

Subtle atomic group discrimination in the RNA minor groove

MAGALI FRUGIER* AND PAUL SCHIMMEL†

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Paul Schimmel, August 14, 1997

ABSTRACT As a problem in molecular recognition and for drug discovery, great interest has developed around the possibility that RNA structures could be discriminated by peptides and other small molecules. Although small peptides have been shown to have the capacity to discriminate specific bulges and loops in RNA molecules, discrimination of double helical regions by a peptide binder has not been reported. Indeed, the most accessible part of an RNA helix is the minor groove, and fundamental stereochemical considerations have suggested that discrimination of at least some base pairs would be difficult in the minor groove. Here we report the design and isolation of a peptide binder that manifests the most subtle kind of discrimination of base pair differences in the RNA minor groove. Functional discrimination of a single atomic group is demonstrated as well as the difference between two different angular orientations of the same group. This report of RNA helix discrimination by a peptide binder suggests a richer potential for RNA minor groove recognition than previously thought.

The common RNA A-form helix is characterized by major and minor grooves that are lined with distinct atomic groups emanating, respectively, from opposite sides of the paired bases. The constellation of atoms within a particular groove is determined by sequence, so that sequence-specific recognition of helices is generally determined by interactions with a binder in one of the grooves. For the common DNA B-form helix, the major groove is wide and will accommodate a polypeptide binding element such as an α -helix. The minor groove is narrow and therefore less accessible, so that sequence recognition of DNA helices is generally determined by major groove interactions. In contrast, for the RNA A-helix, the major groove is narrow (and deep) whereas the minor groove is wide (and shallow) and more accessible for a sequence-specific binder (1).

The minor groove side of the RNA helix presents a special challenge for specific recognition. For example, the 2-amino group of G is a prominent hydrogen bond donor and is a prime site for interaction with an RNA binder. In particular, G-C and G-U base pairs are thought to be difficult to distinguish from their respective transversions (C-G and U-G, respectively) (2, 3). The problem is that the 2-amino group of G lies on the dyad axis of the minor groove of the helix, so that it is in the same location in G-C vs. C-G or in G-U vs. U-G. The relatively infrequent G-U wobble pairs are of particular interest because they play an important role in RNA recognition, such as in the active center of the Tetrahymena self-splicing group I intron (4) and in specific protein-RNA complexes of the translation apparatus (5–7). In some examples, the unpaired 2-amino group of G-U has been shown to be critical (8, 9) and, at the same time, not replaceable by having a U-G pair (10). Whether these represent special examples with complex systems, or

whether the potential for this kind of subtle recognition is a more general and intrinsic feature of RNA recognition systems has not previously been known.

To investigate whether G-U can be distinguished from U-G with a small peptide binder, we used a hairpin microhelix that has a G-U pair at a specific position and selected for peptide binders within a phage display library (Fig. 1A). [This microhelix is based on the acceptor stem of an *Escherichia coli* alanine tRNA and can be charged with alanine in a G-U-dependent manner (11).] The idea was to select peptide-containing phages that bound to the G-U-containing microhelix but did not bind to the same microhelix that replaced G-U with either I-U, G-C, or U-G (Fig. 1B). The I-U pair differs from G-U only in that it lacks the exocyclic 2-amino group, that is, all other atoms in the major and minor grooves are preserved. Thus, a peptide that distinguishes G-U from I-U functionally binds in the minor groove and requires the 2-amino group. Lack of binding to the G-C pair shows that the 2-amino group has to be unpaired and in the wobble position. Discrimination against U-G shows that the angle at which the 2-amino group protrudes into the minor groove is sensed by the peptide.

MATERIALS AND METHODS

RNA-Directed Phage Display Library. An RNA-directed phage display library was designed and constructed. A DNA oligonucleotide coding for 10 degenerated codons placed between two RNA binding motifs was cloned into phagemid pCANTAB 5E (Pharmacia). Transformed TG1 cells (Pharmacia) were rescued by helper phage M13KO7 (Pharmacia), and each phage produced displayed a unique epitope sequence in fusion with that of the gene III protein. The original library contained 2.10^7 different phage clones, which represents about $10^{-4}\%$ of the theoretical complexity of a decapeptide library (20^{20} possibilities).

