A Novel Method to Simulate the Hydrophobic Effect and Its Application to the Ranking of Host/Guest Complexes

Arthur M. Doweyko* and Stephen R. Johnson*

Computer-Assisted Drug Design, Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, New Jersey 08543

Received May 9, 2006

Solvent entropy change is a major factor in driving the association of hydrophobic species in aqueous solutions. We have developed a novel methodology which simulates the solvation of hydrophobic surfaces by water. A system of virtual solvent particles surrounding the solute governed by arbitrarily determined rules provides a means to estimate the degree of order (*Q*) imposed by such solvation. Computed changes in *Q* (dQ) upon complex formation have been found to correlate well with observed binding affinities of host−guest complexes in aqueous solutions. Examples are described which illustrate the ability of dQ calculations to identify the correct ligand pose from a set of decoy complexes as well as provide rank ordering of a set of highly diverse ligand−protein complexes. Comparisons to surface-area-based calculations are discussed. The *Q* methodology holds great promise in the development of predictive structure-based approaches to drug design, as it provides a relatively simple means to estimate the hydrophobic effect.

INTRODUCTION

The free energy change that drives the complexation of a host and a ligand is made up of both enthalpic and entropic factors. The entropic component includes contributions from the ligand, receptor, and solvent system in which the complexation occurs. The entropy change of the solvent represents a significant portion of the total entropy and is often referred to as the *hydrophobic effect*. There have been several recent reviews of the hydrophobic effect.[1−5]

To mathematically model the complexation affinity, one must account for both the enthalpy and entropy accurately. Generally, the enthalpic component is calculated using methods that are largely consistent among researchers.[2] This is not the case for the calculation of the entropic part of the affinity. While there are simple heuristics for the entropy change of the ligand, and perhaps the receptor, the entropy arising from the solvent is typically ignored in ligand−protein complexation.

For small ligands, the hydrophobic effect is believed to arise from the formation of a clathrate structure by water surrounding the ligands. Consequently, the free energy of solvation appears to scale with molecular volume for smaller solutes.[6] For larger ligands, recent simulations indicate that the presence of a hydrophobic surface creates a drying interface with the solvent of more than one molecular diameter of the solvent from the hard sphere surface.[6,7] The entropic change arises from the rearrangement of the hydrogen-bond network of the water surrounding the ligand, receptor, and resulting complex.

We refer the reader to the introduction of Kelly and Mancera[8] for a brief review of scoring functions for the hydrophobic force in protein−ligand interactions. In recent years, several simplified systems intended to replicate many of the fundamental properties of water have been reported.[3,7,9−14] Among these properties are the heat capacity changes brought about by the addition of a solute to the solvent.[9,10] New methods explicitly designed for use in scoring protein−ligand complexes have been reported.[8] Still, most scoring functions used in docking applications continue to rely on the use of buried hydrophobic surface area as a surrogate for solvent entropy changes.

We report a simple, empirical approach to encoding the change in solvent entropy that is consistent with the "Iceberg" model proposed by Frank and Evans[15] and implicitly considers the curvature effects not captured by most surface area metrics.[16−18] In contrast to most models of solvent entropy used in docking and scoring, our approach uses a highly simplified model of the solvent itself. The approach focuses on the degrees of freedom (*Q*s) of association possessed by a dimensionless solvent particle. Upon the addition of a solute, the degrees of freedom of the solvent are reduced as the numbers of solvent neighbors available for association are reduced. The change in the number of neighboring solvent particles upon complexation is utilized as the basis for a simple measure of solvent entropy. This value is further diminished in a distance-dependent way by the proximity of each solvent particle to a hydrophobic atom of the solute. The utilization of this simple system in the scoring of docked poses, affinity prediction, and loop modeling is presented.

METHODS

Surface Area Calculations. Surface area calculations were carried out using Savol.[19] Hydrophobic surface areas were defined as all those not including polar regions surrounding nitrogen and oxygen atoms, or their attached hydrogens. Calculations of solvent-accessible surface areas

* Corresponding author phone: (609) 252-3632 (A.M.D.), (609) 252-3003 (S.R.J.); e-mail: arthur.doweyko@bms.com (A.M.D.), stephen.johnson@bms.com (S.R.J.).

10.1021/ci600188h CCC: $33.50
Published on Web 10/14/2006

utilized a setting of 1.4 Å as the solvent radius except where noted. Typical correlations involving hydrophobic surface areas focused on the difference in hydrophobic surface area obtained between the ligand and the host, and the ligand/host complex. Thus, the buried hydrophobic surface area = $\text{area}_{\text{ligand}} + \text{area}_{\text{host}} - \text{area}_{\text{ligandhost}}$.

***Q* Methodology.** The change in solvent entropy is an important component of the overall binding constant for host/guest complexation. We have developed a simple metric, based on the degrees of freedom of a model solvent particle, that follows simple, defined, and arbitrary association rules. Our metric, like the change in solvent entropy, focuses on the change in degrees of freedom for the model solvent as the solutes go from independent entities to a single bound system.

Each solute molecule (e.g., free ligand, apo receptor, and the bound complex) is placed on a regularly spaced lattice, where the lattice points represent the geometric location of a dimensionless model solvent particle ("*Q* particle") that solvates each solute. While each solute is placed on its own lattice, each lattice is of the same size and extends at least 7 Å beyond the atoms of the largest solute. Calculations discussed here use a grid spacing of 1 Å unless otherwise specified.

By definition, each *Q* particle interacts only with neighboring *Q* particles that are on the same *x*, *y*, or *z* axis (i.e., no "diagonal" interactions are permitted). In this way, the maximum number of degrees of freedom for a solvent particle is six (two neighbors in each of the *x*, *y*, and *z* directions). In the presence of a solute, however, some *Q* particles will not have six neighbors and will consequently have fewer degrees of freedom available to them. In addition, *Q* particles that are near to the surface of a hydrophobic solute atom have a further loss of freedom that is dependent on a coarse stepwise rule:

$$\Delta\text{D.O.F.} = \begin{cases} -3.0, & d < 1 \\ -1.5, & d < 2 \\ -0.6, & d < 2.5 \end{cases}$$

where *d* is the distance in angstroms between the *Q* particle and the van der Waals surface of the hydrophobic solute atom. For the purposes of this calculation, hydrophobic atoms are defined as any C, halogen, S, or amide N. On the basis of these arbitrary rules, the degrees of freedom of a particular solvent particle ranges from −3 to 6. A simple schematic illustrating the *Q* calculation is shown in Figure 1. Figure 2 shows the gradient in the degrees of freedom of the *Q* solvent for a pose of 3-isobutylphenyl acetate. The effect propagates through several solvation layers, consistent with the behavior of water seen in molecular dynamics simulations. The total *Q* of a system is then calculated as the sum of the degrees of freedom of all included *Q* particles. The overall change in *Q* (dQ) is defined as

$$\text{dQ} = (Q_{\text{complex}} + Q_{\text{empty}}) - (Q_{\text{ligand}} + Q_{\text{receptor}}) \quad (1)$$

where Q_{empty} is the sum of the degrees of freedom of a lattice containing no solute molecules.

Pose Generation. Starting structures for protein/ligand complexes were either obtained directly from the Protein Data Bank (PDB; Research Collaboratory for Structural Bioinformatics, RCSB)[15,20] or were part of a triaged subset

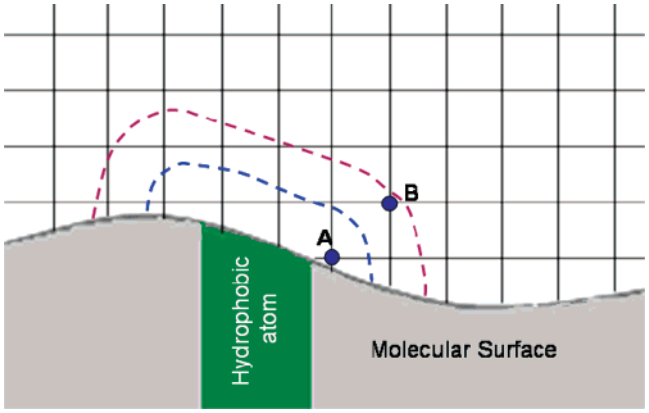

Figure 1. Illustration of the algorithm to calculate degrees of freedom (DOF) in two dimensions. The maximum number of neighboring *Q* particles in two dimensions is four. The green region represents the surface of a hydrophobic atom. The blue dashed line represents the 1 Å distance from the hydrophobic surface, while the purple dashed line is the 2 Å cutoff. In the calculation for point A, of the four possible neighboring solvent particles, two are displaced by the molecular surface. Therefore, DOF(A) = 4 − 2 = 2. The point is within 1 Å of a hydrophobic surface, leading to an adjustment: DOF(A) = 2 − 3 = −1. For point B, all four neighboring *Q* particles are present. DOF(B) = 4; point B is within 2 Å of a hydrophobic surface, therefore, DOF(B) = 4 − 1.5 = 2.5.

