JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Affinity and Specificity of Protein U1A-RNA Complex Formation Based on an Additive Component Free Energy Model

Bethany L. Kormos[1]*, Yulia Benitex[1], Anne M. Baranger[1,2] and David L. Beveridge[1]

[1]*Chemistry Department and Molecular Biophysics Program Wesleyan University Middletown, CT 06459, USA*

[2]*Department of Chemistry University of Illinois at Urbana-Champaign, Urbana IL 61801, USA*

An MM-GBSA computational protocol was used to investigate wild-type U1A-RNA and F56 U1A mutant experimental binding free energies. The trend in mutant binding free energies compared to wild-type is well-reproduced. Following application of a linear-response-like equation to scale the various energy components, the binding free energies agree quantitatively with observed experimental values. Conformational adaptation contributes to the binding free energy for both the protein and the RNA in these systems. Small differences in ΔGs are the result of different and sometimes quite large relative contributions from various energetic components. Residual free energy decomposition indicates differences not only at the site of mutation, but throughout the entire protein. MM-GBSA and *ab initio* calculations performed on model systems suggest that stacking interactions may nearly, but not completely, account for observed differences in mutant binding affinities. This study indicates that there may be different underlying causes of ostensibly similar experimentally observed binding affinities of different mutants, and thus recommends caution in the interpretation of binding affinities and specificities purely by inspection.

Corresponding author

Introduction

The protein U1A is a component of the spliceosome, which excises introns from pre-mRNA prior to gene expression in eukaryotes.[1] Fine control of complex formation between U1A and stem loop 2 (SL2) of U1 snRNA (U1A-RNA) has been studied *via* experiments involving U1A mutants and SL2 base replacements, and observed affinities (ΔG) and specificities ($\Delta\Delta G$) have been reported for the wild-type system and a number of modified forms.[2–24] The results of these experiments have raised interesting and important questions about the nature of U1A-RNA complex formation at the molecular level and the relative contributions of various chemical forces to binding.

U1A binds single-stranded RNA through one of the most common eukaryotic binding domains, the RNA recognition motif (RRM), also known as the RNA binding domain (RBD) or the ribonucleoprotein (RNP) domain.[2] The general RRM structure is comprised of a $\beta\alpha\beta\beta\alpha\beta$ sandwich fold that forms a four-stranded antiparallel β-sheet supported by two α-helices.[3] The N-terminal RRM of U1A binds SL2 of U1 snRNA and two adjacent internal loops in the 3′-untranslated region of its own pre-mRNA with high affinity and specificity.[4–6] All three target sites contain nearly identical sequences in the loop, AUUGCAC closed by a CG base-pair.
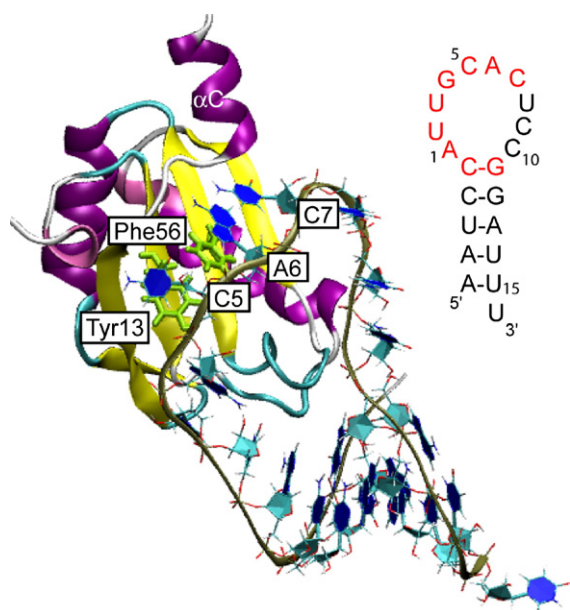
The structure of the N-terminal RRM of U1A (residues 2–117) has been solved by NMR spectroscopy[7] and X-ray crystallography (residues 3–92),[8] and the complex of the N-terminal RRM of U1A (residues 2–97) with SL2 RNA (21 bases) has been

solved by X-ray crystallography,[9] albeit with two surface mutations. The tertiary structure of the U1A-SL2 RNA complex[9] and the secondary structure of SL2 RNA are shown in Figure 1. The C-terminal α-helix (α_C or helix C) of U1A assumes a significantly different orientation free in solution than that in the crystal structure of the complex. α_C is oriented adjacent to the β sheet in the bound form, but lies across the β sheet in the free form (Figure 2), obstructing the approach of the RNA. We refer to the NMR (free) and X-ray (bound) structures of U1A as the "closed" and "open" forms of α_C, respectively. A recent crystal structure of free U1A (residues 6–98) exhibits α_C in the open orientation, which may have been due to the solution conditions[10] or to the shorter construct used in the study compared to the NMR structure.[11] An experimental structure of the free form of SL2 RNA is not available.

Baranger and co-workers investigated the contribution of conserved aromatic amino acids that participate in stacking interactions to the affinity and specificity of the N-terminal RRM of U1A. Based on mutant studies, they found that the highly conserved F56 contributed significantly to binding and that the F56A mutant altered the affinity by more than would be anticipated from a loss in stacking interactions.[12,13] Energetic coupling was found between F56 and A6 of SL2 RNA based on simultaneous modification.[13] Non-polar base isosteres could be substituted for A6 with little loss of binding affinity, evidence that increased hydrophobicity can compensate for the loss of hydrogen-bonding interactions.[14] Taken together, these data indicate an important role for the conserved F56 residue in U1A-RNA binding.
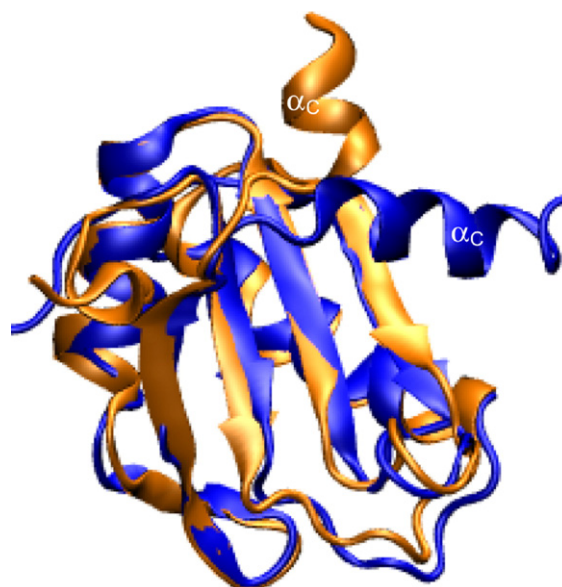
Figure 2. Overlay of the U1A protein in free[7] (blue) and bound[9] (orange, RNA omitted for clarity) forms illustrating the closed and open orientations of α_C, respectively.

Detailed information about chemical forces is not readily obtained from experiment; as such, the interpretation and understanding of observed binding constants involves reference to a theoretical and, in the case of systems of this complexity, a computational model. One such model involves post-processing molecular dynamics (MD) trajectories for a biomolecular complex and its corresponding unbound forms to obtain free energies based on a linearly additive model of terms representing the various chemical forces. The methodology for this genre goes by several names, including the master equation method,[15] end point calculations,[16] free energy component analysis[17] and the MM-PBSA method,[18] which refers to calculations based on a molecular mechanics (MM) configurational energy function combined with Poisson Boltzmann (PB) continuum treatment of solvation with non-polar contributions estimated from solvent accessibility (SA) calculations. A widely used semi-empirical variant of this method utilizes MD trajectories and a Generalized Born (GB) approximation to PB solvation[17,19–21] (MM-GBSA).

The results of MM-PBSA-type and related calculations can provide useful estimates of the contribution of various terms to a binding free energy, but are limited in absolute accuracy due to the propagation of uncertainties in the values of individual contributions. Viable approaches to additive free energy calculations of binding and specificity are due to the laboratories of Honig,[22] Gilson,[23] Sharp,[24] and McCammon,[16] and specific applications to ligand binding have been surveyed by Simonsen *et al.*[25] We have recently reported several studies using component analysis on protein-DNA complexes.[17,19,26] In many cases, trends in binding free

Figure 1. Structure of the N-terminal RRM of U1A bound to SL2 RNA.[9] Inset: nucleic acid sequence of SL2 RNA with nucleotides recognized by U1A for binding highlighted in red.

energies are well represented and a subsequent step of linear regression to rescale the MM-PBSA-type calculated values has proved highly successful.[27,28] A theoretical justification of this has been provided based on linear response (LR) theory and extended linear response methods.[29–31]

A number of theoretical and computational studies on the U1A-RNA system have been reported.[32–45] Reyes and Kollman described MD on U1A binding to both SL2 RNA and internal loop RNA[32] and estimated the unbound forms of both U1A and SL2 RNA to be ~10 kcal/mol lower in energy than the corresponding bound forms, indicating the critical role that adaptation plays on complex formation.[36] The influence of some mutations that both reduce and increase affinity was verified.[35] Olson[37] subsequently reported an MM – non-linear PBSA study on U1A-RNA and provided a semi-quantitative account of specificities for mutants at several different positions. Blakaj *et al.*[38] carried out an early component analysis estimate of $\Delta\Delta G$ for the F56A mutant of U1A-RNA and found that the origin of the large reduction in affinity upon mutation was due to the free energy of the unbound form of U1A.