Affinity Selections for G-U Binders. Peptide motifs that recognize microhelix G-U were purified by affinity through 7 rounds of selection. Rounds 1–4 were positive selections in presence of the microhelix G-U alone, whereas rounds 5–7 were performed in presence of the competitor microhelices. The immunopure immobilized streptavidin agarose (400 μ l of the slurry solution) (Pierce) was pretreated for 1 hour in blocking buffer (100 mM NaHCO₃, pH 8.6/5 mg/ml BSA/0.1 μ g/ml streptavidin/0.02% NaN₃) and washed with 5 ml of TBS (50 mM Tris-HCl, pH 7.5/150 mM NaCl). The biotinylated microhelix G-U (15,000 pmol for selections 1–4, and 125 pmol for the selections 5–7) was incubated with the phage library (10^{10} phages for selections 1 and 2 and 10^6 for selections 3–7), 150 μ M biotin, and pretreated resin in 500 μ l TBS. Incubation conditions varied from 20 hours at 4°C (rounds 1–4) to 1 hr at

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/9411291-4\$2.00/0
PNAS is available online at <http://www.pnas.org>.

*Present address: Institut de Biologie Moléculaire et Cellulaire du Centre National de la Recherche Scientifique, 15, rue René Descartes, 67084 Strasbourg Cedex, France.

†To whom reprint requests should be sent at the present address: The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037. e-mail: schimmel@scripps.edu.

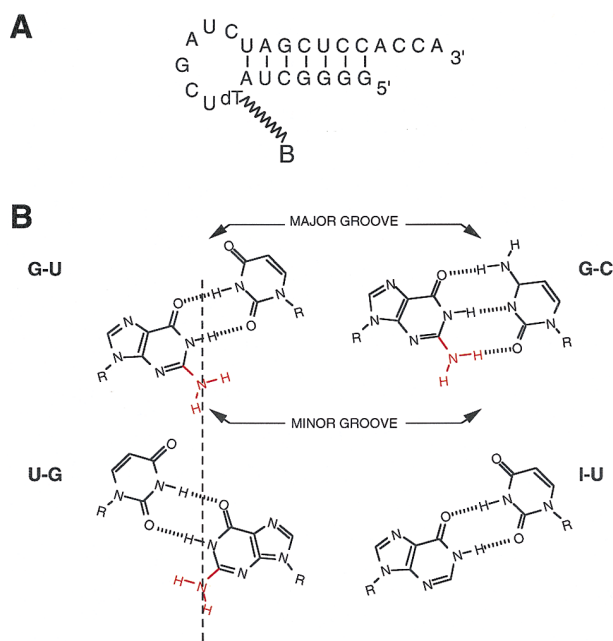


FIG. 1. (A) Sequence and hairpin structure of an RNA microhelix based on the acceptor stem of *E. coli* tRNA^{Ala}. This hairpin is a substrate for aminoacylation with alanine (9). The position in the loop of the biotinylated thymidine is shown (biotin dT phosphoramidite from Glen Research, Sterling, VA). (B) Diagrammatic representation of G-U vs. U-G wobble pairs, showing the position of the 2-amino group on the dyad axis (indicated with a vertical line). The I-U pair is identical to G-U except for the missing 2-amino group. The G-C pair is also shown.

room temperature (rounds 5–7). The columns were washed extensively with 35 ml TBS (and, in addition, 5 ml TBS containing 10 μ g/ml streptavidin for selections 1–4). Specific RNA–phage complexes retained on the column were amplified by infection of *E. coli* TG1 cells. During the competition selections (rounds 5–7) the phage was incubated in the presence of 125 pmol of immobilized microhelix G-U and 1,250 pmol of nonbiotinylated microhelix G-C, I-U, or U-G.