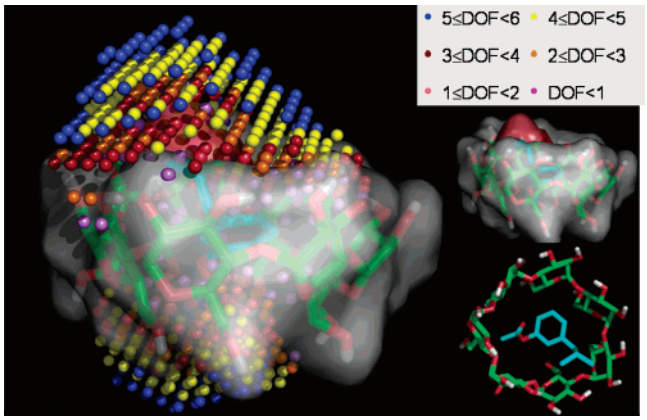

Figure 2. Graphical display of the degrees of freedom of the solvent proximal to the docked ligand in β-cyclodextrin used in a *Q* calculation. The structure shown is 3-isobutylphenyl acetate docked into β-cyclodextrin using QXP/mcdock+. Note that several solvation layers are affected by the presence of a hydrophobic solute.

such as the X-Score set[21] or the GOLD evaluation set.[22] The generation of decoy poses for binding pose analysis, for example, protein/ligand and cyclodextrin/ligand complexes, was carried out with the help of the MCDOCK module within Flo(qxp).[23] Typically, the host molecule or binding site (all residues within 10 Å of the ligand as found in the X-ray structure) is kept rigid while the ligand is flexibly redocked a defined number of times to ensure the generation of docked poses that are reasonably different from one another. Ligand root-mean-square deviation (RMSD) from the X-ray structure was calculated using the RMSLIG program provided as part of the Flo modeling package. In most cases, MCDOCK was carried out four or more times using different ligand starting poses in order to harvest the target number of 100 decoy poses.

Table 1. Observed and Calculated Solvation Energies for Some Alkyl Amines

compound	ΔG_{exp}[a] (kcal/mol)	Q	ΔG_{SCRF}[b] (kcal/mol)	ΔG_{LD}[c] (kcal/mol)	ΔG_{LD}(phob)[c] (kcal/mol)
butylamine	−4.29	22 237.5	−4.1		
ethylamine	−4.5	22 902.2	−4.7	−2.6	2.9
methylamine	−4.56	23 232.8	−5.2	−2.9	2.1
pentylamine	−4.1	21 933.2	−3.9		
propylamine	−4.39	22 549.2	−4.7		
diethylamine	−4.07	22 113.9	−4.1		
dimethylamine	−4.29	22 746.3	−5.1	−1.2	3.3
dipropylamine	−3.66	21 489.7	−4.0		
trimethylamine	−3.23	22 420.4	−4.2	−0.1	4.4

[a] Observed solvation energies from ref 24. [b] Calculated using Jaguar v.6.5, BL3YP/6-31+g**/water.[27] [c] Values from the Langevin Dipoles method from ref 29.

RESULTS

Correlation of *Q* with Solvation Energy. The solvation energies of a series of alkyl amines were collected from the literature.[24] The structure of each compound was modeled in Maestro[25] and minimized as the neutral form using Batchmin[26] with OPLS2005 and the water solvation model. The minimized structures were exported from Maestro and used to calculate *Q*. In addition, the structure of each compound was optimized using Jaguar[27] with the BL3YP/6-31G**+/water basis set. The solvation model was used to calculate the self-consistent reaction field (SCRF) solvation energy for each compound. The observed solvation energy and calculated values for *Q* and the SCRF solvation energies are given in Table 1.

The correlation between the solvation energies of the alkyl amines and *Q* is $r^2 = 0.34$. However, excluding a single compound, trimethylamine, the correlation between *Q* and the solvation energy is $r^2 = 0.88$. This compares favorably to the correlation of the solvation energy with the SCRF calculated values ($r^2 = 0.35$ or $r^2 = 0.55$ excluding trimethylamine). Figure 3 shows the relationships between the observed solvation energy and both the SCRF calculated solvation energies and *Q*. Interestingly, the relationship between *Q* and the solvation energy is fairly linear for both primary and secondary amines. Trimethylamine is a notable outlier using the methodology described here, as it has been in other studies.[28]

The *Q* method is related in some respects to the hydrophobic component of the Langevin Dipoles method of Florian and Warshel.[29] In the Langevin Dipoles method, the electrostatic potential of the Langevin Dipoles within 1.5 Å of the solute surface are used in an empirically derived equation to calculate the hydrophobic component of the solvation energy. The Langevin Dipole results for several of the alkyl amines are also shown in Table 1. While there are some similarities between the two methods, including the use of grid points to represent the solvent, there are a number of important differences as well. First, the *Q* method relies on atom typing rather than electrostatic potentials to determine the hydrophobicity of solvent-exposed atoms. In addition, the Langevin Dipole method focuses only on the first solvation layer when calculating the hydrophobic component of the solvation energy. In contrast, the *Q* method explicitly includes the effect of the solute over several solvation layers.

It is worth noting that there are many additional methods in the literature for assessing hydrophobic solvation, including the Pierotti equation and related scaled particle theory methods,[30−32] the methods of Floris and Tomasi,[33] and the SCF models contributed by Truhlar and Cramer.[34,35] An important difference between the *Q* method and that of Pierotti is that *Q* does not make any assumption regarding the solute shape. The use of a grid in the calculation allows any molecular shape to be handled numerically. In addition, while computing power continues grow, *Q* is computationally accessible even for the large biological systems often of interest in protein/ligand binding.

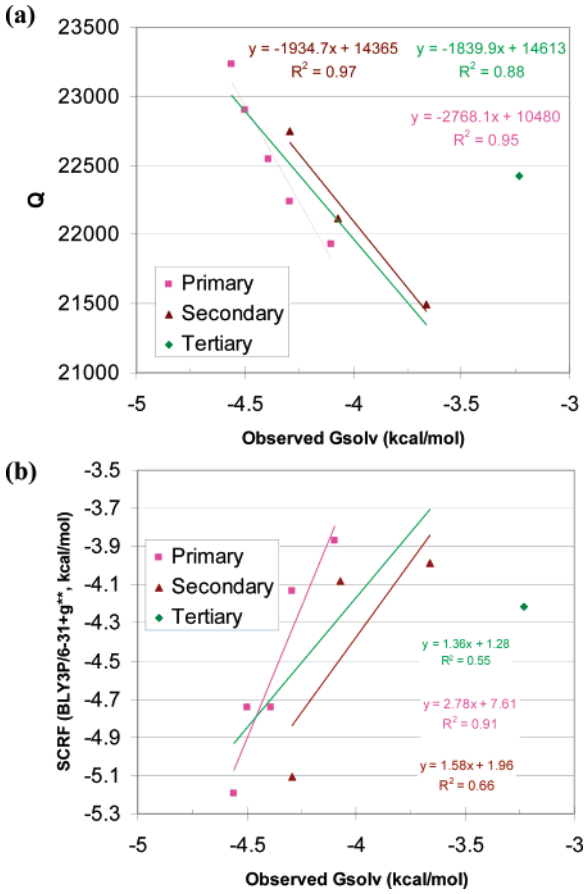

Figure 3. Relationships between the observed alkylamine solvation energy (a) *Q* and (b) solvation energies calculated by SCRF. The green line represents the correlation for the primary and secondary data. The correlations do not include trimethylamine in the calculation.

