

Chapter 2: Differential modes of recognition in N peptide-boxB complexes

Ryan J. Austin[†], Tianbing Xia[‡], Jinsong Ren[‡], Terry T. Takahashi[†], and Richard W. Roberts^{‡*}

(This chapter has been previously published: Biochemistry (2003) 42, 14957-67)

[†] 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

N proteins from bacteriophages λ , P22, and ϕ 21 modulate transcription elongation by binding nascent 'boxB' mRNA hairpins. This RNA recognition is mediated by N-terminal arginine-rich peptide sequences capable of interacting with their cognate boxB RNA targets. Here, we have analyzed the affinity and specificity of the peptide-RNA interactions that modulate this transcriptional switch. To do this, we constructed a series of peptides based on the wild type λ , P22, and ϕ 21 N protein binding domains ranging from 11 to 22 residues and analyzed their interactions with the leftward and rightward boxB RNA hairpin targets for all three phage. Binding constant (K_d) values were determined using RNA hairpins labeled with 2-aminopurine (2AP) and monitoring the fluorescence change as peptide was added. K_d values demonstrate that λ and P22 N peptides bind to their cognate boxB targets with high specificity and show equal affinities for their leftward and rightward hairpins. Surprisingly, ϕ 21 shows very little specificity for its cognate targets. λ and P22 N peptides exhibit differential modes of recognition with specificity conferred by their amino- and carboxy-terminal modules, respectively. We have generated a reciprocal matrix of substituted peptides to examine the contributions of individual residues to specificity. Amino acid coupling analysis supports a binding model where the Arg8 residue of λ peptide acts as a conformational hot-spot, anchoring the induced loop fold of its boxB hairpin target.

Introduction

The life cycle of lambdoid bacteriophage is coordinated by a transcription elongation switch. Late gene expression is triggered when host- and phage-encoded factors recruited to the transcription site modify bacterial RNA polymerase into a termination-resistant form (Reviewed in ref. (1-5)). Under physiological conditions, assembly of this antitermination complex is initiated by specific binding between bacteriophage N protein and cognate *cis*-acting N-utilization sites (nut sites) (6) within the leftward (nut_L) and rightward (nut_R) operons of nascent viral transcripts (7, 8). Phage specificity of this interaction is dictated by an amino-terminal RNA-binding domain of the N protein, termed the N peptide, that recognizes a cognate boxB hairpin sequence within the nut site (9).

Sequence-specificity between N peptide and boxB hairpins among related λ , P22, and ϕ 21 lambdoid bacteriophage, has made the N peptide-boxB complex an excellent system for the study of protein-RNA recognition. Homologous arginine-rich sequences, historically termed arginine-rich motifs (ARMs)(9), within these N peptides are features of a diverse family of prokaryotic and eukaryotic RNA binding proteins. Structural and biochemical studies of the λ , P22, and ϕ 21 N peptide-boxB complexes and other ARM-RNA complexes demonstrate a large degree of induced folding upon binding (10-14). This folding behavior, popularly termed mutually induced fit or adaptive recognition, is a characteristic of peptide-RNA binding interactions (Reviewed in ref. (15-19)).

The NMR structures of λ , P22, and ϕ 21 N peptide-boxB complexes have very similar architectures (Fig. 2.1A)(20-23). The N peptide adopts an α -helix with varying degrees

of bend in the three complexes, ranging from a pronounced 60° tilt in λ , to a nearly linear helix in $\phi 21$. All N peptides bind along the 5' edge of the major groove of their boxB targets. Interactions between the N peptide amino-terminal helix and the A-form helical RNA structure of the boxB stem are very similar. In each complex a conserved alanine residue makes non-polar contacts within the major groove and an R(R/Y)xxRR motif engages the boxB stem in a dense network of hydrogen bonds. In contrast, interactions between the N peptide carboxy-terminal helix and the boxB loop display unique folding variations. The pentaloops of λ and P22 boxB adopt discrete variations of a GNRA-tetraloop fold (24, 25) extruding either their fourth (4-out) or third (3-out) loop bases, respectively, while the $\phi 21$ boxB hexaloop adopts a distinct U-turn fold (Fig. 2.1B). Shape-specific non-polar loop contacts are seen in each complex (23, 26). In λ , a tryptophan residue π -stacks on top of the loop fold; in P22, adjacent alanine and isoleucine residues contact the extruded third base of the loop; and in $\phi 21$, an isoleucine packs against a ribose sugar within the U-turn.

Conserved interactions between the amino-terminal N peptide module and the boxB stem and unique contacts between the carboxy-terminal module and the boxB loop, imply a complementary mode of modular recognition within these complexes. In such a mode, the N peptide amino- and carboxy-terminal modules would be expected to act as nonspecific and specific binding elements, respectively. This recognition scheme predicts that alone, the amino-terminal module of an N peptide will not exhibit binding specificity for its cognate boxB target. Surprisingly, recent experiments from our laboratory contradict this prediction (27). When titrated against hybrid λ , P22, and tetraloop hairpin targets, the λ amino-terminal module retains the specificity of full-length λ N peptide.

Based on these findings, we decided to test the modular nature of N peptide binding specificity among the wild type λ , P22, and ϕ 21 bacteriophage systems.

We measured the binding specificity of three ‘full-length’ N peptides $\lambda_{N22} = \lambda$ N protein residues 1-22 (λ N(1-22)); **P22**_{N21} = P22 N(13-33); and **ϕ 21**_{N22} = ϕ 21 N(11-32) and their respective eleven residue amino-terminal modules $\lambda_{N11} = \lambda$ N(1-11); **P22**_{N11} = P22 N(13-23); and **ϕ 21**_{N11} = ϕ 21 N(11-21) against six boxB RNA hairpin targets, λ boxB_{R,L}, P22boxB_{R,L}, and ϕ 21boxB_{R,L} (Table 2.1). We find that both λ and P22 N peptides bind with high affinity and high specificity in solution, but show little discrimination between leftward and rightward hairpins. In contrast, ϕ 21 N peptide binds its cognate hairpins with weaker affinity and nominal specificity. Our investigations indicate that differential modes of recognition are operating in the λ and P22 N peptide-boxB complexes. Based on these results we have engineered peptides with greatly enhanced affinity for boxB hairpin targets (28).

Results

We began by examining salt-dependence measurements to quantify the electrostatic nature of binding in N peptide-boxB complexes, as well as temperature dependence and isothermal calorimetry measurements to determine binding enthalpies. We supplemented this data with stop-flow kinetic analysis of select complexes. Together, these parameters function as controls for our binding analysis and provide a thermodynamic context for our specificity findings.

Binding in protein-nucleic acid complexes is typically dominated by electrostatic interactions resulting from the negatively charged phosphate backbone of the poly-nucleotide and positively charged amino acids along the protein interface (Reviewed in ref. (33, 40)). This electrostatic force is characterized by the dependence of the association constant (K_{obs}) on monovalent salt concentration $[M^+]$, referred to as the salt-dependence $(-\partial \log K_{\text{obs}})(\partial \log [M^+])^{-1}$ of the interaction. Binding between a charged ligand and linear DNA or RNA, is described by Poisson-Boltzmann counterion condensation theory (Equation 2.2). The relationship between salt-dependence and ligand charge has been shown to accurately quantify electrostatic interactions between short poly-lysine or poly-arginine peptide ligands and uniformly charged linear RNA (41-43). Salt-dependence measurements of the N peptide-boxB complexes studied here offer a metric of electrostatic forces and provide the means to compare *in vitro* binding stabilities of these complexes over a broad range of conditions.

Binding isotherms for N peptide-boxB complexes were collected using RNA hairpins labeled with a 2-aminopurine (2AP) probe and monitoring fluorescence change

as peptide was added (Sample data are shown in Figure 2.2). Dissociation constant (K_d) and salt-dependence (∂/∂) values for N peptide-boxB hairpin complexes are listed in Table 2.2. K_d values are tabulated for a 70 mM $[M^+]$ standard salt buffer. Under these conditions, full-length λ_{N22} and P22_{N21} peptides bind cognate targets λ boxB_{L,R} and P22boxB_{L,R} with low nanomolar and picomolar affinities, respectively (Table 2.2). $\phi 21_{N22}$ binding is significantly weaker, exhibiting 2-orders of magnitude less affinity for its cognate targets $\phi 21$ boxB_{L,R}. A second $\phi 21$ N peptide construct, $\phi 21$ N(1-21), containing ten upstream residues previously implicated in antitermination activity (9), shows no detectable affinity for $\phi 21$ boxB (data not shown). To determine the impact of 2AP probe substitution on binding, we analyzed variations of each boxB RNA target with the 2AP label at distinct loop positions. We find that the effect of probe position on binding free energy and salt dependence is minimal (< 0.5 kcal mol⁻¹), indicating that 2AP does not disrupt electrostatic binding interactions in the N peptide-boxB complexes (Supporting information, Table S2.1). Monovalent salt-dependence data, plotted in Figure 2.3, demonstrate that the relative stabilities of cognate N peptide-boxB complexes are similar between salt concentrations in our standard buffer (70 mM $[M^+]$) and physiologic salt concentrations (150 mM $[M^+]$). K_d values calculated from salt-dependences for λ and P22 complexes are consistent with K_d values calculated from stopped-flow kinetic off-rate (k_{off}) and on-rate (k_{on}) measurements using equation 2.4 (λ_{N22} - λ boxB_R: $K_d = 1.9 \pm 0.7$ nM, $k_{off}/k_{on} = 1.0 \pm 0.2$ nM; P22_{N21}-P22boxB_L: $K_d = 5.0 \pm 2.0$ pM, $k_{off}/k_{on} = 12 \pm 2.0$ pM). The k_{on} rate used in these calculations was determined from the λ complex and is consistent with a diffusion controlled process (λ_{N22} - λ boxB: $k_{on} = 7 \times 10^8$ s⁻¹M⁻¹). Stopped-flow off-rate plots indicate that the relative stability of the P22 complex is due to its slower

dissociation rate (λ_{N22} - λ boxB: $k_{\text{off}} = 0.7 \pm 0.08 \text{ s}^{-1}$; P22_{N21}-P22boxB: $k_{\text{off}} = 0.01 \pm 0.001 \text{ s}^{-1}$).

