

# The RNA-Binding Domain of Bacteriophage P22 N Protein Is Highly Mutable, and a Single Mutation Relaxes Specificity toward $\lambda^{\nabla}$

Alexis I. Coccozaki, Ingrid R. Ghattas, and Colin A. Smith\*

*Department of Biology, American University of Beirut, Beirut, Lebanon*

Received 19 July 2008/Accepted 12 September 2008

**Antitermination in bacteriophage P22, a lambdoid phage, uses the arginine-rich domain of the N protein to recognize boxB RNAs in the nut site of two regulated transcripts. Using an antitermination reporter system, we screened libraries in which each nonconserved residue in the RNA-binding domain of P22 N was randomized. Mutants were assayed for the ability to complement N-deficient virus and for antitermination with P22 boxB<sub>left</sub> and boxB<sub>right</sub> reporters. Single amino acid substitutions complementing P22 N<sup>-</sup> virus were found at 12 of the 13 positions examined. We found evidence for defined structural roles for seven nonconserved residues, which was generally compatible with the nuclear magnetic resonance model. Interestingly, a histidine can be replaced by any other aromatic residue, although no planar partner is obvious. Few single substitutions showed bias between boxB<sub>left</sub> and boxB<sub>right</sub>, suggesting that the two RNAs impose similar constraints on genetic drift. A separate library comprising only hybrids of the RNA-binding domains of P22,  $\lambda$ , and  $\phi$ 21 N proteins produced mutants that displayed bias. P22 N<sup>-</sup> plaque size plotted against boxB<sub>left</sub> and boxB<sub>right</sub> reporter activities suggests that lytic viral fitness depends on balanced antitermination. A few N proteins were able to complement both  $\lambda$  N- and P22 N-deficient viruses, but no proteins were found to complement both P22 N- and  $\phi$ 21 N-deficient viruses. A single tryptophan substitution allowed P22 N to complement both P22 and  $\lambda$  N<sup>-</sup>. The existence of relaxed-specificity mutants suggests that conformational plasticity provides evolutionary transitions between distinct modes of RNA-protein recognition.**

Lambdoid phages  $\lambda$ , P22,  $\phi$ 21, and other phages regulate the expression of delayed early genes by allowing transcription past terminators in the P<sub>left</sub> and P<sub>right</sub> operons. These phages share regulatory mechanisms with  $\lambda$ , but they have uncertain evolutionary relationships that are obscured by recombination among tailed phages (*Caudovirales*) (7, 8). In phage  $\lambda$ , antitermination allows expression of genes regulating the development of lysis or lysogeny (12). The assembly of transcription antitermination complexes in P22,  $\lambda$ , and  $\phi$ 21 is initiated by the binding of viral N proteins to small hairpin boxB RNAs in the nut sites (N utilization) of regulated transcripts (40). These complexes contain N and host factors, including NusA and the transcribing polymerase, allowing transcription to proceed through downstream transcription termination signals. P22,  $\lambda$ , and  $\phi$ 21 exhibit type specificity, where the N protein of one virus cannot complement its absence in a different virus (13, 24).

N proteins recognize their cognate boxB RNAs via arginine-rich domains near their amino termini (21). Their boxBs are hairpin RNAs that have little sequence similarity, yet similar secondary structures (Fig. 1A). Alignment of P22,  $\lambda$ , and  $\phi$ 21 reveals that N protein RNA-binding domains contain four conserved amino acids (Fig. 1B). Protein and RNA sequence differences create specificity; noncognate interactions function poorly. boxBs bind noncognate N peptides poorly in vitro (1, 10, 39). Likewise, noncognate N-nut interactions do not function in vivo (19, 29), and noncognate N proteins do not rescue

N-deficient viruses (13, 24). Compared to the extensive genetic and biochemical work on  $\lambda$  N-boxB, there have been far fewer studies of the corresponding  $\phi$ 21 and P22 interactions. Insight into the  $\phi$ 21 N-boxB complex is limited to biochemical characterization (1) and a nuclear magnetic resonance (NMR) model (9). Previous P22 studies included extensive mutagenesis of P22 boxBs (11), examination of a limited number of P22- $\lambda$  hybrid N proteins (18), and an NMR model of the P22 N peptide bound to P22 boxB<sub>left</sub> (5). Despite the insight provided by the NMR model, the roles of most P22 N residues in recognizing P22 boxBs and the basis of discrimination against  $\lambda$  and  $\phi$ 21 boxBs remain unclear.

The RNA-binding domains of P22,  $\lambda$ , and  $\phi$ 21 N proteins bind as  $\alpha$  helices in the major grooves of the cognate boxBs, making extensive contact with the boxB loop and 5' backbone (9). Interestingly, few base-specific contacts are made between the N proteins and boxB RNAs. P22,  $\lambda$ , and  $\phi$ 21 boxBs adopt similar, yet distinct loop conformations. P22 adopts a 3-out GNRA-like pentaloop (5),  $\lambda$  adopts a 4-out GNRA-like pentaloop (30, 37), and the apical four nucleotides of the  $\phi$ 21 loop adopt a non-GNRA U-turn (9). All of them are structurally similar in having all but one loop base stack on the 3' half. The results of mutagenesis and biochemical studies and NMR data support a model in which the N peptides recognize specific conformations of boxB, with little direct recognition of sequence. The antitermination levels reflect N-boxB affinity, but there are important differences between the host factor dependence of P22,  $\lambda$ , and  $\phi$ 21 N-nut on antitermination (20).  $\lambda$  N antitermination appears to require that  $\lambda$  N Trp18 stacks on the loop of boxB in order to properly bind *Escherichia coli* NusA to achieve full antitermination (41).

The NMR-derived structural model of the P22 N-boxB<sub>left</sub>

\* Corresponding author. Mailing address: Department of Biology, American University of Beirut, P.O. Box 11-0236, Riad El Solh 1107 2020, Beirut, Lebanon. Phone: 961 3 791313, ext. 3887. Fax: 961 1 744 461. E-mail: cs10@aub.edu.lb.

<sup>†</sup> Published ahead of print on 26 September 2008.

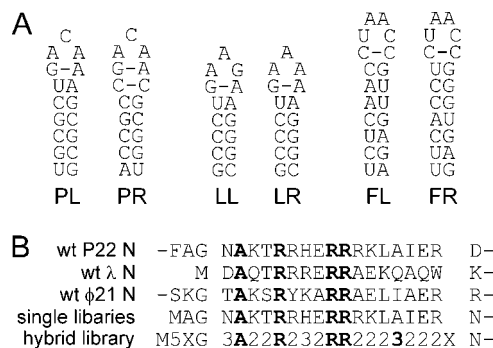


FIG. 1. Comparison of boxB RNAs and N RNA-binding domains of lambdoid phages. (A) The secondary structures of boxBs found in bacteriophages P22,  $\lambda$ , and  $\phi$ 21 from the beginning of the boxB stem. PL, P22 boxB<sub>left</sub>; PR, P22 boxB<sub>right</sub>; LL,  $\lambda$  boxB<sub>left</sub>; LR,  $\lambda$  boxB<sub>right</sub>; FL,  $\phi$ 21 boxB<sub>left</sub>; FR,  $\phi$ 21 boxB<sub>right</sub>. Noncanonical base pairs are indicated by a connecting dash. (B) Alignment of wild-type N proteins and libraries used in this study, with the RNA-binding domains flanked by spaces. Residues indicated by bold type are conserved in the three phage proteins and were not mutated in libraries. Single libraries were made using randomized codons at each nonconserved position in the RNA-binding domain of P22 N. The hybrid library is described by numbers indicating the number of possible amino acids at each position, and in most cases the sets of amino acids include only the corresponding residues from P22,  $\lambda$ , and  $\phi$ 21 at each position; the exception is the residue indicated by the number 3 in bold type, at which an arginine substitutes for the glutamine of  $\lambda$  N. X indicates many amino acids (see Materials and Methods for details).

complex (5) provides limited guidance for understanding the sequence requirements of P22 N protein and the origin of type specificity (Fig. 2A, B, and C). Many contacts are made between the peptide and RNA, but there are few base-specific contacts between amino acid side chains and the RNA (Fig. 2A). Of the residues conserved in P22,  $\lambda$ , and  $\phi$ 21, Ala15 makes hydrophobic contact with boxB bases, and arginines contact backbone phosphates (Fig. 2D). P22 NMR data support a certain role for only one nonconserved P22 N residue, Arg19, and the results of substitution of  $\lambda$  N residues in P22 N (18) suggest that Asn14, Lys16, Arg24, Ala27, Ile28, and Arg30 are important to P22 N function or specificity, yet the structural roles of these residues are uncertain (Fig. 2E). The nonconserved residues likely have important roles in providing both affinity and specificity, including discrimination against noncognate boxBs.

