Structural principles of RNA catalysis in a 2'-5' lariat forming ribozyme

Teresa Carlomagno*, Irene Amata, Luca Codutti, Melanie Falb, Jörg Fohrer, Pawel Masiewicz, Bernd Simon

Structural and Computational Biology Unit, EMBL, Meyerhofstrasse 1, D-69117, Heidelberg, Germany

Table of Contents

Supplementary Methods

Resonance Assignment.

The first necessary step to study the conformation of the 2'-5' AG1 branch forming ribozyme was the assignment of its ^{13}C, ^{15}N and ^{1}H resonances. The extensive resonance overlap, in particular in the ribose region, represented the main obstacle. In general, for large RNAs, ^{13}C,^{15}N (and/or ^{2}H)-selective or unselective labeling of the molecule, in combination with multidimensional NMR experiments, is the solution to the assignment problem.

The resonance assignment of the 2'-5' AG1 branch forming ribozyme was based on the following experiments: 2D ^{13}C/^{15}N-edited HSQCs, 2D HNN-COSY[1,2], 2D imino NOESY, 3D HsCNb/HbCNb[3], 3D HCCH-COSY-TOCSY[4], 3D ^{13}C-edited/^{13}C-filtered NOESY and 3D ^{13}C-edited/^{12}C-filtered NOESY[5] acquired on the following selectively labeled samples: 1) ^{13}C,^{15}N A-labeled RNA; 2) ^{13}C,^{15}N G-labeled RNA; 3) ^{13}C,^{15}N C-labeled RNA; 4) ^{13}C,^{15}N U-labeled RNA; 5) ^{13}C,^{15}N AU-labeled RNA.

Other experiments, such as 3D ^{13}C-edited/^{13}C-filtered base-selective NOESY[6] and 3D HCCH-COSY, were performed as well, but did not contribute substantially to the assignment. We also performed the experiments listed above on other three samples: ^{13}C,^{15}N uniformly labeled RNA; ^{13}C,^{15}N AG-labeled RNA; ^{13}C,^{15}N G-labeled, ^{2}D-AU-labeled RNA. However, these last three samples were of limited utility for assignment, either due to the extensive resonance overlap, as in the case of the uniformly labeled sample, or due to the low concentration of the samples (< 0.1 mM) that prevented the collection of high-quality NOESY spectra. All experiments were acquired at 800 or 900 MHz spectrometers. The use of high field magnets is essential to obtain an acceptable resolution of the RNA resonances. The ^{13}C-HSQC spectra showing the H8,H6/C8,C6 correlation and the H1'-C1' correlation for samples 1-4 are shown in Supplementary Fig. 1.

The ribose protons (H1', H2', H3', H4', H5', H5") assignment was based on correlations detected in 3D HCCH-COSY-TOCSY spectra and confirmed by ^{13}C-edited/^{13}C-filtered NOESY spectra. 82 % of sugar spin systems have been assigned (90 % in the 1-29 stretch at the 5' end of the molecule; 74 % in the 30-59 stretch the 3' end of the

molecule). 25 out of 26 H5 protons have been correlated to the respective H6 protons in pyrimidines. 90% of the A-H2 protons have been assigned by analysis of the 3D ^{13}C-edited/^{13}C-filtered and 3D ^{13}C-edited/^{12}C-filtered NOESY spectra.

In order to resolve residual ambiguities in the assignment, the following eight mutant RNAs were synthesized: G2A; A33G; A40G & U56C; G49C, A50C, U51C, A52U, U53C. Mutant RNAs were selectively ^{13}C,^{15}N A-, U-, C- or G-labeled, depending on the nucleotide type subject to mutation (i.e. the G2A mutant RNA is ^{13}C,^{15}N G-labeled). Thus, single nucleotide substitutions cause the disappearance of the relative resonance peaks in the NMR spectra of the selectively labeled samples. The analysis of the spectra was complicated by the occurrence of chemical shift perturbations for several residues in spatial proximity to the mutated nucleotide. Despite these difficulties, spectral analysis of the mutant RNAs helped to confirm the assignment.

Non-canonical base pair topology.

