Quantitative Analysis of Receptors for Adenosine Nucleotides Obtained via In Vitro Selection from a Library Incorporating a Cationic Nucleotide Analog

Thomas R. Battersby,[†] Darwin N. Ang,[‡] Petra Burgstaller,[§,||] Simona C. Jurczyk,[‡,⊥] Michael T. Bowser,[||] Danielle D. Buchanan,[||] Robert T. Kennedy,[||] and Steven A. Benner*[,‡,||]

Contribution from the Department of Chemistry, Department of Anatomy and Cell Biology, and the Florida Center for Heterocyclic Compounds, University of Florida, Gainesville, Florida 32611, and EraGen, Inc. 12085 Research Drive, Alachua, Florida 32615

Received May 11, 1998. Revised Manuscript Received November 16, 1998

Abstract: 5-(3′′-Aminopropynyl)-2′-deoxyuridine (dJ), a modified nucleoside with a side chain carrying a cationic functional group, was incorporated into an oligonucleotide library, which was amplified using the Vent DNA polymerase in a polymerase chain reaction (PCR). When coupled to an in vitro selection procedure, PCR amplification generated receptors that bind ATP. This is the first example of an in vitro selection generating oligonucleotide receptors where the oligonucleotide library has incorporated a cationic nucleotide functionality. The selection yielded functionalized receptors having sequences differing from a motif known to arise in a standard selection experiment using only natural nucleotides. Surprisingly, both the natural and the functionalized motifs convergently evolved to bind not one, but two ATP molecules cooperatively. Likewise, the affinity of the receptors for ATP had converged; in both cases, the receptors are half saturated at the 3 mM concentrations of ATP presented during the selection. The convergence of phenotype suggests that the outcome of this selection experiment was determined by features of the environment during which selection occurs, in particular, a highly loaded affinity resin used in the selection step. Further, the convergence of phenotype suggests that the optimal molecular phenotype has been achieved by both selections for the selection conditions. This interplay between environmental conditions demanding a function of a biopolymer and the ability of the biopolymer to deliver that function is strictly analogous to that observed during natural selection, illustrating the nature of life as a self-sustaining chemical system capable of Darwinian evolution.

Introduction

In vitro selection is a combinatorial method that generates receptors, ligands, and catalysts from nucleic acid libraries containing as many as 10^{15} different molecules.[1] It exploits a "selection" procedure to separate RNA or DNA molecules with specific binding or catalytic properties from those lacking these properties, the polymerase chain reaction (PCR) to amplify single selected RNA or DNA molecules to give a large number of their descendants, and mutation to allow the descendent molecules to "evolve", improving their binding or catalytic activities.

In addition to its technological value,[2] in vitro selection offers the opportunity to ask general questions about the relation between structure and function in organic molecules. As with any polymer built from a limited set of building blocks, "structure space" in nucleic acids is well defined, with a countable number of possible sequences. This allows us to quantify the probability of finding a receptor or catalyst meeting certain specifications within that space. In many cases, the probability is low, especially when compared with the structure space defined by the 20 amino acid building blocks of polypeptides.[3] For example, to have a 50% chance of holding a single RNA molecule able to catalyze a template-directed ligation reaction by a modest (by protein standards) factor of 10^4, a library of RNA molecules 220 nucleotides in length must contain $\sim 2 \times 10^{13}$ random sequences.[4]

Natural evolution in the biomolecular world is also described as the search for function within a biopolymeric sequence space.[5] As with natural evolutionary experiments, the outcome of an in vitro selection experiment reflects many parameters. These include the details of the selection conditions, the number of selection cycles performed (and hence the amount of structure space explored), and the "searchability" of structure space, an abstract concept that depends on the "ruggedness" of the landscape that relates fitness to structure. Ruggedness is determined by the structure and structural versatility of the biopolymer whose sequence space is being searched.

* To whom correspondence should be addressed at the Department of Chemistry, University of Florida.

[†] EraGen, Inc.

[‡] Department of Anatomy and Cell Biology, University of Florida.

[§] Present address: Aventis Research & Technologies GmbH & Co. KG, 65926 Frankfurt-am-Main, Germany.

[||] Department of Chemistry, University of Florida.

[⊥] Florida Center for Heterocyclic Compounds, University of Florida.

(1) (a) Szostak, J. W. In *Redesigning the Molecules of Life*; Benner, S. A., Ed.; Springer-Verlag: Heidelberg, Germany, 1988; pp 87−114. (b) Joyce, G. F. *Gene* **1989**, *82*, 83−87. (c) Joyce, G. F. *UCLA Symp. Mol. Cell. Biol. New Ser.* **1989**, *94*, 361−371. (d) Irvine, D.; Tuerk, C.; Gold, L. *J. Mol. Biol.* **1991**, *222*, 739−761.

