J Comput Aided Mol Des (2012) 26:577–582
DOI 10.1007/s10822-012-9553-2

Force-field and quantum-mechanical binding study of selected SAMPL3 host-guest complexes

Nobuko Hamaguchi · Laszlo Fusti-Molnar · Stanislaw Wlodek

Received: 1 November 2011 / Accepted: 10 February 2012 / Published online: 25 February 2012

Abstract A Merck molecular force field classical potential combined with Poisson-Boltzmann electrostatics (MMFF/PB) has been used to estimate the binding free energy of seven guest molecules (six tertiary amines and one primary amine) into a synthetic receptor (acyclic cucurbit[4]uril congener) and two benzimidazoles into cyclic cucurbit[7]uril (CB[7]) and cucurbit[8]uril (CB[8]) hosts. In addition, binding enthalpies for the benzimidazoles were calculated with density functional theory (DFT) using the B3LYP functional and a polarizable continuum model (PCM). Although in most cases the MMFF/PB approach returned reasonable agreements with the experiment (±2 kcal/mol), significant, much larger deviations were reported in the case of three host-guest pairs. All four binding enthalpy predictions with the DFT/PCM method suffered 70% or larger deviations from the calorimetry data. Results are discussed in terms of the molecular models used for guest-host complexation and the quality of the intermolecular potentials.

Keywords Binding thermodynamics · Synthetic receptors · Intermolecular potential

Introduction

Synthetic receptors provide valuable model systems for studying the principles of biomolecular recognition because they bind guest molecules using the same physical interactions as proteins. A number of specific receptors have been synthesized that specifically bind different classes of guests—for example, alkali metal cations [1], carbohydrates [2, 3], dipeptides [4], etc. Theoretical studies of such host-guest systems result not only in the elucidation of binding mechanisms but, when confronted with high-quality experimental thermodynamic binding data, can improve computational methods and increase their predictive power. Cucurbit[n]urils [5, 6] are a special class of synthetic receptors that bind species containing both hydrophobic chains and cationic groups. In this study we prospectively predicted thermodynamic binding parameters (ΔG and ΔH) for three cucurbit[n]uril receptors with a small set of guest molecules for which experimental data were generated and presented as part of the SAMPL3 challenge.

N. Hamaguchi · L. Fusti-Molnar · S. Wlodek (✉)
OpenEye Scientific Software, 9 Bisbee Court, Suite D, Santa Fe, NM 87508, USA
e-mail: stan@eyesopen.com

Methods

Two types of computational approaches were used:

- a classical MMFF potential and a Poisson Boltzmann continuum solvent model, hereafter called MMFF/PB
- a quantum mechanical density functional theory (DFT) and polarizable continuum solvent model (PCM), hereafter called DFT/PCM.

MMFF/PB

Initial structures of host-guest complexes were obtained by docking [7] conformations produced by the OpenEye tool Omega [8]. The lowest energy pose was selected by minimizing the docked conformations with MMFF (van der Waals + Coulomb) and the solvent forces using the OpenEye tool Szybki [9]. The binding enthalpy was estimated as

$$\Delta H = E_{h-g} + E_{strain} \quad (1)$$

where $E_{h\text{-}g}$ is the host-guest interaction energy for the most favorable pose and E_{strain} is the guest strain energy obtained as the difference between the intrinsic energies of the bound and free ligand. The first term, $E_{h\text{-}g}$, is the sum of vdW and electrostatic interactions. The latter includes an electrostatic protein-ligand interaction in the presence of a polarizable solvent and the electrostatic desolvation of both host and guest. Desolvation energy may be defined as the loss of the electrostatic interaction energy between the solvent and the ligand (or host) upon binding, so it is calculated as the difference between the ligand (or host) electrostatic interaction with the solvent in the absence and presence of the uncharged partner:

$$E_d = \sum_{i=1} q_i(V_{i,0} - V_i) \quad (2)$$

where q_i are atomic charges for the host and guest, and V_i are electrostatic potentials on atoms for the isolated host and guest in solution (obtained from the solution of the Poisson equation), while $V_{i,0}$ are atomic potentials in the complex in which the partner has zero atomic charges. In all PB calculations, MMFF94 partial atomic charges were used.

