

Native Top-Down Mass Spectrometry of TAR RNA in Complexes with a Wild-Type tat Peptide for Binding Site Mapping

Eva-Maria Schneeberger and Kathrin Breuker*

Abstract: Ribonucleic acids (RNA) frequently associate with proteins in many biological processes to form more or less stable complex structures. The characterization of RNA–protein complex structures and binding interfaces by nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, or strategies based on chemical crosslinking, however, can be quite challenging. Herein, we have explored the use of an alternative method, native top-down mass spectrometry (MS), for probing of complex stoichiometry and protein binding sites at the single-residue level of RNA. Our data show that the electrostatic interactions between HIV-1 TAR RNA and a peptide comprising the arginine-rich binding region of tat protein are sufficiently strong in the gas phase to survive phosphodiester backbone cleavage of RNA by collisionally activated dissociation (CAD), thus allowing its use for probing tat binding sites in TAR RNA by top-down MS. Moreover, the MS data reveal time-dependent 1:2 and 1:1 stoichiometries of the TAR–tat complexes and suggest structural rearrangements of TAR RNA induced by binding of tat peptide.

Interactions between ribonucleic acids (RNA) and proteins are central to many fundamental biological processes, including gene expression and infection by RNA viruses. For a thorough understanding of such interactions, RNA–protein complexes are commonly investigated by nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallography, both of which require relatively large quantities of sample material. Moreover, NMR data interpretation can be complicated by unfavorable conformational dynamics,^[1] and crystallography can become impossible if a complex fails to crystallize properly. Strategies based on (photo)chemical crosslinking^[2] have the advantage that they can be performed in vivo^[3] but can variously suffer from low crosslinking yields, different crosslinking reactivity of different residues, or the formation of intramolecular instead of intermolecular cross-

links.^[4] Moreover, crosslinking reagents generally target specific functional groups such as amines or thiols that may not be present in a binding region, and efficient crosslinking can require pH values that may not be compatible with RNA–protein complex stability.^[5] Although all of the above techniques can provide highly important structural data, each of them requires laborious sample preparation procedures, that is, crystallization, the introduction of heavy isotopes, or optimization of the reaction conditions for the formation of intermolecular crosslinks.

As an alternative to these methods, we explore the potential of native top-down mass spectrometry (MS) using electrospray ionization (ESI)^[6] for the characterization of RNA–protein interactions. Previous studies showed that native ESI can produce gaseous RNA–protein or RNA–ligand complexes,^[7] and in top-down MS experiments using collisionally activated dissociation (CAD), Loo^[8] and Fabris^[7] observed cleavage of covalent mononucleotide phosphate and RNA phosphodiester bonds, respectively, rather than dissociation of the noncovalent bonds of the complexes studied. In agreement with our recent work on KIX protein^[9] and data from the literature,^[10] their findings suggest that in the absence of solvent, the strength of salt bridges can be comparable to that of covalent bonds. By contrast, desolvated duplexes of small interfering RNA that are stabilized by base pairing and stacking instead of salt-bridge interactions were found to dissociate by strand separation rather than backbone bond cleavage.^[11] Coincidentally, electrostatic interactions are common elements of RNA–protein binding; a statistical analysis of crystal structures revealed a high percentage of positively charged amino acid residues in the binding interfaces with the negatively charged RNA.^[12]

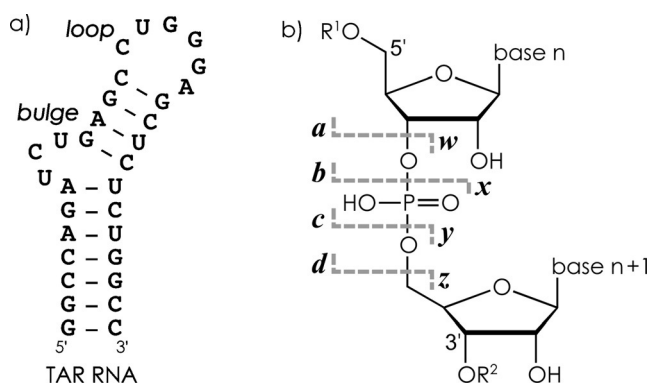
As a model system for our studies, we investigated complexes of 31 nt TAR (transactivation responsive) RNA^[7a] (Scheme 1a) and a basic peptide (H₂N-GRKKRRQRRRPP-NH₂) comprising the arginine-rich binding region of tat (trans-activating) protein from human immunodeficiency virus type 1 (HIV-1). Previous studies showed that MS can be used to probe the effect of relative TAR and tat concentrations on complex stoichiometry.^[7a,13] Solution NMR studies of the HIV-1 TAR–tat complex were generally complicated by broad signals,^[14] but highly converging structures of complexes with peptide mimetics of tat were obtained by Varani and co-workers.^[15]