Arginine-rich motifs occur in important regulatory systems and have become important models of RNA-protein recognition (16). The structural diversity of these motifs raises questions about how recognition strategies evolve. The similarity of sequences, structures, and functions is indisputable evidence that the P22,  $\lambda$ , and  $\phi$ 21 N-boxBs are related, and understanding what evolutionary paths connect them may provide general insight into how new RNA-protein recognition strategies evolve. Neutral theorists contend that there are enough mutants with neutral fitness to create incremental paths to new phenotypes without any intermediate loss of function (28, 34). Intriguingly, workers have found both RNAs and peptides that adopt different conformations in different contexts (11, 33, 38). These chameleon sequences could allow smooth evolutionary transitions between distinct recognition strategies.

In order to determine functional roles of nonconserved

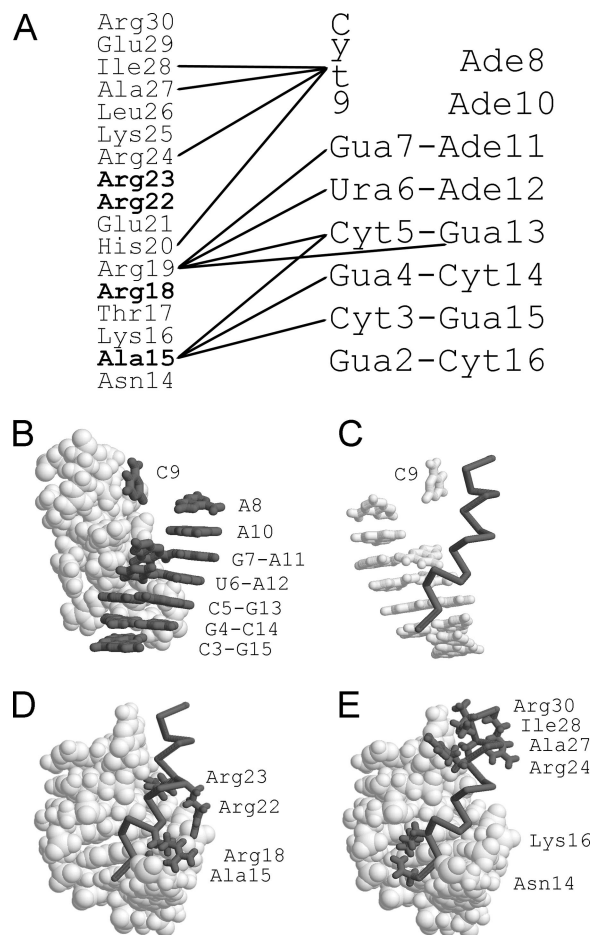


FIG. 2. Observed contacts and NMR model of P22 N peptide bound to P22 boxB<sub>left</sub>. (A) Schematic diagram of the secondary structure of P22 boxB<sub>left</sub> when it is bound to P22 N peptide and the sequence of the P22 N peptide. The lines indicate observed close contacts between amino acid side chains and RNA bases (5). boxB numbering starts from the 5' nucleotide at the base of the stem, and N amino acid numbering starts from the amino-terminal methionine of P22 N. Conserved amino acids are indicated by bold type. Cytosine 9 is extruded from the loop, whereas all other bases are stacked. (B) View of the NMR model from the minor groove, in an orientation similar to that of the schematic diagram in A, with the peptide atoms indicated by light gray spheres and the RNA bases indicated by dark gray sticks. The RNA backbone is not shown for clarity. (C) View of the NMR model from the major groove, where the peptide backbone is indicated by dark gray sticks and the RNA bases are indicated by light gray sticks. Peptide side chains and the RNA backbone are not shown. (D) View from the major groove, with RNA indicated by light gray spheres and the peptide indicated by dark gray sticks. Only the peptide backbone and conserved side chains are shown. (E) Same as panel D, but only nonconserved peptide side chains previously implicated in P22 N function are shown (1, 18).

amino acids of the P22 N RNA-binding domain, we screened 13 single-codon, randomized libraries of P22 N for activity using a plasmid-based  $\beta$ -galactosidase reporter system that reconstitutes antitermination in *E. coli* (19). A library of hybrid N RNA-binding domains of P22,  $\lambda$ , and  $\phi$ 21 was also screened to find sequences that distinguish between boxB<sub>left</sub> and boxB<sub>right</sub>. The relationship between boxB<sub>left</sub> and boxB<sub>right</sub> antitermination and plaque size resulting from the complementation of

clear N<sup>-</sup> virus was also examined. We found that a single R30W substitution and two hybrids are able to complement both P22 and λ N<sup>-</sup> viruses, suggesting that there are multiple evolutionary paths between these similar, yet distinct protein-RNA recognition strategies.

## MATERIALS AND METHODS

**General.** Restriction enzymes and T4 DNA ligase were obtained from Roche (Germany). Bacterial medium components were obtained from Oxoid (United Kingdom). Fine chemicals were obtained from Amersham (United Kingdom), Sigma (United States), and Amresco (United States). Laboratory chemicals were obtained from Acros (Belgium). Disposable plasticware was obtained from Sarstedt (Germany).

**Bacterial strains, plasmids, and bacteriophages.** *Escherichia coli* supporting antitermination, N567 (15), pBR-ptac-N\*λ (19), and lytic N<sup>-</sup> λ phages with immunity regions of λ, P22, and φ21 (phage λimm<sup>22</sup>24am<sup>cl</sup>, phage λ Clear Nam7am53 [18], and phage λimm<sup>21</sup>Nam Clear) were obtained from Naomi Franklin (University of Utah). DH5α cells and control plasmids for the human immunodeficiency virus type 1 (HIV-1) Rev-RRE interaction, pBRN-HIVRev (23) and pAC-HIV-RRE (23), were obtained from Kazuo Harada (Tokyo Gakugei University). P22 boxB reporter plasmids are replacements of λ boxB in the λ nut<sub>left</sub> site by P22 boxB<sub>left</sub> or boxB<sub>right</sub> (11).

**Construction of N fusions and libraries.** All mutant and hybrid RNA-binding domains were expressed such that they replaced the RNA-binding domain of λ N (residues 1 to 19), which are in an NcoI-BsmI cassette. Standard molecular biology procedures were used, and all clones were tested for function and were sequenced using the synthetic insert for confirmation of identity.

To construct the P22 N supplier plasmid pBRNP22N12-30, referred to below as wt P22 N, a double-stranded DNA with NcoI and BsmI sites was designed to contain an initiation codon followed by P22 N residues 12 to 30. It was made by annealing and extending two mutually priming synthetic oligonucleotides, P22N12-30F (5'-GCG CCC ATG GCA GGC AAT GCT AAA ACT CGT CGC CAC GAA CGT CGC-3') and P22N12-30R (5'-GGG ATT TGC ATT CCG CTC AAT CGC CAG TTT ACG GCG ACG TTC GTG GCG ACG-3'). Following primer extension with *Thermus aquaticus* DNA polymerase I, the product was digested with NcoI and BsmI and cloned into the pBR-ptac-N\*λ backbone so that the expressed N protein was a fusion between P22 N residues 12 to 30 (MAGNAKTRRHERRRKLAIER) and λ N residues 19 to 107.

pBRNphi21N1-30, referred to below as wt φ21 N, was constructed by using a similar strategy and synthetic oligonucleotides Phi21N1-30F (5'-GAGCCCATG GTAACCATGTCTGGAAAGAATCCAAAGGTACGGCAAAAAGCCGC TACAAAGCTCGC-3') and Phi21N1-30R (5'-GACTGCTGATTCAGCGCT CGCTCGCAATAAGTTCTGCTCTGCGAGCTTTGTAGCGGCTTTTTG C-3'). The expressed N protein was a fusion between φ21 N residues 1 to 30 (MVTIVWKESKGTAKSRYKARRAELIAERRS) and λ N residues 19 to 107.

