LUND UNIVERSITY

An MM/3D-RISM approach for ligand binding affinities.

Genheden, Samuel; Luchko, Tyler; Gusarov, Sergey; Kovalenko, Andriy; Ryde, Ulf

Published in:
The Journal of Physical Chemistry Part B

DOI:
10.1021/jp101461s

2010

Document Version:
Peer reviewed version (aka post-print)

Link to publication

Citation for published version (APA):
Genheden, S., Luchko, T., Gusarov, S., Kovalenko, A., & Ryde, U. (2010). An MM/3D-RISM approach for ligand binding affinities. *The Journal of Physical Chemistry Part B*, *114*(25), 8505-8516. https://doi.org/10.1021/jp101461s

Total number of authors:
5

Creative Commons License:
Unspecified

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

An MM/3D-RISM approach for ligand-binding affinities

Samuel Genheden,[1] Tyler Luchko,[2,3] Sergey Gusarov,[2]

Andriy Kovalenko [2,4,*] and Ulf Ryde [1,*]

[1] Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124, SE-221 00 Lund, Sweden

[2] National Institute for Nanotechnology, 11421 Saskatchewan Drive, Edmonton, Alberta, T6G 2M9, Canada

[3] Present Address: BioMaPS Institute for Quantitative Biology, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854

[4] Department of Mechanical Engineering, University of Alberta

Correspondence to

Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se, Tel: +46 – 46 2224502, Fax: +46 – 46 2224543

Andriy Kovalenko, E-mail: Andriy.Kovalenko@nrc.ca, Tel: 780-641-1716, Fax: 780-641-1601

2017-03-12

Abstract

We have modified the popular MM/PBSA or MM/GBSA approaches (molecular mechanics for a biomolecule, combined with a Poisson–Boltzmann or generalized Born electrostatic and surface area non-electrostatic solvation energy) by employing instead the statistical-mechanical, three-dimensional molecular theory of solvation (a.k.a. 3D reference interaction site model, or 3D-RISM-KH) coupled with molecular mechanics or molecular dynamics (N. Blinov *et al*, Biophys. J., 2010; T. Luchko *et al*, J. Chem. Theory Comput., 2010). Unlike the PBSA or GBSA semiempirical approaches, the 3D-RISM-KH theory yields a full molecular picture of the solvation structure and thermodynamics from the first principles, with proper account of chemical specificities of both solvent and biomolecules, such as hydrogen bonding, hydrophobic interactions, salt bridges, etc. We test the method on the binding of seven biotin analogues to avidin in aqueous solution and show it to work well in predicting the ligand-binding affinities. We have compared the results of 3D-RISM-KH with four different generalised Born and two Poisson–Boltzmann methods. They give absolute biding energies that differ by up to 208 kJ/mol and mean absolute deviations in the relative affinities of 10–43 kJ/mol.

Key Words: ligand-binding affinity, solvation, Poisson–Boltzmann, generalized Born, MM/GBSA, MM/PBSA, MM/3D-RISM-KH.

Introduction

One of the major challenges for theoretical calculations is to estimate the affinity of a small ligand (L) to a macromolecule (R), i.e. the free energy (ΔG_{bind}) of the reaction

$$L + R \rightleftharpoons LR \quad (1)$$

Many methods have been developed with this aim, ranging from accurate, but time consuming free-energy perturbation (FEP) approaches to fast empirical scoring functions.[1] Much effort has been spent on developing methods in the middle of this scale, i.e. methods that are based on molecular simulations, but are less time-consuming than FEP. Examples of such approaches are the linear-response approximation (LRA),[2] the semi-macroscopic protein-dipole Langevin-dipole approach (PDLD/S),[3,4] the linear interaction energy approach (LIE),[5,6] and the MM/PBSA approach (molecular mechanics combined with Poisson–Boltzmann and surface area solvation).[7]

The latter approach is interesting because it does not contain any empirical parameters and it divides the binding energy into a set of separate and physically well-defined terms. The binding free energy is calculated as the difference in free energy of the three reactants:

$$\Delta G_{bind} = \langle G_{RL} \rangle - \langle G_R \rangle - \langle G_L \rangle \quad , \quad (2)$$

where each free energy is estimated as a sum, according to

$$G = E_{MM} + G_{solv} - TS_{MM} \quad , \quad (3)$$

where E_{MM} is the molecular mechanics gas-phase energy of the reactant, consisting of the internal energy (from bonds, angles, and dihedral angles), as well as the non-bonded electrostatic and van der Waals energies:

$$E_{MM} = E_{int.} + E_{el.} + E_{vdW} \, , \quad (4)$$

G_{solv} is the solvation energy, which is traditionally calculated by solving the Poisson–Boltzmann (PB) equation[8] to get the polar part of the solvation energy (G_{pol}), and by a relation to the solvent accessible surface area (SASA) for the non-polar part (G_{np}). The last term TS_{MM} is the product of the absolute temperature and the entropy, which is calculated from a normal-mode analysis of a truncated system at the molecular-mechanics level. The averages in Eqn. 2 are calculated from a set of snapshots taken from a molecular dynamics simulation to include the effects of dynamics. Each of the three free energies in Eqn. 2 should in principle be calculated from an individual simulation, but it is more common to only simulate the complex and then calculate all three free energies in Eqn. 2 from this simulation.[9,10] Thereby, the precision of the method is improved and the internal MM energy (E_{int}) cancels out.

From Eqn. 3, it is clear that the solvation energy plays an important role in the MM/PBSA estimate, typically being one of the two largest terms, and to a great extent cancelling the electrostatic interactions between the ligand and the receptor. Therefore, several different approaches have been used for this term. In particular, it has been tested if the calculations can be sped up by using the generalized Born (GB) method[11] instead, giving a MM/GBSA approach.[12,13] Alternatively, the polarised continuum model (PCM)[14,15] has also been used, especially in combination with quantum chemical approaches.[16,17]

However, implicit solvation models are phenomenological and thus bear fundamental drawbacks: Hydrogen bonding and hydrophobic interactions are not transferable to new complex systems with features arising from molecular specificities of their constituents, such as various association and steric effects due to structural solvent, co-solvent, salt, buffer, etc. Moreover, the solvent-accessible surface and volume are not well defined, and the entropic term is absent in continuum solvation. It has been demonstrated recently that PB methods are capable of faithfully reproducing polar explicit solvent forces for dilute protein systems.

However, the popular SASA method cannot accurately describe non-polar solvation forces at atomic length scales and may not be accurate or transferable enough for high-resolution modelling studies of protein folding and binding.[18,19,20] Furthermore, the PB solvation energies strongly depend on the atomic radii set chosen.[19] A possible way to address these issues is to supplement the SASA model with additional volume and dispersion integral terms suggested by scaled particle models and Weeks–Chandler–Andersen theory.[20,21]

An attractive alternative based on first-principles statistical mechanics is the three-dimensional molecular theory of solvation (also known as the three-dimensional reference interaction site model, 3D-RISM), complemented with the Kovalenko-Hirata (KH) closure relation.[22,23,24] It starts from an explicit solvent model, but operates with 3D distributions of species in the statistical ensemble, rather than with molecular trajectories, and predicts the solvation structure and thermodynamics of biomolecules. The 3D-RISM-KH theory properly accounts for chemical functionalities of both biomolecules and solvent species (including ligands) by representing both electrostatic and non-polar features of the solvation structure, such as hydrogen bonding, hydrophobicity, salt bridges, structural solvent, etc. Moreover, it analytically yields all the solvation thermodynamics, including the solvation free energy, its energetic and entropic decomposition, as well as the partial molar volume and compressibility. The 3D-RISM-KH theory has been coupled with *ab initio* quantum chemistry methods in a self-consistent description of electronic structure, to obtain optimised geometries and to study chemical reactions in solution.[22,25,26] The KS-DFT/3D-RISM-KH method has been extensively validated against experimental data for solvation thermochemistry, conformational equilibria, tautomerisation energies, and activation barriers for various nanosystems in different solvents.[25,26]

Among many applications, the 3D-RISM-KH theory has been successfully used to predict and explain the driving forces of self-assembly pathways and conformational stability of synthetic nanotubes in different solvents,[27,28,29] solvent-controlled formation and switching of supramolecular chirality of nanotubes,[30] water channels formed and host atoms held inside nanotubes,[29] formation and conformational stability of microtubular architectures in aqueous solution,[31] hydration effects on prions and amyloid aggregates,[32] and effects of mutation on the association thermodynamics and conformational stability of amyloid oligomers.[33]

Recently, the 3D-RISM-KH method has been coupled with MD simulation of biomolecules in the Amber molecular dynamics package.[34] This included a number of accelerating schemes with several cut-offs for the interaction potentials and correlation functions, an iterative guess for the 3D-RISM solutions, as well as extrapolating solvent-induced forces and applying them in large multi-time steps (up to 20 fs) to enable simulation of large biomolecules. The coupled MD/3D-RISM-KH method makes feasible modelling of biomolecular structures of practical interest and thus has tremendous potential for computer-aided drug design. It allows one to study processes on long time-scales, as the solvent dynamics is accounted for statistical-mechanically. It could replace the MM/GBSA or MM/PBSA post-processing, suffering from the empirical treatment of non-polar contributions.[33] The MM/3D-RISM-KH method gives at once 3D maps of binding affinity of relatively small ligand molecules to a biomolecule without any phenomenological approximations.[35] It quickly and accurately yields the solvation structure for biomolecular systems as large and complex as chaperones[36] and predicts binding maps of prion proteins for development of novel inhibitors of prion protein conversion.[37] In this paper, we employ and test the recently introduced MM/3D-RISM-KH method,[36] in which the PBSA solvation model in MM/PBSA analysis of the thermodynamics of the MD trajectories is replaced by the solvation free energies calculated with the 3D-RISM-KH theory. The results are compared to standard MM/PBSA and MM/GBSA calculations. As a test case, we have chosen the binding of seven biotin analogues to protein avidin. This protein is well-characterized by X-ray crystallography[38,39,40,41] and experimental binding affinities are available.[42,43,44] This system has been the subject of several studies with FEP,[45,46] LIE,[46] and MM/PB(GB)SA.[47,47,48,63,49,50,51]

Methods

Preparation of complexes

We have studied the binding of the seven biotin analogues in Figure 1 to avidin. The preparation of the avidin protein (PDB ID 1AVD[40]) and the inhibitors has been described before.[49] The Amber99SB force field[52] was used for the protein atoms and the inhibitors were described with the Amber99 force field[53] with charges[49] calculated with the RESP (restrained electronic potential) method.[54] Each protein–ligand system was immersed in an octahedral box of TIP4P-Ewald[55] water molecules that extended at least 10 Å outside the protein. The Amber99SB force field, especially in combination with TIP4P-Ewald water, has in several investigations been shown to better reproduce experimental structures and dynamic properties than other Amber force fields.[52,56,57,58]

Simulation protocol

The molecular dynamics (MD) simulations were run by the SANDER module in Amber 10.[59] The temperature was kept at 300 K using Langevin dynamics[60] with a collision frequency of 2.0 ps^{-1}. The pressure was kept at 1 atm using a weak-coupling approach,[61] with isotropic position rescaling and a relaxation time of 1 ps**.** Particle-mesh Ewald (PME)[62] with a fourth-order B-spline interpolation and a tolerance of 10^{-5} was used to treat long-range electrostatics. The non-bonded cut-off was 8 Å and the non-bonded pair list was updated every 50 fs. The MD time step was 2 fs. The SHAKE algorithm[63] was used to constrain bond lengths involving hydrogen atoms.

