Supporting information

Target flexibility in RNA-ligand docking modeled by elastic potential grids

Dennis M. Krüger[a], Johannes Bergs[a][b], Sina Kazemi[a], Holger Gohlke*[a][b]

[a]Department of Mathematics and Natural Sciences
Institute of Pharmaceutical and Medicinal Chemistry
Heinrich-Heine-University, Düsseldorf Germany
[b]Department of Biological Sciences
Institute of Cell Biology and Neurosciences
Goethe-University, Frankfurt, Germany

*Universitätsstr.1, 40225 Düsseldorf (Germany), Fax: (+49) 211-8113847, E-mail: gohlke@uni-duesseldorf.de

RNA structure and ligand preparation for docking

The coordinates from the RNA complexes were retrieved from the Protein Data Bank (PDB).[1] All RNA structures of one type were superimposed with respect to their phosphor atoms. In each case, the first structure of a dataset shown in Table S2 was used as the reference. If there was more than one RNA structure in a PDB file due to symmetry, we chose the structure with the lowest average B-factor for docking. For each complex structure, the ligand was removed, as were all water molecules. Ions were considered as heteroatoms and included into the docking. Ligand coordinates were extracted from the Relibase database[2] together with the assignment of Sybyl atom types. For each ligand, hydrogens were added using AutoDock-Tools (ADT) 1.4.5 except for nonpolar hydrogen atoms. The ligands were further prepared by ADT by assigning aromatic carbon atoms and rotatable bonds.

Docking engine, scoring function, and docking experiments

For the evaluation of our approach, we used Autodock4.1.6[3] as a docking engine. As a scoring function, we used DrugScore^{RNA}.[4] Previously, the knowledge-based pair-potentials had been successfully applied as objective function in docking optimizations.[5] For the DrugScore^{RNA} potential grids, a grid spacing of 0.375 Å and a grid dimension of 61x61x61 Å^3 was used. All DrugScore^{RNA} potential grids were centered on the geometrical center of the ligand found in the crystal/NMR structure. We performed three different docking experiments. I) Every ligand was docked into the bound RNA structure from which it had been extracted ("re-docking"). II) For each dataset, all ligands were docked to the *apo* structure ("*apo*-docking"). III) For each dataset, elastic potential grids computed for the *apo* structure were deformed according to RNA movements from the *apo* to a *holo* structure. Subsequently, each ligand was docked into the deformed grids. Since Autodock does not support irregular grids for docking, the deformed grids were mapped back to a regular grid using the interpolation method described in ref. [6]. This last docking experiment is referred to as "docking into deformed grids".

Docking parameters

Docking was performed using the Lamarckian genetic algorithm[7] in Autodock4.1.6. Each docking experiment was repeated 100 times, yielding 100 docked conformations. The following parameters were used for docking: population size: 200; random starting position and conformation; maximal effect of mutation: 2 Å for translation and 50° for rotation; mutation rate: 0.02; crossover rate: 0.80; local search rate: 0.99; maximum number of consecutive successes or failures before a change is made to the stepsize of the local search:[8] 20. Docking simulations were performed with a maximum of 50.000 generations. The number of energy evaluations was chosen depending on the number of rotatable bonds of the

ligand. Up to 10 rotatable bonds, we used $3*10^6$ evaluations; for more than 10 rotatable bonds, we used $30*10^6$ evaluations. Finally, docked conformations were clustered using a tolerance of 1Å rmsd.

Evaluation of docking accuracy

The docking accuracy was evaluated by calculating the rmsd between docked ligand poses and the experimental ligand configuration. As final docking result, the ligand pose found on the first scoring rank of the largest cluster generated by AutoDock was chosen. This follows the idea that native-like structures reside in a broader energy minimum than non-native structures.[9] This then implies that instead of focusing solely on the conformation of best scoring, the best-scoring structure of the largest cluster of structurally related solutions should be considered the most probable pose.[9-11] The results from re-docking, apo docking, and docking to deformed grids are given in Table S3.

