

Analysis of SM8 and Zap TK calculations and their geometric sensitivity

Benjamin A. Ellingson · A. Geoffrey Skillman · Anthony Nicholls

Received: 12 January 2010 / Accepted: 31 March 2010 / Published online: 30 April 2010
© Springer Science+Business Media B.V. 2010

Abstract A prospective study of aqueous solvation energies was done using the SM8 and Zap TK models for a variety of geometries. CM4M charges calculated with M06 and M06-2X were found to yield similar results for the SM8 model. Zap TK calculations were primarily done with AM1BCC charges but limited attempts to use charges derived from DFT showed promise. The OMEGA application quickly generated conformations that performed well with both solvation models, while the use of computationally expensive DFT optimized geometries yielded increased RMSE and MSE. It is shown that increasing levels of gas phase geometry optimization yield increasingly unfavorable solvation energy for single conformer models.

Keywords SAMPL · Solvation · Zap · SM8 · M06 · M06-2X · OMEGA · QUACPAC · SZYBKI · CM4M

Introduction

In this paper, we discuss our submissions to the solvation energy portion of the SAMPL-09 challenge [1]. This is a blind challenge where contributors do not have access to the experimental data until after all of the submissions have been compiled. As Richard Feynman eloquently stated, “The first principal is that you must not fool yourself—and you are the easiest person to fool [2].” It is easy for scientists to convince themselves that obviously they should have done things a

certain way after seeing initial results, whereas a blind challenge prevents a cyclical tuning of a method to a specific data set by only allowing comparison to experimental results at the very end. In addition to the official submissions to the challenge, several retrospective calculations are presented to help explain the unexpected results. The purpose of these calculations is not to do better, which would violate the spirit of a blind challenge, but rather to clearly establish trends that were not known a priori.

The solvation methods discussed in this paper use an implicit solvent model and a single conformer. Zap TK [3] is a Poisson–Boltzmann (PB) solver that models solvent as a dielectric and solute as low dielectric with point charges. Zap TK differs from other PB solvers by using atom-centered Gaussians to create a smooth transition between dielectric regions, which greatly diminishes the numerical instabilities inherent in a grid-based approach [3]. SM8 [4] is a Generalized Born (GB) method, which uses descriptors such as refractive index and bulk surface tension in addition to the dielectric to model a solvent. The GB method does not explicitly solve the Poisson equation and thus is ostensibly faster; however, when combined with conformation and charge calculations, the difference in performance between well implemented PB and GB is trivial.

A single conformer model assumes that the geometric conformation is the same for both gas phase and solution phase. This approximation is known not to be physical; the lowest energy conformation in gas phase and solution phase are rarely similar for all but the most rigid molecules. However, assuming a single conformation allows the vibrational–rotational partition function to reasonably be ignored, which simplifies the problem and appears to yield a fortuitous cancellation of error. The single conformation model has been shown to do well versus a simple multi-conformer model [5].

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9355-3) contains supplementary material, which is available to authorized users.

B. A. Ellingson (✉) · A. G. Skillman · A. Nicholls
OpenEye Scientific Software, Santa Fe, NM, USA
e-mail: ben@eyesopen.com

The SM8 [4] method has done well retrospectively on data from previous SAMPL challenges [6, 7] and the authors hoped to replicate the success prospectively in this challenge. The authors submitted SM8 results for a variety of calculations using the M06 and M06-2X functionals [8]. These were done at MMFF94s [9, 10] gas phase global minima and also at geometries further optimized with the functionals. We found that the more computationally expensive geometries did quite poorly. Also, the SM8 submission from the Minnesota group [11] performed better than our submissions. Further analysis reveals a strong geometric sensitivity of model and that the use of OMEGA [12] geometries was key to their success.

For a baseline calculation, a workflow was created that was extremely fast and required minimal human interaction with the data. This workflow took the lowest energy conformer from the OMEGA application, minimized with SZYBKI [13] using the MMFF94s forcefield, used QUACPAC [14] to assign AM1BCC [15, 16] partial charges, and calculated the solvation energy with Zap TK [3]. This workflow is able to calculate the entire SAMPL dataset in a few seconds on commodity hardware. Surprisingly, this was the best submission from the authors. Retrospective Zap TK calculations were done at a variety of geometries in an attempt to explain this result and it was found that the Zap TK model has the same strong geometric sensitivity as SM8.

