ELSEVIER

Biophysical Chemistry 78 (1999) 183–193

Biophysical Chemistry

Nucleic acid base-pairing and *N*-methylacetamide self-association in chloroform: affinity and conformation

Ray Luo[a,1], Martha S. Head[a,b,1], James A. Given[a], Michael K. Gilson[a,b,*]

[a] *Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, MD 20850, USA*
[b] *National Institute of Standards and Technology, Gaithesburg, MD 20899, USA*

Received 8 October 1998; received in revised form 23 November 1998; accepted 23 November 1998

Abstract

A recently developed computational method, 'mining minima', is used to examine the hydrogen-bonding interactions of nucleic acid base-pairs and of the *N*-methylacetamide homodimer in chloroform. The mining minima algorithm aggressively samples molecular conformations, identifies the most important local minima, and computes their contributions to the overall free energy of the system. Here, the CHARMM 98 parameter set is used for the potential energy and the generalized Born/surface area solvent model is used to account for the influence of the solvent. Good agreement with experiment is obtained for the non-covalent binding affinities of a series of complexes. The computational approach used here is applicable to a range of molecular systems.

1. Introduction

Current methods for computing binding affinities tend to lie at the extremes of a spectrum of complexity. They are either detailed but time-consuming perturbative free energy simulations, or fast but highly simplified energy-component models [1,2]. Between these two extremes lies a broad middle ground that accommodates intermediate models of binding that could capture significant physical detail but that would also be fast. We have recently proposed a new class of 'predominant-states' methods for computing binding affinities that would occupy this middle ground [2]. These predominant-states methods calculate differences in standard chemical potential via sums over the predominant conformations of the reactants and products. Such calculations are made possible by two technical advances.

One advance is the development of efficient algorithms that use the predominant states concept [3–5] to calculate configuration integrals — chemical potentials, in effect — for systems of modest size. One of these algorithms, the mining

* Corresponding author. Tel.: +1 301 738 6217; fax: +1 301 738 6255; e-mail: gilson@indigo14.carb.nist.gov

[1] R. L. and M. S. H. contributed equally to this paper.

[2] Present address: Smithkline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, PA 19406-0939, USA.

0301-4622/99/$ - see front matter
PII: S0301-4622(98)00229-4

minima (MM) method [6], locates stable conformations of a molecular system and computes the free energy of each conformation with a Monte Carlo integration method. The results can be used to compare the stabilities of the various conformations [6]. The free energies of the various conformations also can be combined to yield the overall free energy of the system. Such calculations can be used to compute binding affinities [7], because the difference between the free energy of a molecular complex and the separated molecules is directly related to their binding affinity. Another promising algorithm that also evaluates configuration integrals through analysis of low-energy conformations has recently been described [8].

The second technical advance is the development of fast yet accurate implicit models of the solvent. Implicit models of the solvent allow the influence of solvent upon conformational energetics to be estimated without the use of time-consuming explicit-solvent simulations [2]. Of particular interest here is a class of implicit solvation models [9] in which the solvation energy is separated into a non-polar part and a polar part. The non-polar part is typically treated as linearly related to molecular surface area [10,11], and the electrostatic part is computed by a continuum model [12]. Computationally rapid approximations to the electrostatic solvation energy [13–18] now make it possible to use this class of solvation models in calculations that examine large numbers of molecular conformations [7,8,19–21].

Fast, implicit solvation models combined with the predominant states algorithms described above form a promising approach to the computation of molecular properties. Thus, calculations using the generalized Born (GB) electrostatics model [13,18] with the MM algorithm have yielded good agreement with experiment for the pKa shifts of small molecules in solution [21] and for the strength of solvent-exposed ion-pairs in water [7]. The related method of Kolossvary et al. has yielded good results for the relative binding affinities of similar ligands for a small-molecule receptor [20].

In the present study, the mining minima/generalized Born (MM/GB) method is applied to the non-covalent association of a series of hydrogen-bonding molecules in chloroform. The molecules studied are two purine and two pyrimidine nucleic acid bases, and *N*-methylacetamide. The interactions studied here are of fundamental importance in biochemistry as the hydrogen bond plays a key role in the structure and binding of biomolecules. Similar systems have been examined previously by computational methods that are more detailed than the present method in important respects; a recent publication reviews many of these contributions [22]. One approach has been to model hydrogen-bonded complexes in solution with empirical force-fields and an explicit representation of the solvent [23–25]. The explicit treatment of solvent is expected to be relatively accurate, but it is also time-consuming. It is probably for this reason that such studies have rarely attempted to account for all potentially stable conformations of the complex. Another approach has been to enhance the model of the hydrogen bond by replacing the empirical force field with ab initio electronic structure calculations, and then to reduce the computational burden by using an implicit model of the solvent [22]. This approach also is time-consuming, and the report focuses on a single optimized conformation of the complex [22]. A third approach combines an empirical force field with an implicit model of the solvent [26,27]. This approach increases the computational speed, but the use of detailed finite-difference solutions of the Poisson equation still restricts the analysis to about 10 different conformations of the complex.