We describe herein an application of the MM-GBSA-LR computational protocol to wild-type U1A-RNA and the F56A, F56L and F56W mutants of the protein. The corresponding experimental measurements were obtained locally and described in a series of recent articles by Baranger and co-workers.[12,13,46]

Results

MD simulations of U1A-RNA systems

The MD trajectories computed for this study are summarized in Table 1. They include wild-type U1A-RNA complex, free wild-type U1A in closed and open conformations, and free SL2 RNA, as well as those of three U1A mutants, F56A, F56L and F56W. MD simulations were performed with explicit water and counterions. The root-mean-square deviations (RMSD) of the protein backbone residues Phe8–Lys98 and RNA backbone atoms from the equilibrated structures were calculated over the course of the trajectories and are plotted in Figure 3. The N and C-terminal protein residues were omitted from the RMSD analysis because they are very flexible throughout the MD simulations, causing high RMSD values that are not indicative of any significant structural changes of interest in the proteins or protein-RNA complexes.

The RMSD plots of the wild-type U1A-RNA complex and mutant complexes (Figure 3(a)) are all quite stable over the 0.5–3 ns portion of the trajectories for which each was sampled for the MM-PB(GB)SA calculations. Verification of the wild-type U1A-RNA complex trajectory is provided by comparison to previous MD studies on this system.[32–34,38,42] The average RMSD of 1.1 Å is consistent with these prior studies. Subsequent to this study, the U1A-RNA wild-type complex trajectory was extended to 10 ns and remains stable throughout the course of the trajectory. Thus, results based on the 3 ns trajectory are expected to be very similar to those that would be obtained from a longer simulation. The average RMSD of the F56A mutant complex from the initial structure is 1.4 Å, larger than wild-type and the other mutants (F56L = 1.1 Å, F56W = 1.2 Å). This may indicate loss of stability of the F56A complex, which has been shown experimentally.[47,48] The free U1A simulations in closed form were run for 5 ns. The RMSD plots of the free wild-type, F56L and F56W U1A proteins in the closed conformation stabilize after 2 ns, while the F56A mutant takes 3 ns to stabilize (Figure 3(b)). Comparison to a previous MD study on this system indicates good agreement[40] with an average RMSD value of 2.1 Å. Comparing the average distance between the center of mass of helix C and the center

Table 1. MD simulations performed

	Simulation length (ns)	Portion of simulation selected for analysis (ns)	No. of snapshots for MM-PB(GB)SA/NMode analysis
Free protein			
Wild-type U1A (closed)	5	2–5	150/150
F56A U1A (closed)	5	3–5	100/100
F56L U1A (closed)	5	2–5	150/150
F56W U1A (closed)	5	2–5	150/150
Wild-type U1A (open)	5	3–5	100/100
F56A U1A (open)	3	0.5–3	125/125
F56L U1A (open)	3	0.5–3	125/125
F56W U1A (open)	3	0.5–3	125/125
Free RNA			
Stem loop 2 RNA	32	8–32	1200/120
Protein-RNA complex			
Wild-type U1A-RNA	3	0.5–3	125/125
F56A U1A-RNA	3	0.5–3	125/125
F56L U1A-RNA	3	0.5–3	125/125
F56W U1A-RNA	3	0.5–3	125/125
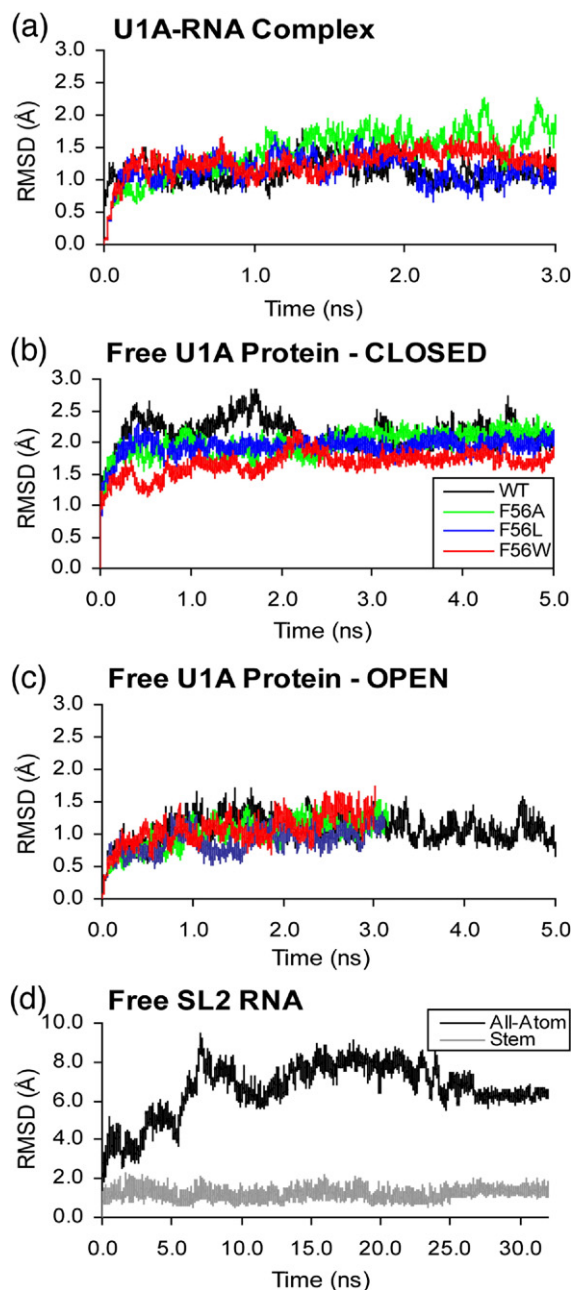
Figure 3. Plots of the RMSD of protein backbone residues Phe8–Lys98 and RNA backbone atoms from the equilibrated structures calculated over the course of the MD trajectories. (a) U1A-RNA complex. (b) Free U1A protein with helix C in the closed orientation. (c) Free U1A protein with helix C in the open orientation. (d) Free SL2 RNA.

of mass of the β-sheet, wild-type and F56W U1A have similar distances (17.1 Å and 16.5 Å, respectively), while the F56A and F56L are shorter (15.1 Å and 15.7 Å, respectively). This indicates that helix C moves to a new space during the MD simulations in the F56A and F56L mutants, which may affect binding.

The free U1A simulations in open form were more stable than those in the closed form, so were run for

3 ns. The simulation of free wild-type U1A in the open form was run for 5 ns to confirm the stability of the open form. The RMSD plots of the free wild-type and F56 mutant U1A proteins in the open conformation stabilize between 0.5–1 ns (Figure 3(c)). Comparison to a previous MD study on this system indicates good agreement[40] with an average RMSD value of 1.1 Å.

The 32 ns simulation of SL2 RNA appears to be the longest MD trajectory performed to date on this system. The all-atom RMSD plot of the free SL2 RNA stabilizes at 8 ns (Figure 3(d)). It can be seen from the plot that the stem residues are quite stable throughout the duration of the simulation and that the flexible loop residues appear to evolve through two substates with average RMSDs of 3.5 Å and 5 Å before reaching a more stable conformation with an average RMSD of 7.5 Å. Previous MD simulations on SL2 RNA[32,33,36,40] have been run between 0.6–5 ns, thus would not have observed these substates. However, an average RMSD of 3.5 Å for the first substate agrees quite well with the results from these previous studies. Additional analysis of this simulation including a complete analysis of the ion and hydration atmospheres will be reported elsewhere.

Analysis of wild-type and F56 mutant U1A-RNA binding free energies using MM-PB(GB)SA methods

MM-PB(GB)SA component analysis was performed on snapshots extracted from the MD simulations stripped of water and ions, then averaged over the snapshots to obtain free energies of binding as described in Materials and Methods. The "separate trajectory" method involves analysis of three different trajectories to obtain the overall free energy of binding: the U1A-RNA complex, the free U1A protein (in either closed or open form) and the free SL2 RNA. The "single trajectory" method obtains all snapshots from the MD simulation of the complex, and obtains the free protein and RNA contributions by separating the complex into its constituent parts. This method is less computationally intensive than the separate trajectory method, but neglects effects from structural adaptation. Solute free energy contributions include the MM energy computed by AMBER and entropy contributions computed using normal mode analysis. Solvent free energy contributions include the electrostatic component computed using PB or GB implicit solvent methods and the non-polar component based on the molecular solvent-accessible surface area (SASA).

MM-GBSA analysis by end points quantitatively agrees with observed U1A-RNA binding free energies after application of a linear-response-type analysis

The snapshots from the MD trajectories selected for MM-PB(GB)SA analysis were determined based

on stabilization of the RMSD plots (Figure 3) and are reported in Table 1. MM-PB(GB)SA binding free energies were calculated as described in Materials and Methods. The electrostatic contribution to the solvation free energy (G^{es} in equation (8)) was calculated using PB as implemented in the *pbsa* program in AMBER 8 as well as a number of GB methods (see Materials and Methods). Upon comparison of the computed binding free energies to the experimental values, the GB^{OBC}(I) method[49,50] provided the most reasonable overall binding free energies in conjunction with normal mode analysis. The correct trends for the binding free energies of mutants are observed by all of the GB methods, however GB^{OBC}(I) performs best, even compared to PB (results not shown). As such, analysis throughout the remainder of this article pertains to the results obtained using the GB^{OBC}(I) method.