P22N12-30 libraries were cloned from synthetic double-stranded NcoI-BsmI fragments like the wt P22 N libraries (Fig. 2B). The inserts were created using four similar strategies, all of which created the same sequence as wt P22 N, with one codon completely randomized. For libraries Asn14X, Lys16X, and Thr17X single-codon randomized oligonucleotides based on CF1 (5'-GCG CCC ATG GCA GGC AAT GCT AAA ACT CGT CGC CAC GAA CGT CGC-3') with unmutated oligonucleotide CR1 (5'-GGG ATT TGC ATT CCG CTC AAT CGC CAG TTT ACG GCG ACG TTC GTG GCG ACG-3') and mutual priming were used. For libraries Arg19X, His20X, and Glu21X single-codon randomized oligonucleotides based on CF2 (5'-GCG CCC ATG GCA GGC AAT GCT AAA ACT CGT CGC CAC GAA CGT CGC-3') with unmutated oligonucleotide CR2 (5'-GGG ATT TGC ATT CCG CTC AAT CGC CAG TTT ACG GCG ACG-3') were used. For libraries Arg24X, Lys25X, Leu26X, Ala27X, Ile28X, and Glu29X unmutated oligonucleotide CF1 and single-codon randomized oligonucleotides based on CR1 were used, and library Arg30X was made by extending primer R30XR (5'-C ATT NNN CTC AAT CGC CAG TTT-3') on single-codon randomized oligonucleotide R30XF (5'-GCG CCC ATG GCA GGC AAT GCT AAA ACT CGT CGC CAC GAA CGT CGC-3'). Most of the resulting double-stranded DNAs were digested with NcoI and BsmI; the only exception was the R30X DNA, in which the BsmI overhang was preformed.

The hybrid library was constructed using a combinatorial strategy involving codon-based mutagenesis, degenerate codons, and mixing of solid support resins during oligonucleotide synthesis (Fig. 1B). The library was designed to include every possible hybrid of P22, λ, and φ21 RNA-binding domains fused to the λ N

activation domain. Because no two degenerate codons could encode only Gln, Ala, and Ile and because λ NQ15R (which aligned with P22 N Ala27) has been reported to function (41), an arginine codon was substituted for λ N Gln1. The hybrid region was flanked by degenerate codons. The sequence of the resulting insert, represented by the coding strand, was GGC CCC ATG GNN NNN GGT (RAT/AMT) GCT MAA WCT CGC (CGT/TAT) (CRT/ARA) GMA CGT CGA (GCC/CGC) RAA (AAA/CTG) (AKA/GCC) (GCA/ATT) SAA (TGG/CGT) NNG AAT GCA GCA AAT CCC-3', and the resulting expressed N proteins were a 22-residue library fused to λ N19-107 with the sequence M(A/D/E/G/V)(X)G(D/N/T)A(K/Q)(S/T)R(R/Y)(H/K/R)(A/E)RR(A/R)(E/K)(E/L)(A/I/R)(A/I)(E/O)(R/W)(A/E/G/K/L/M/P/Q/R/S/T/V/W), where degenerate positions are indicated by parentheses and X indicates all possible amino acids and stop codons. The DNA insert encoding this library was created by NcoI and BsmI digestion of the extension product of two mutually priming, degenerate oligonucleotides. Each degenerate oligonucleotide (HybridF and HybridR) was synthesized using a combinatorial strategy with two concurrent synthesis programs and mixing solid support before and after each divergent codon. Standard synthesizer codes were used for degenerate nucleotides (N = A + C + G + T, R = A + G, Y = C + T, K = G + T, M = A + C, S = G + C, and W = A + T). HybridF was made using HybridF1 and HybridF2 programs (HybridF1, 5'-GGC CCC ATG GNN NNN GGT RAT GCT MAA WCT CGC CGT CRT GMA CGT CGA-3'; HybridF2, 5'-GGC CCC ATG GNN NNN GGT AMT GCT MAA WCT CGC TAT ARA GMA CGT CGA-3'), and HybridR was made using HybridR1 and HybridR2 programs (HybridR1, 5'-GGG ATT TGC TGC ATT CNN CCA TTS TGC TMT TTT TTY GGC TCG ACG TKC AYG ACG GCG A-3'; HybridR2, 5'-GGG ATT TGC TGC ATT CNN ACG TTS AAT GGC CAG TTY GCG TCG ACG TKC TYT ATA GCG A-3'). P22 N R19K and P22 N H20Y inserts were directly constructed by annealing coding and noncoding oligonucleotides creating preformed NcoI and BsmI overhangs.

**DNA preparation and sequencing.** All N-expressing constructs were sequenced with PBRNR2 (5'-GGCTTGCTGTACCATGTG-3') using a BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, United States) under standard conditions. The labeled products were subjected to electrophoresis with an ABI 310 genetic analyzer sequencing system (Applied Biosystems, United States). The entire insert was confirmed using Chromas Lite software from Technelysium (Australia).

**Single-position library screening and X-Gal plate assays.** Competent N567 host cells carrying boxB reporter plasmids (P22 boxB<sub>left</sub>, P22 boxB<sub>right</sub>, or HIV RRE) were transformed with the library or clones of interest and control plasmids, including the wild-type λ N supplier (pBR-ptac-N\*λ), the P22 N fusion supplier (wt P22 N), the φ21 N supplier (wt φ21 N), and the HIV-Rev N fusion supplier (pBRN-HIVRev). Approximately 10 to 100 ng of plasmid per 100 μl of competent cells was transformed by heat shock and plated on tryptone medium plates containing 50 μg/ml ampicillin, 15 μg/ml chloramphenicol, and 80 μg/ml X-Gal (5-bromo-4-chloro-3-indolyl β-D-galactoside), which is a chromogenic substrate of the β-galactosidase reporter protein. All X-Gal plates included 0.05 mM IPTG (isopropyl β-D-thiogalactoside) to induce the *tac* promoters expressing N protein and the reporter transcript. The plates were viewed and scored after 1 day of incubation at 34°C and after a second day of incubation at 24°C.

**Virus complementation assays.** Plaque assays were performed by standard procedures (36), using clear strains of N-deficient λ phage, in which the immunity region was either from phage λ or a replacement from phage P22 or φ21. Overnight cultures of N567 hosting N supplier plasmids were grown in LB with ampicillin at 37°C. Cultures were centrifuged and resuspended in 10 mM MgSO<sub>4</sub> to obtain an optical density at 600 nm of 2.0. Cells were incubated for approximately 30 min with approximately 100 PFU of virus by mixing 0.05 ml cells and 0.05 ml virus in SM (100 mM NaCl, 8 mM MgSO<sub>4</sub>, 50 mM Tris-HCl [pH 7.5], 0.1 g/liter gelatin) along with 1.2 ml tryptone top agar (48°C) and immediately plated on 5-cm tryptone plates. No antibiotic was used in the bottom or top agar, but where indicated below, N expression was induced with 0.3 mM IPTG in the bottom agar. The plates were allowed to set before overnight incubation at 37°C. Plaque diameters were assessed the next morning by superimposition on positive controls, and the results were expressed as percentages of the cognate diameters.

**ONPG solution antitermination assay.** For each N-boxB interaction, representative colonies were picked from X-Gal plates for use in solution assays (32). At least three independent colonies were used for each interaction. For measurement of N-mediated antitermination, cultures were grown overnight at 30°C with aeration in tryptone with 50 μg/ml ampicillin, 15 μg/ml chloramphenicol, and 0.05 mM IPTG. The cells were then permeabilized, the β-galactosidase activity was assayed using *o*-nitrophenol-β-D-galactoside (ONPG), and the β-galactosidase activity was calculated by using the method of Miller (32). The activities were normalized using P22N14-30 for boxB<sub>left</sub> and boxB<sub>right</sub> and RevN for RRE.

TABLE 1. Library antitermination frequencies

P22 N library <sup>a</sup>	X-Gal score (%) <sup>b</sup>	
	P22 boxB <sub>left</sub>	P22 boxB <sub>right</sub>
Asn14X <sup>c</sup>	50	50
Lys16X	13	12
Thr17X	60	60
Arg19X	6	5
His20X <sup>d</sup>	60	40
Glu21X	60	50
Arg24X	40	30
Lys25X	60	50
Leu26X	40	50
Ala27X	40	40
Ile28X	50	50
Glu29X	50	40
Arg30X	70	60
Mixed <sup>e</sup>	40	50

<sup>a</sup> Libraries are indicated by three-letter codes and positions in P22 N, and X indicates complete randomization (see Fig. 2).