ESI approximately 5 min after the preparation of a 1:1 TAR/tat solution produced 24% unbound RNA, that is, (TAR–nH)^{n–}, 42% 1:1 complex, (TAR + tat–nH)^{n–}, and 34% 1:2 complex, (TAR + 2tat–nH)^{n–} (these proportions changed somewhat with pH, see Figure S1 in the Supporting Information), but ESI after approximately 30 min produced

[*] E.-M. Schneeberger, Priv.-Doz. Dr. K. Breuker
Institut für Organische Chemie and Center for Molecular Biosciences
Innsbruck (CMBI), Universität Innsbruck
Innrain 80-82, 6020 Innsbruck (Austria)
E-mail: kathrin.breuker@uibk.ac.at
Homepage: <http://www.bioms-breuker.at>

Supporting information, including experimental details, and the ORCID identification number(s) for the author(s) of this article can be found under:
<http://dx.doi.org/10.1002/anie.201610836>.

© 2016 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



Scheme 1. a) Predicted secondary structure (<http://rna.tbi.univie.ac.at/>)^[16] of the TAR RNA construct studied. b) Nomenclature of fragment ions from RNA backbone cleavage.

mostly the 1:1 complex (Figure 1a and the Supporting Information, Figure S1). This data illustrates that native top-down MS can provide time-resolved information on complex assembly and stoichiometry^[7a,13] that is not generally available by other methods.

CAD of (TAR + tat-13H)¹³⁻ ions (m/z ca. 886; Figure 1b) produced only 0.3% unbound TAR RNA and 0.2% free tat peptide but 26.1% products from RNA backbone cleavage (Scheme 1b), that is, *c* (9.6%), *y* (12.0%), *a* (1.8%), *w* (2.4%), and *i* (internal) fragments from secondary backbone cleavage (0.3%), along with 64.4% undissociated ions and 9.0% nucleobase loss from the latter. Nucleobase loss from fragments was also observed in 17.2%, 5.1%, 72.7%, and 0.4% of all *c*, *y*, *a*, and *w* ions, respectively; the substantially higher value for *a* ions is consistent with the dominant mechanism proposed for 3' C–O cleavage that is initiated by nucleobase loss.^[17] The lower-energy mechanism of phosphodiester backbone bond cleavage into *c* and *y* ions (Supporting Information, Figure S2), on the other hand, does not involve nucleobase loss.^[18] Site-specific yields of *c* and *y* ions were generally higher on the 5'-side of guanosine (Figure 2a), which we have recently attributed to the hydrogen bonding between nucleobase and phosphodiester moieties at the cleavage site prior to dissociation.^[18] Peptide binding somewhat reduces this effect (Supporting Information, Figure S3), consistent with the restricted conformational RNA flexibility in the bound state.

Moreover, the CAD spectrum in Figure 1b shows *c*, *y*, *a*, and *w* fragments from RNA backbone cleavage both with and without the intact peptide attached, the m/z values of which were consistently higher with the peptide attached than without, regardless of the fragment ion mass (Supporting Information, Figure S4). This suggests that some of the net positive charge of the tat peptide is preserved in both the (TAR + tat-13H)¹³⁻ ions and its *c*, *y*, *a*, and *w* fragments and corroborates the hypothesis that electrostatic interactions play a major role in stabilizing the gaseous TAR–tat complex. Importantly, noncovalent bonds between TAR and tat as well as the covalent bonds of tat can apparently be preserved at energies that are sufficiently high to cleave covalent RNA–backbone bonds and disrupt intramolecular RNA interactions such as base pairing in stem regions (Scheme 1a).

