for systems where these effects become of importance (Smiatek et al., 2018; Kohagen et al., 2019; Nandy and Smiatek, 2019; Smiatek, 2019).

However, some processes take place on time and length scales, which are not accessible for atomistic MD simulations. Common examples are the formation of lipid bilayers and polyelectrolyte complexes, polymer and colloidal diffusion, charge transport or large scale DNA translocation (Smiatek and Schmid, 2011; Michalowsky et al., 2017, 2018; Smiatek and Holm, 2018). For the study of these and closely related problems, simple as well as more refined coarse-grained models offer a wide range of applications. Here, coarse-graining means the introduction of effective interaction sites (beads) instead of individual atoms, which reduces the degrees of freedom and thus also the number of necessary computations. In addition, the lower level of detail supports the straightforward use of implicit solvent approaches in combination with larger time steps (Marrink and Tieleman, 2013; Kleinjung and Fraternali, 2014; Onufriev and Case, 2019). Depending on the degree of coarse graining, one can differentiate between simple approaches such as reduced bead-spring models for polymers and advanced or semi coarse-grained methodologies such as iterative Boltzmann inversion or the MARTINI method among others (Reith et al., 2003; Clark et al., 2012; Marrink and Tieleman, 2013; Noid, 2013; McCarty et al., 2014; Rudzinski and Noid, 2014; Dunn and Noid, 2015; Guenza et al., 2018; Smiatek and Holm, 2018). Although advanced coarse-graining approaches are often based on rather mild parameterization procedures, it should be noted that the consideration of effective interaction sites crucially affects the resulting size and the geometry of the molecular species (Vögele et al., 2015a; Michalowsky et al., 2017, 2018). With regard to this point, also coarse-grained methodologies reveal some generic drawbacks, thereby limiting the applicability of these approaches for the thermodynamic analysis of complex solutions.

In terms of a specific example, many biologically relevant solutions, such as in mammalian or bacterial cells, are dense mixtures of various ions, co-solute and co-solvent species including a non-negligible concentration of solute components (Zhou et al., 2008). Among other effects, the individual components of the solution and their thermodynamic properties exert a tremendous influence on the structural stability of the dissolved biological species (Canchi and García, 2013; Smiatek, 2017; Oprzeska-Zingrebe and Smiatek, 2018a). For instance, it was shown (Zhang and Cremer, 2010; Canchi and García, 2013; Sukenik et al., 2013; Oprzeska-Zingrebe et al., 2018) that ions like SCN^- or molecules like urea destabilize DNA or protein structures, whereas the presence of SO_4^{2-}, trimethylamine-N-oxide (TMAO), or ectoine enhances the stability of native macromolecular states. Additionally, many molecular mechanisms are also dominated by intra- and intermolecular hydrogen bonds, polarization mechanisms as well as electrostatic and dispersion interactions. The presence of these mainly short-ranged interactions influences the radial distribution functions, potentials of mean force or the corresponding chemical potentials of the species, so that in the end, for non-negligible concentrations, there are more or less pronounced deviations from ideal solutions (Chandler, 1987; Smiatek, 2014, 2017; Dunn and Noid, 2015; Guenza et al., 2018; Oprzeska-Zingrebe and Smiatek, 2018a). The question now is whether coarse-grained models can reproduce these findings? Of course, one may wonder if the aforementioned properties need to be exactly reproduced, but we will illustrate by means of the following arguments that even slight deviations may have a decisive influence on the thermodynamic properties of the solution.

In more detail, modified interactions like in coarse-grained models under constant pressure p and temperature T result in variations of free energies, as defined by $G = H - TS$ with the enthalpy H and the entropy S, and changes in the chemical potential via $\mu_\alpha = (\partial G / \partial N_\alpha)_{p,T}$ where N_α denotes the number of molecules of species α. Due to changes in the enthalpy, also the corresponding molecular arrangements are affected, which often induces entropic variations as a second-order effect. Furthermore, changes of chemical potentials from reference chemical potential μ_α^0 with the universal gas constant R are directly related to changes in thermodynamic activities $a_\alpha = \exp((\mu_\alpha - \mu_\alpha^0)/RT)$, vapor pressures, solubilities or chemical reaction equilibria, as can be shown by relations from equilibrium thermodynamics and Kirkwood-Buff (KB) theory (Kirkwood and Buff, 1951; Ben-Naim, 2013) . In consequence, it becomes obvious that even slight modifications of molecular interactions may establish a non-negligible variation of relevant thermodynamic properties as it will be discussed in more detail in the following.

