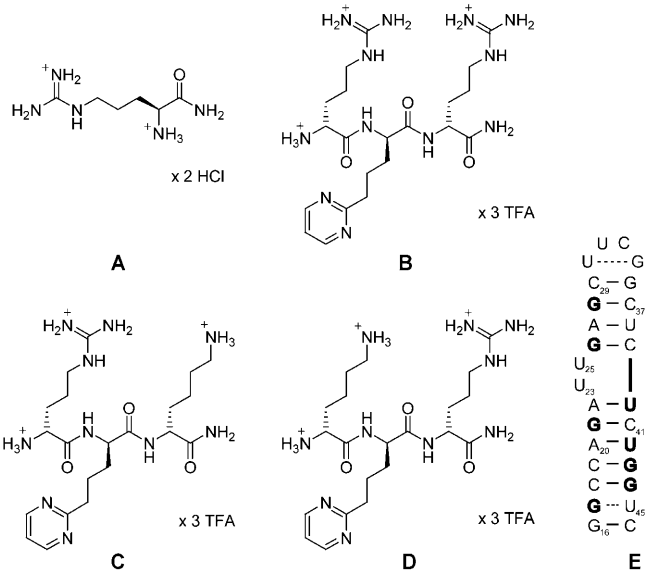
DOI: 10.1002/cbic.200900220

Structures of HIV TAR RNA–Ligand Complexes Reveal Higher Binding Stoichiometries

Jan Ferner,[a] Marcel Suhartono,[b] Sven Breitung,[b] Hendrik R. A. Jonker,[a] Mirko Hennig,[c] Jens Wöhnert,[d] Michael Göbel,*[b] and Harald Schwalbe*[a]

The development of low molecular weight ligands that bind to RNA with high affinity and specificity remains challenging.[1] The trans-activation response element TAR is regarded as an important RNA target to control the replication cycle of the human immunodeficiency virus (HIV). In particular, the internal bulge within the apical hairpin of TAR (**E**′ in Scheme S1 in the Supporting Information) has been investigated widely as a drug target because its interaction with the trans-activator protein (Tat) is essential for the transcription of viral proteins.[2] The hairpin contains a six nucleotide loop and an internal bulge. The size of the bulge is the only difference between HIV-1 and HIV-2 strains. HIV-1 TAR has a trinucleotide bulge with the sequence U23-C24-U25 while HIV-2 TAR lacks the central cytidine.

The bulge functions as a flexible linker that leads to substantial angular fluctuation of the two canonical stems.[3] Initially, the complex of TAR with argininamide (Scheme 1 **A**) served as a mimic of the complex between TAR and Tat and provided important insights into the conformation of TAR in the protein–RNA complex. Argininamide binds with a low millimolar dissociation constant in the bulge region.[4] In HIV-1[5] and HIV-2[6] TAR, argininamide induces the formation of a base triple, which is formed between the bulge residue U23 and the canonical A27-U38 base pair. In conjunction with G26, this region also displays the most contacts to argininamide.[4,7] Complexes of TAR with several further ligands have been investigated including Tat-derived peptides,[8] ligands with heteroaromatic moieties,[9] guanidinium-like ligands,[10] aminoglycosides[11] as well as divalent cations.[12] Almost all ligands bind in the bulge region and therefore compete directly with Tat. Despite different recognition modes, all ligands have been reported to rigidify the interhelical angle.[3a,13] Recently, Zhang et al. showed that the conformation of free TAR represents a dynamic ensemble that samples all possible interhelical angles detected in the various RNA–ligand structures. Ligands therefore select and stabilize free state conformations rather than induce new ones.[14]



Scheme 1. Constitution of the ligands (**A**–**D**) and secondary structure of the HIV-2 TAR RNA (**E**) investigated in this study. Residues of the RNA whose CSPs are presented in Figure 1 are highlighted.