CM4M [17] charges commonly used with the SM8 method were extracted from the output, and these charges were used for Zap TK calculations to see what sort of effect they would have. The previous attempt at using high-level DFT charges for SAMPL was very successful [5]. The results presented below show reason to be hopeful that high-level charges yield better results. Unfortunately, this experiment was done at what is now known to be a poor set of geometries, which prevents us from comparing these results to other methods in a meaningful way.

Methods

Geometries

Several sets of geometries are used in this paper. The geometries distributed with the SAMPL data set were generated with the OMEGA version 2.3.2 application from OpenEye Scientific Software. With default settings, this application uses a modified version of the MMFF94s forcefield where Coulomb terms have been removed. For small molecules, such as those in the SAMPL data set, complete sampling of all the torsion rules is possible. This set of geometries is referred to throughout the paper as the OMEGA set.

The SM8 submission from Ribeiro et al. [11] used the OMEGA geometries except for a further optimization of

glycerol as discussed in their paper found in this issue. This set of geometries is referred to as OMEGA*. Also, an optimization of the OMEGA geometries was done using the MMFF94 forcefield as implemented in the SZYBKI version 1.3.2 application from OpenEye.

Another set of geometries was calculated using 1000 trials of random coordinate embedding, distance geometry relaxation and then MMFF94s optimization. These geometries are assumed to be at the global MMFF94s minimum and are referred to as the MMFF set throughout the paper.

Further optimization was done at both the M06/6-31+G** and M06-2X/6-31+G** levels using the MMFF geometries for input. The 6-31+G** basis set [18] was chosen for optimization in order to replicate the method used in the CM4M paper [17]. All M06 suite calculations were done using the GAMESSPLUS v2008-2 program [19]. GAMESSPLUS v2008-2 is a modified version of the April 11, 2008 (R1) release of GAMESS [20], which adds M06 suite and SM8 functionality as well as other features. The GAMESSPLUS distribution from the University of Minnesota modifies the source code of GAMESS, which is then compiled by the end user.

Cartesian coordinates for the MMFF, M06, and M06-2X sets of geometries are reported in Electronic Supplementary Material.

Charges

AM1BCC charges were calculated using QUACPAC 1.3.1, which includes OpenEye's implementation of the work of Bayly and colleagues. CM4M charges were calculated for the M06 and M06-2X functionals in GAMESSPLUS using the 6-31G* basis set [21]. Charges were calculated at the listed geometry for each data set without further optimization.

Solvation models

Zap TK from the OpenEye version 1.7.0-3 toolkit release was used for Poisson-Boltzmann calculations. The ZAP9 radii [5] were used for Zap TK SAMPL submissions unless otherwise noted. Also discussed in this paper is a submission from Nicholls [22] where the radii have been parameterized for CM4M charges. Default parameters were used, which include an internal dielectric of 1 and an external dielectric of 80. SM8 calculations were done with the GAMESSPLUS v2008-2 program using default parameters.

Statistics

Root mean square error (RMSE), mean unsigned error (MUE), and mean signed error (MSE) are presented for all methods discussed. These statistics were calculated using

the amended values for cyanuric acid and glycerol. Due to missing basis sets for iodine, 5-iodouracil is not discussed in this paper and is left out of all calculated statistics despite being part of the official SAMPL set.

Results and discussion

SM8 calculations

The results for the SM8 calculations along with the given experimental measurements are presented in Table 1. The 1st column containing SM8 calculations done at the OMEGA* geometries is an official SAMPL submission from Ribeiro et al. [11] The other sets of SM8 calculations are official SAMPL submissions from the authors.

Columns 2 and 4 are calculated at MMFF globally optimized geometries using CM4M charges at the M06-2X/6-31G* and M06/6-31G* levels, respectively. Comparing these two sets of calculations show that the M06-2X and M06 functionals yield similar results when used for this purpose. The calculations done with M06 had a slightly higher MSE, 0.68 versus 0.56 kcal/mol, but the two submissions had nearly identical results of 2.20 and 2.21 kcal/mol for RMSE. The solvation energies from these submissions are plotted in Fig. 1, where the systematically higher solvation energies from M06 calculations can be seen.