Steady-state fluorescence measurements were used to determine the temperature-dependence of the cognate λ and P22 complexes at a high salt concentration (420 mM $[\text{M}^+]$). van't Hoff plots of this data demonstrate a linear temperature dependence for both complexes over the range of 20-40° C (Fig. 2.4A). Binding enthalpy (ΔH°) values were calculated from the slope of the van't Hoff plot using equation 2.3. We also directly measured ΔH° for λ and P22 complexes at our standard salt concentration (70 mM $[\text{M}^+]$) using isothermal calorimetry (ITC)(Fig 2.4B). ΔH° and ΔG° values are listed in Table 2.3. At both 420 mM and 70 mM salt, the λ and P22 complexes are stabilized by negative enthalpic contributions. Interestingly, van't Hoff and ITC measurements both show a correlation between the binding stability of the P22 complex and its binding enthalpy, relative to the λ complex ($\Delta\Delta G^\circ \sim \Delta\Delta H^\circ$).

Generally, temperature dependence data can also be analyzed to estimate the heat capacity change at constant pressure (ΔC_p). Heat capacity changes are important in macromolecular folding and binding because negative values correlate with the amount of hydrophobic surface that is buried (44). To address this issue, we examined our van't Hoff data for both P22 and λ complexes using a straight line ($\Delta C_p = 0$; Fig. 2.4) or a plot that took possible changes in ΔC_p into account. To estimate an upper bound for $|\Delta C_p|$ we used the published value of 2200 Å^2 total surface buried from the structurally similar $\phi 21$ N peptide-boxB complex (23) and the relationship between ΔC_p and buried non-polar surface area (ΔA_{np}), $\Delta C_p/\Delta A_{\text{np}} = -0.28 \pm 0.05 \text{ cal mol}^{-1} \text{ K}^{-1} \text{ Å}^{-2}$ (44). We further estimated that at most, 50% of the buried surface is actually hydrophobic. These assumptions

produce an estimate of $-\Delta C_p \sim 310 \text{ cal mol}^{-1} \text{ K}^{-1}$. Even though our van't Hoff data in Figure 2.4 show a high level of precision, they are well fit by a straight line that assumes $\Delta C_p = 0$ as well as by a curved line corresponding to $-\Delta C_p \sim 310$. Our results thus illustrate the challenge of experimentally measuring the hydrophobic burial in RNA-peptide complexes.

Binding Specificity

Both λ_{N22} and P22_{N21} full-length peptides discriminate strongly between their cognate RNA hairpins and other boxB targets: $\Delta\Delta G^\circ \geq 2.5 \text{ kcal mol}^{-1}$ (Table 2.4). In contrast $\phi 21_{N21}$ peptide exhibits significantly less specificity for its cognate hairpin: $\Delta\Delta G^\circ \leq 1.5 \text{ kcal mol}^{-1}$. All three amino-terminal peptide modules bind cognate boxB hairpins with affinities near 1 μM . Interestingly, the specificity of full-length λ_{N22} is largely retained in its amino-terminal module, λ_{N11} : $\Delta\Delta G^\circ > 2.2 \text{ kcal mol}^{-1}$ (Fig. 2.5; Table 2.4). Distinctly, P22_{N11} and $\phi 21_{N11}$ peptides exhibit no specificity for their cognate hairpins. These results indicate that λ and P22 N peptides employ differential modes of recognition with cognate specificity largely attributable to the amino-terminal module in λ and the carboxy-terminal module in P22.

To determine the residues responsible for binding specificity in the λ_{N11} peptide we reciprocally interchanged non-conserved amino acids between the specific λ_{N11} and the nonspecific P22_{N11} peptides. These two peptides vary at 4 of their 11 residues: positions 1, 2, 4, and 8. Thus, a reciprocal matrix of 16 substituted peptides allowed us to explore all combinatorial variations between the λ and P22 amino-terminal peptide sequences (Table 2.5). For convenience in the text we use the subscript notation $\lambda_{N11(\text{MIG})}$

to denote the peptide generated by an M1G reciprocal substitution to λ_{N11} . Conversely, this peptide could also be denoted P22_{N11(N2D,K4Q,H8R)}.

M1G, D2N, and Q4K reciprocal substitutions increase the binding affinity of λ_{N11} for both λ_{boxB} and P22_{boxB} targets, while an R8H mutation abolishes binding. We reasoned that increased peptide affinities for λ_{boxB} arising from P22 residues Asn2 and Lys4 were due to electrostatic forces as both D2N and Q4K substitutions result in a +1 increase in net peptide charge. This prediction was tested by measuring the salt-dependence of peptide- λ_{boxB} binding for each separate reciprocal substitution of λ_{N11} peptide: $\lambda_{N11(M1G)}$, $\lambda_{N11(D2N)}$, and $\lambda_{N11(Q4K)}$. As anticipated, the D2N and Q4K substitutions increased δ/δ values while the M1G substitution had no effect (λ_{N11} - λ_{boxB_R} : $\partial/\partial = 2.6 \pm 0.1$; $\lambda_{N11(M1G)}$ - λ_{boxB_R} : $\partial/\partial = 2.6 \pm 0.1$; $\lambda_{N11(D2N)}$ - λ_{boxB_R} : $\partial/\partial = 2.8 \pm 0.1$; $\lambda_{N11(Q4K)}$ - λ_{boxB_R} : $\partial/\partial = 3.2 \pm 0.2$). To determine if Lys4 was interacting specifically with λ_{boxB} , we mutated this residue to an arginine. The resultant peptide, $\lambda_{N11(Q4R)}$, gives a similar ∂/∂ value, indicating that electrostatic interactions between the Lys4 residue and λ_{boxB} are nonspecific ($\lambda_{N11(Q4R)}$ - λ_{boxB_R} : $\partial/\partial = 3.2 \pm 0.2$).

Free energies in Table 2.5 demonstrate that λ residues Gln4 and Arg8 are largely responsible for the binding specificity of λ_{N11} . To illustrate the specificity of these residues from the ΔG° values in Table 2.5, it is most clear to consider reciprocal substitutions to the nonspecific P22_{N11} peptide. However, the reciprocal matrix does exhibit internal symmetry and similar observations can be made from following reciprocal substitutions to the λ_{N11} peptide, which result in decreased specificity for λ_{boxB} target. The net-specificity ($\Delta\Delta G$) value in Table 2.5 is an absolute difference in

binding free energy corrected for the 0.1 kcal mol⁻¹ nominal specificity of P22_{N11} peptide ($\Delta\Delta G = |\Delta G^\circ(\text{peptide}-\lambda\text{boxB}_R) - \Delta G^\circ(\text{peptide}-\text{P22boxB}_L)| - 0.1 \text{ kcal mol}^{-1}$).

We see in Table 2.5 that the reciprocal substitution K4Q results in a 0.9 kcal mol⁻¹ increase in peptide net-specificity. Inspection of the individual ΔG° values informs us that the Gln4 residue decreases the affinity of P22_{N11} peptide for both λboxB and P22boxB targets (P22_{N11}- λboxB_R : $\Delta G^\circ = -8.2 \text{ kcal mol}^{-1}$; P22_{N11(K4Q)}- λboxB_R : $\Delta G^\circ = -6.9 \text{ kcal mol}^{-1}$; P22_{N11}-P22boxB_L: $\Delta G^\circ = -8.1 \text{ kcal mol}^{-1}$; P22_{N11(K4Q)}-P22boxB_L: $\Delta G^\circ = -5.9 \text{ kcal mol}^{-1}$). This decrease in affinity is smaller for the λboxB target (+1.3 kcal mol⁻¹) than for the P22boxB target (+2.2 kcal mol⁻¹). Therefore, the individual Gln4 residue is responsible for an increase in net peptide specificity towards λboxB (0.9 kcal mol⁻¹). Separately, the individual H8R substitution to P22_{N11} generates 0.3 kcal mol⁻¹ net peptide specificity towards λboxB . When introduced in tandem, the K4Q and H8R substitutions generate a net peptide specificity of 2.1 kcal mol⁻¹. This net-specificity is larger than the sum of the individual net-specificity contributions of Gln4 (0.9 kcal mol⁻¹) and Arg8 (0.3 kcal mol⁻¹) and equal to the net-specificity of the cognate λ_{N11} peptide (2.1 kcal mol⁻¹). In P22_{N11(K4Q,H8R)}, specificity arises from a favorable coupling energy (ΔG_1) exerted between Gln4 and Arg8 toward λboxB target. ΔG_1 is distilled from individual residue contributions to binding free energy using equation 2.5. Gln4 and Arg8 exhibit a 1.0 kcal mol⁻¹ favorable coupling energy toward λboxB target, but exhibit a nominal 0.1 kcal mol⁻¹ coupling energy toward P22boxB target (P22_{N11(K4Q,H8R)}- λboxB_R : $\Delta G_1 = -1.0 \text{ kcal mol}^{-1}$; P22_{N11(K4Q,H8R)}-P22boxB_L: $\Delta G_1 = -0.1 \text{ kcal mol}^{-1}$). In combination, the individual net-specificities of Gln4 (0.9 kcal mol⁻¹) and Arg8 (0.3 kcal mol⁻¹) and the relative coupling energy exhibited between the two residues towards the λboxB hairpin (0.9 kcal

mol⁻¹), account for the total net-specificity of P22_{N11(K4Q,H8R)} peptide (2.1 kcal mol⁻¹). Interestingly, large coupling energies are observed between Arg8 and other reciprocal substitutions suggesting that Arg8 plays a role in folding of the cognate λ complex. A favorable coupling energy ($\Delta G_1 = -0.7$ kcal mol⁻¹) between Arg8 and the carboxy-terminal module of λ N peptide was determined using chimeric-peptides (P22_{N11}- λ boxB_R: $\Delta G^\circ = -8.2$ kcal mol⁻¹; P22_{N11(H8R)}- λ boxB_R: $\Delta G^\circ = -9.9$ kcal mol⁻¹; P22_{N11(λ :12-22)}- λ boxB_R: $\Delta G^\circ = -11.7$ kcal mol⁻¹; P22_{N11(H8R)(λ :12-22)}- λ boxB_R: $\Delta G^\circ = -14.1$ kcal mol⁻¹). In comparison, an unfavorable coupling energy ($\Delta G_1 = +1.1$ kcal mol⁻¹) is observed between Arg8 and the carboxy-module of P22 N peptide, indicating incompatibility of these two elements (P22_{N21}- λ boxB_R: $\Delta G^\circ = -12.1$ kcal mol⁻¹; P22_{N21(H8R)}- λ boxB_R: $\Delta G^\circ = -12.7$ kcal mol⁻¹). Arg8 ΔG_1 values are summarized in Figure 2.6.