Polyacrylamide Affinity Coelectrophoresis. For polyacrylamide affinity coelectrophoresis gels: 5 μ l samples of radio-labeled microhelix G-U (5 nM [⁵-³²P] RNA/250 μ g/ml ribopolyadenosine/10% glycerol/1 \times TBS) were electrophoresed for 80 min (100 V at 4°C) through a 6% acrylamide/bisacrylamide gel (29:1, 1.5 \times 80 \times 100 mm³) containing increasing concentrations of peptide (6 nM to 10 μ M) in 0.5 \times TB (50 mM Tris base/50 mM boric acid). The gels were dried and analyzed on a PhosphorImager (Molecular Dynamics).

Competition Experiments that Delineate Specificity. The specificity of the interaction was determined by competition experiments where the four microhelices (G-U, G-C, I-U and U-G) were tested for their ability to inhibit the formation of the complex. In these experiments, the peptide concentration in the gel [12% acrylamide/bisacrylamide (29:1) in 0.5 \times TB] was uniform (0.1 μ M). The samples, containing 5' labeled microhelix G-U (1 μ M) and increasing concentrations of competitor microhelices (0.625–40 μ M) were electrophoresed for 150 min, 100 V at 4°C.

RESULTS

Scheme for Selections. The general scheme was to pass peptide-containing phages over the column and to remove those that did not bind. The peptide element is presented at the N terminus of the gene III protein, and diverse sequences of the peptide were generated by nucleotide variations within its gene coding region (10). Phages that bound to the column were

eluted and amplified and subjected to another round of selection, and so on. Subsequent elutions of bound phages involved successive passages of solutions of microhelices containing, respectively, the I-U, G-C, and U-G base pairs (11). Phages that were liberated by these elutions were discarded and, ultimately, phages that were bound to the G-U-containing microhelices were retained.

Design of Phage-Encoded Peptide Motif and Selection of G-U-Specific Binders. In our first experiments, a hexapeptide-encoding element in the gene III coding sequence was randomized (12). After multiple rounds of selection with about 6.4×10^6 different sequences, no binders with all of the desired characteristics (*vide supra*) were obtained. Reasoning that the size of the peptide may have been too small and that the library had no bias toward those that bind to RNA, we next created a 10-amino acid variable region that was flanked on either side by generic RGG motifs of 9 amino acids (Fig. 2). The RGG motifs provide for general, nonspecific RNA binding (13, 14). Thus, the rationale was to have a library where many of the peptides bound to RNA and to select for sequences in the central variable region that would confer specificity.

After 7 rounds of selection, using competitors with I-U, G-C, and U-G pairs, two binders were obtained from an initial library pool of about 2×10^7 phages (of 60 clones that were sequenced, these two comprised 45% of all bound phages obtained at this final step in the selections). The sequence of each was determined from the DNA coding region. One peptide contained two cysteines and, because of concerns about aggregation due to sulfhydryl oxidation, this peptide was not investigated further. The sequence of the second peptide—designated as MF β 2—was chemically synthesized (Fig. 2) and then tested directly for RNA binding to the microhelices.

Binding Affinity and Specificity with Chemically Synthesized MF β 2 Peptide. For this purpose, the polyacrylamide coelectrophoresis procedure was used. This procedure is a variation of the agarose coelectrophoresis method used to study protein–RNA interactions (15–17). With this method, a slab gel is used with multiple lanes that contain successively higher concentrations of embedded protein. Labeled RNA is then electrophoresed through the gel and shifted in position according to the degree to which it is bound. The result is a titration curve from which the dissociation constant for the complex can be determined. By using polyacrylamide instead of agarose, smaller peptide complexes can also be detected by this method (17).