Results for further geometry optimizations at the M06-2X/6-31+G** level are given in column 3 and optimizations at the M06/6-31+G** level are presented in column 5. Both of these sets perform significantly worse than the calculations done at the MMFF geometries, with an RMSE of 2.51 kcal/mol for the M06-2X/6-31+G** geometries and 2.78 kcal/mol for the M06/6-31+G** geometries. Also, the MSE rose significantly for both functionals following optimization, with a rise from 0.56 to 1.25 kcal/mol for M06-2X and a rise from 0.68 to 1.54 kcal/mol for M06.

Comparison of SM8 calculations at the OMEGA*, MMFF, and M06-2X/6-31+G** geometries show that the RMSE worsens and MSE increases with increasing optimization. The signed errors of these methods for individual molecules are plotted in Fig. 2. This figure shows that the calculated solvation energy becomes systematically less favorable (higher energy) as optimization increases.

Zap TK calculations

The Zap TK calculations are presented in Table 2. The best submission from the authors is presenting in column 2 with an RMSE of 2.38 kcal/mol. This submission used OMEGA geometries further optimized with SZYBKI, AM1BCC charges assigned by QUACPAC and solvation calculations were done using Zap TK; this only required a few seconds

of computer time on a single core of a Q6600 processor to calculate the entire SAMPL set.

Additional Zap TK calculations were performed retrospectively at several readily available geometries used for other submissions from the authors. These include the OMEGA geometries (column 1), MMFF globally optimized geometries (column 3), and M06-2X geometries (column 4). Columns 1 through 4 were all done using the Zap TK solvation model with AM1BCC charges at increasing levels of geometry optimization. Comparing the MSE of these calculations show that the solvation model predicts less favorable solvation energies with increasing levels of geometry optimization. This is the same geometric sensitivity seen with the SM8 model. The signed errors of individual molecules for columns 1, 3, and 4 are plotted in Fig. 3.

CM4M charges were extracted from the SM8 M06-2X/6-31G* calculation done at the M06-2X/6-31++G** geometry. A Zap TK calculation was done using these charges and is presented in column 5 of Table 2. As expected, this combination performed very poorly, likely due to the fact that ZAP9 radii are parameterized for AM1BCC charges. A reparameterization of 11 atomic radii [22] specifically for the CM4M charges coupled with the Zap TK method led to significantly better results, as presented in column 6. The Zap TK calculation with reparameterized radii did not perform well compared to the best methods; however, this calculation is now known to have been done at a poor set of geometries as discussed in the following section. This calculation had a lower RMSE than the calculation with AM1BCC charges and ZAP9 radii, so improvement was seen and further study is required to determine how this method will perform with better geometries.

Calculated solvation energy less favorable as optimization increases

The OMEGA application, using default options, generates conformations with a version of the MMFF94s forcefield where electrostatic contributions are removed. This removal causes the molecule to not be concerned with self-attraction, which typically yields a conformation that is able to easily interact with its exterior environment. This sort of conformation works well for the primary use of OMEGA, ligand conformations in protein active sites, but also works well for solvent. The best SM8 submission was performed at this geometry.

The second level of optimization (OMEGA plus SZYBKI) adds back in the electrostatic contributions but uses the first OMEGA geometry as the starting point. This allows the molecule to adjust into a better gas phase geometry. However, the molecule may be trapped in the original solvent-friendly well and not allowed to fully relax, which can be seen in the differences between

Table 1 Solvation energies for SM8 calculations along with experimental results in units of kcal/mol

