THE JOURNAL OF PHYSICAL CHEMISTRY A

 VOLUME 108, NUMBER 43, OCTOBER 28, 2004

LETTERS

Surface Generalized Born Method: A Simple, Fast, and Precise Implicit Solvent Model beyond the Coulomb Approximation

Alexey N. Romanov, Sergey N. Jabin, Yaroslav B. Martynov, Alexey V. Sulimov, Fedor V. Grigoriev, and Vladimir B. Sulimov*

Department of Quantum Chemistry, Algodign, LLC, Bolshaya Sadovaya ulitsa, 8a, Moscow 123378, Russia

Received: July 23, 2004; In Final Form: September 9, 2004

We present the surface generalized born (S-GB) method, based on expansion of energy in a surface integral series. The method can be parametrized for simultaneous reproduction of the reaction field energy for small molecules as well as macromolecules. For a set of 195 small molecules, our S-GB model gives a root mean square (rms) of 0.13 kcal/mol relative to the rigorous polarizable continuum model (PCM).

Introduction and Theory

The solvation/desolvation energy is a key component in the characterization of various biological processes, such as protein−ligand interactions and protein folding. Thus, it is not surprising that a broad diversity of solvation models, primarily focused on electrostatic solute charge screening by solvents, exists today. They range from rigorous solutions of proper electrostatic problems to fast empirical models.

One end of this spectrum is represented by the Poisson−Boltzmann (PB) method and the polarizable continuum model (PCM).[1] These approaches provide a rigorous solution to the electrostatic problem of solute charges in a dielectric cavity, formed by the bulk solvent around the solute. The solution can be obtained on a 3D grid or, alternatively, the formalism of surface charges can be exploited (PCM). Being rigorous, these methods are very slow, which precludes their utilization in molecular mechanics and molecular dynamics simulations of large biomolecules. The other end of the spectrum is represented by empirical methods.[2] The high speed of calculation allows their use in docking and MD/MC protein simulations, but the physical origin and accuracy of such methods are questionable.

An interesting compromise in speed and accuracy is the generalized born (GB) approximation.[3,4] This approach reduces the problem of determination of the polarization energy, G^{pol}, of a point charge system in a dielectric cavity to finding the "self-polarization energy", G_i^{self}, of each charge, which means the polarization energy of the charge in the same dielectric cavity as the whole system, but with other charges absent:

$$G^{pol} = -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\sum_{i,j}\frac{q_i q_j}{\left(r_{ij}^{\,2} + \alpha_i\alpha_j \exp\left(-\frac{r_{ij}^{\,2}}{c\alpha_i\alpha_j}\right)\right)^{1/2}} \quad (1)$$

Here, the Born radii of atoms, α_i, are inversely proportional to their self-polarization energy, G_i^{self}:

$$\alpha_i = -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\frac{q_i^{\,2}}{G_i^{self}} \quad (2)$$

and c is an empirical coefficient, which is equal to 4 in Still's original paper[3] or 8 in a recent parametrization;[5] q_i are atomic charges, r_{ij} is the distance between charges q_i and q_j, and ϵ_{in} and ϵ_{out} are the dielectric constants inside and outside the cavity, respectively.

Equation 1, although empirical, gives a surprisingly accurate approximation to the rigorous PB results. It was shown[6] that in

* Corresponding author. E-mail: vladimir.sulimov@algodign.com.

10.1021/jp046721s CCC: $27.50
Published on Web 10/02/2004

the case of the exact Born radii, α_i, defined by eq 2 where the G_i^{self} values were calculated by the rigorous PB method, the GB polarization energy obtained from eq 1 differed from the exact PB polarization energy by <1%.