The complex was first optimised by 500 steps of steepest descent minimisation, keeping all atoms, except water molecules and hydrogen atoms, restrained to their start position with a force constant of 418 kJ/mol/$Å^2$. The minimisation was followed by 20 ps MD equilibration with a constant pressure and the restraining force constant reduced to 214 kJ/mol/$Å^2$. Finally, we started 20 different 300 ps simulations at a constant pressure, but without any restraints, using different random seed numbers for the generation of starting velocities and the final structure was used for the energy calculations (for the ligand Btn2, 25 simulations were run to obtain a precision similar to that of the other ligands). We have recently shown that such an approach with many short independent simulations is more effective than a single long simulation to yield statistically valid and converged results.[51]

Avidin is a tetramer with four binding sites. The full tetramer was treated explicitly in the calculations and the binding energy was calculated for each ligand separately, treating the other three ligands as a part of the protein. Thus, the results are based on 4 ? 20 = 80 energy calculations (100 for Btn2). Prior to the MM/PB(GB)SA calculations, the ligand in the subunit of interest was centred in the octahedral box.

MM/PB(GB)SA calculations

ΔG_{bind} was calculated according to Eqns. 2–4. All terms in Eqn. 4 were calculated with Amber 8[64] with all water molecules stripped off and without any periodic boundary conditions, but with an infinite cut-off. We tested all four GB methods available in Amber, viz. the GB^{HCT} method,[65,66] the GB^{OBC} method, either with the model I or model II parameter sets (i.e. the ?, ?, and ? parameters are either set to 0.8, 0, and 2.91, or to 1.0, 0.8, and 4.85),[67] and the GBn method.[68] The four methods employ different atomic radii as is detailed in the Amber manual:[59] the modified Bondi radii[69] for GB^{HCT}, a second modified Bondi radii set (mbondi2)[67] for the two GB^{OBC} methods, and unmodified Bondi radii for GBn.[70] Default parameters were used for all methods.

In addition, we also tested the Poisson–Boltzmann method.[8] These calculations were performed either by the SANDER routine implemented in Amber 10,[59] or by the stand-alone

software Delphi II.[71] Both calculations employed the Parse radii[72] for all atoms and a grid spacing of 0.5 Å. The dielectric constants in the solute and the solvent were 1 and 80, respectively, and the probe radius was 1.4 Å. In the Delphi calculations, the solute filled 90% of the solvent box, whereas in Amber, the box was three times larger than the solute.

Both the PB and GB methods provide only the polar part of the solvation energy. The non-polar part of the solvation energy was estimated from the solvent-accessible surface area (SASA), according to

$$\boxed{?}G_{np} = \boxed{?}\,\mathrm{SASA} + b \qquad (5)$$

with $\boxed{?}$ = 0.0227 kJ/mol/Å^2 and b = 3.85 kJ/mol.[64] In the Amber implementation of MM/PBSA, the SASA is always calculated with the Bondi atomic radii. However, if Parse radii[72] are used instead, the SASA contribution to the net binding energy changes by only 0.1–0.3 kJ/mol.

The entropy was estimated by a normal-mode analysis of the harmonic frequencies calculated at the MM level. For this calculation, we used our recently described modification to improve the precision:[73] All residues more than 12 Å from any atom in the ligand were deleted and the remaining atoms were minimised, keeping all residues more than 8 Å from the ligand (including all water molecules) fixed, to ensure that the geometry is as close as possible to the original structure. Thereby, the questionable use of a distance-dependent dielectric constant in the standard MM/PBSA approach[7] is avoided. In the frequency calculations, the fixed buffer region was omitted.

3D-RISM-KH calculations

The solvation structure is represented by the probability density $\rho_\gamma g_\gamma(\boldsymbol{r})$ of finding interaction site γ of solvent molecules at 3D space position $\boldsymbol{r}$ around the solute macromolecule/supramolecule, which is determined by the average number density ρ_γ in the solvent bulk times the 3D distribution function (normalized density distribution) $g_\gamma(\boldsymbol{r})$ of solvent site γ. The latter indicates site density enhancement when $g_\gamma > 1$ or depletion when $g_\gamma < 1$, relative to the average density at a distance from the solute molecule in the solvent bulk where $g_\gamma \rightarrow 1$.

The 3D-RISM theory in the so-called hypernetted-chain (HNC) closure approximation was sketched by Chandler and co-workers in their derivation of density functional theory for site distributions of molecular liquids,[74] and then introduced in that way by Beglov and Roux for polar molecules in water solvent.[75] Kovalenko and Hirata derived the 3D-RISM integral equation from the six-dimensional, molecular Ornstein–Zernike integral equation[76] by averaging out the orientation degrees of freedom of solvent molecules, while keeping the orientation of the solute macromolecule described at the three-dimensional level.[22,23] They also developed an analytical treatment of the electrostatic asymptotics of the 3D site correlation functions in solving the 3D-RISM equations.[22,24] This enabled description of ionic species and polar macromolecules for which distortion of the electrostatic asymptotics of either of the correlation functions leads to huge errors in the solvation free energy (even for simple ions and ion pairs in water) while the analytical treatment of the asymptotics restores it to an accuracy of a fraction of a kJ/mol.

The 3D-RISM integral equation for the 3D site correlation functions has the form

$$h_\gamma(\boldsymbol{r}) = \sum_\alpha \int d\boldsymbol{r}'\, c_\alpha(\boldsymbol{r}-\boldsymbol{r}')\chi_{\alpha\gamma}(r') \quad , \qquad (6)$$

where $h_\gamma(\boldsymbol{r})$ is the 3D total correlation function of solvent site γ, which is related to the 3D site

distribution function by $g_\gamma(\boldsymbol{r}) = h_\gamma(\boldsymbol{r}) + 1$, and $c_\gamma(\boldsymbol{r})$ is the 3D direct correlation function, which has the asymptotics of the solute–solvent site-interaction potential, $c_\gamma(\boldsymbol{r}) \sim -u_\gamma(\boldsymbol{r}) / (k_B T)$, where $u_\gamma(\boldsymbol{r})$ is the 3D interaction potential between the whole solute molecule and solvent site γ specified by a molecular force field, and $k_B T$ is the Boltzmann constant times the solution temperature. The site–site susceptibility of pure solvent $\chi_{\alpha\gamma}(r)$ is an input to the 3D-RISM theory, and site indices α and γ enumerate all sites on all sorts of solvent species (in the case of a multicomponent solvent). Another relation between the 3D total and direct correlation functions, called a closure, is necessary to complement the 3D-RISM integral equation, Eqn. 6. The exact closure can be formally expressed as a series in terms of multiple integrals of the combinations of the total correlation functions. However, it is extremely cumbersome, and in practice is replaced with more amenable approximations. Kovalenko and Hirata proposed the closure approximation (3D-KH closure) that automatically applies the 3D-HNC treatment to repulsive cores and other regions of density depletion due to repulsive interaction and steric constraints, and the 3D mean-spherical approximation (3D-MSA) to distribution peaks due to associative forces and other density enhancements, including long-range distribution tales for structural and phase transitions in fluids and mixtures,[76,22]

$$g_\gamma(\boldsymbol{r}) = \begin{cases} \exp(d_\gamma(\boldsymbol{r})) & for \quad d_\gamma(\boldsymbol{r}) \leqslant 0 \\ 1 + d_\gamma(\boldsymbol{r}) & for \quad d_\gamma(\boldsymbol{r}) > 0 \end{cases}, \tag{7}$$

$$d_\gamma(\boldsymbol{r}) = -u_\gamma(\boldsymbol{r})/(k_B T) + h_\gamma(\boldsymbol{r}) - c_\gamma(\boldsymbol{r}),$$

The 3D-KH approximation yields solutions to the 3D-RISM equations for polyionic macromolecules, solid–liquid interfaces, and fluid systems near structural and phase transitions, for which the 3D-HNC approximation is divergent and the 3D-MSA approximation produces non-physical areas of negative density distributions (for the conventional RISM theory,[76] the corresponding site–site version of the KH closure approximation is available and capable of predicting phase and structural transitions in both simple and complex associating liquids and mixtures[22]). The 3D-RISM-KH theory has been proven to be appropriate in numerous cases, including complex liquids and electrolyte solutions,[22,25,26] as well as organic supramolecular[27,28,29,30] and biomolecular[31,32,33,34,35,36,37] systems in solution.

The site–site susceptibilities of pure solvent $\chi_{\alpha\gamma}(r)$ in Eqn. 6 consist of intra- and intermolecular terms,

$$\chi_{\alpha\gamma}(r) = \omega_{\alpha\gamma}(r) + \rho_\alpha h_{\alpha\gamma}(r), \tag{8}$$

where the intramolecular correlation function $\omega_{\alpha\gamma}(r) = \delta_{\alpha\gamma}\,\delta(r) + (1 - \delta_{\alpha\gamma})\,\delta(r - l_{\alpha\gamma})/(4\pi\, l_{\alpha\gamma}^2)$ represents the geometry of solvent molecules with site–site separations $l_{\alpha\gamma}$ specified by the molecular force field, and $h_{\alpha\gamma}(r)$ is the radial total correlation function between sites α and γ enumerating all sites on all sorts of molecules in pure solvent. In advance to the 3D-RISM-KH calculation, $h_{\alpha\gamma}(r)$ of pure solvent is obtained from the dielectrically consistent RISM theory (DRISM)[77] coupled with the KH closure (DRISM-KH),[22] applied to the bulk solution with counter ions at infinite dilution. The bulk solvent susceptibility, Eqn. 8, is then input into the 3D-RISM integral equation, Eqn. 6.

