Design and application of implicit solvent models in biomolecular simulations

Jens Kleinjung[1] and Franca Fraternali[2]

We review implicit solvent models and their parametrisation by introducing the concepts and recent devlopments of the most popular models with a focus on parametrisation *via* force matching. An overview of recent applications of the solvation energy term in protein dynamics, modelling, design and prediction is given to illustrate the usability and versatility of implicit solvation in reproducing the physical behaviour of biomolecular systems. Limitations of implicit modes are discussed through the example of more challenging systems like nucleic acids and membranes.

Addresses
[1] Division of Mathematical Biology, MRC National Institute for Medical Research, The Ridgeway, London NW7 1AA, United Kingdom
[2] Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, London SE1 1UL, United Kingdom

Corresponding author: Fraternali, Franca (franca.fraternali@kcl.ac.uk)

http://dx.doi.org/10.1016/j.sbi.2014.04.003

Introduction

It has been 37 years since the first molecular dynamics (MD) simulations of the protein BPTI *in vacuo* has been published [1]. In the meantime the technique has progressed appreciably in applicability and efficiency. Three of the founders of the field, namely Martin Karplus, Michael Levitt and Arieh Warshel, were awarded the Nobel Prize in Chemistry 2013 for the ''development of multi-scale models of complex chemical systems'' [2,3]. Within a few decades MD simulations have become an essential and complementary tool for the structural and conformational study of biomolecules [4]. Nevertheless, efficient and accurate multi-scale modelling of complex biochemical systems remains a challenging task.

Many central aspects of biomolecular simulations have undergone rapid advancements: the reproducibility of a biomolecular system's representation by classical force fields, the combined treatment of reactions and dynamics *via* QM/MM calculations and the use of coarse-grained models for very large systems.
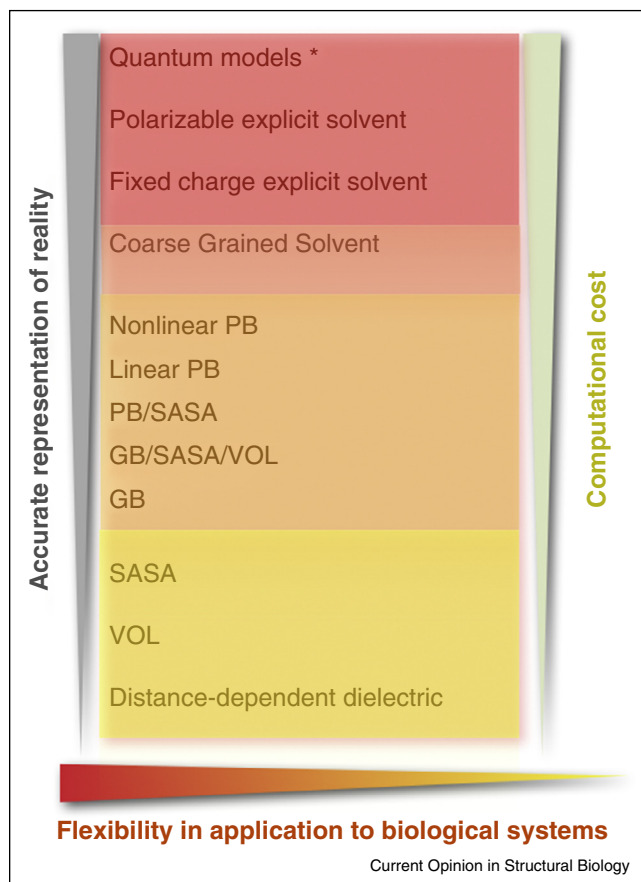
Multi-scale approaches are designed to bridge spatial and temporal scales by merging models with different resolution. The underlying physical descriptions can also vary with scale: first, the quantum mechanical treatment of degrees of freedom at (sub-)atomic scales, second, the classical representation of discrete atoms, molecules or groups of molecules at intermediate scales, and third, the mesoscopic or continuum treatment of entire molecular systems (Figure 1).

A multi-scale model should be able to balance accuracy and efficiency and integrate techniques with different resolution to provide a realistic description of the system's behaviour. A good molecular model can be directly used for experimental testing and/or validation. Interestingly, one of the first methods in the spirit of multiscaling was Warshel's pioneering approach to the development of solvent models [5].

Modelling water is essential for the description of cellular biopolymers, as they have evolved to function in an environment containing predominantly water, enriched with ions, proteins, lipids and sugars. The description of the polyfunctional nature of each of these components is intrinsically difficult and can only be partially formalised for *in silico* applications. Most biomolecular simulations account for the presence of the native (cellular) environment either explicitly, for example by inclusion of water molecules and ions, or implicitly, by approximating the mean force (Potential of Mean Force, PMF) exerted by the external media on the biomolecule. Implicit water models can be considerably faster to compute, because the implicit solvent contributes no or few degrees of freedom to the simulation. However, they neglect specific important features such as hydrogen bond fluctuations at the solute surface, water dipole reorientation in response to conformational changes and bridging water molecules. Generally implicit solvent tends to be a good approximation where it models isotropic and bulk solvent.

To sample efficiently the dynamics of a system with many degrees of freedom or along large time scales, coarse-graining by bundling degrees of freedom into pseudo-atoms is a commonly used technique. This transformation smoothens the underlying energy hyper-surface and speeds up considerably energy and force calculations [6–9]. Typical strategies for multi-scaling are capturing the physical effects of the presence of solvent *via* an intrinsic term that is directly related to the solvation free energy of transfer of a rigid solute from vacuum to

Implicit solvation models.