The present approach uses an empirical force field; a fast, implicit solvation model; and a conformational sampling method that drives rapidly to energetically important low-energy conformations. The computational efficiency of this approach allows it to account for all energetically important conformations of the bound complexes. Each binding reaction examined here requires a matter of hours on a workstation. Another advantage of the present method is that it is clearly linked to the underlying theory of non-covalent binding [2]. It yields standard free energies of binding, referenced to the usual 1 mol/l standard state. As a consequence, the calculations can be compared directly with measured free energies of

binding, unlike most previous calculations on these systems. It is worth noting that the method automatically accounts for the loss in orientational and translational entropy [2], as well as for the positive entropy associated with intermolecular vibrations of the complex [28].

Comparisons with experiment are important for any model of binding, because any computationally tractable model must involve approximations. Comparison with experiment is particularly important for the present method because it is new and because it relies upon the GB model, which is still relatively untested. We therefore study seven binding reactions for which experimental data are available.

2. Methods

2.1. Free energy calculations

2.1.1. Theory

For the non-covalent association of two molecules, A + B $\rightleftarrows$ A:B, the change in Helmholtz free energy at standard concentration is [2]:

$$\Delta A^\circ = -RT\ln\frac{Z_{A:B}}{Z_A Z_B} + RT\ln\frac{8\pi^2\sigma_{A:B}}{C^\circ\sigma_A\sigma_B} \quad (1)$$

where R is the gas constant, T is the temperature, C° is the standard concentration, $\sigma_{A:B}$, σ_A, and σ_B are symmetry numbers [29], and $Z_{A:B}$, Z_A, and Z_B are the configuration integrals of the complex and of the isolated monomers, respectively.

The configuration integrals of the isolated molecules, Z_A and Z_B, can be written in the form [2]:

$$Z = \int e^{-(U+W)/RT}\mathrm{d}\mathbf{r}. \quad (2)$$

Here the Boltzmann factor is given in terms of the vacuum potential energy U and the solvation energy W, both of which depend upon molecular conformation. The integral extends over the full range of the internal coordinates $\mathbf{r}$.

It is assumed here that the probability distribution of bond lengths and bond angles is not significantly affected by complexation. Therefore the integration only extends over free torsion angles for groups larger than a methyl group. In the present study, this means that the molecules are treated as rigid because the only rotatable bonds are for the methyls. Previous computational studies of the nucleic acid bases have reported good results when using this approximation [25,30]. For the complex, the configuration integral extends over the relative position $\mathbf{R}_B$ and orientation ξ_B of molecule B relative to molecule A, for a total of 6 degrees of freedom. The configuration integral for the complex also includes an index function $I(\mathbf{R}_B)$ that defines the complexed state [2,23,31]:

$$Z_{A:B} = \int I(\mathbf{R}_B)e^{-(U+W)/RT}\mathrm{d}\mathbf{R}_B\mathrm{d}\xi_B. \quad (3)$$

The region for which $I(\mathbf{R}_B) = 1$ must include all the relevant low-energy conformations of the complex but no more than a small volume of conformational space in which the two molecules are not effectively bound [2]. Here, an adaptive procedure is used to determine this region, as described below.

The configuration integrals are evaluated with a modified version of the previously described MM method [6]. This method uses the predominant states approximation [32] that the free energy,

$$A^\circ = -RT\ln Z \quad (4)$$

is dominated by a modest number of low energy states. The free energy is therefore computed from the contributions A_j of a finite number M of energy wells j:

$$A^\circ \approx -RT\ln\left(\sum_{j=1}^{M} e^{-A_j/RT}\right) \quad (5)$$

2.1.2. Algorithm

The MM method may be summarized as follows (see Fig. 1). The first step in this iterative algorithm is to find an energy-minimum. (The

energy-minimizer used here has been described previously [6].) The search for energy minima extends over the full range of all degrees of freedom in the integrals, except that the position of molecule B relative to A is restricted to a cubic box centered on A and having a side of length 30 Å. The second step in the algorithm is to evaluate the configuration integral within the energy well around the energy-minimum. This is done with a Monte Carlo sampling method, as previously described [33]. The Monte Carlo sampling region is gradually extended outward from the deepest point of the well until the entire relevant low-energy region has been included in the integral; the stopping criteria are described below. In the third step, the configuration for this energy well is added to the running sum, as shown in Eq. (5). Then the running Boltzmann-averaged system energy over all energy wells ($\langle U+W\rangle$) is checked for convergence; it is considered to have converged if the contributions of five successive energy wells change $\langle U+W\rangle$by less than 0.001 RT. If the average energy has not converged to within this tolerance, then the region that has just been sampled is marked to prevent it from being visited again and a new cycle is begun with the search for a new energy well. However, if the average energy has converged, then all relevant conformations have been accounted for, so the calculation is complete.