Tables

Table S1: RNA-ligand complexes used for parameter training

RNA type	PDB code[a]	Reference
16S *E. coli* rRNA	1BYJ, 1LC4, 1MWL, 1PBR, 1O9M, 2BE0, 2ESI, 2ET3, 2ET4, 2ET5, 1YRJ, *2I2U*	[12-18]
23S *H. marismortui* rRNA	1VQ8, 1YIT, 1K9M, 1KD1, 1NJI, *1FFK*	[19-22]
16S *T. thermophilus* rRNA	1FJG, 2F4S, 2F4T, 2UUB, *1J5E*	[23-25]
Aptamer RNA	1F1T, *1Q8N*	[26, 27]
HIV-1 TAR RNA	1QD3, 1UUD, *1ANR*	[28-30]
Diels-Alder ribozyme	1YLS, *1YKQ*	[31]
Dimerization initiation site of HIV-1 RNA	2FCY, *1XPE*	[32, 33]
Ribosomal decoding A site of 16S rRNA	2O3X, *2FQN*	[34, 35]
Thi-box riboswitch	2HOJ, *2HOO*	[36]

[a] Apo structures are given in italics. These were deformed to the holo structures given in normal print.

Table S2: RNA-ligand complexes used for evaluation of DrugScore^{RNA}

RNA type	PDB code (ligand)[a]	Rotatable bonds[b]	RMSD[c]	Cluster size[e]	Reference
16S *E. coli* rRNA	1BYJ (ge1-ge2-ge3)	6	1.7	65	[12-18, 37, 38]
	1LC4 (toy)	6	1.2	72	
	1MWL (get)	6	1.0	99	
	1O9M (42d-bdg-p24-pa2)	15	3.0	10	
	1PBR (idg-pa1-pa2-pa3)	9	2.3	13	
	1YRJ (am2)	6	3.0	9	
	2BE0 (js5)	15	2.4	16	
	2BEE (js4)	15	2.2	14	
	2ESI (kan)	6	1.3	43	
	2ESJ (liv)	12	3.2	4	
	2ET3 (lll)	6	1.0	64	
	2ET4 (nmy)	9	2.5	6	
	2ET5 (rio)	6	3.1	21	
	2ET8 (xxx)	3	3.1	29	
	2PWT (lha)	20	3.1	1	
	2I2U				
16S *T. thermophilus* rRNA	1FJG (scm, sry)	2, 9[d]	6.5, 1.3[d]	93 / 36	[23-25, 39, 40]
	2F4S (xxx)	3	3.1	12	
	2F4T (ab9)	13	1.6	22	
	2F4U (ab6)	14	3.1	25	
	2UUB (par)	9	2.0	12	
	1J5E				
23S *H. marismortui* rRNA	1K8A (cai)	14	6.1	5	[20-22, 41, 42]
	1K9M (tyk)	13	3.0	18	
	1KD1 (spr)	11	4.9	22	
	1M90 (aca-pha)	9	9.4	30	
	1NJI (clm)	6	6.1	9	
	1VQ8 (sps)	8	1.6	25	
	1YHQ (zit)	7	4.5	12	
	1YI2 (erx)	7	4.9	18	
	1YIJ (tel)	11	5.7	7	
	1YIT (vir, vrs)	1, 6[d]	1.3, 4.8[d]	86 / 70	
	1YJN (cly)	7	4.2	20	
	1YJW (syb)	10	5.6	24	
	2OTJ (13t)	5	4.7	20	
	2OTL (gir)	3	3.8	61	
	1FFK				
Aptamer RNA	1F1T (ros)	3	0.7	62	[26, 27]
	1Q8N (mgr)	5	0.9	100	
Diels-Alder ribozyme	1YLS (dai)	5	1.3	90	[31]
	1YKQ				
Dimerization initiation site of HIV-1 RNA	2FCX (xxx)	3	4.4	60	
	2FCY (nmy)	9	4.8	8	
	2FCZ (rio)	6	5.0	10	
	2FD0 (liv)	12	3.9	1	
	1XPE				
HIV-1 frameshift site RNA	2JUK (g0b)	15	3.6	1	[43, 44]
	1Z2J				
Ribosomal decoding A-site of 16S rRNA	2G5K (am2)	6	4.0	18	[19, 34, 35, 45, 46]
	2O3V (n33)	8	4.6	11	
	2O3W (par)	9	4.9	8	
	2O3X (n30)	6	4.0	16	
	1FYP (par)	9	3.0	12	
	2FQN				