The contribution to ΔG of binding due to entropy ($T\Delta S$) was evaluated with a recently published method [10]. Briefly, the method estimates the configurational entropy of a ligand in solution from basic statistical thermodynamics using the partition function

$$q = q_t \sum_{i=1}^{n_c} e^{-\frac{\epsilon_i}{kT}} q_{iv} q_{ir} \quad (3)$$

where q_t is the translational partition function, the summation is over the number of ligand conformers n_c, and q_{iv} and q_{ir} are the vibrational and rotational partition functions of conformer i, and ϵ_i is the sum of the internal energy and the solvation free energy of conformation i. For a host-bound ligand, the method relies on the assertion that the three translational and three rotational degrees of freedom of a ligand are transformed into six low-frequency vibrational modes of the trapped ligand, so Eq. 3 is reduced to

$$q = q_v \quad (4)$$

for a single binding pose, and to

$$q = \sum_{i=1}^{n_p} \exp\left(-\frac{\epsilon_i}{kT}\right) q_{iv} \quad (5)$$

for n_p binding poses, where the vibrational partition function q_v for a bound ligand with n atoms is calculated over $3n$ vibrational modes. The vibrational frequencies are evaluated from normal-mode analysis of the mass-weighted Hessian (the matrix of potential second derivatives with respect to coordinates). The Hessian is generated by converged quasi-Newton (QN) optimization of a ligand in solution or bound by the host. Solvation entropy of a ligand is split into electrostatic and hydrophobic components. The hydrophobic part is estimated from scaled particle theory (SPT) [11], and the electrostatic part by the differentiation with respect to the temperature of the Sheffield expression [12] for the free energy of ligand solvation.

The method used in the current study is a modification of the original approach [10] in two ways. The first is an improved algorithm for the evaluation of the partial solvation entropy of the bound guest molecule. We have found that eq. 26 in [10] can underestimate the fraction of a bound ligand surface area exposed to the solvent, so it is no longer used. Instead, a direct calculation of the fraction of the exposed ligand surface in the complex is done. The second change is specific to the host-guest systems studied here: the host desolvation entropy (ΔS_{des} in eq. 25 of [10]) is calculated as the fraction of its total solvation entropy, fS_{solv}, where f is taken as the fraction of the host surface buried upon complexation, rather than from the surface area expression $\gamma A_{buried}/T$ where γ is microscopic surface tension and T is temperature. The latter change is introduced for consistency with the calculation of the bound ligand partial solvation entropy, given that there is not a very large difference in size between the partners in the host-guest complex.

DFT/PCM

Due to the long cpu time needed, binding calculations were done for only Host2 (cucurbit[7]uril) and Host3 (cucurbit[8]uril). In this study, conformations of guest molecules generated initially with Omega [8] were further sampled for hydrogen atom positions, optimized with the MMFF94 force field in a vacuum, pruned for duplicates using a geometrical difference criteria of 0.1 Å RMSD, and finally re-optimized in solution with MMFF and PB solvent forces. Docking to the host molecules was done with a new intermolecular force field (still under development at OpenEye). The force field also contains vdW (E_{vdw}) and Fermi repulsion (E_{rep}) terms. The overall intermolecular energy is calculated as

$$E = E_{Coul} + E_{rep} + E_{vdw} \quad (6)$$

where the electrostatic term E_{Coul} is calculated using atomic multipoles obtained by distributed multipole analysis (DMA) [13, 14] as available in GAMESS code [15]. The multipoles were calculated at the DFT B3LYP 6-311G** level. The repulsion term E_{rep} was evaluated as

$$E_{rep} = \left\{ \sum_{ij} n_{ei} n_{ej} O(r_{ij}) F(r_{ij}) \right\}^{1/4} \quad (7)$$

where n_{ei} and n_{ej} are the total numbers of electrons in atoms i and j, $O(r_{ij})$ is the overlap between two normalized s-type Gaussians with exponents η_i and η_j, for the interatomic separation r_{ij}:

$$O(r_{ij}) = \pi^{3/2} \eta_{ij}^{3/2} \exp(-\eta_{ij} r_{ij}^2) \quad (8)$$

$$\eta_{ij} = \eta_i \eta_j / (\eta_i + \eta_j) \quad (9)$$

and $F(r_{ij})$ is a penalty function set to 1 at long interatomic distances and exponentially increases as the interatomic distance decreases. The vdW attractive term E_{vdw} is in the form

$$E_{vdw} = C \sum_{ij} (f_{d,ij} C_6 / r_{ij}^6) \quad (10)$$

where parameter C depends only on the selected charge model. Damping function $f_{d,ij}$ has the form

$$f_{d,ij} = 1/(1 + D(r_{ij}/R_6)^{-12}) \quad (11)$$

where parameter D depends only on the selected charge model. C_6 and R_6 are

$$C_6 = \sqrt{C_{6,i} C_{6,j}} \quad (12)$$

$$R_6 = R_{6,i} + R_{6,j} \quad (13)$$

where parameters $C_{6,i}$ and $R_{6,j}$ depend only on the element and are hybridization independent. All parameters ($C_{6,i}$, $R_{6,j}$ and η_i) have been fitted to reproduce high level CCSD(T) coupled-cluster energies extrapolated to Complete Basis Set database values according to the procedure published by Fusti-Molnar et al. [16].

For the QM calculations on Guest1, the 32 lowest energy docked conformations were selected, while for Guest3 the 16 lowest energy solutions were used. Single point calculations were done for all host-guest complexes with the B3LYP functional using the 6-31G** basis set and the solvent continuum PCM model. The basis set superposition error (BSSE) was calculated at the HF/6-31G** level. All QM calculations were done with GAMESS (version 1 OCT 2010) [15]. Our final results were calculated as Boltzmann averages over the sets of host-guest complexes described above.

Host models

An arbitrary decision was made regarding the charge state of the acyclic cucurbit[4]uril congener host, hereafter called Host1. It contains four carboxylic acid groups, and we assumed that they are aligned in two pairs in which only one carboxyl group in each pair is deprotonated. As a consequence, an intramolecular hydrogen bond is formed between the oxygen atoms of adjacent carboxyl groups as shown on Fig. 1.

Host2 and Host3 are neutral, relatively rigid molecules and no assumption was required regarding their conformation or charge state. Those molecules are powerful binders of cationic species [6]. It was found that in the presence of even small concentrations of alkali metal salts, cucurbit[n]urils bind two metal cations at the center of both portals. We have assumed, therefore, that in a sodium buffer the host binding process displaces Na^+ cations from the host:

$$CB[n](Na^+)_2 + Guest^{2+} \longrightarrow CB[n]Guest^{2+} + 2Na^+ \quad (14)$$

All the calculations for Host2 and Host3 included sodium ions. The guests studied are dications (see below), so the overall charge of the complex remains unchanged.

The crystal structure of cucurbit[8]uril (Host3) reveals the presence of up to eight water molecules inside the cage [17]. In an earlier unpublished study (Hamaguchi N, Wlodek S, 2011, personal communication), we found that neglecting these explicit water molecules bound inside the cucurbit[8]uril leads to a significant overestimation of the binding enthalpies of alkyltrimethylammonium ions using the MMFF/PB method described above. We used that finding in this study when predicting the binding thermodynamics for Host2 and Host3 with the MMFF/PB method, assuming further that water tetramers and water hexamers are present explicitly in Host2 and Host3, respectively. The final model for Host3 is shown in Fig. 2. In the case of DFT/PCM, no explicit water molecules were used in the model in order to reduce the cpu time.