To properly account for electrostatic forces in electrolyte solution with polar solvent and ionic species when evaluating the convolution in the 3D-RISM integral equation, Eqn. 6, by using 3D Fourier transform, the electrostatic asymptotics of all the correlation functions (both the 3D and radial ones) are separated out in the direct and reciprocal space and treated analytically.[74,22,78] For ionic or polar solutes, simply truncating the electrostatic asymptotics leads to catastrophic errors. Bare periodic Ewald treatment of the 3D solute-solvent site

interaction potential $u_\gamma(\boldsymbol{r})$ imposes supercell periodicity on the solute molecule, resulting in huge artefacts in both the solvation structure and thermodynamics, which can be eliminated by using analytical corrections.[74,22] Alternatively,[78] the 3D solute-solvent site interaction potential $u_\gamma(\boldsymbol{r})$ and correlation functions $c_\gamma(\boldsymbol{r})$ and $h_\gamma(\boldsymbol{r})$ are discretized on a 3D linear grid in a non-periodic rectangular box, and their non-periodic electrostatic asymptotics specified analytically and subtracted before applying the 3D fast Fourier transform to the remaining short-range parts and then added back after the transform. The box has to be large enough to to have the short-range parts of the potential $u_\gamma(\boldsymbol{r})$ and correlation functions $c_\gamma(\boldsymbol{r})$ and $h_\gamma(\boldsymbol{r})$ around the macromolecular solute decay at its boundaries. The asymptotics of the 3D direct correlation function is determined by that of the 3D interaction potential, $c_\gamma^{(as)}(\boldsymbol{r}) = -u_\gamma^{(as)}(\boldsymbol{r})/(k_B T)$, and so they cancel out in the closure relation and Eqns. 6-7 are in fact being solved for the short-range part of $c_\gamma(\boldsymbol{r})$. The asymptotics is specified as the potential of the site charges, gauss-smeared with half-width η for its convenient behaviour with the suppressed singularity at $\boldsymbol{r}\to 0$ and Gaussian decay at large $\boldsymbol{k}$,

$$c_\gamma^{(as)}(\boldsymbol{r}) = -\frac{1}{k_B T}\sum_i \frac{Q_i q_\gamma}{|\boldsymbol{r}-\boldsymbol{R}_i|} erf\left(\frac{|\boldsymbol{r}-\boldsymbol{R}_i|}{\eta}\right) \quad , \tag{9a}$$

$$c_\gamma^{(as)}(\boldsymbol{k}) = -\frac{4\pi}{k_B T}\sum_i \frac{Q_i q_\gamma}{k^2}\exp\left(-\frac{1}{4}k^2\eta^2 + i\,\boldsymbol{k}\cdot\boldsymbol{R}_i\right) \quad , \tag{9b}$$

where Q_i and q_γ are partial charges of solute sites i located at $\boldsymbol{R}_i$ in the solute molecule and of solvent sites γ, and the smearing parameter is set about $\eta\approx 1$Å to ensure smoothness of the form 9a within a molecular core and quick decay of the form 9b with k. Substitution of the long-range $c_\gamma^{(as)}(\boldsymbol{k})$ given by Eqn. 9b the 3D-RISM integral equation 6 yields the renormalized, long-range $\tilde{h}_\gamma^{(as)}(\boldsymbol{k}) = \sum_\alpha c_\alpha^{(as)}(\boldsymbol{k})\chi_{\alpha\gamma}(k)$ in a rather involved form not amenable to analytical transformation to the direct space. However, it suffices to introduce a simple analytical form with the same asymptotic behaviour. The electrostatic asymptotics of the 3D total correlation function $h_\gamma(\boldsymbol{r})$ represents the distribution of solvent site charges resulting in dielectric screening of the solute charge density by solvent polar molecules and Debye screening by solvent ions. Thus, it can be readily introduced in terms of the asymptotics of the potential of mean force $w_\gamma(\boldsymbol{r}) = -k_B T\log(h_\gamma(\boldsymbol{r}))$, which is dipole–dipole for a polar solute in polar solvent, ion–dipole for an ionic solute in polar solvent, and Debye-screened ion–ion for an ionic solute in electrolyte solution. The latter cannot be neglected in the case of electrolyte solution at medium and especially low ionic concentration, and the analytical electrostatic asymptotics, which are subtracted from $h_\gamma(\boldsymbol{r})$ before applying the backward 3D-FFT and then added back, are thus specified as

$$h_j^{(as)}(\boldsymbol{k}) = -\frac{4\pi}{\varepsilon k_B T}\sum_i \frac{Q_i q_j}{k^2+\kappa_D^2}\exp\left(-\frac{1}{4}k^2\eta^2 + i\,\boldsymbol{k}\cdot\boldsymbol{R}_i\right) \quad , \tag{10a}$$

$$\begin{aligned} h_j^{(as)}(\boldsymbol{r}) = &-\frac{1}{\varepsilon k_B T}\sum_i \frac{Q_i q_j}{|\boldsymbol{r}-\boldsymbol{R}_i|}\exp\left(-\frac{1}{4}\kappa_D^2\eta^2\right) \\ &\times\frac{1}{2}\left\{\exp\left(-\kappa_D|\boldsymbol{r}-\boldsymbol{R}_i|\right)\left[1-erf\left(\frac{\kappa_D}{2}-\frac{|\boldsymbol{r}-\boldsymbol{R}_i|}{\eta}\right)\right]\right. \\ &\left.-\exp\left(\kappa_D|\boldsymbol{r}-\boldsymbol{R}_i|\right)\left[1-erf\left(\frac{\kappa_D}{2}+\frac{|\boldsymbol{r}-\boldsymbol{R}_i|}{\eta}\right)\right]\right\} \end{aligned} \quad , \tag{10b}$$

where ε is the dielectric constant of the polar solvent (an input parameter in the DRISM theory for bulk solvent), and $\kappa_D = \left(\sum_j 4\pi \rho_j q_j^2 / (\varepsilon k_B T) \right)^{1/2}$ is the inverse Debye length of the electrolyte solution with ionic species j of concentrations ρj. The k=0 singularity in the asymptotics 9b and 10a is omitted by specifying $c_\gamma^{(as)}(\boldsymbol{k}=0)=0$, and $h_j^{(as)}(\boldsymbol{k}=0)=0$ for κD=0; the proper electrostatic asymptotics of $h_\gamma(\boldsymbol{r})$ ae restored when adding the form 10b back to the short-range part of $h_\gamma(\boldsymbol{r})$ after the backward 3D-FFT.

The solvation free energy of the solute molecule in multicomponent solvent is obtained from the 3D-RISM-KH integral equations, Eqns. 6–7, by analytically integrating the Kirkwood's charging formula for the solvation free energy, given by the closed analytical expression[75,22]

$$\mu^{3D-RISM-KH} = k_B T \sum_\alpha \rho_\alpha \int d\boldsymbol{r} \left(\frac{1}{2} h_\alpha^2(\boldsymbol{r}) \Theta(-h_\alpha(\boldsymbol{r})) - \frac{1}{2} h_\alpha(\boldsymbol{r}) c_\alpha(\boldsymbol{r}) - c_\alpha(\boldsymbol{r}) \right), \tag{11}$$

where Θ(x) is the Heaviside step function: $\Theta(x) = \begin{cases} 1 & for \quad x>0 \\ 0 & for \quad x<0 \end{cases}$. The solvation free energy 11 is actually the potential of mean force between the protein parts in solution dependent on protein conformation.

A modification to the form 11 can be obtained in the so-called Gaussian fluctuation (GF) approximation suggested by Chandler and co-workers[79,80] by dropping the term $\frac{1}{2} h_\alpha^2(\boldsymbol{r})$ in the solvation free energy,

$$\mu^{3D-RISM-GF} = k_B T \sum_\alpha \rho_\alpha \int d\boldsymbol{r} \left(-\frac{1}{2} h_\alpha(\boldsymbol{r}) c_\alpha(\boldsymbol{r}) - c_\alpha(\boldsymbol{r}) \right), \tag{12}$$

in which the correlation functions are obtained in the same way, from the 3D-RISM-KH equations 6–7 in the present case. This approximation to the solvation free energy will be denoted below as 3D-RISM-KH-GF. The GF approximation to the site–site RISM-HNC theory showed improvement for the solvation thermodynamics of both non-polar and polar molecules in aqueous solution.[80,81]

Accurate calculation of the integral in the solvation free energy, Eqn. 11, or equally Eqn. 12, requires analytical treatment of the asymptotics of the 3D correlation functions $c_\gamma(\boldsymbol{r})$ and $h_\gamma(\boldsymbol{r})$. In the last term of Eqn 11, the electrostatic asymptotics of $c_\gamma(\boldsymbol{r})$ cancel out upon the summation over the site index γ because of the electroneutrality condition. As the short-range parts of the 3D correlation functions decay at the 3D box boundaries, by subtracting the asymptotics and then adding them back, the integrand in Eqn 11 is split up into short-range and long-range parts, the former integrated numerically over the 3D box volume V and the latter expressed in terms of the analytical asymptotics integrated over the whole space,

$$\begin{aligned} \mu^{3D-RISM-KH} = k_B T \sum_\alpha \rho_\alpha \int_V d\boldsymbol{r} \Bigg(& \frac{1}{2} h_\alpha^2(\boldsymbol{r}) \Theta(-h_\alpha(\boldsymbol{r})) - \frac{1}{2} h_\alpha(\boldsymbol{r}) c_\alpha(\boldsymbol{r}) - c_\alpha(\boldsymbol{r}) \\ & - \frac{1}{2} (h_\alpha^{(as)}(\boldsymbol{r}))^2 \Theta(Q q_j) + \frac{1}{2} h_\alpha(\boldsymbol{r}) c_\alpha(\boldsymbol{r}) \Bigg) \\ + k_B T \sum_\alpha \rho_\alpha \int d\boldsymbol{r} & \left(\frac{1}{2} (h_\alpha^{(as)}(\boldsymbol{r}))^2 \Theta(Q q_j) - \frac{1}{2} h_\alpha(\boldsymbol{r}) c_\alpha(\boldsymbol{r}) \right) \end{aligned} , \tag{13}$$

where $Q=\sum_i Q_i$ is the total charge of the solute molecule, and $q_j=\sum_{\alpha\in j} q_\alpha$ is the total charge of solute ionic species *j* which site γ belongs to. The factor $\Theta(Q\,q_j)$ switches the asymptotics term $(h_\alpha^{(as)}(\boldsymbol{r}))^2$ on for like charges of the solute and solvent ion Q and q_j, and off for unlike ones, according to the term $h_\alpha^2(\boldsymbol{r})$ in the solvation free energy, Eqn. 11, being switched on at long range by the depleted long-range tale and off by the enhanced long-range tale of $h_\alpha(\boldsymbol{r})$. The term $(h_\alpha^{(as)}(\boldsymbol{r}))^2\Theta(Q\,q_j)$ is dropped for the solvation free energy calculated in the 3D-RISM-GF approximation, Eqn. 12. The latter integral of the asymptotics over the whole space in Eqn. 13 is analytically reduced to 1D integrals easy to compute (e.g. by using the Gaussian quadrature),

$$\begin{aligned} & k_B T \sum_\alpha \rho_\alpha \int d\boldsymbol{r} \left(\frac{1}{2}(h_\alpha^{(as)}(\boldsymbol{r}))^2 \Theta(Q\,q_j) - \frac{1}{2} h_\alpha(\boldsymbol{r}) c_\alpha(\boldsymbol{r}) \right) \\ & = \frac{1}{\pi} \sum_\alpha \frac{\kappa_\alpha^2}{\kappa_D} \int_0^1 dx \sum_{ii'} Q_i Q_{i'} J_0(\kappa_D l_{ii'} x) \exp\left(-\frac{1}{2}(\kappa_D \eta\, x)^2 \right) \\ & \qquad \times \left(\frac{x^2}{\varepsilon\,(x^2+(1-x)^2)^2} - \frac{1}{x^2+(1-x)^2} \right) \end{aligned} \quad , \tag{14}$$

where $\kappa_\alpha^2 = 4\pi\,\rho_\alpha q_\alpha q_j/(\varepsilon\,k_B T)$ is the partial contribution of site γ of solution ionic species *j* to the Debye length squared, $J_0(x)$ is the zero-order Bessel function of the first kind, and $l_{ii'}$ are the site separations between the solute partial site charges Q_i.

