Chapter 3: Designed arginine rich RNA-binding peptides with picomolar affinity

Ryan J. Austin†, Tianbing Xia‡, Jinsong Ren‡, Terry T. Takahashi†, and Richard W. Roberts‡*

† Option of Biochemistry and Molecular Biophysics, California Institute of Technology, Pasadena CA 91125

‡ Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena CA 91125.

* Correspondence to: Richard W. Roberts, California Institute of Technology, M/C 147/75 CH, Pasadena, CA, 91125-0001, Phone (626) 395-2321; Fax (626)568-9430;

Abstract

Arginine-rich peptide motifs (ARMs) capable of binding unique RNA structures play critical roles in transcription, translation, RNA trafficking and RNA packaging. Bacteriophage ARMs necessary for transcription antitermination bind to distinct boxB RNA hairpin sequences with a characteristic induced α-helical structure. Characterization of ARMs from lambdoid phages reveals that the dissociation constant of the P22 bacteriophage model-antitermination complex (P22N21-P22boxB) is 200 ± 56 pM in free solution at physiologic concentrations of monovalent cation; significantly stronger than previously determined by gel mobility shift and polyacrylamide gel coelectophoresis, and two orders of magnitude stronger than the tightest known native ARM-RNA interaction at physiological salt. Here, we use a reciprocal design approach to enhance the binding affinity of two separate α helical ARM-RNA interactions; one derived from the native λ phage antitermination complex and a second isolated using mRNA display selection experiments targeting boxB RNA.

Manuscript

Arginine-rich peptide motifs (ARMs) capable of binding unique RNA structures play critical roles in transcription, translation, RNA trafficking and RNA packaging.[1] Bacteriophage ARMs necessary for transcription antitermination[2] bind to distinct boxB RNA hairpin sequences with a characteristic induced α-helical structure.[3] Characterization of ARMs from lambdoid phages reveals that the dissociation constant (K_d) of the P22 bacteriophage model-antitermination complex (P22$_{N21}$-P22boxB) is 200 $\pm$ 56 pM in solution at physiologic concentrations of monovalent cation; significantly stronger than previously determined by gel mobility shift (GMSA)[4] and polyacrylamide gel coelectophoresis (PACE)[5] and two orders of magnitude stronger than the tightest known native ARM-RNA interaction at physiological salt.[6] Here, we use a reciprocal design approach to enhance the binding affinity of two separate α-helical ARM-RNA interactions; one derived from the native λ phage antitermination complex and a second isolated using mRNA display selection experiments targeting boxB RNA.[7]

Steady-state K_D values over a range of salt concentrations were determined by monitoring the change in fluorescence of a 2-aminopurine (2AP) base analog substituted at different adenine positions within the loop of the P22boxB RNA target.[8] At physiological concentrations of monovalent salt (150 mM [M$^+$]) the P22 cognate complex is 100 times (2.7 kcal mol^{-1}) more stable than its counterpart antitermination complex in λ phage (λ_{N22}-λboxB). Kinetic data from stopped-flow measurements indicate that greater stability of the P22$_{N21}$-P22boxB complex relative to the λ_{N22}-λboxB complex results from a slower rate of dissociation (k_{off})(P22$_{N21}$-P22boxB: k_{off} = 0.01 s^{-1}; λ_{N22}-λboxB: k_{off} = 0.7

s^{-1})(Fig. 3.1) while the rate of association (k$_{on}$) is consistent with a diffusion controlled process in both complexes (P22$_{N21}$-P22boxB, λ$_{N22}$-λboxB: k$_{on}$ ~ 7 x 10^{8} s^{-1}M^{-1}).

The NMR solution structures of P22 and λ antitermination complexes reveal adaptation of similar but distinct RNA hairpin folds, extruding either the third base pyrimidine or fourth base purine to generate a highly stable GNRA-fold[9] in the pentaloops of P22boxB and λboxB hairpins, respectively. In both P22$_{N21}$-P22boxB and λ$_{N22}$-λboxB complexes the ARM peptides adopt bent α-helical structures that bind within the major groove of their cognate hairpins.[10] The amino-terminal residues of P22$_{N21}$ and λ$_{N22}$ peptides are highly conserved and make similar contacts with the hairpin stems of their cognate boxB RNA targets.[11]