We will first briefly introduce the most frequently used models and then focus on parametrisations *via* force matching and applications that require fast estimates of the free solvation energy.

Models of implicit solvation
Free energy of solvation
The free energy of solvation is traditionally decomposed into three components that relate to a purely theoretical process: creation of a cavity within the solvent to accommodate the solute molecule (ΔG_{cav}) and embedding of the solute molecule into the cavity leading to van der Waals (ΔG_{vdW}) and electrostatic (ΔG_{ele}) interactions between solute and solvent.

$$\Delta G_{sol} = \Delta G_{cav} + \Delta G_{vdW} + \Delta G_{ele} \qquad (1)$$

Solvent-Accessible Surface Area methods model either the nonpolar terms $\Delta G_{cav} + \Delta G_{vdW}$ or the entire ΔG_{sol} term, depending on their parametrisation. Poisson–Boltzmann and Generalized-Born methods model the ΔG_{ele} term.

Solvent-accessible surface area (SASA) models
The first SASA parametrisations were performed by Eisenberg and McLachlan on the basis of the free energy of transfer of amino acids between octanol and water [10••] and by Ooi *et al.* on seven chemical groups whose thermodynamic solvation parameters were derived from a benchmark set of small molecules [24•].

Implicit solvation models on the basis of SASA assume the interactions between solute and solvent to be proportional to the surface area. Therefore, the free energy of solvation ΔG_{sol} of a solute molecule is described by a mean solvation potential V_{solv}^{SASA}. The term is computed as the product of an atom-specific solvation energy per surface area σ_i^{SASA} and the atomic $SASA_i$, summed over all atoms i.

$$V_{solv}^{SASA}(\vec{r}) = \sum_i \sigma_i^{SASA} \cdot SASA_i(\vec{r}_i) \qquad (2)$$

In practice, σ_i^{SASA} is either used in conjunction with a force field for vacuum simulations in which charged atom groups are parametrised as polar entities with nonzero total charge [11••,25••] or to model exclusively the nonpolar solute–solvent interactions in conjunction with a PB or GB model, in the following denoted as PB/SASA or GB/SASA (see below). To include long-range effects of the solvent on the solute's interior, the SASA solvation model has been further refined by a volume term [26] using a SASA/VOL switch function $g(\vec{r}_i)$

$$V_{solv}^{VOL}(\vec{r}) = \sum_i \sigma_i^{VOL} \cdot g(\vec{r}_i) \cdot \frac{4}{3}\pi R_i^3 \qquad (3)$$

and by recalibrated contributions of cavity formation and van der Waals interactions [27,28]. Several algorithms exist to compute the molecular SASA, spanning from

solution or by parametrising the implicit solvent by matching solute properties like explicit water forces or conformational ensembles [11••,12••]. In recent years, large-scale atomistic simulations in different types of solvent have been performed and the trajectories have been made publicly available [13–15]. These data can be fruitfully used for the parametrisation of coarse-grained models and implicit solvent models.

Some simplified and fast implicit solvation models use a first-shell approximation of the solvent effect, under the assumption that the forces exerted on a solute atom by the solvent are proportional to the solvent-accessible surface area (SASA) of the solute atom. This assumption holds for the nonpolar part of the solute-solvent interactions, which is complemented by a contribution that models electrostatic interactions, on the basis of the approximation that the bulk solvent behaves as a dielectric continuum. The electrostatic term is either integrated into the parametrisation of the SASA model or treated by a separate energy term.

Previous reviews on implicit solvent models have been published in this journal [16–18] and in others [19–23].

exact to analytical approximations. An overview of the analysis of protein geometry can be found in Gerstein and Richards [29]. Fast methods for the SASA computation of proteins and nucleic acids are implemented in POPS [30,31] and POPSCOMP [32], a related application for the SASA analysis of biomolecular complexes. Instead of a SASA term, solely the volume of neighbour groups can be used as a measure of desolvation, as in the EEF1 [33••] and EEF1-SB models [12••].

$$\Delta G_i^{sol} = \Delta G_i^{ref} - \sum_{j \neq i} f_i(\vec{r}_{ij}) \cdot VOL_j, \tag{4}$$

where ΔG_i^{ref} denotes the free energy of the fully solvated group i and the negative term the decrease in solvation free energy due to desolvation by neighbouring groups j.

Poisson–Boltzmann
The Poisson equation relates the scalar electric potential Φ to the charge density ρ.

$$\vec{\nabla} \cdot [\in (\vec{r}) \vec{\nabla} \Phi(\vec{r})] = -\frac{\rho(\vec{r})}{\in_0}, \tag{5}$$

where ε is the local dielectric constant and ε_0 the vacuum permittivity.

In practice the charge density may be obtained by integration of the Coulomb potential. The solution to the Poisson equation for a given charge density reads

$$\Phi(\vec{r}) = \frac{1}{4\pi \in_0 \in} \int \frac{\rho(\vec{r})}{|\vec{r}|} dV(\vec{r}). \tag{6}$$

The Poisson–Boltzmann (PB) equation is a special case of the Poisson equation. At equilibrium the charge density is expected to follow a Boltzmann energy distribution

$$\rho_B(\vec{r}) = \sum_i c_i^b z_i q e^{-z_i q \Phi(\vec{r})/k_B T}, \tag{7}$$

where c_i denotes the bulk concentration of ion i, z_i its valence, q the unit charge, k_B Boltzmann's constant and T the temperature.