However, the problem of finding the correct Born radii, α_i (i.e., self-polarization energies, G_i^{self}), still exists. Of course, they can be calculated by the PB method or the PCM, but this vitiates the point of the GB method and thus only has interest for the calibration of other, faster methods. A more straightforward way is to use the following approximate expression:

$$G_i^{\text{self}} = -\frac{1}{8\pi}\left(\frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}}\right)\int_V \bar{D}^2\, dV \tag{3}$$

where $\bar{D} = q\bar{r}/|r|^3$ is the electric displacement vector, approximated by the Coulomb field of the ith atom, and the integration is performed over the volume outside the dielectric cavity. Equation 3 is exact for spherical cavities. For other cavities, it is only an approximation, but it works quite well and is often referred to as the Coulomb approximation for α_i:

$$\frac{1}{\alpha_i} = \frac{1}{4\pi}\int_V \frac{1}{r^4}\, dV \tag{4}$$

From the computational point of view, it is reasonable to change the integration volume, V, in eq 4 and integrate over the volume, W, inside the dielectric cavity, excluding the volume in the sphere of radius R_i^{es} over the ith atom, which gives

$$\frac{1}{\alpha_i} = \frac{1}{R_i^{\text{es}}} - \frac{1}{4\pi}\int_W \frac{1}{r^4}\, dV \tag{5}$$

Here, the R_i^{es} radius is the electrostatic radius for the particular atom. The set of different R_i^{es} values for each atom type can be used to build the dielectric cavity in the PB and GB methods.

Hence, the Born radii can be calculated by eq 5 using numerical volume integration, as was done by Scarci et al.[7] Several attempts were made to substitute numerical volume integration by analytical techniques, involving summation over atoms, but the results usually were less satisfactory.

Ghosh et al.[8] used the Green formula

$$\int_V \nabla\bar{A}\, dV = \int_S (\bar{A}\,\bar{n})\, dS \tag{6}$$

to reformulate eq 4 in terms of surface integration instead of volume integration:

$$\frac{1}{\alpha_i} = \frac{1}{4\pi}\int_S \frac{(\bar{r}\,\bar{n})}{r^4}\, dS \tag{7}$$

The advantage of this surface GB method (S-GB) is that surface numerical integration is usually faster than volume integration; also, the surface integration scales better with increasing system size, which is important in the case of large macromolecules.

Several attempts were made to go beyond the Coulomb approximation. Ghosh et al.[8] used an empirical correction in their S-GB method to enhance the correspondence between the Coulomb approximation (eq 4) and PB-calculated radii, α_i. The most successful approach was introduced by Lee et al.[5] They used what they call a "correction term" to improve the Coulomb approximation result (eq 4):

$$\frac{1}{\alpha_i - \alpha_0} = -\frac{1}{4\pi}\int_V \frac{1}{r^4}\, dV + P\left(\frac{1}{4\pi}\int_V \frac{1}{r^5}\, dV\right)^{1/2} \tag{8}$$

Here, the integration volume is outside the cavity, α_0 is a small offset which slightly enhances the accuracy, and the constant $P \approx 2(2)^{1/2}$ is needed to obtain the exact result in the case of a spherical cavity. In fact, the term

$$P\left(\frac{1}{4\pi}\int_V \frac{1}{r^5}\, dV\right)^{1/2}$$

is not actually a correction, because its value is comparable with the unmodified value. The Coulomb term

$$\frac{1}{4\pi}\int_V \frac{1}{r^4}\, dV$$

even has the opposite sign compared to the unmodified eq 4. Thus, eq 8 has no clear origin as a Coulomb approximation correction. Nonetheless, this approach gives the best correspondence of calculated α_i with the exact PB values (correlation coefficient, 0.9996), but the reason for such high accuracy is not clear. The original method of Lee et al.[5] was formulated and implemented in the CHARMM package (version c30b2, GBMV methods) using a numerical grid volume integration and several analytical approaches to define the standard molecular volume.

On the other hand, as pointed out above, the surface integration technique possesses some advantages, and it would be interesting to formulate and test a method with equivalent or higher accuracy compared to the volume method of eq 8, but based on surface integration, which we do here.