HIV-1 TAR RNA	1ARJ (arg)	5	4.0	14	[28-30, 47, 48]
	1LVJ (pmz)	5	3.3	52	
	1QD3 (bdg-ccy-rib-idg)	9	2.2	17	
	1UTS (p13)	10	7.0	10	
	1UUD (p14)	12	2.4	21	
	1UUI (p12)	7	6.1	12	
	1ANR				
Phenylalanin tRNA	1I9V (nmy)	9	4.3	6	[49, 50]
	1EVV				
	1EHZ				
Thi-box riboswitch	2HOJ (tpp)	8	2.8	11	[36]
	2HOM (tps)	6	2.4	26	
	2HOO (bft)	12	4.3	5	

[a] The Relibase-ID[2] of the ligand is given in parenthesis. For apo structures, no ligand is specified.
[b] The number of ligand bonds considered rotatable in docking.
[c] Results for re-docking into the holo structure. Rmsd between the docking pose found on the first scoring rank of the largest cluster and the native solution. In Å.
[d] Two ligands are given, see brackets in the second column.
[e] Size of the largest cluster where the docking pose was found on the first scoring rank.

Table S3: Results for docking into apo and holo structures and deformed grids.[a]

RNA type	PDB ID	Apo[b]	Holo[c]	Def. mean [d]	Def. single [e]	Rot. bonds[f]	Resolution [g]	Charged groups [h]	
								+	-
16S *E. coli* rRNA	1BYJ	**2.3**	**1.7**	**1.8**	1.9	6	-	5	0
	1LC4	4.8	**1.2**	**1.5**	1.7	6	2.5	5	0
	1MWL	4.4	**1.0**	**1.4**	1.7	6	2.4	4	0
	1PBR	2.8	**2.3**	**2.5**	2.9	9	-	5	0
	2BE0	3.1	**2.4**	**2.3**	3.6	15	2.6	5	0
	2BEE	2.8	**2.2**	2.9	3.0	15	2.6	6	0
	2ESI	4.5	**1.3**	5.3	5.6	6	3.0	4	0
	2ET3	4.8	**1.0**	**1.2**	1.1	6	2.8	5	0
	2ET4	**2.5**	**2.4**	**2.5**	3.5	9	2.4	6	0
16S *T. thermophilus* rRNA	2F4T	3.9	**1.6**	3.9	3.8	13	3.0	5	1
	2UUB	5.3	**2.0**	5.1	5.0	9	2.8	5	0
Aptamer RNA	1F1T	**1.6**	**0.7**	**0.9**	0.9	3	2.8	0	0
23S *H. marismortui* rRNA	1VQ8	7.0	**1.6**	5.8	7.0	8	2.2	0	0
	1YIT	3.9	**1.3**	5.1	5.1	1	2.8	0	0
HIV-1 TAR RNA	1QD3	**2.4**	**2.2**	**2.1**	2.3	9	-	6	0
	1UUD	6.3	**2.4**	4.5	4.6	12	-	3	0
Thi-box riboswitch	2HOM	3.2	**2.4**	3.0	3.2	6	2.8	2	1

[a] Rmsd between the docking pose found on the first scoring rank of the largest cluster and the native solution. In Å. Values ≤ 2.5 Å are depicted in bold.
[b] Apo docking. PDB codes for the respective apo structures are given in Table S2.
[c] Re-docking.
[d] Docking into deformed grids using $k/G = 30.0$ for each structure.
[e] Docking into deformed grids using complex-specific k/G values as depicted in Figure S1.
[f] The number of rotatable ligand bonds considered in docking.
[g] In Å. No resolution value is given in case of an NMR structure.
[h] Number of positively (+) and negatively (-) charged functional groups in the ligand, assuming pH = 7 when determining protonation states.

