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

Dukagjin M. Blakaj, Kevin J. McConnell, David L. Beveridge,* and Anne M. Baranger*

Contribution from the Chemistry Department and Molecular Biophysics Program, Wesleyan University, Middletown, Connecticut 06459

Received August 21, 2000

Abstract: Molecular dynamics (MD) simulations and free energy component analysis have been performed to evaluate the molecular origins of the 5.5 kcal/mol destabilization of the complex formed between the N-terminal RNP domain of U1A and stem loop 2 of U1 snRNA upon mutation of a conserved aromatic residue, Phe56, to Ala. MD simulations, including counterions and water, have been carried out on the wild type and Phe56Ala peptide-stem loop 2 RNA complexes, the free wild type and Phe56Ala peptides, and the free stem loop 2 RNA. The MD structure of the Phe56Ala−stem loop 2 complex is similar to that of the wild type complex except the stacking interaction between Phe56 and A6 of stem loop 2 is absent and loop 3 of the peptide is more dynamic. However, the MD simulations predict large changes in the structure and dynamics of helix C and increased dynamic range of loop 3 for the free Phe56Ala peptide compared to the wild type peptide. Since helix C and loop 3 are highly variable regions of RNP domains, this indicates that a significant contribution to the reduced affinity of the Phe56Ala peptide for RNA results from cooperation between highly conserved and highly variable regions of the RNP domain of U1A. Surprisingly, these structural effects, which are manifested as cooperative free energy changes, occur in the free peptide, rather than in the complex, and are revealed only by study of both the initial and final states of the complexation process. Free energy component analysis correctly accounts for the destabilization of the Phe56Ala-stem loop 2 complex, and indicates that ∼80% of the destabilization is due to the loss of the stacking interaction and ∼20% is due to differences in U1A adaptation.

Introduction

The ribonucleoprotein (RNP) domain is a ubiquitous RNA binding domain that recognizes single-stranded RNA in various structural contexts with a wide range of affinities and specificities.[1] It is unclear how the arrangement of functional groups on the basic scaffold of the RNP domain is modulated to recognize specifically widely divergent RNA target sites. Highly conserved residues that contact RNA are assumed to be important for affinity and variable regions for specificity, but recent work argues against a clear distinction between the roles of conserved and variable sequences.[2,3] Baranger and co-workers have recently reported experimental measurements comparing the binding of the N-terminal RNP domain of U1A and a mutant, Phe56Ala, to stem loop 2 of U1 snRNA.[4] The large destabilization of the complex that resulted from mutation of the highly conserved Phe56 suggested this system to be an interesting prototype for studies of the roles of conserved residues in RNA−protein complexes. In this article, we report molecular dynamics (MD) simulations and a free energy component analysis on the binding of the N-terminal RNP domain of U1A and the Phe56Ala peptide to RNA. The calculations describe the observed destabilization and suggest that coupling between the conserved aromatic residue and variable regions in the free peptide contributes, along with stacking interactions in the complex, to the stability of the U1A−RNA complex. Our calculations also suggest that electrostatics are favorable to complexation, although not the dominant source of stability.

Background

The RNP domain consists of a four stranded antiparallel β-sheet supported by two α-helices. Highly conserved residues that contact RNA in RNP domains are found on the surface of the β-sheet. Three of these highly conserved residues are aromatic and stack with RNA bases in structurally characterized RNP−RNA complexes.[5−9] The N-terminal RNP domain of U1A, a component of the U1 snRNP, binds stem loop 2 of U1 snRNA with exceptionally high affinity and specificity (Figure 1).[10,11] To probe the contribution of one of the conserved

(1) Varani, G.; Nagai, K. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 407−445.
(2) Kranz, J. K.; Hall, K. B. *J. Mol. Biol.* **1998**, *275*, 465−481.
(3) Kranz, J. K.; Hall, K. B. *J. Mol. Biol.* **1999**, *285*, 215−231.
(4) Nolan, S. J.; Shiels, J. C.; Tuite, J. B.; Cecere, K. L.; Baranger, A. M. *J. Am. Chem. Soc.* **1999**, *121*, 8951−8952.
(5) Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. *Nature* **1994**, *372*, 432−438.
(6) Avis, J. M.; Allain, F. H.-T.; Howe, P. W. A.; Varani, G.; Nagai, K.; Neuhaus, D. *J. Mol. Biol.* **1996**, *257*, 398−411.
(7) Price, S. R.; Evans, P. R.; Nagai, K. *Nature* **1998**, *394*, 645−650.
(8) Handa, N.; Nureki, O.; Kuimoto, K.; Kim, I.; Sakamoto, H.; Shimura, Y.; Muto, Y.; Yokoyama, S. *Nature* **1999**, *398*, 579−585.
(9) Deo, R. C.; Bonanno, J. B.; Sonenberg, N.; Burley, S. K. *Cell* **1999**, *98*, 835−845.
(10) Tsai, D. E.; Harper, D. S.; Keene, J. D. *Nucleic Acids Res.* **1991**, *19*, 4931−4936.