Sampling within an energy well proceeds as follows. For each energy well, an initial integral is carried out for a small hypercube of conformational space around the energy minimum. The integration region is then extended outward incrementally in one direction, and an additional integral is done for the resulting strip of configuration space on one side of the central hypercube. The integral over this strip is added to the running sum for the energy well. The size of the sampling region is thus incremented along one dimension at a time to form a growing hyperrectangle in conformational space. Torsional and orientational angles are incremented in steps of 10°, and the range of positions of molecule B relative to A (translation) is incremented in 1 Å steps. The stopping criteria for this process are as follows. During sampling of energy well j, the

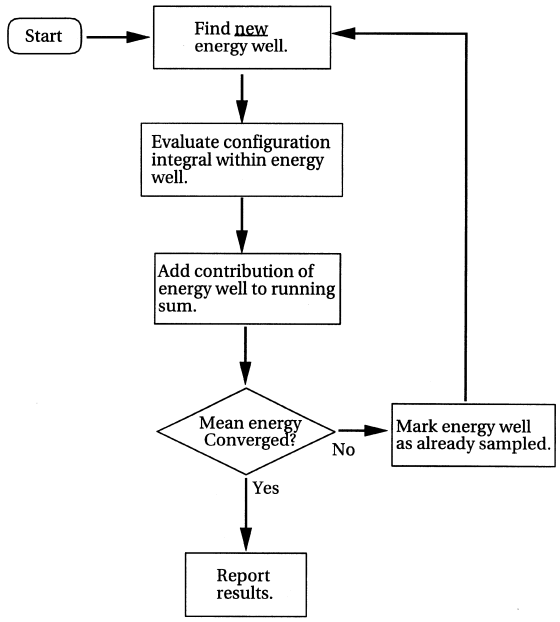

Fig. 1. Flow chart of the Mining Minima algorithm used in this study.

Boltzmann-weighted average $\langle U+W\rangle_j$ is accumulated. When an increment in the size of the well along a particular dimension produces a change in $\langle U+W\rangle_j$ of less than 0.001 RT, the well is no longer extended along that dimension. Extension along a particular dimension is also stopped if the quantity $U+W$ begins to fall along that axis, indicating entry into a new energy well. Sampling of a given energy well is completed when extension of the sampling region is closed off in all dimensions by application of one of these two stopping criteria.

Basing the convergence criterion upon the convergence of $\langle U+W\rangle$ in effect sets the index function in $Z_{A:B}$ to 1 for conformations that contribute significantly to this average and 0 for all other conformations. This automatically causes the integral for the bound complex to encompass the low-energy part of conformational space, but very little conformational space in which the two molecules are not stably bound. Thus, the domain over which $I(\mathbf{R}_B)=1$ depends, appropriately, upon the range of the attractive forces between the two molecules [2].

Our initial implementation of the MM method

sampled only over dihedral angles [6]. However, the present implementation also samples isotropically over the position and orientation of one molecule with respect to the other, as required to compute the free energy of the complex (Eq. (3)) [2]. The use of isotropic sampling makes including an explicit Jacobian determinant unnecessary. In the present study, numerical integration is unnecessary for the isolated molecules, because they are treated as rigid. This means that their configuration integrals simply equal $e^{-(U+W)/RT}$, where U and W are computed for the single rigid conformation.

A complication results from the fact that the experimental data are based upon shifts in infrared (IR) absorptions due to the formation of intermolecular hydrogen bonds [34]. This method is expected to detect hydrogen-bonded conformations of the complexes, but not stacked conformations. The MM/GB method allows us to separate the contributions of hydrogen-bonded and stacked conformations to the chemical potential of the complexes. The computed binding energies tabulated in the Results section include only the contributions from hydrogen-bonded conformations. However, the results with stacked conformations are also analyzed and discussed.

A complete binding calculation for the molecules considered here requires 4–6 h of processor time on an SGI R10000 workstation. All calculations use a locally modified version of the program UHBD [35].

2.1.3. Energy model

The energy in the configuration integrals is separated into a potential energy, $U(\mathbf{r})$ and a solvation energy, $W(\mathbf{r})$ [2]. Here, the potential energy is computed with the all-hydrogen CHARMM parameter set as of February 1998 (A. D. MacKerell, personal communication) [36,37] where available. The molecular dielectric constant is set to 1 because this is the value for which the atomic parameters are optimized.

The solvation energy is computed with the generalized Born/Surface area model [13,18], which is essentially an approximation to the Poisson-Boltzmann(PB)/γ model [10]. Both models compute the solvation energy of a molecule in a given conformation as the sum of a non-polar work ΔG_{np} and an electrostatic work ΔG_{el}:

$$\Delta G_{solv} = \Delta G_{np} + \Delta G_{el} \quad (6)$$

The non-polar work is the work of forming a non-polar cavity in the solvent having the shape of the solute and having van der Waals interactions with the solvent. It is taken to be linearly dependent upon the solvent-accessible surface area, as discussed below. The electrostatic solvation energy is the work of transferring the atomic charges into this cavity, and is taken to equal the $\frac{1}{2}\Sigma_i^N q_i \phi_{i,rf}$, where N is the number of atoms, q_i is the partial charge of atom i, and $\phi_{i,rf}$ is the reaction field at atom i due to the solvent. The solvent dielectric constant is set to that of chloroform, 4.8. The cavity radius of each atom is set to the mean of the solvent probe radius and the atom's σ parameter from the Lennard-Jones part of the CHARMM force field [21]. The radius of the chloroform solvent probe is taken to be 2.5 Å.