Model	SM8	SM8	SM8	SM8	SM8	Expt
CM4M charges	M06-2X	M06-2X	M06-2X	M06	M06	
Geometry	OMEGA*	MMFF	M06-2X	MMFF	M06	
4-Nitroaniline	-10.48	-9.95	-9.84	-10.58	-10.66	-9.45
5-Bromouracil	-16.39	-16.29	-15.06	-15.96	-14.67	-18.17
5-Chlorouracil	-16.59	-16.59	-15.23	-16.23	-14.80	-17.74
5-Flurouracil	-17.38	-17.35	-15.71	-16.95	-15.15	-16.92
5-Trifluoromethyluracil	-16.52	-17.07	-14.89	-16.65	-14.26	-15.46
6-Chlorouracil	-14.24	-14.65	-13.18	-14.34	-12.73	-15.83
Acetylsalicylic acid	-11.93	-10.92	-10.09	-10.78	-9.79	-9.94
Butyl paraben	-8.42	-8.30	-8.08	-8.31	-8.08	-8.72
Caffeine	-11.90	-12.47	-12.26	-11.93	-11.46	-12.64
Cyanuric acid	-20.32	-21.59	-18.44	-21.00	-17.71	-18.06
D-Glucose	-25.29	-19.69	-18.38	-19.78	-16.97	-25.47
Diffunisal	-13.96	-8.42	-7.92	-8.24	-7.46	-9.40
Dimethyl methylphosphonate	-9.53	-6.21	-5.35	-6.18	-5.17	
D-Xylose	-21.71	-16.18	-14.97	-16.22	-14.76	-20.52
Ethyl paraben	-9.50	-8.64	-8.50	-8.65	-8.44	-9.20
Flurbiprofen (racemic)	-9.03	-7.72	-7.55	-7.49	-7.22	-8.42
Glycerol	-17.04	-11.54	-10.93	-11.57	-10.87	-13.43
Hexachlorobenzene	-0.23	-0.20	-0.13	-0.21	-0.14	-2.33
Hexachloroethane	-0.04	-0.05	-0.01	-0.05	-0.02	-1.41
Hexamethyl phosphoramidate	-7.47	-7.07	-4.82	-6.85	-4.38	
Ibuprofen (racemic)	-6.88	-5.64	-5.34	-5.50	-5.17	-7.00
Isothiazole	-6.22	-6.18	-5.98	-6.09	-5.98	
Ketoprofen (racemic)	-12.56	-11.71	-11.01	-11.55	-10.75	-10.78
Methyl dimethylphosphinate	-9.79	-6.68	-5.83	-6.61	-5.68	
Methyl paraben	-9.85	-9.01	-8.76	-9.02	-8.69	-9.51
Methyl phenyl sulfone	-13.55	-13.10	-13.29	-13.13	-13.15	
Methyl phenyl sulfoxide	-11.53	-11.26	-11.13	-11.05	-10.58	
Naproxen	-10.90	-9.54	-9.09	-9.31	-8.67	-10.21
Octafluorocyclobutane	3.60	3.56	3.87	3.54	4.03	3.43
Oxazole	-5.94	-5.96	-5.83	-5.66	-5.54	
Pentachloronitrobenzene	-1.73	-1.34	-0.99	-1.24	-0.89	-5.22
Phthalimide	-11.99	-12.21	-11.04	-12.06	-10.96	-9.61
Propyl paraben	-9.24	-8.42	-8.25	-8.42	-8.20	-9.37
Sulfolane	-13.26	-13.22	-14.09	-13.29	-13.86	-8.61
Thiazole	-4.86	-4.89	-4.69	-4.66	-4.38	
Trifluoromethyl phenyl sulfone	-8.32	-7.71	-8.04	-7.83	-8.19	
Trifluoromethyl phenyl sulfoxide	-7.63	-7.25	-7.11	-7.24	-6.98	
Trimethyl orthotrifluoroacetate	-1.67	-1.84	-1.16	-1.85	-0.94	-0.8
Trimethyl phosphate	-7.46	-5.42	-5.73	-5.46	-5.41	-8.7
Uracil	-16.76	-16.91	-15.70	-16.52	-15.16	-16.59
MUE	1.41	1.66	1.82	1.69	2.07	
MSE	-0.45	0.56	1.25	0.68	1.54	
RMSE	1.89	2.21	2.51	2.20	2.78	

Calculations at the OMEGA* geometries were submitted by Riberia et al. and the remaining SM8 calculations were submitted by the authors. CM4M charges used the 6-31G* basis set and DFT geometries used the 6-31+G** basis set. Molecules without experimental data are from the investigatory data set