These results support a role for Gln4 and Arg8 in the adaptive recognition and discrimination of the λ boxB and P22boxB hairpin loop sequences. The largely conserved λ boxB_R and P22boxB_L targets tested with our reciprocal matrix of peptides, however, have sequence-variability in both their loop (nucleotide 8) and stem (base pair 3,13) regions. Thus it remained conceivable that Gln4 and Arg8 were discriminating the boxB stem sequences of λ and P22. To test this possibility we measured K_d 's for the λ _{N11} and P22_{N11(K4Q,H8R)} peptides against a substituted series of hairpins: λ boxB_R, λ boxB_{R(C3G,G13C)}, λ boxB_{R(A8C)}, and λ boxB_{R(C3G,G13C)(A8C)} (note that λ boxB_{R(C3G,G13C)(A8C)} = P22boxB_L hairpin). Binding free energies of these complexes demonstrate that peptides recognize the loop and not the stem sequence of λ boxB_R (Supporting Information, Table S2.2)

The distinct loop sequences of λ (A8) and P22 (C8) hairpins are expected to result in the unique loop folds of these targets. However, it is likely that different λ and P22 loop folds transmit structural variability to their respective stem folds as well. Therefore, the highly specific Gln4 residue could be discriminating between λ and P22 boxB structures through any of its observed hydrogen bond contacts with the conserved loop-base G6 and stem-bases U5 and G12 in the λ N peptide-boxB structure (21, 22). We have investigated these interactions with NMR spectra of substituted λ_{N11} - λ boxB_R complexes. The reciprocal substitution Q4K causes a 0.3 ppm downfield shift in the G6 imino-peak of the λ_{N11} - λ boxB_R complex, as well as smaller shifts in the imino-peaks of U5 and G12 bases (Fig. 2.7). The complementary substitution K4Q to the P22_{N11} peptide also results in shifts of the U5 and G12 imino-resonances, although the G6 imino-peak is not resolved in the P22_{N11}- λ boxB_R and P22_{N11(K4Q)}- λ boxB_R complexes (data not shown). The data support that the Gln4 residue of λ recognizes both the loop and stem of λ boxB target.

Unlike Gln4, the specific Arg8 residue of λ N peptide interacts exclusively with the loop of λ boxB (21, 22). The Arg8 guanidinium group contacts the irregular backbone of the loop between phosphates pA9 and pA10. In the unique 4-out pentaloop fold of λ boxB, these two negatively charged phosphates are packed close to one another, generating unfavorable electrostatic forces, which are likely mitigated by the presence of the positively charged guanidinium group of Arg8. From the NMR structure we can predict a model for Arg8 recognition consistent with our coupling analysis, in which the Arg8 residue specifically stabilizes a 4-out pentaloop fold.

Binding Mechanism

We next measured binding-induced structural changes within the RNA hairpin loops by monitoring the change in environment of the 2AP probe. 2AP fluorescence is quenched by base stacking and enhanced by exposure to aqueous solvent, but is insensitive to base pairing and other hydrogen bonding interactions (45, 46). Titration of N peptide into boxB RNA results in a change in fluorescence reported here as a fraction, $F = F_C/F_R$, where F_R and F_C are fitted values for the fluorescence of RNA alone and the RNA-peptide complex, respectively (Equation 2.1). 2AP fluorescence changes in the λ and P22 cognate complexes are consistent with their NMR structures. 2AP bases at either loop position 2 (2AP-2) or 3 (2AP-3) of the λ boxB_R hairpin show stacking in the λ_{N22} - λ boxB_R complex while 2AP at loop position 4 (2AP-4) is exposed to solution in agreement with a 4-out pentaloop fold with Trp18 stacking on loop base 2 (λ_{N22} - λ boxB_R: $F_{2AP-2} = 0.30$, $F_{2AP-3} = 0.20$, $F_{2AP-4} = 2.40$). In the P22_{N21}-P22boxB_L complex, 2AP-2 is exposed to solvent and 2AP-4 is stacked in the loop corresponding to a 3-out pentaloop fold (P22_{N21}-P22boxB_L: $F_{2AP-2} = 1.80$, $F_{2AP-4} = 0.60$, Fig. 2.8A). Changes in the fluorescence signal of 2AP at positions 3 and 4 of the ϕ 21boxB hairpin upon ϕ 21_{N21} peptide binding are less dramatic, but consistent with the NMR structure of the complex (23), in which nucleotides A8 (2AP-3) and A9 (2AP-4) continue the 3' base stack of the ϕ 21 boxB hexaloop (ϕ 21boxB: $F_{2AP3} = 1.30$, $F_{2AP4} = 0.60$). This folding pattern is distinct from the GNRA-like pentaloops of λ boxB and P22boxB cognate complexes.

Strikingly, the non-cognate complexes λ_{N22} -P22boxB_L and P22_{N21}- λ boxB_R exhibit loop structures dictated by peptide identity. λ_{N22} -P22boxB_L complex adopts a λ -like 4-out pentaloop conformation (λ_{N22} -P22boxB_L: $F_{2AP-2} = 0.30$; $F_{2AP-4} = 3.50$), while non-cognate

P22_{N21}-λboxB_R complex exhibits an intermediate change in fluorescence (P22_{N21}-λboxB_R: F_{2AP-2} = 1.60; F_{2AP-4} = 1.00) (Fig. 2.8B). NMR imino-spectra of the two complexes reveal absence of the sheared G6 imino-peak and broadening of the Trp18 imino-peak in the λ_{N22}-P22boxB_L complex, consistent with multiple loop conformations in both non-cognate structures (Fig. 2.8C).

Reciprocal substitution of N peptide residue 8 dramatically alters the fluorescence signatures of these non-cognate complexes. The Arg8 substituted P22 N peptide induces a 4-out, λ-like loop in the λboxB (P22_{N21(H8R)}-λboxB_R: F_{2AP-4} = 2.00). In contrast, the λ-like 4-out pentaloop fluorescence of λ_{N22}-P22boxB_L is abrogated by the His8 substitution (λ_{N22(R8H)}-P22boxB_L: F_{2AP-4} = 1.00)(Fig. 2.8B). We observe a similar result from reciprocal substitution of the P22_{N11λN12-22}-P22boxB_L chimera complex (P22_{N11λN12-22}-P22boxB_L: F_{2AP-2} = 1.30, F_{2AP-4} = 1.10; P22_{N11(H8R)λN12-22}-P22boxB_L: F_{2AP-2} = 0.40; F_{2AP-4} = 3.70). Fluorescence data for the substituted amino-terminal 11mer peptide complexes exhibits a trend comparable to the full-length peptides (Supporting information, Table S2.3). From the fluorescence and NMR data, we conclude that N peptide sequence influences bound hairpin-loop structure and that the presence of Arg8 can confer 4-out pentaloop stacking to both λ and P22 sequences.

Discussion

Thermodynamic characterization of cognate complexes

Using a set of 2AP labeled RNAs we have investigated the thermodynamic and kinetic properties of N peptide-boxB complexes from the lambdoid family of bacteriophage. Equilibrium and kinetic data indicate that the P22 cognate complex is roughly 3 kcal mol⁻¹ more stable than the λ cognate complex under standard buffer conditions. ITC and temperature dependence data demonstrate a correlation between the binding stability of the P22 complex and its binding enthalpy relative to the λ complex ($\Delta\Delta G^\circ \sim \Delta\Delta H^\circ$). Additionally, salt-dependence measurements indicate a greater electrostatic binding potential in the P22 complex (P22_{N21}-P22boxB_L: $\partial/\partial = 5.2 \pm 0.3$; λ_{N22} - λ boxB_R: $\partial/\partial = 3.5 \pm 0.2$). The relationship between the relative enthalpic binding contribution to the P22 complex and its relative electrostatic potential, however, is not evident from this correlation. Typically, salt-dependences are characteristic of entropic forces, attributable to the release of counterions from the RNA upon formation of a complex (47). These entropic effects drive binding events in nonspecific, polycationic peptide-RNA interactions (41-43). Distinctly, the P22 and λ N peptide-boxB complexes exhibit large unfavorable binding entropies, consistent with highly ordered, mutually induced folding in ARM-RNA complexes (16).

Electrostatic interactions between basic residues in the peptide and the phosphate backbone of the RNA are presumably a major component of binding free energy in the N peptide-boxB complexes. A quantitative analysis of salt-dependence measurements based on Poisson-Boltzmann counterion-condensation models (Equation 2.2) is limited by

irregular charge distributions and end-effects in the short hairpin complexes (48). These effects complicate structural calculations of the ψ value, which is a function of the linear distance between phosphates. Structurally predicted ψ values for ds DNA (0.88) and ds RNA (0.9) have been experimentally measured for the nonspecific binding of oligo-arginine and oligo-lysine to nucleotide duplexes ($\psi = 0.9(\pm 0.05)$)(34, 43). Interestingly, we observe a correlation between salt-dependence and net peptide charge consistent with this ψ value in most N peptide-boxB complexes studied ($\psi = 0.9(\pm 0.1)$); complexes deviating from this relationship are indicated in supporting information, Table S2.1). Thus, the binding ΔG° attributable to electrostatic interactions in the N peptide-boxB complexes is quantitatively similar to the charge-dependent binding ΔG° of nonspecific complexes. This effect may be attributable to the small size of these complexes, as most charges on an N peptide bind within or adjacent to the atmosphere of thermodynamically bound counterions surrounding the RNA. It should be noted that within the range of this correlation, the salt-dependences of individual amino acid substitutions are not strictly additive (data not shown), consistent with previous investigations of compact protein-RNA complexes (49).

The N peptide-boxB complexes span a wide range of binding affinities with the P22 cognate complex exhibiting roughly 6 kcal mol⁻¹ greater stability than the ϕ 21 cognate complex. The biological function of these variable stabilities remains unclear. Previous mutagenic studies of λ phage (51,52) and recent combinatorial investigations from our laboratory (50, 51) indicate that the structural context and conformation of the N peptide-boxB complex, not its binding stability, dictates the function of the transcriptional switch. These structural properties are likely transmitted to function

through binding interactions with other components of the antitermination complex, which are conserved in the various phage. Structural mimicry in the λ , P22, and ϕ 21 N peptide-boxB architectures supports this notion (23, 26).

Binding Specificity

λ and P22 N peptides exhibit differential modes of recognition with specificity largely attributable to their amino- and carboxy-terminal modules, respectively. Sequence identity in the arginine-rich amino-terminal modules of these peptides has allowed for a comprehensive examination of the amino acid determinants of λ N peptide specificity. Coupling analysis using a reciprocal matrix of substituted peptide sequences has identified two residues, Gln4 and Arg8, which are instrumental to the specificity of λ N peptide. Tandem substitution of these two residues into the nonspecific P22_{N11} amino-terminal module results in a peptide with λ boxB specificity. The versatile specificities of N peptides resulting from discrete amino acid substitutions shown here are consistent with previous observations that the arginine-rich motif provides a flexible framework for recognition of RNA structures (52, 53).