For illustrative purposes, we develop our arguments for a binary solution under isobaric-isothermal conditions with two components, including only solvent (index 1) and co-solvent (index 3) species. It has to be noted that the corresponding expressions change for different ensembles and higher-component mixtures, such that we here focus on one of the simplest examples (Smith, 2006). In KB theory, the derivative of the chemical potential of the co-solvent μ_3 is defined as

$$\frac{1}{RT} \left(\frac{\partial \mu_3}{\partial \ln \rho_3} \right)_{T,p} = \left(\frac{\partial \ln a_3}{\partial \ln \rho_3} \right)_{T,p} = \frac{1}{1 + \rho_3 (G_{33} - G_{31})}, \quad (1)$$

where ρ_3 denotes the number density of co-solvent species and G_{33} and G_{31} the corresponding KB integrals. A detailed explanation of KB integrals, their relation to radial distribution functions and their central meaning in KB theory can be found in the literature (Kirkwood and Buff, 1951; Ben-Naim, 2013; Smiatek, 2017; Oprzeska-Zingrebe and Smiatek, 2018a). For our considerations, it is sufficient to know that the KB integrals rely on radial distribution functions and represent excess volumes, which can be transformed into excess particle numbers $N_{\alpha\beta}^{xs} = \rho_\beta G_{\alpha\beta}$ for arbitrarily chosen components β around species α. With regard to this definition, Equation (1) can also be written as

$$\left(\frac{\partial \ln a_3}{\partial \ln \rho_3} \right)_{T,p} = \frac{1}{1 + \left(N_{33}^{xs} - (\rho_3/\rho_1) N_{31}^{xs} \right)} \quad (2)$$

with the excess number of solvent N_{31}^{xs} and co-solvent molecules N_{33}^{xs} in combination with the corresponding number densities ρ_1 and ρ_3. In terms of implicit solvent approaches with a continuum dielectric background, it follows that $N_{31}^{xs} = 0$ by definition, which implies that Equation (2) approaches the outcomes of experiments and atomistic models only under nearly ideal conditions with $\rho_3 \rightarrow 0$ at infinite dilution. Further deviations can be observed for large and spherical coarse-grained solvent beads such that the resulting excess volumes are often not correctly reproduced (Vögele et al., 2015a), which implies a significant influence on bulk thermodynamic properties like solubilities or isothermal compressibilities (Pierce et al., 2008; Smiatek et al., 2018).

Noteworthy, also the transfer free energies in ternary mixtures between the co-solvent "3" and the solute "2" as defined by $G^\dagger = N_{23}^{xs} - (\rho_3/\rho_1) N_{21}^{xs}$ rely on accurate values for the number densities and the excess numbers of molecules (Smiatek, 2017; Oprzeska-Zingrebe and Smiatek, 2018b) Otherwise, the thermodynamic affinity between the considered species is crucially affected. In order to highlight some further inconsistencies, it can be shown that also the chemical equilibrium between distinct chemical states in coarse-grained models differs from experimental values and atomistic approaches. In contrast to the chemical equilibrium constant K_0 in presence of a neat solute-solvent mixture, the modified chemical equilibrium constant K^* for denatured or native protein or DNA states (Oprzeska-Zingrebe and Smiatek, 2018a,b) or for associated and dissociated ion pairs (Krishnamoorthy et al., 2018) in presence of low co-solvent concentrations reads (Oprzeska-Zingrebe et al., 2019)

$$K^* = K_0 \exp(\Delta N_{23}^{xs}) \quad (3)$$

with $\Delta N_{23}^{xs} = N_{23}^{xs}(d) - N_{23}^{xs}(n)$ where d denotes the denatured and n the native state (Oprzeska-Zingrebe et al., 2019). With regard to the previous equation, a different value of ΔN_{23}^{xs} as obtained from the coarse-grained simulations when compared to the atomistic

model or experimental values ($\Delta N_{23,\text{exp}}^{\text{xs}}$) modifies the chemical equilibrium constant $K^* \neq K_{\text{exp}}^*$ and also the free energy difference in accordance with $\Delta G^* = -RT \ln K^* \neq \Delta G_{\text{exp}}^*$.