Formulation and Implementation of S-GB Method

We propose the Surface GB method, based on the calculation of surface integrals of type

$$I_n = \left(\int_S \frac{(\bar{r}\,\bar{n})}{r^n}\, dS\right)^{1/(n-3)}, \quad n \geq 4 \tag{9}$$

These integrals have the proper dimension (inverse length), so we can attempt to expand the self-polarization energy of unit charge as a sum of such integrals:

$$G_i^{\text{self}} = -\left(\frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}}\right)\left(A_0 + \sum_{n=4}^{\infty} A_n I_n\right) \tag{10}$$

Then, the G_i^{self} values are used to calculate α_i by eq 2. Certainly, the summation in eq 10 should be stopped after a reasonable number of terms. The A_n coefficients can be determined by fitting the results from the S-GB calculations of eq 10 to the exact PCM solution on the same surface.

We utilize our own procedure to build the solvent excluded surface (SES)[9] used in S-GB and PCM calculations. This procedure employs secondary rolling of the surface[10] and guarantees smoothing of the surface, which is necessary to calculate the PCM energy for molecules of an arbitrary size (including proteins). It is known that the SES provides the closest fit of results obtained by continuum solvation methods to experimental results.[5]

We also use our own iteration procedure to find the solution of PCM equations. This procedure is similar to one described by Chudinov et al.[11] and is applied to surface elements (tesserae) constructed on the basis of a triangular grid built on the SES. It was shown that the PCM solution error depends linearly on the triangle size (the length of the longest triangle side), which is an adjustable parameter. We choose the triangle size to be

TABLE 1: Correlation of G_i^{self}(PCM) with Individual Integrals, I_n, and with Energy from eq 10[a]

n	correlation of I_n with G_i^{self}(PCM)	correlation of G_i^{self}(S-GB) with G_i^{self}(PCM)	maximal deviation of G_i^{self}(S-GB) from G_i^{self}(PCM), kcal/mol	rms deviation of G_i^{self}(S-GB) from G_i^{self}(PCM), kcal/mol
4	−0.981 83	0.981 83	11.65	2.27
5	−0.999 19	0.999 60	2.10	0.34
6	−0.997 26	0.999 63	2.42	0.32
7	−0.991 16	0.999 70	1.75	0.29
8	−0.984 92	0.999 70	1.75	0.29
9	−0.979 36	0.999 70	1.75	0.29

[a] The G_i^{self}(S-GB) values are calculated using the A_n coefficients optimized for the first 600 atoms in the PDB file of HIV integrase and using the R_i^{es} radii of Bordner et al.[12]

TABLE 2: Parameters and Performance of Several Models for G_i^{self} Calculation Based on Linear Combinations of Various I_n Integrals[a]

						rms for G_i^{self}, kcal/mol, calculated using the R_i^{es} radii from:	
model no.	A_4	A_5	A_6	A_7	A_0	Bondi[13]	Bordner et al.[12]
1	−0.739	13.526	−19.992	13.113	1.166	0.59	0.32
2	1.881	*	*	*	−0.131	4.35	2.71
3	−0.449	6.326	*	*	1.921	0.69	0.39
4	0.437	*	5.315	*	0.483	0.75	0.43
5	0.732	*	*	4.911	−0.411	0.96	0.56
6	*	5.119	*	*	1.488	1.13	0.62
7	*	3.118	2.697	*	1.185	0.65	0.37
8	*	3.904	*	1.895	−0.136	0.64	0.37
9	*	*	13.064	−7.137	1.825	0.74	0.42

[a] An asterisk indicates that the corresponding integral was not used in the model. The units of A_0 are kcal/mol, and the units of A_4-A_7 are (kcal·nm)/mol. The rms was calculated over the first 400 atoms in the PDB file of HIV integrase (Bondi set of R_i^{es}) or over all 2374 atoms of the HIV integrase catalytic domain (Bordner et al. set of R_i^{es}).

0.02 nm. Arbitrary shift of the 0.02 nm triangular grid along the SES results in solution variations within 0.3 kcal/mol for large molecules (∼0.02% of the total molecule reaction field energy). We believe that this accuracy is sufficient for most PCM applications. In the remainder of this paper, all PCM and GB solutions are made on the SES grid with a triangle size 0.02 nm. Other parameters, used in building the SES, are either Bordner et al.[12] or Bondi's[13] set of atomic radii, R_i^{es}, and values of 0.14 and 0.04 nm for the first and second rolling radii,[9] respectively.