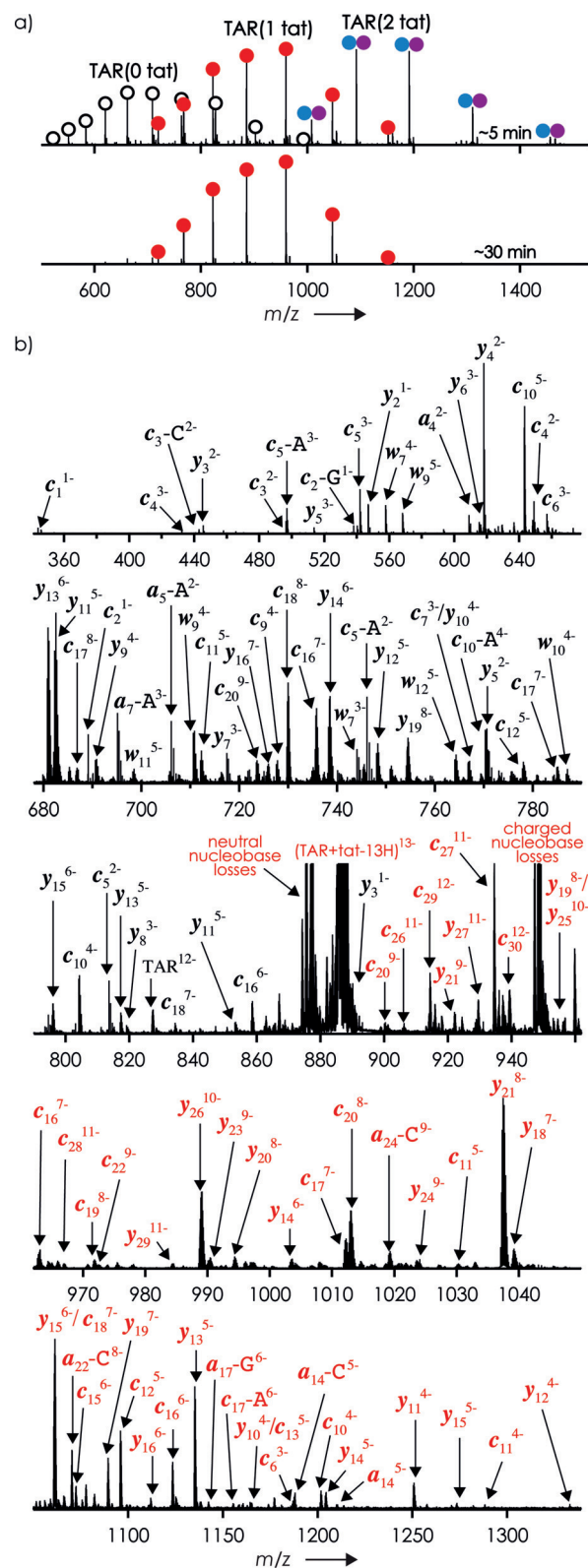


Figure 1. a) ESI spectra of TAR and tat (2 μ M each) in 9:1 H₂O/CH₃OH at pH 7.7 (adjusted by addition of piperidine, final concentration 1.125 mM) sprayed ca. 5 (top) and 30 min (bottom) after preparation of the solution. Empty circles, (TAR–nH)ⁿ⁻; red circles, (TAR + tat–nH)ⁿ⁻; blue/purple circles, (TAR + 2tat–nH)ⁿ⁻ ions. b) CAD spectrum of (TAR + tat–13H)¹³⁻ ions at 120.9 eV laboratory frame energy (36% dissociation), TAR fragments with (red) and without (black) tat attached are labeled.