The original published parameters of the GB model were ajdusted to optimize agreement with detailed finite difference (FD) solutions of the Poisson equation for a varied training set of molecules [18]. Here, the GB model is adapted to further optimize the agreement with FD for the base pairs, by the following method. For each base pair, the most stable hydrogen-bonded and stacked conformations identified with the MM/GB method are used as representatives of the complex. The electrostatic solvation energies of these conformations, of these complexes and also of the separated bases are calculated with both GB and FD and the change in electrostatic solvation energy upon assembly of the complex is computed. If GB agreed perfectly with FD, both would give the same change in the electrostatic solvation energy. In fact, the deviations of GB relative to FD range from -3.59 to 3.34 kJ/mol, depending upon the base pair. In order to correct for the difference between GB and FD, these deviations were subtracted from the binding affinities computed with GB and the Mining Minima method. Correcting GB in this way improves the

agreement of the computed binding affinities with experiment.

As noted above, the work of forming a non-polar cavity, ΔG_{np} is estimated with a term that is linearly dependent upon the solvent accessible surface area; thus,

$$\Delta G_{np} = aA_{SA} + b \tag{7}$$

where A_{SA} is the solvent accessible surface area. Previous studies have shown that good agreement can be obtained with this approximation; see, for example, [10,38]. Values of the slope and offset in the PB/γ and GB/SA models are usually established empirically by relating the measured solvation energies of linear alkanes to computed surface areas for the solvent of interest [10,38,39]. However, we were unable to find measured solvation energies of alkanes in chloroform for linear alkanes other than octane. We therefore rely upon solvation data for several series of monofunctional molecules with alkane chains of varying length.

Four series of molecules are examined: aliphatic alcohols, carboxylic acids, amines, and esters; their measured solvation energies in chloroform are compiled elsewhere [39]. For each molecule, atomic partial charges and Lennard–Jones parameters were generated by Quanta 4.1 [40] with the charge-templates method and default charge-smoothing. The cavity radii are set as described above. The electrostatic part of the solvation energy of each molecule, ΔG_{el} is then computed by FD solutions of the Poisson equation with a molecular dielectric constant of 1 and a solvent dielectric constant of 4.8. The non-polar part of the solvation energy is then computed for each molecule as $\Delta G_{solv} - \Delta G_{el}$, and solvent-accessible surface areas are also computed. Linear least-square fitting is used for each series to obtain the slope and intercept in Eq. (7). We find that the slopes for the four series of molecules agree to within 1.5%. The mean slope, -0.0686 kJ/mol $-$ Å^2, is used here to compute the affinities of the base pairs. Note that the negative sign of a causes the non-polar solvation term to favor the exposure of molecular surface area to chloroform and thus opposes formation of a complex. This is in contrast to binding when water is the solvent.

The fitted values of the offset, b, for the four series of molecules deviate significantly from 0, varying from 15.89 to 26.75 kJ/mol. The nonzero value of b should, at least in principle, influence the computed binding affinities [38]. Thus, the change in non-polar solvation energy when molecules A and B bind will be estimated, from Eq. (7), as

$$\Delta\Delta G_{np} = a\Delta A_{SA} - b \tag{8}$$

where ΔA_{SA} is the change in surface area upon binding. This shows that the positive offsets found for chloroform are predicted to contribute to the binding free energy of A and B. The offset is therefore an important parameter in the model. However, the large range found for b makes it difficult to settle upon a single value for general use. The value of 25.1 kJ/mol yields good agreement with the binding energies considered here, and is also within the range arrived at empirically, as just described. This value is therefore used to arrive at the standard free energies of binding reported in the Results section. However, it is not clear that this value will be generally applicable and we regard the problem of the offset as one that merits further attention in future work. Fortunately, the need to select a value of the offset may be avoided by focusing upon the relative binding free energies of various complexes. The Results section presents both standard free energies of binding that include the 25.1 kJ/mol offset, and relative binding free energies that do not depend upon the value of this unsettled parameter.

It is worth noting that, although a and b depend upon the radii assigned to atoms of the solvent and to the radius of the solvent probe, the solvation energies computed with Eq. (7) are insensitive to these radii if they are used consistently in both the fitting process and in the calculation of solvation energy.

Computing the surface area, A_{SA}, is somewhat time-consuming. Here, we speed the calculations by using the fact that this quantity varies only weakly with conformation for either hydrogen-

bonded or stacked conformations. The changes in surface area upon binding are pre-computed for the low-energy hydrogen-bonded and stacked conformations that were also used to adjust the GB calculations (see above). These calculations yield two values of $\Delta\Delta G_{np}$ for each base-pair: one that is applicable to hydrogen-bonded conformations of the complex, and one that is applicable to stacked conformations. These values are used to adjust the computed free energy of each conformation generated by the mining minima method. This avoids recomputing the surface areas hundreds of thousands of times during the mining minima calculations. The non-polar solvation term favors the hydrogen-bonded conformations over stacked conformations by approximately 6 kJ/mol.