The identification of the topology for the 4 non-canonical A˙A and G˙A base pairs of helix 1-27 was attempted through band selective HNN-COSY experiments[1] applied to a uniformly ^{13}C,^{15}N labeled, truncated RNA comprising nucleotides 1 to 29. The choice of this truncated RNA to represent the 5'-terminal half of the 2'-5' AG1 lariat forming ribozyme is justified by the fact that the 1-27 and the 32-59 stretches of the 2'-5' AG1 lariat forming ribozyme are structurally independent in the absence of magnesium (see Results section). The HNN-COSY experiment detects hydrogen bonds between imino or amino groups and acceptor nitrogens. Briefly, in our version of the experiment, the magnetization starts from either the H8 or the H2 of adenosine and is transferred through $^{2}J_{HN}$ scalar couplings to either the N7 or the N1/N3 of the base ring, respectively. A further transfer step through the hydrogen bond mediated $^{2}J_{NN}$ scalar coupling brings the magnetization to the N6(N2) of the base paired adenosine(guanosine). The magnetization is then transferred back to H8/H2 for detection. A second magnetization transfer pathway uses the intra-ring $^{2}J_{N1N6}$ coupling and generates a correlation from the H2 atom of Adenosine, over the N1, to the N6 atom of the same ring. Both 2D HNN-COSY and 3D HCNN-COSY experiments were recorded. Unfortunately, the extensive overlap of both the N1 and N6 resonances of adenosines and the presence of two magnetization transfer pathways made the assignment of the spectra impossible. Nevertheless, one piece of information can be derived from the 3D HCNN-COSY

experiments: the N7 atoms of adenosines or guanosines are not involved in hydrogen bonds, as no correlation is observed in the 3D HCNN-COSY for H8-C8 atom pairs.

To infer the geometry of the non-canonical base pairs, we separated them in two groups (A8·A20 and A7·C21) and (G2·A26, A3·A25 and A4·A24), consisting of neighboring base pairs. *Determination of the base pair geometry in the (A8·A20 and A7·C21) group*: we performed structure calculation of the 2'-5' AG1 lariat forming ribozyme without imposing any restraints on hydrogen bonds between A8 and A20 or A7 and C21, but imposing co-planarity of the base-paired rings. 9 of the lowest 20 energy structures formed an A8·A20 base pair. The NOEs drove the geometry of this base pair to cis WC-WC with a clear preference for the A8-N1 – A20-N6 hydrogen bond (75% of the structures) over the A8-N6 – A20-N1 variant. Similarly, the NOEs determined the geometry of the A7·C21 base pair as cis WC-WC with a hydrogen bond between A7-N6 and C21-N3. *Determination of the base pair geometry in the (G2·A26, A3·A25 and A4·A24) group.* Analogously to what described for the previous group, structure calculations for the 2'-5' AG1 lariat forming ribozyme were performed without imposing any hydrogen bond between G2 and A26, A3 and A25 or A4 and A24, but restraining the co-planarity of the rings involved in base pairs. 13 of the lowest 20 energy structures formed an A4·A24 base pair. In 50% of the structures the A4·A24 base pair involved the N1 of A4 as acceptor in the cis WC-WC geometry, while in the other 50% the A4·A24 base pair was cis WC-SE with hydrogen bonding A4-N6 – A24-N3. This second geometry cannot be easily accommodated in a helix; in fact, in all structures with the A4·A24 base pair in the cis WC-SE geometry, neither the base pair A3·A25 nor the base pair G2·A26 were formed. The mutant analysis (Fig. 2), however, indicates that helix H1 is continuous, with single canonical base pairs at positions 4-24, 3-25 and 2-26 supporting catalysis. This allows us to exclude the geometry of the A4·A24 base pair that is incompatible with helical stacking and to assign this base pair to the cis WC-WC geometry with hydrogen bond A4-N1 – A24-N6. In the next structure calculation, the geometry of only the A3·A25 and G2·A26 base pairs was left unrestrained. The NOEs drove the G2·A26 base pairs unambiguously to the cis WC-WC geometry with G2-N1 – A26-N1 and G2-C6 – A26-N6 hydrogen bonds. Similarly, for the A3·A25 base pair, the structures converged to the cis WC-WC geometry with 60% having the hydrogen bond between A3-N6 – A25-N1 and the remaining 40% having the hydrogen bond between A3-N1 – A25-N6. To further distinguish between the two cis WC-WC geometries of the A3·A25 base pair, we performed one structure calculation run with each of the geometries. The structures with

hydrogen bond between A3-N6 – A25-N1 had a lower total energy and less NOE violations than those with the A3-N1 – A25-N6 hydrogen bond, clearly confirming the first pattern.