<sup>b</sup> Library plasmids were transformed into P22 boxB<sub>left</sub> and P22 boxB<sub>right</sub> reporter cells and plated on tryptone agar with 50 mM IPTG and 80 μg/ml X-Gal. The approximate percentages of colonies with activity greater than the background activity are indicated. With the exception of the Asn14X and His20X libraries, the activities for positive libraries ranged from low to wild-type levels.

<sup>c</sup> The distribution of activities in the Asn14 library tended toward remarkably lower activity.

<sup>d</sup> The distribution of activities in the His20 library included an estimated 5% of the members with activity well above the activity of unmutated P22 N.

<sup>e</sup> The libraries were mixed using equal masses and screened as a population.

**Visualization of the structure.** Protein Explorer (31) was used to view the solution state NMR models of the structure of P22 N peptide-P22 boxB<sub>left</sub> (Protein Data Bank accession number 1A4T) (5), λ N peptide-λ boxB<sub>right</sub> (Protein Data Bank accession number 1QFQ) (37), and φ21 N peptide-φ21 boxB<sub>right</sub> (Protein Data Bank accession number 1NYB) (9).

## RESULTS

Available genetic (11, 15), biochemical (2), and NMR (5, 9, 30, 37) evidence suggests that the P22, λ, and φ21 N proteins recognize boxB conformation, with little direct recognition of RNA sequence. The roles of only a few P22 N residues in boxB recognition and specificity are well understood based on the available P22 N-boxB NMR data (5) and mutagenesis data (18). We first constructed plasmids expressing P22 N and φ21 N RNA-binding domains fused to the λ N activation domain (wt P22 N and wt φ21 N, respectively). These plasmids and the λ N plasmid specifically complement viruses lacking the cognate N protein. The roles of the conserved residues in the three viral N-boxB structures have been found to be similar (9), and these residues cannot play a role in type specificity. To assess the mutability of P22 N residues and to reveal their structural roles, we constructed 13 single-substitution libraries for each nonconserved residue in the RNA-binding domain of P22 N (Fig. 1B). Each library contained an unbiased collection of 64 codons at the targeted position.

**The mutability of P22 N residues ranges from invariant to unrestricted.** To estimate the mutability of each residue, plasmid libraries were transformed into P22 boxB<sub>left</sub> and boxB<sub>right</sub> reporter cells, and their abilities to support antitermination were visualized by inspection of colonies grown on solid media containing the chromogenic substrate X-Gal, which was transformed into a blue pigment by the β-galactosidase reporter gene (Table 1). The proportion and intensity of blue colonies

provided an indication of the tolerance of each residue to mutation. We estimated from screening wt P22 N and wt φ21 N clones (constructed with the same strategy used for the libraries) that oligonucleotide errors created nonfunctional clones in 20 to 40% of the transformants; therefore, completely mutable positions should produce 60 to 80% blue colonies. These data suggest that the mutability of P22 N RNA-binding domain residues ranges from invariant to unrestricted. Although large differences in the proportion of active clones in the libraries were apparent, no bias within any library was obvious for boxB<sub>left</sub> and boxB<sub>right</sub> reporters, with the exception of His20X. Notably, we observed that when inactive colonies were disregarded, the distribution of activity varied between libraries; for the Asn14X library few colonies were as active as wild-type colonies, and for the His20X library about 5% of the total colonies had activity higher than the wild-type activity. For two basic residue libraries, Lys16X and Arg19X, there were very few active clones. Arg19X colonies appeared to be either white or as blue as wild-type P22 N colonies, whereas Lys16X colonies exhibited a range of activity. Several positions appeared to be very mutable; in particular, the Thr17, Glu21, Lys25, and Arg30 residues in most library members appeared to be active. The mixed-library results suggest that most of the 247 possible single-substitution mutants of P22 N expressed by these libraries of the P22 N RNA-binding domain were functional.

**Mutations complementing P22 N<sup>-</sup> virus support specific structural roles.** From each library, colonies with activities that were close to and greater than wild-type activity were selected using boxB<sub>left</sub> and boxB<sub>right</sub> reporters. N supplier plasmids were separated from reporter plasmids, retransformed into fresh bacteria, and tested to determine their abilities to complement an N-deficient λ strain dependent on P22 N-nut antitermination. Clones able to complement P22 N<sup>-</sup> virus were sequenced, and plaque diameters were compared to those of wt P22 N (Fig. 3). Most libraries yielded active mutants; the only exception was Arg19X, for which only Arg codons were found. Wild-type residues were not isolated at Lys16, Thr17, and E29, presumably because wild-type codons are less active than mutant codons or are relatively rare among active clones.

Consistent with their orientation away from boxB, we found that Thr17, Glu21, Lys25, Leu26, Glu29, and Arg30 were replaced by dissimilar residues, including residues with different charges, without a major loss of function (Fig. 3). In particular, Thr17 accepted a diverse assortment of residues. The loss of some function suggests that this residue has only an indirect or minor effect on N-boxB binding. Previous mutagenesis data (18) and the recovery of only alanine and serine residues from library Ala27X supports the NMR model, where the residue occupies a small recess against Cyt9.

Only arginine codons were found in active clones from library Arg19X. The very low proportion and wild-type activity of positive colonies seen with P22 boxB reporters suggested that this position is immutable without a complete loss of activity. The immutability of Arg19 was expected in light of NMR data (5), which showed that this residue has an unambiguous and certainly critical role, binding to Gua7 of the noncanonical G:A pair and the phosphate backbone. Because we sequenced only five clones and could not rule out the possibility that lysine might function at this position, we con-

30	R	4/8	RK <sub>F</sub> W
29	E	0/3	SL
28	I	3/10	ILMTV <sub>C</sub> HS
27	A	3/6	AS
26	L	1/3	LAK
25	K	2/7	KQRS
24	R	3/5	R <sub>L</sub>
23	R		
22	R		
21	E	1/3	ELI
20	H	1/22	HFW
19	R	5/5	R
18	R		
17	T	0/7	SWYAR <sub>S</sub>
16	K	0/4	R
15	A		
14	N	1/5	NTS <sub>C</sub>

FIG. 3. Summary of active single amino acid substitutions in P22 N. The wild-type sequence of the RNA-binding domain of P22 N is indicated on the left by using position numbers and single-letter codes. The number of wild-type residues observed in the total number sequenced for each library (wild type/total) are indicated in the middle. Conserved positions (Ala15, Arg18, Arg22, and Arg23) were not examined using mutagenesis. Single mutations supporting viral replication are indicated on the right, and the heights represent plaque diameters of 100%, 75%, 50%, and 25% and less. When wild-type sequences were recovered, they were grouped with the library of origin.

structed and tested the R19K mutant and found that it was inactive.

Surprisingly, we recovered only arginine codons from library Lys16X, presumably by inadvertently selecting only the most active colonies. The active arginine mutant and the low proportion of positive clones in this library suggest that a basic residue is required at this position. Without exhaustive screening, we cannot rule out the possibility that constructs with nonbasic substitutions function, although we predict that they have low activity. Previous mutagenesis data indicated that constructs with the equivalent  $\lambda$  glutamine residue at this position can function weakly (18).

The aliphatic chain of Arg24 appears in the NMR model as packing against His20 and the ribose of the extruded loop nucleotide Cyt9, possibly stabilizing the essential 3-out GNRA-like conformation of boxB (5). Franklin (18) has reported that a P22- $\lambda$  N hybrid with alanine at this position has weak activity. The guanidinium group of Arg24 is poorly defined and has no clear role in the NMR model. Our finding that the R24L mutant functions, although weakly, supports the hypothesis that the hydrophobic interaction between the aliphatic chain of Arg24 and His20 and the ribose of Cyt9 is important, but it does not clarify the role of the guanidinium group. No phosphate appears to be within reach, and there is no data supporting an interaction with two nearby planar groups, Cyt9 and His20. We imagine that the guanidinium positive charge may increase nonspecific affinity, and we expect that a construct with a lysine substitution would be functional.

Asn14 mutants exhibit reduced activity. We found only small hydrophilic substitutions, suggesting that Asn14 has a specific role. Ile28 is defined well by NMR data, packing against the

face of Cyt9, and mutagenesis data suggest that it is functionally important (18). We found that this residue can be mutated to similar residues capable of packing against Cyt9. P22 boxB can tolerate mutation of the extruded base at boxB position 9 with little loss of activity (11), and the role of Ile28 is most likely to recognize the 3-out GNRA-like loop conformation. Presumably, any residue with a hydrophobic face can play the same role.