Computed MM-GBSA binding free energies for wild-type and F56 mutant U1A-RNA systems using both the single trajectory and separate trajectory methods are reported in Table 2. Numbers are reported using either the closed or open form of free U1A protein in the separate trajectory method. ΔGs relative to wild-type are indicated in parentheses for each system. In all cases, the computed free energies of binding for the U1A mutants to SL2 RNA follow the trend of the experimental binding free energies compared to the wild-type system. However, the binding free energies are overestimated by MM-GBSA, a common result that has been seen in other studies.[20,36] The standard errors (σ) reported in Table 2 are comparable to those seen in other similar MM-GBSA studies[20,36] and are generally smaller than the $\Delta\Delta G$s for the mutants compared to the wild-type system. This indicates that the trend in binding free energies compared to the experimental data is upheld upon taking into account the standard error for each system.

The differences in the binding free energies between the single trajectory and separate trajectory methods stems from conformational adaptation. The free energies of the bound conformations of U1A protein and SL2 RNA from the single trajectory method are expected to be higher than those of the free conformations from the separate trajectories. As such, the free energies of binding computed from separate trajectories should be smaller than those calculated from the single trajectory

method. This is indeed observed in the binding free energies in Table 2.

Comparing the binding free energies of the closed and open forms of U1A obtained from the separate trajectory method, ΔG for the closed form is more negative for the wild-type and all of the mutants. In general, the trend in binding free energies compared to the experimental values is better for the open form of U1A primarily because of the ΔG for the F56L mutant. However, in terms of absolute binding free energies, the F56A and F56L mutants bind very weakly in the open form and, in fact, a positive ΔG indicates that the F56A mutant should not bind SL2 RNA. For these reasons, and because the NMR structure of the free U1A protein in solution is the closed form, we chose to base the remaining analysis on the closed form of U1A.

The trend in the binding free energies of wild-type and mutant U1A-RNA was reproduced very well by MM-GBSA compared to experiment, though the absolute binding free energies were overestimated. As MM-GBSA is empirical in nature, the various energy components may be scaled to obtain more accurate estimates of free energies. The four systems in this study have experimentally known binding free energies, which were used to train the data set. The solute and solvent electrostatic energy components tend to dominate the energy components, so the energy components were factored into electrostatic terms and non-electrostatic terms, as shown in equations (11) and (12). Multiple regression analysis was then applied to optimize the α, β and γ parameters to derive a LR-like equation:

$$\Delta G = 0.2796(\Delta ES) + 0.25837(\Delta NES) + 2.65244 \quad (1)$$

The optimal amount by which the ES and NES terms should be scaled are quite similar (0.2796 and 0.25837, respectively), suggesting that multiplication by a constant may have been a sufficient choice. However, several different methods were explored, including different numbers of parameters, and this equation gave the best fit and lowest error compared to the experimental data.

The results from the LR-type analysis in equation (1) are shown in the LR column in Table 2 and reproduce the experimental binding free energies quite well. The binding free energy predicted for the F56L U1A mutant with SL2 RNA remains more favorable

Table 2. Calculated average binding free energy (kcal/mol)

| | Single trajectory | | Separate trajectory | | | | | | |
| | | | U1A closed | | | U1A open | | | |
	GB^{OBC}(I)[a]	σ[b]	GB^{OBC}(I)[a]	σ[b]	LR[a]	GB^{OBC}(I)[a]	σ[b]	MP2/6-311G(d,p)[a]	EXPT[c]
Wild-type	−62.1 (0.0)	3.7	−57.7 (0.0)	3.5	−12.3 (0.0)	−23.1 (0.0)	3.4	1.7 (0.0)	−12.7±0.3 (0.0)
F56A	−43.8 (18.3)	4.0	−37.4 (20.3)	3.8	−6.8 (5.5)	2.7 (25.9)	3.5	4.0 (2.4)	−7.2±0.3 (5.5)
F56L	−52.5 (9.6)	3.7	−52.3 (5.4)	3.6	−10.7 (1.6)	−0.4 (22.7)	3.3	2.4 (0.7)	−8.6±0.3 (4.1)
F56W	−68.9 (−6.8)	3.6	−55.2 (2.4)	3.7	−11.4 (0.9)	−19.4 (3.8)	3.3	2.4 (0.7)	−12.7±0.3 (0.0)

[a] $\Delta\Delta G$s relative to wild-type are reported in parentheses.
[b] Standard error of mean values.
[c] Experimental binding free energies determined by Baranger and co-workers.[12,13]

than is seen in the experimental results; however, the trend in the mutant binding free energies compared to wild-type remains correct compared to the experimental results. The α, β and γ parameter values derived here may be applicable to future MM-GBSA calculations on F56 U1A mutants.

MM-GBSA analysis by chemical forces indicates different underlying causes contributing to experimentally observed binding free energies of different F56 U1A mutants

MM-GBSA-LR provided excellent results computing binding free energies for wild-type and several F56 U1A mutants with SL2 RNA compared to the experimental results. The advantage to using an energy component technique such as MM-GBSA-LR is that the resultant data can be further analyzed to obtain information that cannot be obtained from the experimental data. One such technique involves analysis of the contributions of the various energy components for the systems of interest.

The analysis in Figure 4 is based on the separate trajectory method with the free U1A protein in the closed conformation. The individual energy components contributing to the overall binding free energy of U1A-RNA are shown in Figure 4(a). The overall free energy of binding is dominated by the solute and solvent electrostatic components. These components are of opposite sign and nearly equal value for each of the systems. This indicates that differences in the solute electrostatics in the systems are generally compensated for by equal and opposite differences in the electrostatic contribution to the solvation free energy. Other components may, of course, contribute to this compensation, as well. In addition, this analysis indicates that very large contributions from the different energy components sum to an overall small binding free energy for wild-type U1A-RNA and the F56 mutants, indicated by the TOTAL label in Figure 4(a).

In Figure 4(a), the component contributions appear to have quite similar values between wild-type and the F56 mutants, due to the large energy values of the electrostatic component contributions. This issue was remedied by plotting the same data of the F56 mutants relative to wild-type (Figure 4(b)). Figure 4(b) provides a clearer picture of the contributions to the overall binding free energy of the F56 U1A-RNA mutants compared to the wild-type system. The F56A U1A mutant has the weakest binding free energy for SL2 RNA compared to the other mutants. According to Figure 4(b), this poor binding is due to less favorable solute ELE and VDW contributions compared to wild-type that cannot be compensated for by a slightly more favorable VAL and GB-ES. The F56L U1A-RNA mutant also has poor binding compared to wild-type. Figure 4(b) indicates that the energy components for the F56L U1A-RNA mutant are for the most part quite similar to wild-type. However, an unfavorable contribution from GB-ES cannot be completely compensated for by a
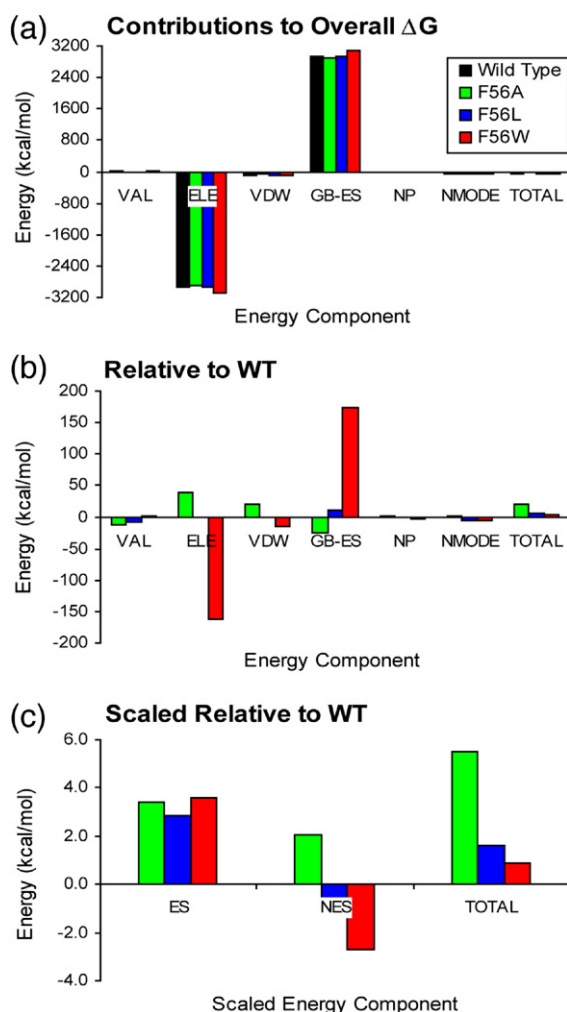
Figure 4. Energy components (kcal/mol) contributing to the overall free energies of binding for wild-type and F56 mutant U1A-RNA systems. VAL, ELE and VDW are the solute valence, electronic and van der Waals energy contributions, respectively; GB-ES and NP are the electrostatic and non-polar contributions to the solvation free energy, respectively; NMODE is the entropy contribution to the free energy; TOTAL is the overall free energy from the summation of the energy components. (a) Absolute energies for wild-type, F56A, F56L and F56W systems. (b) Relative energies of F56 mutants compared to wild-type. (c) Electrostatic and non-electrostatic energy contributions of F56 mutants compared to wild-type scaled according to equation (10).

favorable contribution from the VAL component compared to wild-type. The F56W U1A-RNA mutant has the largest differences in energy components compared to the wild-type system, yet has the most similar overall binding free energy. The unfavorable GB-ES component contribution almost perfectly compensates for the favorable ELE and VDW contributions.