The principle of our design was to use peptide sequence information from the P22$_{N21}$-P22boxB complex to increase the stability of the λ$_{N22}$-λboxB complex without disrupting its unique loop-binding structure. To this end, we focused on non-conserved residues within the amino-terminal helical segment of λ$_{N22}$ that interact primarily with the λboxB hairpin stem in the cognate complex. Non-conserved residues within the amino-terminal helix of λ$_{N22}$ were substituted with reciprocal residues from P22$_{N21}$. The binding affinities of these substituted peptides for λboxB at 150 mM and 70 mM salt are shown in Figure 3.2. Three reciprocal substitutions (M1G, D2N, Q4K) in the λ$_{N22}$ peptide increase the stability of the λ$_{N22}$-λboxB substituted complex by 1.8 kcal mol^{-1} (Table 3.1). Substitution of Lys4 with an arginine residue further increases binding stability by 0.7 kcal mol^{-1}, in line with previous *in vitro* evolution experiments showing a glutamine to arginine change at this position (J.E.B and R.W.R personal communication) and with work indicating that arginine oligomers bind RNA more tightly than those made of

lysine.[11] Similarly, application of the M1G, D2N, Q4R substitution set to the peptide 11-36, a previously selected high-specificity binder of λboxB[7], increases the binding stability of this peptide by 2.0 kcal mol^{-1}.

In addition to providing a binding signal, the relative fluorescence (F)[13] of 2AP in the peptide-boxB complex confers information about the structure of the bound RNA. A decrease in fluorescence (F < 1) results from base stacking of the 2AP probe while an increase in fluorescence (F > 1) results from exposure of 2AP to aqueous solvent.[14] Titration of the native λ_{N22}-λboxB complex gives a fluorescence signature consistent with the known solution structure of the complex. A large decrease in fluorescence is observed for 2AP label at positions 2 (2AP-2: F = 0.31) and 3 (2AP-3: F = 0.22) of the λboxB loop consistent with Trp18 stacking and base stacking interactions, while a large increase in fluorescence is observed for loop position 4 of λboxB (2AP-4: F = 2.35), indicative of base extrusion into solvent. This native structure is maintained in the substituted peptide complexes (2AP-2: F = 0.32-0.36; 2AP-3: F = 0.12-0.22; 2AP-4: F = 2.36-2.60). Stacking of Trp18 on the λboxB loop is also observed in NMR imino-spectra of the substituted peptide complexes.[15] Thermodynamic and kinetic measurements, together with fluorescence and NMR data, demonstrate that our reciprocal substitutions stabilize the adopted loop structure of the unique λ_{N22}-λboxB native complex (Supplemental information).

Interestingly circular dichroism measurements indicate that substituted peptides bind λboxB with less α-helical character than native λ_{N22}, in large part attributable to the D2N substitution, which decreases helicity of the bound peptide by 13%. We speculate

that this decrease in helicity results from loss of macrodipole stabilization by the negatively charged Asp2 residue.

Salt dependence data ($\delta/\delta = (-\delta\log K_{obs})(\delta\log[M^+])^{-1}$) are useful in investigating the electrostatic contributions of different amino-acid residues in binding events.[16] D2N and Q4K reciprocal substitutions increase δ/δ values in the λ_{N22}-λboxB complexes, demonstrating that favorable binding affinity exhibited from these substitutions has an electrostatic component (Table 3.1). Conversely, M1G and K4R replacements do not significantly affect δ/δ values, indicating that enhanced binding affinity from these substitutions is not purely electrostatic in nature.

We have determined that the native P22$_{N21}$-P22boxB complex and engineered λ_{N22}-λboxB and 11-36-λboxB complexes bind with picomolar affinity at physiological salt. These pairs are the most stable peptide-RNA complexes reported to our knowledge. Such examples provide evidence that short peptides can be found that bind unique RNA hairpins with affinity rivaling macromolecular complexes, providing a potential means to control RNA-dependent cellular processes through chemical methods. The reciprocal engineering approach presents a complementary tool to combinatorial or computational design techniques[17] and should be of particular utility where all the space of possible sequences greatly exceeds the total number of sequences that may be generated and examined.