2.2. Molecular models and force-field

Initial all-atom coordinates are generated for the nucleic acid bases and *N*-methylacetamide (NMA) with Quanta 4.1 [40]. Each individual molecule is then energy-minimized with the full CHARMM 98 vacuum energy function, by means of the Newton–Raphson method [33] in version 26 of program CHARMM [36]. The minimizations are terminated when the energy gradient changes by less than 4.0×10^{-5} kJ/mol Å per step. Binding free energies are computed with the molecules fixed in these optimized conformations.

The MM method was used to find the most stable conformations of the nucleic acid base pairs in vacuo as a preliminary study of the new nucleic acid force-field parameters. The interaction energies ($U + W$) of the most stable hydrogen-bonded and stacked conformations of each base-pair are listed in Table 1. These agree well with results previously obtained with this force-field (A.D. MacKerell, personal communication). Note that a hydrogen-bonded conformation is the most stable in each case, except for AA, for which a stacked conformation is found to be the most stable by 8.4 kJ/mol. This result, which is somewhat unexpected, stems from parameter choices that were driven by the goal of obtaining accurate heats of sublimation of base crystals (A.D. MacKerell, personal communication). As described below, our calculations with this force-field indicate that, in chloroform, a stacked conformation is again more stable than any hydrogen bonded conformation. This presumably reflects the same compromise in the parameterization. Identifying the hydrogen-bonded conformations of the AA pair in chloroform required imposing a more stringent convergence criterion in the MM calculations than was used for the other complexes.

Table 1
Potential energies (kJ/mol) of the most stable hydrogen-bonded and stacked conformations of the base pairs found with the MM method in vacuo, with CHARMM 98 force-field parameters. Energies are relative to the summed energy of the separated molecules in vacuo. The MM calculation converged without identifying any stacked conformations of CG

Pair	H-Bonded	Stacked
A–A	−45.6	−51.4
A–U	−55.6	−51.4
U–U	−46.4	−43.9
C–C	−77.3	−66.9
C–G	−103.3	(none)
G–G	−89.5	−78.6

3. Results and Discussion

3.1. Analysis of the bound conformations

The MM algorithm used here aggressively samples the conformations of the system under study and computes the free energy of each stable conformation that is found. This procedure avoids missing stable conformations that might not otherwise be noted. It also makes it possible to test the hypothesis that the free energy of a molecular system with a modest number of degrees of freedom will be dominated by a few important conformations [2,32]. If this is true, then the calculation of free energy is essentially complete once these conformations are identified and accounted for. This section examines the free energy distribution and the predominant conformations of the bound complexes in chloroform.

The MM algorithm finds hydrogen-bonded conformations to be the most stable for five of the six base-pairs examined and for NMA. In these cases, the algorithm identifies all the hydrogen-bonding

patterns that are expected to be energetically important. For example, Fig. 2 illustrates the four most stable conformations of the AU base-pair. These correspond to the classic Watson-Crick, reverse Watson-Crick, Hoogsteen, and reverse Hoogsteen conformations [41]. That hydrogen-bonded conformations are generally the most stable in chloroform is consistent with the experimental data on base-pairs in chloroform [41].

Only for the AA base-pair does the MM/GB calculation identify stacked conformations as more stable than hydrogen-bonded conformations in chloroform. This result is traceable to the unexpected dominance of stacked conformations in vacuum for AA with this force field. In chloroform, hydrogen bonding becomes still weaker than in vacuo because the dielectric constant of chloroform is 4.8. Therefore stacked conformations dominate even more in solution. The implications of this result are discussed in the section on binding affinities.

If it is true that a few low-energy conformations dominate the free energy of these complexes, then the free energy sum in Eq. (5) should converge once these are accounted for. This is the case for the CG pair, as illustrated in a graph of cumulative free energy vs. number of energy minima (Fig. 3): the calculation is essentially converged after the first energy-minimum is included. However, for the AU base-pair, more energy-minima must be included before the free energy converges (Fig. 3), indicating that the conformational distribution of this complex is less restricted than that of GC. For NMA dimers, the convergence pattern resembles that of the CG base-pair.

3.2. Calculated vs. measured binding affinities

Table 2 shows good agreement between computed and measured standard free energies of binding for the bases and for NMA. The largest deviation of computation from experiment (IR measurements) is 3.3 kJ/mol — less than 1