10.1021/ja005538j CCC: $20.00
Published on Web 02/20/2001

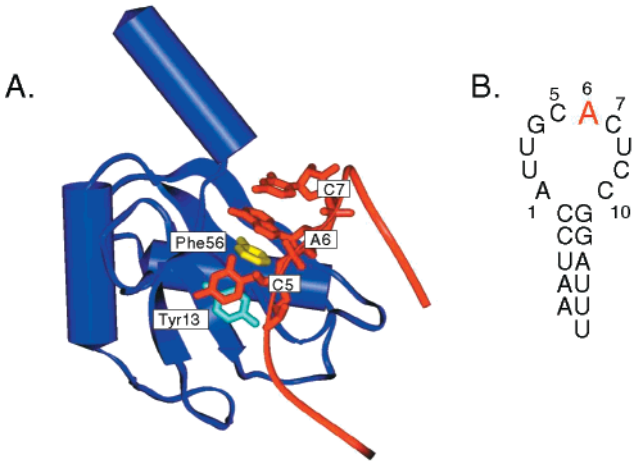

Figure 1. (A) Diagram of the complex formed between the N-terminal RNP domain of U1A and stem loop 2 of U1 snRNA from the X-ray cocrystal structure with four additional C-terminal residues modeled.[5] Only a portion of stem loop 2, in red, is shown. (B) Stem loop 2 of U1 snRNA. The adenine that stacks with Phe56 is in red.

stacking interactions to the stability of the U1A−RNA complex, Baranger and co-workers mutated one of these residues, Phe56, to Ala, Leu, His, Trp, and Tyr.[4] Mutation of Phe56 to Ala resulted in a surprisingly large, 5.5 kcal/mol, destabilization of the complex. The size of this destabilization reflects the importance of stacking interactions in RNA recognition, but the observed magnitude exceeds that ordinarily ascribed to stacking effects.[12−14]

Previous MD studies of this complex have been reported.[15−19] Reyes and Kollman, using AMBER, showed the protein−RNA complex was stable in MD and the protein−RNA interface exhibited reduced thermal motions.[16,18] Free energy component analysis was performed on U1A binding to stem loop and internal loop RNAs, using estimates of internal enthalpies and entropies from the AMBER and Poisson−Boltzmann/solvent accessibility (PBSA) estimates of the free energy of solvation.[17] Good agreement with calculated binding affinities was reported, although a potentially significant translational and rotational entropy change was neglected. MD simulations on the U1A−stem loop 2 complex using CHARMM were reported by Tang and Nilsson.[19] Their results showed similar dynamic behavior of the system and demonstrated that both the loop and stem regions of the RNA become more ordered on binding. However, Reyes and Kollman concluded that electrostatics is unfavorable to complexation, whereas Tang and Nilsson identified electrostatics as the "dominant source of RNA−protein interaction energy."

Calculations

The calculations involved in this study are (a) MD simulations on the complex formed between the N-terminal RNP domain of U1A and stem loop 2 of U1 snRNA including counterions and water, (b) corresponding MD simulations on the uncomplexed N-terminal RNP domain of U1A and stem loop RNA, and (c) free energy component analysis of the U1A−RNA crystal structure.[20] All MD simulations were carried out using AMBER 5.0,[21] the parm95 AMBER force field developed by Cornell et al.,[22] and the particle mesh Ewald (PME) treatment of long-range forces.[23] The simulation cell in each case was comprised of solute, counterions, and TIP3P water. The MD protocol applied here closely parallels that described recently elsewhere;[24] specifics with regard to the molecules treated in this study are noted in the presentation of results below.