An analogous strategy was used to define the geometry of the non-canonical base pair A35·A44. The structures, calculated without imposing any base pair topology, converged to the trans Hoogsteen-WC geometry with hydrogen bond between A35-N6 – A44-N1 in 70% of the low energy conformers.

Structure calculations.

Structures were calculated using the Aria 1.2/CNS 1.1 set-up[7,8]. 1234 unambiguous and 105 ambiguous NOE distances were categorized as weak (2.0 - 5.5 Å), medium (2.5 - 4.0 Å) or strong (1.8 - 3.0 Å) (Table 1). The ribose conformation of nucleotides 1-12, 17-27, 32-34, 39-41, 45-47 and 55-57 of helix H1, H2 or H3 was restrained to the *C3'-endo* range, as indicated by the analysis of the chemical shifts of the C1', C4' and C5' carbons[9]. The dihedral angles α, β, ε and ζ were restrained to A-form helix ranges 300º ± 30º, 180º ± 30º, -135º ± 30º and 300º ± 30º, respectively for nucleotides 5, 10-12, 17-19, 23, 32-34, 38-41, 45, 46, 55-57, involved in canonical WC base pairs, and loosely to the allowed ranges 180º ± 150º, 180º ± 110º, -125º ± 75º and 180º ± 150º, respectively, for all other nucleotides. The dihedral angle γ was restrained to the *gauche*$^+$ range for nucleotides involved in canonical base pairs only. The χ angles of 39 nucleotides were restricted to the *anti* conformation on the basis of the intensities of the intranucleotide H8-H1' (Pu), and H6/H5-H1' (Py) NOEs.

Hydrogen bonds of WC base pairs were detected in HNN correlations and in NOESY experiments.

During the calculations, hydrogen bonds were maintained by distances restraints, while planarity was enforced through weak planarity restraints (5 kcal mol^{-1} Å^{-2}).

200 structures were calculated in one iteration without using the automated assignment or the distance calibration options of Aria 1.2. The simulated annealing (SA) protocol starts with a high-temperature torsion angle simulated annealing phase with 100,000 steps at 20,000 K (time step of 25 fs). This is followed by a torsion angle dynamic cooling phase from 20,000 K to 2000 K in 20,000 steps and by two cartesian dynamic cooling phases with a time step of 3 fs (from 2000 K to 1000 K in 100,000 steps and from 1000 K to 50 K in 280,000 steps, respectively). The final ensemble of 20 structures was refined in a shell of water molecules[10–12].

The final structures showed no NOE (> 0.5 Å) or dihedral angle (> 3°) violations. They were analysed using MolMol[13]. Figures were prepared with Pymol (http://www.pymol.org).

The model of the active state was calculated in a similar manner as the structure of the inactive state. In addition to the NMR derived restraints defining the structure of helix H1 and of the 3'-terminal pseudoknot, hydrogen bonds restraints were added between A48-N6 and N7 and A26-N7 and N6, respectively; coplanarity was imposed for the base rings of A26 and A48. A distance restraint of 2.5 ± 0.5 Å was imposed between the G1-αP and the A48-2'-O. NOEs stemming from A31 and G49 were eliminated to allow for rearrangements of the relative position of the 5'-terminal helix and the 3'-terminal pseudoknot in the active state.

Positioning of the Mg^{2+} ions.

Molecular Dynamics of the 2'-5' AG1 lariat-forming ribozyme were prepared using the AmberTools 11.0 suite and run using AMBER 11[14]. The system in analysis consists of a molecule of ribozyme whose charges are counter-balanced by K^+ ions. The molecular dynamics were set in explicit TIP3P water[15] forming an octahedral layer of 12 Å around the RNA. We used the F99SB force-field[16] to which guanosine triphosphate parameters were added[17,18]. Ion parameters were used as in the AMBER force field. Prior to the dynamics, the K^+ ions were randomized in different runs using independent randomization seeds, so that inter-ions or RNA-ions distances were > 5 Å. K^+ ions diffused in the RNA molecule during 1.0 ns of Molecular Dynamics simulation at constant pressure (1 atm) and temperature (300 K) using a Langevin thermostat. During the simulation a restraint mask of 0.5 kcal mol^{-1} $Å^{-2}$ was imposed onto the RNA. In addition, a weak distance restraint with average value of 2.9 Å was imposed between G1-Pα and A48-2'-O with force varying between 0.01 and 0.2 kcal mol^{-1} $Å^{-2}$. The cutoff particle-mesh Ewald for the non-bonded long-range electrostatics was set to 10 Å. The trajectories were analyzed by means of the bio3d package[19] implemented in the software R[20]. The ions that were found to reside for more than 50% of the simulation time closer than 3.5 Å to the ribozyme were collected from all runs. From them, 8 ions occupying conserved sites in all runs were selected, corresponding to a final concentration of ca. 30 mM, and mutated to Mg^{2+} ions for subsequent Molecular Dynamics. MD simulations were set as before, starting from the last equilibrated frame from the K^+ dynamics runs. Magnesium ions parameters and force field modification