In the NMR model His20 is described as part of a hydrophobic surface facing Cyt9 and possibly interacting with a backbone phosphate via a weak hydrogen bond. A previous mutagenesis study revealed no functional role for His20 (18). In contrast, we found only the aromatic residues His, Phe, and Trp in our screening analysis. Close examination of the NMR model reveals that the imidazole ring of His20 makes only partial contact with the aliphatic chain of Arg24 and the ribose of Cyt9 and that the majority of its aromatic ring faces an internal void (5). The NMR data provide no support for stacking against either of the two closest planar partners, Cyt9 and Arg24. When the colonies with the highest activity were chosen, only H20F and H20W mutants were found. The H20F substitution suggests that contact with the phosphate backbone by His20 is unimportant. The presence of only aromatic residues suggests strongly that His20 stacks against a planar partner. We made the H20Y construct and found that it produced half-size P22 N<sup>-</sup> plaques.

**Few single mutants distinguish between boxB<sub>left</sub> and boxB<sub>right</sub>**  
The unresolved roles of Asn14, His20, and Arg24 prompted us to consider whether these residues are more important in binding boxB<sub>right</sub> than in binding boxB<sub>left</sub>. boxB<sub>left</sub> and boxB<sub>right</sub> have identical loops but different base pairs at each end of their stems (Fig. 1A). Residues contacting these bases or subtly stabilizing a specific bound boxB conformation could create bias. We noted that the largest bias between boxB<sub>left</sub> and boxB<sub>right</sub> was observed when the His20X library was screened, although sequencing colonies with high activity revealed the same collection of aromatic mutants for both screens. We selected mutants from each library for quantitative assays of the function on boxB<sub>left</sub> and boxB<sub>right</sub> (Table 2).

We found that very few substitutions cause as much as a twofold bias between the two reporters. The strongest bias was found for Asn14C, which is within reach to interact with boxB<sub>left</sub> and boxB<sub>right</sub> stem differences (Fig. 1A and 2E). We found that H20W had activity higher than that of the wild-type sequence and that it was biased in favor of boxB<sub>right</sub>. These data suggested that an aromatic residue at this position has an important and definable role, but it is doubtful that this role can be clarified without biophysical studies.

**A library of basic domain hybrids reveals biased N proteins.**  
Previously, we observed that  $\lambda$  N activates P22 boxB<sub>left</sub> reporters strongly and boxB<sub>right</sub> reporters weakly, demonstrating that bias is possible (11). Reasoning that a library of hybrid RNA-binding domains might contain more constructs with bias, we designed a library of N proteins in which almost all possible hybrids of P22,  $\lambda$ , and  $\phi$ 21 RNA-binding domains were fused to  $\lambda$  activation domain (Fig. 1B). So that the hybrid library could be synthesized with the available two-column oligonucleotide synthesizer, an arginine codon was substituted for  $\lambda$  N Gln15 (which aligns with P22 N Ala27).  $\lambda$  NQ15R has been reported to function in antitermination at wild-type levels (41).

TABLE 2. Antitermination activities of single substitutions

P22N fusion <sup>a</sup>	% of cognate function <sup>b</sup>		
	P22 boxB <sub>left</sub>	P22 boxB <sub>right</sub>	HIV RRE <sup>c</sup>
P22 N	100 ± 9	100 ± 8	10 ± 3
λ N <sup>d</sup>	2,100 ± 200	7 ± 1.1	10 ± 3
φ21 N	12.3 ± 0.6	8.7 ± 0.8	13 ± 2
RevN	5.5 ± 0.9	4.9 ± 0.4	100 ± 13
N14C	77 ± 2	18 ± 1.3	4.0 ± 0.2
N14S	57 ± 5	39.4 ± 0.8	5.0 ± 0.11
N14T	152 ± 5	70 ± 3	6.7 ± 0.8
K16R	160 ± 14	170 ± 70	6.2 ± 0.7
T17G	90 ± 11	80 ± 12	5.0 ± 0.7
T17W	177 ± 6	110 ± 30	7.2 ± 0.8
T17Y	74 ± 3	73 ± 1.6	4.9 ± 0.4
H20F	110 ± 18	200 ± 50	3.7 ± 0.10
H20W	200 ± 20	400 ± 30	3.5 ± 0.15
E21L	240 ± 16	240 ± 17	6.8 ± 0.4
R24L	160 ± 16	80 ± 20	6.5 ± 0.5
K25Q	84 ± 7	60 ± 20	6.7 ± 0.4
K25S	50 ± 15	46 ± 9	7 ± 1.6
L26A	224 ± 8	300 ± 30	7.6 ± 0.4
A27S	70 ± 17	70 ± 15	5.8 ± 0.4
I28C	39 ± 2	28 ± 6	6.0 ± 0.4
I28H	34 ± 6	28 ± 3	6.4 ± 1.2
I28L	161 ± 5	177 ± 8	3.7 ± 0.18
I28M	150 ± 60	190 ± 40	6.7 ± 0.4
E29L	106 ± 7	83 ± 7	6.8 ± 0.6
E29S	170 ± 20	90 ± 17	7.4 ± 0.5
R30F	27 ± 4	20 ± 1.7	11 ± 1.5
R30K	100 ± 7	73 ± 4	5.8 ± 0.8

<sup>a</sup> P22 N mutants proteins are indicated by single-letter codes for wild-type and mutant residues separated by the position. RevN is a fusion of the HIV-1 Rev RNA-binding domain with the λ N activation domain and was used as a heterologous control; it activates the HIV-1 RRE boxB reporter.

<sup>b</sup> Library plasmids were transformed into P22 boxB<sub>left</sub>, P22 boxB<sub>right</sub>, and HIV RRE reporter cells, and cultures were assayed to determine reporter gene β-galactosidase activity with ONPG after overnight growth at 30°C in tryptone medium and induction with 50 μM IPTG.

<sup>c</sup> HIV-1 RRE, replacing λ boxB<sub>left</sub>, was used as a heterologous control for nonspecific activation.

<sup>d</sup> Wild-type λ N-boxB interactions in this reporter system resulted in approximately 100-fold-higher levels of antitermination than P22 N-boxB, and λ N activated P22 boxB<sub>left</sub> about 20% as well as λ boxBs but did not activate P22 boxB<sub>right</sub> above the background level.

Less than 600 of approximately 160,000 hybrid library transformants plated exhibited activity on boxB<sub>left</sub> or boxB<sub>right</sub> reporters. Colonies active on boxB<sub>left</sub> or boxB<sub>right</sub> were separately pooled and rescreened to examine differential activity on boxB reporters. Approximately one-half of the pool of transformants originally selected on boxB<sub>left</sub> showed bias. In sharp contrast, less than 2% of the pool selected on boxB<sub>right</sub> showed bias. Repeated attempts to find clones with bias toward boxB<sub>right</sub> identified a few weakly biased sequences. Selected clones were sequenced and characterized to determine their antitermination activities and abilities to complement P22 N<sup>-</sup> virus (Table 3). Some clones were able to complement P22 N<sup>-</sup> virus only when N expression was induced. Considering the 14 unique sequences, bias for boxB<sub>left</sub> was more common than bias for boxB<sub>right</sub>, for which only two sequences had significant bias. The sequences of hybrids able to support P22 replication provided additional support for some findings for the single-substitution libraries. Residues Asn14, Lys16, Arg19, and Ala27 were found only as P22 variants. Interestingly, hybrids with H20R and H20K mutations were active, although none of them were as active as the wild type. Arginine is not an aro-

matic amino acid, but its conjugated planar surface can engage in similar stacking interactions, and the aliphatic portion of lysine can pack against planar surfaces. R24A mutants are active, although not fully. Unexpectedly, an R24P mutant was found, although it was only weakly active. Analysis of these sequences did not yield much insight into the origin of bias between boxB<sub>left</sub> binding and boxB<sub>right</sub> binding, possibly because of compensatory effects of multiple substitutions.

**Some sequences exhibit relaxed specificity.** Comparison of mutations tolerated in P22 N (Fig. 3 and Table 3) to mutations tolerated in λ N (18, 19) suggested that there is at least one residue that is compatible with both viruses at each position, although sometimes the function is reduced. Indeed, Franklin (18) described a few P22-λ N hybrids able to complement both λ and P22 viruses lacking N, albeit weakly. Unfortunately, no mutagenesis of the φ21 N-boxB interaction has been reported. Using complementation of P22 N<sup>-</sup>, λ N<sup>-</sup>, and φ21 N<sup>-</sup> viruses in plaque assays, we tested N proteins shown in Fig. 3 and Table 3, as well as NF3862 (18), a previously described, relaxed-specificity hybrid (Table 4). Only the results obtained with control proteins and with the sequences producing plaques with P22 N<sup>-</sup> and λ N<sup>-</sup> are described in Table 4. When induced, all N fusions, including RevN, complemented φ21 N<sup>-</sup>; thus, we do not believe that complementation of φ21 N<sup>-</sup> by induced N protein is related to N-boxB specificity. Increased N expression from plasmids by induction of the pTac promoter with IPTG is known to strongly reduce type specificity (22).