The MM/3D-RISM-KH results are obtained by replacing the PB or GB polar and SASA non-polar solvation terms in Eqn. 3 by the solvation free energy from Eqn. 11. The latter is obtained by solving the 3D-RISM-KH integral equations 6-7 at the selected arrangements of the biomolecule for snapshots taken along the MD trajectory, with water molecules stripped off. The ligand of interest was centered in the octahedral box before the MM/3D-RISM-KH calculations in the same way as for MM/PB(GB)SA.

The 3D-RISM calculations were performed with a modified version of Amber 10, as detailed in Ref. 34, coupled with the Kovalenko-Hirata (KH) closure.[22,22] A modified SPC/E[82] water model was used for solvent–solvent and solute–solvent calculations, as detailed in Ref. 34. Specifically, a Lennard-Jones (LJ) potential was added to the hydrogen sites of the SPC/E model with two different sets of parameters. The SPC/E model with a small-radius LJ potential on the hydrogen sites, σ_H=0.4000 Å and ε_H=0.04600 kcal/mol[83], is denoted as PR-SPC/E. A large-radius LJ potential, σ_H=1.1658 Å and ε_H=0.01553 kcal/mol,[34] applied to SPC/E is denoted cSPC/E. The cSPC/E parameters were chosen such that

$$\frac{\sigma_O}{2} = \frac{\sigma_H}{2} + B_{OH} \tag{15}$$

where σ_O and σ_H are the oxygen and hydrogen site diameters and B_{OH} is the bond length between the two sites, and

$$\epsilon_H = 0.1\,\epsilon_O \tag{16}$$

where ε_H and ε_O are the hydrogen and oxygen well depths. For all calculations, the cSPC/E model was set to a physical density 0.997 g/cm^3 and dielectric constant 78.38, corresponding to the ambient conditions of temperature T=298.15 K and pressure 1 bar.

The 3D-RISM-KH solvent distributions were solved iteratively on a 3D grid. Adjustable parameters affecting the accuracy, precision, and computational cost of the calculation include solvent box size, grid spacing for the solvent box, residual tolerance for the solution, and cut-off distance used for solute–solvent interactions. The final calculations, compared to experiment and PB(GB)SA methods, used a solvation box at least 24 Å larger than the solute in any direction (12 Å buffer) with a grid spacing of 0.5 Å and a solute–solvent interaction

cut-off of 12 Å. The solvent distribution was self-consistently solved to a residual tolerance of 10^{-5}. Additional calculations to explore sensitivity to these parameters used a 20 Å buffer and cut-off, residual tolerances of 10^{-3}, and a grid spacing of 0.25 Å.

As 3D-RISM calculates the total solvation free energy, for direct comparison to PB and GB methods it is useful to decompose this into polar and non-polar contributions. This is accomplished through calculating the solvation free energy of an otherwise identical solute with no partial charges, giving us ΔG_{np}. It corresponds to the quantity obtained by an SASA calculation. The free energy of solvent polarization is then $\Delta G_{pol} = \Delta G_{solv} - \Delta G_{np}$. This procedure was carried out for both the KH and GF solvation free energies.

Quality descriptors

To quantify the performance of the various methods, we use three different estimates: the correlation coefficient between the predicted and experimental data (r^2), the predictive index (PI),[84] and the mean absolute deviation (MAD) from the best correlation line with a unity slope (i.e. after the subtraction of the mean signed difference, trMAD). These measures are quite meaningless without estimates of their statistical uncertainty. They were obtained by simple simulations: Each biotin analogue was assigned a random number from a normal distribution, with the mean and standard deviation obtained for the estimates of ΔG_{bind}.[51] We then calculated trMAD, r^2 and PI, and repeated this procedure 10 000 times. The standard deviations within these three sets are reported as the standard error of the quality descriptor. Throughout this paper, all reported statistical uncertainties are standard errors of the mean, i.e. the standard deviation divided by the square root of the number of estimates.

Results and Discussion

Solvation free energy and the other energy terms

We have studied the binding of the seven biotin analogues in Figure 1 to avidin. The three first ligands have a total charge of –1 *e*, whereas the other four are neutral. We have used the MM/PBSA approach and tested the effect of replacing the PB+SASA solvation method with either 3D-RISM-KH, 3D-RISM-KH-GF or the four GB methods available in Amber 10. The results of the calculations are collected in Table 1. The results are based on 4 × 20 = 80 energy calculations for each ligand (100 for Btn2) and the standard errors of the mean are given in Table 2.

The first three terms in Tables 1 and 2, ΔE_{el}, ΔE_{vdW}, and $T\Delta S_{MM}$, are common to all methods tested here. Among these, $\Delta E_{el} > \Delta E_{vdW} > T\Delta S_{MM}$, except for Btn4 ($\Delta E_{vdW} > \Delta E_{el}$) and Btn7 ($T\Delta S_{MM} > \Delta E_{vdW}$). In particular, ΔE_{el} is much larger than the other terms for the three charged ligands. For the standard errors $\Delta E_{el} > T\Delta S_{MM} > \Delta E_{vdW}$ for all ligands. This is much thanks to the new entropy method,[73] which reduces the standard error of the $T\Delta S_{MM}$ estimate by a factor of ~4, thereby ensuring that it no longer limits the precision of the method. The standard error of ΔE_{el} is about twice as big for the charged ligands than for the neutral ones (5–6 kJ/mol, compared to 2–3 kJ/mol).

The electrostatic component (ΔE_{el}) is to a great extent cancelled by the polar part of the solvation energy (ΔG_{pol}), especially for the charged ligands. As specified in the Methods section, we have calculated this component with eight different methods, the four GB methods available in Amber 10,[59] PB calculated both with Amber[59] and Delphi,[71] and two variants of the 3D-RISM-KH approach (with and without the GF correction). Owing to the large difference of the solvation energy for the charged and neutral ligands, we need to analyse these two groups separately. For the charged ligands, the two RISM methods, as well

as GB^{HCT} and GB^{OBC1} give nearly identical results (within 10 kJ/mol). However, the other three methods differ much more (by up to 161 kJ/mol) in a systematic manner, GBn < GB^{OBC2} < PB^{Amber} < others < PB^{Delphi}. The correlation coefficients between the results of the various methods are lowest for the GBn method (from 0.62 for GBn/GB^{HCT} and GBn/GB^{OBC1} to 0.80 for GBn/3D-RISM-KH), whereas they are 0.80–1.00 for 3D-RISM-KH (highest for 3D-RISM-KH/GB^{OBC2}).

For the neutral ligands, the variation is slightly smaller (up to 131 kJ/mol) and less systematic, but in general GB^{OBC2} < GBn < GB^{HCT} < GB^{OBC1} < 3D-RISM-KH < PB^{Amber} < PB^{Delphi}. The correlation coefficients are lowest for the PB^{Delphi} method (from 0.91 for PB^{Delphi}/GB^{HCT} to 0.96 for PB^{Delphi}/GBn), whereas for 3D-RISM-KH, they are 0.96 for 3D-RISM-KH/GBn, 0.98 for 3D-RISM-KH/GB^{OBC2}, and 0.99 for the others. The GF correction has only a minimal influence on the polar 3D-RISM-KH results (< 4 kJ/mol), giving r^2 > 0.999 for both the neutral and charged ligands compared to 3D-RISM-KH.

On the other hand, for the non-polar solvation energy, there is a qualitative difference between the SASA and 3D-RISM-KH estimates. For all ligands, ΔG_{np}(SASA) < 0 and therefore favourable. In addition the values for the different ligands are very similar, –11 to –21 kJ/mol. On the other hand, the 3D-RISM-KH estimates are always positive (unfavourable), 3–5 times larger, and with a more pronounced variation among the seven ligands, as can be seen in Figure 2. In addition, the GF correction has a rather large influence on the results, making the energies 14–42 kJ/mol more positive. The correlation coefficients between the two RISM methods is 0.92 and those between SASA and the two 3D-RSIM-KH methods are 0.89 and 0.64 with and without the GF correction, respectively (a negative correlation).

The standard error of the polar solvation energy estimates are similar for all eight solvation methods for the charged ligands, 4–5 kJ/mol for the charged ligands and 1–3 kJ/mol for the neutral ligands, i.e. slightly smaller than for the electrostatic contribution (Table 2). For the neutral ligands, the standard error is somewhat larger for 3D-RISM-KH and PB^{Delphi} than for the other methods, but by less than 1 kJ/mol on average. For the non-polar solvation energy, the difference is even larger, because the standard error of the SASA method is ~0.03 kJ/mol for all ligands, whereas it is 1–2 kJ/mol for the 3D-RISM-KH method.

Binding free energy

The calculated binding free energies of each ligand (ΔG_{bind}) are also included in Table 1 and they are graphically compared to the experimental data[43] in Figure 3. It can be seen that all methods show a fair correlation to the experimental trend, with correlation coefficients *(r^2)* ranging from 0.59±0.03 for GB^{HCT} to 0.93±0.02 for PB^{Delphi} (Table 3). However, it can also be seen that the absolute binding affinities vary quite extensively, with differences of up to 208 kJ/mol. To a great extent, the differences between are caused by a constant offset, as can be seen in Figure 3. Such a constant offset has been observed before when different solvation models are used,[85,86] and they indicate that relative binding energies should be much more accurate than absolute ones.

This is supported by the results in Table 3: The mean absolute deviations (MAD) of the calculated binding affinities from the experimental results for the various methods is poor for all methods except PB^{Delphi} (16 kJ/mol), ranging from 37 kJ/mol for 3D-RISM-KH to 93 kJ/mol for GBn. However, if the results are translated by the average signed error, the MAD ranges from 10±1 kJ/mol for PB^{Delphi} to 43±1 kJ/mol for GBn (this estimate is called trMAD in Table 3 and it corresponds to the best correlation line with a slope of unity). For most methods, this is similar to what has been obtained before with MM/PB(GB)SA for this system, 8 kJ/mol with PB^{Delphi},[48] 5–25 kJ/mol with PB^{Delphi} and various force fields and simulation methods,[49] 19 kJ/mol with GB^{OBC2},[49] and 14.8±0.3 kJ/mol with GB^{OBC1},[51] especially considering the low precision of the results in the former two studies,[48,49] (the

standard error of the trMAD estimate was probably not better than ~6 kJ/mol because only 20 snapshots were used, taken from a single trajectory).[51]

The correlation between the binding affinities obtained with the various methods range between 0.24 for GB^{HCT}/3D-RISM-KH-GF and 0.99 for GB^{HCT}/GB^{OBC1}. The GF correction for 3D-RISM-KH has quite a large influence on the results, 15–44 kJ/mol, giving a correlation of 0.95 (but no correlation for the neutral ligands, $r^2 = 0.02$). However, the two variants give similar trMADs of 16 and 17 kJ/mol.