Phenyl Barrel. A series of 13 benzene molecules manually placed ca. 0.5 Å apart within a hydrophobic cavity made up by benzene molecules served as a model to study the

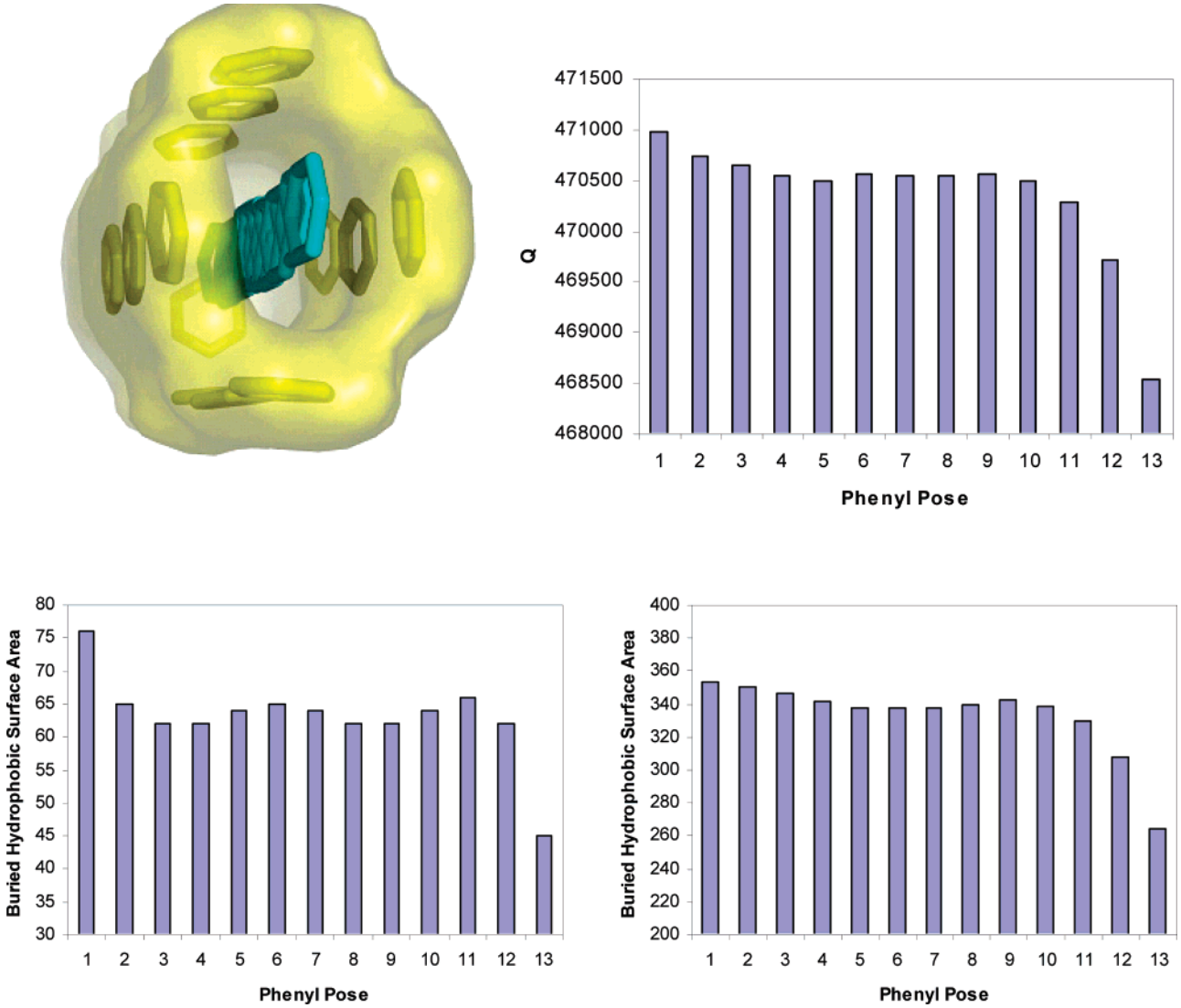

Figure 4. Phenyl barrel model (upper left panel). Benzene poses are numbered consecutively from the deepest (1) to the outermost (13). Upper right panel: the highest *Q* value (corresponding to the highest solvent entropy) is associated with pose 1. The lower left and right panels represent buried hydrophobic surface area calculations using a probe radius of 0.5 and 1.4 Å, respectively.

sensitivity of *Q* and surface area calculations in identifying the most buried benzene pose (Figure 4). In this series, poses 1−13 were numbered in the order of most buried to least buried. Values of *Q* were determined for each complex (benzene + phenyl barrel). In addition to the *Q* analysis, the buried hydrophobic surface area was calculated for each complex using a standard 1.4 Å probe radius and a smaller 0.5 Å probe radius. As can be seen from the bar graphs depicted in Figure 4, the most buried phenyl pose (1) is easily identified by its highest ranked *Q* value. Interestingly, the buried hydrophobic surface area calculations were found to be affected by the choice of probe size, to the extent that, when using the standard probe size of 1.4 Å, the most buried pose is not strongly set apart from any of poses 2−10, all of which occupy some part of the hydrophobic tunnel. However, the use of a smaller probe size (0.5 Å) was found to enhance the buried hydrophobic surface area signal for pose 1.

Binding Pose. Several crystal structures available from the RCSB were used to investigate the utility of the *Q* methodology to identify best binding poses for protein−ligand complexes. The results from three such analyses are presented here and illustrated in Figure 5. The three complexes chosen represent ligands with significantly different nonpolar surface areas (expressed as np/tot = nonpolar/total surface area). Compounds **1**, **2**, and **3** represent ligands taken from RCSB entries 1FKG, 1AEC, and 1ROB, respectively, and were calculated to be decreasingly hydrophobic with np/tot = 0.91, 0.55, and 0.38, respectively. As described previously, in each case, 100 decoy poses were generated, with each complex yielding a dQ value. Plots of dQ versus RMSD (relative to the protein−ligand X-ray structure) illustrate the association of high dQ values with poses having the least RMSD, that is, poses most closely resembling the X-ray structures. Flo (Eass) energetics (kilojoules per mole) are provided as a reference typical of a molecular-mechanics-based set of calculations. Flo utilizes an Amber force field; Eass is defined as the total contribution of interactions and strain energies for the protein−ligand complex.

Because the hydrophobic effect is strongly correlated to hydrophobic surface areas, it is expected that similar results will be obtained by using buried hydrophobic surface areas as the key metric pointing to the best pose. Initial investiga-