(2) For some examples, see (a) Wiegand, T. W.; Janssen, R. C.; Eaton, B. E. *Chem. Biol.* **1997**, *4*, 675−683. (b) Zhang, B. L.; Cech, T. R. *Nature* **1997**, *390*, 96−100.

(3) Johnsson, K.; Allemann, R. K.; Widmer, H.; Benner, S. A. *Nature* **1993**, *365*, 530−532.

(4) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411−1418.

(5) (a) Benner, S. A.; Ellington, A. D. *CRC Crit. Rev. Biochem.* **1988**, *23*, 369−426. (b) Benner, S. A. *Chem. Rev.* **1989**, *89*, 789−806.

10.1021/ja9816436 CCC: $18.00
Published on Web 10/13/1999

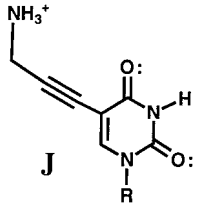

Figure 1. 5-(3″-Aminopropynyl)-2′-deoxyuridine ("dJ", R= 2′-deoxyribose), a modified nucleoside with a side chain carrying a cationic functional group.

One approach to understand the interaction between these parameters involves altering the nature of the building blocks used by the biopolymer, incorporating nonstandard[6] or functionalized[7] derivatives of nucleic acids in in vitro selection experiments. Technical obstacles hinder this approach. Even modest discrimination by a polymerase against an unnatural functionalized nucleotide may lead to a selective loss of the functionality during PCR amplification. While polymerases are known to incorporate nucleotides carrying neutral functionality[8] and two in vitro selection experiments have incorporated such nucleotides,[9] it has proven especially difficult to amplify oligonucleotides containing nucleotide analogues carrying positively charged functionality, the functionality most notably missing from standard DNA and RNA.

We report here the first example of incorporation of a positively charged functional group into an in vitro selection experiment and the first quantitative comparison of a selection experiment with standard biopolymers and functionalized biopolymers of any kind. The results provide a remarkable example of convergent evolution of physical behavior in two molecular systems, as well as a view of the landscapes that relate structure and behavior in macromolecules.

Results

5-(3″-Aminopropynyl)-2′-deoxyuridine[10] (Figure 1), trivially designated as "dJ", was prepared by coupling *N*-propynyltrifluoroacetamide[11,12] and 2′-deoxy-5-iodouridine with the aid of a palladium catalyst and converted to the corresponding triphosphate following a procedure of Ludwig and Eckstein.[13] The side chain ammonium ion, with an expected pK_a of ~9.5, carries a positive charge at physiological pH. In preliminary experiments, dJTP was examined as a substrate for a range of thermostable polymerases. As has been observed with previous experiments with nonstandard bases,[14] different polymerases behaved quite differently with respect to the unnatural modification. Taq polymerase, for example, incorporates dATP opposite dJ in the template but stops rather than incorporating dJTP opposite dA in a template. Tth and Tfl polymerases pause when encountering dJ both in a template and as a triphosphate. Vent polymerase, however, incorporates both dJTP opposite dA and dATP opposite dJ with only modest amounts of pausing. Vent polymerase was therefore chosen as a tool for incorporating dJ into a PCR experiment.

PCR experiments were then run to show that Vent is able to amplify oligonucleotides containing the positively charged nucleobase dJ. Parallel PCR experiments were run with natural deoxynucleoside triphosphate mixes and mixes having TTP replaced by dJTP (data not shown). In both cases, amplification was seen, with the dJTP supporting PCR amplification only slightly less well than TTP with the Vent polymerase.

With Vent polymerase enabling PCR amplification with dJTP, two in vitro selection experiments were run in parallel to select for receptors that bind the adenosine derivatives ATP, ADP, and AMP, one using a library built from natural nucleotides T, dA, dC, and dG (and therefore exploring "natural structure space") and the other using a library built from dJ, dA, dC, and dG (therefore exploring "unnatural structure space"). These experiments followed closely the recipe of Huizenga and Szostak, who did this selection with natural nucleotides some time ago.[15] Nine rounds of selection, each separated by 10−14 cycles of amplification, were performed by passing successively enriched libraries through a column with immobilized ATP, ADP, and AMP. Products were cloned (with T replacing dJ), and a set of clones was sequenced.

A total of 17 sequences were recovered from the in vitro selection experiment performed with standard nucleotides (Figure 2). All appeared to be derived from different ancestors. A total of 39 sequences were recovered from the dJ library (Figure 2). These represented 30 ancestral sequences, with sequences 401 and 409; 4011, 4013, and 4014; 504, 505, 509, 5010, and 5015; 602 and 609; and 6010 and 6015 forming five individual families.