Considering the low uncertainties in the quality estimates in Table 3, we can conclude with good confidence that for this system, the PB^{Delphi} method gives the best result: It is the only method that gives reasonable absolute affinities (MAD = 16 kJ/mol). Moreover, it gives the smallest trMAD (10 ? 1 kJ/mol) and the largest correlation coefficient ($r^2 = 0.93$? 0.02). Finally, it gives an excellent predictive index[84] of 0.99, although four other methods (3D-RISM-KH, PB^{Amber}, GB^{OBC2}, and GBn) also give similar PI. The other methods give worse and more varying results: GB^{OBC1}, GB^{HCT} and both 3D-RISM-KH methods give trMADs of 12–17 kJ/mol, whereas PB^{Amber}, GB^{OBC2}, and GBn methods give very poor trMADs (>30 kJ/mol). On the other hand, PB^{Amber}, 3D-RISM-KH, GB^{OBC2}, and GBn gave correlation coefficients of 0.85–0.91.

The poor trMADs of the GB^{OBC2} and GBn methods are quite unexpected, because these two methods are newer developments of the GB^{HCT} method, and therefore could be expected to be better.[67,68] Moreover, the GBn method gave appreciably better solvation energies than the other three GB methods, when compared to a set of 20 neutral molecules,[87] although it was never calibrated towards such data.[68] The reason for this unexpected behaviour is most likely that our test set involves both charged and neutral molecules. Improved results and correlations can be expected if the net charge of the ligands is the same. With our restricted test set, we obtain excellent results for the three charged ligands for all methods, trMAD = 3–9 kJ/mol (PB^{Amber} best, GBn worst), but somewhat worse results for the four neutral ligands, 4–22 kJ/mol (3D-RISM-KH best, GB^{HCT} worst).

Still, this difference is larger than expected. A recent systematic study of various solvation methods indicated that if only similar ligands with the same charge are compared, nearly all implicit solvation models yielded the same relative solvation energies within 2–5 kJ/mol for three series of drug-like molecules (including the present set of biotin analogues).[87] Examination of the present results shows that the variation of the average solvation energies of the isolated ligands among the five PB and GB solvation models in this study is only slightly larger, 2–5 kJ/mol for the charged ligands, but 5–8 kJ/mol for the neutral ligands. However, it is clear that the large variation of ? G_{bind} among the various methods is not primarily caused by differences for the ligand, but rather by differences for the protein binding pocket (which do not cancel between the complex and the free protein).

Likewise, the large difference between the results obtained with the PB method implemented in Delphi and in Amber is surprising. The former gave excellent results even for the absolute binding energies, whereas both the absolute and relative results for PB^{Amber} are poor. A similar difference has been observed recently for another system, but in that case Amber gave the better results.[86] The difference in the PB energies is ~100 kJ/mol for the charged ligands and 14–64 kJ/mol for the neutral ones. Thus, there is a perfect correlation for the charged ligands ($r^2 = 1.00$), whereas $r^2 = 0.95$ for the neutral ones. The differences are mainly caused by the proteins, whereas the isolated ligands give more similar results with the two methods.

Variations of the 3D-RISM method

The 3D-RISM method provides a first-principles, statistical-mechanics approach to solvation calculations. However, it involves a number of approximations. The basic assumption of RISM is the decomposition of the direct correlation function into partial site

contributions.[88] This is inherent in RISM theory and is beyond the scope of this study. The quality of the results from RISM then depends on the explicit solvent model and the form of the closure used to solve the equations.

In this study we have used the KH closure, which is a combination of the HNC and MSA closures and inherits much of their behaviours. The properties of HNC and KH closures with respect to polar and non-polar solutes in liquid water has received much attention.[54,58,59,60,22,24] While HNC and KH closures compare very well to explicit calculations for polar and charged solutes, they have generally had difficulty with non-polar solutes. For example, HNC provides the wrong qualitative behaviour for increasing solvent size. However, the high quality of the solvent polarization free energy and the analytic expression for the solvation free energy make the KH closure the best choice at this time.

Finally, the pair-potential (force field) for the solvent (and solute) determines the thermodynamic properties of the solvent and the quality of the result, and any shortcomings in the solvent model (e.g. SPC/E water) will affect the accuracy of the results. Furthermore, not all pair-potentials can be used directly for 3D-RISM. In the case of SPC/E, an LJ potential is required for the hydrogen site. As demonstrated in Ref.54 and in Table 4, variations in the hydrogen LJ parameters have a pronounced effect on the polar component of the solvation free energy, although they have little impact on the non-polar component.

The sensitivity of the 3D-RISM results was tested by varying parameters from those given in the Methods section. These parameters are known work well for molecular dynamics simulations[54] and were used as an initial set to test the sensitivity to changes in grid spacing, solvent cut-off and box size, residual tolerance and solvent potential. Table 4 shows a subset of the parameters tested.

Notably, almost all of the binding energies were identical within the estimated error limits, except those involving the PR-SPC/E water model. Even in the latter case, the non-polar contribution was nearly identical to that with cSPC/E water, showing that all significant differences come from the polar contribution. Differences in the polar component were expected, because this was the motivation in creating the cSPC/E variant.

Minor differences in non-polar solvation free energy were observed when the solvation cut-off and box size were changed. Calculations with a 20-Å solvent buffer were generally in good agreement with calculations with a 12-Å buffer, but the former gave slightly higher non-polar contribution to the binding free energy. however, the polar contributions were identical. Given that the total binding free energies are in good agreement, the extra calculation time required for the larger box is not justified.

Decreasing the grid spacing to 0.25 Å was not extensively tested owing to the computational burden involved. As the number of grid point increases by a factor of eight, calculating the bind free energy of a single complex at 0.25 Å grid spacing requires more time than computing all seven pairs at 0.5 Å grid spacing. For the one example tested, this increase in resolution gave no observable difference in the binding free energy.

Increasing the residual tolerance by a factor of 100 (10^{-3}) had no significant impact on the results. Differences in binding energies are no more than 1 kJ/mol when compared to the default tolerance of 10^{-5} and are well within the statistical uncertainty. A nearly 40% reduction in computational effort was also observed. While a higher tolerance produces acceptable results in this case, a larger study with a corresponding lower standard error may require a low residual tolerance.

Results for the GF correction are given in Table 1. The use of the GF correction was intended to modify the non-polar component of the solvation free energy to bring it into better agreement with results from experiment and explicit models. The correction is applied to the calculation of the solvation free energy using the solvent distributions calculated from a standard closure (KH in this case). Generally, the GF correction lowers the non-polar contribution to the solvation energy for a solute. This is the case here (data not shown) with the solvation free energy being lower for each of the ligand, receptor and complex. However,

using Eqn. 2, the differences in solvation free energy due to the GF correction increases the calculated total binding free energy, making this correction unsuitable for the MM/3D-RISM approach to binding.

Timing

Each 3D-RISM-KH calculation (six calculations for a single ligand in one snapshot) took on average 5.4 CPU hours, using a single core of a dual-core AMD Opteron 2.8 GHz machine. Of these, 2.3 h comes from the calculation of the non-polar part and 3.0 h from the total binding free energy. Of course, calculating the non-polar contribution is only required if one is interested in decomposing the total solvation free energy into polar and non-polar contributions. Calculating only the total binding free energy reduces the computational burden by ~40%. As the GF solvation free-energy correction requires the solvent distribution calculated with the KH closure, both the KH solvation free energy and the GF correction were calculated from the same solvent distribution. The GF correction required approximately 1% of the total computation time. The PB solvation calculation took 3 minutes on a Intel Core 2 Due 2.4 GHz machine, whereas the GB calculation took less than a minute and the SASA calculation took only a few seconds. Thus, the 3D-RISM calculations are much more time consuming than the other methods, but the total time is not prohibitively high (in total 73 CPU days for this project) and the calculations are both trivially parallel (3 ? 7 ? 4 ? 20–25 independent calculations) and the 3D-RISM-KH code implemented in Amber is parallelised.

Conclusions

In this paper, we have applied the MM/3D-RISM-KH method, with the 3D molecular theory of solvation replacing the PB(GB)SA continuum solvation in evaluating the solvation thermodynamics along a set of MD trajectories. Test calculations on the binding of seven biotin analogues to avidin indicate that the MM/3D-RISM-KH approach gives results of a comparable accuracy to those of the more traditional MM/PBSA and MM/GBSA approaches: It gives a predictive index that is as good as for the best methods, being equal to 1. The correlation coefficient, 0.90±0.02, is similar to that of the best MM/PBSA method (with PB^{Delphi}), 0.93±0.02. Likewise, the trMAD is only slightly worse (17±1 kJ/mol compared to 10±1 kJ/mol), whereas the absolute MAD is appreciably worse (37±1 kJ/mol compared to 16±1). We have ensured the calculations are converged with respect to the parameters of the method, in particular, the box size. It is too preliminary, however, to draw final conclusions regarding the relative accuracy of these solvation approaches, as the comparison to experiment we made involved several approximations in the present calculations, including the treatment of conformational entropy and the use of the same conformation for the protein and ligand apart and for those in the complex.

Interestingly, there is a qualitative discrepancy for the non-polar solvation calculated by the 3D-RISM-KH and SASA methods. Whereas the latter is always small and favourable for the binding, the 3D-RISM-KH prediction is 3–5 times larger and always unfavourable. This difference might be attributed to a possible failure of SASA to correctly predict the non-polar contributions, as has been documented recently.[18,19,20,21] On the other hand, the HNC closure, and therefore the KH one, does have known problems,[76] which constitutes a motivation for new closure developments, some of the attempts introducing empiric bridge corrections for small organic molecules in water (see citations in Ref. 22). A behaviour similar to that of 3D-RISM has been observed for also for the non-polar energies obtained with the polarised continuum method.[17,89]

We also show that the polar solvation energies, and therefore also the calculated absolute binding affinities strongly depend on the implicit solvation model used. We have compared

the 3D-RISM-KH results with those of the Poisson–Boltzmann model calculated by Amber and Delphi and the four generalized Born models available in the Amber 10 software. We show that the absolute binding affinities differ by up to 208 kJ/mol, although most of this variation comes from a simple translation of the affinities. However, even if only relative affinities are considered, the trMADs vary from 10 to 43 kJ/mol. To reduce this variation further, it is necessary to consider only ligands with the same net charge. If this is done for our test set, we obtain trMADs of 3–8 kJ/mol for the charged ligands and 4–22 kJ/mol for the neutral ones. Interestingly, the best GB results are obtained with the old GB^{HCT} and GB^{OBC1} methods, and not by the newer GBn method, which gave better results for absolute hydration energies of small neutral molecules.[87] It is likely that other slower and more accurate GB methods, e.g. GBMV and GBR6 (not available in Amber), will give results that are closer to those of PB.[90,91,92] Apparently, the MM/P(G)BSA approach is very sensitive to the continuum-solvation model and that there is no consensus in the predictions of different method, which of course is a serious problem with the approach. It should also be noted the calculated trMADs are not very impressive: The null hypothesis that all the seven ligands have the same affinity gives trMADs of 20 kJ/mol for the full set, and 12 and 8 kJ/mol for the charged and neutral ligands, respectively.