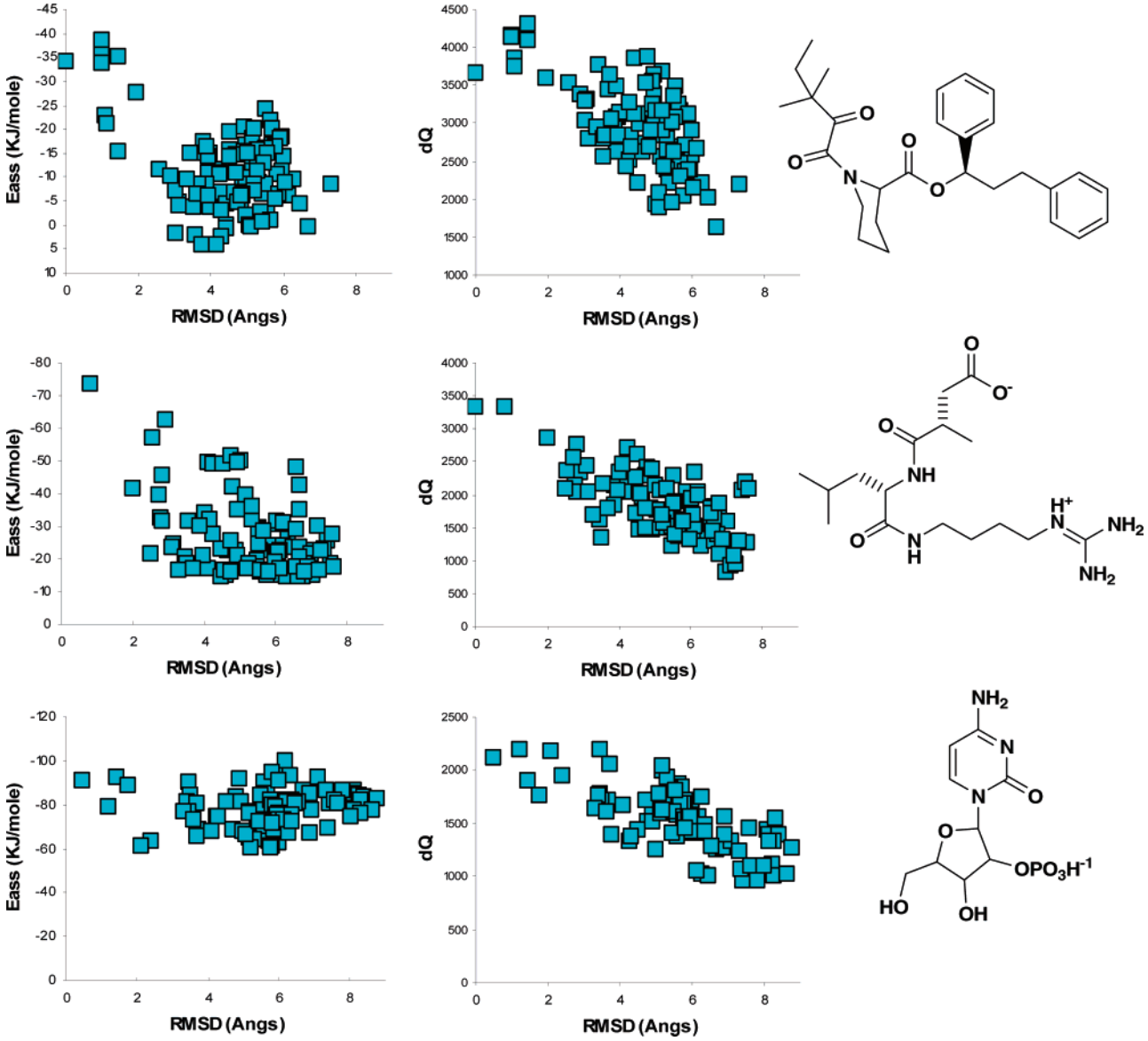

Figure 5. Three examples of pose selection using ligands with different ratios of nonpolar to total surface areas. Top panel: FK506 binding protein (1FKG) using a ligand with np/tot = 0.91. Middle panel: actinidin hydrolase (1AEC) using a ligand with np/tot = 0.55. Bottom panel: cytidylate/ribonuclease (1ROB) using a ligand with np/tot = 0.38. A total of 100 decoy poses for each ligand were evaluated by comparing the dQ and RMSD (X-ray) for each.

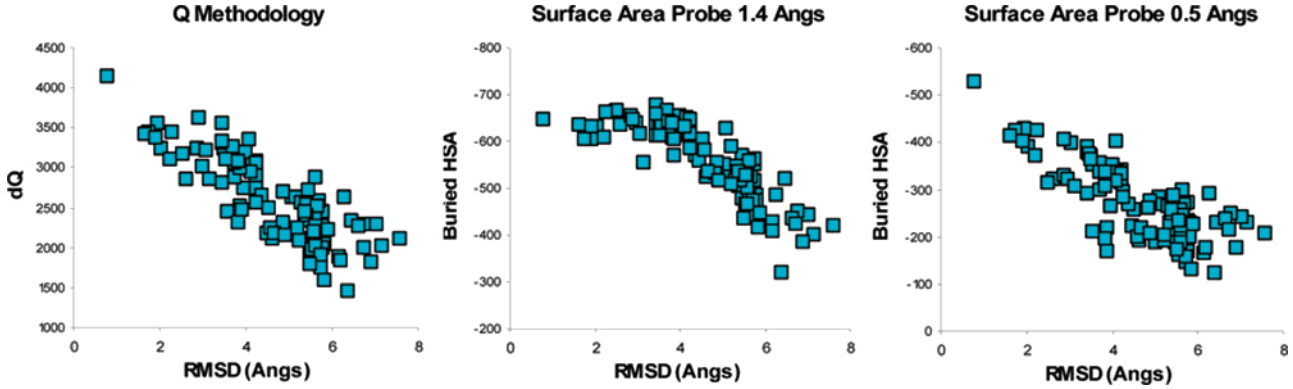

Figure 6. Binding pose analysis of the ligand in 1FKI. dQ results clearly point to the preferred ligand pose (left), while buried hydrophobic surface area calculations using a 1.4 Å probe fail to identify the best pose (center). The recalculation of buried hydrophobic surface areas using a 0.5 Å probe highlights the correct pose having the largest value (right).

tions using the standard probe size of 1.4 Å (water molecule radius) to provide solvent-accessible surface areas indicated a poor performance as illustrated for the example in Figure 6 (RCSB entry 1FKI). The use of a smaller probe size (0.5 Å) yielded results comparable to those obtained with the *Q* analysis. The reasons for this observation may lie in the

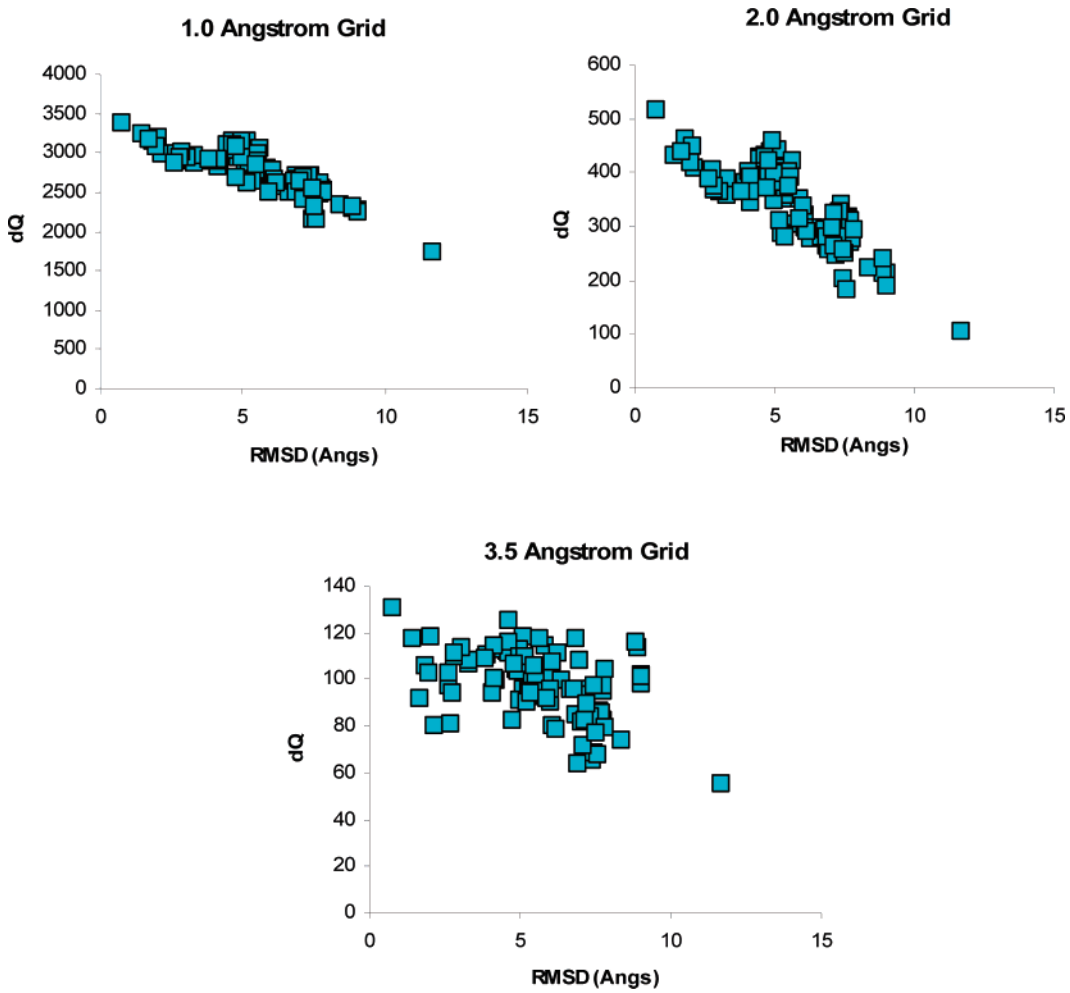

Figure 7. Effect of grid spacing on dQ binding pose identification using the TMP/DHFR example (1DYR). The correct binding pose is most easily distinguishable at a small grid spacing, 1.0−2.0 Å, while at larger spacing intervals, the dQ values are attenuated and differences between poses smaller.

inability of solvent-accessible surface area changes to accurately account for the hydrophobic effect when computed using a probe size of 1.4 Å radius.