Figures

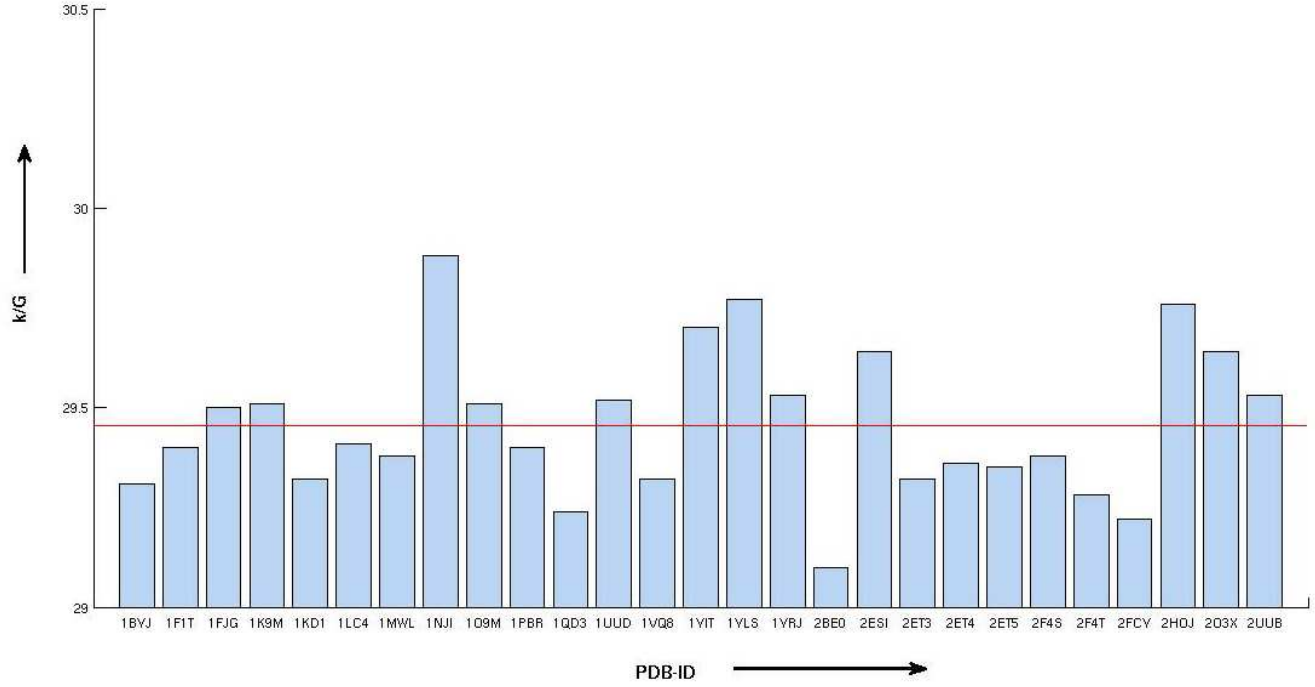

Figure S1: Ratio *k*/*G* obtained during the parameterization procedure for different training runs depicted on the abscissa. For each RNA type (Table S1), the PDB codes of the starting RNA structures are given. The average value of *k*/*G* = 29.45 ± 0.19 is depicted as a red line.