Based on NMR and fluorescence measurements reported here and insight from the solution structures of N peptide-boxB complexes (20-23), we propose a model for cognate recognition by λ N peptide. In the λ cognate complex, Arg8 acts as a conformational hot-spot, anchoring a 4-out pentaloop fold through electrostatic or bifurcated arginine-fork (54, 55) contacts with the tightly packed pA9 and pA10 phosphates of the boxB backbone. This unique 4-out pentaloop fold transmits structural

changes to the loop-closing sheared base-pair and hairpin stem, which are recognized by the Gln4 residue of λ N peptide.

Our model offers a simple mechanistic rationale for the discrimination of non-cognate P22boxB hairpin by the λ_{N11} peptide that is consistent with the collected fluorescence data. The energetic penalty for non-cognate binding of λ_{N11} is largely incurred from rearrangement of the 3-out, GA(C)AA GNRA-like pentaloop of P22boxB into a 4-out, GAC(A)A loop conformation, or distribution of conformations. Such a rearrangement would alter the stable GN(Y)RA structure into a GNY(R)A structure (R = purine base, Y = pyrimidine base), nullifying purine stacking interactions and destabilizing the heterogeneous GNRA hydrogen bonding network (25). Coupling between the Arg8 residue and the carboxy-terminal module of λ N peptide (Fig. 2.6) is presumably mediated through the hairpin's 4-out pentaloop fold, which justifies the observed weak binding affinity of P22_{N11(λ :12-22)} and $\lambda_{N11(P22:12-21)}$ modular chimeric peptides for boxB targets (data not shown).

One advantage of modularly organized binding proteins with independent binding domains is that modules can be swapped or arrayed to construct desired binding specificities and affinities. This approach has previously utilized helix-turn-helix (HTH), zinc-finger, and ARM modules with some success (Reviewed in ref. (56-59)). Simple swapping of amino- and carboxy-terminal modules in the N peptide system fails to generate peptides with high affinity for boxB targets due to coupling effects between these modules, which are likely transmitted through the boxB RNA folding framework. Reciprocal substitution of individual residues in tandem with modular swapping,

however, has allowed us to engineer N peptide hybrids with greatly enhanced binding affinities (28).

In addition to their utility in protein design methods, peptide binding modules are also used in selection techniques to investigate specific interactions and reduce the sequence-space of combinatorial binding searches (59). mRNA-display selection experiments from our laboratory employing the λ_{N11} peptide module to search for boxB hairpin binders have isolated amino acid substitutions D2N and Q4K (60), identical to tight-binding substitutions characterized from our reciprocal analysis (Table 2.5). This result suggests that reciprocal engineering approaches may be of particular utility in reducing the sequence-space of future combinatorial binding searches.

Experimental Procedures

RNA and Peptide Synthesis.

Peptides were generated on an Applied Biosystems 432A peptide synthesizer using solid phase, Fmoc 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. These peptides were quantified using an extinction coefficient of $\epsilon = 1.197 \times 10^3 \text{ M}^{-1}\text{cm}^{-1}$ for tyrosine.

Unlabeled RNA hairpins (λ boxB_{R15} and P22boxB_{L15}) were synthesized by *in vitro* transcription using T7 RNA polymerase (29). 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 260 nm wavelength.

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

Steady-State Fluorescence Measurements.

Steady-state fluorescence measurements were conducted as previously reported, by monitoring the change in fluorescence of a 2AP probe incorporated in boxB RNA (27). Titrations were performed on a Shimadzu Spectrofluorophotometer at 20° C with Excitation/Emission wavelengths at 310/370 nm. Peptides were titrated into a constantly stirred solution of 2AP labeled RNA hairpin (20-2000 nM). Binding buffer contained a

variable concentration of KOAc (15-500 mM) at pH 7.5. All binding constants reported in the text are for a standard 20 mM Tris•OAc buffer condition, containing 50 mM KOAc (70 mM [M⁺]) unless indicated otherwise. Individual binding constants were fit to a single-site association model (Equation 2.1) by non-linear least squares regression using the computer program DynaFit (30). Variables [P], [R], and [C] represent the equilibrium concentrations of peptide, RNA target, and peptide-RNA complex, respectively. [P]₀ and [R]₀ are the initial concentrations of peptide and RNA target in solution. The dissociation constant (K_d), free RNA fluorescence (F_R), and fluorescence of the peptide-RNA complex (F_C) were fit as free parameters for all titrations. Free peptide fluorescence (F_P) was set to zero for all peptides except λ_{N22}, which contains a tryptophan residue. Fluorescence values reported in the text were determined from fits at standard conditions unless stated otherwise.

$$F_T = ([P]_0 - [C])F_P + ([R]_0 - [C])F_R + [C]F_C \quad (\text{Eq. 2.1})$$

$$[C] = \frac{1}{2} ([P]_0 + [R]_0 + K_d - \{([P]_0 + [R]_0 + K_d)^2 - 4[P]_0[R]_0\}^{1/2})$$

Binding constants required a change in fluorescence of greater than 10% for reliable fitting. Salt-dependence plots were determined from five or more binding constants within a range of salt concentrations (35-520 mM KOAc + Tris•OAc). These plots deviated from linearity at low salt conditions (\leq 50 mM [M⁺]), consistent with previous accounts in the literature (31-33).

Binding between a charged ligand and linear DNA or RNA is described by Poisson-Boltzmann counterion condensation theory (Equation 2.2) where [M⁺] is the concentration of monovalent cation, Z is the net charge of ligand, and ψ is the number of counterions thermodynamically bound to the poly-nucleotide per phosphate (34, 35).

$$(\partial \log K_{obs}) (\partial \log [M^+])^{-1} = -Z\psi \quad (Eq. 2.2)$$

Steady-state temperature dependence measurements were used to calculate thermodynamic parameters of the cognate complexes following equation 2.3, where R is the gas constant and T is the absolute temperature.

$$\Delta G^\circ = -RT \ln K_{obs} = \Delta H^\circ - T\Delta S^\circ \quad (Eq. 2.3)$$

Stopped-Flow Fluorescence Measurements.

Stopped-flow fluorescence experiments were conducted following the procedures of Lacourciere et al. (36). Measurements were performed at 20° C under standard buffer conditions (20 mM Tris•OAc, 50 mM KOAc, pH 7.5) using a stopped-flow device from Applied Photophysics (Surrey, U.K.) in two-syringe mode. Dissociation rate constant (k_{off}) values were determined for interactions by infusing labeled complex with both 100x and 500x concentration of unlabeled competitor boxB at time zero. Dissociation constants were calculated from rate constants using equation 2.4.

$$K_d = K_{obs}^{-1} = (k_{off})(k_{on})^{-1} \quad (Eq. 2.4)$$

Coupling Energy.

The free energy of two individual binding components X and Y can be related to the free energy of the pair X,Y by equation 5, where ΔG_1 is the coupling energy (37, 38).

$$\Delta \Delta G_{(X,Y)} = \Delta \Delta G_{(X)} + \Delta \Delta G_{(Y)} + \Delta G_1 \quad (Eq. 2.5)$$

NMR Spectroscopy.

NMR samples were prepared in NMR buffer: 50 mM NaCl, 0.5 mM EDTA, 10 mM phosphate, pH 6 in H₂O:D₂O (90:10, vol:vol) as previously described (27). NMR spectra were collected at 15° 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 (39). Assignments were based on reported work (10, 21, 22).

Isothermal Calorimetry.

Calorimetric data were obtained using a MicroCal isothermal microtitration calorimeter. Serial additions of RNA were injected into a large excess of peptide solution at standard buffer conditions. Specifically, a stock solution of 40 μM RNA was titrated into a 1.38 ml solution of 60 μM peptide within the reaction cell, usually making 15 injections of 10 μL. Integration of each titration peak and normalization for the number of moles of ligand added provided a direct and model-free estimate of the binding enthalpy. Heat of dilution was determined by injections of RNA into binding buffer alone and was subtracted from the enthalpy to give a corrected value for binding (ΔH°).

Acknowledgements

We would like to thank Jeffrey Barrick for his suggestions and Dr. Adam Frankel for his comments on the manuscript. This work was supported by the NSF (9876246), NIH (RO1 60416), and Beckman Foundation. R.W.R. is as an Alfred P. Sloan Foundation Research Fellow.

References

1. Das, A. (1993) *Annu Rev Biochem* 62, 893-930.
2. Greenblatt, J., Nodwell, J. R., and Mason, S. W. (1993) *Nature* 364, 401-6.
3. Roberts, J. W. (1993) *Cell* 72, 653-5.
4. Friedman, D. I., and Court, D. L. (1995) *Mol Microbiol* 18, 191-200.
5. Weisberg, R. A., and Gottesman, M. E. (1999) *J Bacteriol* 181, 359-67.
6. Salstrom, J. S., and Szybalski, W. (1978) *J Mol Biol* 124, 195-221.
7. Franklin, N. C. (1985) *J Mol Biol* 181, 75-84.
8. Franklin, N. C. (1985) *J Mol Biol* 181, 85-91.
9. Lazinski, D., Grzadzilska, E., and Das, A. (1989) *Cell* 59, 207-18.
10. Su, L., Radek, J. T., Hallenga, K., Hermanto, P., Chan, G., Labeots, L. A., and Weiss, M. A. (1997) *Biochemistry* 36, 12722-32.
11. Tan, R., and Frankel, A. D. (1995) *Proc Natl Acad Sci U S A* 92, 5282-6.
12. Cilley, C. D., and Williamson, J. R. (1997) *RNA* 3, 57-67.
13. Van Gilst, M. R., Rees, W. A., Das, A., and von Hippel, P. H. (1997) *Biochemistry* 36, 1514-24.
14. Mogridge, J., Legault, P., Li, J., Van Oene, M. D., Kay, L. E., and Greenblatt, J. (1998) *Mol Cell* 1, 265-75.
15. Weiss, M. A., and Narayana, N. (1998) *Biopolymers* 48, 167-80.
16. Draper, D. E. (1999) *J Mol Biol* 293, 255-70.
17. Patel, D. J. (1999) *Curr Opin Struct Biol* 9, 74-87.
18. Frankel, A. D. (2000) *Curr Opin Struct Biol* 10, 332-40.
19. Williamson, J. R. (2000) *Nat Struct Biol* 7, 834-7.
20. Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A., and Patel, D. J. (1998) *Nat Struct Biol* 5, 203-12.
21. Legault, P., Li, J., Mogridge, J., Kay, L. E., and Greenblatt, J. (1998) *Cell* 93, 289-99.
22. Scharpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S., and Rosch, P. (2000) *Eur J Biochem* 267, 2397-408.
23. Cilley, C. D., and Williamson, J. R. (2003) *RNA* 9, 663-676.
24. Heus, H. A., and Pardi, A. (1991) *Science* 253, 191-4.
25. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) *J Mol Biol* 264, 968-80.
26. Weiss, M. A. (1998) *Nat Struct Biol* 5, 329-33.
27. Barrick, J. E., Takahashi, T. T., Ren, J., Xia, T., and Roberts, R. W. (2001) *Proc Natl Acad Sci U S A* 98, 12374-8.
28. Austin, R. J., Xia, T., Ren, J., Takahashi, T. T., and Roberts, R. W. (2002) *J. Am. Chem. Soc.* 124, 10966-10967.
29. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res* 15, 8783-98.
30. Kuzmic, P. (1996) *Anal Biochem* 237, 260-73.
31. deHaseth, P. L., Gross, C. A., Burgess, R. R., and Record, M. T., Jr. (1977) *Biochemistry* 16, 4777-83.