One single-substitution mutant of P22 N, R30W, was able to complement both P22 N<sup>-</sup> and λ N<sup>-</sup> viruses without induction. Arg30 has no clear function in the P22 N-boxB NMR model and is relatively mutable (Fig. 3). The λ Trp18 equivalent plays a critical role in the λ N-boxB interaction, stacking onto the boxB loop, and this interaction has been implicated in recruiting host factor NusA in a specific binding mode to allow full λ antitermination. Concordantly, two of our hybrids with tryptophan at this position exhibited relaxed specificity, but this occurred only under induction conditions. The weak ability of NF3862 to complement P22 N<sup>-</sup> virus is not surprising, because it does not have two residues that have been found to be important for P22 N function, the basic residue Lys16 and the aromatic residue His20. NF3862 had only background antitermination on P22 boxB<sub>right</sub> and boxB<sub>left</sub> reporters (data not shown). Likewise, the poor ability of NF3862 to complement λ N is explained by the absence of a residue equivalent to the critical λ Trp18 residue.

## DISCUSSION

We have shown that the RNA-binding domain of P22 N is highly mutable. The data inform and have minimal conflict with the NMR model (5) and are consistent with the importance of Asn14, Lys16, Arg24, Ala 27, and Ile28, as previously suggested by the findings for a limited series of mutants (1, 18). New findings include the importance of His20 and the gain of activity with K16R, H20W, E21L, L26A, and I28M. The limited mutability of Asn14, His20, and Arg24 suggests that important aspects of the interaction are still not understood. Our results suggest that P22 boxB<sub>left</sub> and boxB<sub>right</sub> impose similar constraints on P22 N. We found that the single substitution R30W and two P22-λ-φ21 hybrids are able to complement

TABLE 3. Antitermination activity and P22 N<sup>-</sup> complementation of hybrids

N fusion <sup>a</sup>	Sequence <sup>b</sup>	% of cognate function		P22 plaque size (%) <sup>c</sup>	
		P22 boxB <sub>left</sub>	P22 boxB <sub>right</sub>	Uninduced conditions	With induction
P22 N	MAGNAKTRRHERRRKLAIER	100 ± 9	100 ± 8	100	100
λ N	MDAQTRRRERRAEKQAOQW	2,100 ± 200	7 ± 1.1	0	0
φ21 N	MVTIVWKESKGTAKSRYKARRAELIAERRS	12.3 ± 0.6	8.7 ± 0.8	0	0
RevN <sup>d</sup>	MATROARRNRRRRWRRAAA	5.5 ± 0.9	4.9 ± 0.4	0	0
II493-1	MDSGNAKSRRRERRAKLAAKWW	109 ± 7	2.8 ± 0.3	0	25
II493-16	MAKGNASRRRARRAEKAIQWV	20 ± 4	3.8 ± 0.3	0	25
II592-5	MAYGNAKSRRHARRRKKAIERS	123 ± 9	330 ± 30	100	100
II592-8	MAVGNAKTRRRERRAKKAIQRL	130 ± 17	60 ± 15	75	100
II592-13	MATGNAKSRRHARRREKAIERM	27 ± 6	53 ± 5	100	100
II592-14	MVRGNAKTRRHERRAKLAAQRQ	17 ± 2	16 ± 3	25	75
II592-15	MVSGNAKTRRRERRAEKAIERW	25 ± 3	26 ± 2	50	75
II592-17	MAPGNAKTRRRARRAEKAIQWL	81 ± 6	25 ± 7	Pin	75
II592-19	MGPNAKSRRRERRRKKAAERV	47 ± 3	22 ± 3	50	75
II592-20	MGPNAKSRRRARRAKKAIQRW	170 ± 30	37 ± 5	50	100
II592-21	MAVGNAKTRRRARRAEKAIQRS	140 ± 11	89 ± 9	25	75
II592-22	MGVNAKSRRKARRAEKAIQRV	110 ± 14	33 ± 4	50	100
II592-23	MAHGNAKTRRRARRAEKAIERG	67 ± 6	50 ± 4	25	100
II601-3	MAKGNASRRHARRPEKAIQRM	60 ± 20	75 ± 6	0	25

<sup>a</sup> N proteins are indicated as described in Table 2, footnote a, or by the laboratory stock numbers of hybrid library clones.

<sup>b</sup> The sequences of fusions until the λ N activation domain are indicated, and conserved residues are indicated by bold type.

<sup>c</sup> The sizes of P22 N<sup>-</sup> plaques under uninduced conditions and after induction of N protein expression with 50 μM IPTG are expressed as percentages of the wild-type size. Pin indicates barely visible.

<sup>d</sup> RevN is a fusion of the RNA-binding domain of HIV-1 Rev to the activation domain of λ N, and it was used as a heterologous control for nonspecific complementation.

both P22 N<sup>-</sup> and λ N<sup>-</sup> viruses, suggesting that evolutionary transitions between these two distinct RNA-protein interactions are facile. No mutants were found to bridge the specificity between P22 and φ21 N.

**P22 N structural roles informed by mutagenesis.** The mutability of Thr17, Glu21, Lys25, Leu26, Glu29, and Arg30, all of which face away from RNA, supports the conclusion that these residues have no specific structural role. Some mutations increase antitermination activity (Table 2), suggesting that they can modulate N-boxB affinity, if only indirectly. Despite the typically interchangeable properties of threonine and serine, Thr17 is considered a nonconserved residue in the N RNA-binding domains. This view is bolstered by the lack of any definable role for this position in any N-boxB complex, in

contrast to the findings for the conserved Ala15, Arg18, Arg22, and Arg23 residues (9). Its high mutability indicates that conservation merely implies conservation of function.

The structural role of Arg19 is clearly defined by NMR data (Fig. 2A). A similar interaction has been reported for the HIV-1 Tat-TAR complex (6), where this residue is involved in the critical protein-RNA recognition interaction. We found that K16R has activity higher than that of the wild-type sequence with boxB reporters, supporting the specific contact suspected based on NMR results but not observed by NMR. Interestingly, examination of the NMR model suggested that the arginine mutant could contact Gua13 and the phosphate similar to Arg19 at Gua7, and this could explain the higher activity of K16R.

TABLE 4. Relaxed-specificity sequences

N fusion <sup>a</sup>	Sequence	Plaque diam (%) <sup>b</sup>					
		P22		λ		φ21	
		Uninduced conditions	With induction	Uninduced conditions	With induction	Uninduced conditions	With induction <sup>c</sup>
wt P22	MAGNAKTRRHERRRKLAIER	100	100	0	0	0	100
wt λ	MDAQTRRRERRAEKQAOQW	0	0	100	100	0	100
wt φ21	MVTIVWKESKGTAKSRYKARRAELIAERRS	0	0	0	0	100	100
RevN	MATROARRNRRRRWRRAAA	0	0	0	0	0	100
NF3862	MDAQTRRRERRRKLAIER	Pin	25	Pin	25	0	100
R30W	MAGNAKTRRHERRRKLAI EW	50	75	25	50	0	100
II493-1	MDSGNAKSRRRERRAKLAAKWW	0	25	0	75	0	100
II493-17	MAPGNAKTRRRARRAEKAIQWL	Pin	75	0	Pin	0	100

<sup>a</sup> N proteins are indicated as described in Tables 2 and 3, footnotes a. Hybrid library clones are indicated by laboratory stock numbers, and NF3862 is a relaxed-specificity hybrid provided by N. Franklin (18).

<sup>b</sup> The sizes of N<sup>-</sup> plaques are expressed as percentages of the wild-type size, and Pin indicates barely visible plaques. Plaques were observed under uninduced conditions and after induction of N protein expression with 50 μM IPTG.

<sup>c</sup> When induced, all N proteins complemented φ21; such relaxed specificity after induction is not thought to be independent of N-boxB interactions (22).