Eq. (5) applied to the charge density ρ_B yields the PB equation.

$$\vec{\nabla} \cdot [\in (\vec{r}) \vec{\nabla} \Phi(\vec{r})] = -\rho_B(\vec{r}) - \sum_i c_i^b z_i q e^{-z_i q \Phi(\vec{r})/k_B T} \tag{8}$$

Being a differential equation, it is usually solved numerically. Given the exponential term is small enough $(e^{-z_i q \Phi(\vec{r})/k_B^T} \ll 1)$, it can be approximated by the linear expression $(1 - z_j q \Phi(\vec{r})/k_B T)$, yielding the linearised PB equation.

$$\vec{\nabla} \cdot [\in (\vec{r}) \vec{\nabla} \Phi(\vec{r})] = -\rho_B(\vec{r}) + \sum_i c_i^b z_i^2 \frac{q^2 \Phi(\vec{r})}{k_B T} \tag{9}$$

Generalized Born
The Generalized Born (GB) equation was introduced by Still *et al.* [34••]. The Born model is a solution of the PB equation (see previous section) for a charge in the centre of an ideally spherical solute with radius α and internal dielectric ε_{int} in a solvent with dielectric ε_{ext}:

$$\Delta G_{solv} = -\frac{1}{2} \left(\frac{1}{\in_{int}} - \frac{1}{\in_{ext}} \right) \frac{q^2}{\alpha}. \tag{10}$$

In the GB method a pseudo-ideal situation is emulated locally in non-ideal solutes (like biomolecules) by variation of α. By fixing the (computed or measured) value of ΔG_{solv} in Eq. (10), the undetermined radius α can be varied as control parameter to match the given ΔG_{solv}. The resulting 'effective Born radius' α adjusts the local screening to an optimal value modulo approximations of the GB theory. The effective Born radii are used for the computation of the GB pair terms between charged atoms or atom groups i and j:

$$\Delta G_{solv} = -\frac{1}{2} \left(\frac{1}{\in_{int}} - \frac{1}{\in_{ext}} \right) \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j e^{-r_{ij}^2/4\alpha_i \alpha_j}}}. \tag{11}$$

Zhu *et al.* [35] implemented different GB methods and tested those in conjunction with the force fields GROMOS and OPLSAA.

Several GB methods have recently been implemented in the GROMACS package [36–38]. The GBn method of Mongan *et al.* [39•], which includes a volume correction for regions of interstitial solvent exclusion, is available in the AMBER package. An efficient parallelised GB/SASA algorithm has been implemented as a hybrid GPU/CPU application in the package NAMD [40].

Combined PB/SASA, GB/SASA and related methods
Many biomolecular applications require a precision that demands combined PB/SASA and GB/SASA methods with complementary free energy contributions. In a recent study [41•] the methods HTC [37] and OBC [38] implemented in AMBER and the methods GBMV [42], GBSW [43] and FACTS (Fast Analytical Continuum Treatment of Solvation) [44•] implemented in CHARMM were compared.

Parametrisation *via* force matching
Several options exist for the parametrisation of implicit solvents: first, matching the values of a reference method as in the parametrisation of the GB model to the more precise results of the PB model [36]; second, matching the known (free) energy differences of given reference states, for example the transfer free energies of small solutes [10••] or other relevant physico-chemical observables; third, reproducing known ensembles of the solute or the solvent, which implicitly reproduces the free energy landscape under the constraint that the conformational space has been represented realistically [25••,45•]; fourth, matching the solvation forces of molecular trajectories in explicit solvent [11••,12••, 46].

The availability of large-scale simulation trajectories provides statistically meaningful force distributions that can be utilised to derive robust solvation parameters. A force-matching parametrisation (fourth above) is given in the following analytical formula for the determination of an atom-specific solvation parameter σ_i^{SASA} for a SASA model [11••] (see also Eq. (2)):

$$\sigma_i^{SASA} = -\frac{\partial A_i / \partial r_i}{|\partial A_i / \partial r_i|^2} \cdot \langle f_i^{exp/} \rangle \qquad (12)$$

The term $\partial A_i / \partial r_i$ denotes the distance-dependent change of the exposed surface area and serves as projection direction for the explicit solvation force $\langle f_i^{exp/} \rangle$. In practice, two sets of simulations are required to perform the parametrisation: a short Langevin simulation in implicit solvent to extract the $\partial A_i / \partial r_i$ values and a long conformationally constrained simulation in explicit water to obtain the solvation forces, both using identical starting conformations to ensure correct projection geometry.

In an ensemble-matching parametrisation (third above), Bottaro et al. [46] used replica exchange MD simulations of the helical peptide $(AAQAA)_3$ and the hairpin peptide GB1. Two parameters, a charge screening parameter and a backbone torsion term, were parametrised by minimising the Kullback–Leibler divergence between the configurational ensembles in explicit and implicit solvent. Because the Kullback–Leibler divergence is an information-theoretic metric, it is applicable to any type of distribution and therefore easily generalizable to other properties or combined properties.

An entirely different approach is the design of solvent-free coarse-grained PMFs, where the solvent contributions are included in the potential by integrating out solvent degrees of freedom during the parametrisation process [46,47]. The applied multiscale coarse-graining (MS-CG) method [48] fits observed forces in atomistic simulations by spline functions along a radial mesh connecting atoms. By employing a variational force-matching method, the residual between the effective atomistic and (to be parametrised) coarse-grained forces are minimised, thereby determining the linear fitting parameters of the coarse-grained force field.

Use of energy terms on the basis of implicit solvation
We will focus here on the use of fast evaluation of the solvation energy term in protein dynamics, modelling, design and prediction.