were based on reference[21]. The simulations aimed at equilibrating the system with the Mg^{2+} ions were run for a total of 0.62 ns at constant pressure (1 atm) and temperature (298 K) using a Langevin thermostat. The equilibration was divided in 14 steps gradually[22] releasing a restraint mask from heavy atoms and backbone from 10 to 0.001 kcal mol^{-1} $Å^{-2}$. A distance restraint with average value of 3.0 Å was imposed between G1-Pα and A48-2'-O with force of 5 kcal mol^{-1} $Å^{-2}$ during the simulation; furthermore, co-planarity (15 kcal mol^{-1} $Å^{-2}$) and distance restraints (5 kcal mol^{-1} $Å^{-2}$) were imposed between bases A48 and A26. The cutoff particle-mesh Ewald for the non-bonded long-range electrostatics was set to 10 Å.

Supplementary Tables

Table S1.
Overview of non-canonical base pairs in the 2'-5' AG1 lariat forming ribozyme, determined by NMR data, as explained in the Supplementary Methods.

Base pair	Base pair type	Hydrogen bonds
G2·A26	*cis* WC-WC	G2-N1 ⋯ A26-N1
		G2-O6 ⋯ A26-N6
A3·A25	*cis* WC-WC	A3-N6 ⋯ A25-N1
A4·A24	*cis* WC-WC	A4-N1 ⋯ A24-N6
A7·C21	*cis* WC-WC	A7-N6 ⋯ C21-N3
A8·A20	*cis* WC-WC	A8-N1 ⋯ A20-N6
A35·A44	*trans* Hoogsteen-WC	A35-N6 ⋯ A44-N1
		A35-N7 ⋯ A44-N6

Table S2.

Mean helical parameters and major groove widths for the 5'-terminal helix 1-27, averaged over the 20 lowest energy structures after water refinement.

Base pair	Inclination (°)	Base	Rise	Base position	Major groove width
G1-C27	4±6	G1/G2	3.2±0.2		
G2-A26	12±6	G2/A3	3.00±0.08		
A3-A25	11±5	A3/A4	3.1±0.1	A3	11.6±0.6
A4-A24	7±5	A4/G5	3.1±0.2	A4	10.9±0.6
G5-C23	6±5	G5/A6	3.13±0.05	G5	9.0±0.7
A6-U22	3±4	A6/A7	3.03±0.09	A6	7.8±0.6
A7-C21	0.4±3	A7/A8	3.4±0.1	A7	7.4±0.5
A8-A20	5±3	A8/G10	2.8±0.2	A8	8.1±0.5
G10-C19	15±3	G10/G11	3.1±0.1	G10	10.2±0.6
G11-C18	11±4	G11/C12	3.27±0.06		
C12-G17	15±5				