The binding of the MF β 2 peptide to the G-U-containing microhelix showed a smooth titration curve consistent with 1:1 complex formation and characterized by a K_d of 300 nM at pH 7.5 (Fig. 3A). When challenged with specific competitors containing different substitutions of the G-U pair, only the wild-type G-U-containing sequence was an efficient competitor (Fig. 3B). In particular, the ability to discriminate G-U from I-U showed that the peptide has a functional requirement

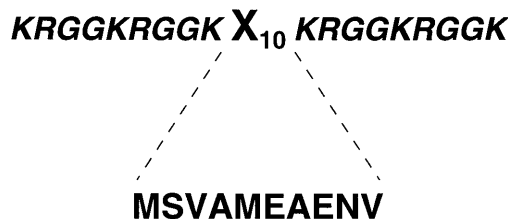


FIG. 2. Design of the peptides in the library. The peptides contain a 9-amino acid RGG motif that flanks both sides of a 10-amino acid variable segment designated X₁₀. Although the library has a theoretical complexity of 10^{13} , the practical limit that was screened consisted of about 2×10^7 peptides. The sequence of the selected MF β 2 peptide in the X₁₀ segment is also shown. In addition, during selections, one R was changed to H in one RGG motif.

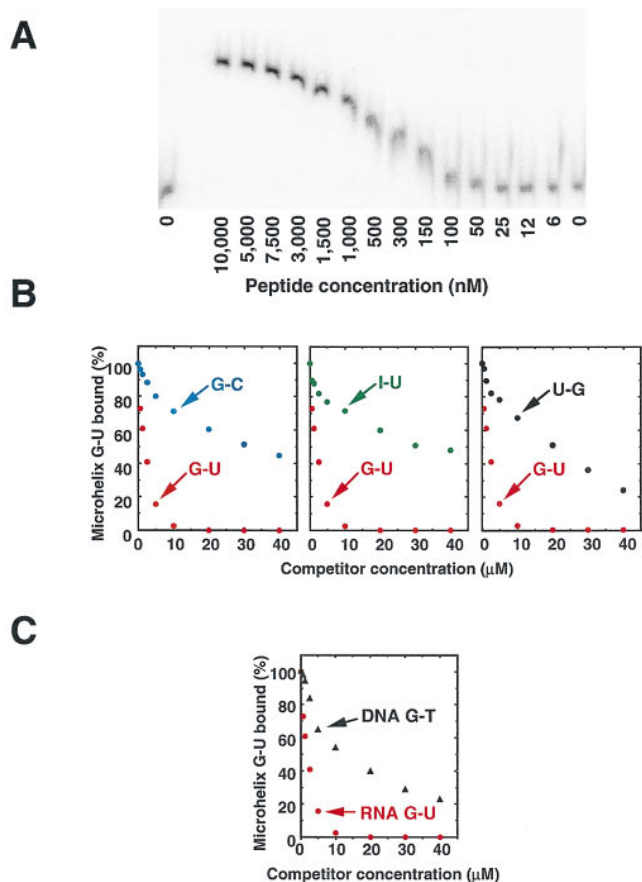


FIG. 3. Polyacrylamide coelectrophoresis of the peptide binder MFβ2 with and without specific competitors at pH 7.5, 4°C. (A) A peptide concentration gradient was established across the different lanes of the gel, as indicated. [5'-³²P]RNA was electrophoresed into the peptide-containing gel, and the RNA was shifted according to the extent of binding. (B) Competition assays with unlabeled microhelices containing G-U, I-U, G-C, and U-G base pairs. Only the G-U-containing microhelix is an effective competitor. (C) Competition assays with a DNA microhelix containing a G-T base pair.

for the 2-amino group in the minor groove, whereas the discrimination against G-C showed that a free amino group is required. Lastly, discrimination against U-G established that the peptide was sensitive to the angle at which the free amino group protrudes into the minor groove. We estimate that the discrimination against U-G is 20- to 25-fold. It is even higher against the other pairs which were no better than nonspecific single-stranded RNA oligonucleotides (data not shown).