References

1. Weiss, M. A.; Narayana, N. Biopolymers 1998, 48, 167-180.
2. Lazinski, D.; Grzadzielska, E.; Das, A. Cell 1989, 59, 207-218.
3. (a)Van Gilst, M. R.; Rees, W. A.; Das, A.; von Hippel, P. H. Biochemistry 1997, 36, 1514-1524.(b) Frankel, A. D. Curr Opin Struct Biol 2000, 10, 332-340.
4. Tan, R.; Frankel, A. D. Proc Natl Acad Sci U S A 1995, 92, 5282-5286.
5. Cilley, C. D.; Williamson, J. R. RNA 1997, 3, 57-67.
6. Lacourciere, K. A.; Stivers, J. T.; Marino, J. P. Biochemistry 2000, 39, 5630-5641.
7. Barrick, J. E.; Takahashi, T. T.; Ren, J.; Xia, T.; Roberts, R. W. Proc Natl Acad Sci U S A 2001, 98, 12374-12378.
8. Peptides and unlabeled RNA were synthesized as described in supplemental information. P22boxBL: gGCGCU GACAA AGCGCc; λboxBR: gGCCCU GAAAA AGGGCc. Loop residues are underlined.
9. Heus, H. A.; Pardi, A. Science 1991, 253, 191-194.
10. (a) Su, L.; Radek, J. T.; Hallenga, K.; Hermanto, P.; Chan, G.; Labeots, L. A.; Weiss, M. A. Biochemistry 1997, 36, 12722-12732. (b) Mogridge, J.; Legault, P.; Li, J.; Van Oene, M. D.; Kay, L. E.; Greenblatt, J. Mol Cell 1998, 1, 265-275. (c) Cai, Z.; Gorin, A.; Frederick, R.; Ye, X.; Hu, W.; Majumdar, A.; Kettani, A.; Patel, D. J. Nat Struct Biol 1998, 5, 203-212. (d) Legault, P.; Li, J.; Mogridge, J.; Kay, L. E.; Greenblatt, J. Cell 1998, 93, 289-299. (e) Scharpf, M.; Sticht, H.; Schweimer, K.; Boehm, M.; Hoffmann, S.; Rosch, P. Eur J Biochem 2000, 267, 2397-2408.
11. (a) Weiss, M. A. Nat Struct Biol 1998, 5, 329-333. (b) Patel, D. J. Curr Opin Struct Biol 1999, 9, 74-87. (c) Weisberg, R. A.; Gottesman, M. E. J Bacteriol 1999, 181, 359-367.
12. (a) Mascotti, D. P.; Lohman, T. M. Biochemistry 1992, 31, 8932-8946. (b) Mascotti, D. P.; Lohman, T. M. Biochemistry 1997, 36, 7272-7279.
13. F = FF/FO; FO = fluorescence of free RNA; FF = fluorescence of RNA following titration with peptide.
14. (a) Millar, D. P. Curr Opin Struct Biol 1996, 6, 322-326. (b) Rachofsky, E. L.; Osman, R.; Ross, J. B. Biochemistry 2001, 40, 946-956. (c) Menger, M.; Eckstein, F.; Porschke, D. Biochemistry 2000, 39, 4500-4507. (d) Rist, M. J.; Marino, J. P. Nucleic Acids Research 2001, 29, 2401-2408.
15. Trp18 imino peak shifts from 10.0 ppm in free substituted peptides to 9.2 ppm in peptide λboxB complexes.
16. (a) Draper, D. E. J Mol Biol 1999, 293, 255-270. (b) GuhaThakurta, D.; Draper, D. E. J Mol Biol 2000, 295, 569-580.
17. Cheng, A. C.; Calabro, V.; Frankel, A. D. Current Opinion in Structural Biology 2001, 11, 478-484.
18. The PyMOL molecular Graphics System (2002) DeLano Scientific, S. C., CA, USA. http://www.pymol.org.