In our investigations, we were interested whether a ligand could occupy two binding sites present in the ensemble of free state structures. Although many TAR–ligand complexes are formed with 1:1 stoichiometry, one of the first structural studies of TAR–ligand complexes in 1995 provided evidence that the tightly binding peptide ligand ADP-1 binds to two distinct sites of HIV-1 TAR.[15] Very recently, a cyclic aminoglycoside analogue was shown to bind TAR with RNA/ligand-stoichiometry of 2:3.[11b]

Here, we report the structural characterization of complexes between TAR and novel peptidic ligands developed by Göbel and co-workers. These compounds block Tat–TAR association in vitro and attenuate HIV proliferation in cell cultures.[16] The tripeptides contain two D-arginines flanking a non-natural amino acid with a heteroaromatic side chain and two variants with D-lysine replacing D-arginine (Scheme 1 **B**–**D**). The ligands were designed to contain a heteroaromatic amino acid to provide hydrophobic stacking interactions in addition to the elec-

[a] *J. Ferner, Dr. H. R. A. Jonker, Prof. Dr. H. Schwalbe*
Institut für Organische Chemie und Chemische Biologie
Zentrum für Biomolekulare Magnetische Resonanz (BMRZ)
Johann Wolfgang Goethe-Universität Frankfurt am Main
Max-von-Laue-Strasse 7, 60438 Frankfurt am Main (Germany)
Fax: (+49) 69-798-29515
E-mail: schwalbe@nmr.uni-frankfurt.de

[b] *Dr. M. Suhartono, S. Breitung, Prof. Dr. M. Göbel*
Institut für Organische Chemie und Chemische Biologie
Johann Wolfgang Goethe-Universität Frankfurt am Main
Max-von-Laue-Strasse 7, 60438 Frankfurt am Main (Germany)

[c] *Prof. Dr. M. Hennig*
Department of Biochemistry and Molecular Biology,
Medical University of South Carolina
173 Ashley Avenue, PO Box 250509, Charleston, SC 29425 (USA)

[d] *Prof. Dr. J. Wöhnert*
Institut für Molekulare Biowissenschaften
Zentrum für Biomolekulare Magnetische Resonanz (BMRZ)
Johann Wolfgang Goethe-Universität Frankfurt am Main
Max-von-Laue-Strasse 9, 60438 Frankfurt am Main (Germany)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.200900220.

trostatic attraction exerted by the two arginine side chains. Low micromolar IC_{50} values have been observed in competition experiments with a dye-labeled Tat-peptide. The affinities to TAR RNA result from specific binding to the bulge region in combination with weaker interactions in at least one alternative binding site. Our NMR data provide a structural model and reveal that ligands in excess, with similar constitution to **B** can bind with higher stoichiometries in line with previous mass spectrometry data.[17] Our findings of higher binding stoichiometries are not unprecedented; in fact, similar ligands have been reported to exhibit higher binding stoichiometries at ligand excess.[9d,e] Structural models, however, have not been reported.

We performed a titration of argininamide **A** to HIV-2 TAR (**E**) that resulted in continuous chemical shift perturbations (CSPs) of the imino ^{1}H NMR resonances around the bulge up to a saturation level and indicated binding in the fast exchange limit (Figure 1 A); residues further away show marginal CSPs upon addition of ligand (Figure S1 A). The imino resonance of residue U40 is not detectable in the free RNA but can be detected upon addition of argininamide due to acquired protection against exchange with the solvent water. In agreement with Brodsky et al.,[7] argininamide binds directly below the bulge, selects a single RNA conformation from the dynamic ensemble of structures in the apo state, and stabilizes the A22-U40 base pair and the base triple of U23-A27-U38.