The randomized sequences obtained from the selection with natural nucleotides were approximately 69 nucleotides in length, virtually identical to the length of the initial random pool. In contrast, the average length of the randomized region resulting from selection with the dJ library was only 25 nucleotides. Further, the natural library yielded randomized regions containing 25% T, while the dJ library yielded randomized regions containing 14% dJ. These results suggest that Vent, although satisfactory as a polymerase for performing these experiments, nevertheless selects against dJ and that further work developing new polymerases that incorporate functionalized and nonstandard nucleotides is warranted. Work to improve polymerases is ongoing in these laboratories.[16]

The sequences were then examined to see whether they resembled the sequences reported by Huizenga and Szostak, who also selected a DNA molecule that could bind to adenosine derivatives.[15] Their paper does not report a sample of the

(6) (a) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322−8323. (b) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33−37. (c) Krosigk, U.; von, Benner, S. A. *J. Am. Chem. Soc.* **1995**, *117*, 5361−5362. (d) Vögel, J. J.; Benner, S. A. *J. Am. Chem. Soc.* **1994**, *116*, 6929−6930. (e) Vögel, J. J.; Altorfer, M. M.; Benner, S. A. *Helv. Chim. Acta* **1993**, *76*, 2061−2069. (f) Vögel, J. J.; von Krosigk, U.; Benner, S. A. *J. Org. Chem.* **1993**, *58*, 7542−7547.

(7) (a) Heeb, N. V.; Benner, S. A. *Tetrahedron Lett.* **1994**, *35*, 3045−3048. (b) Robertson, M. P.; Miller, S. L. *Science* **1995**, *268*, 702−705. (c) Dewey, T. M.; Zyzniewski, M. C.; Eaton, B. E. *Nucleosides Nucleotides* **1996**, *15*, 1611−1617. (d) Kodra, J.; Benner, S. A. *Synth. Lett.* **1997**, 939−940.

(8) (a) Langer, P. R.; Waldrop, A. A.; Ward, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 6633−6637. (b) Prober, J. M.; Trainor, G. L.; Dam, R. J.; Hobbs, F. W.; Robertson, C. W.; Zagursky, R. J.; Cocuzza, A. J.; Jensen, M. A.; Baumeister, K. *Science* **1987**, *238*, 336−341.

(9) (a) Latham, J. A.; Johnson, R.; Toole, J. J. *Nucleic Acids Res.* **1994**, *22*, 2817−2822. (b) Tarasow, T. M.; Tarasow, S. L.; Eaton, B. E. *Nature* **1997**, *389*, 54−57.

(10) (a) Hobbs, F. W., Jr. *J. Org. Chem.* **1989**, *54*, 3420−3422. (b) Casalnuovo, A. L.; Calabrese, J. C. *J. Am. Chem. Soc.* **1990**, *112*, 4324−4330.

(11) Cruickshank, K. A.; Stockwell, D. L. *Tetrahedron Lett.* **1988**, *29*, 5221−5224.

(12) Meyer, R. B., Jr. in *Protocols for Oligonucleotide Conjugates: Synthesis and Analytical Techniques*; Agrawal, S., Ed.; Humana Press: Totowa, New Jersey; pp 77−80.

(13) Ludwig, J.; Eckstein, F. *J. Org. Chem.* **1989**, *54*, 631−635.

(14) (a) Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329−6333. (b) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hübscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308−1313.

(15) Huizenga, D. E.; Szostak, J. W. *Biochemistry* **1995**, *34*, 656−665.

(16) (a) Lutz, M.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 499−504. (b) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1149−1152.

Sequences recovered from selection of standard DNA to bind to ADP

201 TCTAGAGTATGCGGTAGTGTTAGTTGGAGTATTTCCGGGGTTGGGGTGAGGAGGAAATATTCCGGTCGGGTCATTGTGTTGCTCGGCCCGAAGAAGCCT
202 TCTAGAGTATGCGGTAGCGCGGGAACGTTGCCGGGGGAGCTGTGTCTGCGCGGAGGAGGCGCTTGGAGTTCCCGGCTGTTTGGTTATCCGAAGAAGCTT
203 TCTAGAGTATGCGGTAGAAGCCCCNTCTNGAGCGGGATGANCATGNTGCAGCATGATGTAGTCTGTNTATNNNNGTGGGAATACNCNTGACCGAAGAAGCTT
204 TCTAGAGTATGCGGTAGTGGCAAGCTGTCTGTTGGAGAGCGGGGGAGCAGTCGCGGATGAGCGTAAACAATGTGTAGGCGGCTTGGTCCGAAGAAGCTT
206 TCTAGAGTATGCGGTAGCTGGGTATCTTGTGACGACTAAATTCCGGGGGAGGGCGCACATGCACGGCGGAGGTATTTTGTGTGGCAACCGAAGAAGCTT
207 TCTAGAGTATGCGGTAGACTGGCATGTGACGCGGGGGAAGGCCTGCCGGGTCTGGGGGAGCAGGTGCGTGTGGTCTGATGTTGATTGCCGAAGAAGCTT
208 TCTAGAGTATGCGGTAGTTCGTCCCTGCACCGAAGAA(G)CTT
209 TCTAGAGTATGCGTTAGTTACGGCGGAGCTTGTTCATCTANGCGGGGGAGTGTTCCGNTTACTACGTGTCTTTGGCATTAGCTTTACCCGAAGAAGCTT
2010 TCTAGAGTATGCGGTAGCCCGCTGGGAAGCCCCGGGGGATGTTAAGGAGGAGGGGCGATCCCTACGGATTGCCGGCTAACCACGCACCCGAAGAAGCTT