Figures and Tables

Figure 3.1 Stopped flow measurement of P22 cognate complex. Stopped-flow measurement of P22N21-P22boxB complex koff infused with 100x concentration of unlabeled P22boxB at time zero (data shown in gray, fit shown in black) Experiments were performed at 20º C in 50 mM KOAc, 20 mM Tris-OAc buffer, pH 7.5.
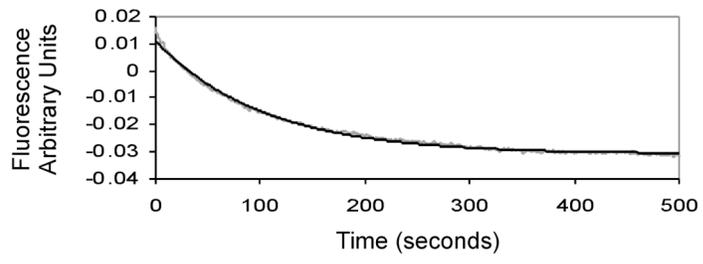
Figure 3.2 reciprocal mutant binding affinities. (A) Structural models of P22N21-P22boxB (blue) and λN22 (M1G, D2N, Q4K)-λboxB (red). The flipped third base in P22boxB interfaces with P22N21 and the second base in λboxB stacks with Trp18 of λN22 in the cognate complexes. **(B)** Sequence alignment of P22N21 (blue), λN22 (red), and substituted λN22 peptides; conserved residues are shown in black, non native residues are shown in gray. Binding constants for peptides against P22boxB 2AP-2 (blue) and λboxB 2AP-2 (red) hairpins at 150 and 70 mM salt were determined from salt dependences as described in Supplemental Information.

A
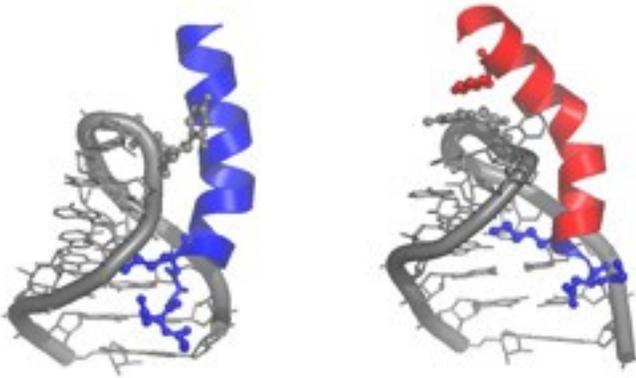
B

			Kd (150) (nM)	Kd (70) (pM)
P22$_{N21}$	GNAKTRRHERR	RKLAIERDTIgy	0.20	5
λ$_{N22}$	MDAQTRRRERR	AEKQAQWKAAN	22	1,900
λ$_{N22}$(N2)	MNAQTRRRERR	AEKQAQWKAAN	2.6	170
λ$_{N22}$(N2,K4)	MNAKTRRRERR	AEKQAQWKAAN	2.2	97
λ$_{N22}$(N2,R4)	MNARTRRRERR	AEKQAQWKAAN	0.56	23
λ$_{N22}$(G1,N2,K4)	GNAKTRRRERR	AEKQAQWKAAN	0.88	34
λ$_{N22}$(G1,N2,R4)	GNARTRRRERR	AEKQAQWKAAN	0.30	12
11-36	MDAQTRRRERR	AMERATLPQVLgy	27	3,000
11-36(G1,N2,R4)	GNARTRRRERR	AMERATLPQVLgy	0.89	33

Table 3.1. Thermodynamic Parameters and Helicity of Peptides [a]

peptide	$\Delta\Delta G$ (kcal mol^{-1})[b]	δ/δ[c]	α-helicity (%)[d]
P22$_{N21}$	-2.7	5.2 ± 0.2	65
λ_{N22}	0	3.2 ± 0.2	57
$\lambda_{N22(N2)}$	-1.2	3.5 ± 0.1	44
$\lambda_{N22(N2,K4)}$	-1.3	4.1 ± 0.2	38
$\lambda_{N22(N2,R4)}$	-2.1	4.2 ± 0.2	41
$\lambda_{N22(G1,N2,K4)}$	-1.8	4.3 ± 0.2	52
$\lambda_{N22(G1,N2,R4)}$	-2.5	4.2 ± 0.2	44

[a] Data are listed for cognate complexes (P22N21-P22boxB, λN22-λboxB, λN22(N2)-λboxB, etc.). [b] Free energy values ($\Delta\Delta G$) were calculated from binding constants at 150 mM [M$^+$]; ΔG = -10.3 kcal mol^{-1}. [c] Salt dependence values δ/δ = $(\delta \log Kobs)(\delta \log[M^+])^{-1}$ were determined from five or more measurements over a range of 50-500 mM KOAc. [d] Helicities were calculated from circular dichroism measurements in 10 mM phosphate buffer .