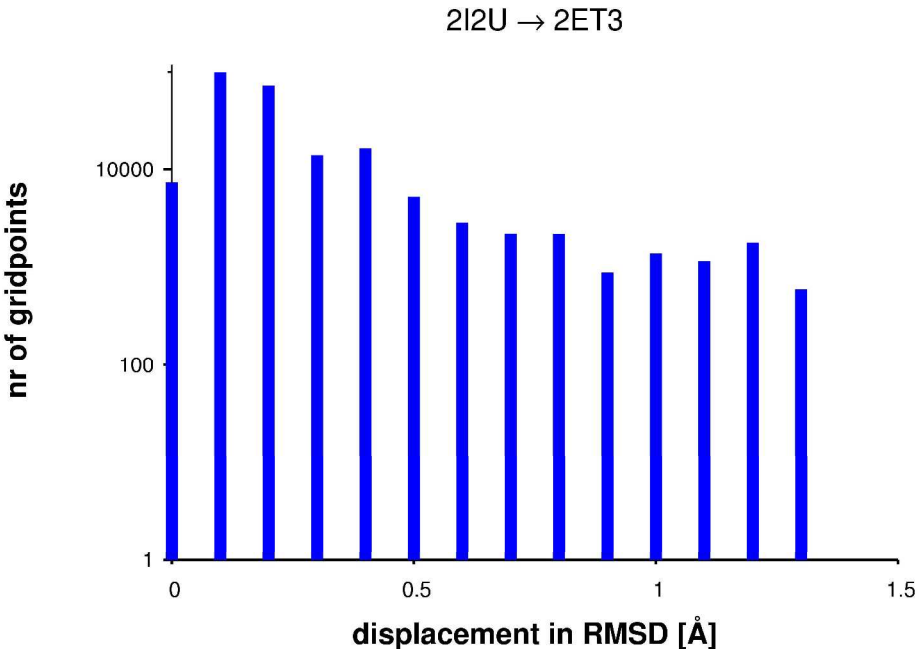

Figure S2: Frequency distribution of grid point displacements obtained for potential grids that were deformed from the apo structure (PDB code: 2I2U) to the bound structure (PDB code: 2ET3) of 16S rRNA.

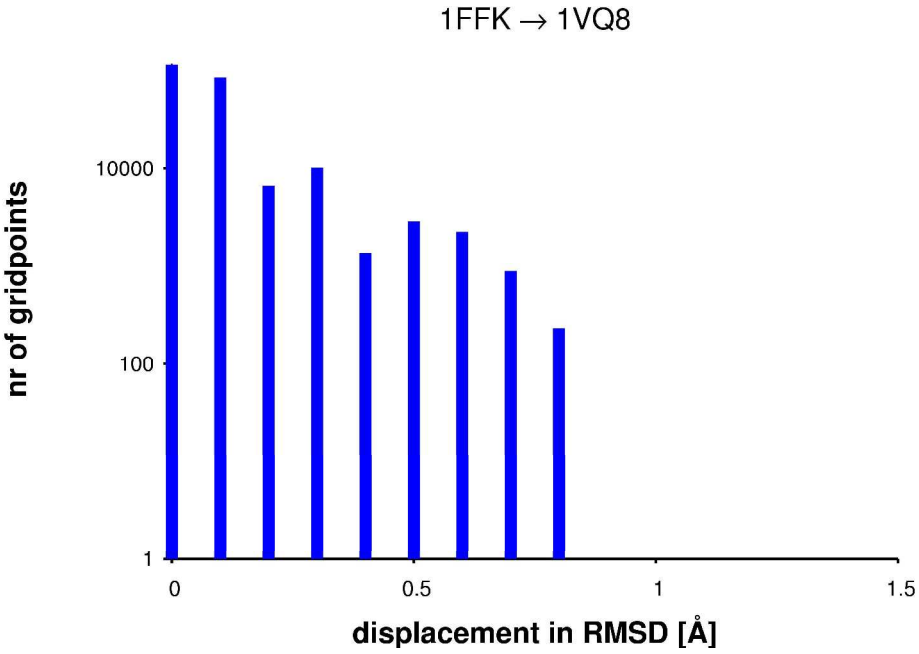

Figure S3: Frequency distribution of grid point displacements obtained for the potential grids that were deformed from the apo structure (PDB code: 1FFK) to the bound structure (PDB code: 1VQ8) of 23S rRNA.