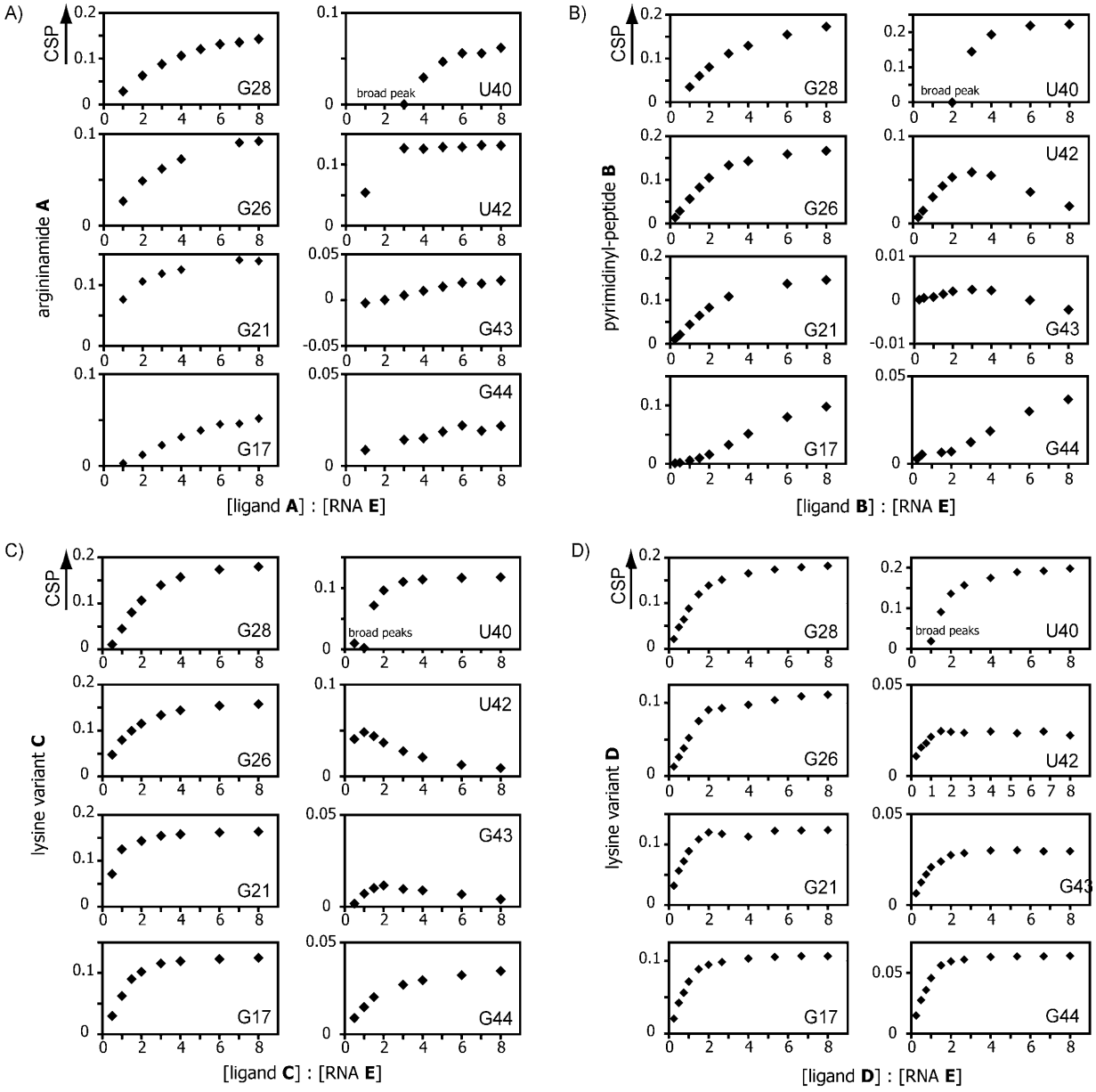


Figure 1. Chemical shift perturbations (CSPs) of the imino protons of RNA **E** as a function of the [ligand]/[RNA] ratio for the ligands A) argininamide, B) the pyrimidinyl-peptide and C) and D) the lysine variants.