Protein modelling and design
Implicit solvent models like GB/SA coupled to energy minimisation have been proven to be more effective than MD in water [49] on a large-scale dataset with decoy models [50•]. Scheraga and co-workers applied their ECEPP05/SA force field combined to Monte-Carlo runs

to discriminate native-like from non-native conformations in a set of decoys taken from the test set of the Rosetta@home all-atom decoys [51]. This approach identified successfully near-native conformations with a performance comparable with other existing physics-based scoring functions with computationally more expensive solvent models.

The Analytical Generalized Born plus Non-Polar (AGBNP) implicit solvent model has been shown to be effective in the prediction of the native conformations of medium-long loops (9–13 residues) for proteins with low sequence identity. The nonpolar solvation free energy predictor implemented in AGBNP with appropriate correction terms was crucial to attain the best prediction accuracy [49,52]. Recently, the group of Lavery [53] employed a simple geometrical measure, circular variance, to reflect residue burial (value range [0,1]) by the spatial distribution of neighbouring residues. This measure was used to build a very fast and effective model of protein solvation that was used to discriminate between native and decoy conformations.

In protein design implicit solvation is the method of choice for sidechain placement and mutant assessment. Lopes and Simonson [54] compared a parametrised SASA model [25••] to two GB models and showed that the former predicts the correct sign and order of magnitude of the stability change of mutants with a marginally higher error compared to the more accurate GB method. The same group developed a combined Coulomb/Accessible Surface Area (CASA) implicit solvent model, implemented it within the CHARMM19 force field and found that this combination was suitable for the computational engineering of ligands and proteins [55]. The group of Mayo has developed a Poisson–Boltzmann (FDPB) model ([56] and references therein) with a simplified surface term that is solely dependent on the identity and conformation of the protein backbone. This method has led to stable designed proteins in vitro and in silico.

Protein dynamics: conformational equilibria and protein folding
Most of the implicit solvation models implemented in MD packages are routinely tested for their ability to reproduce conformational equilibria in explicit solvent and the cognate ensemble average properties (rmsd, rmsf, radius of gyration, exposed and buried atomic areas) [25••,26,32,33••,57]. Nevertheless, some challenging tests for implicit solvation can be found in seemingly simpler applications, like the reproduction of the conformational sampling of small dipeptides [58]. The group of McCammon reported the conformational sampling of alanine dipeptide in solution [45•] performed by means of the extended reference interaction site model (XRISM) integral equation theory. This approach is more detailed than a GB/SASA model, because the properties of the

hydration structure can refine the PMF. The obtained Ramachandran plot is remarkably close to the one derived from simulations in explicit water and the typically observed solution minima basins are all densely populated. This model allows for the inclusion of interesting features like salt concentration effects, that is for the modelling of physiological conditions.

Significant progress has been made in the folding studies of small structures containing β-hairpins [59] or fragments from a series of very well characterised proteins [60] by means of replica exchange MD in GB/SASA solvent. A similar success has been achieved by Zagrovic and Pande within the Folding@home project [61]. Pande's group has also successfully folded [62] *ab initio* small fast-folder proteins using a GB/SASA/AMBERf99 parametrisation. In addition to revealing mechanisms of the folding process, these simulations sampled native-like and near-native conformations that show heterogeneity in their hydrophobic packing, mostly due to alternative side chain arrangements.

While β-sheets and β-hairpins seem to be easier to fold in implicit solvent, some inefficacies of the GB/SASA model in dealing with the folding equilibrium of α-helices have been highlighted by Nymeyer and García [63], where unphysical nucleation parameters were observed for implicit solvent models with respect to explicit water.

Despite the outlined progress, folding simulations in implicit solvent remain challenging for single helical structures or small protein folds that do not have a defined hydrophobic core. In a conceptually novel development, Duan *et al.* [64] have coupled discrete on-the-fly charge fitting schemes (Hydrogen Bond-specific Charge, HBC) with AMBER GB/SASA for the folding of a helical 17-residue peptide in a single 16 ns trajectory.

Challenging applications: large-scale screening, binding free energy, nucleic acids and membranes

Large-scale screening

Implicit solvation models play a central role in large-scale screening applications like the combinatorial assembly of protein-protein complexes and the estimation of the free energy of complexation. SASA calculations, if appropriately parametrised, provide a semi-quantitative estimate of the desolvation free energy upon protein complexation [32] and allow for interface prediction [65–67]. In docking score functions for protein complex prediction, the solvation term is often embedded in the electrostatic term [68], otherwise it is implemented as a separate implicit solvation term [33••] as in the Rosetta Docking program [69].

Binding free energy

A challenging application for explicit and implicit solvation is the calculation of accurate binding free energies. The group of Essex has performed extensive comparisons between explicit solvation simulations and a GBSA implicit solvent model [70]. Despite some initial success [71], later results showed that explicit solvation leads to more accurate results [72].

Implicit-solvent based alchemical perturbation gives access to the configurational entropy, resulting in improved prediction of the binding free energy [73].

Overall, MD simulations in implicit solvent represent a viable route for medium to high-throughput computational drug discovery. In large-scale screening studies for structure-based drug design, even quite simple continuum models have been used successfully [74].

Nucleic acids

Nucleic acids are a difficult target for the parametrisation of implicit solvent, due to their highly charged backbone and the required modelling of associated counter-ions. Most of the currently available implicit solvent models are unsuitable for MD simulations of structure and dynamics of nucleic acids. Instead, docking methods applied to RNA molecules have included implicit solvent models such as GBSA, PBSA and DOCK6 [75] and have devised useful strategies for effective structure-based RNA drug design.