References

[1] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *Arch Biochem Biophys* **1978**, *185*, 584.
[2] M. Hendlich, A. Bergner, J. Gunther, G. Klebe, *J Mol Biol* **2003**, *326*, 607.
[3] R. Huey, G. M. Morris, A. J. Olson, D. S. Goodsell, *J Comput Chem* **2007**, *28*, 1145.
[4] P. Pfeffer, H. Gohlke, *J Chem Inf Model* **2007**, *47*, 1868.
[5] C. A. Sotriffer, H. Gohlke, G. Klebe, *J Med Chem* **2002**, *45*, 1967.
[6] S. Kazemi, D. M. Kruger, F. Sirockin, H. Gohlke, *ChemMedChem* **2009**, *4*, 1264.
[7] G. M. Morris, Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. & Olson, A. J. , *Journal of Computational Chemistry* **1998**, *19*, 1639.
[8] F. J. Solis, R. J.-B. Wets, *Math. Oper. Res.* **1981**, *6*, 19.
[9] D. Shortle, K. T. Simons, D. Baker, *Proc Natl Acad Sci U S A* **1998**, *95*, 11158.
[10] Y. Zhang, J. Skolnick, *J Comput Chem* **2004**, *25*, 865.
[11] C. Wendel, H. Gohlke, *Proteins* **2008**, *70*, 984.
[12] B. Francois, R. J. Russell, J. B. Murray, F. Aboul-ela, B. Masquida, Q. Vicens, E. Westhof, *Nucleic Acids Res* **2005**, *33*, 5677.
[13] S. Yoshizawa, D. Fourmy, J. D. Puglisi, *Embo J* **1998**, *17*, 6437.
[14] Q. Vicens, E. Westhof, *Chem Biol* **2002**, *9*, 747.
[15] R. J. Russell, J. B. Murray, G. Lentzen, J. Haddad, S. Mobashery, *J Am Chem Soc* **2003**, *125*, 3410.
[16] D. Fourmy, M. I. Recht, S. C. Blanchard, J. D. Puglisi, *Science* **1996**, *274*, 1367.
[17] B. Francois, J. Szychowski, S. S. Adhikari, K. Pachamuthu, E. E. Swayze, R. H. Griffey, M. T. Migawa, E. Westhof, S. Hanessian, *Angew Chem Int Ed Engl* **2004**, *43*, 6735.
[18] V. Berk, W. Zhang, R. D. Pai, J. H. Cate, *Proc Natl Acad Sci U S A* **2006**, *103*, 15830.
[19] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 905.
[20] J. L. Hansen, J. A. Ippolito, N. Ban, P. Nissen, P. B. Moore, T. A. Steitz, *Mol Cell* **2002**, *10*, 117.
[21] T. M. Schmeing, A. C. Seila, J. L. Hansen, B. Freeborn, J. K. Soukup, S. A. Scaringe, S. A. Strobel, P. B. Moore, T. A. Steitz, *Nat Struct Biol* **2002**, *9*, 225.
[22] D. Tu, G. Blaha, P. B. Moore, T. A. Steitz, *Cell* **2005**, *121*, 257.
[23] A. P. Carter, W. M. Clemons, D. E. Brodersen, R. J. Morgan-Warren, B. T. Wimberly, V. Ramakrishnan, *Nature* **2000**, *407*, 340.
[24] J. B. Murray, S. O. Meroueh, R. J. Russell, G. Lentzen, J. Haddad, S. Mobashery, *Chem Biol* **2006**, *13*, 129.
[25] A. Weixlbaumer, F. V. t. Murphy, A. Dziergowska, A. Malkiewicz, F. A. Vendeix, P. F. Agris, V. Ramakrishnan, *Nat Struct Mol Biol* **2007**, *14*, 498.
[26] C. Baugh, D. Grate, C. Wilson, *J Mol Biol* **2000**, *301*, 117.
[27] J. Flinders, S. C. DeFina, D. M. Brackett, C. Baugh, C. Wilson, T. Dieckmann, *Chembiochem* **2004**, *5*, 62.
[28] F. Aboul-ela, J. Karn, G. Varani, *Nucleic Acids Res* **1996**, *24*, 3974.
[29] C. Faber, H. Sticht, K. Schweimer, P. Rosch, *J Biol Chem* **2000**, *275*, 20660.
[30] B. Davis, M. Afshar, G. Varani, A. I. Murchie, J. Karn, G. Lentzen, M. Drysdale, J. Bower, A. J. Potter, I. D. Starkey, T. Swarbrick, F. Aboul-ela, *J Mol Biol* **2004**, *336*, 343.
[31] A. Serganov, S. Keiper, L. Malinina, V. Tereshko, E. Skripkin, C. Hobartner, A. Polonskaia, A. T. Phan, R. Wombacher, R. Micura, Z. Dauter, A. Jaschke, D. J. Patel, *Nat Struct Mol Biol* **2005**, *12*, 218.
[32] E. Ennifar, P. Dumas, *J Mol Biol* **2006**, *356*, 771.
[33] E. Ennifar, J. C. Paillart, A. Bodlenner, P. Walter, J. M. Weibel, A. M. Aubertin, P. Pale, P. Dumas, R. Marquet, *Nucleic Acids Res* **2006**, *34*, 2328.
[34] J. Kondo, A. Urzhumtsev, E. Westhof, *Nucleic Acids Res* **2006**, *34*, 676.