Sequences recovered from selection of standard DNA to bind to ATP

301 TCTAGA-GTATGCGGTAGCGGGGGAGCGTGATGTCAATACCATCTGCGGGGGAGTCTGGCTTACTTAGCGACTAGTGTCCGTGGGTCTCCGAAGAAGCTT
302 TCTAGA-GTATGCGGTAGAATTGGACATAATGCGGGGGAGTGTGCGGCGGAGCATGTGTTCAGGTTGCGTGATGTAGACATTTGGGGACCGAAGAAGCTT
304 TCTAGA-GTATGCGGTAGACGGTCAGGTCTCCGTTAGACGTCCATGCGGAAGGAGTGTAATGGGGGGAGTCATAAGTGGGTATAGTGGCCGAAGAAGCTT
305 TCTAGA-GTATGCGGTAGGCGTGTTAGGGATATGCCCCGACAGAGGGAGGAGTTTCACGGGGGACGATGCCGTGGGTTCTTTCCCGTCCGAAGAAGCTT
306 TCTAGAAGTATGCGGTAGCTCGCACGATGCTTTAAATTAGCCGGGGGAGTTGTTTATGGAGGAGCGCTGATTTAATACTGGGCTAAACCCGAAGAAGCTT
307 TCTAGA-GTATGCGGTAGTGTCCGAAGAA(G)CTT
309 TCTAGA-GTATGCGGTAGAGAGGGCCATCAGTGGTCCAGTGGCGGGGGAGGCTTAAACCGGAGGAGCTCACATCTAAAGACTGCTTTACCGAAGAAGCTT
3010 TCTAGA-GTATGCGGTAGATGTTCTCGGGGGAGTGTATGCGGGGGTGAGAAGGCTATGCGTATTGCGGTAACGGCTTATGGGTGGCCACCGAAGAAGCTT

Sequences recovered from selection of dJ-functionalized DNA to bind to AMP

401 GGTCGTCTAGAGTATGCGGTAGGAACGJCAGJGGGGGGAGCAJAJGAJGJGAJACGCGACCCGAAGAAGCJJGGCCCAJG
402 GGTCGTCTAGAGTATGCGGTAGGAGJGCCAACGGGAGGAGCGJGACCGAAGAAGCJJGGCCCAJG
403 GGTCGTCTAGAGTATGCGGTAGJJAJGGGGAGJACAGGGGACGGCGGAAJACCGAAGAAGCJJGGCCCAJG
404 GGTCGTCTAGAGTATGCGGTAGAJCGGGGGGAGGCACGGGGGGAGAJGJACCGAAGAAGCJJGGCCCAJG
405 GGTCGTCTAGAGTATGCGGTAGJJACJCGJGCGCJGACCGAAGAAGCJJGGCCCAJG
406 GGTCGTCTAGAGTATGCGGTAGGAGCAGGGCGGGGGGAGCACACJCCGAAGAAGCJJGGCCCAJG
407 GGTCGTCTAGAGTATGCGGTAGGJGCAGGGCGGGNNGAGCACNJCCGAAGAAGCJJGGCCCAJG
408 GGTCGTCTAGAGTATGCGGTAGGAGJJCACGGGGGGAGCAJGGAJCGCCGAAGAAGCJJGGCCCAJG
409 GGTCGTCTAGAGTATGCGGTAGGAACGJCAGJGGGGGGAGCAJAJGGJGJGAJACGCGACCCGAAGAAGCJJGGCCCAJG
4010 GGTCGTCTAGAGTATGCGGTAGGAGGJACGGGGGGAGCJAJACCGAAGAAGCJJGGCCCAJG
4011 GGTCGTCTAGAGTATGCGGTAGJCACGGAGGAJGJAAGGGGGAGJGAGACGJACCGAAGAAGCJJGGCCCAJG
4012 GGTCGTCTAGAGTATGCGGTAGGCGCGGAGGAAJJGAJAGJGGGGGAGCGCJGAJCCGAAGAAGCJJGGCCCAJG
4013 GGTCGTCTAGAGTATGCGGTAGJCACGGAGGAJGJAAGGGGGAGJGAGACGJACCGAAGAAGCJJGGCCCAJG
4014 GGTCGTCTAGAGTATGCGGTAGJCACGGAGGAJGJAAGGGGGAGJGAGACGJACCGAAGAAGCJJGGCCCAJG
4015 GGTCGTCTAGAGTATGCGGTAGJGGGGGGAJGJJAGCGJAGNJAGGCGAJGJGCCJGJJGGCCGAAGAAGCJJGGCCCAJG