In conclusion, we have shown that the MM/3D-RISM-KH method is a promising method to calculate ligand-binding affinities from first-principles statistical mechanics, without the fundamental drawbacks of continuum solvation approaches involving empirical parametrisation of interactions. The popular PB(GB)SA models poorly define the dielectric cavity shape, inadequately treat non-polar interactions, miss the solvent entropic term, and are not transferable to new complex systems with features arising from molecular specificities of their constituents. This includes hydrogen bonding and hydrophobic interactions, association and steric effects due to structural solvent, co-solvent, salt, buffer, and other solution species like ligands.

Acknowledgements

This investigation was supported by grants from the Swedish research council and from the Research school in pharmaceutical science. T.L, S.G. and A.K. were supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada and by the National Research Council (NRC) of Canada. Calculations were supported by computer resources of Lunarc at Lund University, HPC2N at Umeå University, and by the Centre of Excellence in Integrated Nanotools (CEIN) at the University of Alberta. We also thank Dr Nikolay Blinov for useful discussions.

References

1 Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644-2676.
2 Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Prot. Eng.* **1992**, *5*, 215-228.
3 Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. *Proteins: Struct. Funct. Genet.* **2000**, *39*, 393-407.
4 Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochim. Biophys. Acta* **2006**, *1764*, 1647-1676.
5 Åqvist, J.; Medina, C.; Samuelsson, J. E. *Prot. Eng.* **1994**, *7*, 385-391.
6 Hansson,T.; Marelius, J.; Åqvist J. *J. Comput.-Aided Mol. Design* **1998**, *12*, 27-35.
7 Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee., T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889-897.
8 Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301-332.
9 Foloppe, N.; Hubbard, R. *Curr. Med. Chem.* **2006**, *13*, 3583-3608.
10 Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67-74.
11 Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
12 Srinivasan, J.; Cheatham, T.E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401-9409.
13 Massova, I; Kollman, P. A. Perspect. *Drug Discov. Design* **2000**, *18*, 113-135.
14 Miertus, S; Scrocco, E.; Tomasi, *J. Chem. Phys.* **1981**, *55*, 117-129.
15 Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999-3093.
16 Fedorov, D.G.; Kitaura, K.; Li, H.; Jensen, J. H.; Gordon, M. S. *J. Comput. Chem.* **2006**, *27*, 976-985.
17 Söderhjelm, P.; Kongsted, J:; Ryde, U. *J. Chem. Theory Comput.* **2010**, ASAP; DOI: 10.1021/ct9006986.
18 Wagoner, J.; Baker, N. A. *J. Comput. Chem.* **2004**, *25*, 1623–1629.
19 Case, D. A.; Gohlke, H. *J. Comput. Chem.* **2004**, *25*, 238–250.
20 Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. *J. Am. Chem. Soc.* **2003**, *125*, 9523–9530.
21 Wagoner, J.; Baker, N. A. *Proc. Natl. Acad. Sci.* USA **2006**, *103*, 8331–8336.
22 A. Kovalenko, Three-dimensional RISM theory for molecular liquids and solid-liquid interfaces, in: *Molecular Theory of Solvation*, F. Hirata (ed.). Series: *Understanding Chemical Reactivity*, vol.24, (Kluwer, Dordrecht, **2003**) pp.169-275, (and references therein).
23 Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999**, *110*, 10095–10112.
24 Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2000**, *112*, 10391–10417.
25 Gusarov, S.; Ziegler, T.; Kovalenko, A. *J. Phys. Chem. A* **2006**, *110*, 6083-6090.
26 Casanova, D.; Gusarov, S.; Kovalenko, A.; Ziegler, T. *J. Chem. Theory Comput.* **2007**, *3*, 458-476.
27 Moralez, J. G.; Raez, J.; Yamazaki, T.; Motkuri, R. K.; Kovalenko, A.; Fenniri, H. *J. Am. Chem. Soc. Communications* **2005**, *127*, 8307-8309.
28 Tikhomirov, G.; Yamazaki, T.; Kovalenko, A.; Fenniri, H. *Langmuir* **2008**, *24*, 4447-4450.
29 Yamazaki, T.; Fenniri, H.; Kovalenko, A. *ChemPhysChem* **2010**, *11*, 361-367.
30 Johnson, R. S.; Yamazaki, T.; Kovalenko, A.; Fenniri, H. *J. Am. Chem. Soc.* **2007**, *129*, 5735-5743.
31 Drabik, P.; Gusarov, S.; Kovalenko, A. *Biophys. J.* **2007**, *92*, 394-403.
32 Yamazaki, T.; Blinov, N.; Wishart, D.; Kovalenko, A. *Biophys. J.* **2008**, *95*, 4540-4548.
33 Blinov, N.; Dorosh, L.; Wishart, D.; Kovalenko, A. *Biophys. J.* **2010**, *98*, 282-296.
34 Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. *J. Chem. Theory Comput.* **2010**, *6*, 607-624.

35 Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. *J. Am., Chem. Soc.* **2009**, *131,*12430-12440.
36 Stumpe, T. M.; Pande, V.; Blinov, N.; Kovalenko, A. **2010**, submitted.
37 Blinov, N.; Perez Pineiro, R.; Bjorndahl, T.; Wishart, D.; Kovalenko, A. to be published.
38 Weber, P. C.; Ohlendorf, D. H.; Wendolowski, J. J.; Salemme, F. R. *Science* **1989**, *243*, 85-88.
39 Weber, P. C.; Wendolowski, J. J.; Pantoliano, M. W.; Salemme, F. R. *J. Am. Chem. Soc.* **1992**, *14*, 3197-3200.
40 Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M. *J. Mol. Biol.* **1993**, *231*, 698-710.
41 Livnah, O.; Bayer, E. A.; Wilchek, M.; Sussman, J. L. *Prot. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5076-5080.
42 Green, N. M. *Biochem*. J. **1966**, *101*, 774-790.
43 Green, N. M. *Adv. Protein. Chem.* **1975**, *29*, 85-133.
44 Green, N. M. *Methods Enzymol.* **1990**, *184*, 51-67.
45 Miyamoto, S.; Kollman, P. A. *Proteins: Struct. Funct. Genet.* **1993**, *16*, 226-245.
46 Wang, J.; Dixon, R.; Kollman P. A. *Proteins, Struct. Funct. Genet.* **1999**, *34*, 69-81.
47 Kuhn, B.; Gerber, P.; Schultz-Gash, T.; Stahl, M. *J. Med. Chem.* **2005**, *48*, 4040-4048.
48 Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786-3791.
49 Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde U. *J. Med. Chem.*, **2006**, *49*, 6596-6606.
50 Brown, S. P.; Muchmore, S. W. *J. Chem. Inf. Model.* **2006**, *46*, 999-1005.
51 Genheden, S.; Ryde, U. *J. Comput. Chem.*, **2010**, *31*, 837-846.
52 Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct. Bioinform.* **2006**, *65*, 712-725.
53 Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem*. **2000**, *21*, 1074-1074.
54 Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
55 Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665-9678.
56 Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961-975.
57 Koller, A. N.; Schwalbe, H.; Gohlke, H. Biophys. *J. Biophys. Lett.* **2008**, *108*, L04-L06.
58 Wong, V.; Case, D. A. *J. Phys. Chem. B.* **2008**, *112*, 6013-6024.
59 Case, D.A.; Darden, T.A.; T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C.Walker,W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, Seabra, G.; I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin,and P.A. Kollman (2008), AMBER 10, University of California, San Francisco.
60 Wu, X.; Brooks, B.R. *Chem. Phys. Lett.* **2003**, *381*, 512-518.
61 Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A.; Haak, J.R. *J. Chem. Phys*., **1984**, *81*, 3684–3690.
62 Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
63 Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.
64 Case, D. A.; Darden, T. A.; T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, B. Wang, D. A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui. P. Beroza, C. Schafmeister, J. W. Caldwell, W. Ross, S.; Kollman, P. A.; (2004) AMBER 8, University of California, San Francisco.
65 Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824-19839.
66 Tsui, V.; Case, D. A. *Biopolymers* **2001**, 122*,* 2489-2498.
67 Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383-394.
68 Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. J. *Chem. Theory Comput.* **2007**, *3*, 156-169.

69 Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D.; *J. Chem. Theory Comput.* **2006**, *2*, 128-139
70 Bondi, A., *J. Phys. Chem.* **1964**, 68, 441-451.
71 Rocchia, W.; Alexov, E.: Honig. B. *J. Phys. Chem.* B **2001**, *105*, 6507-6514.
72 Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978-1988.
73 Kongsted, J.; Ryde, U. *J. Comp.-Aided Mol. Des.*, **2009**, *23*, 63-71.
74 Chandler, D.; McCoy, J.; Singer, S. *J. Chem. Phys.* **1986**, *85*, 5971-5982.
75 Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 7821-7826.
76 Hansen, J.-P.; McDonald, I. *Theory of Simple Liquids,* 3rd edn. Elsevier, Amsterdam, the Netherlands, **2006**.
77 Perkyns, J. S.; Pettitt, B. M. *J. Chem. Phys.* **1992**, *97*, 7656–7666.
78 Kaminski, J. W.; Gusarov, S.; Wesolowski, T. A.; Kovalenk, A. *J. Phys. Chem. A,* **2010**, in press.
79 Chandler, D.; Sing, Y.; Richardson, D. M. *J. Chem. Phys.* **1984**, *81*,1975–1982.
80 Chandler, D.; Ichiye, T. *J. Phys. Chem.* **1988**, *92*, 5257–5261.
81 Lee, P. H.; Maggiora, G. M. *J. Phys. Chem.* **1993**, *97*, 10175–10185.
82 Berendsen, H.; Grigera, J.; Straatsma, T. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
83 Pettitt, B. M.; Rossky, P. *J. Chem. Phys.* **1982**, *77*, 1451–1457.
84 Pearlman, D. A.; Charifson, P. S. *J. Med. Chem.* **2001**, *44*, 3417-3423.
85 Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238-250.
86 Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. *J. Comput. Chem.* **2010**, *31*, 797-810.
87 Kongsted, J.; Söderhjelm, P.; Ryde, U. *J. Comp.-Aided Mol. Design* **2009**, 23, 395-409.
88 Hirata. F. Theory of molecular liquids. In: *Molecular Theory of Solvation*, F. Hirata (ed.). Series: *Understanding Chemical Reactivity*, vol.24. Kluwer Academic Publishers: Dordrecht, **2003**; pp 1-60.
89 Söderhjelm, P.; Kongsted, J.; Genheden S.; Ryde, U. *Interdiscip. Sci. Comput. Life Sci.*, **2010**, *2*, 21-37.
90 Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1348-1356.
91 Tjong, H.; Zhou, H.-X. *J. Phys. Chem. B* **2007**, *111*, 3055-3061.
92 Liu, H.-Y.; Grinter, S. Z.; Zou, X. *J. Phys. Chem. B* **2009**, *113*, 11793-11799.