Grid Spacing. As the *Q* methodology is grid-based, it was of interest to see what effect grid spacing had on the pose-selection process. Using the trimethoprim/dihydrofolate reductase (TMP/DHFR) example (RCSB entry 1DYR), dQ values were calculated with grid spacings ranging from 1.0 to 3.5 Å. The results are shown in Figure 7. While dQ values were highest for poses closest to the X-ray structure, as expected, larger grid spacing was found to result in smaller dQ changes with a relative increase in signal "noise" brought about by a decrease in the absolute dQ signal.

The Encysting Effect. The case of trimethoprim binding to DHFR (1DYR) is an example for which the buried hydrophobic surface area does not correctly identify poses similar to the X-ray structure, as illustrated in Figure 8. Once again, 100 decoy poses were evaluated using dQ and buried hydrophobic surface area values. In this case, surface area calculations yielded a significant deviation in the form of a series of poses (RMSD = 4−6 Å) which scored higher than pose 1 (X-ray, relaxed, at RMSD = 0.6 Å). A close inspection of one such example, pose 33, revealed that it was inverted and sealed the mouth of the binding site. In this encysted state, the solvent-accessible surface area calculation was blind to much of the unoccupied binding site, resulting in an over-estimation of the buried surface area. At smaller probe sizes, the effect is attenuated, because more of the unoccupied binding site is accounted for.

The Mesa Effect. A second potentially critical difference between the *Q* methodology and surface area calculations may be related to a relative response effect between two surfaces, or the mesa effect. A comparison of these two metrics is shown in Figure 9, wherein the dQ values and buried surface areas are calculated for the approach of a benzene molecule toward a hydrophobic surface made up of other benzene molecules. For the case of the buried hydrophobic surface area, no metric is realized until a van der Waals contact is made at about 7 Å, followed by a relatively linear curve reflecting the continual intersection of the van der Waals radii of benzene and the benzene plane (up to 1 Å). For the same series of poses, dQ values follow a nonlinear response, increasing as the distance nears the same 1 Å endpoint. This is a consequence of the gradient calculation discussed previously. Thus, the model system demonstrates that a nonlinear response surface exists for the dQ approach, with a greater relative metric as indicated by the larger slope in Figure 9 as the two hydrophobic surfaces collapse toward each other. Such a condition may represent a distinct advantage of the dQ simulation over surface area calculations in providing a special sensitivity to interactions between complementary and flat hydrophobic surfaces. The

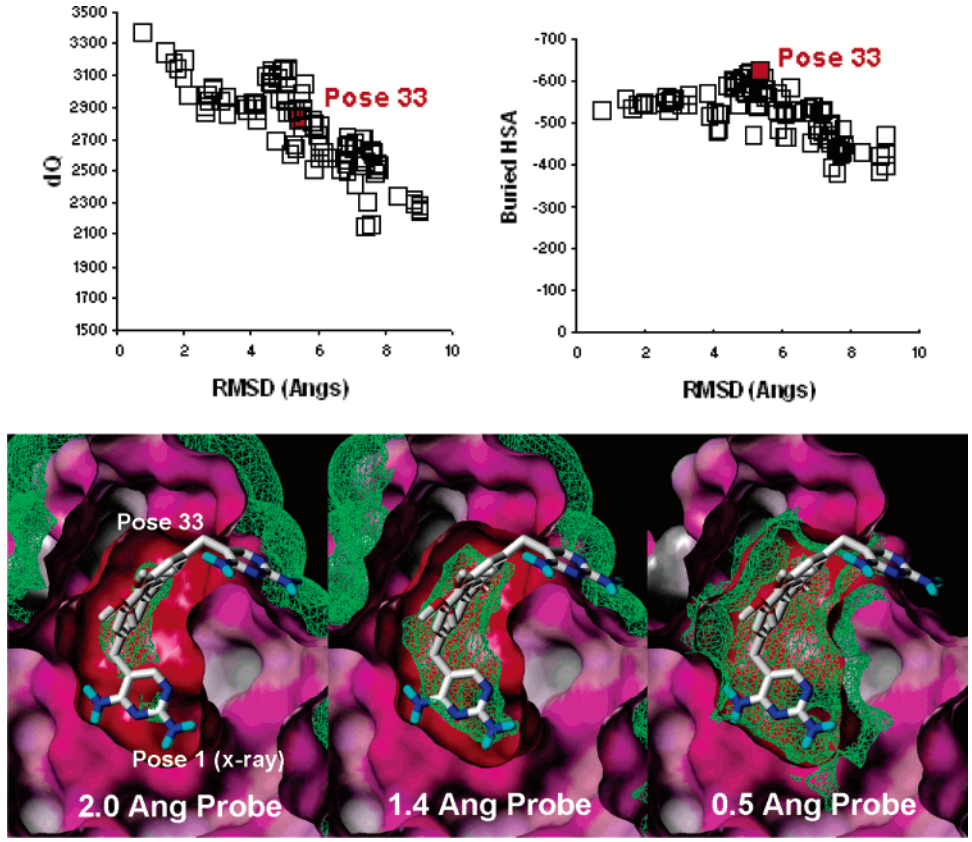

Figure 8. TMP/DHFR (1DYR) binding pose selection results. dQ and buried hydrophobic surface area analyses were conducted using a grid spacing of 1 Å and a probe size of 1.4 Å, respectively (top). TMP poses 1 (X-ray) and 33 are shown (bottom) with a mesh reflecting protein surface areas accounted for at each probe size.

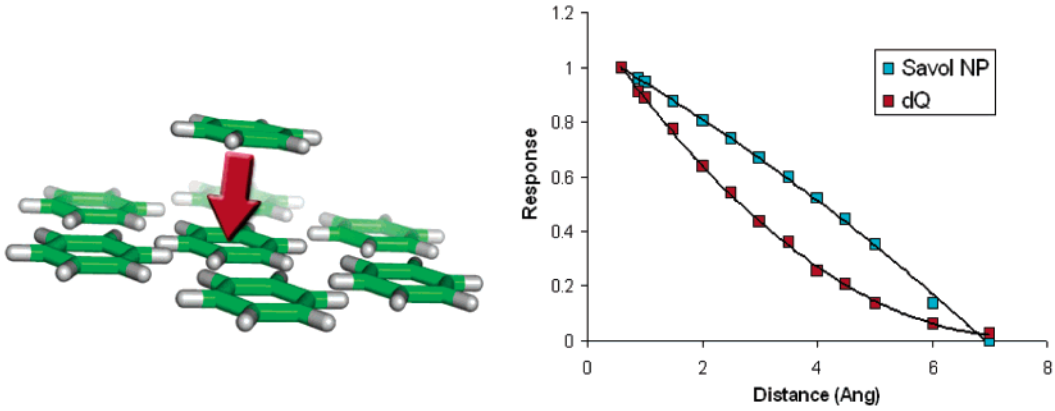

Figure 9. Relative response surface analysis comparing buried surface area calculations to *Q* methodology for a model system composed of a hydrophobic mesa (a benzene plane) and an approaching benzene molecule.

binding of a ligand to elastase (RCSB entry 2EST) represents an illustration of the hydrophobic mesa effect (Figure 10). In the X-ray structure, the pendant trifluoromethylphenyl group represents a planar hydrophobe belonging to the ligand which is found to lie along a relatively flat and exposed hydrophobic bed on the enzyme. After the generation of 100 decoy poses, a dQ analysis conducted without a gradient calculation afforded a number of high-scoring poses ranging up to a 5 Å RMSD from the X-ray pose (left panel). When the gradient terms were utilized, the analysis pointed to only several poses ranging 1−2 Å in RMSD (right panel). These results underscore the importance of the gradient effect when addressing flat and exposed hydrophobic surfaces in protein binding sites.

Ligand/Protein Complexes. As binding pose prediction for modestly nonpolar compounds appeared to be well-handled by the *Q* methodology, it was of interest to see whether the dQ metrics could correlate with observed ligand−protein binding affinities. To this end, a series of 60 complexes was chosen from the *X*-score set (100 complexes, ref 21), selected to provide a broad range of K_i values (10^{-1} to 10^{-10} M). For each complex, dQ was calculated for both the original coordinates and a relaxed set of coordinates (Flo minimization of the ligand within the rigid protein). Each Flo-minimized complex yielded an Eass (kilojoules per mole) value which is an Amber force-field-based metric accounting for ligand, protein, and complex electrostatic and other nonbonded interaction energetics.