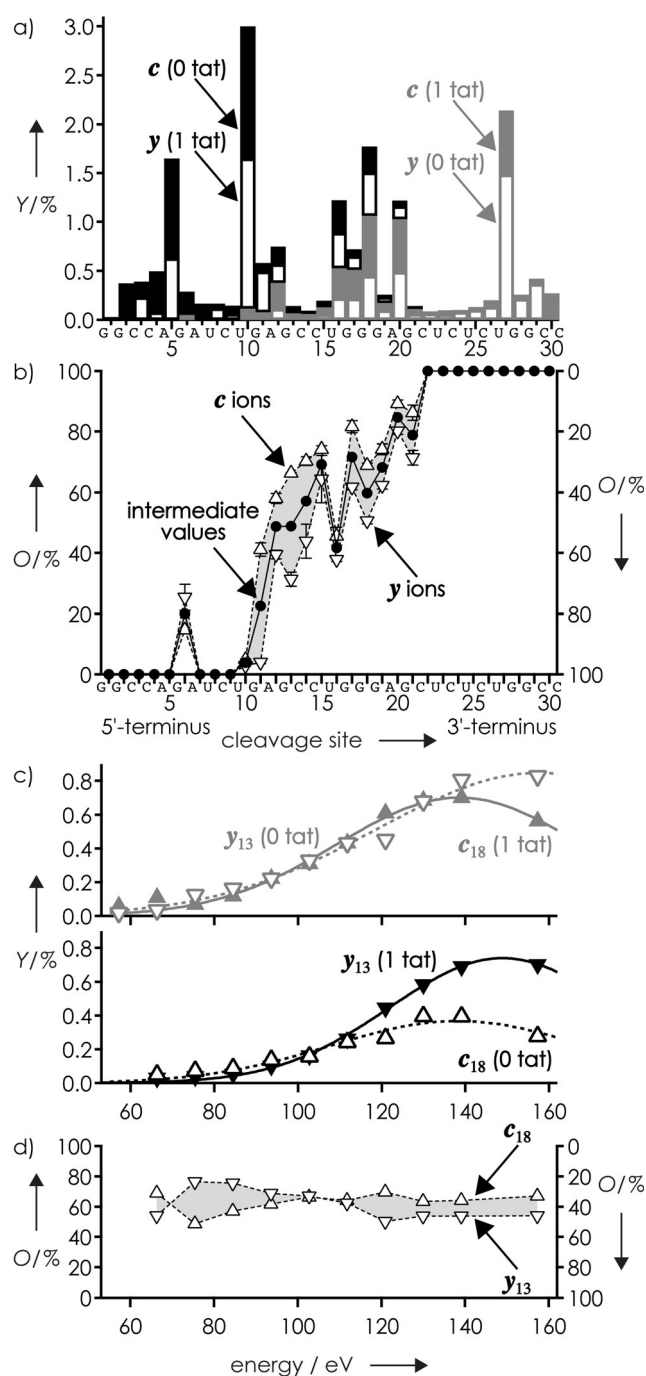


Figure 2. For CAD of (TAR + tat-13H)¹³⁻ ions, a) yield (Y) of *c* and *y* fragments without (0 tat) and with (1 tat) peptide attached from spectrum in Figure 1 and b) average level of tat occupancy (O) of *c* (Δ, left axis) and *y* (∇, right axis) fragments (from triplicate measurements under identical experimental conditions at 120.9 eV, error bars show standard deviation of the mean) and intermediate values (●, calculated as $(O(c) - O(y) + 100)/2$) versus cleavage site. c) Yield of complementary *c*₁₈(1 tat) and *y*₁₃(0 tat) fragments (top) and complementary *c*₁₈(0 tat) and *y*₁₃(1 tat) fragments (bottom) and d) occupancy values for *c*₁₈ and *y*₁₃ versus energy used for CAD; lines are meant to guide the eye.

Figure 2b illustrates the level of tat occupancy of the *c* and *y* fragments of TAR RNA for each cleavage site, calculated as $100 \cdot Y(c \text{ or } y \text{ with tat}) / (Y(c \text{ or } y \text{ with tat}) + Y(c \text{ or } y \text{ without$

tat)); within error limits, these were the same for ESI circa 5 and 30 min after preparation of the solution. Zero occupancy was found for *c*₁-*c*₅, *c*₇-*c*₉, *y*₂-*y*₉ (*y*₁ was not detected) and 100% occupancy for the complementary *y*₃₀-*y*₂₆, *y*₂₄-*y*₂₂, *c*₃₀-*c*₂₂ fragments, suggesting that TAR residues 1-5, 7-9, and 22-31 did not, or at least not strongly, interact with tat. Occupancy values other than 0% or 100% were found for all other sites, for example, from cleavage at site 18, $68.9 \pm 0.8\%$ for *c*₁₈ and $49.4 \pm 1.0\%$ for its complement *y*₁₃. This can be rationalized by the interactions of tat peptide with more than one residue of TAR RNA and the breaking by CAD of only some of them according to the strength and number of interactions with each fragment. In the above example, the higher fraction of *c*₁₈ fragments with tat peptide attached compared to that of *y*₁₃ suggests an overall higher stability of tat interactions with *c*₁₈ than with *y*₁₃, but both are sufficiently strong to prevent loss of tat during the backbone cleavage of TAR.