Results

Predicted values of free energy of binding using our MMFF/PB method are shown in Table 1 and Fig. 3 for all

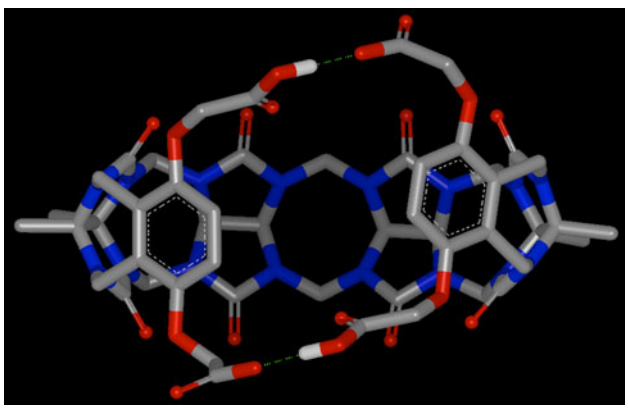


Fig. 1 Adopted model for Host1 with a total charge of −2 with two hydrogen bonds between the carboxylic groups aligned in two pairs

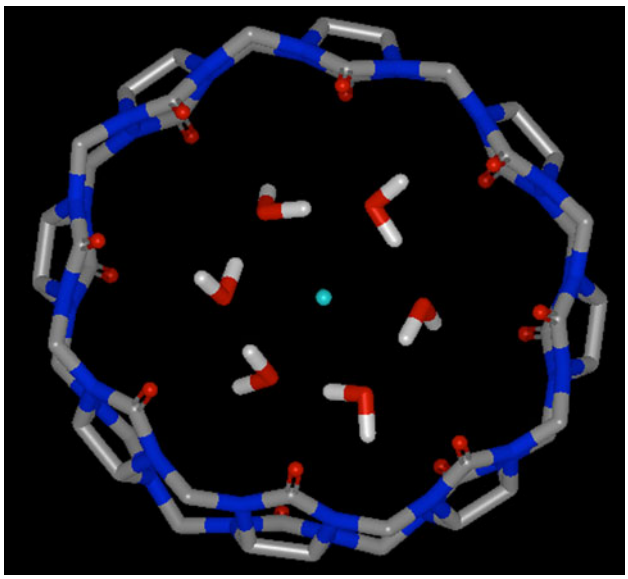


Fig. 2 Cucurbit[8]uril (Host3) with two Na^+ cations at the cavity portals (*blue sphere*) and six explicit water molecules inside the cage, adopted for MMFF/PB calculations in this study

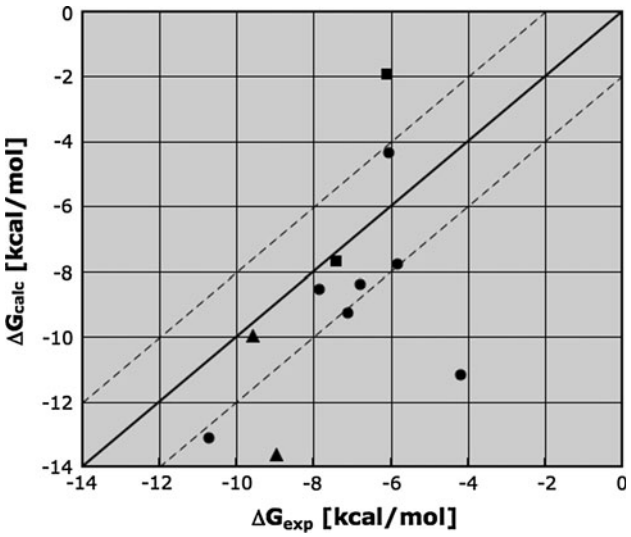


Fig. 3 Predicted (with our MMFF/PB method) versus experimental free energies of binding for Host1 (*circles*), Host2 (*squares*) and Host3 (*triangles*). The *solid diagonal line* represents ideal behavior, while the *broken lines* are offset by ±2 kcal/mol. Total RMS error is 3.1 kcal/mol

Table 1 Experimental and predicted ΔG of binding (in kcal/mol) using our MMFF/PB method

Host/guest	Experimental	Predicted
1/1	−5.84	−7.8
1/2	−7.1	−9.3
1/3	−6.8	−8.4
1/4	−4.17	−11.2
1/5	−6.06	−4.3
1/6	−10.72	−13.1
1/7	−7.85	−8.5
2/1	−6.12	−1.9
2/2	−7.43	−7.7
3/1	−9.6	−10.0
3/2	−8.99	−13.6