Sequences recovered from selection of dJ-functionalized DNA to bind to ADP

502 GGTCGTCTAGAGTATGCGGTAGCGGCJGGJGCCCGAAGAAGCJJGGCCCAJG
503 GGTCGTCTAGAGTATGCGGTAGGAJGAGAJGGGJAGGGACCGAAGAAGCJJGGCCCAJG
504 GGTCGTCTAGAGTATGCGGTAGGGAGGACGGGGACGGGGGGGAGACCGCACCGAAGAAGCJJGGCCCAJG
505 GGTCGTCTAGAGTATGCGGTAGGGAGGACGGACGGGGGAGAGACCGCACCGAAGAAGCJJGGCCCAJG
506 GGTCGTCTAGAGTATGCGGTAGAACGJJGJGGGGGAGJAJACCCGAAGAAGCJJGGCCCAJG
509 GGTCGTCTAGAGTATGCGGTAGGGAGGACGGGGACGGGGGGGAGACCGCACGAAGAAGCJJGGCCCAJG
5010 GGTCGTCTAGAGTATGCGGTAGGGAGGACGGGGACGGGGGGGAGACCGCACCGAAGAAGCJJGGCCCAJG
5011 GGTCGTCTAGAGTATGCGGTAGGJACCGAAGAAGCJJGGCCCAJG
5012 GGTCGTCTAGAGTATGCGGTAGGGGGGGGCGGGACCGAAGAAGCJJGGCCCAJG
5013 GGTCGTCTAGAGTATGCGGTAGAGCGJJCGCACGCGGGGGAGCAGAJJCJAGCCGAAGAAGCJJGGCCCAJG
5014 GGTCGTCTAGAGTATGCGGTAGGCJAGNCCGJCCGAAGAAGCJJGGCCCAJG
5015 GGTCGTCTAGAGTATGCGGTAGGGAGGACGGGGACGGGGGGGAGACCGCACCGAAGAAGCJJGGCCCAJG

Sequences recovered from selection of dJ-functionalized DNA to bind to ATP

602 GGTCGTCTAGAGTATGCGGTGGAGCGAJAGCGGGGGAGGAJACACCAJGGJGACCGAANAAGCJJGGCCCAJG
604 GGTCGTCTAGAGTATGCGGTAGNGGCGJJACNCAGJGACJGJCACGAAGAANCJJGGCCCAJG
607 GGTCGTCTAGAGTATGCGGTAGGAGGGACGGGGGGAGCAJCGCJGCAJGJJCCGAAGAAGCJJGGCCCAJG
608 GGTCGTCTAGAGTATGCGGTAGJGCAGJGGGGGAGGAACGGCGGAAJGGCACCGAAGAACCJGGCCCAJG
609 GGTCGTCTAGAGTATGCGGTAGAGCGGJAGCGGGGGAGGAJACACCAJGGJGACCGAAGAAGCJJGGCCCAJG
6010 GGTCGTCTAGAGTATGCGGTAGGAGCCAJGCGGGGGGAGCAJAACCGJCAGACCGAAGAAGCJJGGCCCAJG
6011a GGTCGTCTAGAGTATGCGGTAGAGJCGGJGACJCJJAJGJGGJCGACCGAAGAAGCJJGGCCCAJG
6011b GGTCGTCTAGAGTATGCGGTAGGAGCGJGCGGGGGGAGCAJAJJJAGAGGCGJCCGAAGAAGCJJGGCCCAJG
6012 GGTCGTCTAGAGTATGCGGTAGGAGGGGGCGGGAGGAGCAJJJCJGGGJGCCGAAGAAGCJJGGCCCAJG
6013 GGTCGTCTAGAGTATGCGGTAGGAGGGGCCCGGGGGAGCAJAJGJACCGAAGAAGCJJGGCCCAJG
6014 GGTCGTCTAGAGTATGCGGTAGGAGCCAJGCGGGGGGAGCAJAJJJGCAGGCGJCCGAAGAAGCJJGGCCCAJG
6015 GGTCGTCTAGAGTATGCGGTAGGAGCCAJGCGGGGGGAGCGJAACCGJCAGACCGAAGAAGCJJGGCCCAJG

Figure 2. Sequences generated from in vitro selection starting from a natural library and from a library containing dJ. Underlined regions are derived from the primer or primer binding regions.

sequences generated by the selection directly. Huizenga and Szostak did, however, report a single 42mer sequence derived from the selection, generated 43 unique sequences from it using mutation-prone PCR followed by a selection, and analyzed these for evidence of a conserved motif among those that continued to bind adenosine derivatives. Lin and Patel examined by NMR the solution structure of a 25mer designed from the sequences obtained by mutation-prone PCR, and to which had been added an additional base pair on the stem;[17] the structure is shown schematically in Figure 3.