The added occupancy values of complementary *c* and *y* fragments, however, generally exceeded or fell below the expected 100%; for cleavage at site 18, the values for *c*₁₈ and *y*₁₃ add up to 118%. The discrepancies of the added occupancy values from 100% ranged from 2% at site 10 to 37% at site 11, with an average value of 17%, and can be attributed to different stabilities of the *c* and *y* fragments, with and without peptide attached, against secondary backbone cleavage. To exemplify this point, yields of complementary *c*₁₈(1 tat) and *y*₁₃(0 tat) fragments (top) and *c*₁₈(0 tat) and *y*₁₃(1 tat) fragments (bottom) are illustrated in Figure 2c as a function of the energy used for CAD. Apparently, *c*₁₈ fragments undergo secondary backbone cleavage more readily than *y*₁₃ fragments, and *c*₁₈(1 tat) showed higher stability than *c*₁₈(0 tat). Because secondary backbone cleavage of a *c* or *y* fragment produces both an internal *i* fragment and a smaller *c* or *y* fragment, it not only decreases the yield of larger *c* or *y* fragments (Figure 2c) but also increases the yield of smaller *c* or *y* fragments (Supporting Information, Figure S5).^[19] For the determination of the occupancy values, spectra obtained at lower CAD energy (at which fewer products from secondary backbone cleavage are observed)^[19] should thus be more accurate; on the other hand, both the signal-to-noise level and sequence coverage generally increase with increasing energy (Supporting Information, Figure S2). Full sequence coverage in CAD of (TAR + tat-13H)¹³⁻ ions was obtained at energies greater than or equal to 111.8 eV, and the yields of the internal fragments were less than 1% at energies below 130 eV, with 120.9 eV (Figure 1b, Figure 2a,b) just in between.

The occupancy values of *c* and *y* fragments increased almost monotonically with increasing number of residues (Figure 2b); however, some variation outside of the error limits, even when only intermediate values were considered, was nonetheless observed. For example, the tat occupancy of *c*₆ was circa 15%, but 0% for *c*₇, *c*₈, and *c*₉, and that of the complementary fragments circa 75% for *y*₂₅ but 100% for *y*₂₄, *y*₂₃, and *y*₂₂; likewise, the occupancy values at site 16 stand out. The variations in occupancy values, however, do not appear to be related to the net charge of the fragment ions, as these showed a steady increase with increasing number of residues (Supporting Information, Figure S4). We instead attribute

this phenomenon to charge redistribution within fragment ions by proton transfer according to Coulombic repulsion,^[18] after backbone cleavage but before complex dissociation, that can weaken or strengthen individual interactions differently in different fragment ions, which in turn affects the competition between complementary *c* and *y* fragments for *tat*. In this scenario, an electrostatic interaction of residue 6 of *c*₆ with *tat* can be sufficiently strong to compete with the interactions between the complementary fragment *y*₂₅ and *tat*, whereas in the *c*₇, *c*₈, and *c*₉ ions, negative charge will move away from residue 6 to the terminal residues 7, 8, and 9, respectively,^[18] thus weakening the electrostatic interaction of residue 6 with *tat*. This does not exclude the possibility of *tat* interactions with residues 7, 8, and 9 but suggests that they must be substantially weaker than that between *tat* and residue 6 of *c*₆.

Nevertheless, the data in Figure 2b clearly show that in the 1:1 TAR–*tat* complex, *tat* binds to residues 6 and 10–22 of TAR RNA. The color coding of residues 6 and 9–22 (red) and the more weakly interacting residue 10 grouped with the possibly weakly interacting residues 7–9 (orange) of the ground state structure of TAR in complex with a peptide mimetic (green)^[15b] shows excellent agreement between MS and NMR data (Figure 3a), even though the *tat* peptide

studied here was linear and that in the NMR structure was cyclic. In the same manner, data from CAD MS of (TAR + 2*tat*–12H)^{12–} ions (Supporting Information, Figures S6 and S7) revealed the sites of *tat* binding to TAR in the 1:2 TAR–*tat* complex (Figure 3b). In the 1:1 complex, *tat* binds to both the predicted loop, bulge, and upper 5′-stem region, whereas in the 1:2 complex, one *tat* binds to the loop and upper 3′-stem region (blue), and the other to the bulge and 5′-stem region (violet). Consistent with the rather ill-defined NMR structure of free TAR,^[20] our data suggest a dynamic and sufficiently open TAR RNA structure to which two *tat* peptides can initially bind, followed by rearrangements into a more compact TAR structure along with ejection of one peptide.