91

Supplemental Information

RNA and Peptide Synthesis

Peptides were generated on an Applied Biosystems 432A peptide synthesizer using solid phase, F-Moc chemistry. Crude peptides were deprotected by TFA/ethanedithiol/thioanisole treatment and purified on a C-18 reverse phase HPLC column to a final purity greater than 95% (MALDI-TOF, Analytical C-18 HPLC). Peptides without a naturally occurring tryptophan or tyrosine residue were synthesized with a carboxy-terminal Gly-Tyr tag for quantification purposes.

Unlabeled RNA hairpins (λboxB$_{R15}$ and P22boxB$_{L15}$) were synthesized by *in vitro* transcription using T7 RNA polymerase.[1] The RNA was purified by 20% urea-PAGE, desalted on a NAP column (Amersham Pharmacia), and freeze-dried. RNA was quantified by UV absorption at a wavelength of 260 nm.

Labeled RNA hairpins containing 2-amino purine (2AP) at loop position 2 (2AP-2), 3 (2AP-3), or 4 (2AP-4) were constructed by automated RNA synthesis using either 2-aminopurine-TOM-CE phosphoramidite or 2'-O-methyl 2-aminopurine phosphoramidite (Glen Research, Sterling, VA).

Steady-State Fluorescence Measurements

Measurements were conducted following the procedures of Barrick et al.[2] Titrations were performed on a Shimadzu Spectrofluorophotometer at 20° C with Excitation/Emission wavelengths at 310/370 nm . Peptides were titrated iteratively into a constantly stirred solution of 2AP labeled RNA hairpin (20-200 nM RNA). Binding

buffer contained 20 mM Tris-OAc, with a variable concentration of KOAc (15 mM-500 mM) at pH 7.5. Binding Constants were calculated for a one step binding mechanism by nonlinear least squares regression using the computer program DynaFit.[3] All isotherms were fit with < 10% uncertainty. Relative 2AP fluorescence change (F) is reported as a fraction; $F = (F_F/F_O)$ where F_O and F_F are the initial and final fluorescence respectively. Binding constants required a change in fluorescence of greater than 10% for reliable fitting. Salt dependence values $(-\delta logK_{obs})(\delta log[M^+])^{-1}$ were determined from five or more binding constants within a range of salt concentrations (50-500mM KOAc).

Stopped-Flow Fluorescence Measurements

Experiments were conducted following the procedures of Lacourciere et al.[4] Measurements were performed at 20° C under standard buffer conditions (20 mM Tris-OAc, 50 mM KOAc, pH 7.5) using a stop-flow device from Applied Photophysics (Surray, UK) in two-syringe mode. Fluorescence excitation was performed at 310 nm and emmision was measured with a filter cutoff > 360 nm. The association rate (k_{on}) was determined for the λ_{N22}-λboxB complex to be 7×10^8 s^{-1}M^{-1}. A similar rate was observed for the P22$_{N21}$-P22boxB complex. Dissociation rate values (k_{off}) were determined for each interaction by infusing labeled complex with both 100x and 500x concentration of unlabeled competitor boxB at time zero.

Circular Dichroism Spectroscopy

Spectra were recorded on an Aviv 62 DS CD spectrofluorimeter in 10 mM potassium phosphate buffer at 20° C. Spectra for the peptide-RNA complexes were

determined by subtracting spectra with the equivalent concentrations of RNA and of peptide from the spectra of the complex. Helix content was calculated from ellipticity at 222 nm (θ_{222}), using $-40,000(1-2.5/n)$ and 0 deg cm^2 dmol^{-1} as values for $\theta_{helix222}$ and $\theta_{coil222}$ respectively (n is the number of amino acids in the peptide).[5] Helicity = $(\theta_{observed})/(\theta_{helix} + \theta_{aromatic})$. The contribution of Trp was assumed to be -2300 ± 600 ($\theta_{aromatic}$) while the GY tag of the P22$_{N21}$ was assumed to contribute insignificant helicity and aromaticity to the CD spectrum.[6]

NMR Spectroscopy

Samples were prepared in NMR buffer: 50 mM NaCl, 10 mM phosphate, pH 6 in H$_2$O:D$_2$O (90:10, vol:vol) as previously described.[2] NMR spectra were collected at 25°C on a Varian INOVA 600-MHz spectrometer. A modified double gradient echo Watergate solvent-suppression pulse sequence was used to suppress the solvent peak.[7] Assignments were based on reported work.[8]