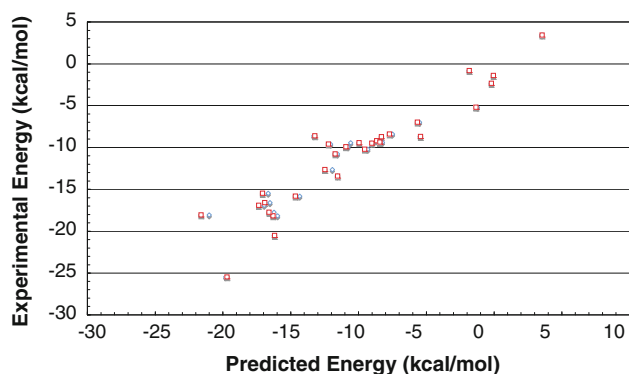


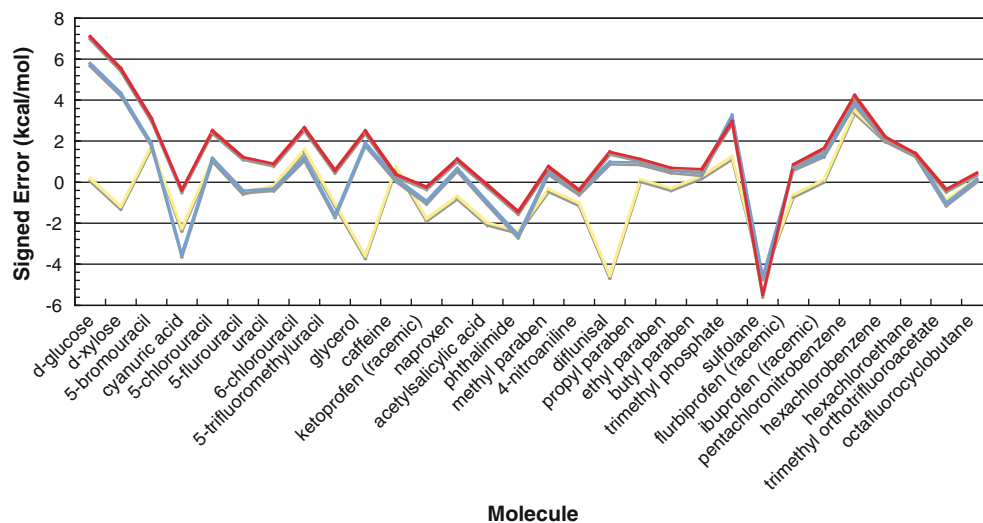
Fig. 1 Comparison of experimental energy with predicted energy of M06 and M06-2X using SM8 solvation model in kcal/mol with CM4M charges at MMFF geometries. M06 calculations are represented by blue diamonds and M06-2X calculations are represented by red squares

columns 2 and 3 in Table 2. This second level of optimization proved to be a useful compromise and the best Zap TK submission was performed at this geometry. It is noteworthy that no SM8 submissions were performed at this geometry and the performance of such a calculation is worthy of further study.

The third level of optimization involved a global search for the lowest energy conformation using the full MMFF94s forcefield. This can be thought of as the best gas phase geometry according to the MMFF94s forcefield. The gas phase geometry assumes that there are no exterior forces and causes self-attraction to play a large role in the optimal geometry.

The fourth level of optimization attempted by the authors was a full DFT optimization from the MMFF starting point using both the M06 and M06-2X functionals. It is not required in general that DFT will optimize to a geometry that is even less favorable to solvation than a forcefield,

Fig. 2 Signed error in kcal/mol of SM8 with CM4M charges using M06-2X/6-31G* at various geometries. Yellow line denotes OMEGA geometries with reoptimized glycerol, blue line denotes MMFF geometries, and red line denotes M06-2X/6-31+G** geometries. Molecules ordered by increasing solvation energy from left to right



since they are different models and the results could be reversed depending on parameterization. However, for the case of the M06 suite and the MMFF94s forcefield, we find that the M06 and M06-2X functionals yield the least favorable geometries for both the SM8 and Zap TK solvation models. In particular, low barrier torsions such as hydroxyl rotations can point towards a negatively charged region of the molecule, which is an electrostatic advantage in the gas phase but is a disadvantage in solution. Additionally, ring systems with multiple low energy conformations can adjust to accommodate the optimal orientation of polar groups. This is highlighted in Figs. 4 and 5, where the OMEGA and M06-2X geometry are compared for D-glucose. The largest differences between the OMEGA and M06-2X solvation energies are in D-glucose, D-xylose, glycerol, and diflunisal, all of which have multiple hydroxyl rotors.

The differences in calculated solvation energy with regard to geometry are consistent with the findings of Mobley et al. [23]. That study also found that the best vacuum phase geometry had a positive MSE when calculating solvation energy. The study of Mobley et al. found that the best solution phase geometry had a small negative MSE. We did not do a rigorous search for the best solution phase geometry; however, the set of OMEGA geometries also yield a small negative MSE, which is consistent with a good geometry in solution phase. Mobley et al. did not study the effect of increasing levels of quantum mechanical geometry optimization, which we find to yield increasingly large positive errors in calculated solvation energy.