[35] J. Kondo, M. Hainrichson, I. Nudelman, D. Shallom-Shezifi, C. M. Barbieri, D. S. Pilch, E. Westhof, T. Baasov, *Chembiochem* **2007**, *8*, 1700.
[36] T. E. Edwards, A. R. Ferre-D'Amare, *Structure* **2006**, *14*, 1459.
[37] Q. Han, Q. Zhao, S. Fish, K. B. Simonsen, D. Vourloumis, J. M. Froelich, D. Wall, T. Hermann, *Angew Chem Int Ed Engl* **2005**, *44*, 2694.
[38] J. Kondo, K. Pachamuthu, B. Francois, J. Szychowski, S. Hanessian, E. Westhof, *ChemMedChem* **2007**, *2*, 1631.
[39] D. E. Brodersen, W. M. Clemons, Jr., A. P. Carter, R. J. Morgan-Warren, B. T. Wimberly, V. Ramakrishnan, *Cell* **2000**, *103*, 1143.
[40] B. T. Wimberly, D. E. Brodersen, W. M. Clemons, Jr., R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch, V. Ramakrishnan, *Nature* **2000**, *407*, 327.
[41] J. L. Hansen, P. B. Moore, T. A. Steitz, *J Mol Biol* **2003**, *330*, 1061.
[42] S. J. Schroeder, G. Blaha, J. Tirado-Rives, T. A. Steitz, P. B. Moore, *J Mol Biol* **2007**, *367*, 1471.
[43] D. W. Staple, S. E. Butcher, *J Mol Biol* **2005**, *349*, 1011.
[44] D. W. Staple, V. Venditti, N. Niccolai, L. Elson-Schwab, Y. Tor, S. E. Butcher, *Chembiochem* **2008**, *9*, 93.
[45] J. Kondo, B. Francois, A. Urzhumtsev, E. Westhof, *Angew Chem Int Ed Engl* **2006**, *45*, 3310.
[46] S. R. Lynch, J. D. Puglisi, *J Mol Biol* **2001**, *306*, 1023.
[47] F. Aboul-ela, J. Karn, G. Varani, *J Mol Biol* **1995**, *253*, 313.
[48] M. Lerch, V. Gafner, R. Bader, B. Christen, G. Folkers, O. Zerbe, *J Mol Biol* **2002**, *322*, 1117.
[49] H. Shi, P. B. Moore, *Rna* **2000**, *6*, 1091.
[50] L. Jovine, S. Djordjevic, D. Rhodes, *J Mol Biol* **2000**, *301*, 401.