An interesting review of the relationship between force field parametrisation (namely AMBER and CHARMM) and GB implicit solvent methods in simulating DNA and in extracting characteristic structural features is given by Gaillard and Chase [76•]. The work shows that GB methods have considerably improved over the last years and are approaching explicit solvation if judged by structural and dynamical features of the resulting ensembles. Developments in the design of GB/SASA and PB/SASA methods for RNA MD simulation by Liu *et al.* show that combining PB with a Langevin–Debye Model for the induced dielectric response of solvent and counter-ions improves the structural stability of the solute [77••]. The authors suggest that further improvement can be obtained by deriving the dielectric function with an iterative approach and by including explicitly more information about the topology and charge distribution of the solute.

Membranes

One of the most challenging design problem for implicit solvent models is posed by the lipid-membrane environment, mostly due to the anisotropic nature of the medium and the associated contrasting hydrophobic/charged protein–lipid interactions depending on the relative positioning of the solute and the membrane. Ben-Tal, one of the pioneers, described the thermodynamics of an α-helix insertion while representing the membrane as a simple weakly dielectric slab [78]. Lazaridis' IMM1 model [79]

uses a generalization of the EEF1 solvation model [33[••]] to a membrane environment.

Implicit membrane models on the basis of the GB theory for ionic solvation have also been employed [43], extending previous work by the Brooks group [80], where a smoothing function was used to improve the numeric behaviour of the volume integration. Similarly, Spassov *et al.* [81[•]] considered the membrane to be part of the protein interior and analytic corrections to the Coulomb field term designed for soluble proteins were implemented. However, the assumption that the dielectric character of the membrane and of the protein are identical is rather unsatisfactory. Feig and coworkers developed a procedure to handle multiple dielectric environments [82[••],83], the heterogeneous dielectric Generalized Born (HDGB) model, and applied it to implicit membrane modelling. Consequently, the membrane representation reproduces correctly the chemical heterogeneity of the membrane–water interface; a series of dielectric slabs, rather than just two regions, were used in the model. The PB method is used by many research groups to describe the membrane–protein association thermodynamics. A detailed discussion is provided in the review by Grossfield [84].

Conclusions

Implicit solvation is gaining momentum in computational modelling and simulations of biomolecules due to the rapid progress in the development of multi-scale approaches and large-scale analysis of protein structures for use in protein design and targeted drug design. The diversity of models and implementations reflects the various requirements for the balance between time complexity of the computation and accuracy of the results. Details of the interactions between biomolecules and water carry valuable information for structural analysis [85,86]. A desirable feature of future implicit models solvent would be the possibility to preserve some of this information.

Working through the literature about implicit solvation, one cannot fail to note the absence of benchmark sets. It is common practice in most areas of computational biology to test novel methods on established benchmarks, for example the BaliBase benchmark for sequence alignment and the CASP targets for protein structure prediction. A benchmark set for implicit solvation would facilitate the cross-comparison between methods and provide a quantitative picture of the relationship between accuracy and computational expense. With the availability of repositories for MD trajectories, one could devise the development of reference test sets for controlled benchmarking of novel solvent models or parametrisations.

In the wake of genomics and proteomics, disease-related simulations and predictions can find a place as a diagnostic tool set, for example in the analysis of mutational effects and their influence on disease traits. Given the complexity of biomolecular structures and interactions, providing consistently reliable results within the diagnostic time constraints of a few days is beyond the scope of virtually any current computational method. It is however foreseeable that implicit solvent treatment will play a central role in such methods, because of their flexible applicability and tunable performance in multi-scale approaches to biomolecular characterisation and design.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. McCammon JA, Gelin BR, Karplus M: **Dynamics of folded proteins**. *Nature* 1977, **267**:585-590.

2. Jorgensen WL: **Foundations of biomolecular modeling**. *Cell* 2013, **155**:1199-1202.

3. Hodak H: **The Nobel Prize in Chemistry 2013 for the development of multiscale models of complex chemical systems: a tribute to Martin Karplus, Michael Levitt and Arieh Warshel**. *J Mol Biol* 2014, **426**:1-3.

4. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glättli A, Hünenberger PH et al.: **Biomolecular modeling: goals, problems, perspectives**. *Angew Chem (Intl Ed)* 2006, **45**:4064-4092.

5. King G, Warshel A: **A surface constrained all-atom solvent model for effective simulations of polar solutions**. *J Chem Phys* 1989, **91**:3647.

6. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ: **The MARTINI coarse-grained force field: extension to proteins**. *J Chem Theor Comp* 2008, **4**:819-834.

7. Kar P, Gopal SM, Cheng YM, Predeus A, Feig M: **PRIMO: a transferable coarse-grained force field for proteins**. *J Chem Theor Comp* 2013, **9**:3769-3788.

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

9. Riniker S, van Gunsteren WF: **Mixing coarse-grained and fine-grained water in molecular dynamics simulations of a single system**. *J Chem Phys* 2012, **137**:044120.

10. •• Eisenberg D, McLachlan AD: **Solvation energy in protein folding and binding**. *Nature* 1986, **319**:199-203.
The linear correlation between SASA and solvation free energy of acetyl amino-acid amides is shown. This relationship is the foundation of SASA-based implicit solvent models.

11. •• Kleinjung J, Scott WRP, Allison JR, van Gunsteren WF, Fraternali F: **Implicit solvation parameters derived from explicit water forces in large-scale molecular dynamics simulations**. *J Chem Theor Comp* 2012, **8**:2391-2403.
Parametrisation of a SASA model using an analytical force-matching approach. The large-scale study was performed on 188 topologically diverse protein structures.