In Table 1, we present the correlation of individual integrals, I_n, as well as the values of S-GB G_i^{self} calculated from eq 10 relative to the PCM-calculated self-polarization energies for the first 600 atoms in the PDB file (1QS4 code in the Protein Data Bank)[14] of a typical protein, HIV integrase. The A_n coefficients were determined by a least-squares procedure to minimize the difference between the PCM and S-GB results over the 600 integrase atoms using Bordner et al.[12] radii, R_i^{es}.

We choose to take the first four integrals in eq 10, because more integrals do not appear to increase the accuracy (see Table 1). We call this the four-integral model. The values of the coefficients A_0 and A_4-A_7 calculated as described above are shown in the first data line of Table 2.

The coincidence of the exact self-polarization energy with the four-integral model is shown in Figure 1. For comparison, in Figure 2, we present the results for only I_4 integral fitting, which is equivalent to the Coulomb approximation. In these calculations, we use the R_i^{es} radii from Bondi.[13] It is clear that the four-integral model gives much better results and also eliminates the nonlinear behavior seen with the Coulomb approximation. This nonlinear behavior is a characteristic feature of the Coulomb approximation, manifested in all molecular systems, because the Coulomb approximation cannot simultaneously fit the cases of buried atoms with low G_i^{self} values and atoms exposed to solvent with high G_i^{self} values.

It is also interesting to test our S-GB model for a particular force field and R_i^{es} radii parametrization. We chose the recent

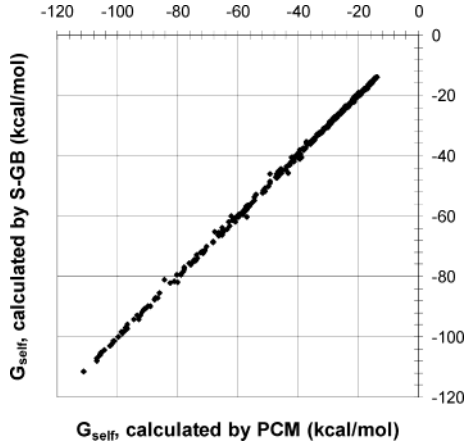

Figure 1. Plot of the rigorous atomic self-polarization energies (calculated by the PCM) vs the four-integral S-GB values for the first 400 atoms in the PDB file of HIV integrase. Partial atomic charges are 1 au, and the R_i^{es} radii are taken from Bondi.[13] The dielectric coefficient is $\epsilon_{in} = 1$ inside the cavity and is $\epsilon_{out} = 78.39$ outside. The coefficients A_0 and A_4-A_7 used to calculate the S-GB values by eq 10 are listed in the first data line of Table 2. The correlation coefficient is 0.999 736. The maximal deviation is 3.42 kcal/mol, and the rms is 0.59 kcal/mol.

parametrization of Bordner et al.[12] for the MMFF force field.[15] This parametrization is a good case, because it is derived by a variant of the PCM and formulated for the popular MMFF force field, which is applicable to a wide variety of organic molecules. Unlike the previous examples, the Bordner et al. parametrization utilizes an ϵ_{in} value that is different from unity. The results of the S-GB four-integral model for the self-polarization energy of HIV integrase atoms are shown in Figure 3. The correlation coefficient is practically the same as that in Figure 1, while the root mean square (rms) and maximal deviation are approximately 2-fold less, which is expected because in this case the S-GB polarization energy corresponds not to transferring the molecule to vacuum but to media with $\epsilon_{in} = 2.21$. Importantly,