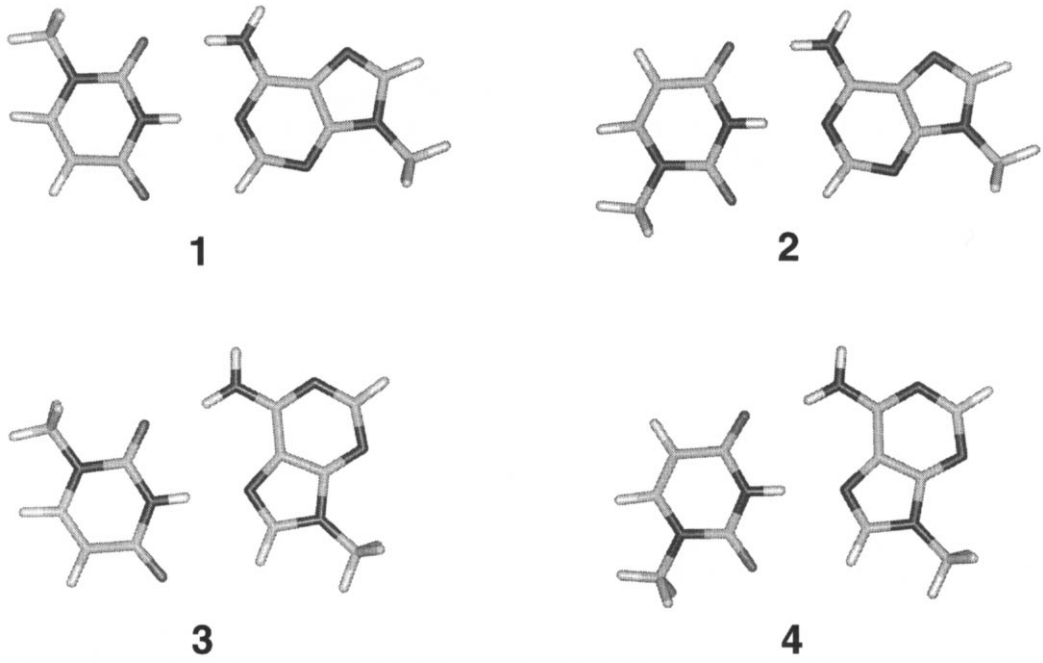

Fig. 2. The four hydrogen-bonded minima of A–U pair. (1) Reversed Watson-Crick; (2) Normal Watson-Crick; (3) Reversed Hoogsteen; and (4) Normal Hoogsteen.

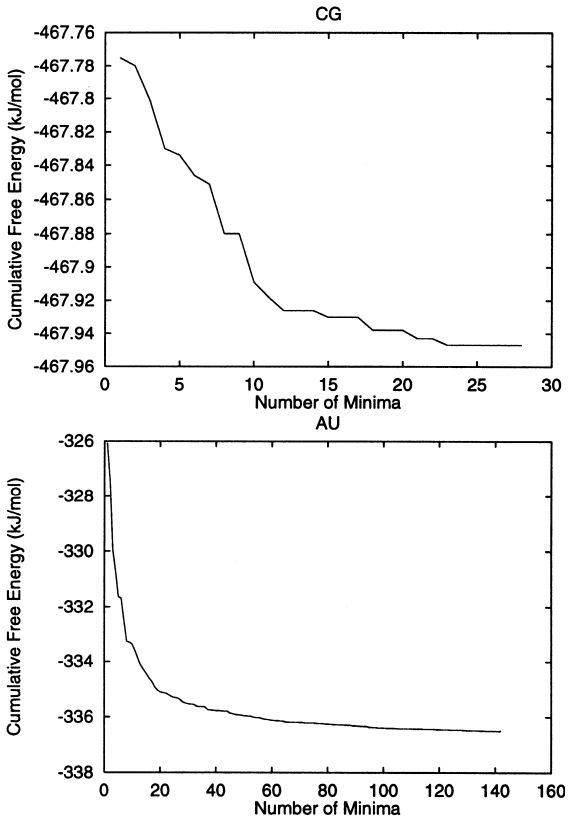

Fig. 3. Cumulative free energy vs. number of minima sampled for C–G and A–U pairs.

kcal/mol — for NMA. The root-mean-square deviation (RMSD) between the computed and measured free energies is 2.4 kJ/mol.

Although the standard free energies of binding agree rather well with experiment, they include a contribution from the 25 kJ/mol offset in the non-polar solvation term. The physical justification for including this contribution, and for using this particular value of it, are not clear-cut, as discussed in Methods and in a previous publication [38]. The offset in effect contributes a fixed −25 kJ/mol to any binding reaction in chloroform, independent of chemical formula or change in surface area. This might be correct: forming a single solute cavity in chloroform may indeed be significantly easier than forming two solute cavities. However, questions remain even if one accepts this view. For example, it is not clear how to handle the case of a flexible molecule whose two ends associate with each other in chloroform. These issues are important because the offset is large for chloroform and because the GB/SA model is already in use for this common solvent [20,42]. The issue is less important in the case of water, for which the offset is only 3.6 kJ/mol [10].

The issue of the offset can be circumvented by examining relative binding free energies, because the contribution of the offset cancels in this case. Accordingly, computed and experimental relative binding free energies were computed from the data in Table 2. There are 21 relative energies among the seven cases. The RMSD of all the computed vs. measured energies is 3.6 kJ/mol, or less than 1 kcal/mol.