Figure 4(c) reports the electrostatic and non-electrostatic energy contributions as well as the overall free energies of binding for the F56 mutants compared to the wild-type system. The electrostatic and non-electrostatic energy components were scaled

using the α and β parameters, respectively, from equation (1). This graph indicates that the F56A system has the lowest binding free energy of all of the mutants because both the ES and NES contributions are unfavorable compared to wild-type. This is in contrast to the F56L and F56W systems, which indicate unfavorable ES contributions that are compensated for by favorable NES contributions compared to the wild-type system. For the F56W system, the ES and NES contributions are nearly equal in magnitude and opposite in sign, whereas for the F56L system, the favorable NES contribution is not large enough to compensate completely for the unfavorable ES contribution.

Figure 5 plots the energy components of SL2 RNA from the single trajectory method relative to wild-type. This enables comparison of the SL2 RNA energetics when bound to wild-type and F56 U1A mutants. Though these energies cannot be scaled by the optimized parameters in equation (1) because they are not ΔGs, they can still contribute to an understanding of differences being imposed on the RNA structure by the mutation at position 56 in U1A. As with the energy components contributing to the overall ΔGs (Figure 4), the most significant differences are in the ELE and GB-ES components compared to wild-type, which compensate for each other with nearly equal and opposite values. However, SL2 RNA bound to F56A U1A has the opposite trend compared to wild-type U1A than when it is bound to F56L or F56W. The VAL energy component is unfavorable for SL2 RNA bound to all of the mutant proteins compared to wild-type, while the VDW energy component is favorable. Overall, SL2 RNA bound to each of the mutant proteins is unfavorable energetically compared to wild-type, as indicated by the results in the TOTAL column.

MM-GBSA analysis on a residual basis indicates energetic differences throughout F56 mutant systems rather than localized at the site of mutation

Energy decomposition on a per residue basis was performed for the wild-type and F56 mutant free
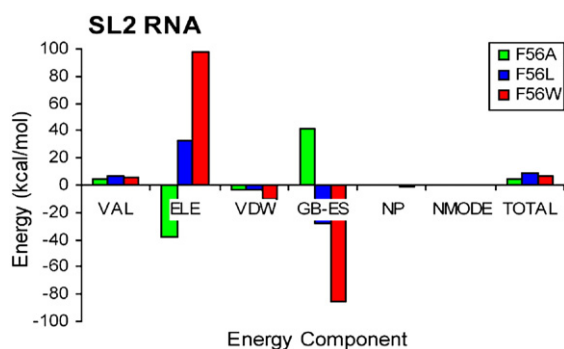
U1A proteins in the closed form, free SL2 RNA and U1A-RNA complexes. The ES and NES terms for each residue were collected in both the bound and free forms, followed by scaling based on equation (1) and calculation of ΔG for each residue. The entropy term complicates matters for this analysis, as it is a collective property of the entire system and cannot be decomposed on a residual basis as may the other energy components. Entropy was, however, originally included in the linear regression analysis that was used to obtain the parameters in equation (1). In this computation, the approximation is made that it is acceptable to use the α, β and γ parameters from equation (1) to scale the residual ES and NES terms, since change in entropy was similar for all of the systems.

Figure 6 plots the scaled energy decomposition on a residual basis for the F56 mutant systems relative to wild-type. Differences in ΔΔG greater than ±0.5 kcal/mol are highlighted in blue. Here, we concentrate mainly on the residues between Thr11–Lys98, as the N and C-terminal residues (residues Ala2–His10 and Gly99–Val102, respectively) are unstructured and quite mobile in the simulations and are thus expected to vary compared to wild-type. As may be seen from the plots, the energy differences are distributed over the entire protein as well as the RNA. The site of mutation, position 56, indicates a significant difference in ΔG compared to wild-type for all three mutant systems studied. However, this energy difference is unfavorable for the F56A and F56L mutants, which have unfavorable binding compared to wild-type, while the energy difference from the F56W mutant at this position is favorable compared to wild-type. The majority of the significant perturbations occur in the loop 3 – β3 and loop 6 – helix C regions of the protein, and the nucleotides involved in binding in the RNA loop region. In addition, the differences in ΔG for the RNA nucleotides are generally unfavorable in the mutant systems.

Ab initio interaction energetics estimate the extent to which stacking energetics contribute to U1A-RNA binding free energies

Ab initio calculations were performed on U1A-RNA model systems obtained from the average MD structures at the MP2/6-311G(d,p) level of theory to obtain an estimate of the contribution of stacking interactions in U1A-RNA binding. The results of these calculations can be found in Table 2. The F56A mutant has the weakest interaction energy compared to wild-type (2.3 kcal/mol), which may be expected due to the loss of stacking interactions. However, the F56L and F56W mutants have very similar interaction energies compared to wild-type (0.7 kcal/mol), which is interesting since the Trp mutant maintains stacking while the Leu mutant does not. Hydrophobic interactions may help to compensate for the loss of stacking for the Leu mutant. These interaction energies may provide an estimate of the degree to which a loss of
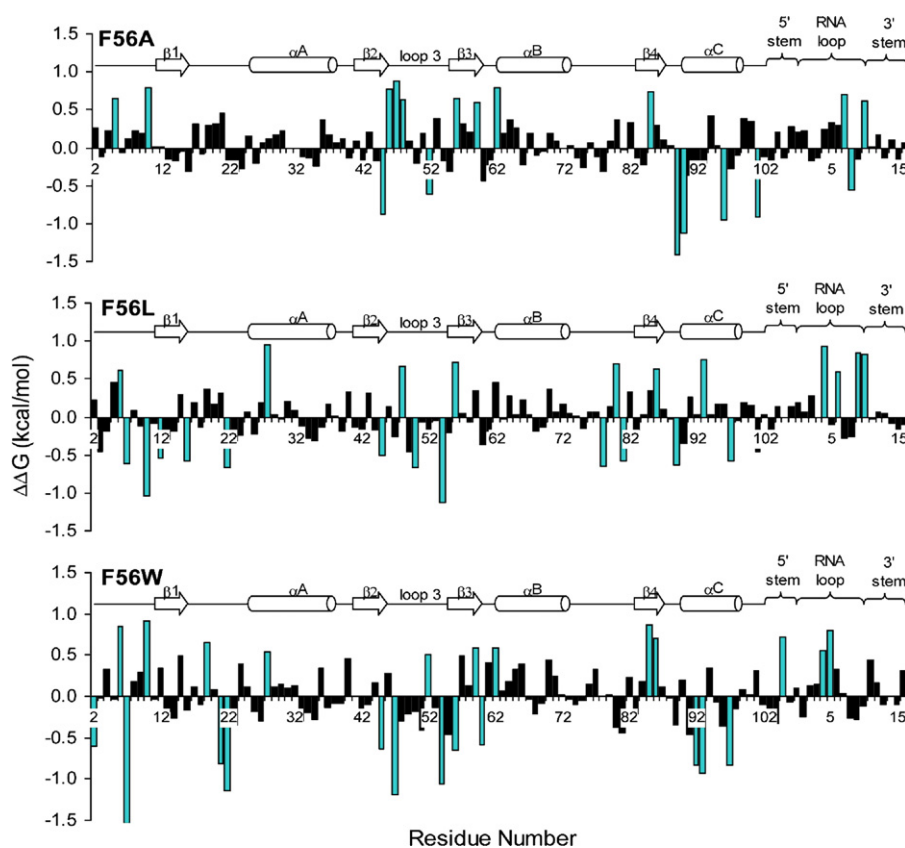

Figure 5. Unscaled energy components (kcal/mol) contributing to the free energy of SL2 RNA bound to F56 U1A mutants relative to wild-type obtained from the single trajectory method. Energy components are defined in the legend to Figure 4 and in the text.

Figure 6. Per residue free energy decomposition of F56 mutants compared to wild-type ($\Delta\Delta G$, kcal/mol) scaled according to equation (10). The secondary structure of U1A residues (2–102) and SL2 RNA (−5 to 16) is indicated. Residues with $\Delta\Delta G$ values ±0.5 kcal/mol compared to wild-type are highlighted in blue. The truncated value for Arg7 in the F56W mutant is −1.55 kcal/mol.

stacking interactions affects binding in the F56 U1A mutants.