32. Lundback, T., and Hard, T. (1996) *J. Phys. Chem.* 100, 17690-17695.
33. Record, M. T., Jr., Ha, J. H., and Fisher, M. A. (1991) *Methods Enzymol* 208, 291-343.
34. Record, M. T., Jr., Lohman, M. L., and De Haseth, P. (1976) *J Mol Biol* 107, 145-58.
35. Record, M. T., Jr., Anderson, C. F., and Lohman, T. M. (1978) *Q Rev Biophys* 11, 103-78.
36. Lacourciere, K. A., Stivers, J. T., and Marino, J. P. (2000) *Biochemistry* 39, 5630-41.
37. Carter, P. J., Winter, G., Wilkinson, A. J., and Fersht, A. R. (1984) *Cell* 38, 835-40.
38. Ackers, G. K., and Smith, F. R. (1985) *Annu Rev Biochem* 54, 597-629.
39. Liu, W., Okajima, K., Murakami, K., Harada, N., Isobe, H., and Irie, T. (1998) *J Lab Clin Med* 132, 432-9.
40. Record, M. T., Jr., Zhang, W., and Anderson, C. F. (1998) *Adv Protein Chem* 51, 281-353.
41. Mascotti, D. P., and Lohman, T. M. (1992) *Biochemistry* 31, 8932-46.
42. Mascotti, D. P., and Lohman, T. M. (1993) *Biochemistry* 32, 10568-79.
43. Mascotti, D. P., and Lohman, T. M. (1997) *Biochemistry* 36, 7272-9.
44. Spolar, R. S., Ha, J. H., and Record, M. T., Jr. (1989) *Proc Natl Acad Sci U S A* 86, 8382-5.
45. Millar, D. P. (1996) *Curr Opin Struct Biol* 6, 322-6.
46. Rachofsky, E. L., Osman, R., and Ross, J. B. (2001) *Biochemistry* 40, 946-56.
47. Lohman, T. M., and Mascotti, D. P. (1992) *Methods Enzymol* 212, 400-24.
48. Draper, D. E. (1995) *Annu Rev Biochem* 64, 593-620.
49. GuhaThakurta, D., and Draper, D. E. (2000) *J Mol Biol* 295, 569-80.
50. Xia, T., Frankel, A., Takahashi, T. T., Ren, J., and Roberts, R. W. (2003) *Nat Struct Biol* in press.
51. Xia, T., Becker, H.-C., Wan, C., Frankel, A., Roberts, R. W., and Zewail, A. H. (2003) *Proc Natl Acad Sci U S A* in press.
52. Harada, K., Martin, S. S., and Frankel, A. D. (1996) *Nature* 380, 175-9.
53. Harada, K., Martin, S. S., Tan, R., and Frankel, A. D. (1997) *Proc Natl Acad Sci U S A* 94, 11887-92.
54. Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) *Science* 252, 1167-71.
55. Tao, J., and Frankel, A. D. (1992) *Proc Natl Acad Sci U S A* 89, 2723-6.
56. Chandrasegaran, S., and Smith, J. (1999) *Biol Chem* 380, 841-8.
57. Segal, D. J., and Barbas, C. F., 3rd. (2000) *Curr Opin Chem Biol* 4, 34-9.
58. Wolfe, S. A., Nekludova, L., and Pabo, C. O. (2000) *Annu Rev Biophys Biomol Struct* 29, 183-212.
59. Cheng, A. C., Calabro, V., and Frankel, A. D. (2001) *Current Opinion in Structural Biology* 11, 478-84.
60. Barrick, J. E., and Roberts, R. W. (2002) *Protein Sci* 11, 2688-96.
61. Garcia-Garcia, C., and Draper, D. E. (2003) *J Mol Biol* 331, 75-88.

Figures and Tables

Figure 2.1 Structural and schematic models of λ , P22, and ϕ 21 N peptide-boxB complexes. (A) Structural models of phage λ , P22, and ϕ 21 N peptide-boxB hairpin complexes viewed from the major groove (20, 22, 23)(models are adapted from Weiss (26)). In all complexes the N peptide is shown as a ribbon with amino- and carboxy-terminal modules in black and gray, respectively. Homologous hydrophobic interactions occur between the boxB hairpin stem and conserved alanine residues within the N peptide amino-terminal module (10). Distinct hydrophobic interactions appear between the boxB hairpin loops and carboxy-terminal modules of the N peptides. In λ , a tryptophan residue stacks on the boxB loop (21); in P22, non-polar alanine and isoleucine residues interact with an extruded pyrimidine (20); and in ϕ 21, an isoleucine residue packs against a ribose sugar within the hairpin U-turn (23). Models were generated with PyMol (<http://www.pymol.org>). (B) Schematic representation of λ , P22, and ϕ 21 boxB hairpin loops viewed from the minor groove (diagrams are adapted from solution structures (20, 21, 23)). In phage λ and P22, the bound pentaloops adopt stable GNRA tetraloop folds (24) by extruding either loop base 4 (4-out) in the λ complex, or loop base 3 (3-out) in the P22 complex. Boxed letters in black and gray denote bases and amino acids, respectively. Open circles with loop position numbers denote ribose sugars and solid black circles represent backbone phosphates. Solid black rectangles indicate stabilizing interactions.

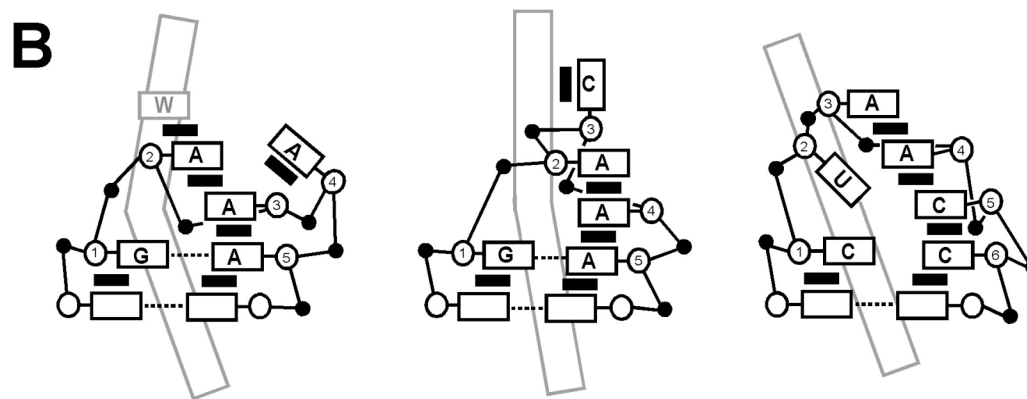
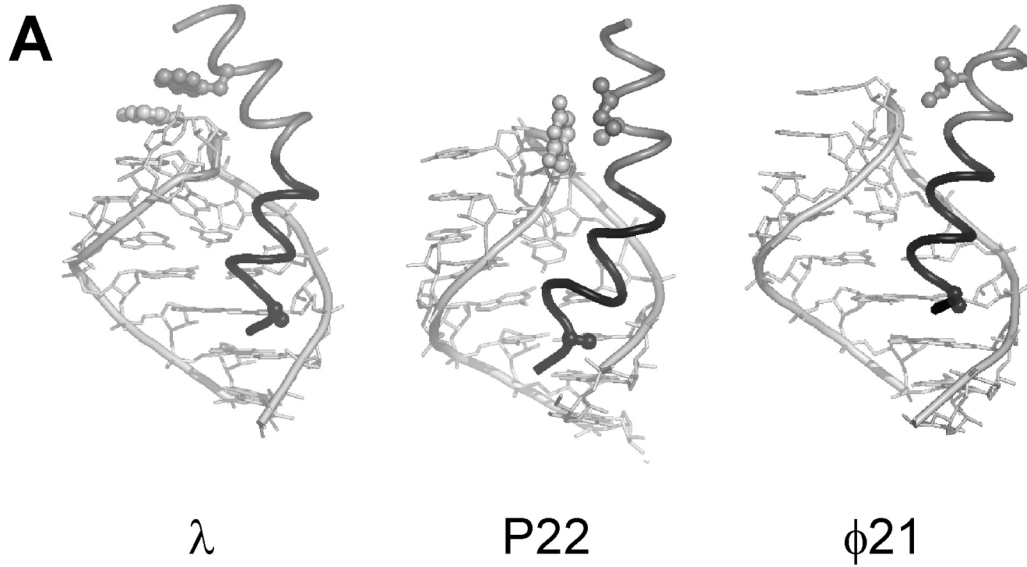


Figure 2.2 Fluorescence measurements of λ_{N22} - λ boxB_R (2AP-2) complex. (A) Steady-state titration of λ boxB_R(2AP-2) RNA with λ_{N22} peptide at 200 mM KOAc. Plot includes raw data (open circles), data corrected for tryptophan fluorescence of free peptide (open squares), and the K_d fit (black line). **(B)** Salt-dependence for λ_{N22} - λ boxB_R(2AP-2) over a range of 150-500 mM KOAc. **(C)** Stopped-flow timed measurement of the λ_{N22} - λ boxB_R(2AP-2) complex dissociation rate infused with 100x concentration of unlabeled λ boxB_R at time zero (data in gray; k_{off} fit is black line). **(D)** Dependence of the pseudo first-order rate constant on peptide concentration for the association rate of λ_{N22} with λ boxB_R(2AP-2) RNA.