12. •• Bottaro S, Lindorff-Larsen K, Best RB: **Variational optimization of an all-atom implicit solvent force field to match explicit**

solvent simulation data. *J Chem Theor Comp* 2013, **9**:5641-5652.

Force-matching parametrisation on the basis of the Kullback–Leibler divergence between configurational ensembles of an α-helix and a β-strand peptide in explicit and implicit solvent (EFF1 in CHARMM). The method is tested using NMR data of intrinsically disordered proteins.

13. Tai K, Murdock S, Wu B, Ng MH, Johnston S, Fangohr H, Cox SJ, Jeffreys P, Essex JW, Sansom MS: **BioSimGrid: towards a worldwide repository for biomolecular simulations**. *Org Biomol Chem* 2004, **2**:3219-3221.

14. Meyer T, D'Abramo M, Hospital A, Rueda M, Ferrer-Costa C, Pérez A, Carrillo O, Camps J, Fenollosa C, Repchevsky D *et al.*: **MoDEL (Molecular Dynamics Extended Library): a database of atomistic molecular dynamics trajectories**. *Structure* 2010, **18**:1399-1409.

15. Candotti M, Pérez A, Ferrer-Costa C, Rueda M, Meyer T, Gelp'i JL, Orozco M: **Exploring early stages of the chemical unfolding of proteins at the proteome scale**. *PLoS Comput Biol* 2013, **9**:e1003393.

16. Feig M, Brooks CL: **Recent advances in the development and application of implicit solvent models in biomolecule simulations**. *Curr Opin Struct Biol* 2004, **14**:217-224.

17. Baker NA: **Improving implicit solvent simulations: a Poisson-centric view**. *Curr Opin Struct Biol* 2005, **15**:137-143.

18. Chen J, Brooks CL, Khandogin J: **Recent advances in implicit solvent-based methods for biomolecular simulations**. *Curr Opin Struct Biol* 2008, **18**:140-148.

19. Dill KA, Truskett TM, Vlachy V, Hribar-Lee B: **Modeling water, the hydrophobic effect, and ion solvation**. *Annu Rev Biophys Biomol Struct* 2005, **34**:173-199.

20. Tomasi J, Mennucci B, Cammi R: **Quantum mechanical continuum solvation models**. *Chem Rev* 2005, **105**:2999-3093.

21. Prabhu N, Sharp K: **Protein–solvent interactions**. *Chem Rev* 2006, **106**:1616-1623.

22. Onufriev A: **Implicit solvent models in molecular dynamics simulations: a brief overview**. *Annu Rep Comput Chem* 2008, **4**:125-137.

23. Ren P, Chun J, Thomas DG, Schnieders MJ, Marucho M, Zhang J, Baker NA: **Biomolecular electrostatics and solvation: a computational perspective**. *Quart Rev Biophys* 2012, **45**:427-491.

24. Ooi T, Oobatake M, Némethy G, Scheraga HA: **Accessible**
• **surface areas as a measure of the thermodynamic parameters of hydration of peptides**. *Proc Natl Acad Sci USA* 1987, **84**:3086-3090.

The authors present a SASA-based computation of the free energy and free enthalpy of hydration of acetyl amino-acid amides.

25. Fraternali F, van Gunsteren W: **An efficient mean solvation force**
•• **model for use in molecular dynamics simulations of proteins in aqueous solution**. *J Mol Biol* 1996, **256**:939-948.

An analytical approximation to the SASA solvation model is introduced and parametrised. The method is shown to improve the accuracy of MD simulations in GROMOS. The same parametrisation was later used by the Caflish group in conjunction with the CHARMM force field.

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

27. Gallicchio E, Levy RM: **AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling**. *J Comput Chem* 2004, **25**:479-499.

28. Zacharias M: **Continuum solvent modeling of nonpolar solvation: improvement by separating surface area dependent cavity and dispersion contributions**. *J Phys Chem A* 2003, **107**:3000-3004.

29. Gerstein M, Richards FM: **Protein geometry: volumes, areas, and distances**. *Int Tables Cryst* 2001, **22**:531-539.

30. Fraternali F, Cavallo L: **Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome**. *Nucleic Acids Res* 2002, **30**:2950-2960.

31. Cavallo L, Kleinjung J, Fraternali F: **POPS: a fast algorithm for solvent accessible surface areas at atomic and residue level**. *Nucleic Acids Res* 2003, **31**:3364-3366.

32. Kleinjung J, Fraternali F: **POPSCOMP: an automated interaction analysis of biomolecular complexes**. *Nucleic Acids Res* 2005, **33**:W342-W346.

33. Lazaridis T, Karplus M: **Effective energy function for proteins in**
•• **solution**. *Proteins* 1999, **35**:133-152.

Introduction of the EEF1 model, which computes the reduction of the solvation free energy proportional to the deshielding volume of neighbouring groups.

34. Still WC, Tempczyk A, Hawley RC, Hendrickson T: **Semianalytical**
•• **treatment of solvation for molecular mechanics and dynamics**. *J Am Chem Soc* 1990, **112**:6127-6129.

Introduction of the GB equation as a heuristic solution to the free energy of solvation. The numerical computation of Born radii and the excellent fit with explicit solvent calculations are described.

35. Zhu J, Alexov E, Honig B: **Comparative study of generalized Born models: Born radii and peptide folding**. *J Phys Chem B* 2005, **109**:3008-3022.