In conclusion, we have shown that native top-down MS provides time-resolved information on TAR–*tat* complex stoichiometry and reveals *tat* binding sites at the single-residue level of TAR RNA. Apparently, the electrostatic interactions between TAR RNA and the basic *tat* peptide in the gas phase are sufficiently strong to survive at CAD energies that cleave phosphodiester backbone bonds of RNA. Native top-down MS requires only small amounts of sample without the need for laborious preparation procedures, allows for the separate study of species that differ in *m/z*, simultaneously provides extensive sequence information (Supporting Information, Tables S1 and S2),^[21] and is a promising approach for probing RNA–protein binding, especially in cases in which conventional methods for structural probing are not applicable.

Experimental Section

Experiments were performed on a 7 T Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker, Austria) equipped with an ESI source, a linear quadrupole for ion isolation, and a collision cell for CAD.

Acknowledgements

Funding was provided by the Austrian Science Fund (FWF): Y372 and P27347 to K.B.

Conflict of interest

The authors declare no conflict of interest.

Keywords: binding sites · collisionally activated dissociation · mass spectrometry · proteins · RNA

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 1254–1258
Angew. Chem. **2017**, *129*, 1274–1278

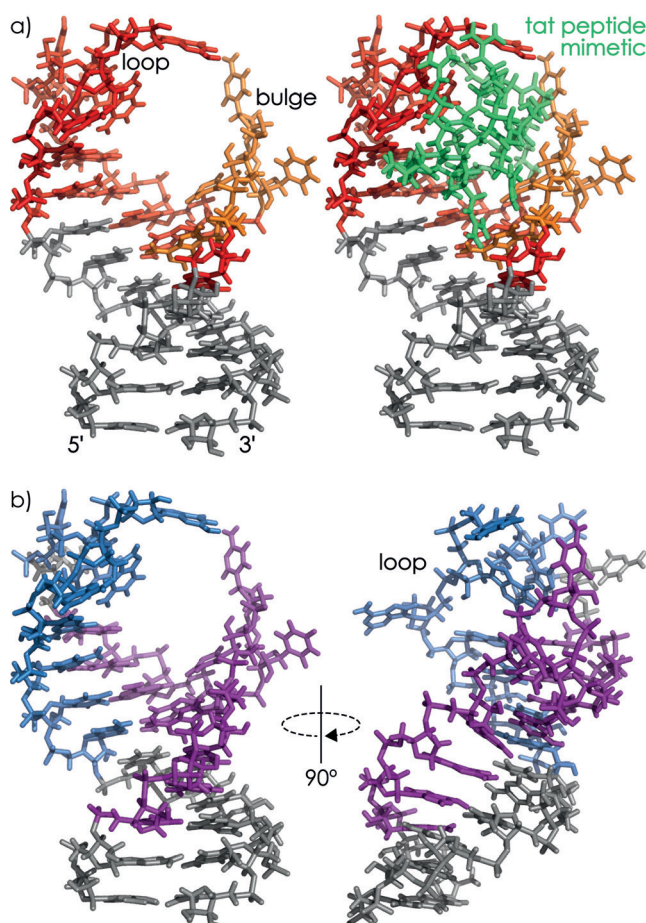


Figure 3. NMR ground state structure of TAR RNA in complex with peptide mimetic (PDB entry 5J0M)^[15b] with *tat* binding sites from MS data color-coded (see main text) for a) 1:1 and b) 1:2 TAR–*tat* complex.

- [1] B. Fürtig, C. Richter, J. Wöhnert, H. Schwalbe, *ChemBioChem* **2003**, *4*, 936–962.
- [2] H. Steen, O. N. Jensen, *Mass Spectrom. Rev.* **2002**, *21*, 163–182.
- [3] C. A. McHugh, P. Russell, M. Guttman, *Genome Biol.* **2014**, *15*, 203.
- [4] A. Sinz, C. Arlt, D. Chorev, M. Sharon, *Protein Sci.* **2015**, *24*, 1193–1209.