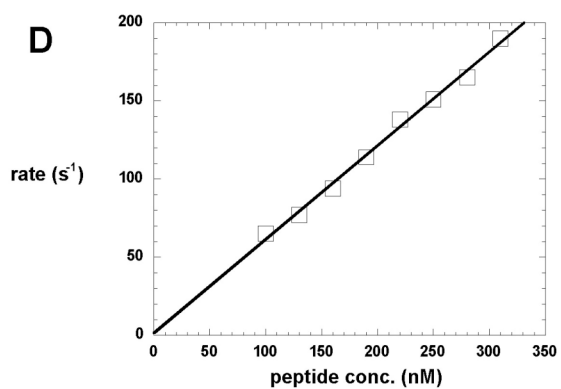
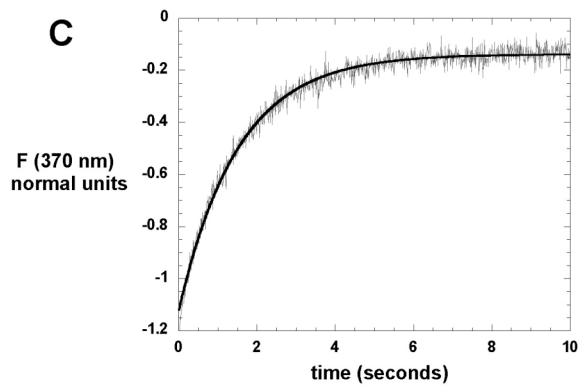
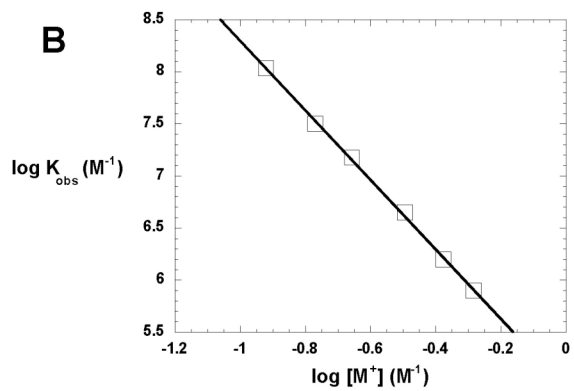
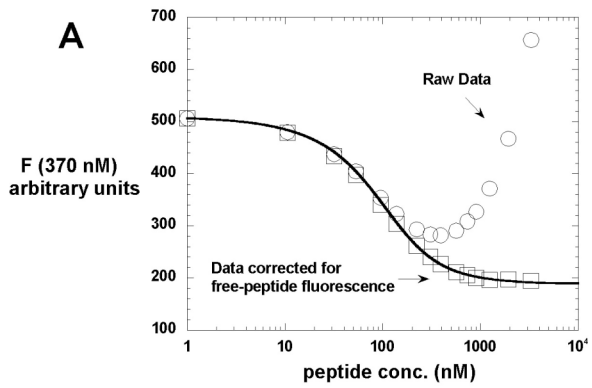


Figure 2.3 Salt dependence plots for λ , P22, and ϕ 21 cognate complexes.

Salt-dependence plots for λ (open), P22 (black), and ϕ 21 (gray) cognate complexes (dotted lines with triangle data and full lines with circle data represent N peptide-boxB_L and -boxB_R complexes, respectively). Relative binding affinities of the various complexes are maintained between standard buffer (70 mM [M⁺]) and physiologic salt concentrations (150 mM [M⁺]).

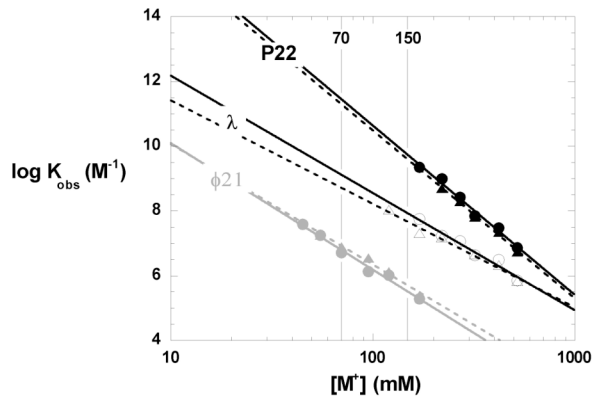


Figure 2.4 van't Hoff plots and ITC data for λ and P22 cognate complexes. (A) van't Hoff plots of steady-state fluorescence measurements for the λ_{N22} - λ boxB_R and P22_{N21}-P22boxB_L complexes at 400 mM KOAc. (B) Primary ITC data for the serial addition of P22boxB_L RNA into excess P22_{N21} peptide at standard conditions.

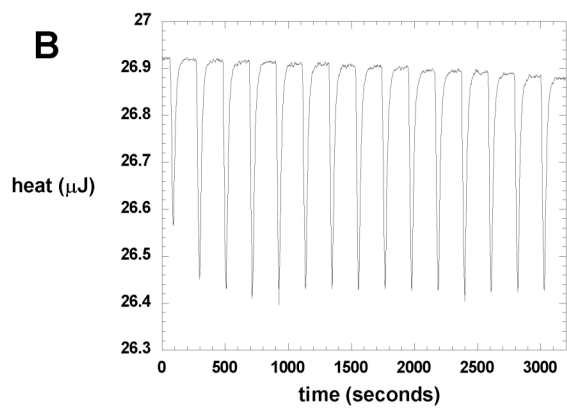
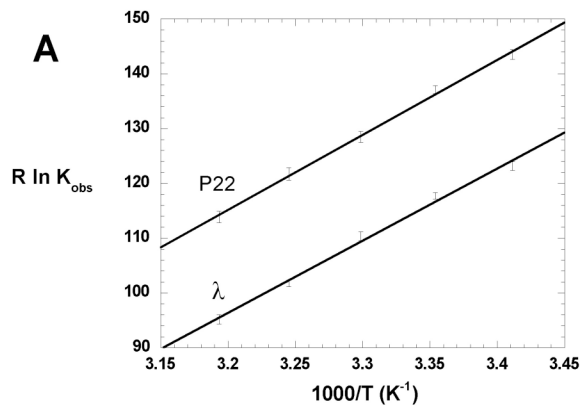


Figure 2.5 Free energy values for λ and P22 complexes. ΔG° values for λ and P22 full-length and amino-terminal peptide modules. The amino-terminal 11 residues of λ N peptide exhibit binding specificity similar to the full-length λ_{N22} peptide, while the P22 amino-terminal 11 residues bind with minimal discrimination between cognate and non-cognate targets.

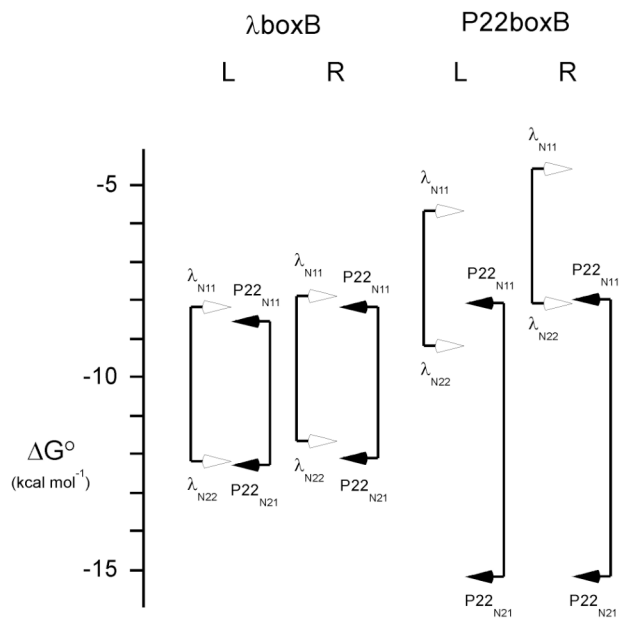


Figure 2.6 Coupling energies. Favorable coupling energies ($-\Delta G_1$) reported in kcal mol⁻¹, for λ N peptide residues bound to cognate hairpin. ΔG_1 values were determined from reciprocal substitution of variable residues within the P22_{N11} peptide. ΔG_1 values among amino-terminal residues were found to be internally consistent with ΔG_1 values from complementary substitutions to the λ_{N11} peptide (± 0.1 kcal mol⁻¹). The ΔG_1 between Arg8 and the carboxy-terminal module of λ N peptide was calculated from the binding free energies of P22_{N11}, P22_{N11(H8R)}, and P22_{N11(H8R)(λ :12-22)} peptides against λ boxB_R hairpin.

Favorable coupling energies (kcal mol⁻¹)

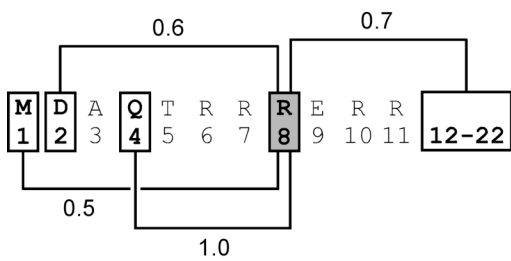


Figure 2.7 NMR spectra of λ_{N11} and $\lambda_{N11(Q4K)}$. NMR spectra of λ_{N11} - λ boxB_R complexes. Imino-proton peaks for the four base pairs of the boxB hairpin stem and the sheared G:A base pair are labeled. Reciprocal substitution Q4K of the λ_{N11} peptide shifts λ boxB_R stem (U5, G12) and sheared base-pair (G6) imino-proton peaks in the $\lambda_{N11(Q4K)}$ - λ boxB_R complex.

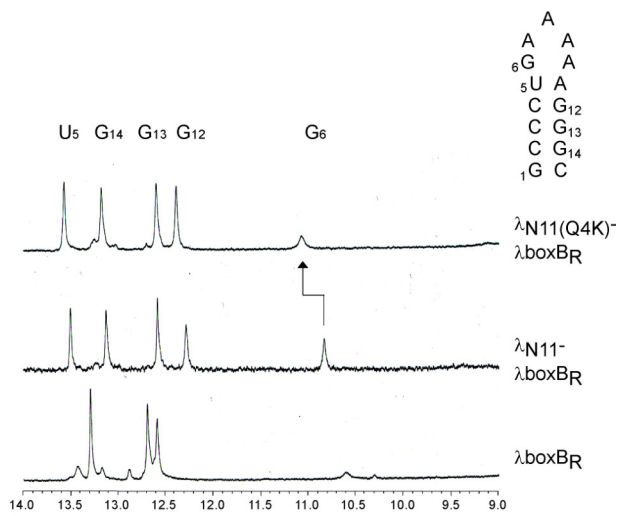


Figure 2.8 Fluorescence and NMR measurements of complexes. (A) Fluorescence spectra for the titration of 2AP-labeled P22boxB_L targets with cognate P22_{N21}. Black spectra represent free hairpin; progressively lighter spectra indicate serial addition of peptide. An increase in fluorescence of 2AP is caused by exposure to aqueous solvent; a decrease in fluorescence is the result of base stacking (45, 46). Titrations of P22_{N21} into labeled P22boxB_L hairpins exhibit increased fluorescence of 2AP-2 and decreased fluorescence of 2AP-4 consistent with the solution structure of cognate complex (20). (B) Relative change in fluorescence (F_{2AP-4}) of λ boxB_R and P22boxB_L targets upon binding of cognate peptides and $\lambda_{N22(R8H)}$ and P22_{N21(H8R)}} reciprocal substituted peptides. Binding to λ boxB_R and P22boxB_L hairpin is indicated in open bars and black bars, respectively. Fluorescence signatures of non-cognate complexes are dramatically influenced by the identity of residue eight. The P22- λ boxB_R complex showed no change in 2AP-4 fluorescence. F_{2AP-4} values for the $\lambda_{N22(R8H)}$ -P22boxB_L complex were determined at 25 mM [M⁺]. (C) NMR spectra of boxB complexes. Four base pair imino-protons, ranging from 12-14 ppm, are identifiable from the boxB stem in both λ boxB_R and P22boxB_L complexes. Positions of the sheared G6 imino-peak and the Ala3 amide proton peak are further downfield in P22_{N22} peptide complexes. Spectra were recorded at 15° C.