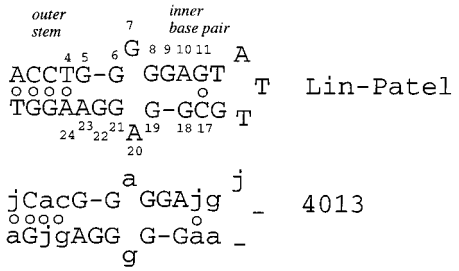

Figure 3. A schematic illustrating the fold determined by Lin and Patel[17] for a DNA molecule prepared by Huizenga and Szostak,[15] and how sequence 4013, selected from the dJ-functionalized DNA library for binding to AMP, fits. Conserved Watson−Crick pairing is indicated by o. Nucleotides present in the Lin-Patel structure are in upper case; nucleotides different in Sequence 4013 are in lower case.

From this combination of experiments, including both the structure and the pattern of sequence conservation, we defined a "Lin−Patel−Huizenga−Szostak motif" in terms of the following features: i. G5, A10, G18, and A23 are absolutely invariant and were conserved in the Huizenga−Szostak mutagenesis studies for effective receptors. ii. G6, G8, G9, G19, G21, and G22 are nearly always guanosine. In the Huizenga-Szostak mutagenesis studies, only 12 of the 264 positions sampled failed to have a G at these positions, and all of these substituted G by A. The NMR structure shows that G6-G21 and G8-G19 form two reverse Hoogsteen base pairs. G9 and G22 bind to the AMP ligand in the Lin−Patel structure. iii. The bases adjacent to G5 and A23 always form a Watson−Crick base pair (or, in some cases, a G−T wobble). In the Huizenga−Szostak mutagenesis studies, only 4 of the 44 examples sampled were these positions not Watson−Crick complementary and, of these, three were GT pairs. These features were used to define the Lin-Patel-Huizenga-Szostak motif. In addition, several sequence features frequently were found. iv. G7 and A20 are usually purines, with one position G and the other A. In the Huizenga−Szostak mutagenesis studies, only 2 of the 88 positions mutated to pyrimidines. v. The bases at positions A10 and G18, first position of the inner loop, can usually form a Watson−Crick base pair, or a GT wobble. In the Huizenga−Szostak mutagenesis studies, 8 of the 44 examples had a mismatch at this position; of these, three were GT pairs. In the NMR structure, a Watson−Crick base pair was formed. vi. The outer stem frequently could form two, and sometimes more, Watson−Crick pairs. In the Huizenga−Szostak mutagenesis studies, 25 of the 44 sequences had three or more of these. Of the remainder, five had the potential for forming two Watson−Crick pairs plus one GT pair, 10 could form Watson−Crick pairs in two of the three positions, while four could form one Watson−Crick pair plus one GT wobble. Huizenga and Szostak note that one base pair was sufficient to detect ligand interaction on the ATP column, but three potential base pairs were required for optimal affinity.

Of the 17 sequences obtained from the unfunctionalized library in this work, sequences 202, 204, 206, 207, 209, 2010, 301, 302, 305, 306, 309, and 3010 (12 in total, 71%) met these criteria to an acceptable degree (Figure 4). The remaining sequences (201, 203, 208, 304, and 307) from the standard selection did not obviously fit the Lin−Patel−Huizenga−Szostak motif. Sequence 204 has a T at a position requiring a G; sequence 3010 is missing the A at position 23.

A similar analysis was then applied to the sequences selected from the dJ library. This analysis suggested that the Lin−Patel−Huizenga−Szostak fold was likely to be adopted by five of these (Figure 5) (sequences 4011, 4012, 4013, 4014, and 608). Sequences 4011, 4013, and 4014 are identical and presumably arose from a single ancestral sequence in the original library. They are 31 nucleotides long, and share 31 nucleotides in sequence. As there are $\sim 4 \times 10^{18}$ possible 31mers, it is unlikely that these three sequences were found three times in the initial population. Sequence 4012 apparently arose independently. All presumably adopt the Lin−Patel−Huizenga−Szostak fold. They all contain the two G consensus regions, and all have the potential to form the inner and outer Watson−Crick base pairs required by the fold. In addition, sequence 404 may possibly adopt a Lin−Patel−Huizenga−Szostak fold, even though a part of the second G-rich region is not canonical and it contains a non-GT stem mismatch in the outer stem positions closest to the conserved regions (not observed in any of the Huizenga−Szostak binders). An inner Watson−Crick pair is possible, as is an outer stem joined by three external Watson−Crick base pairs (Figure 5).