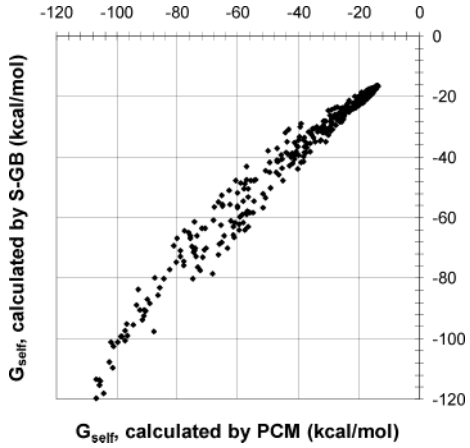

Figure 2. Plot of the rigorous atomic self-polarization energies (calculated by the PCM) vs the one-integral S-GB values (Coulomb approximation) for the first 400 atoms in the PDB file of HIV integrase. Partial atomic charges are 1 au, and the R_i^{es} radii are taken from Bondi.[13] The dielectric coefficient is $\epsilon_{in} = 1$ inside the cavity and is $\epsilon_{out} = 78.39$ outside. The coefficients A_0 and A_4 in eq 10 are listed in the second data line of Table 2. The correlation coefficient is 0.989 984. The maximal deviation is 14.10 kcal/mol, and the rms is 4.352 kcal/mol.

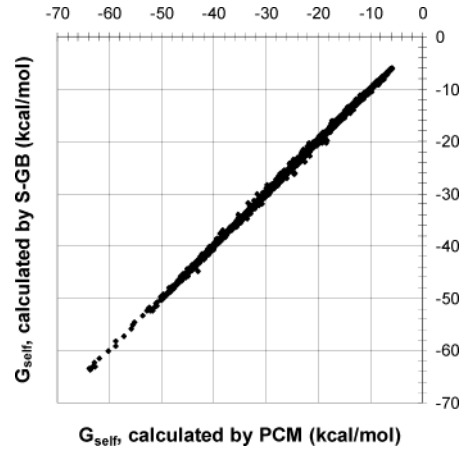

Figure 3. Plot of the rigorous atomic self-polarization energies (calculated by the PCM) vs the four-integral S-GB values for the atoms of the HIV integrase catalytic domain (2374 atoms). Partial atomic charges are 1 au, and the R_i^{es} radii are those of Bordner et al.[12] The dielectric coefficient is $\epsilon_{in} = 2.21$ inside the cavity and is $\epsilon_{out} = 80.0$ outside. The coefficients A_0 and A_4-A_7 are the same as those for Figure 1. The correlation coefficient is 0.999 707. The maximal deviation is 1.90 kcal/mol, and the rms is 0.319 kcal/mol.

the coefficients A_0 and A_4-A_7 used for this calculation were the same as those previously optimized for the SES built using the Bondi R_i^{es} radii. Hence, these results suggest that the four-integral model has universal properties, and its performance does not depend on any particular R_i^{es} radii parametrization, the media dielectric constant, or the method of building the SES.

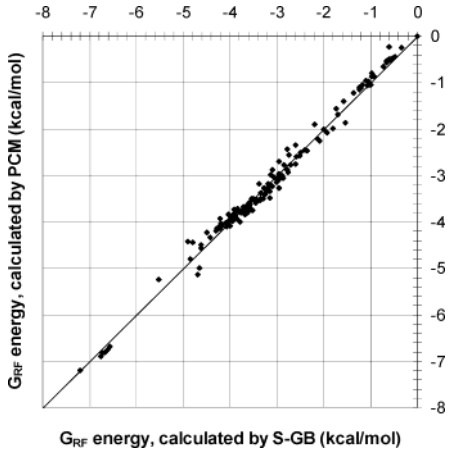

Figure 4. Deviations of S-GB energies from rigorous PCM reaction field energies for a set of 195 small molecules. The R_i^{es} radii are taken from Bordner et al.[12] The dielectric coefficient is $\epsilon_{in} = 2.21$ inside the cavity and is $\epsilon_{out} = 80.0$ outside. The coefficients A_0 and A_4-A_7 are the same as those for Figure 1. The empirical coefficient, c, in eq 1 is 8. The correlation coefficient is 0.997, and the rms is 0.13 kcal/mol. The grid-based volume integration method of Lee et al.[5] gives a correlation coefficient of 0.983 and a rms of 0.79 kcal/mol on the same set.