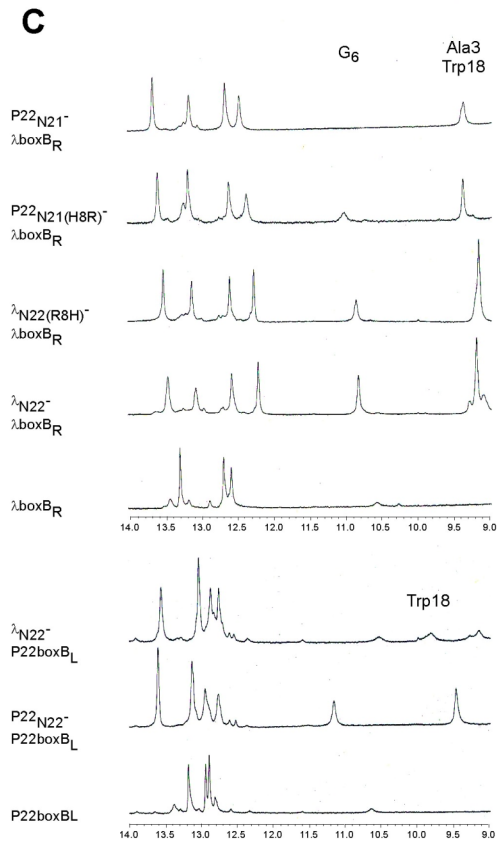
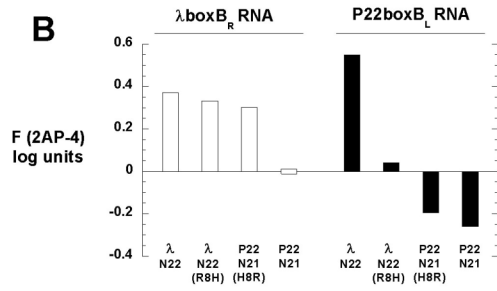
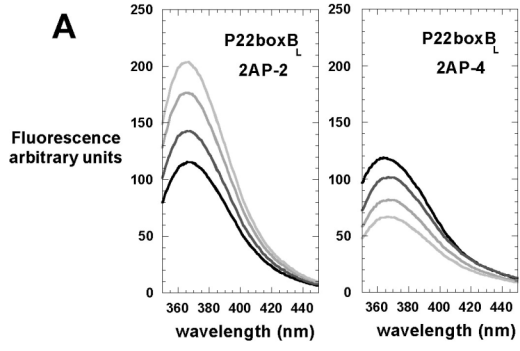


Table 2.1. N peptide and boxB RNA sequences ^a

| Peptide | Sequence |
|-----------------------------|---------------------------|
| λ_{N11} | MDAQTRRRERR gY |
| λ_{N22} | MDAQTRRRERR AEKQAQWKAAN |
| P22 _{N11} | GNAKTRRHERR gY |
| P22 _{N21} | GNAKTRRHERR RKLAIERDTI gY |
| $\phi 21_{N11}$ | GTAksRYKARR gY |
| $\phi 21_{N22}$ | GTAksRYKARR AELIAERRSNE |
| RNA | Sequence |
| λ boxB _L | gGCCCU GAAGA AGGGCc |
| λ boxB _R | gGCCCU GAAAA AGGGCc |
| P22boxB _L | gGCGCU GACAA AGCGCc |
| P22boxB _R | gACCGCC GACAA CGCGGUc |
| $\phi 21$ boxB _L | gCUCAAC CUAACC GUUGAGc |
| $\phi 21$ boxB _R | gCACCU CUAACC GGGUGc |

^a Sequence of bacteriophage N peptides. The sequences of λ_{N22} , P22_{N21}, and $\phi 21_{N22}$ “full-length” peptides correspond to λ , P22, and $\phi 21$ N protein residues 1-22, 13-33, and 11-32, respectively (non-native residues are in lower-case).

Table 2.2. Peptide-RNA binding constants (K_d) and salt-dependence values (∂/∂)^{a,b}

| N peptide | λ RNA | | P22 RNA | | $\phi 21$ RNA | |
|--------------------|---|---------------------|---|---------------------|---|---------------------|
| | K_d (nM) | ∂/∂ | K_d (nM) | ∂/∂ | K_d (nM) | ∂/∂ |
| | <u>A</u> G G A U A C G C G C G G C g c | | <u>A</u> A G A U A C G C G C G G C g c | | <u>A</u> A U C C C C G A U A U C G U A C G g c | |
| | boxB _L -2 | | boxB _R -2 | | boxB _L -2 | |
| | K_d (nM) | ∂/∂ | K_d (nM) | ∂/∂ | K_d (nM) | ∂/∂ |
| λ_{N22} | † 0.8 | 3.6 | † 1.9 | 3.2 | 260 | 3.3 |
| P22 _{N21} | † 0.6 | 4.3 | † 0.9 | 4.1 | † 0.005 | 5.2 |
| $\phi 21_{N22}$ | 1,300 | 4.0 | 1,800 | 4.5 | 800 | 3.5 |
| | | | | | nd | |
| λ_{N11} | 650 | 2.6 | 1,200 | 2.6 | 54,000 | |
| P22 _{N11} | 390 | 3.7 | 750 | 3.9 | 840 | 3.9 |
| $\phi 21_{N11}$ | 1,100 | 5.1 | 1,800 | 5.3 | 1,100 | 4.9 |
| | | | | | 1,500 | 3.4 |
| | | | | | 3,700 | 4.5 |
| | | | | | 9,100 | 3.4 |
| | | | | | 340 | 4.3 |
| | | | | | 160 | 3.9 |
| | | | | | 120 | 3.7 |
| | | | | | 5,000 | 3.2 |
| | | | | | 320 | 5.1 |
| | | | | | C G | C G |
| | | | | | A U | A U |
| | | | | | C G | C G |
| | | | | | g c | g c |
| | | | | | nd ^c | nd |
| | | | | | 16,000 | 13,000 |
| | | | | | 2,200 | |

^a Binding constants were determined for standard conditions: 20° C; 50 mM KOAc, 20 mM Tris•Oac, pH 7.5. Individual isotherms were fit to a one-step reaction with less than 10% error. Hairpin base positions substituted with 2AP are underlined. ^b All salt-dependence values ($\partial/\partial = (-\partial \log K_{obs}) / (\partial \log [M^+])^{-1}$) were determined from five or more steady state binding constants within a range of 5 to 500 mM KOAc. ^c nd values were not determined due to weak binding or weak binding signal.

† Binding constants determined by extrapolation from salt-dependence plot; errors of calculated salt-dependence values against kinetic calculations ($k_{off}/k_{on} = K_d$) were less than 0.5 kcal mol⁻¹.

Table 2.3. Thermodynamic Parameters ^a

| complex | van't Hoff (420 mM [M ⁺]) | | ITC (70 mM [M ⁺]) | |
|-----------|---|---|--|---|
| | ΔG° (kcal mol ⁻¹) | ΔH° (kcal mol ⁻¹) | ΔG° ^b (kcal mol ⁻¹) | ΔH° (kcal mol ⁻¹) |
| λ | -8.6 | -31.5 ± 1.5 | -11.7 | -31.6 ± 2.0 |
| P22 | -9.9 | -32.8 ± 0.9 | -15.2 | -35.4 ± 1.5 |
| | $\Delta\Delta G^\circ = 1.3$ | $\Delta\Delta H^\circ = 1.3$ | $\Delta\Delta G^\circ = 3.5$ | $\Delta\Delta H^\circ = 3.8$ |

^a ΔG° values and van't Hoff ΔH° values were calculated from steady-state measurements using equation 2.3. ^b ΔG° values were calculated from fluorescence measurements.

Table 2.4. Binding free energy ΔG° (kcal mol⁻¹)^a

| peptide | λ RNA | | P22 RNA | | ϕ 21 RNA | |
|--------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | boxB _L 2 | boxB _R 2 | boxB _L 2 | boxB _R 4 | boxB _L 3 | boxB _R 3 |
| λ_{N11} | -8.2 | -7.9 | -5.7 | -4.6 | nd | nd |
| λ_{N22} | -12.2 | -11.7 | -9.2 | -8.1 | -6.7 | -7.1 |
| P22 _{N11} | -8.6 | -8.2 | -8.1 | -8.0 | -6.4 | -6.5 |
| P22 _{N21} | -12.3 | -12.1 | -15.2 | -15.2 | -8.6 | -8.7 |
| ϕ 21 _{N11} | -7.9 | -7.6 | -8.0 | -7.8 | -7.2 | -7.5 |
| ϕ 21 _{N22} | -7.8 | -7.6 | -8.1 | nd | -9.1 | -9.2 |

^a ΔG° values at standard conditions (293.15 K) were calculated from steady-state measurements using equation 2.3.