Table 1. Results of the various solvation models for the binding of the seven biotin analogues to avidin. All energies are in kJ/mol. For comparison, the experimental binding energies are also given.[18]

	Btn1	Btn2	Btn3	Btn4	Btn5	Btn6	Btn7
E_{el}	-1273.0	-1271.6	-1249.3	-163.4	-106.8	-74.2	-105.3
E_{vdW}	-155.7	-155.3	-142.8	-205.0	-136.0	-136.3	-57.6
TS_{MM}	-97.7	-104.6	-99.5	-98.2	-81.3	-74.9	-65.8
G_{np}(3D-RISM-KH-GF)	83.3	80.4	81.8	130.1	91.4	95.8	46.7
G_{np}(3D-RISM-KH)	47.7	44.4	46.8	88.5	62.1	67.6	32.5
G_{np}(SASA)	-16.8	-16.8	-16.8	-21.1	-16.3	-16.1	-10.5
G_{pol}(3D-RISM-KH-GF)	1226.3	1233.9	1202.9	199.2	124.3	103.6	91.0
G_{pol}(3D-RISM-KH)	1223.0	1230.1	1199.6	196.5	122.2	102.0	90.0
G_{pol}(PB^{Amber})	1175.5	1185.8	1159.4	217.3	134.9	114.4	85.3
G_{pol}(PB^{Delphi})	1274.2	1283.7	1260.3	281.3	149.0	157.4	118.2
G_{pol}(GB^{HCT})	1224.6	1239.9	1203.3	171.9	107.6	83.5	84.3
G_{pol}(GB^{OBC1})	1223.0	1237.3	1203.3	181.8	107.4	88.3	85.6
G_{pol}(GB^{OBC2})	1151.3	1157.3	1135.7	150.0	78.1	67.1	64.2
G_{pol}(GBn)	1125.8	1122.6	1112.9	174.7	79.8	72.5	68.8
G_{bind}(3D-RISM-KH-GF)	-21.4	-8.1	-7.8	59.1	54.2	63.7	40.6
G_{bind}(3D-RISM-KH)	-60.3	-47.9	-46.3	14.8	22.9	34.0	25.4
G_{bind}(PB^{Amber})	-172.2	-153.4	-150.1	-74.0	-42.8	-37.4	-22.3
G_{bind}(PB^{Delphi})	-73.5	-55.4	-49.2	-10.0	-28.7	5.6	10.7
G_{bind}(GB^{HCT})	-123.2	-99.3	-106.1	-119.4	-70.1	-68.3	-23.2
G_{bind}(GB^{OBC1})	-124.8	-101.8	-106.3	-109.6	-70.4	-63.5	-21.8
G_{bind}(GB^{OBC2})	-196.5	-181.7	-173.9	-141.5	-99.6	-84.6	-43.2
G_{bind}(GBn)	-222.0	-216.4	-196.7	-116.7	-97.9	-79.3	-38.6
G_{bind}(Exp)	-85.4	-59.8	-58.6	-36.8	-34.3	-20.9	-18.8

Table 2. Standard error of the mean over the 80 snapshots (100 for Btn2) for the various energy terms in the binding energy of the seven biotin analogues to avidin. All energies are in kJ/mol.

	Btn1	Btn2	Btn3	Btn4	Btn5	Btn6	Btn7
E_{el}	5.6	5.0	6.2	2.8	2.8	2.1	2.0
E_{vdW}	1.4	1.2	1.6	1.3	1.2	1.1	0.9
TS_{MM}	1.9	1.8	1.9	2.2	1.5	1.7	1.1
G_{np}(3D-RISM-KH-GF)	1.3	1.2	1.4	1.4	1.5	1.3	0.8
G_{np}(3D-RISM-KH)	1.4	1.3	1.5	1.5	1.7	1.4	0.9
G_{np}(SASA)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G_{pol}(3D-RISM-KH-GF)	4.6	4.6	5.4	2.4	2.1	2.4	1.6
G_{pol}(3D-RISM-KH)	4.6	4.6	5.4	2.4	2.1	2.3	1.6
G_{pol}(PB^{Amber})	4.4	3.8	4.8	2.3	1.4	1.6	1.1
G_{pol}(PB^{Delphi})	5.0	4.1	5.2	2.7	1.8	2.2	1.9
G_{pol}(GB^{HCT})	4.7	4.4	5.4	2.3	1.5	1.6	1.2
G_{pol}(GB^{OBC1})	4.7	4.4	5.3	2.3	1.4	1.7	1.2
G_{pol}(GB^{OBC2})	4.2	4.1	4.9	1.9	1.0	1.3	0.8
G_{pol}(GBn)	4.4	4.2	5.1	2.2	1.1	1.6	0.8
G_{bind}(3D-RISM-KH-GF)	3.1	2.8	2.8	2.9	2.6	2.5	1.7
G_{bind}(3D-RISM-KH)	3.2	2.9	2.9	3.0	2.7	2.6	1.7
G_{bind}(PB^{Amber})	3.3	3.1	2.9	3.0	2.7	2.3	1.8
G_{bind}(PB^{Delphi})	3.3	2.8	2.7	3.1	2.9	2.8	2.1
G_{bind}(GB^{HCT})	2.4	2.3	2.3	2.9	2.1	2.0	1.6
G_{bind}(GB^{OBC1})	2.5	2.5	2.4	2.9	2.3	2.0	1.7
G_{bind}(GB^{OBC2})	3.0	2.9	2.8	2.8	2.7	2.2	1.7
G_{bind}(GBn)	3.8	3.4	3.2	2.7	3.2	2.2	1.8

Table 3. Quality descriptors of ΔG_{bind}, compared to experiments, calculated with the various solvation methods. The four descriptors are MAD (mean absolute difference), trMAD (MAD, after subtraction of the mean signed difference), the correlation coefficient (r^2), and the predictive index.[84] For each descriptor, except MAD, the standard error is also given in the second column.

	MAD		trMAD		r^2		PI	
ΔG_{bind}(3D-RISM-KH-GF)	70.7	1.0	16.3	1.0	0.80	0.02	0.87	0.02
ΔG_{bind}(3D-RISM-KH)	36.8	1.0	17.4	1.1	0.90	0.02	0.99	0.02
ΔG_{bind}(PB^{Amber})	48.2	1.0	36.4	1.0	0.91	0.01	1.00	0.01
ΔG_{bind}(PB^{Delphi})	16.3	1.0	9.7	1.0	0.93	0.02	0.99	0.00
ΔG_{bind}(GB^{HCT})	42.1	0.8	14.6	0.9	0.59	0.03	0.85	0.06
ΔG_{bind}(GB^{OBC1})	40.5	0.9	12.3	0.9	0.69	0.03	0.85	0.03
ΔG_{bind}(GB^{OBC2})	86.8	1.0	30.4	0.6	0.85	0.01	1.00	0.00
ΔG_{bind}(GBn)	93.3	1.1	43.3	1.1	0.89	0.01	0.99	0.03

Table 4. The sensitivity of the 3D-RISM-KH results with respect to some parameters. The default parameter set consists of 0.5 Å grid spacing, 12 Å solvent buffer cut-off and a 10^{-5} residual tolerance with cSPC/E water. Parameter sets indicate differences from the default parameters. Binding energies are given in kJ/mol and total CPU time in hours. Standard errors in the least significant digit are given brackets.

	Btn1	Btn2	Btn3	Btn4	Btn5	Btn6	Btn7	Total Time
				ΔG_{solv}				
PR-SPC/E	1310 (5)	1319 (5)	1285 (6)	305 (3)	203 (3)	179 (2)	133 (2)	2934
Default	1271 (5)	1275 (5)	1246 (5)	285 (3)	184 (3)	170 (2)	122 (2)	1754
10^{-3} Tolerance	1270 (5)	1274 (5)	1246 (5)	284 (3)	184 (3)	170 (2)	122 (2)	1097
20 Å Solvent Buffer	1275 (5)	1278 (5)	1250 (5)	290 (3)	188 (3)	173 (2)	124 (2)	2322
0.25 Å Grid Spacing	1271 (5)	- -	- -	- -	- -	- -	- -	1991
				ΔG_{pol}				
PR-SPC/E	1265 (5)	1277 (5)	1240 (6)	219 (3)	141 (2)	112 (3)	101 (2)	4754
Default	1223 (5)	1230 (5)	1200 (5)	196 (2)	122 (2)	102 (2)	90 (2)	3110
10^{-3} Tolerance	1222 (5)	1229 (5)	1199 (5)	195 (2)	121 (2)	102 (2)	90 (2)	1928
20 Å Solvent Buffer	1223 (5)	1230 (5)	1200 (5)	196 (2)	122 (2)	102 (2)	90 (2)	4108
				ΔG_{np}				
PR-SPC/E	46 (1)	42 (1)	45 (2)	86 (2)	61 (2)	67 (1)	32 (1)	1819
Default	48 (1)	44 (1)	47 (2)	89 (2)	62 (2)	68 (1)	32 (1)	1357

Figure 1. The seven biotin analogues used in this study. a) Btn1 (biotin), b) – g) Btn2–Btn7.