Supplemental References

1. (a) Milligan, J. F.; Uhlenbeck, O. C. Methods Enzymol 1989, 180, 51-62; (b) Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. Nucleic Acids Res 1987, 15, 8783-8798.
2. Barrick, J. E.; Takahashi, T. T.; Ren, J.; Xia, T.; Roberts, R. W. Proc Natl Acad Sci U S A 2001, 98, 12374-12378.
3. Kuzmic, P. Anal Biochem 1996, 237, 260-273.
4. Lacourciere, K. A.; Stivers, J. T.; Marino, J. P. Biochemistry 2000, 39, 5630-5641.
5. Chakrabartty, A.; Schellman, J. A.; Baldwin, R. L. Nature 1991, 351, 586-588.
6. (a) Chakrabartty, A.; Kortemme, T.; Padmanabhan, S.; Baldwin, R. L. Biochemistry 1993, 32, 5560-5565; (b) Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. Protein Sci 1994, 3, 843-852.
7. Liu, W.; Okajima, K.; Murakami, K.; Harada, N.; Isobe, H.; Irie, T. J Lab Clin Med 1998, 132, 432-439.
8. (a) Su, L.; Radek, J. T.; Hallenga, K.; Hermanto, P.; Chan, G.; Labeots, L. A.; Weiss, M. A. Biochemistry 1997, 36, 12722-12732; (b) Legault, P.; Li, J.; Mogridge, J.; Kay, L. E.; Greenblatt, J. Cell 1998, 93, 289-299.

Supplemental Figures

Figure S3.1 Helicity of reciprocal mutants. Circular dichroism measurements for λ_{N22} and reciprocal mutants in free solution (**A**) and in complex with λboxB (**B**). Measurements were taken in 10mM phosphate buffer at 20 °C. Concentration of peptide in free solution was 10 μM. Peptide-RNA complexes were prepared at 6:5 μM stoichiometry.
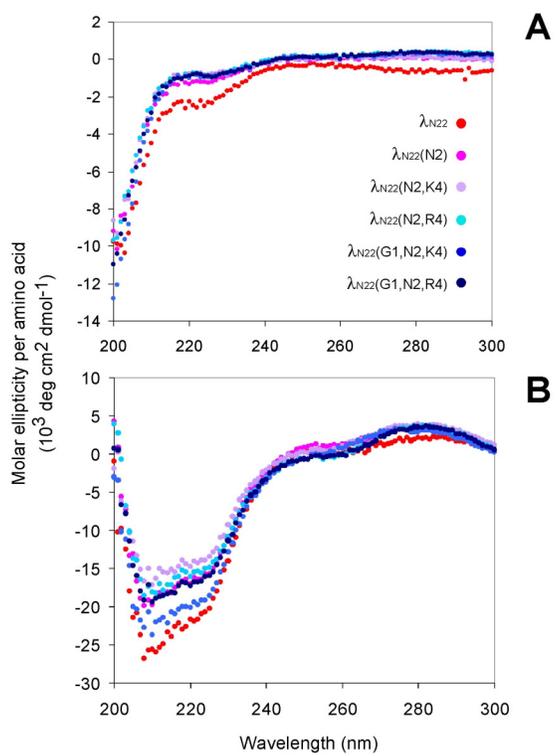
Figure S3.2 NMR spectra of reciprocal mutant complexes. NMR imino-spectra of

λ_{N22}-λboxB native and reciprocal complexes. The Trp18 imino-peak shifts from 10.0 ppm

in the free peptides to 9.2 ppm in the λboxB comlex. Samples were prepared in 50 mM

NaCl, 10 mM phosphate buffer, pH 6 and titrated to 1:1 stoichiometry at 25°C.

A

A   A

6G   A

5U A

C G12

C G13

C G14

1G C

U5 G14 G13 G12   G6   Ala3,Trp18

λ_{N22}-λboxB

λ_{N22}(N2)-λboxB

λ_{N22}(N2,K4)-λboxB

λ_{N22}(N2,R4)-λboxB

λ_{N22}(G1,N2,K4)-λboxB

λ_{N22}(G1,N2,R4)-λboxB

λ_{N22} peptide

λboxB RNA

13.5 13.0 12.5 12.0 11.5 11.0 10.5 10.0 9.5 9.0

ppm