Structures of guest molecules are depicted in Fig. 4

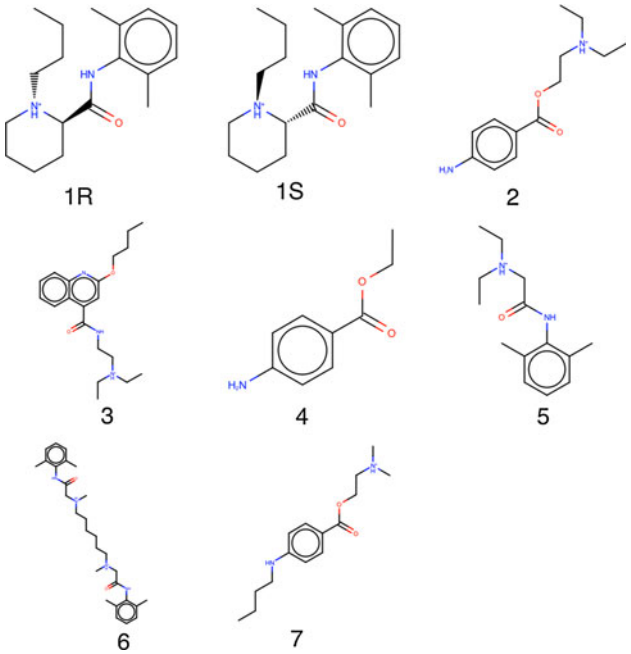


Fig. 4 Structures of guest molecules bound by Host1

three host and guest molecules. Structures of guest molecules bound by Host1 are depicted in Fig. 4. All but three data points (for Host1/Guest4, Host2/Guest1 and Host3/Guest2) are around ±2 kcal/mol or closer to the experimental values. The total RMS error is 3.1 kcal/mol, with a low correlation coefficient ($R^2 = 0.26$) of broad confidence limits (0.02–0.72), apparently due to the above-mentioned significant deviations for three data points. The Kendall tau rank correlation coefficient is 0.38.

Predicted data for the binding enthalpies for Host2 and Host3 using DFT/PCM method are shown in Table 2.

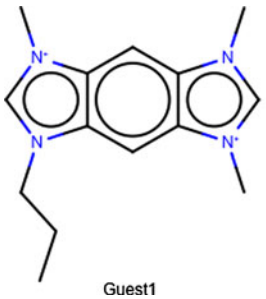
Discussion

Given the simplicity of our MMFF/PB method, the predicted free energies for Host1 might be considered satisfactory except for Guest4, for which the predicted binding is 2.7 times stronger than measured. Guest4 is the drug benzocaine, which has a pKa of 2.5 [18]—so at a neutral pH its charge is most likely zero. All the other guest

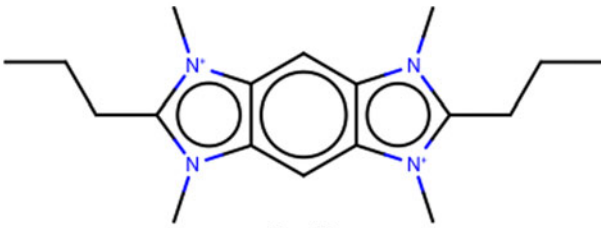
Table 2 Experimental and predicted ΔH of Host2 and Host3 (in kcal/mol) using the DFT/B3LYP method

Host/guest	Experimental	Predicted
2/1	−8.5	−2.5
2/2	−3.8	−10.5
3/1	−7.2	−13.0
3/2	−4.2	−8.1

the binding for this system, with no binding interaction at all calculated with the MMFF/PB method. Retrospective analysis of this particular case reveals that the reason for the lack of binding predictions is the strain energy of the guest, most likely accumulated in the propyl chain that is mostly encapsulated inside the cage of the host molecule. The calculated guest strain is about 3 kcal/mol higher than in the case of Guest2 bound by Host2.