Conclusions

The best submissions for both SM8 and Zap TK were found in the lowest energy well from OMEGA. The best

Table 2 Solvation energies for Zap TK calculations in units of kcal/mol

Model	Zap TK	Zap TK	Zap TK	Zap TK	Zap TK	Zap TK, opt radii	
Charges	AM1BCC	AM1BCC	AM1BCC	AM1BCC	CM4M	CM4M	
Geometry	OMEGA	OMEGA	SZYBKI	MMFF	M06-2X	M06-2X	
4-Nitroaniline	-10.12	-9.86		-10.21	-10.22	-13.14	-11.28
5-Bromouracil	-15.08	-15.17		-15.36	-13.96	-15.32	-15.67
5-Chlorouracil	-15.19	-15.23		-15.41	-14.04	-16.36	-16.30
5-Flurouracil	-14.79	-14.87		-14.83	-13.40	-16.85	-17.02
5-Trifluoromethyluracil	-15.03	-15.39		-15.51	-13.81	-16.53	-16.83
6-Chlorouracil	-13.14	-13.46		-13.51	-12.13	-12.73	-12.69
Acetylsalicylic acid	-12.17	-10.82		-10.73	-10.27	-9.35	-8.79
Butyl paraben	-8.43	-8.36		-8.45	-8.42	-7.34	-8.19
Caffeine	-13.47	-14.09		-13.95	-13.73	-12.73	-11.04
Cyanuric acid	-19.40	-20.14		-20.42	-17.63	-16.78	-16.37
D-Glucose	-30.40	-25.33		-19.78	-18.71	-13.67	-19.48
Diffunisal	-14.25	-13.35		-8.01	-7.75	-4.86	-5.44
Dimethyl methylphosphonate	-13.61	-13.37		-12.42	-12.15	-10.51	-6.05
D-Xylose	-23.73	-15.95		-16.43	-15.29	-10.40	-15.51
Ethyl paraben	-9.40	-8.84		-9.00	-8.98	-7.40	-8.31
Flurbiprofen (racemic)	-9.57	-8.82		-8.23	-8.19	-5.28	-5.70
Glycerol	-18.33	-12.54		-12.14	-11.40	-7.80	-11.32
Hexachlorobenzene	1.83	1.86		1.88	2.01	2.30	2.66
Hexachloroethane	1.61	1.58		1.58	1.68	1.91	2.20
Hexamethyl phosphoramidate	-18.38	-17.69		-17.30	-15.38	-9.21	-5.96
Ibuprofen (racemic)	-7.89	-6.75		-6.63	-6.54	-4.57	-4.79
Isothiazole	-4.70	-4.73		-4.67	-4.52	-7.36	-7.92
Ketoprofen (racemic)	-13.13	-11.64		-12.23	-11.58	-9.91	-9.09
Methyl dimethylphosphinate	-14.78	-14.80		-13.05	-12.69	-11.92	-7.13
Methyl paraben	-9.81	-9.21		-9.30	-9.27	-7.83	-8.79
Methyl phenyl sulfone	-12.82	-12.51		-12.66	-12.80	-12.35	-9.01
Methyl phenyl sulfoxide	-9.04	-8.82		-8.87	-8.86	-13.29	-9.57
Naproxen	-11.62	-10.09		-10.13	-10.03	-7.42	-8.08
Octafluorocyclobutane	2.43	2.43		2.43	2.42	2.46	2.89
Oxazole	-6.39	-6.45		-6.38	-6.59	-5.26	-5.27
Pentachloronitrobenzene	1.10	1.26		1.24	1.39	-0.15	1.32
Phthalimide	-10.40	-10.56		-10.49	-9.76	-9.49	-8.78
Propyl paraben	-9.00	-8.54		-8.58	-8.55	-6.99	-8.12
Sulfolane	-12.35	-12.35		-12.30	-13.50	-13.45	-9.39
Thiazole	-4.64	-4.67		-4.58	-4.59	-4.35	-4.44
Trifluoromethyl phenyl sulfone	-4.81	-6.15		-6.34	-6.57	-5.36	-3.46
Trifluoromethyl phenyl sulfoxide	-6.34	-4.62		-4.67	-4.63	-6.96	-4.66
Trimethyl orthotrifluoroacetate	-2.54	-2.47		-2.48	-2.69	-1.68	-2.34
Trimethyl phosphate	-12.48	-11.93		-11.95	-12.68	-10.25	-5.87
Uracil	-15.21	-15.48		-15.44	-14.24	-16.37	-15.82
MUE	2.22	1.77		1.87	2.22	2.81	2.22
MSE	-0.46	0.40		0.72	1.23	1.94	1.80
RMSE	2.76	2.38		2.49	2.98	3.89	2.75