For the newly designed ligand **B**, a more complex titration behavior is observed (Figure 1 B). The CSPs are substantially larger (by a factor of two) than those observed with argininamide. Interestingly, the NMR chemical shift response changes in the course of titration. At intermediate [ligand]/[RNA] ratio, the CSPs are the largest. Up to a ratio of 2:1, the observed shift changes resemble the argininamide titration profile. At higher [ligand]/[RNA] ratios, the so far unaffected signals of the imino protons of residues G17 and G44 start to change. Concomitantly, the imino resonances of U42 and G43 start to move into the opposite direction. A similar behavior is observed for the CSPs of the H5/H6 cross peaks detected in TOCSY spectra during the titration (Figure S2). While the residues in and around the bulge (U23, U25, C39 and U40) change their chemical shifts of H5 and H6 significantly at the beginning of the titration, resonances of residues C18, C41, U42, C45 either start to change or the CSPs redirect at a [ligand]/[RNA] ratio above 2:1. Taken together, this indicates the formation of a distinct 1:1 complex at intermediate [ligand]/[RNA] ratios and occupation of a second, weaker binding site for ligand **B** within the lower stem of TAR at higher [ligand]/[RNA] ratios.

In order to calculate the structure of the complex, we investigated a TAR construct containing a UUCG-tetraloop (Scheme 1 **E**) due to its favourable NMR characteristics. The ligand binding characteristics of this RNA remain unaltered. Structure calculations assuming a 1:1 complex resulted in two different structural models in which the ligand either binds in the major groove of the bulge similar to argininamide[4] and ligand rbt203 from Davis et al.[9d] or on the opposing side similar to the binding site of aminoglycosides such as neomycin B.[11a] In both structural models, the ligand is in close proximity to the nucleobase of residue U23 to which it shows strong intermolecular NOEs. The orientation of this bulge residue is different in the two models (Figure S3). For both 1:1 complexes, however, nine out of 48 intermolecular distance restraints localized in the lower stem are severely violated and the convergence rate of the structure calculations is low. We further analysed the subnanosecond dynamics of the bulge by using ^{13}C relaxation data.[18] The dynamics revealed that only nucleobase U25 remains flexible upon addition of the ligand while U23 is as rigid as the stem residues (Figure S4). Since neither doubling of the resonances nor peculiar dynamics of residue U23 could be detected, we conclude that the solution structure at saturation levels of ligand can neither be described by both structures nor by a rapid conformational equilibrium between the two structures.

Therefore, we performed NMR-based structural calculations assuming a 2:1 complex stoichiometry in agreement with our previous mass spectrometry results of related compounds.[17] In the ternary complex residue U23 is located in the major groove close to the base pairs G26-C39 and A27-U38 (Figure 2). However, no evidence for the formation of the base triple U23-A27-U38 could be obtained experimentally (data not shown). For the first ligand molecule with a larger number of stronger intermolecular NOEs, binding around the nucleobase of U23 is observed with the N-terminal Arg1 side chain below and the C-terminal Arg3 side chain above the plane of the base. This arrangement of the guanidinium groups is very similar to the 1:1 TAR-complex with rbt203.[9d] The N-terminal arginine is structurally well defined, and its guanidinium group is in close proximity to the Hoogsteen side of G26 to which it can form hydrogen bonds with the acceptor atoms N7 and O6. The more flexible C-terminal arginine shows transient interactions with the nucleobase of A27 and therefore prevents the contact of U23 and A27. This observation is in agreement with the absence of a base triplet in the ternary complex. The central pyrimidine residue points out to the solution and does not interact with the RNA. The second ligand molecule is localized in the major groove of the bottom stem along the pyrimidine stretch U40-C41-U42 but its structure is less defined. The intermolecular contacts are exclusively mediated by the two arginine residues. Occupation of this second binding site has been previously reported for other TAR complexes with structurally unrelated ligands.[11b,15]

