Rhyme or reason: RNA–arginine interactions and the genetic code

Robin D Knight and Laura F Landweber

Theories about the origin of the genetic code require specific recognition between nucleic acids and amino acids at some stage of the code's evolution. A statistical analysis of arginine-binding RNA aptamers now offers the opportunity to test such interactions and provides the strongest support for an intrinsic affinity between any amino acid and its codons.

Address: Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544-1003, USA.

Correspondence: Laura F Landweber
E-mail: lfl@princeton.edu

Chemistry & Biology September 1998, **5**:R215–R220
http://biomednet.com/elecref/10745521005R0215

Introduction

The selection of RNA molecules (aptamers) that bind amino-acid ligands has made theories about the origin of the genetic code at last testable. The genetic code assigns similar amino acids to similar codons (Figure 1) [1]. This could be a result of selection to minimize the effect of mutations [2,3] or translation error [4], or of codon concession by metabolic precursors to related derivatives [5,6]. Alternatively, if similar amino acids interact favorably with similar RNA sequences, the observed relationships in the genetic code could have a chemical, rather than an adaptive, basis [7]. If any codon assignments arose from specific binding between amino acids and short RNA motifs, then directed evolution might be able to recapitulate such interactions *in vitro*.

Various authors have suggested that the original amino-acid-binding motifs could have been the actual codons [8] or some transform of them [4], such as the anticodon [9] or codon–anticodon duplexes [10]. Recent support for the codon–amino-acid pairing hypothesis comes from a specific interaction between arginine and its codons; the guanosine-binding site of self-splicing group I introns also binds arginine, and a conserved arginine codon confers this specificity [11].

Another possibility is that the original amino-acid recognition took place at the tRNA acceptor stem rather than the anticodon [12]. Such recognition could occur, for example, by a stereochemical interaction at a 'C4N' (complex of four nucleotides) [13], which consists of the three nucleotides at the 5′ end (assumed to be identical in sequence to the anticodon) together with the 'discriminator base' (the nucleotide immediately preceding the invariant CCA at the 3′ end of the tRNA). This model is consistent with evidence that the acceptor-stem domain and the anticodon domain of tRNA molecules might have independent evolutionary histories [14–18]. The various sites that have been suggested as the primitive binding sites are shown in Figure 2.

Each of the stereochemical hypotheses predicts that specific short RNA motifs will be found at sites that bind amino acids. Consequently, aptamers (nucleic-acid molecules selected to bind specific ligands) [19–21] that recognize amino acids should contain these sequences at their binding sites. RNA aptamers have been isolated to several amino acids, but most research has focused on RNA aptamers that bind arginine [22–27] because free arginine can mimic the natural interaction of HIV Tat peptides with TAR RNA [28].