The second, fifth, and sixth columns were official SAMPL submissions from the authors. The remaining calculations in this table were done retrospectively. CM4M charges were calculated using M06-2X/6-31G*. M06-2X geometries were calculated with the 6-31+G** basis set. Molecules without experimental data are from the investigatory data set

Fig. 3 Signed error in kcal/mol of Zap TK using AM1BCC charges at various geometries. *Yellow line* denotes OMEGA geometries, *blue line* denotes MMFF geometries, and *red line* denotes M06-2X/6-31+G** geometries. Molecules ordered by increasing solvation energy from left to right

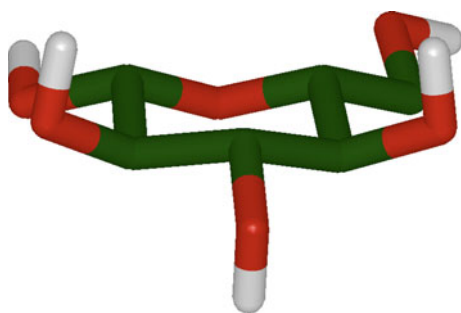
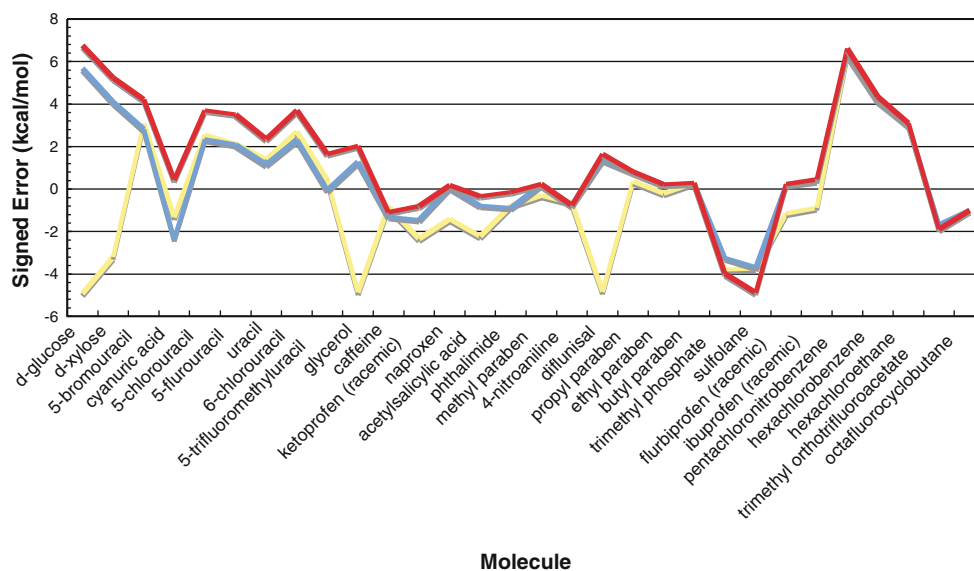


Fig. 4 Geometry for D-glucose generated by OMEGA

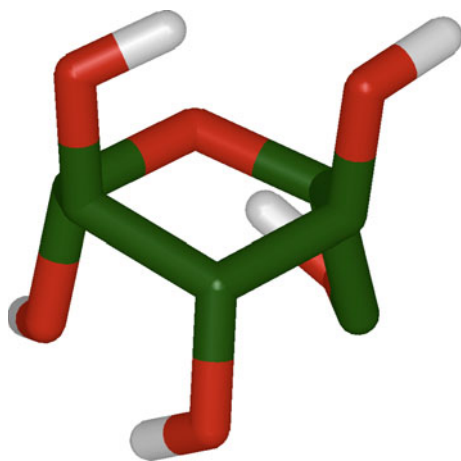


Fig. 5 Geometry for D-glucose optimized with M06-2X/6-31+G**

Zap TK submission benefited from a further MMFF94s optimization of this well, while similar optimization was not attempted for any of the SM8 submissions. Global optimization of MMFF94s and expensive DFT geometry optimizations were detrimental to performance of single

conformer solvation energy calculations with the SM8 and Zap TK methods.