In consequence, incorrect sizes and geometries as well as simplified interactions or inaccurately parameterized coarse-grained interaction sites may induce significant deviations and spurious artifacts. A recent article revealed that specifically the number of interaction sites is of crucial importance (Dunn and Noid, 2015). Noteworthy, most deviations are only relevant for small molecular species like organic solvent molecules or ions, whereas significant improvements of coarse-grained models for polymers were recently reported (McCarty et al., 2014; Dunn and Noid, 2015; Vögele et al., 2015a,b; Guenza et al., 2018; Michalowsky et al., 2018).

In terms of these challenges, why should one use coarse-grained models at all? To answer this question, one should keep in mind that everything should be made as simple as possible, but not simpler. As already discussed, deviations between atomistic and coarse-grained models are mainly relevant for small molecular or ionic species where coarse-graining means a significant change of size and geometry. With regard to this point, it was recently shown that improvements in the parameterization strategy, the functional form of the interaction potentials as well as the consideration of polarizabilities in coarse-grained models increase the validity of the results (Noid, 2013; Rudzinski and Noid, 2014; Dunn and Noid, 2015; Michalowsky et al., 2017, 2018; Zeman et al., 2017; Guenza et al., 2018; Uhlig et al., 2018). With regard to this point, variations in thermodynamic properties become even visible for united- and all-atom models which highlights the importance of accurately parameterized molecular structures and interaction sites (Markthaler et al., 2017). Nevertheless, if the key features of interest can be reproduced through reduced models, nothing stands in the way of using these approaches. Otherwise, one must always be aware that uncontrollable artifacts may occur. In consequence, one may always keep the limits of the individual models in mind, such that the applicability of the approaches for certain research questions should be carefully reviewed.

AUTHOR CONTRIBUTIONS

EO-Z and JS wrote, reviewed, and edited all versions of this article.

ACKNOWLEDGMENTS

REFERENCES

Ben-Naim, A. (2013). *Statistical Thermodynamics for Chemists and Biochemists.* Berlin: Springer Science and Business Media.

Canchi, D. R., and García, A. E. (2013). Cosolvent effects on protein stability. *Ann. Rev. Phys. Chem.* 64, 273–293. doi: 10.1146/annurev-physchem-040412-110156

Chandler, D. (1987). *Introduction to Modern Statistical Mechanics.* Oxford: Oxford University Press.

Clark, A. J., McCarty, J., Lyubimov, I. Y., and Guenza, M. G. (2012). Thermodynamic consistency in variable-level coarse graining of polymeric liquids. *Phys. Rev. Lett.* 109:168301. doi: 10.1103/PhysRevLett.109.168301

Dommert, F., Wendler, K., Berger, R., Delle Site, L., and Holm, C. (2012). Force fields for studying the structure and dynamics of ionic liquids: a critical review of recent developments. *ChemPhysChem* 13, 1625–1637. doi: 10.1002/cphc.201100997

Dunn, N. J., and Noid, W. G. (2015). Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *J. Chem. Phys.* 143:243148. doi: 10.1063/1.4937383

Guenza, M., Dinpajooh, M., McCarty, J., and Lyubimov, I. (2018). Accuracy, transferability, and efficiency of coarse-grained models of molecular liquids. *J. Phys. Chem. B* 122, 10257–10278. doi: 10.1021/acs.jpcb.8b06687

Kirkwood, J. G., and Buff, F. P. (1951). The statistical mechanical theory of solutions. I. *J. Chem. Phys.* 19, 774–777. doi: 10.1063/1.1748352

Kleinjung, J., and Fraternali, F. (2014). Design and application of implicit solvent models in biomolecular simulations. *Curr. Opin. Struct. Biol.* 25, 126–134. doi: 10.1016/j.sbi.2014.04.003