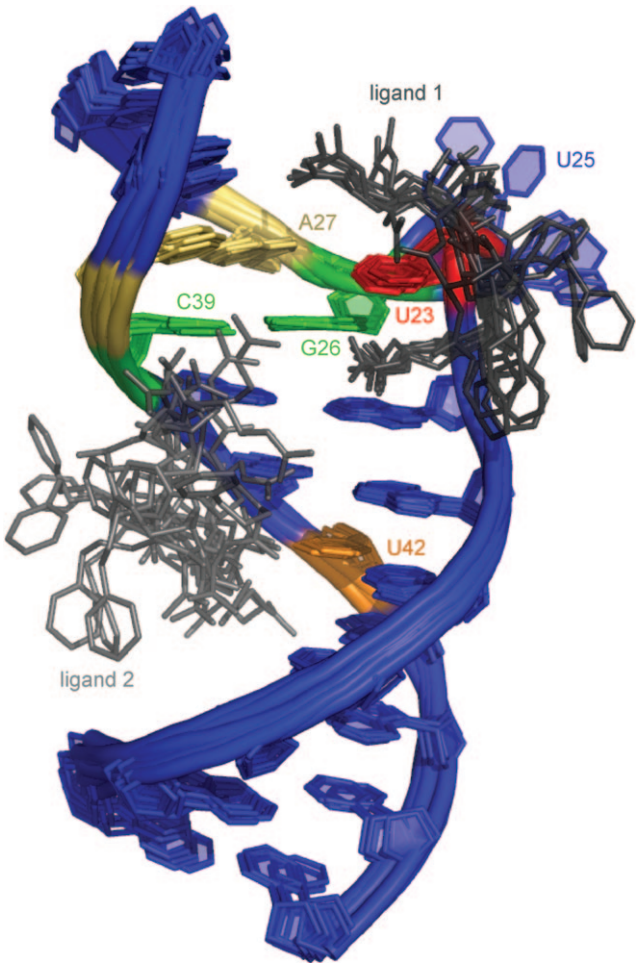


Figure 2. Ten best structures of the ternary complex (red = U23, green = G26-C39, yellow = A27-U38, orange = U42, black = ligand 1, gray = ligand 2).

Based on these results and in order to test whether specificity could be improved without compromising affinity, two additional tripeptides were synthesized and investigated. Replacement of the C-terminal Arg3 with lysine (Scheme 1 **C**) improved the binding affinity reducing the IC_{50} value by a factor of four. Importantly, it showed identical patterns of CSPs during the

NMR titration when compared to **B**. In particular, U42 and G43 represent the region where the effects of the two binding sites are accentuated. Since the influence of the second ligand binding occurs already at a lower [ligand]/[RNA] ratio (Figure 1 C), the first, higher-affinity binding site in the bulge is saturated at an earlier stage: in agreement with the observed lower IC_{50}, both affinity and specificity for the first binding site are increased. The replacement of the N-terminal arginine with a lysine (Scheme 1 **D**) leads to a slight increase of the IC_{50} value and does not show the characteristics in the NMR titration as observed with ligands **B** or **C**. This observation points to the importance of an N-terminal arginine residue for higher affinity to the binding site in the bulge.

From a structural point of view, the role of the heteroaromatic side chain in the central residue of the tripeptide remains undefined. The lack of intermolecular NOEs is likely due to the absence of a persistent direct contact of the heteroaromatic side chain with the target RNA. Substitution with amino acids containing other aromatic side chains (for example, phenanthrene or pyrazinyl) results in similar binding affinities and influences on the imino resonance shifts but leads to a broadening of the imino NMR signals (Supporting Information, Figure S5).