Free energy component analysis considers the free energy of the complex and unbound constituents as the sum of terms identified with the various chemical and thermodynamic forces including solvation.[25] The theoretical basis of this approach and the approximations involved in component analysis have been discussed elsewhere on the basis of statistical mechanics.[26,27] The treatment of free energy via component analysis invokes assumptions regarding additivity and involves individual terms treated by a set of plausible theoretical and semiempirical estimates. This analysis offers the material advantage of being readily decomposable into contributions of terms identifiable with valence, van der Waals, electrostatic, and hydrophobic forces. While quantitative in nature, the results, considering the approximations involved, are best utilized as a basis for qualitative analysis of a binding problem and consideration of trends across related systems.

Component analysis has been applied to the U1A−RNA complex by Reyes and Kollman as described above.[17] Recent applications of this methodology in the form used herein have been reported for the Eco RI endonuclease complex,[28] the λ repressor-operator,[29] and a study of some 40 protein DNA complexes based on crystal structures.[27] In calculating the binding free energy of the U1A−RNA complex, internal energies of protein and RNA are computed with use of the MD force field.[22] Solvation free energy is treated by the method of Generalized Born−Solvent Accessibility (GBSA),[30] with the modifications suggested by Jayaram et al.[31] The GBSA method has been demonstrated to give agreement within 5% of observed solvation free energies for a large number of small molecules and ions, including prototypes of the sugars, phosphate ions, and nucleotide bases of RNA.[31] Related studies on protein−protein interactions,[32] ligand binding of proteins,[33] and protein folding have been performed recently.[34] Free energy component analysis as applied herein includes both counterion reorganization and entropy terms and follows the computational protocols described previously unless otherwise noted.[27,29] Full details of this and all other calculations referred to in this article are described by Blakaj.[35]

(11) Hall, K. B. *Biochemistry* **1994**, *33*, 10076−10088.
(12) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X. F.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 2213−2222.
(13) Friedman, R. A.; Honig, B. *Biophys. J.* **1995**, *69*, 1528−1535.
(14) Burkard, M. E.; Kierzek, R.; Turner, D. H. *J. Mol. Biol.* **1999**, *290*, 967−982.
(15) Hermann, T.; Westhof, E. *Nat. Struct. Biol.* **1999**, *6*, 540−544.
(16) Reyes, C. M.; Kollman, P. A. *RNA* **1999**, *5*, 235−244.
(17) Reyes, C. M.; Kollman, P. A. *J. Mol. Biol.* **2000**, *297*, 1145−1158.
(18) Reyes, C. M.; Kollman, P. A. *J. Mol. Biol.* **2000**, *295*, 1−6.
(19) Tang, Y.; Nilsson, L. *Biophys. J.* **1999**, *77*, 1284−1305.
(20) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Am. Chem. Soc.* **1998**, *120*, 10629−10633.
(21) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. *AMBER: Version 5*, 5.0 ed.; University of California: San Francisco, 1997.
(22) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(23) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(24) Kombo, D. C.; Young, M. A.; Beveridge, D. L. *Biopolymers* **2000**, *53*, 596−605.
(25) Ajay; Murcko, M. A. *J. Med. Chem.* **1995**, *38*, 4953−4967.
(26) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(27) Jayaram, B.; McConnell, K.; Dixit, S. B.; Beveridge, D. L. *J. Comput. Chem.* **2001**, in press.
(28) Jayaram, B.; McConnell, K. J.; Dixit, S. B.; Beveridge, D. L. *J. Comput. Phys.* **1999**, *151*, 333−357.
(29) Kombo, D.; Jayaram, B.; McConnell, K.; Beveridge, D. *Mol. Simul.* **2001**, in press.
(30) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(31) Jayaram, B.; Sprous, D.; Beveridge, D. L. *J. Phys. Chem.* **1998**, *102*, 9571.
(32) Proloff, N.; Windemuth, A.; Honig, B. *Protein Sci.* **1997**, *6*, 1293−1301.
(33) Wang, L.; Eriksson, M. A. L.; Pitera, J.; Kollman, P. A. *ACS Symp. Ser.* **1999**, *719*, 37−52.
(34) Simmerling, C.; Lee, M.; Ortiz, A.; Kolinski, A.; Skolnick, J.; Kollman, P. *J. Am. Chem. Soc.* **2000**, *122*, 8392−8402.
(35) Blakaj, D. M.S. Thesis, Wesleyan University: Middletown, CT, 2001.