Guest1



Guest2

molecules studied here that complex with Host1 have tertiary amino groups, so in a sodium buffer they are protonated with a charge of +1 or +2 in the case of Guest6, which has two such groups. The neutrality of Guest4 suggests that in a sodium buffer, the host molecule binds a Na^+ cation in addition to the Guest4 molecule. Restrospectively performed calculations in which a single sodium ion was placed inside the host cage resulted in a dramatically different free energy of binding: −3.4 kcal/mol, which agrees very well with the experimental value of −4.2 kcal/mol (Table 1). The overall correlation between the experimental and calculated data dramatically increases: the Kendall tau rank correlation coefficient increases from 0.38 to 0.64 and R^2 jumps from 0.26 to 0.70; in addition, its confidence limits have moved from 0.02–0.72 to 0.23–0.92. We conclude therefore that the MMFF/PB method correctly predicts binding thermodynamics for Host1/Guest4, provided that the model system is correctly defined.

Another large deviation from the experiment is found for Host2/Guest1 binding. Our predicted free energy of binding with the MMFF/PB method is 3.2 times too weak (−6.1 vs. −1.9 kcal/mol). Experimental data for Host2 and Host3 were derived from microcalorimetry measurements resulting in −8.5 kcal/mol for ΔH and −2.4 kcal/mol for $T\Delta S$. Our predicted values with the fast classical method are 0.2 and 2.1 kcal/mol, respectively. Restrospective analysis identified another pose, which produced −3.3 and 1.9 kcal/mol, respectively, still significantly underestimating ΔH of the binding for this system. The quantum mechanical calculation with the B3LYP/PCM model predicted only −2.5 kcal/mol for the enthalpy change upon binding. Both methods significantly underestimate

This strong disagreement with the experiment is probably due to the deficiency of the MMFF94 potential. It is important to note that our single-point B3LYP/PCM calculations were done using the geometry optimized with the MMFF/PB method, so they were affected indirectly by the MMFF94 potential as well.

The superposition of errors in the enthalpy and entropy of binding calculations for Host3/Guest2 with the MMFF/PB method led to a relatively large error of 52% compared with the experiment. As mentioned above, all other predictions done with this method were within or very close to ± 2 kcal/mol from the experiment.

The data in Table 2 show significant errors between the predicted ΔH of binding with the DFT/B3LYP/PCM method for all four investigated systems. The case of Host2/Host1, for which the binding enthalpy was largely overestimated, was already discussed above. In the remaining three cases, predicted binding enthapies are overestimated by 80–175%. Among a number of possible sources of error, there are two that we think might play an important role in current calculations:

- Deficiency of the dispersion energy correction in the B3LYP functional
- Lack of explicit solvent molecules for ionic systems.

The first issue has been recently recognized by Civalleri et al. [19] in the area of crystal structure prediction. These authors have proposed a special damping function for scaling the dispersion contribution to the intermolecular energy in crystalline urea when calculated with DFT/B3LYP. In order to test if the dispersion term in B3LYP is indeed a source of major error, we have performed retrospective single point calculations for randomly selected

single structures of the four investigated complexes using different DFT functionals, specifically TPSS, PBE and B97D. We have found that those functionals decrease the calculated binding energy with respect to B3LYP on average 3.6, 5.2 and 7.8 kcal/mol, respectively. This result suggests that the dispersion energy term in B3LYP is not a source of major error in the calculated binding energies for the investigated complexes.

The second issue has recently been addressed by a number of researchers by augmenting the continuum solvent model with explicit solvent molecules [20–22]. Bryantsev et al. found that neglecting the cluster structure of explicit water molecules solvating cations leads to an error in the calculated solvation energy >10 kcal/mol for H^+ and >30 kcal/mol for Cu^{2+}. Both Guest1 and Guest2 molecules are doubly charged species, so the effects of neglecting explicit water molecules might be of particular importance. Large errors in calculated solvation free energies of ionic species using a continuum PB solvent model have also been recognized by our company, OpenEye. For example, the ZAP continuum PB solvent model [23] overestimates the solvation free energy of Na^+ [24] by about 12.3 kcal/mol (Word M, Nicholls A, 2011, personal communication). Recently developed OpenEye technology for probing the area around the solvated species with a single explicit water molecule, called SZMAP [25], corrects the solvation free energy of ionic solutes to a very good agreement with the experiment. In the case of Na^+ it brings the solvation free energy to agreement within 0.5 kcal/mol (Word M, Nicholls A, 2011, personal communication). We hope to adapt this technology to improve the prediction of binding in host-guest systems.