Although the role of Asn14 is not defined by NMR data (5), previous genetic (18) and biochemical (3) data indicated that this residue is preferred to aspartate by both P22 and  $\lambda$  boxBs. Interestingly, N14C exhibits a strong bias toward boxB<sub>left</sub> (Table 2). The NMR model indicates that Asn14 is within reach of Gua2, which is replaced by Cyt2 in boxB<sub>right</sub> (Fig. 1A). This suggests that Asn14 has a discrete role, perhaps interacting with Gua2, and that the two boxBs may impose different constraints on N genetic drift. This study, which examined only the single-substitution mutants able to form plaques, would have missed strongly biased mutants that cannot form plaques. All hybrids, even the boxB<sub>right</sub>-biased mutant, contain Asn14, implying that if this position can cause bias, the available residues from  $\lambda$  and  $\phi$ 21, Asp2 and Thr12, do not cause bias (Table 3).

The role of His20 remains uncertain and intriguing. We found only planar residues at this position, but in one case a hybrid had a lysine residue (Fig. 3 and Table 3). The P22 N H20W mutant had activity that was two- and fourfold higher than the wild-type activity with boxB reporters (Table 2), strongly suggesting that there is a planar partner surface, yet no planar partner surface was apparent. We speculate that biophysical studies of the P22 N-boxB<sub>right</sub> complex would illuminate the role of His20. Indeed, the intrinsic fluorescence of tryptophan may allow the H20W mutant to be useful for biophysical studies of N-boxB interactions, such as the studies done for  $\lambda$  N-boxB (41).

**Bias between boxB<sub>left</sub> and boxB<sub>right</sub>.** The relative lack of bias between boxB<sub>left</sub> and boxB<sub>right</sub> reporter activities observed suggests that P22 N protein binds the two boxBs very similarly. In contrast,  $\lambda$  N had very high activity with our P22 boxB<sub>left</sub> reporter and minimal activity with our boxB<sub>right</sub> reporter (Table 2), most likely because  $\lambda$  N forces a  $\lambda$  boxB conformation on P22 boxB<sub>left</sub> (1). Although there was no sequence pattern evident in hybrids that correlated with either boxB<sub>left</sub> or boxB<sub>right</sub> reporter activities (Table 3), strong activity with boxB<sub>left</sub> could arise from a hybrid N forcing a  $\lambda$ -like, 4-out GRNA-like loop conformation on P22 boxB<sub>left</sub>. Bias would then result from P22 boxB<sub>right</sub> being unable to adopt a  $\lambda$  conformation, as a result of indirect thermodynamic effects of the C:C pair or stem differences, both of which have been implicated in  $\lambda$  N discrimination between P22 boxB<sub>left</sub> and boxB<sub>right</sub> (11). Our recovery of very few hybrids biased toward P22 boxB<sub>right</sub> supports this hypothesis (Table 3).

**Viral replication correlates to unbiased antitermination.** We also examined the relationship between boxB<sub>left</sub> and boxB<sub>right</sub> antitermination activity and plaque size using a clear  $\lambda$  strain regulated by P22 N-nut antitermination. We plotted the plaque diameter of P22 N<sup>-</sup> virus complemented by the mutant N clones described in Tables 2 and 3 as a function each clone's antitermination activity with boxB<sub>left</sub> and boxB<sub>right</sub> reporters (Fig. 4). Although weakly antiterminating N mutants correlated with small or absent plaques, the reporter activity did not accurately predict plaque size. Importantly, all full-size plaques had low bias, suggesting that balance in leftward and rightward antitermination is important for viral fitness. The lack of a strong correlation between antitermination activity and plaque size most likely reflects the very different conditions used in the two assays. The reporter system measured reporter gene accumulation in saturated cultures grown with continuous induction at 30°C. In contrast, plaques were assessed using lawns

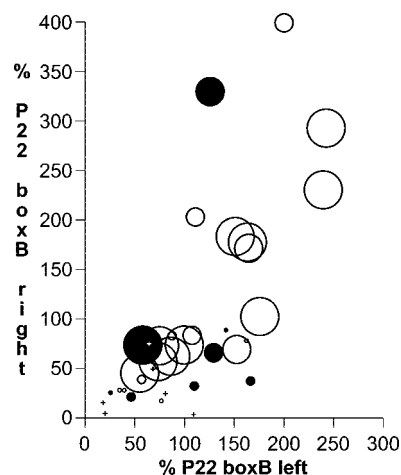


FIG. 4. P22 N<sup>-</sup> plaque size as a function of antitermination activity on P22 boxB<sub>left</sub> and boxB<sub>right</sub>. Plaque sizes and antitermination activities of sequences shown in Tables 2 and 3 are plotted. Single substitutions are indicated by open circles, and hybrids are indicated by filled circles. The activities in boxB<sub>left</sub> and boxB<sub>right</sub> reporter cells are expressed as percentages of the wild-type P22 N activity. The circle diameters are proportional and indicate plaque sizes of 100% (largest), 75%, 50%, 25%, and barely visible. Hybrid sequences unable to produce plaques (from Table 3 only) are indicated by plus signs.

grown without induction at 37°C. Nonetheless, full-size plaques were found only when boxB<sub>left</sub> antitermination and boxB<sub>right</sub> antitermination were nearly balanced, which could have provided incremental selective pressure for boxB<sub>left</sub> and boxB<sub>right</sub> to maintain similar binding affinities.

**Relaxed N-boxB specificity.** The high mutability of many positions in P22 N raises the question of whether P22,  $\lambda$ , and  $\phi$ 21 N specificity is the result of complex effects of multiple mutations or the result of specific intolerance to nontype residues at critical positions. Our finding that there is a moderately active, single-substitution P22 N RNA-binding domain mutant and our finding that there are additional weakly active, relaxed-specificity hybrids suggest that there may be many hybrids with relaxed specificity (Table 4). Indeed, the P22 N R30W mutant shows that type specificity can be relaxed without a severe loss of function for either partner by a single substitution.

We found no mutant capable of bridging P22- $\phi$ 21 N type specificity, although such mutants may exist. No mutagenesis study examining  $\phi$ 21 N has been described.  $\phi$ 21 N Tyr17 appears to contact  $\phi$ 21 boxB in an important interaction similar to the critical interaction of Arg19 of P22 N. Our data and the NMR model indicate that tyrosine should be unable to play the role of Arg19 in P22 N. Nonetheless, it is possible that P22 N Arg19 could play the role of Tyr17 in  $\phi$ 21 N or that compensatory effects of other mutations could create sequences that bridge type specificity.

**Transiting type specificity.** The distinct, yet structurally similar P22,  $\lambda$ , and  $\phi$ 21 N-boxB interactions are doubtless related evolutionarily. What mutational paths through sequence space might connect these three distinct solutions to the problem of binding a protein to an RNA? Relaxed-specificity N or boxB mutants represent possible transitions between discrete recognition strategies. Our data suggest that there may be many



relaxed-specificity hybrids between P22 and  $\lambda$  N. From a biophysical perspective, the induced-fit nature of RNA-protein interactions (35) allows plasticity, where the adopted conformation depends on context. Such conformational plasticity may be particularly common in arginine-rich peptide-RNA interactions (4). Induced fit suggests that there is binding of discrete conformations rather than a smooth continuum of specific recognition strategies. There is evidence that  $\lambda$  boxB samples discrete conformations, including the bound conformation, when it is unbound (27). Indeed, the results of a recent NMR study of HIV TAR RNA support the idea that proteins sometimes merely capture conformations that are being sampled by motional modes of RNAs (42).

Relaxed-specificity mutants are able to participate in multiple strategies and thereby may provide evolutionary routes between discrete modes of interaction without severe loss of fitness. The ability of N proteins to recognize boxBs with few base-specific contacts suggests that the type specificity of N is primarily the result of being unable to bind more than one boxB loop conformation. In contrast, type specificity in boxBs appears to be limited by the thermodynamics of assuming different bound conformations, with only indirect effects of sequence (1, 11). The ability of  $\lambda$  N to force P22 boxB<sub>left</sub> to adopt a  $\lambda$  conformation illustrates one mechanism by which relaxed specificity can be achieved. The reciprocal mechanism of adopting both P22 and  $\lambda$  boxB loop conformations appears to be employed by relaxed-specificity boxBs (11). Relaxed-specificity sequences may be able to acquire specificity simply by mutations that discriminate against one target without affecting binding to another, as has been seen in HIV RRE (26).