In DNA helices, the minor groove is narrower than in RNA and, therefore, less accessible to a minor groove binder. We tested the binding of the MFβ2 peptide to a DNA microhelix containing the 2-amino group in a G-T base pair. As would be expected for a minor groove peptide binder, the DNA microhelix containing a G-T base pair was only a weak competitor of the G-U-containing RNA microhelix (Fig. 3C). These results further demonstrated the high selectivity of the peptide binder.

The MFβ2 peptide was also fused to an unrelated protein—glutathione-S-transferase. This fusion protein had the same G-U-specific discrimination as seen with the free MFβ2 peptide. Using a filter binding assay, we determined that 1.08 mol of fusion protein were bound per RNA microhelix (data not shown.) Thus, the MFβ2 peptide can confer G-U-specific binding on a foreign protein.

Determining Essential Residues for G-U-Specific Interaction. To determine which of the 10 amino acids within MFβ2 were essential for the G-U-specific interaction, ran-

dom mutations were introduced into the peptide coding sequence with a frequency of about one substitution per peptide (Fig. 4A). The phage selection was again carried out, and G-U-specific binders were isolated and sequenced. This analysis showed that in the sequence MSVAMEAENV, the serine and first glutamic (shown in bold) were conserved among the specific binders. Starting with this pool, another round of selection was carried out, keeping fixed the aforementioned serine and glutamic acid residues. The specific binders in this new pool had two additional conserved residues and had the general sequence XSXAXEXXNX. If this peptide element forms an α-helix, then the functional hydrophilic groups of S, E, and N lie on the same side of the helix, whereas A (a strong helix former) is on the opposite side. Consistent with this possibility, the decapeptide segment was strongly predicted to form an α-helix (Fig. 4B) (18) and, in addition, we detected α-helix content (25%) in the MFβ2 peptide by circular dichroism measurements (data not shown).

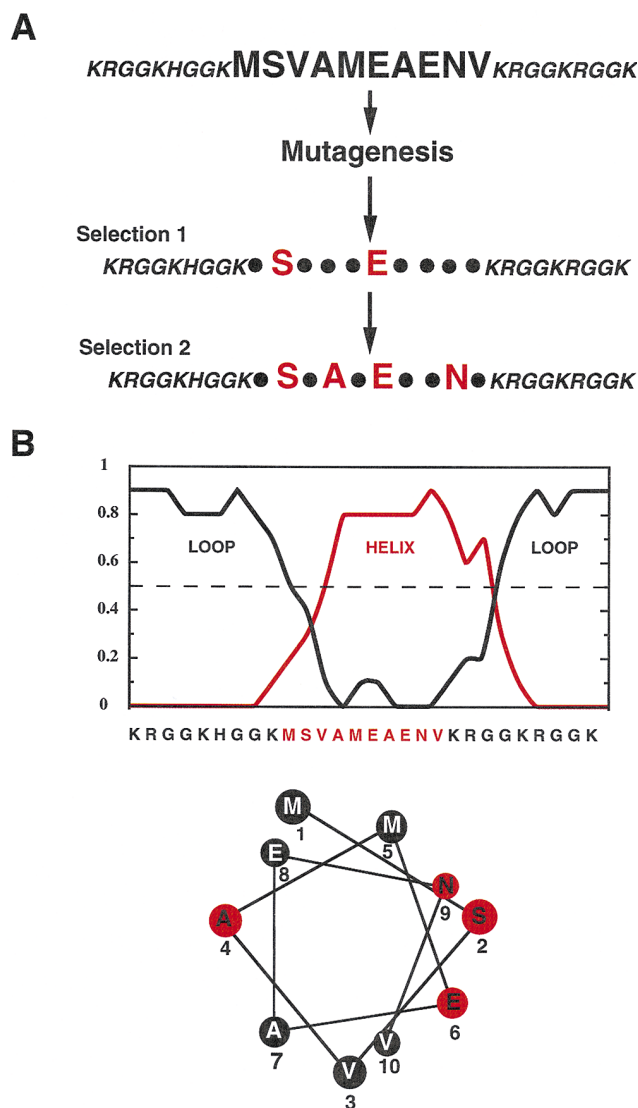


FIG. 4. Mutagenesis of the 10 amino acid element in the MFβ2 peptide and selection of G-U-specific binders. (A) In the first round of selection, G-U-specific binders were obtained that conserved 2 of the 10 positions in the peptide. In the second round, two additional residues were conserved in the G-U-specific binders. (B) Helix prediction (18) and helical wheel projection of the predicted central helical element in the MFβ2 peptide.