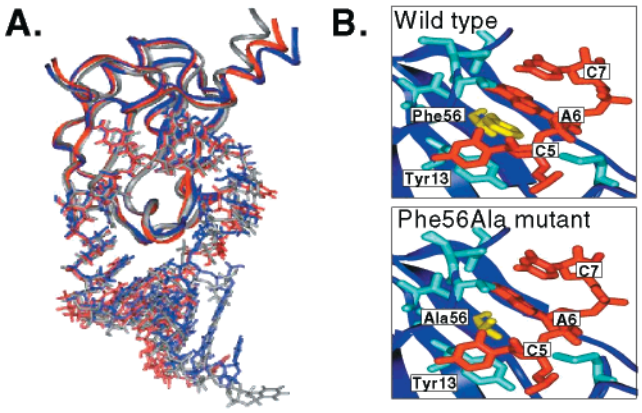

Figure 2. (A) A superposition of the peptide backbones of the average MD structures of the RNA complexes of the wild type (red) and Phe56Ala mutant (blue) peptides and the X-ray cocrystal structure of the wild type stem loop 2 complex (gray).[5] (B) Close-up of the average MD structures of the wild type peptide-RNA complex (top) and the Phe56Ala peptide−RNA complex (bottom).

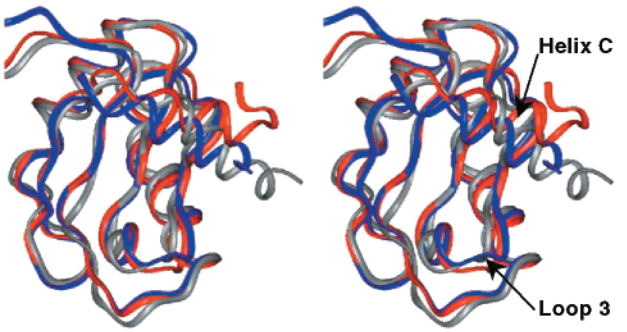

Figure 3. Stereoview of the superposition of the peptide backbones of the average MD structures of the wild type (red) and Phe56Ala (blue) peptides and the NMR structure (gray).[6]

Results and Discussion

MD simulations were performed on stem loop 2 RNA, the wild type and Phe56Ala peptides, and their respective complexes with stem loop 2 RNA. Initial structures for the simulations of the Phe56Ala and wild type peptide−RNA complexes and for stem loop 2 were based on the X-ray cocrystal structure, with four additional C-terminal amino acids added to produce a structure comprised of amino acids 2−102.[5] The NMR structure of the free peptide (amino acids 2−117) was used as the initial structure in MD for the free wild type and Phe56Ala peptides, removing the C-terminal 15 amino acids and incorporating the mutations Tyr31His and Gln34Arg to correspond to the crystal structure sequence.[6] MD was based on AMBER and the parm.94 force field,[22] including TIP3P water, sodium ions, and chloride ions at a concentration of added salt of 250 mM, matching the experimental conditions under which the binding constants were measured.[4] All MD simulations had stable trajectories over 3 ns based on the examination of root-mean-square deviations as a function of time (not shown).

MD on the wild type and Phe56Ala peptide−RNA complexes showed no significant structural changes (Figure 2A), with the space created by the Phe56Ala mutation taken up by water molecules. The calculated structure of the wild type−RNA complex is similar to those obtained in other studies.[15,16,19] The average MD structure of the backbones of residues 5−95 was just 0.55 and 0.47 Å from the cocrystal structure for the wild type and Phe56Ala peptide−RNA complexes, respectively. Even in the region local to the Phe56Ala mutation, the position and orientation of the side chains in the wild type and Phe56Ala peptides is almost identical (Figure 2B), consistent with experiments that suggest the hydrogen bonding network around the adenine (A6) that stacks with Phe56 is maintained in the Phe56Ala peptide−RNA complex.[4]

For the free stem loop 2, the MD structure is significantly different from that in the complex, 2.39 Å distant in heavy atom RMSD. The unpaired nucleotide bases of the loop region point toward the interior in the MD model for free RNA, whereas in the complex these bases splay to contact the β-sheet. Noted previously in MD simulations,[17,19] we find this adaptation to be similar in wild type and Phe56Ala complexes and therefore neglected it in our $\Delta\Delta G$ calculations.