As described in Methods, the results in Table 2 exclude stacked conformations because the IR binding measurements are not expected to detect stacking (see Methods). However, other measurement techniques, notably calorimetry, are expected to detect all forms of the complex. It is therefore of interest to examine the computed binding affinities when all conformations of the complexes are included. Including stacked conformations produces significant changes only for the AA and CC base-pairs. For CC, the binding free energy becomes more negative by 1.1 kJ/mol,

Table 2
Computed and experimental standard free energies of binding (kJ/mol) in chloroform at temperature 25°C. Computed energies include only hydrogen-bonded conformations of the complexes (see text). The standard state is a hypothetical ideal 1 M chloroform solution. The computed results vary by less than 0.5 kJ/mol when different random number seeds are used in the MM algorithm. ΔG°_{IR} is the mean of the cited IR absorption measurements [45–47]. ΔG°_{calor} lists the available calorimetric measurements [43]. The last column lists the estimated random errors of the measurements as listed in the published papers. No estimate of systematic errors is available

Pair	ΔG°_{calc}	ΔG°_{IR}	ΔG°_{calor}
A–A	−4.5	−2.8 ± 0.2	−6.4 ± 0.4
A–U	−12.5	−11.6 ± 0.5	−11.5 ± 0.3
U–U	−6.3	−4.5 ± 0.3	−7.4 ± 0.6
C–C	−11.7	−9.3 ± 0.1	
C–G	−24.2	−27.6 ± 0.3	
G–G	−14.6	−17.2 ± 0.1	
NMA-NMA	−5.9	−2.6 ± 0.1	

slightly degrading the agreement with the IR measurements. For the AA base pair, including stacked conformations leads to a standard free energy of binding of -7.6 kJ/mol, which deviates markedly from the IR results, but agrees rather well with the calorimetric result, -6.4 ± 0.4 kJ/mol [43]. This observation raises the possibility that AA may indeed stack significantly in chloroform. If stacking does indeed compete significantly with hydrogen bonding, then the IR study that monitored only hydrogen bonding [34] should have detected a deviation from ideal two-state behavior. This would have been manifest as a non-linearity in the plot of IR absorbance vs. total concentration of A (Eq. (4) in the experimental paper [34]). The text states that the plot was linear, but the goodness of fit cannot be evaluated quantitatively because the raw data are not listed in the paper. Interestingly, however, the van't Hoff plot for AA is clearly non-linear, suggesting that more than one binding reaction occurs. Thus, the authors regard the AA binding data as 'less reliable' than their results for UU and AU. Further experimental work would be needed to ascertain whether these presumed alternative binding reactions involve AA stacking.

4. Conclusions

The computationally efficient MM/GB method yields rather good agreement with measured binding free energies of hydrogen-bonded complexes in chloroform. Interestingly, the combination of calorimetric and computational results for AA hint that stacked conformations compete significantly with hydrogen bonded conformations in this case, despite the low polarity of the solvent. The calculations rapidly identify the most important binding modes and determine the contribution that each one makes to the overall free energy of binding. This contrasts with perturbative free energy simulations, which normally require the user to guess in advance which conformation is thermodynamically the most important, and which therefore risk missing important binding modes. The present calculations are fast enough that they should permit statistically meaningful validation studies of the method to be carried out.

We also find that the non-polar solvation term for chloroform has a large offset of ~ 20 kJ/mol that complicates the calculation of binding affinities. This issue, which has been discussed previously in the case of alkane solvents [38], can be sidestepped by focusing on relative binding affinities, for which the offset cancels. However, the use of implicit solvent models for non-aqueous solvents is an important topic for future study. Note that the issue is less critical for aqueous solvents, because their offset is small [10].

The present computational approach should be useful for computing the binding affinities of molecules more complex than those examined here. Indeed, the MM algorithm has been shown to converge for isolated molecules with up to 18 rotatable bonds [6,44] and for salt-bridging side-chains in helix-forming peptides [7]. A related algorithm has given good results for larger systems [20]. Continued development of these algorithms should yield methods of computing binding affinities that are readily tested and that will be useful in a variety of practical applications. It is worth mentioning that the algorithm in Fig. 1 is quite general and could be implemented with any number of different algorithms for finding energy minima and evaluating their configuration integrals [2].

Acknowledgements

We thank Prof Alex D. MacKerel for providing us the latest CHARMM parameters for nucleic acids. We are grateful to Dr David S. Nunn for helpful conversations. This work was supported by the National Institutes of Health (GM54053 to MKG) and the National Institute of Standards and Technology. M.S.H. was supported by a National Research Council Research Associateship. Certain commercial equipment or materials are identified in this paper in order to specify the methods adequately. Such identification does not imply recommendation or endorsement by the