Discussion

MM-PBSA and related techniques involve post-processing of MD trajectories to obtain free energy estimates by summing gas phase (MM) and solvent (PB/GB and SA) energetic contributions. Such calculations are pairwise additive, and recent studies have proposed a link between dynamical structure and functional energetics *via* detailed energy breakdowns into contributions from solute and solvent.[17,20,21] The advantage of this class of free energy calculations is the rapidity with which estimates can be obtained on a wide range of individual cases. There are, however, problems to be contended with as well. First and foremost is that additivity *per se* is an approximation: free energies are in general not additive, since the entropic contribution is a collective property of an entire system.[51,52] This is inherent in describing MM-PBSA as "phenomenological," i.e. a method postulated on the basis of plausible intuition as opposed to something rigorously derived from a full statistical mechanics partition function. In addition, calculations of this genre are susceptible to both errors of omission and commission. Errors of omission involve the neglect of particular energetic contributions, such as

counterion effects and configurational entropy in this study, while errors of commission refer to uncertainties in the estimates of terms included in a model, i.e. the use of a constant dielectric in calculating the solvation free energy or the use of a harmonic potential in the normal mode calculation.

The MM-GBSA adaptation energies, detailed in Table 3, are of considerable interest in the U1A-RNA system due to the conformational changes that take place in both the U1A protein and SL2 RNA. The adaptation energies indicate that free U1A in the open form and free SL2 RNA are less stable in their bound forms than in their free forms, while the same is not true for U1A in the closed form. MM-GBSA also predicts that the open form of U1A is energetically more stable than the closed form, which would suggest the structure of the free 2–102 U1A construct would be in an open form instead of a closed form. These discrepancies may be due to the use of an implicit solvent[20] or the neglect of certain terms. In addition, they may be due to the difference in the length of the U1A protein constructs used. The experimental NMR structure is comprised of residues 2–117,[7] while the construct used in this study and the experimental binding studies of the F56 mutants is comprised of residues 2–102.[12,13,46] It has been observed by NMR that helix C of the 2–102 U1A construct is more dynamic than that of the 2–117 construct,[11] and that these motions are on the

Table 3. Conformational adaptation energies (kcal/mol)

	Single trajectory		Separate trajectory		Adaptation energy
		σ[a]		σ[a]	
Free U1A (closed)					
Wild-type	−4216.39	2.30	−4200.80	2.09	−15.59
F56A	−4206.42	2.37	−4188.34	2.89	−18.08
F56L	−4238.23	2.45	−4209.61	2.28	−28.62
F56W	−4211.28	2.23	−4197.67	2.54	−13.61
Free U1A (open)					
Wild-type	−4216.39	2.30	−4235.34	2.24	18.95
F56A	−4206.42	2.37	−4228.46	2.09	22.04
F56L	−4238.23	2.45	−4261.49	2.05	23.26
F56W	−4211.28	2.23	−4233.51	2.08	22.23
Free SL2 RNA					
Wild-type	−4608.26	1.47	−4628.29	0.58	20.03
F56A	−4603.79	1.68	−4628.29	0.58	24.50
F56L	−4599.49	1.46	−4628.29	0.58	28.80
F56W	−4601.03	1.44	−4628.29	0.58	27.26

[a] Standard error of mean values.

micro-to millisecond timescale, a longer timescale than can be observed using MD. Dynamical motions of helix C on a 2–3 ns time scale have also been reported by time-resolved fluorescence anisotropy studies (Baranger, Knee *et al.*, unpublished results) NMR,[11] and MD.[40] These short timescale motions may contribute to dynamical processes on longer timescales. The effect of the length of U1A construct on the dynamics of helix C may be indicated by a crystal structure of the 98 residue construct of U1A reported with helix C in the open position.[10] No interconversion between the closed and open forms of free U1A has been observed on the timescale of our simulations, and there is currently no estimate of the energy barrier for interconversion between the two forms.

Mutation of F56 in U1A to Ala or Leu results in a 5.5 or 4.1 kcal/mol loss in binding free energy compared to wild-type, which is a much more detrimental effect on binding than mutation to Trp. Since Trp retains the aromaticity lost upon mutation of Phe, while Ala and Leu do not, it is an obvious assumption to attribute the loss in binding free energy with the Ala and Leu mutants to the loss of stacking interactions upon binding SL2 RNA. However, previous studies have indicated smaller interactions between aromatic amino acids and nucleobases, on a scale of 1–3 kcal/mol.[53–59] MP2/6-311G(d,p) interaction energies calculated for model systems of these F56 U1A mutants derived from MD average structures suggest that a loss of stacking is worth 2.3 and 0.7 kcal/mol in the F56A and F56L systems, respectively (hydrophobic interactions in the Leu mutant may partially compensate for loss of the aromatic interaction). These estimates suggest a similar contribution from stacking in this system compared to earlier estimates that do not completely account for the differences in binding free energies. In contrast, stacking is maintained by the Trp mutant, yet it has an interaction energy

difference similar to the Leu mutant compared to wild-type.

If the experimental binding free energies in Table 2 could completely be accounted for by the stacking interactions between the residue at position 56 and A6 of SL2 RNA, then it might be expected that largest difference in the MM-GBSA component analysis of the F56 mutants would be in the VDW energy component.[60] This, however, is not the case. Figure 4(b) shows that, compared to wild-type, the VDW component is unfavorable for F56A, favorable for F56W and very similar for F56L. This result supports the idea that, though stacking contributes to the differences in the binding free energies in the mutants, it is not the only factor. Further analysis of Figure 4(b) and (c) indicates that there are significant differences between the wild-type and F56 mutants in most of the energy components, leading to a complex view of the overall contributions to binding in this system. Though the overall binding free energy is similar for all of the systems, the mutants obtain the overall binding free energies by different mixtures of the various energy terms. This is not necessarily obvious upon inspection of the experimental binding free energies, and indicates different contributing factors to binding for the different F56 mutants.

Inspection of the residual free energies of the F56 mutants compared to wild-type (Figure 6) indicates that energetic differences are not localized to the site of the mutation. Rather, the single mutation at position 56 creates differences in residual free energies that are distributed throughout the mutant proteins. This also supports the idea that stacking is not the only factor contributing to the differences in the binding free energies of the F56 mutants. We do note, however, that the main significant differences in the residual free energies are in regions of the protein that have been shown to be involved in conformational adaptation[61,62] and in regions important for binding (loop 3 – β3 and loop 6 – helix C).[2] There is also significant compensation in the free energies of the residues in the protein, in that significant unfavorable contributions at sites in the protein are compensated for by favorable contributions at other sites.

Analysis of the SL2 RNA energetics also lends support to the idea that stacking is not the sole contributor to differences in the experimental binding free energies in the F56 mutant systems. Figure 5 indicates that energetics of SL2 RNA are less favorable when bound to each of the F56 mutant U1A proteins compared to the wild-type system. However, SL2 RNA bound to F56L and F56W mutants is more unfavorable than when bound to the F56A mutant. If stacking was the sole contributor to the differential binding free energies of the mutants, it would be expected that the energy of SL2 RNA bound to F56A U1A would be the least favorable compared to wild-type. This idea can be explored further by analysis of the energy decomposition of SL2 RNA on a nucleotide basis (Figure 6). If stacking was the only contributor to the differences in binding

free energies of the F56 U1A mutants, it would be expected that a large unfavorable ΔG compared to wild-type would appear at the A6 nucleotide (which stacks with Phe56 in wild-type U1A) for the Ala and Leu mutants, but not for the Trp mutant. Though all of the mutants do have unfavorable ΔGs compared to wild-type at the A6 position, unfavorable ΔGs are not localized specifically to that position, and are seen throughout most of the RNA.

Insight into U1A-RNA binding can be gained from additional analysis of SL2 RNA in these systems. Figure 5 indicates that SL2 RNA bound to the F56A mutant is more favorable than when bound to the F56L and F56W mutants. Since the F56A mutant has the weakest binding free energy compared to the mutants, this result suggests that the poor binding free energy of the F56A system is likely due more to differences in the protein than to differences in the RNA upon binding, which has been suggested previously.[38] Energy decomposition of SL2 RNA on a nucleotide basis in the mutant systems (Figure 6) indicates that the majority of nucleotides have unfavorable differences in ΔG compared to wild-type. This observation corresponds to the data in Table 3, which show that the adaptation energy for SL2 RNA is higher upon binding to the F56 U1A mutants than it is upon binding to wild-type. Overall, these results indicate that SL2 RNA binding to wild-type is more favorable than binding to any of the F56 mutants.

Conclusions

MM-GBSA was used in conjunction with normal mode analysis to successfully account for experimental binding free energies of the U1A protein for SL2 RNA for wild-type and several F56 protein mutants. Average energies were obtained from samples of snapshots extracted from MD simulations. The trend in the computed binding free energies for the U1A-RNA mutants was in agreement with the experimental values, though absolute binding free energies were overestimated. A linear-response-like equation was used to scale the various energy components to obtain binding free energies in agreement with the experimental values. *Ab initio* calculations were performed on model systems to estimate contributions of stacking to binding.