Supplementary Figures

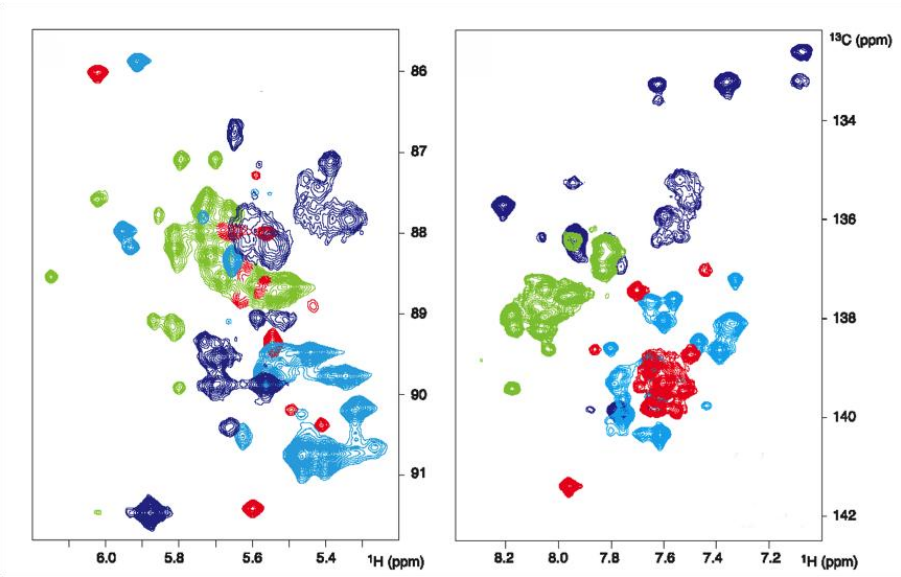

Figure S1.

Overlay of ^{13}C-HSQC spectra of the differentially labeled samples of the 2'-5' AG1 lariat forming ribozyme. Left panel: H1'-C1' ribose resonances; right panel: H6-C6 (Py) / H8-C8 (Pu) base resonances. Peaks relative to ^{13}C/^{15}N selective U-, A-, C- and G-labeled RNAs are depicted in, red, green, light blue and dark blue, respectively.

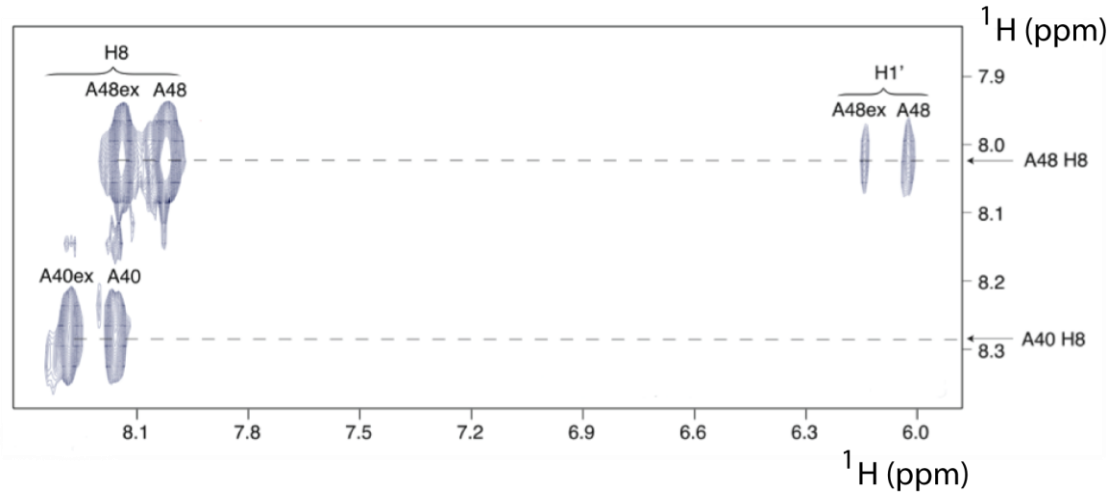

Figure S2.

2D slice from a ^{13}C-edited/^{13}C-filtered NOESY spectrum run for the ^{13}C/^{15}N selective A-labeled RNA. Diagonal signals of A40 and A48 and relative exchange peaks of the two detectable conformations.

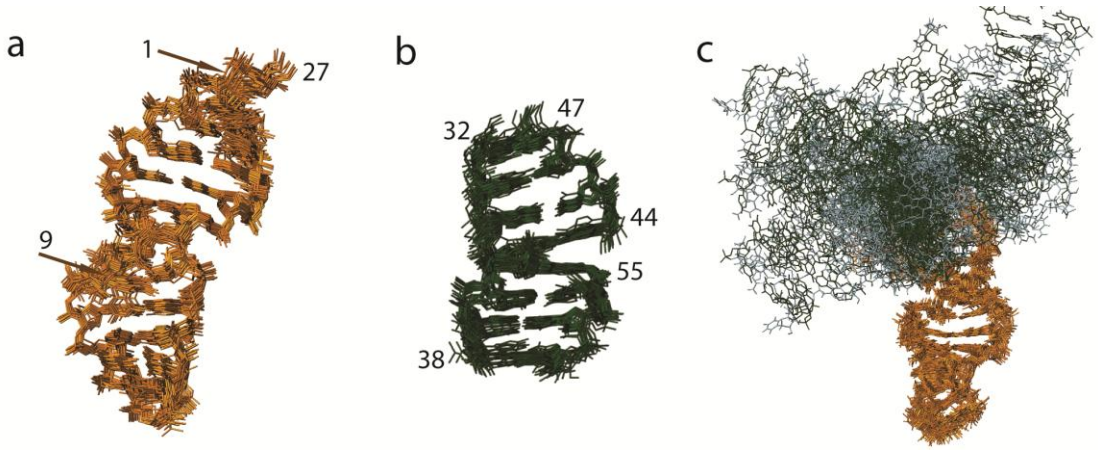

Figure S3.