We conclude that the high affinity of our novel TAR-ligands is based on two properties:

1) The terminal residues clasp around the nucleobase of residue U23 in the bulge capable to form hydrogen bonds to the nucleobases of G26 and A27. A C-terminal lysine is favored since its side chain amine group can bind to the N7 of A27 but experiences less repulsion from the adenine amine group compared to a guanidinium group of an arginine. Additionally, cation–π interactions between the arginine guanidinium group and nucleobases are likely determinants that stabilize this binding. The central amino acid acts as spacer. Its aromatic side chain has an indirect influence on the complexation by its effects on ligand solubility, dynamics and probably entropic contributions to binding.
2) Ligands with arginine side chains bind to at least one additional site on TAR. This site is localized along the pyrimidine stretch U40-C41-U42 directly below the bulge adjacent to residue G43 and has previously been observed in the context of ligand-bound HIV-1 and HIV-2 TAR.[11b,15] The specificity is surprising since this region is expected to adopt a standard double helical A-RNA structure. It is interesting to note that Varani and co-workers have recently detected subnanosecond dynamics for the stem-residue G43 in free HIV-1 TAR,[19] a dynamical feature that might be linked to the lower stems capability to bind additional ligand molecules.

In conclusion, NMR spectroscopy revealed two binding sites of varying affinity for peptidic ligands containing non-natural amino acids on TAR RNA. We wish to discuss our findings here in the context of the netropsin-DNA complex by Wemmer and co-workers, a seminal finding changing further development of specific DNA binders.[20] The NMR data in this case revealed a 2:1 binding stoichiometry in the ligand–DNA complex, different to proposed models at the time. In our case, we find that the bulge region of TAR RNA represents the major determinant for selectivity; affinity, however, can be further increased by additional positive charges targeting the lower stem of TAR. The detection of multiple binding sites, although some are weak, may inspire novel medicinal chemistry and enable the exploitation of fragment-based ligand design to target HIV TAR RNA.[21]

Experimental Section

Titrations of RNA (**E** and **E′**) with the ligands were monitored by 1D ^{1}H NMR with RNA (150 μM) in a buffer (at pH 6.2) containing K_3PO_4 (25 mM) and KCl (50 mM) and ligand (up to eightfold excess).

For resonance assignment and structure determination, the following NMR experiments were measured in a RNA sample (**E**, 0.65 mM) with eightfold excess of ligand **B**: normal and constant time ^{1}H,^{13}C HSQC, ^{1}H,^{15}N HSQC, ^{2}J ^{1}H,^{15}N HSQC,[22] HNN-COSY,[23] H5NN-COSY,[24] 2D H(C)N,[25] 3D ^{13}C-edited NOESY-HSQC, 3D ^{15}N-edited NOESY-HSQC, 3D HCCH-COSY, 3D HCCH-TOCSY,[26] 3D forward-directed HCC-TOCSY-CCH E.COSY,[27] 2D ^{15}N-edited CPMG-NOESY,[28] 2D ^{1}H^{1}H-NOESY.[29] To cross-validate the structures, residual dipolar couplings (RDC) measured with IPAP-HSQCs and with pf1 phages as aligning medium were used.[30] All experimental data were processed with Topspin 1.3 (Bruker, Germany) and analyzed with Sparky 3.114 (T. D. Goddard and D. G. Keller, UCSF, USA).

Calculations of the RNA–complex structure were performed with CNS 1.1[31] by using the ARIA 1.2 setup and protocols.[32] The dna-rna-allatom force field was used with OPLS parameters[33] and restraints are summarized in Table S1. Modelling of the complexes with 1:1 and 1:2 stoichiometries (RNA/ligand) were achieved using a high ambiguity driven docking approach with the program HADDOCK 2.1.[34] The ambiguous interaction restraints (AIRs) were defined from intermolecular NOEs which were classified by their intensity into either strong (distance closer than 4 Å), medium (distance closer than 5 Å) and weak (distance closer than 6 Å).

Acknowledgements

H.S. is member of the DFG-funded Center of Excellence: Macromolecular Complexes. The work was supported by the state of Hessen (BMRZ), the Sonderforschungsbereich 579 "RNA–Ligand-Interactions" and the European Commission (SpineII and eNMR).