In the free peptide, the Phe56Ala mutation results in changes in the structure and dynamics of helix C (residues 90−98), a variable region known to be important for the high specificity of U1A for its target site (Figure 3).[36] The root-mean-square deviation (RMSD) of the average MD structure, with respect to the NMR structure of the backbone residues 5−95, was 1.44 and 1.78 Å for the wild type and Phe56Ala peptides, respectively. Kollman and co-workers also reported MD simulations on the wild type peptide, obtaining a similar RMSD, and noted the flexibility of both helix C and loop 3 in their simulated structure.[17] In the NMR structure, helix C contacts residues on the surface of the β-sheet, while these contacts are absent in the complex.[6] Mutation of Tyr13Phe on the surface of the β-sheet results in a significant loss in binding in the context of the 2−102 peptide (2.7 kcal/mol) and an even greater loss in binding affinity, 3.6 kcal/mol, in the context of the 2−95 peptide that lacks a portion of helix C.[2] Thus, there is thermodynamic coupling between the conserved residues of the β-sheet and the variable residues in the terminal helix and the 2−95 peptide is more sensitive to changes in the sequence of the β-sheet than is the 2−102 peptide. Therefore, any changes in the interactions between helix C and the β-sheet seen in MD of the Phe56Ala peptide should affect RNA affinity.

Parsing the MD energy with respect to constituents revealed that nonbonding interactions between amino acids in helix C and the remainder of the peptide are stronger in the Phe56Ala peptide than in the wild-type peptide. This result is supported by a comparison of the thermal B-factors for residues in helix C, which indicate that helix C shows a reduced (∼40%) dynamic range of motion in the Phe56Ala mutant peptide compared with the wild type. The most dramatic changes are seen at Ile94, Lys96, and Met97. These results suggest, if the enthalpy contribution is dominant, that disruption of the stronger nonbonded interactions in the free Phe56Ala peptide may contribute to its poor affinity for RNA. The MD simulations also suggest that changes in the dynamics of loop 3 result from the Phe56Ala mutation. The thermal B-factors of the MD structures of the free peptides and the RNA−peptide complexes indicate that residues in loop 3 are 60% more flexible in Phe56Ala than in the wild type peptide, both when free and when bound to stem loop 2. Previous NMR experiments showed mutation of residues on the β-sheet (Tyr13Phe, Phe56Tyr, or Gln54Glu) resulted in increased dynamics in loop 3 in the free peptide.[3] Because a conformational restriction of loop 3 is required to reach the final complexed structure, the increased flexibility of loop 3 in the Phe56Ala peptide may also contribute to low RNA affinity.[37,38]

(36) Scherly, D.; Kambach, C.; Boelens, W.; van Venrooij, W. J.; Mattaj, I. W. *J. Mol. Biol.* **1991**, *1991*, 577−584.
(37) Lu, J.; Hall, K. B. *Biochemistry* **1997**, *36*, 10393−10405.
(38) Mittermaier, A.; Varani, L.; Muhandiram, D. R.; Kay, L. E.; Varani, G. *J. Mol. Biol.* **1999**, *294*, 967−979.

To examine this hypothesis further, we performed free energy component analysis of the binding of wild type and Phe56Ala peptides to stem loop 2 RNA. The observed destabilization of the complex on Phe56Ala mutation is reproduced, albeit overestimated. In both complexes, we find the contribution from electrostatics to be net favorable to binding, considering both intramolecular and solvation components. However, the dominant term favoring complexation is van der Waals interactions, i.e., shape complementarity and nonelectrostatic components of the solvation energy. Our calculations predict successfully the lower RNA binding energy of the Phe56Ala peptide; the destabilization can be apportioned as ∼78% due to stacking and ∼22% due to U1A adaptation.

Conclusions

The MD simulations suggest that the mutation of a highly conserved aromatic residue to alanine in U1A changes the structure and dynamics of two of the most variable regions of RNP domains, helix C and loop 3. The altered structure of the free Phe56Ala peptide and the missing stacking interaction in the complex are responsible for the low RNA affinity of the Phe56Ala peptide. These results argue against the assumption that conserved regions of proteins in protein−nucleic acid complexes provide affinity, whereas variable regions provide specificity. Instead, intricate cooperation between conserved and variable residues in both the free peptide and the complex enables the high affinity and specificity of binding.[2] In fact, cooperation between conserved and variable sequences may be a general mechanism to achieve specific, high-affinity RNA recognition.

Acknowledgment. Funding was provided by NIH to A.M.B., GM-56857, and to D.L.B., GM-37909. A.M.B. is an Alfred P. Sloan Research Fellow. K.J.M. was supported by a NIH Training Grant in Molecular Biophysics (GM-08271).

JA005538J