36. Qiu D, Shenkin PS, Hollinger FP, Still WC: **The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii**. *J Phys Chem A* 1997, **101**:3005-3014.

37. Hawkins GD, Cramer CJ, Truhlar DG: **Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium**. *J Phys Chem* 1996, **100**:19824-19839.

38. Onufriev A, Bashford D, Case DA: **Exploring protein native states and largeâscale conformational changes with a modified generalized born model**. *Proteins* 2004, **55**:383-394.

39. Mongan J, Simmerling C, McCammon JA, Case DA,
• Onufriev A: **Generalized Born model with a simple, robust molecular volume correction**. *J Chem Theor Comp* 2007, **3**:156-169.

A correction to the excluded solvent model is described, leading to improved effective Born radii.

40. Tanner DE, Phillips JC, Schulten K: **GPU/CPU algorithm for generalized Born/solvent-accessible surface area implicit solvent calculations**. *J Chem Theor Comp* 2012, **8**:2521-2530.

41. Knight JL, Brooks CL: **Surveying implicit solvent models for**
• **estimating small molecule absolute hydration free energies**. *J Comp Chem* 2011, **32**:2909-2923.

A comparison between 9 implicit solvation models applied to a database of 499 small organic molecules. The results show overall a good correlation between the computed hydration free energies.

42. Lee MS, Feig M, Salsbury FR, Brooks CL: **New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations**. *J Comp Chem* 2003, **24**:1348-1356.

43. Im W, Lee MS, Brooks CL: **Generalized Born model with a simple smoothing function**. *J Comp Chem* 2003, **24**:1691-1702.

44. Haberthür U, Caflisch A: **FACTS: fast analytical continuum**
• **treatment of solvation**. *J Comp Chem* 2008, **29**:701-715.

A measure of solvent displacement is computed on the basis of analytical functions for each van der Waals radius and a sigmoidal function is used to fit effective Born radii to the free solvation energy.

45. Ishizuka R, Huber GA, McCammon JA: **Solvation effect on the**
• **conformations of alanine dipeptide: integral equation approach**. *J Phys Chem Lett* 2010, **1**:2279-2283.

The extended reference interaction site model (XRISM) integral equation theory is coupled to Brownian dynamics and is applied to the conformational sampling of the model dipeptide alanine. Salt effects on the conformational sampling are included in the PMF calculation. The results match closely the corresponding explicit water simulation map.

46. Cao Z, Dama JF, Lu L, Voth GA: **Solvent free ionic solution models from multiscale coarse-graining**. *J Chem Theor Comp* 2013, **9**:172-178.

47. Hills RDJ, Lu L, Voth GA: **Multiscale coarse-graining of the protein energy landscape**. *PLoS Comput Biol* 2010, **6**:e1000827.

48. Izvekov S, Voth GA: **Solvent-free lipid bilayer model using multiscale coarse-graining**. *J Phys Chem B* 2009, **113**:4443-4455.

49. Friesner RA, Abel R, Goldfeld DA, Miller EB, Murrett CS: **Computational methods for high resolution prediction and refinement of protein structures**. *Curr Opin Struct Biol* 2013, **23**:177-184.

50. Chopra G, Summa CM, Levitt M: **Solvent dramatically affects
• protein structure refinement**. *Proc Natl Acad Sci* 2008, **105**:20239-20244.
The authors demonstrate on a set of 75 proteins and 729 near-native decoys that implicit solvation helps in minimising structures towards their native conformations.

51. Arnautova YA, Vorobjev YN, Vila JA, Scheraga HA: **Identifying native-like protein structures with scoring functions based on all-atom ECEPP force fields, implicit solvent models and structure relaxation**. *Proteins* 2009, **77**:38-51.

52. Felts AK, Gallicchio E, Chekmarev D, Paris KA, Friesner RA, Levy RM: **Prediction of protein loop conformations using the AGBNP implicit solvent model and torsion angle sampling**. *J Chem Theor Comp* 2008, **4**:855-868.

53. Ceres N, Pasi M, Lavery R: **A protein solvation model based on residue burial**. *J Chem Theor Comp* 2012, **8**:2141-2144.

54. Lopes A, Alexandrov A, Bathelt C, Archontis G, Simonson T: **Computational sidechain placement and protein mutagenesis with implicit solvent models**. *Proteins* 2007, **67**:853-867.

55. am Busch MS, Lopes A, Amara N, Bathelt C, Simonson T: **Testing the Coulomb/accessible surface area solvent model for protein stability, ligand binding, and protein design**. *BMC Bioinformatics* 2008, **9**:148.

56. Marshall SA, Vizcarra CL, Mayo SL: **One- and two-body decomposable Poisson–Boltzmann methods for protein design calculations**. *Protein Sci* 2005, **14**:1293-1304.

57. Chen J, Im W, Brooks CL: **Refinement of NMR structures using implicit solvent and advanced sampling techniques**. *J Am Chem Soc* 2004, **126**:16038-16047.

58. Kleinjung J, Bayley PM, Fraternali F: **Leap-dynamics: efficient sampling of conformational space of proteins and peptides in solution**. *FEBS Lett* 2000, **470**:257-262.

59. Kalgin IV, Caflisch A, Chekmarev SF, Karplus M: **New insights into the folding of a β-sheet miniprotein in a reduced space of collective hydrogen bond variables: application to a hydrodynamic analysis of the folding flow**. *J Phys Chem B* 2013, **117**:6092-6105.