Keywords: HIV TAR ligands · NMR spectroscopy · peptides · RNA structures · RNA

[1] a) H. Schwalbe, M. W. Göbel, G. Wess, *ChemBioChem* **2003**, *4*, 915–917; b) J. R. Thomas, P. J. Hergenrother, *Chem. Rev.* **2008**, *108*, 1171–1224.
[2] a) B. Berkhout, R. H. Silverman, K. T. Jeang, *Cell* **1989**, *59*, 273–282; b) J. Karn, *J. Mol. Biol.* **1999**, *293*, 235–254.
[3] a) K. T. Dayie, A. S. Brodsky, J. R. Williamson, *J. Mol. Biol.* **2002**, *317*, 263–278; b) Q. Zhang, A. C. Stelzer, C. K. Fischer, H. M. Al-Hashimi, *Nature* **2007**, *450*, 1263–1267.
[4] J. Tao, A. D. Frankel, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2723–2726.
[5] J. D. Puglisi, R. Tan, B. J. Calnan, A. D. Frankel, J. R. Williamson, *Science* **1992**, *257*, 76–80.
[6] A. S. Brodsky, H. A. Erlacher, J. R. Williamson, *Nucleic Acids Res.* **1998**, *26*, 1991–1995.
[7] A. S. Brodsky, J. R. Williamson, *J. Mol. Biol.* **1997**, *267*, 624–639.