Differences in binding free energies between the single and separate trajectory methods indicate that conformational adaptation is important. Individual component analysis provides insight into the different overall free energies of binding for the F56 mutants. Though overall binding free energies of the F56 U1A mutants are very similar, they arise from different combinations of favorable and unfavorable energetic component contributions. Free energy decomposition on a residual basis indicates differences throughout the entire protein, not only at the site of mutation, and especially in areas of the protein shown to be important for binding. MM-GBSA analysis and *ab initio* calculations indicate that

differences in mutant binding free energies are not completely accounted for by differences in stacking interactions. Taken together, these results indicate that there may be different underlying causes of ostensibly similar experimentally observed binding affinities of different mutants, and thus recommend caution in the interpretation of binding affinities and specificities purely by inspection.

Materials and Methods

MD simulations

The starting structure for the MD simulation of the wild-type U1A-RNA complex was based on the X-ray cocrystal structure of the N-terminal RRM of U1A bound to SL2 of U1 snRNA solved at 1.92 Å resolution,[9] PDB ID:[63] 1URN. Biological unit 2 was chosen for the initial structure as it contains the most complete structural information for SL2 RNA. The U1A protein was extended from the crystal structure construct to obtain a structure containing residues 2–102 to match that used in the binding affinity experiments, and two point mutations (H31Y and R36Q) were introduced to revert the protein to the wild-type sequence. Details of these procedures have been described.[40] Point mutations were introduced at the Phe56 position in this structure to create the F56A, F56L and F56W mutant complexes. This extended U1A-RNA complex was then separated into its constituent protein and nucleic acid parts, which served as starting structures for the MD simulations of wild-type U1A in the open form and the free SL2 RNA, respectively. Point mutations were introduced at the Phe56 position in this U1A structure to create the F56A, F56L and F56W mutants of free U1A in the open conformation. The free SL2 RNA used in the previously reported biochemical studies was a 25mer;[12,13,46] however, the SL2 RNA in the X-ray cocrystal structure of the complex and in other MD studies was a 21mer (Figure 1). As previous studies have shown that the shorter SL2 RNA stem does not affect U1A binding,[53] the 21mer RNA was used in the calculations described here.

The starting structure for the MD simulation of the wild-type U1A in the closed form was based on the NMR solution structure of the unbound protein,[7] PDB ID:[63] 1FHT. Model 5 was selected as the representative structure based on analysis of the 43 NMR structures by *NMRCLUST 2.1*.[64] The U1A NMR structure was truncated from residues 2–117 to residues 2–102 to match the construct used in the binding affinity experiments. Point mutations were introduced at the Phe56 position in this structure to create the F56A, F56L and F56W mutant forms of free U1A in the closed conformation.

MD simulations were performed using the AMBER[65,66] suite of programs with the parm96[67] force field. The systems were solvated in a box of explicit TIP3P[68] water molecules that extended a minimum distance of 12 Å from the solute atoms. Neutralizing Na$^+$ or Cl$^-$ ions were added to each system, after which Na$^+$ and Cl$^-$ ions were added to each system to obtain a salt concentration of 250 mM, in accordance with the conditions of the binding affinity experiments. The ions were randomized around the solute such that the ions could be no closer than 5 Å to the solute and no closer than 3 Å to each other. Additional details of the minimization and MD procedures were described by Pitici *et al.*[40] The results pre-

sented here are based on a production runs of 3 ns to 32 ns (see Table 1).

Methodology

Free energies of binding were estimated using MM-PBSA or MM-GBSA in AMBER 8[66] according the methods described by Gohlke and Case,[20] using an additive free energy component analysis approach. This analysis follows a similar protocol established by Beveridge, Baranger *et al.* in studies of the U1A-RNA complex,[38,40,43] the EcoRI endonuclease complex,[17] the λ repressor-operator[26] and a number of protein-DNA complexes.[19]

The overall objective of this method is to calculate the absolute binding free energy for the non-covalent association of any two molecules, A and B, in solution, i.e.:

$$[A]_{\text{aq}} + [B]_{\text{aq}} \Leftrightarrow [A^*B^*]_{\text{aq}^*} \qquad (2)$$
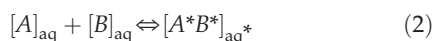
where $[A]_{\text{aq}}$ refers to the dynamical structure of molecule A free in solution, $[B]_{\text{aq}}$ refers to the dynamical structure of molecule B free in solution, and $[A^*B^*]_{\text{aq}^*}$ represents the complex formed from molecules A and B, taking into account any structural changes ($[A^*]$ and $[B^*]$) and solvent reorganization (aq^*) that may occur upon complex formation.

In principle, this calculation requires three MD simulations, one each on $[A]_{\text{aq}}$, $[B]_{\text{aq}}$, and $[A^*B^*]_{\text{aq}^*}$. Gohlke and Case[20] refer to this protocol as "separate trajectories" or "three trajectories" (3T). One may also make the (questionable) approximation that structural adaptation is negligible and draw the trajectories for $[A]_{\text{aq}}$ and $[B]_{\text{aq}}$ from the single MD carried out on $[A^*B^*]_{\text{aq}^*}$, simply by separating the complex into its constituent parts, $[A^*]_{\text{aq}^*}$ and $[B^*]_{\text{aq}^*}$. This is referred to as the "single trajectory" (1T) method.

The binding free energy for the non-covalent association of two molecules may be written as:

$$\Delta G(A + B \rightarrow AB) = G_{AB} - G_A - G_B. \qquad (3)$$

The free energy of any molecule X may be divided into a contribution from the solute and a contribution from the solvent:

$$G(X) = G^{solute}(X) + G^{solvent}(X) \qquad (4)$$

The free energy contribution from the solute may be expressed as:

$$G^{solute}(X) = U(X) - TS(X) \qquad (5)$$

where

$$U(X) = \langle E(X) \rangle = \langle E_{\text{val}}(X) + E_{\text{ele}}(X) + E_{\text{vdW}}(X) \rangle \qquad (6)$$

and

$$S(X) = S_{\text{trans}} + S_{\text{rot}} + S_{\text{vib}} \qquad (7)$$

Here, $\langle E(X) \rangle$ represents the average MM energy of snapshots obtained from the MD simulation(s), which is comprised of internal energy contributions from bonds, angles and torsions (E_{val}), the electronic energy contribution (E_{ele}) and the van der Waals distance-dependent interaction energy (E_{vdW}), S_{trans} and S_{rot} are the entropic contributions from translational and rotational motion, respectively, obtained from classical statistical mechanics, and S_{vib} is the entropic contribution from vibrational motion, obtained using the normal mode analysis module

nmode implemented in AMBER 8. Note that this protocol neglects any small changes in volume upon complex formation as well as configurational entropy contributions from side-chain reorganization.

The quasi-harmonic analysis[69,70] was used as an alternative method of computing the entropy for this study. The length of the MD simulations were not extensive enough to provide convergent behavior (data not shown), which has been shown previously.[20,70] In addition, the quasi-harmonic vibrational entropy is highly dependent on the superposition of the structures used for analysis.[20] As such, we used results from the normal-mode analysis to estimate entropic contributions in this study.

The free energy contribution from the solvent may be expressed as:

$$G^{solvent}(X) = G^{es}(X) + G^{np}(X) \qquad (8)$$

where G^{es} is the electrostatic contribution and G^{np} is the non-polar contribution. In this study, the electrostatic contribution was calculated using PB as implemented in the *pbsa* program in AMBER 8 or one of several GB programs. GB is an analytic approximation to the PB method for solving the electrostatic contribution to the solvation free energy. The GB methods compared in this study are the Hawkins, Cramer, Truhlar pairwise generalized Born model[71,72] with parameters described by Tsui and Case[73] (GB^{HCT}, called by IGB = 1 in AMBER 8), the modified GB model developed by Onufriev, Bashford and Case, which has been optimized with two different α, β and γ parameter sets[49,50] (GB^{OBC}(I), called by IGB = 2 in AMBER 8 and GB^{OBC}(II), called by IGB = 5 in AMBER 8), and the GB and modified GB parameterizations derived by Jayaram, Sprous and Beveridge[74] (GB^{JSB}, called by IGB = 3 in AMBER 7 and MGB^{JSB}, called by IGB = 4 in AMBER 7). The non-polar contribution to the solvation free energy is calculated by:

$$G^{np}(X) = \gamma SASA(X) + b \qquad (9)$$

where γ is a surface tension parameter, set to 0.005 for PB[75] and 0.0072 for GB,[76] $SASA(X)$ is the solvent-accessible surface area of molecule X, computed with the program *molsurf*,[77] and b is a parameterized value, set to 0.86 for PB[75] and 0.0 for GB.[76]

Free energy decomposition was performed on a per-residue basis. Solute free energies and the electrostatic contribution to the solvation free energy were computed according to the methods described earlier in this section, while the non-polar contribution to the solvation free energy was determined using the LCPO method.[78]