Kohagen, M., Uhlig, F., and Smiatek, J. (2019). On the nature of ion-stabilized cytosine pairs in DNA i-motifs: the importance of charge transfer processes. *Int. J. Quant. Chem.* 119:e25933. doi: 10.1002/qua.25933

Krishnamoorthy, A. N., Holm, C., and Smiatek, J. (2018). The influence of co-solutes on the chemical equilibrium - a Kirkwood-Buff theory for ion pair association-dissociation processes in ternary electrolyte solutions. *J. Phys. Chem. C* 122, 10293–10392. doi: 10.1021/acs.jpcc.7b12255

Markthaler, D., Zeman, J., Baz, J., Smiatek, J., and Hansen, N. (2017). Validation of trimethylamine-n-oxide (TMAO) force fields based on thermophysical properties of aqueous TMAO solutions. *J. Phys. Chem. B* 121, 10674–10688. doi: 10.1021/acs.jpcb.7b07774

Marrink, S. J., and Tieleman, D. P. (2013). Perspective on the MARTINI model. *Chem. Soc. Rev.* 42, 6801–6822. doi: 10.1039/c3cs60093a

McCarty, J., Clark, A., Copperman, J., and Guenza, M. (2014). An analytical coarse-graining method which preserves the free energy, structural correlations, and thermodynamic state of polymer melts from the atomistic to the mesoscale. *J. Chem. Phys.* 140:204913. doi: 10.1063/1.4875923

Michalowsky, J., Schäfer, L. V., Holm, C., and Smiatek, J. (2017). A refined polarizable water model for the coarse-grained MARTINI force field with long-range electrostatic interactions. *J. Chem. Phys.* 146:054501. doi: 10.1063/1.4974833

Michalowsky, J., Zeman, J., Holm, C., and Smiatek, J. (2018). A polarizable MARTINI model for monovalent ions in aqueous solution. *J. Chem. Phys.* 149:163319. doi: 10.1063/1.5028354

Nandy, A., and Smiatek, J. (2019) Mixtures of LiTFSI and urea: ideal thermodynamic behavior as key to the formation of deep eutectic solvents? *Phys. Chem. Chem. Phys.* 21, 12279–12287. doi: 10.1039/C9CP01440C

Noid, W. G. (2013). "Systematic methods for structurally consistent coarse-grained models," in *Biomolecular Simulations. Methods in Molecular Biology, Vol. 924*, eds L. Monticelli and E. Salonen (Totowa, NJ: Humana Press), 487–531.

Onufriev, A. V., and Case, D. A. (2019). Generalized Born implicit solvent models for biomolecules. *Annu. Rev. Biophys.* 48, 275–296. doi: 10.1146/annurev-biophys-052118-115325

Oprzeska-Zingrebe, E. A., Kohagen, M., Kästner, J., and Smiatek, J. (2019). Unfolding of DNA by co-solutes: insights from Kirkwood–Buff integrals and transfer free energies. *Europ. Phys. J. Special Top.* 227, 1665–1679. doi: 10.1140/epjst/e2019-800163-5

Oprzeska-Zingrebe, E. A., Meyer, S., Roloff, A., Kunte, H. J., and Smiatek, J. (2018). Influence of compatible solute ectoine on distinct DNA

structures: thermodynamic insights into molecular binding mechanisms and destabilization effects. *Phys. Chem. Chem. Phys.* 20, 25861–25874. doi: 10.1039/C8CP03543A

Oprzeska-Zingrebe, E. A., and Smiatek, J. (2018a). Aqueous ionic liquids in comparison with standard co-solutes – differences and common principles in their interaction with protein and DNA structures. *Biophys. Rev.* 10, 809–822. doi: 10.1007/s12551-018-0414-7

Oprzeska-Zingrebe, E. A., and Smiatek, J. (2018b). Preferential binding of urea to single-stranded DNA structures:a molecular dynamics study. *Biophys. J.* 10:809. doi: 10.1016/j.bpj.2018.02.013