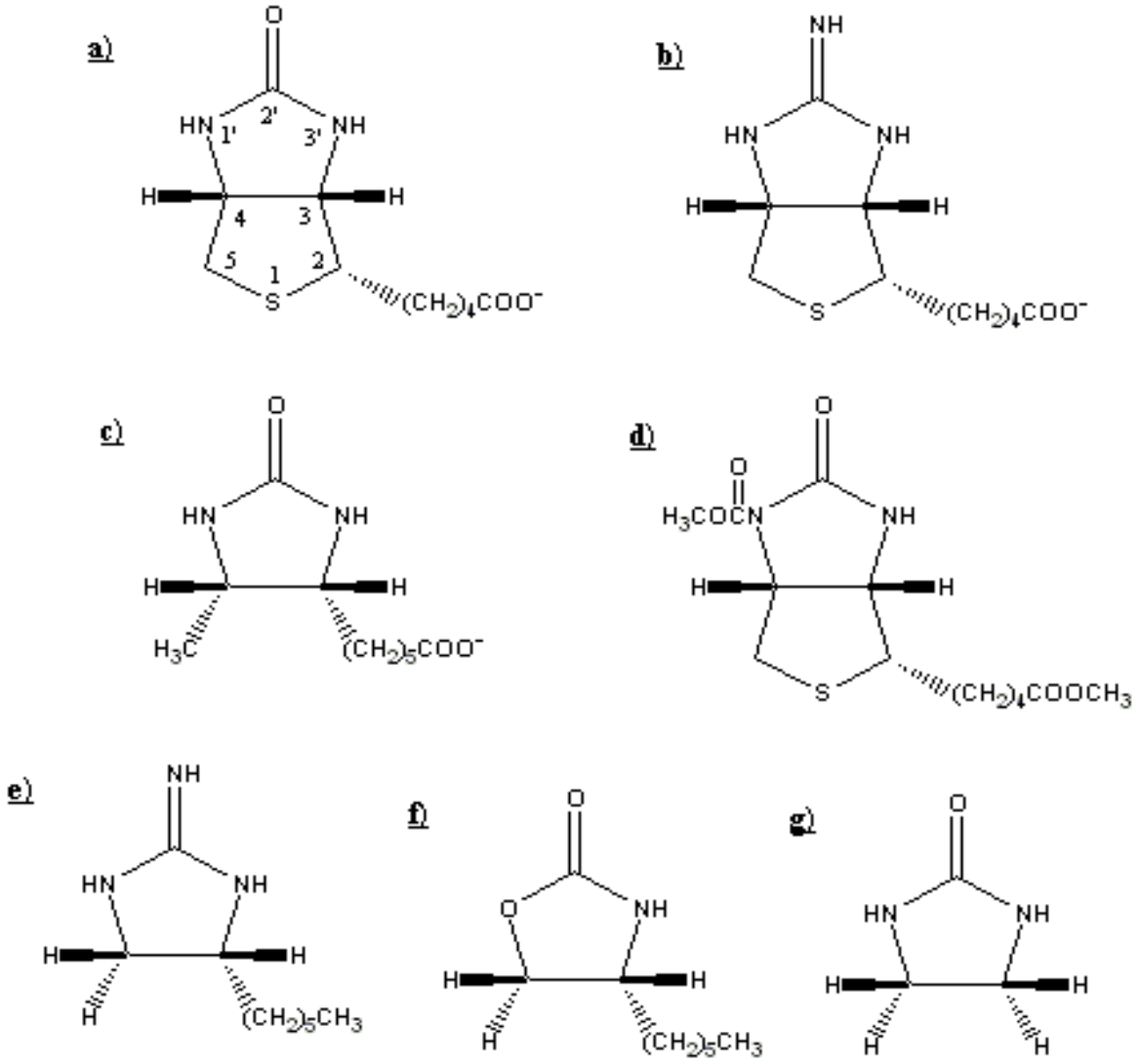


Figure 2. The non-polar solvation energy (ΔG_{np}) for the various ligands and methods.

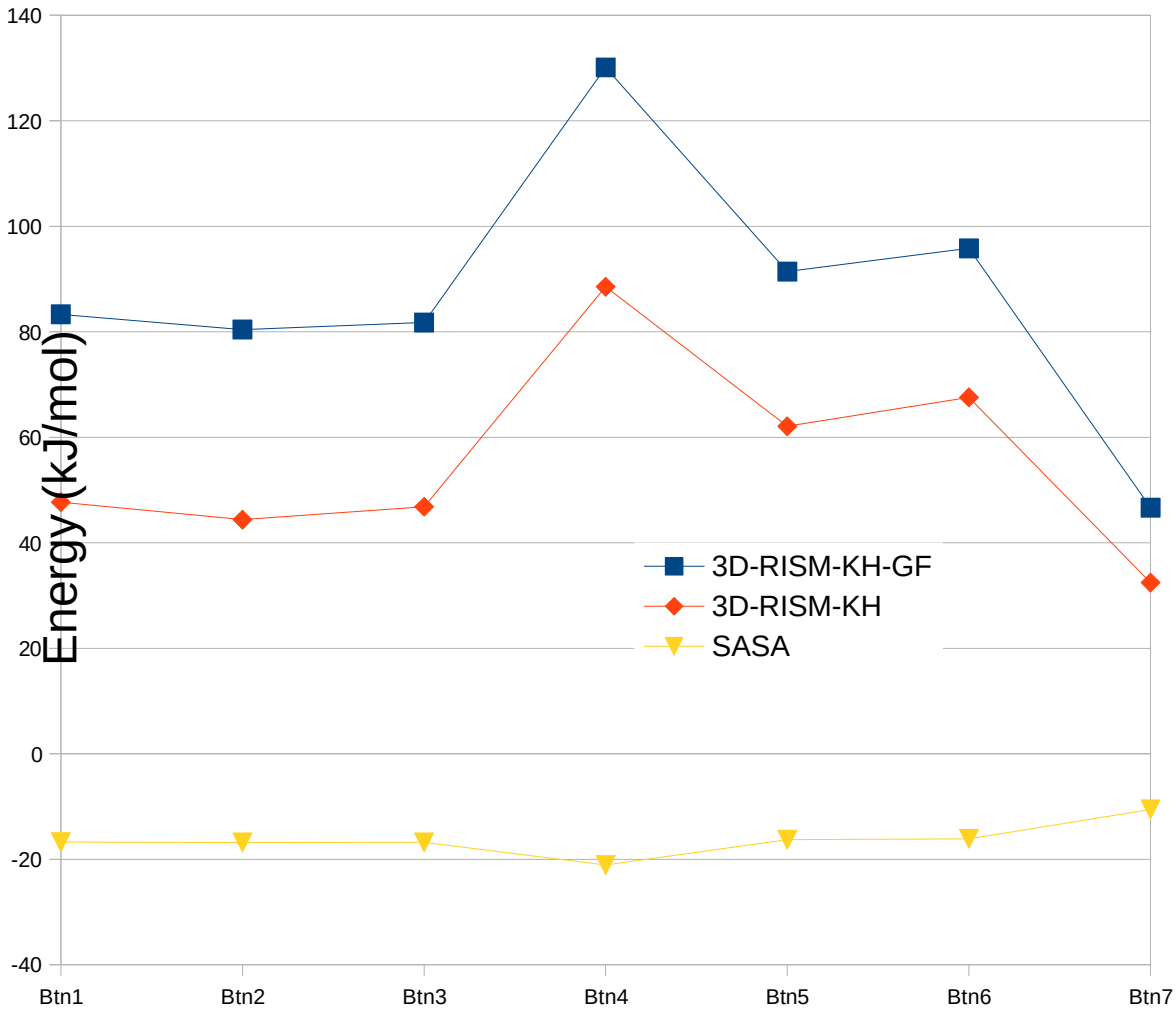
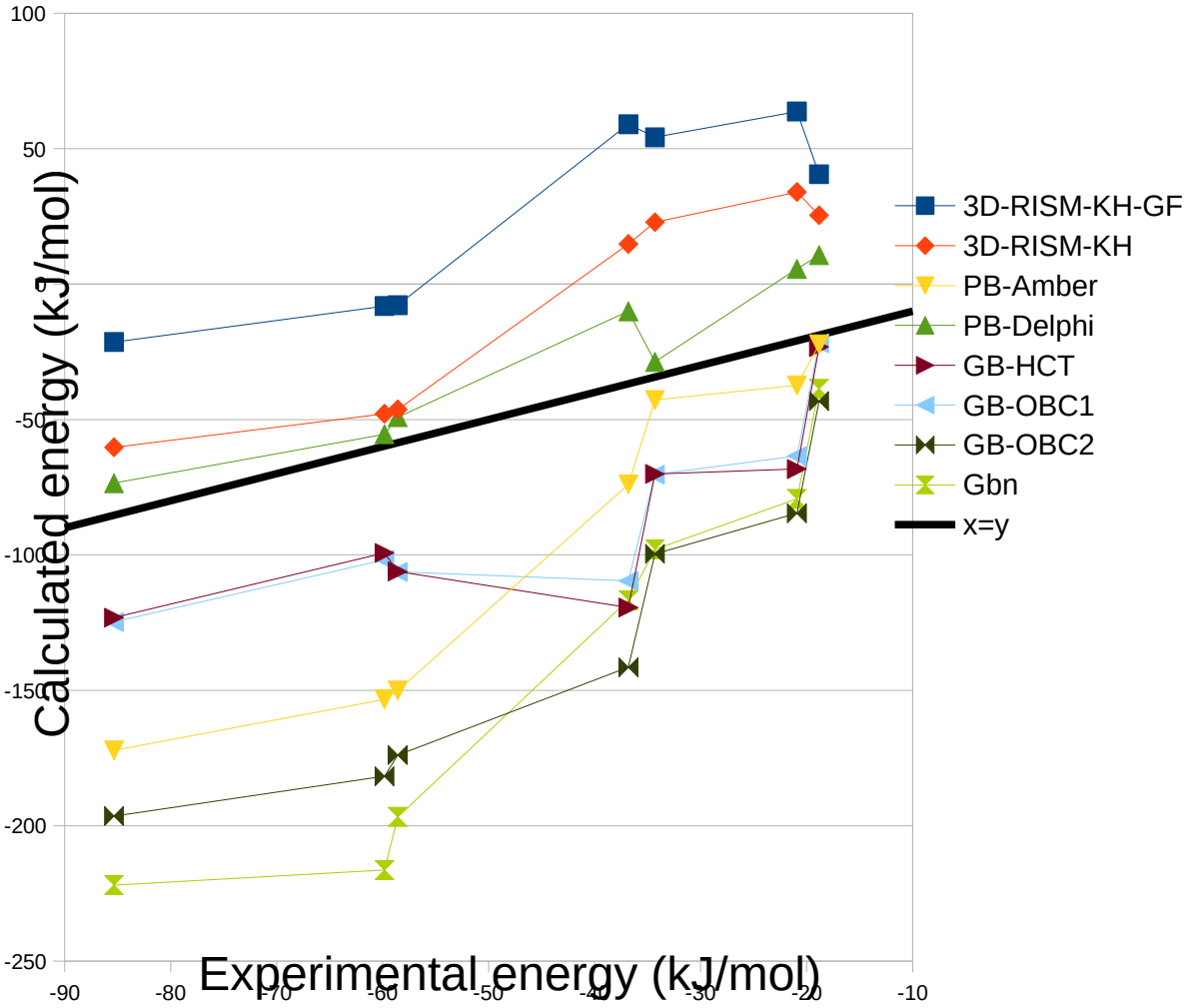


Figure 3. The calculated binding energies for the various methods, plotted against the experimental binding energies.



TOC graphics. Binding affinities of seven biotin analogues to avidin has been estimated with the MM/PBSA approach using 3D-RISM-KH, two Poisson–Boltzmann, and four generalised Born solvation methods, giving results that differ by up to 208 kJ/mol in absolute terms and and mean absolute deviations in the relative affinities of 10–43 kJ/mol.