MM-PB(GB)SA-LR calculations

Snapshots for MM-PB(GB)SA analysis were collected from the MD simulations once every 20 ps over the stable portion of computed trajectories, as determined from convergence of RMSD plots (Table 1). The same snapshots were used for normal mode analysis except for the free SL2 RNA trajectory, for which snapshots were collected every 200 ps. MM-PB(GB)SA and normal mode analysis were performed on each snapshot and averaged to obtain energy components and overall free energies of binding. Energy components were subsequently scaled using a LR-like equation:

$$\Delta G = \alpha(\Delta ES) + \beta(\Delta NES) + \gamma \qquad (10)$$

where α, β and γ are weight factors derived from a multiple regression analysis and ΔES and ΔNES represent the

electrostatic and non-electrostatic energy components, respectively, derived from MM-GBSA:

$$\Delta ES = \Delta ele + \Delta GB \qquad (11)$$

$$\Delta NES = \Delta val + \Delta vdW + \Delta np - T\Delta S \qquad (12)$$

Ab initio calculations

To probe the stacking energetics involved in U1A-RNA binding, *ab initio* calculations were performed between A6 of SL2 RNA and Phe56 of U1A, as well as the Ala, Leu, and Trp mutants. Bound forms were obtained using the average structures from MD simulations, followed by deletion of all atoms but the ones of interest. The ribose of A6 and the C^{α} of the amino acids were then replaced with a methyl group. Interaction energies were calculated using Gaussian 03[79] at the MP2/6-311G(d,p) level of theory[80–86] in the presence of solvent using a Polarizable Continuum model (PCM).[87–99] The free systems were calculated in water ($\varepsilon = 78.39$) to mimic a solvated environment, and the bound forms were calculated in diethyl ether ($\varepsilon = 4.335$) to mimic the hydrophobic core environment in the U1A-RNA complex. Reported interaction energies include (gas-phase) basis set superposition error (BSSE) corrections.[100]

Acknowledgements

References

1. Stark, H. & Lührmann, R. (2006). Cryo-electron microscopy of spliceosomal components. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 435–457.
2. Maris, C., Dominguez, C. & Allain, F. H.-T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131.
3. Varani, G. & Nagai, K. (1998). RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 407–445.
4. Hall, K. B. (1994). Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemisty*, **33**, 10076–10088.
5. Tsai, D. E., Harper, D. S. & Keene, J. D. (1991). U1-snRNP-A selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucl. Acids Res.* **19**, 4931–4936.
6. Boelens, W. C., Jansen, E. J., van Venrooij, W. J., Stripecke, R., Mattaj, I. W. & Gunderson, S. I. (1993). The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell*, **72**, 881–892.
7. Avis, J. M., Allain, F. H.-T., Howe, P. W. A., Varani, G., Nagai, K. & Neuhaus, D. (1996). Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.* **257**, 398–411.
8. Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
9. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
10. Rupert, P. B., Xiao, H. & Ferre-D'Amare, A. R. (2003). U1A RNA-binding domain at 1.8 A resolution. *Acta Crystallog. sect. D*, **59**, 1521–1524.
11. Mittermaier, A., Varani, L., Muhandiram, D. R., Kay, L. E. & Varani, G. (1999). Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J. Mol. Biol.* **294**, 967–979.
12. Nolan, S. J., Shiels, J. C., Tuite, J. B., Cecere, K. L. & Baranger, A. M. (1999). Recognition of an essential adenine at a protein-RNA interface: comparison of the contributions of hydrogen bonds and a stacking interaction. *J. Am. Chem. Soc.* **121**, 8951–8952.
13. Shiels, J. C., Tuite, J. B., Nolan, S. J. & Baranger, A. M. (2002). Investigation of a conserved stacking interaction in target site recognition by the U1A protein. *Nucl. Acids Res.* **30**, 550–558.
14. Tuite, J. B., Shiels, J. C. & Baranger, A. M. (2002). Substitution of an essential adenine in the U1A-RNA complex with a non-polar isostere. *Nucl. Acids Res.* **30**, 5269–5275.
15. Ajay, A. & Murcko, M. A. (1995). Computational methods to predict binding free energies in ligand-receptor complexes. *J. Med. Chem.* **38**, 4953–4967.
16. Swanson, J. M., Henchman, R. H. & McCammon, J. A. (2004). Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **86**, 67–74.
17. Jayaram, B., McConnell, K., Dixit, S. B. & Beveridge, D. L. (1999). Free energy analysis of protein-DNA binding: the EcoRI endonuclease- DNA complex. *J. Comput. Phys.* **151**, 333–357.
18. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L. *et al.* (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897.
19. Jayaram, B., McConnell, K., Dixit, S. B., Das, A. & Beveridge, D. L. (2002). Free-energy component analysis of 40 protein-DNA complexes: a consensus view on the thermodynamics of binding at the molecular level. *J. Comput. Chem.* **23**, 1–14.
20. Gohlke, H. & Case, D. A. (2004). Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex Ras-Raf. *J. Comput. Chem.* **25**, 238–250.
21. Gohlke, H., Kiel, C. & Case, D. A. (2003). Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the

Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* **330**, 891–913.

22. Proloff, N., Windemuth, A. & Honig, B. (1997). On the calculation of binding free energies using continuum methods: application to MHC Class I protein-peptide interactions. *Protein Sci.* **6**, 1293–1301.

23. Gilson, M. K., Given, J. A., Bush, B. L. & McCammon, J. A. (1997). The statistical thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **72**, 1047–1069.

24. Novotny, J., Bruccoleri, R. E., Davis, M. & Sharp, K. A. (1997). Empirical free energy calculations: a blind test and further improvements to the method. *J. Mol. Biol.* **268**, 401–411.

25. Simonson, T., Archontis, G. & Karplus, M. (2002). Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* **35**, 430–437.

26. Kombo, D. C., Jayaram, B., McConnell, K. J. & Beveridge, D. L. (2002). Calculation of the affinity of the λ repressor-operator complex based on free energy component analysis. *Mol. Simul.* **28**, 187–211.

27. Zhou, Z. & Madura, J. D. (2004). Relative free energy of binding and binding mode calculations of HIV-1 RT inhibitors based on Dock-MM-PG/GS. *Proteins: Struct. Funct. Bioinf.* **57**, 403–493.

28. Zhou, Z., Bates, M. & Madura, J. D. (2006). Structure modeling, ligand binding, and binding affinity calculation (LR-MM-PBSA) of human heparanase for inhibition and drug design. *Proteins: Struct. Funct. Bioinf.* **65**, 580–592.

29. Lee, F. S., Chu, Z.-T., Bolger, M. B. & Warshel, A. (1992). Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein Eng.* **5**, 215–228.

30. Aqvist, J., Medina, C. & Samuelsson, J. E. (1994). A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7**, 385–391.

31. Carlson, H. A. & Jorgensen, W. L. (1995). An extended linear resonse method for determining free energies of hydration. *J. Phys. Chem.* **99**, 10673–10677.

32. Reyes, C. M. & Kollman, P. A. (1999). Molecular dynamics studies of U1A-RNA complexes. *RNA*, **5**, 235–244.

33. Tang, Y. & Nilsson, L. (1999). Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.* **77**, 1284–1305.

34. Hermann, T. & Westhof, E. (1999). Simulations of the dynamics at an RNA-protein interface. *Nature Struct. Biol.* **6**, 540–544.

35. Reyes, C. M. & Kollman, P. A. (2000). Investigating the binding specificity of U1A-RNA by computational mutagenesis. *J. Mol. Biol.* **295**, 1–6.

36. Reyes, C. M. & Kollman, P. A. (2000). Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.* **297**, 1145–1158.

37. Olson, M. A. (2001). Calculations of free-energy contributions to protein-RNA complex stabilization. *Biophys. J.* **81**, 1841–1853.

38. Blakaj, D. M., McConnell, K. J., Beveridge, D. L. & Baranger, A. M. (2001). Molecular dynamics and thermodynamics of protein-RNA interactions: mutation of a conserved aromatic residue modifies stacking interactions and structural adaptation in the U1A-stem loop 2 RNA complex. *J. Am. Chem. Soc.* **123**, 2548–2551.

39. Showalter, S. A. & Hall, K. B. (2002). A functional role for correlated motion in the N-terminal RNA-binding domain of human U1A protein. *J. Mol. Biol.* **322**, 533–542.

40. Pitici, F., Beveridge, D. L. & Baranger, A. M. (2002). Molecular dynamics simulation studies of induced fit and conformational capture in U1A-RNA binding: do molecular substates code for specificity? *Biopolymers*, **65**, 424–435.

41. Showalter, S. A. & Hall, K. B. (2004). Altering the RNA-binding mode of the U1A RBD1 protein. *J. Mol. Biol.* **335**, 465–480.

42. Showalter, S. A. & Hall, K. B. (2005). Correlated motions in the U1 snRNA stem/loop 2:U1A RBD1 complex. *Biophys. J.* **89**, 2046–2058.

43. Zhao, Y., Kormos, B. L., Beveridge, D. L. & Baranger, A. M. (2006). Molecular dynamics simulation studies of a protein-RNA complex with a selectively modified binding interface. *Biopolymers*, **81**, 256–269.

44. Kormos, B. L., Baranger, A. M. & Beveridge, D. L. (2006). Do collective atomic fluctuations account for cooperative effects? Molecular dynamics studies of the U1A-RNA complex. *J. Am. Chem. Soc.* **128**, 8992–8993.