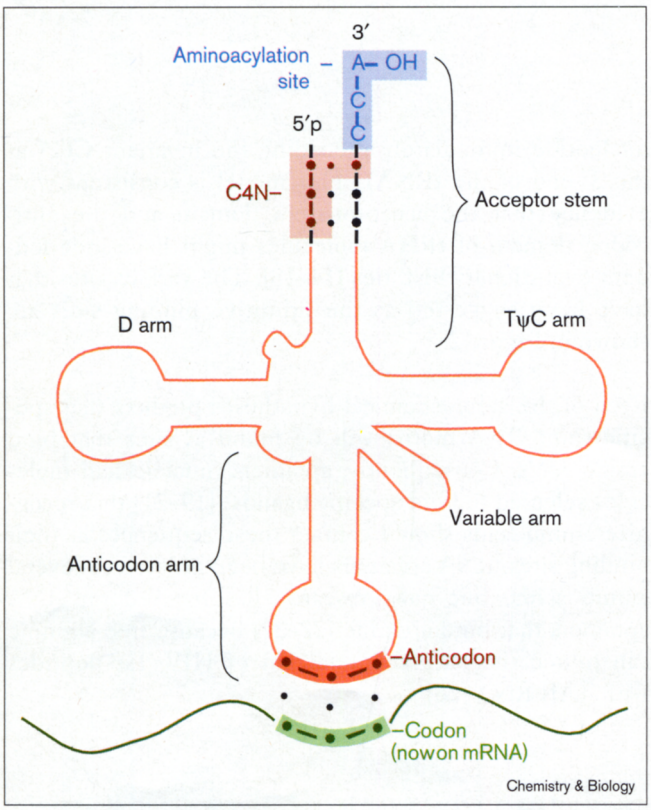
Figure 1

	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
	UUA	Leu	UCA	Ser	UAA	TER	UGA	TER
	UUG	Leu	UCG	Ser	UAG	TER	UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Chemistry & Biology

The 'universal' genetic code. Similar colors reflect similar sidechain composition: purple, hydrophobic; cyan, aromatic; green, hydroxyl containing; red, acidic; orange, amide; blue, basic; yellow, sulfur containing. Tyrosine is aromatic and contains a hydroxyl group, and so is intermediate between green and cyan; similarly, histidine is aromatic and basic. Intensity of color reflects molecular volume [41]. In general, amino acids with similar properties are linked by single-base mutations.

Figure 2



Structure of modern tRNA and mRNA, showing the codon, anticodon and C4N in their present positions. Note that in contemporary organisms the anticodon never contacts the amino acid directly (although it is sometimes important for recognition by the correct aminoacyl-tRNA synthetase).

As with the group I intron, another anecdotal but compelling example of a specific interaction between an amino acid and its codons comes from an arginine aptamer produced by randomization and reselection from an aptamer for the closely related amino acid citrulline [25]. Arginine differs from citrulline only by one moiety: arginine has an imino group where citrulline has a carbonyl group (Figure 3). The arginine aptamer differs from the citrulline aptamer by precisely three point substitutions, which together create two new arginine codons (Figure 4). Nucleotides in both of these arginine codons surround the amino-acid sidechain, forming a set of hydrogen bonds that interact directly with the amino acid [29].

To test whether there is a statistical association between binding sites and codons, it is necessary to have structural data for several independent sequences that bind the same amino acid. Because structural data are available for five phylogenetically unrelated arginine aptamers selected in four experiments in three different laboratories under different conditions [23,26,27,29], we tested whether the arginine-binding sites of these RNA aptamers contain a statistical excess of any of the predicted motifs (codons or anticodons). This allowed us to test the validity of several stereochemical hypotheses. Because we performed many individual tests on the same data, and because statistical results are often controversial in this field, we chose an arbitrary, low cutoff of $P = 0.01$ for statistical significance in any test.

Statistical analysis of aptamers

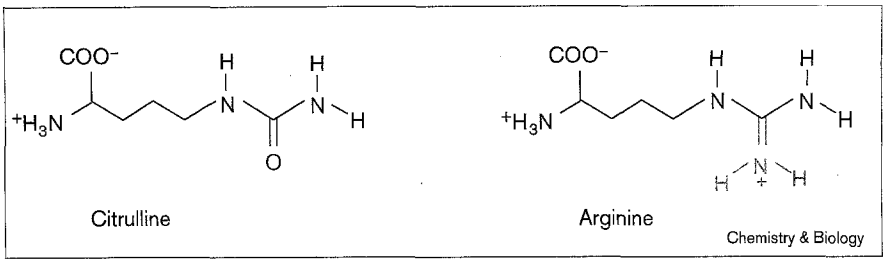
If one or more of the stereochemical hypotheses were true, nucleotides at arginine-binding sites would be expected to participate in arginine codons and/or anticodons. We analyzed 2×2 tables partitioning nucleotides into 'binding site' and 'nonbinding site' classes and into 'motif' and 'nonmotif' classes, depending on whether they are present in the ligand-binding site or in a particular nucleotide-sequence motif, respectively. 'Binding site' nucleotides are either shown by nuclear magnetic resonance (NMR) to form the ligand-binding pocket [29], or are implicated in binding by chemical-modification assays [23,26,27] (Table 1a–e). The latter included both chemical modification protection/deprotection assays (in which nucleotides are protected from chemical modification and/or enzymatic cleavage only when arginine is bound), and assays that measure the extent to which base modification impairs binding. 'Motif' nucleotides participate in the appropriate triplet(s) in any of the three possible reading frames. A statistical association between the two properties would indicate that distribution of the motifs is nonrandom with respect to arginine-binding sites. We used the G test for independence with Williams's correction for continuity [30] to detect interaction between the two variables. Where predictions were directional, the calculated p values were halved (and subtracted from unity when the deviation was in the opposite direction from that predicted) to reflect the fact that deviations in only one direction were relevant.