Conclusion

Current results demonstrate the weakness of the MMFF94 potential for predicting the thermodynamics of binding. In spite of a number of successful predictions, a couple of disastrous failures (particularly visible in the case of the Host2/Guest1 system) make the MMFF94 force field unreliable for binding predictions. Quantum chemical methods at the popular DFT/B3LYP level with a common solvent continuum model (PCM) do not generate high quality predictions either. The most likely reason is the necessity to include explicit solvent molecules in the binding process.

Acknowledgments We thank Anthony Nicholls, Geoff Skillman, Mike Word, Chris Bayly and Matt Geballe for the encouragement to participate in the SAMPL3 challenge and for numerous discussions and suggestions. In particular, we want to acknowledge Chris for his advice regarding the charge state of Host1 and Mike Word for detailed discussions about the importance of augmenting continuum solvent models with explicit solvent molecules.

References

1. Wall SD, Meadows E, Barbour L, Gokel G (2000) Proc Natl Acad Sci 97:6271
2. Pal A, Bérubé M, Hall DG (2110) Angew Chem Int Ed 49:1492
3. Mazik M, Hartmann A (2010) Beilstein J Org Chem 6(9). doi:10.3762/bjoc.6.9
4. Castilla A, Conn M, Ballester P (2010) Beilstein J Org Chem 6(5). doi:10.3762/bjoc.6.5
5. Rekharsky M, Inoue Y (2007) Netsu Sokutei 34:232
6. Ma D, Zavalij Y, Isaacs L (2010) J Org Chem 75:4786
7. Fred2.2.5 (2010) OpenEye Scientific Software, Inc., Santa Fe
8. Omega2.4.3 (2006) OpenEye Scientific Software, Inc., Santa Fe
9. Szybki1.5.0 (2010) OpenEye Scientific Software, Inc., Santa Fe
10. Wlodek S, Skillman A, Nicholls A (2010) J Chem Theory Comput 6:2140
11. Pierotti R (1976) Chem Rev 76:717
12. Grant J, Pickup B, Sykes M, Kitchen C, Nicholls A (2007) Chem Phys Lett 441:163
13. Stone A (1981) Chem Phys Lett 83:233
14. Stone A (2005) J Chem Theory Comput 1:1128
15. Schmidt M, Baldridge K, Boatz J, Elbert S, Gordon M, Jensen J, Koseki S, Matsunaga N, Nguyen K, Su S, Windus T, Dupuis M, Montgomery J (1993) J Comput Chem 14:1347
16. Molnar LF, He X, Wang B, Merz K (2009) J Chem Phys 131:065102
17. Kim J, Jung I, Kim S, Lee E, Kang J, Sakamoto S, Yamaguchi K, Kim K (2000) J Am Chem Soc 122:540
18. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo A, Wishart D (2011) Nucleic Acids Res 39:D1035
19. Civalleri B, Zicovich-Wilson C, Valenzano L, Ugliengo P (2008) CrystEngComm 10:405
20. Kelly C, Cramer C, Truhlar D (2006) J Phys ChemA 110:2493
21. Bryantsev V, Diallo M, Goddard-III W (2008) J Phys ChemB 112:9709
22. da Silva E, Svendsena H, Merz K (2009) J Phys ChemA 113:6404
23. Grant J, Pickup B, Nicholls A (2001) J Comput Chem 22:608
24. Burgess J (1978) Metal ions in solution. Ellis Horwood Ltd., Chichester
25. Szmap1.0.0 (2010) OpenEye Scientific Software, Inc., Santa Fe