45. Kormos, B. L., Baranger, A. M. & Beveridge, D. L. (2007). A study of collective atomic fluctuations and cooperativity in the U1A-RNA complex based on molecular dynamics simulations. *J. Struct. Biol.* **157**, 500–513.

46. Zhao, Y. & Baranger, A. M. (2003). Design of an adenosine analogue that selectively improves the affinity of a mutant U1A protein for RNA. *J. Am. Chem. Soc.* **125**, 2480–2488.

47. Katsamba, P. S., Myszka, D. G. & Laird-Offringa, I. A. (2001). Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA. *J. Biol. Chem.* **276**, 21476–21481.

48. Law, M. J., Chambers, E. J., Katsamba, P. S., Haworth, I. S. & Laird-Offringa, I. A. (2005). Kinetic analysis of the role of the tyrosine 13, phenylalanine 56 and glutamine 54 network in the U1A/U1 hairpin II interaction. *Nucl. Acids Res.* **33**, 2917–2928.

49. Onufriev, A., Bashford, D. & Case, D. A. (2000). Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B*, **104**, 3712–3720.

50. Onufriev, A., Bashford, D. & Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins: Struct. Funct. Bioinf.* **55**, 383–394.

51. Dill, K. A. (1997). Additivity principles in biochemistry. *J. Biol. Chem.* **272**, 701–704.

52. Smith, P. E. & van Gunsteren, W. F. (1994). When are free energy components meaningful? *J. Phys. Chem.* **98**, 13735–13740.

53. Kranz, J. K., Lu, J. & Hall, K. B. (1996). Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain. *Protein Sci.* **5**, 1567–1583.

54. LeCuyer, K. A., Behlen, L. S. & Uhlenbeck, O. C. (1996). Mutagenesis of a stacking contact in the MS2 coat protein-RNA complex. *EMBO J.* **15**, 6847–6853.

55. Deardorff, J. A. & Sachs, A. B. (1997). Differential effects of aromatic and charged residue substitutions in the RNA binding domains of the yeast poly(A)-binding protein. *J. Mol. Biol.* **269**, 67–81.

56. Anderson, E. M., Halsey, W. A. & Wuttke, D. S. (2003). Site-directed mutagenesis reveals the thermodynamic requirements for single-stranded DNA recog-

nition by the telomere-binding protein Cdc13. *Biochemisty*, **42**, 3751–3758.

57. Butterfield, S. M., Cooper, W. J. & Waters, M. L. (2005). Minimalist protein design: a β-hairpin peptide that binds ssDNA. *J. Am. Chem. Soc.* **127**, 24–25.

58. Huang, C.-Y., Hsu, C.-H., Sun, Y.-J., Wu, H.-N. & Hsiao, C.-D. (2006). Complexed crystal structure of replication restart primosome protein PriB reveals a novel single-stranded DNA-bonding mode. *Nucl. Acids Res.* **34**, 3878–3886.

59. Zeeb, M., Max, K. E. A., Weininger, U., Löw, C., Sticht, H. & Balbach, J. (2006). Recognition of T-rich single-stranded DNA by the cold shock protein Bs-CspB in solution. *Nucl. Acids Res.* **34**, 4561–4571.

60. Friedman, R. A. & Honig, B. (1995). A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys. J.* **69**, 1528–1535.

61. Williamson, J. R. (2000). Induced fit in RNA-protein recognition. *Nature Struct. Biol.* **7**, 834–837.

62. Leulliot, N. & Varani, G. (2001). Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemisty*, **40**, 7947–7956.

63. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.

64. Kelley, L. A., Gardner, S. P. & Sutcliffe, M. J. (1996). An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. *Protein Eng.* **9**, 1063–1065.

65. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., III & DeBolt, S. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1–41.

66. Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E. *et al.* (2004). *AMBER 8.* University of California, San Francisco.

67. Kollman, P. A., Dixon, R., Cornell, W., Fox, T., Chipot, C. & Pohorille, A. (1997). The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In *Computer Simulation of Biomolecular Systems* (Wilkinson, A., Weiner, P. & van Gunsteren, W. F., eds), vol. 3, pp. 83–96. Kluwer, London.

68. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–936.

69. Schlitter, J. (1993). Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Letters*, **215**, 617–621.

70. Dixit, S. B., Andrews, D. Q. & Beveridge, D. L. (2005). Induced fit and the entropy of structural adaptation in the complexation of CAP and lambda repressor with cognate DNA sequences. *Biophys. J.* **88**, 3147–3157.

71. Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Letters*, **246**, 122–129.

72. Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. (1996). Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **100**, 19824–19839.

73. Tsui, V. & Case, D. A. (2001). Theory and applications of the generalized Born solvation model in macro-

molecular simulations. *Biopolymers (Nucl. Acid Sci.)*, **56**, 257–291.

74. Jayaram, B., Sprous, D. & Beveridge, D. L. (1998). Solvation free energy of biomacromolecules: parameters for a modified generalized Born model consistent with the AMBER force field. *J. Phys. Chem. B*, **102**, 9571–9576.

75. Sitkoff, D., Sharp, K. A. & Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978–1988.

76. Still, W. C., Tempczyk, A., Hawley, R. C. & Hendrickson, T. (1990). Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127–6129.

77. Connolly, M. L. (1983). Analytical molecular surface calculation. *J. Appl. Crystallog.* **16**, 548–558.

78. Weiser, J., Shenkin, P. S. & Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **20**, 217–230.

79. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R. *et al.* (2004). *Gaussian 03.* Gaussian, Inc., Wallingford, CT.

80. Møller, C. & Plesset, M. S. (1934). Note on an approximation treatment for many-electron systems. *Phys. Rev.* **46**, 618–622.

81. Head-Gordon, M., Pople, J. A. & Frisch, M. J. (1988). MP2 energy evaluation by direct methods. *Chem. Phys. Letters*, **153**, 503–506.

82. Sæbø, S. & Almlöf, J. (1989). Avoiding the integral storage bottleneck in LCAO calculations of electron correlation. *Chem. Phys. Letters*, **154**, 83–89.

83. Frisch, M. J., Head-Gordon, M. & Pople, J. A. (1990). A direct MP2 gradient method. *Chem. Phys. Letters*, **3**, 275–280.

84. Frisch, M. J., Head-Gordon, M. & Pople, J. A. (1990). Semi-direct algorithms for the MP2 energy and gradient. *Chem. Phys. Letters*, **166**, 281–289.

85. Head-Gordon, M. & Head-Gordon, T. (1994). Analytic MP2 frequencies without fifth-order storage. Theory and application to bifurcated hydrogen bonds in the water hexamer. *Chem. Phys. Letters*, **220**, 122–128.

86. Hehre, W. J., Radom, L., Schleyer, J. A. & Pople, P. v. R. (1986). *Ab Initio Molecular Orbital Theory.* Wiley, New York.

87. Miertš, S., Scrocco, E. & Tomasi, J. (1981). Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **55**, 117–129.

88. Miertš, S. & Tomasi, J. (1982). Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. *Chem. Phys.* **65**, 239–245.

89. Cossi, M., Barone, V., Cammi, R. & Tomasi, J. (1996). *Ab initio* study of solvated molecules: a new implementation of the polarizable continuum model. *Chem. Phys. Letters*, **255**, 327–335.

90. Mennucci, B. & Tomasi, J. (1997). Continuum solvation models: a new approach to the problem of solute's charge distribution and cavity boundaries. *J. Chem. Phys.* **106**, 5151–5158.

91. Mennucci, B., Cances, E. & Tomasi, J. (1997). Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: theoretical bases, computational implementation, and numerical applications. *J. Phys. Chem. B*, **101**, 10506–10517.

92. Cammi, R., Mennucci, B. & Tomasi, J. (1999). Second-order Møller-Plesset analytical derivatives

for the polarizable continuum model using the relaxed density approach. *J. Chem. Phys. A*, **103**, 9100–9108.

93. Cossi, M., Barone, V. & Robb, M. A. (1999). A direct procedure for the evaluation of solvent effects in MC-SCF calculations. *J. Chem. Phys.* **111**, 5295–5302.

94. Cammi, R., Mennucci, B. & Tomasi, J. (2000). Fast evaluation of geometries and properties of excited molecules in solution: a Tamm-Dancoff model with application to 4-dimethylaminobenzonitrile. *J. Phys. Chem. A*, **104**, 5631–5637.

95. Cossi, M. & Barone, V. (2000). Solvent effect on vertical electronic transitions by the polarizable continuum model. *J. Chem. Phys.* **112**, 2427–2435.

96. Cossi, M. & Barone, V. (2001). Time-dependent density functional theory for molecules in liquid solutions. *J. Chem. Phys.* **115**, 4708–4717.

97. Cossi, M., Rega, N., Scalmani, G. & Barone, V. (2001). Polarizable dielectric model of solvation with inclusion of charge penetration effects. *J. Chem. Phys.* **114**, 5691–5701.

98. Cossi, M., Scalmani, G., Rega, N. & Barone, V. (2002). New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* **117**, 43–54.

99. Cossi, M., Rega, N., Scalmani, G. & Barone, V. (2003). Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **24**, 669–681.

100. Boys, S. F. & Bernardi, F. (1970). The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **19**, 553–566.

Edited by D. Case