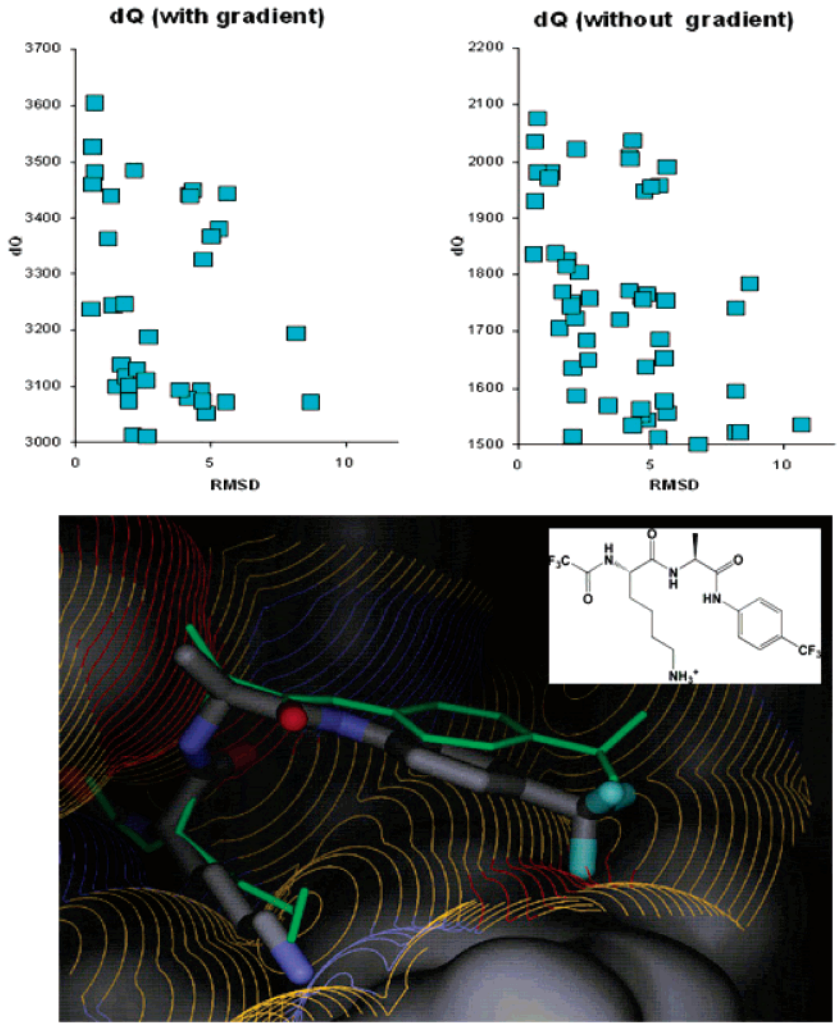

Figure 10. Upper panel: dQ analyses with and without the use of a gradient response to a hydrophobic surface, illustrating the advantage of a gradient response in identifying poses of the ligand (inset) similar to the X-ray structure (using 100 ligand poses docked onto elastase, 2EST): the highest scoring poses are found within a ca. 2 Å RMSD with gradient. Lower panel: illustration of the best pose identified by the gradient version of the *Q* methodology wherein the hydrophobic mesa is occupied by a planar trifluoromethylanilide ring system.

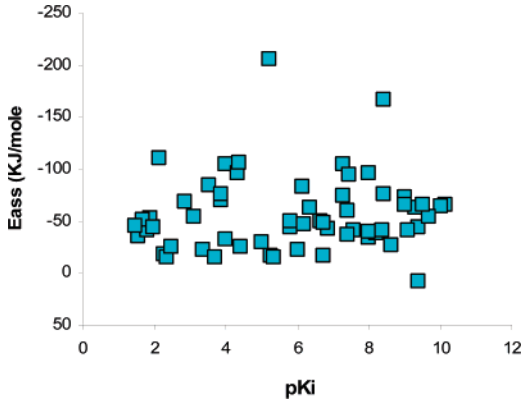

Figure 11. Flo(Eass) scores for 60 ligand/protein complexes (X-ray structures) plotted against observed binding affinities. Eass scores are based on computed interaction energies between the ligand and protein, as well as for the ligand and protein strain. No significant correlation was found.

These values plotted against the binding affinity (as $-\log K_i$ or pK_i) yielded no discernible correlation (Figure 11).

When the 60-complex set was subjected to *Q* analysis, a significant correlation ($r^2 = 0.42$) emerged, which was slightly better than the corresponding analysis based on the buried nonpolar surface area (Figure 12). Parsing the analysis using ligand surface polarity yielded an expected result, namely, that nonpolar ligands yielded dQ values which correlated best with observed binding affinity (np/tot > 0.7, $r^2 = 0.69$), and those with essentially polar surfaces were correlated the least (np/tot < 0.5, $r^2 = 0.16$). This is consistent with the hypothesis that the hydrophobic effect plays a significant role in the binding affinity of relatively nonpolar ligands.

Interestingly, an *X*-score evaluation for the same set of nonpolar ligand−protein complexes (np/tot > 0.7) yielded a correlation of $r^2 = 0.66$. [21] These complexes represent a subset of the original 100. In the case of the *X*-score study, they were also part of the training set wherein the scoring paradigm employed the buried hydrophobic surface area as a key term. In the present case, as the *Q* methodology does not rely on parametrization using training sets, the correlation of dQ values to observed binding affinities of $r^2 = 0.69$ essentially reflects a process independent of training set bias or dependency.

Loop Modeling. Yet another extension of the *Q* methodology includes applications involving protein loop prediction. An example of such an application is given in the case

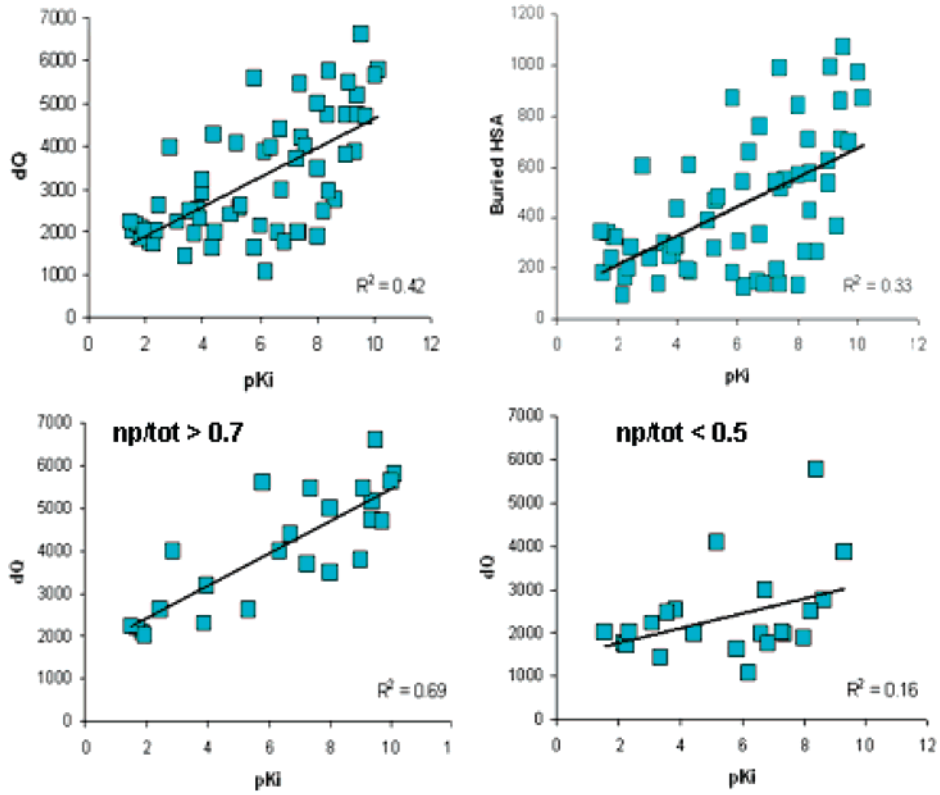

Figure 12. Comparison of dQ (upper left) and buried nonpolar surface area, buried HSA, (upper right) metrics correlated to observed binding affinities. dQ correlation to pK_i was found to be more strongly correlated for ligands with a greater proportion of hydrophobic surface area, as is illustrated in the bottom two graphs.