- [5] M. M. Hanna, *Methods Enzymol.* **1989**, *180*, 383–409.
- [6] a) H. Zhang, W. D. Cui, J. Z. Wen, R. E. Blankenship, M. L. Gross, *Anal. Chem.* **2011**, *83*, 5598–5606; b) H. L. Li, P. Wongkongkathep, S. L. Van Orden, R. R. O. Loo, J. A. Loo, *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 2060–2068; c) H. L. Li, J. J. Wolff, S. L. Van Orden, J. A. Loo, *Anal. Chem.* **2014**, *86*, 317–320.
- [7] a) K. A. Sannes-Lowery, P. F. Hu, D. P. Mack, H. Y. Mei, J. A. Loo, *Anal. Chem.* **1997**, *69*, 5130–5135; b) K. A. Sannes-Lowery, H. Y. Mei, J. A. Loo, *Int. J. Mass Spectrom.* **1999**, *193*, 115–122; c) D. R. Benjamin, C. V. Robinson, J. P. Hendrick, F. U. Hartl, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7391–7395; d) S. A. Hofstadler, R. H. Griffey, *Chem. Rev.* **2001**, *101*, 377–390; e) N. Hagan, D. Fabris, *Biochemistry* **2003**, *42*, 10736–10745; f) K. B. Turner, N. A. Hagan, A. S. Kohlway, D. Fabris, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1402–1411; g) D. Fabris, K. B. Turner, A. S. Kohlway, N. A. Hagan, *Biopolymers* **2009**, *91*, 283–296; h) K. M. Keller, M. M. Breeden, J. M. Zhang, A. D. Ellington, J. S. Brodbelt, *J. Mass Spectrom.* **2005**, *40*, 1327–1337.
- [8] a) S. Yin, J. A. Loo, *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 899–907; b) S. Yin, Y. M. Xie, J. A. Loo, *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1199–1208.
- [9] a) M. Schennach, E. M. Schneeberger, K. Breuker, *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1079–1088; b) K. Breuker, S. Brüscheweiler, M. Tollinger, *Angew. Chem. Int. Ed.* **2011**, *50*, 873–877; *Angew. Chem.* **2011**, *123*, 903–907.
- [10] a) X. Barril, C. Aleman, M. Orozco, F. J. Luque, *Proteins Struct. Funct. Genet.* **1998**, *32*, 67–79; b) E. F. Strittmatter, E. R. Williams, *Int. J. Mass Spectrom.* **2001**, *212*, 287–300.
- [11] T. Y. Huang, J. Liu, X. R. Liang, B. D. M. Hodges, S. A. McLuckey, *Anal. Chem.* **2008**, *80*, 8501–8508.
- [12] M. Treger, E. Westhof, *J. Mol. Recognit.* **2001**, *14*, 199–214.
- [13] N. Morgner, H. D. Barth, B. Brutschy, U. Scheffer, S. Breitung, M. Göbel, *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1600–1611.
- [14] D. J. Patel, *Curr. Opin. Struct. Biol.* **1999**, *9*, 74–87.
- [15] a) A. Davidson, K. Patora-Komisarska, J. A. Robinson, G. Varani, *Nucleic Acids Res.* **2011**, *39*, 248–256; b) A. N. Borkar, M. F. Bardaro, C. Camilloni, F. A. Aprile, G. Varani, M. Vendruscolo, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 7171–7176.
- [16] I. L. Hofacker, *Nucleic Acids Res.* **2003**, *31*, 3429–3431.
- [17] S. Schürch, *Mass Spectrom. Rev.* **2016**, *35*, 483–523.
- [18] C. Riml, H. Glasner, M. T. Rodgers, R. Micura, K. Breuker, *Nucleic Acids Res.* **2015**, *43*, 5171–5181.
- [19] M. Taucher, U. Rieder, K. Breuker, *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 278–285.
- [20] F. Aboul-ela, J. Karn, G. Varani, *Nucleic Acids Res.* **1996**, *24*, 3974–3981.
- [21] M. Taucher, K. Breuker, *Angew. Chem. Int. Ed.* **2012**, *51*, 11289–11292; *Angew. Chem.* **2012**, *124*, 11451–11454.

Manuscript received: November 6, 2016

Final Article published: December 21, 2016