To ensure that any observed association was due to sequence rather than composition, we also tested triplets in which the first two bases were permuted. Thus, for the arginine codon classes CGN and AGR (N, any nucleotide; R, purine), we tested the non-arginine codons GCN and GAR, which have identical nucleotide compositions but different sequences.

As a further negative control, we performed the same statistical tests on all the RNA aptamers to ligands other than arginine for which published NMR structures were available (Table 1f–j). The aptamers to AMP [31], FMN [32], citrulline [29,33], tobramycin [34] and theophylline [35] would not be expected to show the same binding-motif associations as those for arginine, unless such associations arise from biases in composition that are general to binding sites in all RNA aptamers. NMR determines binding sites more accurately than chemical probing; we

Figure 3

Structures of arginine and citrulline.



therefore restricted our analysis of non-arginine aptamers to those that had been examined using this technique.

Purine bias

Binding sites of all aptamers are strongly purine biased. Although purines make up 54% of arginine and 58% of non-arginine aptamer sequences, they constitute 78% and 72% of their respective binding sites. Because one class of arginine codons, AGR, consists entirely of purines, such a purine bias could produce a spurious association between binding sites and arginine codons. Purine bias should also result in excess AAR, GAR, and GGR purine triplets at binding sites in arginine aptamers, however. We observed no such excess ($P >> 0.01$) in any case.

Only arginine aptamers overrepresent the arginine set

Arginine aptamers contain far more arginine codons at binding sites than expected by chance (Table 2; $G = 20.2$; $P = 3.4 \times 10^{-6}$). Of the 32 nucleotides at binding sites of arginine aptamers, 23 (72%) participate in arginine codons. In contrast, of the 59 nucleotides at binding sites of non-arginine aptamers, only 17 (29%) participate in arginine codons. The probability that a base in a random sequence is in an arginine codon is bounded by the relationship (6 arginine codons/64 total codons) × 3 reading frames = 28%, although simulation of a Markov process using the base frequencies in the arginine aptamers indicates that the true probability is closer to 34%. As expected, aptamers to ligands other than arginine show no association between arginine codons and arginine-binding sites ($G = 0.14$, $P >> 0.01$).

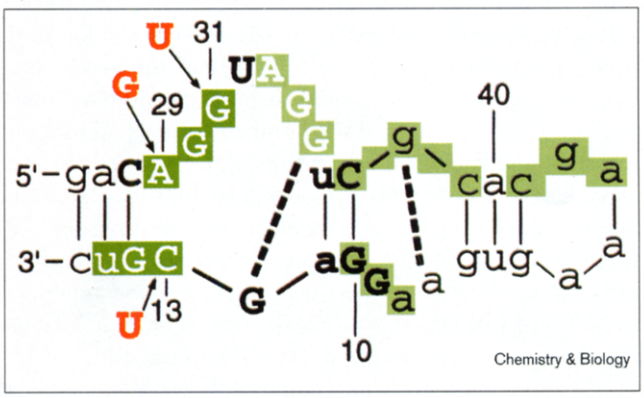
Arginine aptamers overrepresent only arginine codons

If the overrepresentation of codons at arginine-binding sites were due to composition rather than to sequence, then permutations of the codons (i.e. GCN for CGN and GAR for AGR) should be similarly overrepresented. This is not the case (see above). The set of arginine codons showed a much higher association with arginine-binding sites ($G = 20.2$; $P = 3.4 \times 10^{-6}$) than did the codon set of any other amino acid. The next highest match was proline ($G = 4.63$; $P > 0.01$). We also found no association between arginine-binding sites and the arginine anticodons NCG and YCU ($G = 0.19$; $P >> 0.01$; Y, pyrimidine).