Neutral and nearly neutral theories of evolution (28, 34) assert that incremental mutation paths connect distinct phenotypes without loss-of-fitness intermediates. The results of computational studies support the existence of such paths between discrete RNA secondary structures (17, 25), but our understanding of neutral paths between protein phenotypes is less advanced (14). In the case of N-boxB interactions, the coevolution of protein and RNA would expand the number of neutral paths. Recombination occurs between lambdoid phages (7, 8), directly sampling a more diverse sequence space than incremental mutation would sample and likely allowing access to otherwise inaccessible or genetically distant relaxed specificity sequences. Additionally, the enormous population of viruses may allow transient reduced-fitness intermediates to occur. Finally, the reduced type specificity that occurs due to increased expression of N (22) may allow many weakly active, relaxed-specificity mutants to be viable.

#### ACKNOWLEDGMENTS

We gratefully acknowledge funding from the American University of Beirut University Research Board and the William and Flora Hewlett Foundation for funding a research leave of C.A.S.

We express our heartfelt thanks to Naomi Franklin for providing irreplaceable strains, to Kazuo Harada for hosting C.A.S. in his laboratory, and to the group of André Megarbané for excellent sequencing services. This work benefited from access to the Central Research Science Laboratory at the American University of Beirut.

#### REFERENCES

1. Austin, R. J., T. Xia, J. Ren, T. T. Takahashi, and R. W. Roberts. 2003. Differential modes of recognition in N peptide-boxB complexes. *Biochemistry* **42**:14957–14967.

2. Barrick, J. E., and R. W. Roberts. 2003. Achieving specificity in selected and wild-type N peptide-RNA complexes: the importance of discrimination against noncognate RNA targets. *Biochemistry* **42**:12998–13007.
3. Barrick, J. E., T. T. Takahashi, A. Balakin, and R. W. Roberts. 2001. Selection of RNA-binding peptides using mRNA-peptide fusions. *Methods* **23**:287–293.
4. Bayer, T. S., L. N. Booth, S. M. Knudsen, and A. D. Ellington. 2005. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA* **11**:1848–1857.
5. Cai, Z., A. Gorin, R. Frederick, X. Ye, W. Hu, A. Majumdar, A. Kettani, and D. J. Patel. 1998. Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. *Nat. Struct. Biol.* **5**:203–212.
6. Calnan, B. J., B. Tidor, S. Biancalana, D. Hudson, and A. D. Frankel. 1991. Arginine-mediated RNA recognition: the arginine fork. *Science* **252**:1167–1171.
7. Campbell, A. 1994. Comparative molecular biology of lambdoid phages. *Annu. Rev. Microbiol.* **48**:193–222.
8. Casjens, S. R. 2005. Comparative genomics and evolution of the tailed-bacteriophages. *Curr. Opin. Microbiol.* **8**:451–458.
9. Cille, C. D., and J. R. Williamson. 2003. Structural mimicry in the phage  $\phi$ 21 N peptide-boxB RNA complex. *RNA* **9**:663–676.
10. Cille, C. D., and J. R. Williamson. 1997. Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA* **3**:57–67.
11. Coczaki, A. I., I. R. Ghattas, and C. A. Smith. 2008. Bacteriophage P22 antitermination boxB sequence requirements are complex and overlap with those of  $\lambda$ . *J. Bacteriol.* **190**:4263–4271.
12. Court, D. L., A. B. Oppenheim, and S. L. Adhya. 2007. A new look at bacteriophage  $\lambda$  genetic networks. *J. Bacteriol.* **189**:298–304.
13. Dambly, C., and M. Couturier. 1971. A minor Q-independent pathway for the expression of the late genes in bacteriophage lambda. *Mol. Gen. Genet.* **113**:244–250.
14. DePristo, M. A., D. M. Weinreich, and D. L. Hartl. 2005. Missense mean-derings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **6**:678–687.
15. Doelling, J. H., and N. C. Franklin. 1989. Effects of all single base substitutions in the loop of boxB on antitermination of transcription by bacteriophage  $\lambda$ 's N protein. *Nucleic Acids Res.* **17**:5565–5577.
16. Draper, D. E. 1999. Themes in RNA-protein recognition. *J. Mol. Biol.* **293**:255–270.
17. Fontana, W., and P. Schuster. 1998. Continuity in evolution: on the nature of transitions. *Science* **280**:1451–1455.
18. Franklin, N. C. 2004. Morphing molecular specificities between Arm-peptide and NUT-RNA in the antitermination complexes of bacteriophages  $\lambda$  and P22. *Mol. Microbiol.* **52**:815–822.
19. Franklin, N. C. 1993. Clustered arginine residues of bacteriophage  $\lambda$  N protein are essential to antitermination of transcription, but their locale cannot compensate for boxB loop defects. *J. Mol. Biol.* **231**:343–360.
20. Franklin, N. C. 1985. "N" transcription antitermination proteins of bacteriophages  $\lambda$ ,  $\phi$ 21 and P22. *J. Mol. Biol.* **181**:85–91.
21. Franklin, N. C. 1985. Conservation of genome form but not sequence in the transcription antitermination determinants of bacteriophages  $\lambda$ ,  $\phi$ 21 and P22. *J. Mol. Biol.* **181**:75–84.
22. Franklin, N. C., and J. H. Doelling. 1989. Overexpression of N antitermination proteins of bacteriophages  $\lambda$ , 21, and P22: loss of N protein specificity. *J. Bacteriol.* **171**:2513–2522.
23. Harada, K., S. S. Martin, and A. D. Frankel. 1996. Selection of RNA-binding peptides in vivo. *Nature* **380**:175–179.
24. Hilliker, S., and D. Botstein. 1976. Specificity of genetic elements controlling regulation of early functions in temperate bacteriophages. *J. Mol. Biol.* **106**:537–566.
25. Huynen, M. A., P. F. Stadler, and W. Fontana. 1996. Smoothness within ruggedness: the role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA* **93**:397–401.
26. Iwazaki, T., X. Li, and K. Harada. 2005. Evolvability of the mode of peptide binding by an RNA. *RNA* **11**:1364–1373.
27. Johnson, N. P., W. A. Baase, and P. H. von Hippel. 2005. Low energy CD of RNA hairpin unveils a loop conformation required for  $\lambda$ N antitermination activity. *J. Biol. Chem.* **280**:32177–32183.
28. Kimura, M. 1991. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc. Natl. Acad. Sci. USA* **88**:5969–5973.
29. Lazinski, D., E. Grzadziska, and A. Das. 1989. Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell* **59**:207–218.
30. Legault, P., J. Li, J. Mogridge, L. E. Kay, and J. Greenblatt. 1998. NMR structure of the bacteriophage  $\lambda$  N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**:289–299.
31. Martz, E. 2002. Protein Explorer: easy yet powerful macromolecular visualization. *Trends Biochem. Sci.* **27**:107–109.

32. **Miller, J. H.** 1992. A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, Plainview, NY.
33. **Minor, D. L., Jr, and P. S. Kim.** 1996. Context-dependent secondary structure formation of a designed protein sequence. *Nature* **380**:730–734.
34. **Ohta, T.** 2003. Origin of the neutral and nearly neutral theories of evolution. *J. Biosci.* **28**:371–377.
35. **Patel, D. J.** 1999. Adaptive recognition in RNA complexes with peptides and protein modules. *Curr. Opin. Struct. Biol.* **9**:74–87.
36. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
37. **Scharpf, M., H. Sticht, K. Schweimer, M. Boehm, S. Hoffmann, and P. Rosch.** 2000. Antitermination in bacteriophage  $\lambda$ . The structure of the N36 peptide-boxB RNA complex. *Eur. J. Biochem.* **267**:2397–2408.
38. **Smith, C. A., V. Calabro, and A. D. Frankel.** 2000. An RNA-binding chameleon. *Mol. Cell* **6**:1067–1076.
39. **Tan, R., and A. D. Frankel.** 1995. Structural variety of arginine-rich RNA-binding peptides. *Proc. Natl. Acad. Sci. USA* **92**:5282–5286.
40. **Weisberg, R. A., and M. E. Gottesman.** 1999. Processive antitermination. *J. Bacteriol.* **181**:359–367.
41. **Xia, T., A. Frankel, T. T. Takahashi, J. Ren, and R. W. Roberts.** 2003. Context and conformation dictate function of a transcription antitermination switch. *Nat. Struct. Biol.* **10**:812–819.
42. **Zhang, Q., X. Sun, E. D. Watt, and H. M. Al-Hashimi.** 2006. Resolving the motional modes that code for RNA adaptation. *Science* **311**:653–656.