The combination of CM4M charges with the ZAP9 radii produced unsurprisingly poor results; however, the CM4M charges with reparameterized radii outperformed AM1BCC charges with ZAP9 radii at the same geometries. Additionally, other DFT charge models, which have done well in previous SAMPL challenges, were not attempted at the same geometry. Therefore, further studies involving better geometries and direct comparison to other DFT charge models are required in order to properly judge the performance of CM4M charges with Zap TK.

For solvation energy calculations, as is often the case in the field of computational chemistry, increased computational cost quickly leads to diminishing returns. While high-level gas-phase geometry optimizations are shown to be detrimental, high-level charges do in fact yield improved results. However, these high-level charges are several orders of magnitude more expensive than AM1BCC charges. The best SM8 calculation with DFT charges had an RMSE of 1.89 kcal/mol and the best Zap TK calculation with AM1BCC charges had an RMSE of 2.38 kcal/mol.

References

- Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) The SAMPL2 blind prediction challenge: introduction and overview. doi:10.1007/s10822-010-9350-8
- Feynman RP (2005) The pleasure of finding things out: the best short works of Richard P. Feynman. Basic Books, New York
- Toolkits 1.7.0-3 (2009) OpenEye Scientific Software, Inc., Santa Fe
- Cramer CJ, Truhlar DG (2008) A universal approach to solvation modeling. *Acc Chem Res* 41:760–768

- Nicholls A, Wlodek S, Grant JA (2009) The SAMP1 solvation challenge: further lessons regarding the pitfalls of parametrization. *J Phys Chem B* 113:4521–4532
- Chamberlin AC, Cramer CJ, Truhlar DG (2008) Performance of SM8 on a test to predict small-molecule solvation free energies. *J Phys Chem B* 112:8651–8655
- Marenich AV, Cramer CJ, Truhlar DG (2009) Performance of SM6, SM8, and SMD on the SAMPL1 test set for the prediction of small-molecule solvation free energies. *J Phys Chem B* 113:4538–4543
- Zhao Y, Truhlar DG (2007) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* 120:215–241
- Halgren TA (1996) Merck molecular force field. I, II, III, IV, and V. *J Comput Chem* 17:490–641
- Halgren TA (1999) MMFF VI. MMFF94s option for energy minimization studies. *J Comput Chem* 20:720–729
- Ribeiro RF, Marenich AV, Cramer CJ, Truhlar DG (2010) Prediction of SAMPL2 aqueous solvation free energies and tautomeric ratios using the SM8, SM8AD, and SMD solvation models. doi:[10.1007/s10822-010-9333-9](https://doi.org/10.1007/s10822-010-9333-9)
- Omega 2.3.2 (2009) OpenEye Scientific Software, Inc., Santa Fe
- Szybki 1.3.2 (2009) OpenEye Scientific Software, Inc, Santa Fe
- QuacPac 1.3.1 (2008) OpenEye Scientific Software, Inc, Santa Fe
- Jakalian A, Bush BL, Jack DB, Bayly CI (2000) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J Comput Chem* 21:132–146
- Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 23:1623–1641
- Olson RM, Marenich AV, Cramer CJ, Truhlar DG (2007) Charge model 4 and intramolecular charge polarization. *J Chem Theory Comput* 3:2046–2054
- Frisch MJ, Pople JA, Binkley JS (1984) Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J Chem Phys* 80:3265
- Higashi M, Marenich A, Olson R, Chamberlin AC, Pu J, Kelly CP et al (2008) GAMESSPLUS version 2008-2, University of Minnesota, Minneapolis
- Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH et al (1993) General atomic and molecular electronic structure system. *J Comput Chem* 14:1347–1363
- Hehre WJ, Radom L, Schleyer PR, Pople JA (1986) *Ab initio molecular theory*. Wiley, New York
- Nicholls A, Wlodek S, Grant JA (2010) SAMPL2 and continuum modeling. doi:[10.1007/s10822-010-9334-8](https://doi.org/10.1007/s10822-010-9334-8)
- Mobley DL, Dill KA, Chodera JD (2008) Treating entropy and conformational changes in implicit solvent simulations of small molecules. *J Phys Chem B* 112:938–946