No other codon set binds arginine as well as arginine codons

To ensure that the strong association we observed between arginine-binding sites and arginine codons was not a fluke, we tested for possible association between arginine-binding sites and all possible codon sets. We defined a codon set as a 'family box' consisting of all four third-position variants of one codon fixed at the first two positions, together with a 'doublet' consisting of two fixed bases followed by either pyrimidine or either purine. This describes all actual six-codon sets: for instance, serine is UCN + AGY. Out of 480 possible codon sets, the actual arginine set (CGN + AGR) has the highest G of 20.2 ($P = 3.4 \times 10^{-6}$). The next highest set (AGN + CGR), which contains four of six arginine codons, gave a degree of association almost two orders of magnitude less improbable ($G = 13.1$; $P = 1.5 \times 10^{-4}$). The highest association between codons and arginine-binding sites for sets not including arginine codons was for purine-rich GAN + GGY ($G = 5.78$; $P = 0.008$), which is not significant because 378 such codon sets were tested (the probability

Figure 4



Secondary structure of an arginine aptamer derived from a citrulline aptamer by three nucleotide substitutions (arrows); all occur within two new arginine codons (dark green boxes). Note that creation of the first arginine codon requires substitutions at both the first and third positions. Four additional arginine codons are shown as light green boxes. Essential nucleotides are in boldface; nucleotides selected in all isolates in uppercase. Dashed lines indicate noncanonical pairs. Adapted from [29].

Table 1

Sequences and nucleotides forming binding sites of arginine and non-arginine aptamers.

(a) gacAGGuAgGucgcacgaaagugaaGgaGCguc

(b) gggagcucagaauaaacgcucaacccgacagaucggcAaCgCCnuguuuucgacangAgACaccgauccugcaccaaagcuucc

(c) augauAAAccgAugcugggcgAuucuccugaaguagggagagAguugucauguauggg

(d) gggagaauucccgcggcuggugcgugcaggacGUcGAucgaauccGccugcaGugcacggcuccc

(e) gggagaauucccgcgcagcGGUcGAaaucgucaugugcacugcuacugcagugcacggcuccc

(f) ggguugGGAAGAAacuguggcacuucggugccAGcaaccc

(g) gacGGUuAgGucgcacgaaagugaaGgaGUguc

(h) ggcguguAGGAUAugcuucggcaGAAGGacacgcc

(i) ggcacgagGUUUAGCUACAcucgugcc

(j) ggcgaUACCagccgaaaaggcccuugGCaGcgcuc

Sequences of aptamers used in this analysis for **(a–e)** arginine [23,26,27,29], **(f)** AMP [31], **(g)** citrulline [29], **(h)** FMN [32], **(i)** tobramycin [34] and **(j)** theophylline [35]. Capital letters denote nucleotides implicated in binding, as determined by NMR or by chemical and enzymatic probing. Nucleotides participating in CGN arginine codons are green; those participating in AGR arginine codons are red.

that at least one trial would give a result this extreme is greater than 0.01).