CONCLUSION

These data demonstrate that, contrary to reasonable expectations (2, 3), the most subtle discrimination in the RNA minor groove can be achieved with a relatively small peptide binder. By way of contrast to earlier work, this work reports the isolation of a peptide binder to an RNA helix. Previous studies demonstrated small molecule or peptide recognition of bulges and loops (19–21), but not of anything so subtle as a directional sensitivity to a 2-amino group in a helix. Possibly, this directional sensitivity is dependent on the positional location of U-specific atomic groups that are differently located in G-U vs. U-G. In the particular example studied here, a G-U pair was investigated because these pairs play a prominent role in RNA recognition systems. This prominence may be due to the relatively infrequent occurrence of G-U pairs, making them attractive as specificity-determining sites. The comparative ease of isolation of a specific binder in these studies demonstrates that there is no inherent limitation to minor groove recognition of these wobble pairs.

We thank Prof. Jaime Williamson (Massachusetts Institute of Technology) and Alan Frankel (University of California, San Francisco) for their critical comments on the manuscript and Dr. Lluís Ribas de Pouplana (Massachusetts Institute of Technology) for helpful discussions. This work was supported by a grant from the National Foundation for Cancer Research. M.F. was supported by a fellowship from the Association pour la Recherche sur le Cancer (1994–1995).

1. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Advanced Texts in Chemistry (Springer, New York).

2. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 804–808.
3. Steitz, T. A. (1990) *Q. Rev. Biophys.* **23**, 205–280.
4. Strobel, S. A. & Cech, T. R. (1995) *Science* **267**, 675–679.
5. Hou, Y.-M. & Schimmel, P. (1988) *Nature (London)* **333**, 140–145.
6. McClain, W. H. & Foss, K. (1988) *Science* **240**, 793–796.
7. Batey, R. T. & Williamson, J. R. (1996) *J. Mol. Biol.* **261**, 550–567.
8. Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. & Schimmel, P. (1991) *Science* **253**, 784–786.
9. Strobel, S. A. & Cech, T. R. (1996) *Biochemistry* **35**, 1201–1211.
10. Musier-Forsyth, K., Shi, J.-P., Henderson, B., Bald, R., Furste, J. P., Erdmann, V. A. & Schimmel, P. (1995) *J. Am. Chem. Soc.* **117**, 7253–7254.
11. Francklyn, C. & Schimmel, P. (1989) *Nature (London)* **337**, 478–481.
12. Scott, J. K. & Smith, G. P. (1990) *Science* **249**, 386–390.
13. Kiledjian, M. & Dreyfuss, G. (1992) *EMBO J.* **11**, 2655–2664.
14. Burd, C. G. & Dreyfuss, G. (1994) *Science* **265**, 615–621.
15. Lim, W. A., Sauer, R. T. & Lander, A. D. (1991) *Methods Enzymol.* **208**, 196–210.
16. Gale, A. J. & Schimmel, P. (1995) *Biochemistry* **34**, 8896–8903.
17. Cillery, C. D. & Williamson, J. R. (1997) *RNA* **3**, 57–67.
18. Rost, B. & Sander, C. (1993) *J. Mol. Biol.* **232**, 584–589.
19. Tan, R., Chen, L., Buettner, J. A., Hudson, D. & Frankel, A. D. (1993) *Cell* **73**, 1031–1040.
20. Tan, R. & Frankel, A. D. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 5282–5286.
21. Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996) *Science* **273**, 1547–1551.