We summarize the performance of our model and several other models in Table 2. Here, the first model is our four-integral model; the second model is the Coulomb approximation (I_4 integral only); model 3 utilizes the I_5 integral in addition to I_4 and is thus the surface correlate of the volume method of Lee et al.[5] mentioned above; and model 5 corresponds to another variant presented by Lee et al.[5] It is clear in model 3 that the I_5 integral is not just a correction to the Coulomb integral, I_4, because the A_5 coefficient is much larger than the A_4 coefficient. Several other models based on linear combinations of just two integrals are also shown in the table. Interestingly, the most accurate of all the two-integral models, model 8, does not include the Coulomb integral at all.

It is important to go further and compare the whole reaction field energy calculated by our four-integral S-GB model and the more rigorous PCM. This comparison for a set of 195 small molecules[12] is shown in Figure 4. The difference between the S-GB model and the PCM (rms of 0.13 kcal/mol) is much smaller than the difference of the PCM from experimental results (rms of 0.716 kcal/mol[12]). The S-GB method is also about 5-fold more accurate than the volume method of Lee et al.[5] on this set (rms of 0.79 kcal/mol relative to the PCM). Thus, our S-GB method gives excellent results for small molecules, despite the fact that it is parametrized on a macromolecule (HIV integrase protein). It therefore seems that eq 10 is universal and the performance of the four-integral S-GB model does not depend on cavity shape.

Along these lines, Table 3 presents the reaction field energies and their differences for two systems of a protein, ligand, and complex, again using the values A_0 and A_4-A_7 derived from

TABLE 3: Solvent Reaction Field Energy Calculated with Various Methods for Protein−Ligand Binding

complex	energy, calculated for:	PCM energy, kcal/mol	S-GB energy, kcal/mol	Still analytical GB energy, kcal/mol
benzamidine−trypsin	complex G_c	−2165.58	−2098.04	−1584.51
	ligand G_l	−69.98	−69.92	−69.44
	protein G_p	−2105.66	−2037.12	−1529.21
	difference: $G_c - (G_l + G_p)$	10.06	9.00	14.4
4-aminobenzamidine−trypsin	complex G_c	−2180.65	−2112.40	−1575.57
	ligand G_l	−84.90	−84.50	−71.29
	protein G_p	−2103.63	−2033.57	−1529.11
	difference: $G_c - (G_l + G_p)$	7.88	5.67	24.83

HIV integrase. These energies play an important role in studying protein−ligand interactions, and a maximal coincidence with the rigorous PCM results is highly desirable. It is clear that our S-GB method performs much better than the traditional analytical approach.

Conclusions

The fast and precise surface generalized born four-integral method has been presented. Although its theoretical justification is not completely clear and is currently under investigation, this method performs very well in a variety of systems. Its accuracy should allow the S-GB method to be used instead of the very slow PCM and PB method in the modeling of various biochemical processes where the solvation energy plays an important role.

Acknowledgment. We are grateful to E. Nikitina (Institute of Applied Mechanics, Russian Academy of Sciences) for calculations using the CHARMM package (version c30b3) and C. Queen for careful review of this manuscript.

References and Notes

(1) Tomasi, J.; Mennucci, B.; Cammi, R. Molecules in solution. In *Handbook of Molecular Physics and Quantum Chemistry*; Wilson, S., Ed.; Wiley: Chichester, U.K., 2003; p 299.
(2) Stouten, P. F. W.; Froemmel, C.; Nakamura, H.; Sander, C. *Mol. Simul.* **1993**, *10*, 97.
(3) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(4) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(5) Lee, M. S.; Salsbury, F. R.; Brooks, C. L., III. *J. Chem. Phys.* **2002**, *116*, 10606.
(6) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297.
(7) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. A* **1997**, *101*, 8098.
(8) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983.
(9) Jabin, S.; Martynov Y. *Algo-Surf: the program to construct the solvent excluded surface*; Algodign, LLC: Moscow, 2003.
(10) Vorobjev, Y. N.; Hemans, J. *Biophys. J.* **1997**, *73*, 722.
(11) Chudinov, G. E.; Napolov, D. V.; Basilevsky, M. V. *Chem. Phys.* **1992**, *160*, 41.
(12) Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A. *J. Phys. Chem. B* **2002**, *106*, 11009.
(13) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
(14) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13040.
(15) Halgren, T. A. *J. Comput. Chem.* **1997**, *17*, 490.