Monte Carlo simulations

Finally, we tested the appropriateness of the G test, which requires independent observations, because the probability that a nucleotide participates in a codon or binding site is influenced by its surrounding nucleotides. We ran Monte Carlo simulations that randomly generated sequences of the same length as the actual aptamers, and tested whether the observed values of G matched the probabilities obtained from the standard G test. In the weakest condition, the randomized aptamers were constrained to the same base frequencies overall as the actual aptamers and the binding-site positions were randomized. Under these conditions, only 1 of the 100,000 randomized aptamer sets gave a G value higher than the actual set, indicating $P \approx 10^{-5}$. In the most stringent condition, first, nucleotides were assigned with probabilities equal to their actual frequencies in binding sites and elsewhere in the aptamers, to account for compositional bias, and second, binding sites were fixed to their actual positions in the aptamers, to account for spatial correlation within binding sites. Under these conditions, 124 of the arginine aptamers showed positive codon-binding site associations at least as strong as that observed for arginine aptamers. In contrast, 48,985 of the non-arginine aptamers showed associations at least as strong as those observed for non-arginine aptamers. Thus we conclude that only arginine aptamers have significant bias in favor of arginine codons at their binding sites, and that the true probability of association between arginine codons and binding sites might be closer to 1×10^{-3} than 6×10^{-6} (as calculated without accounting for composition bias). Even in the most stringent tests, however, the association between arginine codons and arginine-binding sites remained significant ($P << 0.01$). This analysis also revealed that the true probability of association between AGR codons and binding sites for ligands other than arginine is much greater than 0.01, as 8356 of the 100,000 randomizations of non-arginine codons gave positive associations at least as strong as that observed. (Because of the strong purine bias, AGR codons tended to associate with non-arginine-binding sites [$G = 15.3$; $P = 4.6 \times 10^{-5}$]. GGR codons showed a similar association [$G = 6.47$; $P = 0.005$], when the binding-site nucleotide biases were not taken into account.)

Conclusions

These results show a clear association between both arginine codon classes (CGN and AGR) and regions of RNA molecules that bind arginine. We found no association between arginine-binding sites and arginine anticodons, and no stronger associations between arginine-binding sites and any other possible set of six codons conforming to the 4 + 2 rule of a family box and doublet (see above).

Table 2

Arginine codon/binding-site frequencies for arginine and non-arginine aptamers.

	Codon	nt in Codon*	nt not in Codon*	G	P
Arginine aptamers	Binding	23	9	20.2	3.4×10^{-6}
	Not binding	83	190		
Others	Binding	17	42	0.14	0.35
	Not binding	29	82		

Tests for association between codons and binding sites were directional. *The number of nucleotides involved in arginine codons need not be a multiple of three, because some codons overlap. nt, nucleotide.

Because the C4N hypothesis states that the anticodon forms part of the binding complex, these results provide evidence against both it and the codon–anticodon double-helix hypothesis. If *in vitro* selection protocols mimic an appropriate evolutionary environment, and if aptamer selections are influenced by the same chemical interactions that led to codon assignments, then these results (for the one amino acid for which sufficient data are available) support the hypothesis that amino acids can interact specifically with RNA sequences that contain their cognate codons [36]. If the present-day arginine codons preserve original assignments determined by stereochemical affinity between amino acids and RNA, then their positions in the genetic code could even be considered molecular fossils of an ancient chemical determinism.

Arginine is an unusual amino acid in that the guanidino moiety of its sidechain closely mimics the hydrogen-bonding face of guanine. Furthermore, its positive charge allows it to participate in electrostatic interactions with the phosphate backbone of RNA that are not available to other amino acids. Specific RNA aptamers have been selected to hydrophobic amino acids [37–39], however, and thus interactions other than ionic and hydrogen-bonding must contribute to amino-acid recognition, although there are insufficient structural data to perform statistical tests on these other aptamers. The selection experiments were carried out in three different laboratories and from independent starting pools, so the resulting aptamers should be an unbiased representation of sequence space. Furthermore, attempts to select aptamers against positively charged lysine have been unsuccessful [25], despite the potential for electrostatic interactions similar to those used by arginine aptamers.

Alternatively, it has been suggested that, rather than being primitive, arginine could have been a relatively late addition to the standard repertoire of amino acids [40]. Thus, arginine could have captured those codons for which it has greatest affinity. If so, we conjecture that amino acids incorporated earlier might not show such an association between their codons and binding sites. This hypothesis can only be tested with structural data for multiple aptamers to amino acids besides arginine.