60. Ho BK, Dill KA: **Folding very short peptides using molecular dynamics**. *PLoS Comp Biol* 2006, **2**:e27.

61. Zagrovic B, Sorin EJ, Pande V: **Beta-hairpin folding simulations in atomistic detail using an implicit solvent model**. *J Mol Biol* 2001, **313**:151-169.

62. Voelz VA, Bowman GR, Beauchamp K, Pande VS: **Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1à''39)**. *J Am Chem Soc* 2010, **132**:1526-1528.

63. Nymeyer H, Garcìa AE: **Simulation of the folding equilibrium of α-helical peptides: a comparison of the generalized Born approximation with explicit solvent**. *Proc Natl Acad Sci USA* 2003, **100**:13934-20109.

64. Duan LL, Mei Y, Zhang D, Zhang QG, Zhang JZH: **Folding of a helix at room temperature is critically aided by electrostatic polarization of intraprotein hydrogen bonds**. *J Am Chem Soc* 2006, **132**:11159-11164.

65. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M: **A Bayesian networks approach for predicting protein–protein interactions from genomic data**. *Science* 2003, **302**:449-453.

66. Ogmen U, Keskin O, Aytuna AS, Nussinov R, Gursoy A: **PRISM: protein interactions by structural matching**. *Nucleic Acids Res* 2005, **33**:W331-W336.

67. Keskin O, Nussinov R, Gursoy A: **PRISM: protein–protein interaction prediction by structural matching**. *Methods Mol Biol* 2008, **484**:505-521.

68. Chen R, Li L, Weng Z: **ZDOCK: an initial-stage protein-docking algorithm**. *Proteins* 2003, **52**:80-87.

69. Lyskov S, Gray JJ: **The RosettaDock server for local protein–protein docking**. *Nucleic Acids Res* 2008, **36**:W233-W238.

70. Taylor RD, Jewsbury PJ, Essex JW: **FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function**. *J Comp Chem* 2003, **24**:1637-1656.

71. Michel J, Verdonk ML, Essex JW: **Protein–ligand binding affinity predictions by implicit solvent simulations: a tool for lead optimization?** *J Med Chem* 2006, **49**:7427-7439.

72. Michel J, Essex JW: **Hit identification and binding mode predictions by rigorous free energy simulations**. *J Med Chem* 2008, **51**:6654-6664.

73. Yang T, Wu JC, Yan C, Wang Y, Luo R, Gonzales MB, Dalby KN, Ren P: **Virtual screening using molecular simulations**. *Proteins* 2011, **79**:1940-1951.

74. Vitalis A, Caflisch A: **Micelle-like architecture of the monomer ensemble of alzheimer's amyloid-β peptide in aqueous solution and its implications for aβ aggregation**. *J Mol Biol* 2010, **403**:148-165.

75. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID: **DOCK combining techniques to model RNA-small molecule complexes**. *RNA* 2009, **15**:1219-1230.

76. Gaillard T, Case DA: **Evaluation of DNA force fields in implicit
• solvation**. *J Chem Theor Comp* 2011, **7**:3181-3198.
Influences of the choice of force field and the implicit solvation model applied to DNA simulations on the quality of results are evaluated, particularly with respect to DNA structural parameters and in comparison to explicit simulations. The study highlights advances and pitfalls in the modelling of DNA in solution.

77. Liu Y, Haddadian E, Sosnick TR, Freed KF, Gong H: **A novel
•• implicit solvent model for simulating the molecular dynamics of RNA**. *Biophys J* 2013, **105**:1248-1257.
Novel and effective approach to implicit solvent description for RNA combining the Langevin–Debye (LD) model and the Poisson–Boltzmann (PB) equation to calculate the combined electrostatic screening induced by the dielectric response of both solvent and the monovalent counter ions.

78. Ben-Tal N, Honig B: **Helix–helix interactions in lipid bilayers**. *Biophys J* 1996, **71**:3046-3050.

79. Lazaridis T: **Effective energy function for proteins in lipid membranes**. *Proteins* 2003, **52**:176-192.

80. Im W, Feig M, Brooks CL: **An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins**. *Biophys J* 2003, **85**:2900-2918.

81. Spassov VZ, Yan L, Szalma S: **Introducing an implicit membrane
• in generalized Born/solvent accessibility continuum solvent models**. *J Phys Chem B* 2002, **106**:8726-8738.
First application of the GB/SASA approach to membraneous environment with the solvent-accessible surface approximation including the effect of the membrane (SA/IM model).

82. Tanizaki S, Feig M: **A generalized Born formalism for
•• heterogeneous dielectric environments: application to the implicit modeling of biological membranes**. *J Chem Phys* 2005, **122**:1247-1256.
A new formalism, the heterogeneous dielectric generalised Born (HDGB), is introduced to explicitly handle multiple dielectric environments and is

applied to membrane modelling. The heterogeneity of chemical group at the membrane–water interface is explicitly taken in account by modelling the membrane as a series of dielectric slabs.

83. Tanizaki S, Feig M: **Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model**. *J Phys Chem B* 2006, **110**:548-556.

84. Grossfield A: **Implicit modeling of membranes**. *Curr Top Membr* 2008, **60**:131-157.

85. De Simone A, Dodson GG, Verma CS, Zagari A, Fraternali F: **Prion and water: tight and dynamical hydration sites have a key role in structural stability**. *Proc Natl Acad Sci USA* 2005, **102**:7535-7540.

86. Fornili A, Autore F, Chakroun N, Martinez P, Fraternali F: **Protein–water interactions in MD simulations: POPS/POPSCOMP solvent accessibility analysis, solvation forces and hydration sites**. *Methods Mol Biol* 2012, **819**:375-392.