Pierce, V., Kang, M., Aburi, M., Weerasinghe, S., and Smith, P. E. (2008). Recent applications of Kirkwood-Buff theory to biological systems. *Cell. Biochem. Biophys.* 50, 1–22. doi: 10.1007/s12013-007-9005-0

Reith, D., Pütz, M., and Müller-Plathe, F. (2003). Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* 24, 1624–1636. doi: 10.1002/jcc.10307

Rudzinski, J. F., and Noid, W. G. (2014). Investigation of coarse-grained mappings via an iterative generalized Yvon–Born–Green method. *J. Phys. Chem. B* 118, 8295–8312. doi: 10.1021/jp501694z

Smiatek, J. (2014). Osmolyte effects: impact on the aqueous solution around charged and neutral spheres. *J. Phys. Chem. B* 118, 771–782. doi: 10.1021/jp410261k

Smiatek, J. (2017). Aqueous ionic liquids and their influence on protein conformations: an overview on recent theoretical and experimental insights. *J. Phys. Condens. Matter* 29:233001. doi: 10.1088/1361-648X/aa6c9d

Smiatek, J. (2019). Enthalpic contributions to solvent–solute and solvent–ion interactions: electronic perturbation as key to the understanding of molecular attraction. *J. Chem. Phys.* 150:174112. doi: 10.1063/1.5092567

Smiatek, J., Heuer, A., and Winter, M. (2018). Properties of ion complexes and their impact on charge transport in organic solvent-based electrolyte solutions for lithium batteries: insights from a theoretical perspective. *Batteries* 4:62. doi: 10.3390/batteries4040062

Smiatek, J., and Holm, C. (2018). "From the atomistic to the macromolecular scale: distinct simulation approaches for polyelectrolyte solutions," in *Handbook of Materials Modeling*, eds W. Andreoni and S. Yip (Heidelberg; New York, NY: Springer), 1–15.

Smiatek, J., and Schmid, F. (2011). Mesoscopic simulations of electroosmotic flow and electrophoresis in nanochannels. *Comput. Phys. Commun.* 182, 1941–1944. doi: 10.1016/j.cpc.2010.11.021

Smith, P. E. (2006). Chemical potential derivatives and preferential interaction parameters in biological systems from Kirkwood-Buff theory. *Biophys. J.* 91, 849–856. doi: 10.1529/biophysj.105.078790

Sukenik, S., Sapir, L., Gilman-Politi, R., and Harries, D. (2013). Diversity in the mechanisms of cosolute action on biomolecular processes. *Farad. Discuss.* 160, 225–237. doi: 10.1039/C2FD20101A

Uhlig, F., Zeman, J., Smiatek, J., and Holm, C. (2018). First-principles parametrization of polarizable coarse-grained force fields for ionic liquids. *J. Chem. Theory Comput.* 14, 1471–1486. doi: 10.1021/acs.jctc.7b00903

Vögele, M., Holm, C., and Smiatek, J. (2015a) Coarse-grained simulations of polyelectrolyte complexes: MARTINI models for poly(styrene sulfonate) and poly(diallyldimethylammonium). *J. Chem. Phys.* 143:243151. doi: 10.1063/1.4937805

Vögele, M., Holm, C., and Smiatek, J. (2015b). Properties of the polarizable MARTINI water model: a comparative study for aqueous electrolyte solutions. *J. Mol. Liquids* 212:103. doi: 10.1016/j.molliq.2015.08.062

Zeman, J., Uhlig, F., Smiatek, J., and Holm, C. (2017) A coarse-grained polarizable force field for the ionic liquid 1-butyl-3-methylimidazolium hexafluorophosphate. *J. Phys. Condens. Matt.* 29:504004. doi: 10.1088/1361-648X/aa99c4

Zhang, Y., and Cremer, P. S. (2010). Chemistry of Hofmeister anions and osmolytes. *Ann. Rev. Phys. Chem.* 61, 63–83. doi: 10.1146/annurev.physchem.59.032607.093635

Zhou, H. X., Rivas, G., and Minton, A. P. (2008). Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* 37, 375–397. doi: 10.1146/annurev.biophys.37.032807.125817