Most of the sequences arising from the dJ selection do not obviously fit the Lin−Patel−Huizenga−Szostak motif. Many do, however, display features in common, whether they are obtained for binding to AMP, ADP, or ATP (Figure 6). A new motif can be defined relative to these features as having (i) a stretch of six G's followed by an A, (ii) a preceding GGA, (iii) the potential for Watson−Crick base pairs in inside stems just after the GGA and immediately before the GGGGGGA sequences, and (iv) the potential for Watson−Crick base pairs in outside stems four nucleotides before the GGA and immediately following the GGGGGGA sequence.

When selected for affinity to ADP, the dJ-library also yielded a receptor entirely lacking dJ in the randomized region and containing only three dJ-nucleotides in the segment of the receptor arising through copying of the 3′-primer binding site. It is represented by sequences 504, 505, 509, 5010, and 5015. All five of these sequences appear to have arisen from a single ancestor, as they are 28 nucleotides long and differ by a single nucleotide substitution (Figure 7). This sequence does not fit well the Lin−Patel−Huizenga−Szostak motif. It has three G-rich segments, not the two found in the Patel−Huizenga−Szostak motif. Further, the sequence lacks both the standard internal Watson−Crick potential pair as well as the external pairing that might generate the outer stem. This may be an as-yet undescribed motif for a DNA receptor for adenosine derivatives.

To explore the affinity of the natural and functionalized receptors, two sequences containing a representative Lin−Patel−Huizenga−Szostak motif (CTACCTGGGGGAGCATTGGGGAGGAAGGTAGCCGTGCGAAAA, Sequence A and GTGCTTGGGGGAGTATTGCGGAGGAAAGCGGCCCTGCTGAAG, Sequence B) and a sequence containing dJ (Sequence 409, Figure 2) were synthesized with 3′-fluorescent tags and their binding to ATP examined. Huizenga and Szostak reported that DNA molecules having the motif contained in Sequences A and B bound adenosine with an equilibrium disassociation constant (K_d) of $\sim$6−8 μM. We re-examined the binding of ATP to Sequence A and Sequence B using affinity capillary electrophoresis,[18] and found a sigmoidal binding curve characteristic of a 1:2 complex in both cases (Figure 8). No evidence could be found for a significant contribution of the 1:1 complex between the receptor and ATP at thermodynamic equilibrium. A termolecular complex was also observed by Lin and Patel in the NMR structure of the Huizenga−Szostak receptor with

(17) Lin, C. H.; Patel, D. J. *Chem. Biol.* **1997**, *4*, 817−832.

(18) Chu, Y. H.; Avila, L. Z.; Gao, J. M.; Whitesides, G. M. *Acc. Chem. Res.* **1995**, *28*, 461−468.