[8] a) K. M. Weeks, D. M. Crothers, *Cell* **1991**, *66*, 577–588; b) M. J. Churcher, C. Lamont, F. Hamy, C. Dingwall, S. M. Green, A. D. Lowe, P. J. G. Butler, M. J. Gait, J. Karn, *J. Mol. Biol.* **1993**, *230*, 90–110; c) Z. Athanassiou, R. L. A. Dias, K. Moehle, N. Dobson, G. Varani, J. A. Robinson, *J. Am. Chem. Soc.* **2004**, *126*, 6906–6913.
[9] a) L. Dassonneville, F. Hamy, P. Colson, C. Houssier, C. Bailly, *Nucleic Acids Res.* **1997**, *25*, 4487–4492; b) N. Gelus, F. Hamy, C. Bailly, *Bioorg. Med. Chem.* **1999**, *7*, 1075–1079; c) K. E. Lind, Z. Du, K. Fujinaga, B. M. Peterlin, T. L. James, *Chem. Biol.* **2002**, *9*, 185–193; d) B. Davis, *J. Mol. Biol.* **2004**, *336*, 343–356; e) A. I. H. Murchie, *J. Mol. Biol.* **2004**, *336*, 625–638.
[10] A. Lapidot, E. Ben-Asher, M. Eisenstein, *FEBS Lett.* **1995**, *367*, 33–38.
[11] a) C. Faber, H. Sticht, K. Schweimer, P. Rösch, *J. Biol. Chem.* **2000**, *275*, 20660–20666; b) D. Raghunathan, V. M. Sánchez-Pedregal, J. Junker, C. Schwiegk, M. Kalesse, A. Kirschning, T. Carlomagno, *Nucleic Acids Res.* **2006**, *34*, 3599–3608.
[12] a) J. A. Ippolito, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9819–9824; b) T. E. Edwards, S. T. Sigurdsson, *Biochem. Biophys. Res. Commun.* **2003**, *303*, 721–725.
[13] M. Zacharias, P. J. Hagerman, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6052–6056.
[14] Q. Zhang, X. Sun, E. D. Watt, H. M. Al-Hashimi, *Science* **2006**, *311*, 653–656.
[15] F. Aboul-ela, J. Karn, G. Varani, *J. Mol. Biol.* **1995**, *253*, 313–332.
[16] a) A. Krebs, V. Ludwig, J. Pfizer, G. Dürner, M. W. Göbel, *Chem. Eur. J.* **2004**, *10*, 544–553; b) V. Ludwig, A. Krebs, M. Stoll, U. Dietrich, J. Ferner, H. Schwalbe, U. Scheffer, G. Dürner, M. W. Göbel, *ChemBioChem* **2007**, *8*, 1850–1856.
[17] N. Morgner, H.-D. Barth, B. Brutschy, U. Scheffer, S. Breitung, M. W. Göbel, *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1600–1611.
[18] a) E. Duchardt, H. Schwalbe, *J. Biomol. NMR* **2005**, *32*, 295–308; b) J. Ferner, A. Villa, E. Duchardt, E. Widjajakusuma, J. Wöhnert, G. Stock, H. Schwalbe, *Nucleic Acids Res.* **2008**, *36*, 1928–1940.
[19] M. F. Bardaro Jr, Z. Shajani, K. Patora-Komisarska, J. A. Robinson, G. Varani, *Nucleic Acids Res.* **2009**, *37*, 1529–1540.
[20] a) M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586–7590; b) B. H. Geierstanger, J. P. Jacobsen, M. Mrksich, P. B. Dervan, D. E. Wemmer, *Biochemistry* **1994**, *33*, 3055–3062.
[21] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1996**, *274*, 1531–1534.
[22] V. Sklenar, R. D. Peterson, M. R. Rejante, J. Feigon, *J. Biomol. NMR* **1994**, *4*, 117–122.
[23] a) A. J. Dingley, S. Grzesiek, *J. Am. Chem. Soc.* **1998**, *120*, 8293–8297; b) K. Pervushin, A. Ono, C. Fernandez, T. Szyperski, M. Kainosho, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14147–14151.
[24] J. Wöhnert, A. J. Dingley, M. Stoldt, M. Görlach, S. Grzesiek, L. R. Brown, *Nucleic Acids Res.* **1999**, *27*, 3104–3110.
[25] V. Sklenar, R. D. Peterson, M. R. Rejante, J. Feigon, *J. Biomol. NMR* **1993**, *3*, 721–727.
[26] A. Pardi, E. P. Nikonowicz, *J. Am. Chem. Soc.* **1992**, *114*, 9202–9203.
[27] a) H. Schwalbe, J. P. Marino, S. J. Glaser, C. Griesinger, *J. Am. Chem. Soc.* **1995**, *117*, 7251–7252; b) J. P. Marino, H. Schwalbe, S. J. Glaser, C. Griesinger, *J. Am. Chem. Soc.* **1996**, *118*, 4388–4395; c) S. J. Glaser, H. Schwalbe, J. P. Marino, C. Griesinger, *J. Magn. Reson. Ser. B* **1996**, *112*, 160–180.
[28] L. Mueller, P. Legault, A. Pardi, *J. Am. Chem. Soc.* **1995**, *117*, 11043–11048.
[29] a) G. Varani, I. Tinoco, Jr., *Q. Rev. Biophys.* **1991**, *24*, 479–532; b) K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
[30] a) F. Tian, H. M. Al-Hashimi, J. L. Craighead, J. H. Prestegard, *J. Am. Chem. Soc.* **2001**, *123*, 485–492; b) M. Ottiger, F. Delaglio, A. Bax, *J. Magn. Reson.* **1998**, *131*, 373–378.
[31] A. T. Brünger, *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54*, 905–921.
[32] J. P. Linge, S. I. O'Donoghue, M. Nilges, *Methods Enzymol.* **2001**, *339*, 71–90.
[33] a) W. L. Jorgensen, J. Tirado Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666; b) J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.
[34] a) C. Dominguez, R. Boelens, A. M. J. J. Bonvin, *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737; b) S. J. de Vries, A. D. J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A. M. J. J. Bonvin, *Proteins Struct. Funct. Bioinf.* **2007**, *69*, 726–733.

Received: April 8, 2009
Published online on May 14, 2009

Please note: Minor changes have been made to this manuscript since its publication in *ChemBioChem* Early View. The Editor