Prospects

These results imply the specific prediction that arginine codons will play prominent roles in other RNA–arginine interactions, and that some of the other 19 amino acids will bind RNA sequences that contain their codons. More general conclusions await the structural analysis of aptamers to other amino acids. For example, Majerfeld and Yarus [38] recently reported that isoleucine aptamers contain isoleucine codons at their isoleucine-binding sites. Because isoleucine has an aliphatic sidechain, electrostatic interactions cannot be involved in this case. If the codon-binding-site relationship holds true for other amino acids, then it becomes likely that this intrinsic affinity limited the set of chemically accessible genetic codes. The application of statistical tests to further RNA–amino-acid interactions will ultimately indicate the relative importance of chance and necessity in determining the present form and content of the genetic code.

Acknowledgements

This research is supported in part by NSF grant MCB-9604377 to L.F.L., a Burroughs Welcome Fund New Investigator in Molecular Parasitology. We thank Jannette Carey, Michael Famulok, Stephen Freeland, Andrew Ellington and members of the Landweber Lab for comments and suggestions.

References

1. Woese, C.R. (1965). On the evolution of the genetic code. *Proc. Natl Acad. Sci. USA*, **54**, 1546-1552.
2. Sonneborn, T.M. (1965). Degeneracy of the genetic code: extent, nature, and genetic implications. In *Evolving Genes and Proteins.* (Bryson, V. & Vogel, H.J, eds), pp. 377-297, Academic Press, New York.
3. Zuckerkandl, E. & Pauling, L. (1965). Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins.* (Bryson, V. & Vogel, H.J, eds), pp. 97–166, Academic Press, New York.
4. Woese, C.R. (1967). *The Genetic Code: The Molecular Basis for Genetic Expression.* Harper & Row, New York.
5. Crick, F.H.C. (1968). The origin of the genetic code. *J. Mol. Biol.* **38**, 367-379.
6. Wong, J.T.-F. (1975). A co-evolution theory of the genetic code. *Proc. Natl Acad. Sci. USA*, **72**, 1909-1912.
7. Woese, C.R., Dugre, D.H., Saxinger, W.C. & Dugre, S.A. (1966). The molecular basis for the genetic code. *Proc. Natl Acad. Sci. USA*, **55**, 966-974.
8. Pelc, S.R. & Welton, M.G.E. (1966). Stereochemical relationship between coding triplets and amino acids. *Nature* **209**, 868-872.
9. Dunnill, P. (1966). Triplet nucleotide–amino acid pairing: a stereochemical basis for the division between protein and nonprotein amino acids. *Nature* **210**, 1267-1268.
10. Alberti, S. (1997). The origin of the genetic code and protein synthesis. *J. Mol. Evol.* **45**, 352-358.
11. Yarus, M. (1989). Specificity of arginine binding by the *Tetrahymena* intron. *Biochemistry* **28**, 980-988.
12. Hopfield, J.J. (1978). Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. *Proc. Natl Acad. Sci. USA* **75**, 4334-4338.
13. Shimizu, M. (1982). Molecular basis for the genetic code. *J. Mol. Evol.* **18**, 297-303.
14. de Duve, C. (1988). The second genetic code. *Nature* **333**, 117-118.
15. Noller, H.F. (1993), On the origin of the ribosome: coevolution of subdomains of tRNA and rRNA. In *The RNA World.* (Gesteland, R.F. & Atkins, J.F., eds), pp. 137-156, Cold Spring Harbor Laboratory Press, New York.
16. Schimmel, P., Giege, R., Moras, D. & Yokoyama, S. (1993). An operational genetic code for amino acids and possible relationship to genetic code. *Proc. Natl Acad. Sci. USA* **90**, 8763-8768.
17. Maizels, N. & Weiner, A.M. (1994). Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc. Natl Acad. Sci. USA* **91**, 6729-6734.
18. Dick, T.P. & Schamel, W.W.A. (1995). Molecular evolution of transfer RNA from two precursor hairpins: Implications for the origin of protein synthesis. *J. Mol. Evol.* **41**, 1-9.
19. Ellington, A.D. & Szostak, J.W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822.
20. Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510.
21. Robertson, D.L. & Joyce, G.F. (1990). Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467-468.
22. Connell, G.J., Illangsekare, M. & Yarus, M. (1993). Three small ribooligonucleotides with specific arginine sites. *Biochemistry* **32**, 5497-5502.
23. Connell, G.J. & Yarus, M. (1994). RNAs with dual specificity and dual RNAs with similar specificity. *Science* **264**, 1137-1141.