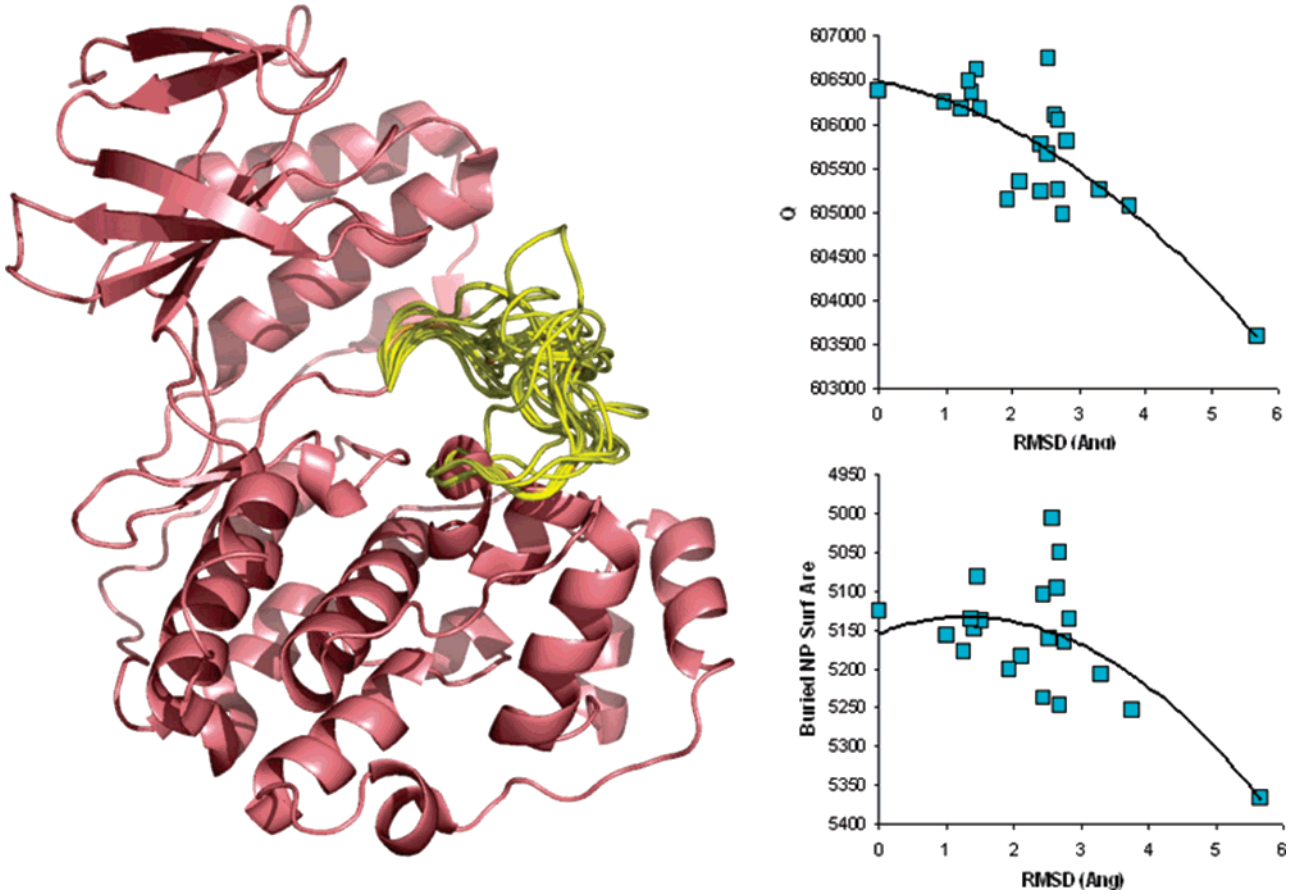

Figure 13. Loop pose analyses using the *Q* approach (upper panel) and buried hydrophobic surfaces (lower panel) applied to 20 poses of the p38 kinase (1DI9) activation loop (residues 170−185).

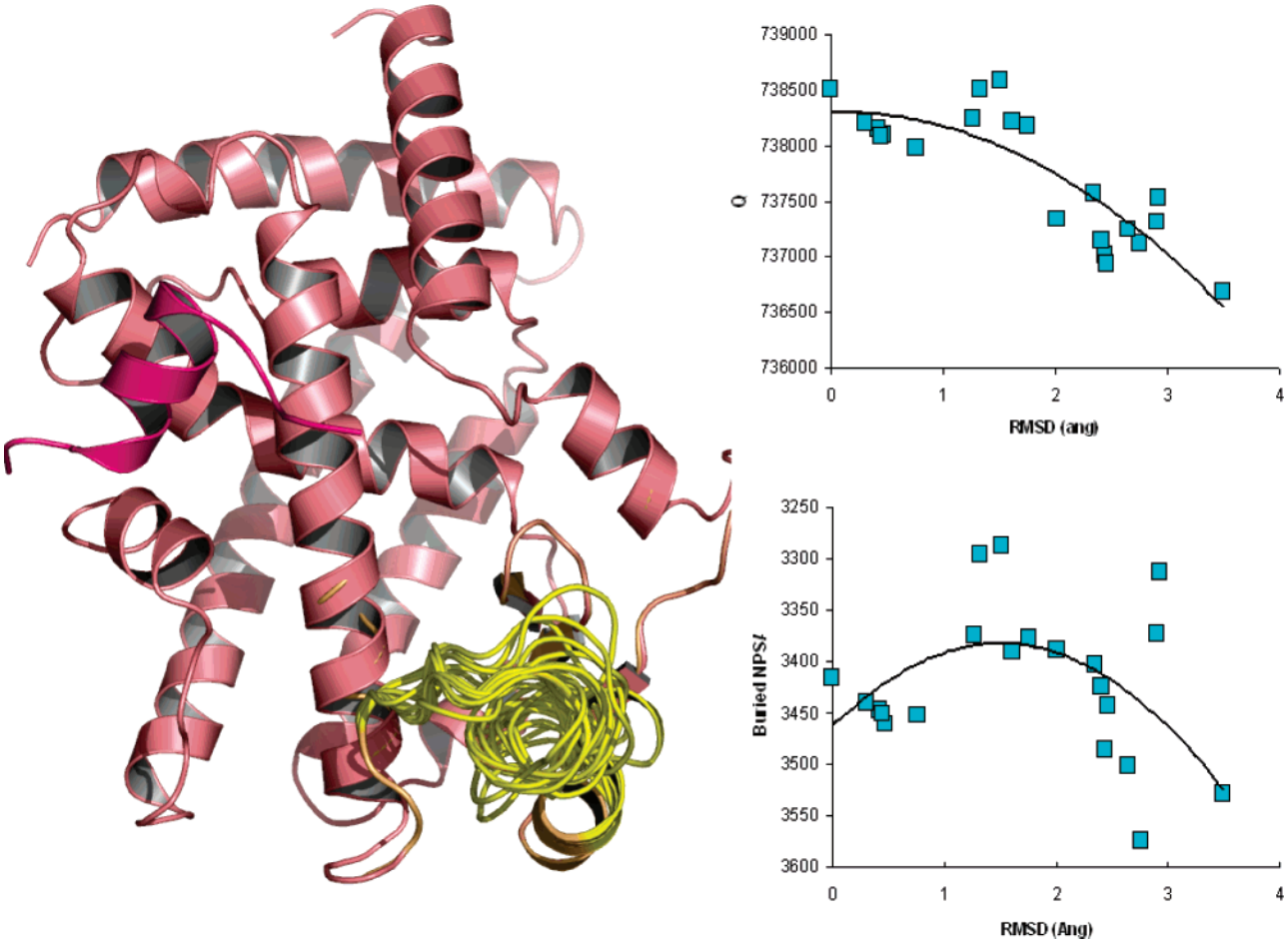

Figure 14. Loop pose analyses using the *Q* approach (upper panel) and buried hydrophobic surfaces (lower panel) applied to 20 poses of the PPARγ (1K74) H2−H3 loop (residues 257−270).

of the p38 activation loop (residues 170−185) using a published structure as a reference (RCSB entry 1DI9). The generation of 20 potential loop configurations was done artificially using a combination of manual deflections of portions of the loop followed by minimization (Flo/qxp), as well as simple molecular dynamics perturbation of the loop followed by minimization. The analysis of the X-ray structure and the 20 artificial loop poses using *Q* and buried hydrophobic surface areas is shown in Figure 13. While the buried hydrophobic surface area approach provides some reasonable trends toward poses with low RMSD errors, there remains a significant high-scoring portion with a RMSD > 2 Å. In the case of *Q* analysis, only one high-scoring pose was > 2 Å, while all other high-scoring poses were at RMSD < 2 Å. Similar results were obtained for a PPARγ (RCSB entry 1K74) example using the H2−H3 loop (residues 257−270) shown in Figure 14.