Superposition of the 18 lowest energy structures of the 2'-5' AG1 lariat forming ribozyme. a. 5' helical segment comprising nucleotides 1-27 (heavy atom root mean square deviation, RMSD = 0.83 Å). Besides the terminal base pair, the position of the bulged out, disordered G9 is indicated. **b.** The two short helical segments, which are part of the 3'-terminal pseudoknot, comprising nucleotides 32-35, 38-41, 44-47 and 55-58 (heavy atom RMSD = 0.97 Å). The two helical segments are stacked upon each other. **c.** Ensemble of the 18 lowest energy structure of the full 2'-5' AG1 lariat forming ribozyme superimposed on the segment 1-27. The RNA folds in two separate domains comprising nucleotides 1-27 (helical segment, Fig. 3b) and 32-58 (pseudoknot, Fig. 3d). The relative position of the two domains is variable. Color code: nucleotides 1-27, orange; 32-35, 38-41, 44-47 and 55-58, green; all others, gray.

Supplementary References

(1) Hennig, M.; Williamson, J. R. *Nucleic Acids Res.* **2000**, *28*, 1585–1593.

(2) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293–8297.

(3) Sklenar, V.; Dieckmann, T.; Butcher, S. E.; Feigon, J. *J. Magn. Reson.* **1998**, *130*, 119–124.

(4) Hu, W. D.; Kakalis, L. T.; Jiang, L. C.; Jiang, F.; Ye, X. M.; Majumdar, A. *J. Biomol. NMR* **1998**, *12*, 559–564.

(5) Zwahlen, C.; Legault, P.; Vincent, S. J. F.; Greenblatt, J.; Konrat, R.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119*, 6711–6721.

(6) Brutscher, B.; Boisbouvier, J.; Kupče, E.; Tisné, C.; Dardel, F.; Marion, D.; Simorre, J.-P. *J. Biomol. NMR* **2001**, *19*, 141–151.

(7) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallographica Section D Biological Crystallography* **1998**, *54*, 905–921.

(8) Linge, J. P.; Habeck, M.; Rieping, W.; Nilges, M. *Bioinformatics* **2003**, *19*, 315–316.

(9) Ohlenschläger, O.; Haumann, S.; Ramachandran, R.; Görlach, M. *J. Biomol. NMR* **2008**, *42*, 139–142.

(10) Linge, J. P.; Williams, M. A.; Spronk, C. A. E. M.; Bonvin, A. M. J. J.; Nilges, M. *Proteins* **2003**, *50*, 496–506.

(11) Ferner, J.; Suhartono, M.; Breitung, S.; Jonker, H. R. A.; Hennig, M.; Wöhnert, J.; Göbel, M.; Schwalbe, H. *Chembiochem* **2009**, *10*, 1490–4.

(12) Nilges, M.; Bernard, A.; Bardiaux, B.; Malliavin, T.; Habeck, M.; Rieping, W. *Structure* **2008**, *16*, 1305–12.

(13) Koradi, R.; Billeter, M.; Wüthrich, K. *Journal of Molecular Graphics* **1996**, *14*, 51–55.

(14) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; others *University of California, San Francisco* **2010**, *142*.

(15) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.

(16) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–25.

(17) Meagher, K. L.; Redman, L. T.; Carlson, H. A. *J. Comput. Chem.* **2003**, *24*, 1016–25.

(18) Cannon, J. F. *J. Comput. Chem.* **1993**, *14*, 995–1005.

(19) Grant, M. B. *Bioinformatics* **2006**, *22*, 2695–2696.

(20) Team R Core; Core, T. R. *R: A Language and Environment for Statistical Computing*; 2012.

(21) Allner, O.; Nilsson, L.; Villa, A. *J. Chem. Theory Comput.* **2012**, *8*, 1493–1502.

(22) Hashem, Y.; Auffinger, P. *Methods (San Diego, Calif.)* **2009**, *47*, 187–97.