National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

[1] Ajay, M.A. Murcko, J. Med. Chem. 38 (1995) 4953–14496.
[2] M.K. Gilson, J.A. Given, B.L. Bush, J.A. McCammon, Biophys. J. 72 (1997) 1047–1069.
[3] B. Lesyng, W. Saenger, Carbohydr. Res. 133 (1984) 187–197.
[4] K.B. Lipkowitz, D.A. Demeter, R. Zegarra, R. Larter, T. Darden, J. Am. Chem. Soc. 110 (1988) 3446–3452.
[5] K.B. Lipkowitz, R. Zegarra, J. Comput. Chem. 10 (1989) 595–602.
[6] M.S. Head, J.A. Given, M.K. Gilson, J. Phys. Chem. 101 (1997) 1609–1618.
[7] R. Luo, L. David, H. Hung, J. Devaney, M.K. Gilson, J. Phys. Chem., in press.
[8] I. Kolossvary, J. Phys. Chem. A 101 (1997) 9900–9905.
[9] B. Honig, K. Sharp, A.-S. Yang, J. Phys. Chem 97 (1993) 1101–1109.
[10] D. Sitkoff, K.A. Sharp, B. Honig, J. Phys. Chem. 98 (1994) 1978–1988.
[11] K.A. Sharp, A. Nicholls, R. Friedman, B. Honig, Biochemistry 30 (1991) 9686–9697.
[12] M.K. Gilson, K.A. Sharp, B.H. Honig, J. Comp. Chem. 9 (1988) 327–335.
[13] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, J. Am. Chem. Soc. 112 (1990) 6127–6129.
[14] M.K. Gilson, B. Honig, J. Comput.-Aided Mol. Des. 5 (1990) 5–20.
[15] M. Schaefer, C. Froemmel, J. Mol. Biol. 216 (1990) 1045–1066.
[16] G.D. Hawkins, C.J. Cramer, D.G. Truhlar, Chem. Phys. Lett. 246 (1995) 122–129.
[17] M. Scarsi, J. Apostolakis, A. Caflisch, J. Phys. Chem. 101 (1997) 8098–8106.
[18] D. Qiu, P.S. Shenkin, F.P. Hollinger, W.C. Still, J. Phys. Chem. 101 (1997) 3005–3014.
[19] D.Q. McDonald, W.C. Still, J. Am. Chem. Soc. 116 (1994) 11550–11553.
[20] I. Kolossvary, J. Am. Chem. Soc. 119 (1997) 10233–10234.
[21] R. Luo, M.S. Head, J. Moult, M.K. Gilson, J. Am. Chem. Soc. 120 (1998) 6138–6146.
[22] N. Bental, I.A. Topol, A. Yang, S.K. Burt, B. Honig, J. Phys. Chem. 101 (1997) 450–457.
[23] W.L. Jorgensen, J. Am. Chem. Soc. 111 (1989) 3770–3771.
[24] S.F. Sneddon, D.J. Tobias, C.L. Brooks III, J. Mol. Biol. 209 (1989) 817–820.
[25] J. Pranata, W.L. Jorgensen, Tetrahedron 47 (1991) 2491–2501.
[26] T. Simonson, A.T. Brunger, J. Phys. Chem. 98 (1994) 4683–4694.
[27] K. Osapay, W.S. Young, D. Bashford, C.L. Brooks III, D.A. Case, J. Phys. Chem. 100 (1996) 2698–2705.
[28] I.Z. Steinberg, H.A. Scheraga, J. Biol. Chem. 238 (1963) 172–181.
[29] Hill, T.L., An Introduction to Statistical Thermodynamics, Dover, New York, 1986.
[30] W.L. Jorgensen, J. Am. Chem. Soc. 111 (1989) 3770–3771.
[31] M.-C. Justice, J.-C. Justice, J. Sol. Chem. 5 (1976) 543–561.
[32] M.K. Gilson, Proteins Struct. Funct. Genet. 15 (1993) 266–282.
[33] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, London, 1986.
[34] Y. Kyogoku, R.C. Lord, A. Rich, J. Am. Chem. Soc. 89 (1967) 496–504.
[35] M.E. Davis, J.D. Madura, B.A. Luty, J.A. McCammon, Comput. Phys. Commun. 62 (1991) 187–197.
[36] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, J. Comput. Chem. 4 (1983) 187–217.
[37] A.D. MacKerell, Jr., J. Wiókiewicz-Kuczera, M. Karplus, J. Am. Chem. Soc. 117 (1995) 11946–11975.
[38] D. Sitkoff, N. Ben-Tal, B. Honig, J. Phys. Chem. 100 (1996) 2744–2752.
[39] D.J. Giesen, C.C. Chambers, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. 101 (1997) 2061–2069.
[40] Quanta 4.1 Molecular Simulations Inc. San Diego, CA.
[41] W. Saenger, Principles of Nucleic Acid Structure, Springer-Verlag, New York, 1984.
[42] F. Mahamadi, N.G.J. Richards, W.C. Guida, et al., J. Comput. Chem. 11 (1990) 440.
[43] J.S. Binford, Jr., D.M. Holloway, J. Mol. Biol. 31 (1968) 91–99.
[44] L. David, R. Luo, M.S. Head, M.K. Gilson, J. Phys. Chem., in press.
[45] Y. Kyogoku, R.C. Lord, A. Rich, Biochim. Biophys. Acta 179 (1969) 10–17.
[46] P. Carmona, M. Molina, A. Lasagabaster, R. Escobar, A.B. Altabef, J. Phys. Chem. 97 (1993) 9519–9524.
[47] S.E. Krikorian, J. Phys. Chem. 86 (1982) 1875–1881.