24. Yarus, M. (1993), An RNA-amino acid affinity. In *The RNA World.* (Gesteland, R.F. & Atkins, J.F., eds), pp. 205-217, Cold Spring Harbor Laboratory Press, New York.
25. Famulok, M. (1994). Molecular recognition of amino acids by RNA-aptamers: an L-citrulline binding RNA motif and its evolution into an L-arginine binder. *J. Am. Chem. Soc.* **116**, 1698-1706.
26. Tao, J. & Frankel, A.D. (1996). Arginine-binding RNAs resembling TAR identified by *in vitro* selection. *Biochemistry* **35**, 2229-2238.
27. Geiger, A., Burgstaller, P., von der Eltz, H., Roeder, A. & Famulok, M. (1996). RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. *Nucleic Acids Res.* **24**, 1029-1036.
28. Tao, J. & Frankel, A.D. (1992). Specific binding of arginine to TAR RNA. *Proc. Natl Acad. Sci. USA* **89**, 2723-2726.
29. Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, F. (1996). Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* **272**, 1343-1346.
30. Sokal, R.R. & Rohlf, F.J. (1995). *Biometry: the Principles and Practice of Statistics in Biological Research.* (3rd edn). W.H. Freeman and Company, New York.
31. Jiang, F., Kumar, R.A., A, J.R. & Patel, D.J. (1996). Structural basis of RNA folding and recognition in an AMP–RNA aptamer complex. *Nature* **382**, 183-186.
32. Fan, P., Suri, A.K., Fiala, R., Live, D. & Patel, D.J. (1996). Molecular recognition in the FMN–RNA aptamer complex. *J. Mol. Biol.* **258**, 480-500.
33. Burgstaller, P., Kochoyan, M. & Famulok, M. (1995). Structural probing and damage selection of citrulline- and arginine-specific RNA aptamers identify base positions required for binding. *Nucleic Acids Res.* **23**, 4769-4776.
34. Jiang, L., Suri, A.K., Fiala, R. & Patel, D.J. (1997). Saccharide–RNA recognition in an aminoglycoside antibiotic–RNA aptamer complex. *Chem. Biol.* **4**, 35-50.
35. Zimmermann, G.R., Shields, T.P., Jenison, R.D., Wick, C.L. & Pardi, A. (1998). A semiconserved residue inhibits complex formation by stabilizing interactions in the free state of a theophylline-binding RNA. *Biochemistry* **37**, 9186-9192.
36. Yarus, M. (1998). Amino acids as RNA ligands: a direct-RNA-template theory for the code's origin. *J. Mol. Evol.* **47**, 109-117.
37. Majerfeld, I. & Yarus, M. (1994). An RNA pocket for an aliphatic hydrophobe. *Nat. Struct. Biol.* **1**, 287-292.
38. Majerfeld, I. & Yarus, M. (1998). Isoleucine:RNA sites with essential coding sequences. *RNA* **4**, 471-478.
39. Zinnen, S. & Yarus, M. (1995). An RNA pocket for the planar aromatic sidechains of phenylalanine and tryptophane. *Nucleic Acids Symp. Ser.* **33**, 148-151.
40. Jukes, T.H. (1973). Arginine as an evolutionary intruder into protein synthesis. *Biochem. Biophy. Res. Commun.* **53**, 709-714.
41. Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science* **185**, 862-865.