CONCLUSIONS

We have described a simple method for assessing the contribution of changes in solvent entropy on the ΔG upon host/guest complexation. The approach considers the change in orientational degrees of freedom of a self-associating arbitrary solvent incurred upon the complexation of solute molecules. Solvent points proximal to hydrophobic atoms of the solute lose additional degrees of freedom on the basis of a distance-dependent function intended to replicate the continuum of solvent ordering observed in the first few solvation layers of water upon the introduction of a hydrophobic solute.[3,7,15]

The utility of the method was demonstrated in the selection of poses similar to the crystallographic bound pose and in the ranking of inhibitors according to binding affinity. These results suggest that the use of a simple system such as that described here could make a substantial impact on virtual screening applications. Additionally, the approach shows early promise in the scoring of flexible protein loops. Both of these applications will be the subject of future research.

Qualitatively, the results reported here are similar in many respects to those from buried hydrophobic surface area calculations. As the utility of buried hydrophobic surface area in the description of the hydrophobic effect is well-established in the literature, this outcome is both expected and encouraging. The *Q* approach, however, holds certain distinctions from the buried surface area that are of note. In particular, the method appears less sensitive to the grid spacing than the buried surface area is to probe size. Indeed, contrary to the established use of a water radius of 1.4 Å, we found a probe radius of 0.5 Å was optimal for the calculation of the buried hydrophobic surface area. We do not believe that this has been widely reported in the literature. There have been several extensions to buried surface area approaches that consider the radius of curvature of the solute surface.[36] The approach described here addresses this implicitly. Furthermore, the *Q* methodology has an advantage over surface area for solvent-exposed binding sites (the

hydrophobic "mesa"), in which solvent ordering over several solvation layers likely plays an important role in binding. Finally, the *Q* method is distinct from buried surface area in that it considers change in the solvent entropy from the solvent perspective, as opposed to a solute-based heuristic. It is our opinion that an approach that addresses solvent entropy in such a manner holds greater promise for future improvement in the accurate assessment of free energy changes upon ligand binding.

REFERENCES AND NOTES

(1) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature (London)* **2005**, *437*, 640−647.
(2) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E. Free Energy of Ligand Binding to Protein: Evaluation of the Contribution of Water Molecules by Computational Methods. *Curr. Med. Chem.* **2004**, *11*, 3093−3118.
(3) Dill, K. A.; Truskett, T. M.; Vlachy, V.; Hribar-Lee, B. Modeling Water, the Hydrophobic Effect, and Ion Solvation. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 173−199, 174 plates.
(4) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. A View of the Hydrophobic Effect. *J. Phys. Chem. B* **2002**, *106*, 521−533.
(5) Widom, B.; Bhimalapuram, P.; Koga, K. The Hydrophobic Effect. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3085−3093.
(6) Huang, D. M.; Chandler, D. The Hydrophobic Effect and the Influence of Solute−Solvent Attractions. *J. Phys. Chem. B* **2002**, *106*, 2047−2053.
(7) Jensen, M. O.; Mouritsen, O. G.; Peters, G. H. The Hydrophobic Effect: Molecular Dynamics Simulations of Water Confined between Extended Hydrophobic and Hydrophilic Surfaces. *J. Chem. Phys.* **2004**, *120*, 9729−9744.
(8) Kelly, M. D.; Mancera, R. L. A New Method for Estimating the Importance of Hydrophobic Groups in the Binding Site of a Protein. *J. Med. Chem.* **2005**, *48*, 1069−1078.
(9) Xu, H.; Dill, K. A. Water's Hydrogen Bonds in the Hydrophobic Effect: A Simple Model. *J. Phys. Chem. B* **2005**, *109*, 23611−23617.
(10) Gallagher, K. R.; Sharp, K. A. A New Angle on Heat Capacity Changes in Hydrophobic Solvation. *J. Am. Chem. Soc.* **2003**, *125*, 9853−9860.
(11) Eads, C. D. Lattice Models for Thermodynamics of Complex Aqueous Mixtures. *Abstracts of Papers, 229th ACS National Meeting*, San Diego, CA, March 13−17, 2005; COLL-003.
(12) Imai, T.; Hirata, F. Hydrophobic Effects on Partial Molar Volume. *J. Chem. Phys.* **2005**, *122*, 094509/094501−094509/094506.
(13) Buzano, C.; Pretti, M. Hydrophobic Effect in a Lattice Model of Aqueous Solutions. *J. Chem. Phys.* **2003**, *119*, 3791−3799.
(14) Sitnikov, G. V.; Nechaev, S. K.; Taran, M. D. A Quantitative Mean-Field Theory of the Hydrophobic Effect of Neutral and Charged Molecules of Arbitrary Geometry. *J. Exp. Theor. Phys.* **2005**, *101*, 962−977.
(15) Frank, H. S.; Evans, M. W. Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *J. Chem. Phys.* **1945**, *13*, 507−532.
(16) Wallqvist, A.; Berne, B. J. Molecular Dynamics Study of the Dependence of Water Solvation Free Energy on Solute Curvature and Surface Area. *J. Phys. Chem.* **1995**, *99*, 2885−2892.
(17) Gibb, C. L. D.; Gibb, B. C. Well-Defined, Organic Nanoenvironments in Water: The Hydrophobic Effect Drives a Capsular Assembly. *J. Am. Chem. Soc.* **2004**, *126*, 11408−11409.
(18) Scarsi, M.; Majeux, N.; Caflisch, A. Hydrophobicity at the Surface of Proteins. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 565−575.
(19) *SAVOL*; Optive Research: Austin, TX.
(20) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(21) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.
(22) Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(23) McMartin, C.; Bohacek, R. S. QXP: Powerful, Rapid Computer Algorithms for Structure-Based Drug Design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333−344.
(24) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute-Water Clusters. *J. Chem. Theory Comput.* **2005**, *1*, 1133−1152.
(25) *Maestro*, version. 7.5; Schrödinger, LLC: New York.
(26) *MacroModel*, version 9.1; Schrödinger, LLC.: New York.
(27) *Jaguar*, version 6.5; Schrödinger, LLC.: New York.
(28) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B. A New Model For Calculation of Solvation Free Energies: Correction of Self-Consistent Reaction Field Continuum Dielectric Theory for Short-Range Hydrogen-Bonding Effects. *J. Phys. Chem.* **1996**, *100*, 11775−11788.
(29) Florian, J.; Warshel, A. Langevin Dipoles Model for ab Initio Calculations of Chemical Processes in Solution: Parametrization and Application to Hydration Free Energies of Neutral and Ionic Solutes and Conformational Analysis in Aqueous Solution. *J. Phys. Chem. B* **1997**, *101*, 5583−5595.
(30) Pierotti, R. A. A Scaled Particle Theory of Aqueous and Nonaqueous Solutions. *Chem. Rev.* **1976**, *76*, 717−726.
(31) Stillinger, F. H.; Cotter, M. A. Free Energy in the Presence of Constraint Surfaces. *J. Chem. Phys.* **1971**, *55*, 3449−3458.
(32) Stillinger, F. H. Structure in Aqueous Solutions of Nonpolar Solutes from the Standpoint of Scaled-Particle Theory. *J. Solution Chem.* **1973**, *2*, 141−158.
(33) Floris, F.; Tomasi, J. Evaluation of the Dispersion Contribution to the Solvation Energy. A Simple Computational Model in the Continuum Approximation. *J. Comput. Chem.* **1989**, *10*, 616−627.
(34) Cramer, C. J.; Truhlar, D. G. AM1-SM2 and PM3-SM3 Parameterized SCF Solvation Models for Free Energies in Aqueous Solution. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629−666.
(35) Cramer, C. J.; Truhlar, D. G. PM3-SM3: A General Parameterization for Including Aqueous Solvation Effects in the PM3 Molecular Orbital Model. *J. Comput. Chem.* **1992**, *13*, 1089−1097.
(36) Whitesides, G. M. Reconciling and Magnitude of the Microscopic and Macroscopic Hydrophobic Effects. *Chemtracts: Org. Chem.* **1992**, *5*, 271.

CI600188H