Table 2.5. Reciprocal mutant binding free energies ^a

| Peptides (11mers) | RNA (2AP-2) | | Net-Specificity ^b | | |
|--------------------------|---|----------------------|---|-----|--------|
| | ΔG° kcal mol ⁻¹ | | $\Delta\Delta G$ kcal mol ⁻¹ | | |
| | λ boxB _R | P22boxB _L | | | |
| λ _{N11} | MDA QTRRRERRgy | -7.9 | -5.7 | 2.1 | ██████ |
| | GDA QTRRRERRgy | -8.2 | -6.0 | 2.1 | ██████ |
| | MNA QTRRRERRgy | -9.2 | -6.9 | 2.2 | ██████ |
| | MDA KTRRRERRgy | -8.8 | -7.5 | 1.2 | ██ |
| | MDA QTRRHERRgy | -4.5 | nd ^c | | |
| | MNA KTRRRERRgy | -9.8 | -8.9 | 0.8 | ██ |
| | MNA QTRRHERRgy | -6.1 | nd | | |
| | MDA KTRRHERRgy | -6.1 | -5.8 | 0.2 | █ |
| | GDA QTRRHERRgy | -4.9 | nd | | |
| | GDA KTRRRERRgy | -8.6 | -7.7 | 0.8 | ██ |
| | GNA QTRRRERRgy | -9.6 | -7.4 | 2.1 | ██████ |
| | GNA KTRRRERRgy | -9.9 | -9.5 | 0.3 | █ |
| | GNA QTRRHERRgy | -6.9 | -5.9 | 0.9 | ██ |
| | GDA KTRRHERRgy | -6.3 | -6.2 | 0 | |
| MNA KTRRHERRgy | -7.6 | -7.4 | 0.1 | █ | |
| P22 _{N11} | GNA KTRRHERRgy | -8.2 | -8.1 | 0 | |

^a ΔG° values at standard conditions (293.15 K) were calculated from steady-state measurements using equation 2.3. ^b The net-specificity ($\Delta\Delta G$) is the absolute free energy contribution to λ boxB_R binding specificity from λ N peptide residues. This value is corrected for the nominal 0.1 kcal mol⁻¹ specificity of P22_{N11} ($\Delta\Delta G = |\Delta G^\circ(\text{peptide-}\lambda\text{boxB}_R) - \Delta G^\circ(\text{peptide-P22boxB}_L)| - 0.1 \text{ kcal mol}^{-1}$) (λ N peptide residues are denoted in **boldface**). ^c ΔG° value could not be determined due to weak binding or weak binding signal.

Supplemental Tables

Table S2.1. Thermodynamic parameters ^a

| RNA | N Peptide | | | | | |
|--|--------------------|-----------------|--------------------|--------------------|-----------------|-----------------|
| | λ_{N11} | λ_{N22} | P22 _{N11} | P22 _{N21} | $\phi 21_{N11}$ | $\phi 21_{N22}$ |
| Dissociation Constant (nM) | | | | | | |
| $\lambda\text{boxB}_L\text{-2}$ | 650 ± 385 | 0.76 ± 0.36 † | 390 ± 140 | 0.6 ± 0.07 † | 1,100 ± 1,000 | 1,300 ± 650 |
| $\lambda\text{boxB}_L\text{-3}$ | | 0.99 ± 0.41 † | | | | 1,800 ± 360 ‡ |
| $\lambda\text{boxB}_R\text{-2}$ | 1,200 ± 260 | 1.9 ± 0.69 † | 750 ± 280 | 0.85 ± 0.28 † | 1,800 ± 920 | 1,800 ± 710 |
| $\lambda\text{boxB}_R\text{-3}$ | | 1.0 ± 0.34 † | | | | 1,900 ± 210 ‡ |
| $\lambda\text{boxB}_R\text{-4}$ | | 1.2 ± 0.59 † | | | | |
| P22 $\text{boxB}_L\text{-2}$ | | 260 ± 81 | 840 ± 710 | 0.005 ± 0.002 † | 1,100 ± 540 | 800 ± 380 |
| P22 $\text{boxB}_L\text{-4}$ | 55,000 ± 1,900 ‡ | 110 ± 96 | 750 ± 80 ‡ | 0.004 ± 0.002 † | | |
| P22 $\text{boxB}_R\text{-2}$ | | 530 ± 320 | | | | |
| P22 $\text{boxB}_R\text{-4}$ | 313,000 ± 33,000 ‡ | 820 ± 710 | 970 ± 390 | 0.005 ± 0.002 † | 1,500 ± 860 | |
| $\phi 21\text{boxB}_L\text{-3}$ | | 9,100 ± 5,400 | 16,000 ± 1,500 ‡ | 340 ± 240 | 3,700 ± 3,200 | 160 ± 95 |
| $\phi 21\text{boxB}_L\text{-4}$ | | 9,900 ± 1500 | | | | 200 ± 180 |
| $\phi 21\text{boxB}_R\text{-3}$ | | 5,000 ± 1,800 | 13,000 ± 1,900 ‡ | 320 ± 210 | 2,800 ± 220 ‡ | 120 ± 80 |
| $\phi 21\text{boxB}_R\text{-4}$ | | 4,200 ± 800 | | | 1,200 ± 70 ‡ | 65 ± 52 |
| Salt-Dependence ($-\partial \log K_{\text{obs}})(\partial \log [M^+])^{-1}$) | | | | | | |
| $\lambda\text{boxB}_L\text{-2}$ | 2.6 ± 0.2 | 3.6 ± 0.3 | 3.7 ± 0.1 | 4.3 ± 0.1 | 5.1 ± 0.4 | 4.0 ± 0.2 |
| $\lambda\text{boxB}_L\text{-3}$ | | 3.3 ± 0.3 | | | | |
| $\lambda\text{boxB}_R\text{-2}$ | 2.6 ± 0.1 | 3.2 ± 0.2 | 3.9 ± 0.1 | 4.1 ± 0.2 | 5.3 ± 0.2 § | 4.5 ± 0.2 § |
| $\lambda\text{boxB}_R\text{-3}$ | | 3.3 ± 0.2 | | | | |
| $\lambda\text{boxB}_R\text{-4}$ | | 3.0 ± 0.2 § | | | | |
| P22 $\text{boxB}_L\text{-2}$ | | 3.3 ± 0.1 | 3.9 ± 0.3 | 5.2 ± 0.3 | 4.9 ± 0.2 | 3.5 ± 0.2 |
| P22 $\text{boxB}_L\text{-4}$ | | 3.1 ± 0.3 § | | 5.0 ± 0.3 | | |
| P22 $\text{boxB}_R\text{-2}$ | | 3.6 ± 0.2 | | | | |
| P22 $\text{boxB}_R\text{-4}$ | | 3.4 ± 0.2 | 3.6 ± 0.1 | 5.2 ± 0.2 | 3.4 ± 0.2 § | |
| $\phi 21\text{boxB}_L\text{-3}$ | | 3.4 ± 0.2 | | 4.3 ± 0.2 | 4.5 ± 0.3 | 3.9 ± 0.2 |
| $\phi 21\text{boxB}_L\text{-4}$ | | | | | | 3.6 ± 0.3 |
| $\phi 21\text{boxB}_R\text{-3}$ | | 3.2 ± 0.1 | | 5.1 ± 0.3 | | 3.7 ± 0.3 |
| $\phi 21\text{boxB}_R\text{-4}$ | | | | | | 3.8 ± 0.3 |

^a Dissociation constants listed for standard conditions: 50 mM KOAc, 20 mM Tris•OAc; 20° C; pH 7.5. K_d values were calculated from salt-dependence data and deviate < 15% from experimental values determined at standard conditions.

† Values listed were determined by extrapolation from salt-dependence data and were found to be within 0.5 kcal mol⁻¹ agreement with K_d values calculated from stopped-flow kinetic data ($k_{\text{off}}/k_{\text{on}}$).

‡ Error estimates indicate the precision of single fits.

§ Data deviates from the Poisson-Boltzmann salt-dependence correlation ($\partial \log K_{\text{obs}})(\partial \log [M^+])^{-1} = -0.9(\pm 0.1)z$.

Table S2.2. BoxB RNA substitutions ^a

| | A A A G A U A C G C G C G G C g c λ boxB _R | A A A G A U A C G G C C G G C g c λ boxB _R (C3G,G3C) | C A A G A U A C G C G C G G C g c λ boxB _R (A8C) | C A A G A U A C G G C C G G C g c λ boxB _R (C3G,G3C) (A8C) 'P22boxB _L ' |
|---|--|---|--|--|
| | ΔG° (kcal mol ⁻¹) | | | |
| λ _{N11} MDAQTRRRERRgy | -7.9 | -8.3 | -5.7 | -5.7 |
| P22 _{N11(K4Q,H8R)} GNA Q TRRR R ERRgy | -9.6 | -9.9 | -7.4 | -7.4 |

^a Free energies listed for standard conditions: 50 mM KOAc, 20 mM Tris•OAc; 20° C; pH 7.5. Substituted residues in **bold**

Table S2.3. Reciprocal mutant peptides, thermodynamic and fluorescent parameters ^a

| Peptides (11mers) | | RNA | | | | | |
|-------------------|---------------|------------------------|-------|-------------|----------------------|-------|-------|
| | | λboxB_R | | | P22boxB _L | | |
| | | K_d | F | | K_d | F | |
| | | | 2AP-2 | 2AP-2 | | 2AP-4 | 2AP-2 |
| λ_{N11} | MDAQTRRRERRgy | 1,200 ± 32 | 1.50 | 2.20 | 54,000 ± 1,500 | 0.90 | 1.30 |
| | GDAQTRRRERRgy | 730 ± 3.6 | 1.60 | | 28,000 ± 2,500 | 0.90 | 1.40 |
| | MNAQTRRRERRgy | 120 ± 10 | 1.60 | | 6,600 ± 370 | 0.90 | 1.70 |
| | MDAKTRRRERRgy | 250 ± 12 | 1.70 | | 2,300 ± 300 | 0.90 | 1.60 |
| | MDAQTRRHERRgy | 410,000 ± 53,000 | 1.70 | | nd | | |
| | MNAKTRRRERRgy | 45 ± 2.6 | 1.70 | | 210 ± 18 | 0.90 | 1.90 |
| | MNAQTRRHERRgy | 23,000 ± 380 | 1.70 | | nd | | |
| | MDAKTRRHERRgy | 28,000 ± 560 | 1.70 | | 43,000 ± 4,100 | 1.20 | 0.80 |
| | GDAQTRRHERRgy | 200,000 ± 11,000 | 2.00 | | nd | | |
| | GDAKTRRRERRgy | 360 ± 9 | 1.60 | | 1,700 ± 180 | 0.90 | 1.50 |
| | GNAQTRRRERRgy | 66 ± 3.8 | 1.60 | | 2,700 ± 240 | 0.90 | 1.70 |
| | GNAKTRRRERRgy | 41 ± 5.7 | 1.80 | | 79 ± 8.6 | 0.90 | 1.80 |
| | GNAQTRRHERRgy | 7,000 ± 140 | 1.70 | | 36,000 ± 2,500 | 1.10 | 1.00 |
| | GDAKTRRHERRgy | 17,000 ± 590 | 1.60 | | 22,000 ± 1,700 | 1.20 | 0.90 |
| MNAKTRRHERRgy | 1,900 ± 48 | 1.70 | | 2,800 ± 110 | 1.20 | 0.90 | |
| $P22_{N11}$ | GNAKTRRHERRgy | 750 ± 55 | 1.70 | 1.80 | 840 ± 51 | 1.20 | 0.90 |