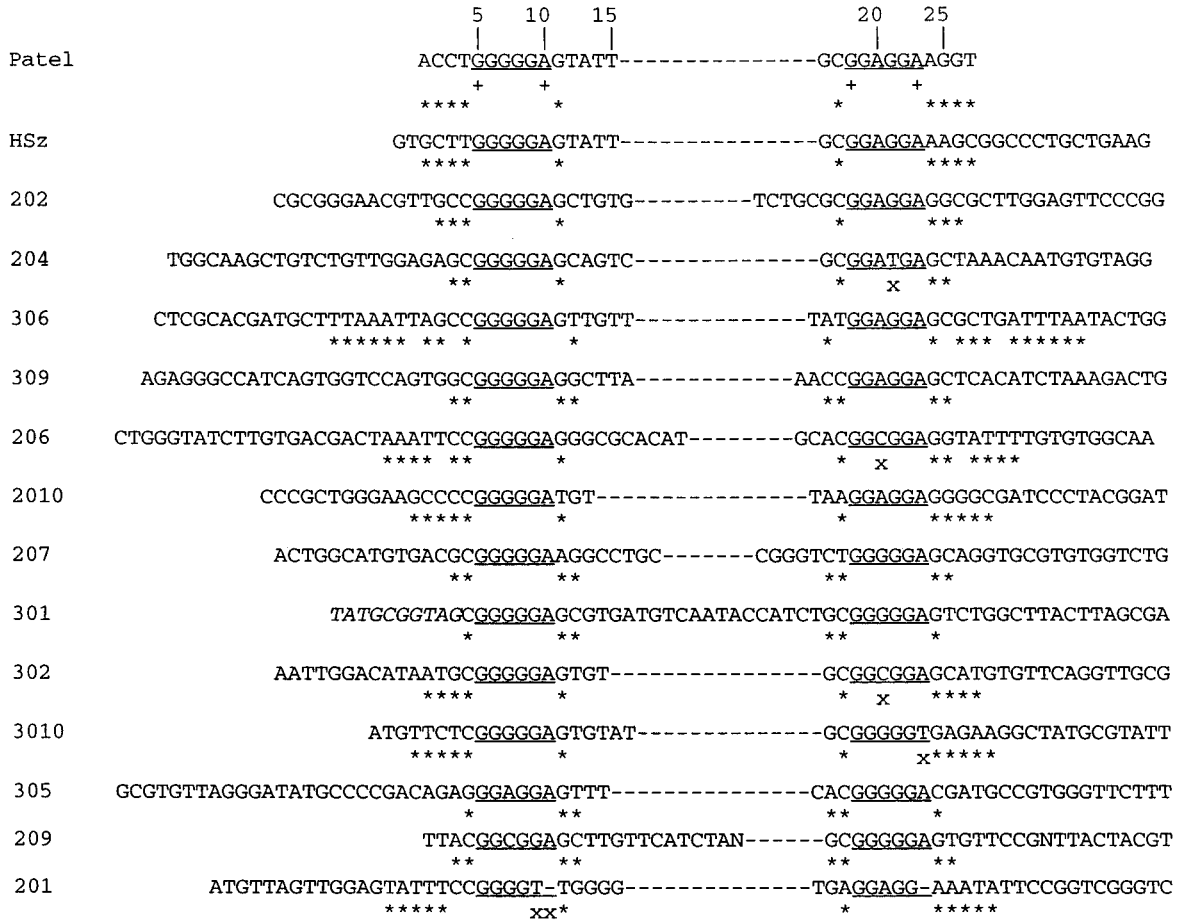

Figure 4. Analysis of the sequences obtained from the unfunctionalized library to identify sequences that fit the Lin−Patel−Huizenga−Szostak motif.[15,17] Positions above + are absolutely conserved in the Lin−Patel−Huizenga−Szostak motif. Underlined regions are the G-rich segments that are largely conserved. Positions above x violate the motif. Segments forming in the inner and outer stem base pairs are indicated by a * below the sequence. Italicized regions are derived from primer or primer-binding regions.

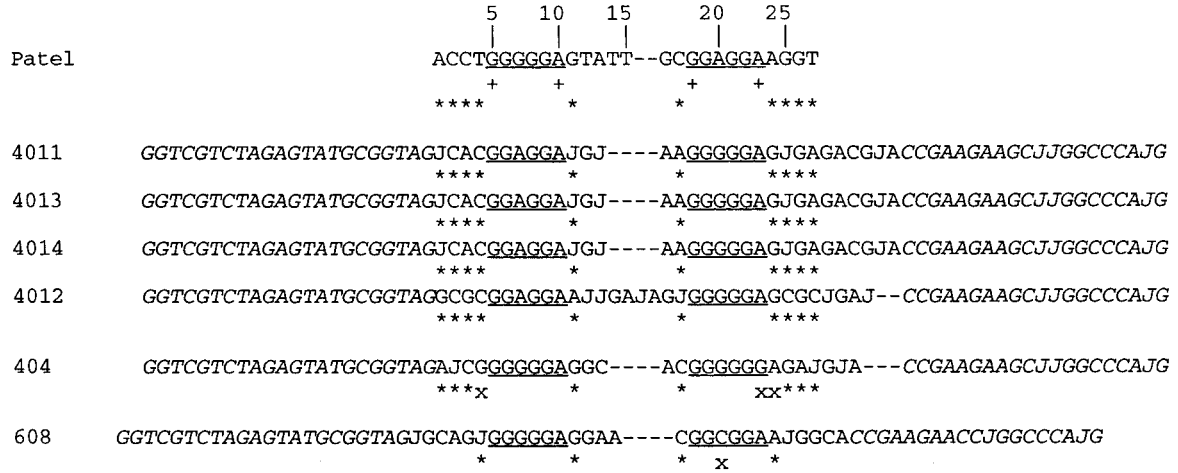

Figure 5. A comparison of the Lin−Patel−Huizenga−Szostak motif with sequences from the dJ selection that might display some of its features. Underlined regions are the G-rich motifs. Segments forming in the inner and outer stem base pairs are indicated by a * below the sequence. Positions above x violate the motif (note especially the violations in sequence 404). Italicized regions are derived from primer or primer-binding regions. Sequences 4011, 4013, and 4014 are identical and appear to be derived from a single sequence in the pool.

AMP.[17] These authors also found no concentration of AMP where a bimolecular complex could be observed.

This termolecular complex implies that the equilibrium constants describing the binding of adenosine derivatives to Sequence A and Sequence B must have units of M^2, not M. Fitting of the binding data using the method of Bowser and Chen[19] gave affinity constants of $9 \pm 2 \times 10^{-6}$ and $15 \pm 2 \times 10^{-6}$ M^2 (note the units) for Sequence A and Sequence B, respectively, different from the $(6-8) \times 10^{-6}$ M (note the units) disassociation constant reported for adenosine in the literature.[15,20]

Remarkably, the dJ-containing Sequence 409 also forms a termolecular complex with ATP, as shown by a sigmoidal binding curve (Figure 8). This implies that the equilibrium constant must again have units of M^2, not M. Again, no

(19) Bowser, M. T.; Chen, D. D. Y. *Anal. Chem.* **1998**, *70*, 3261−3270.

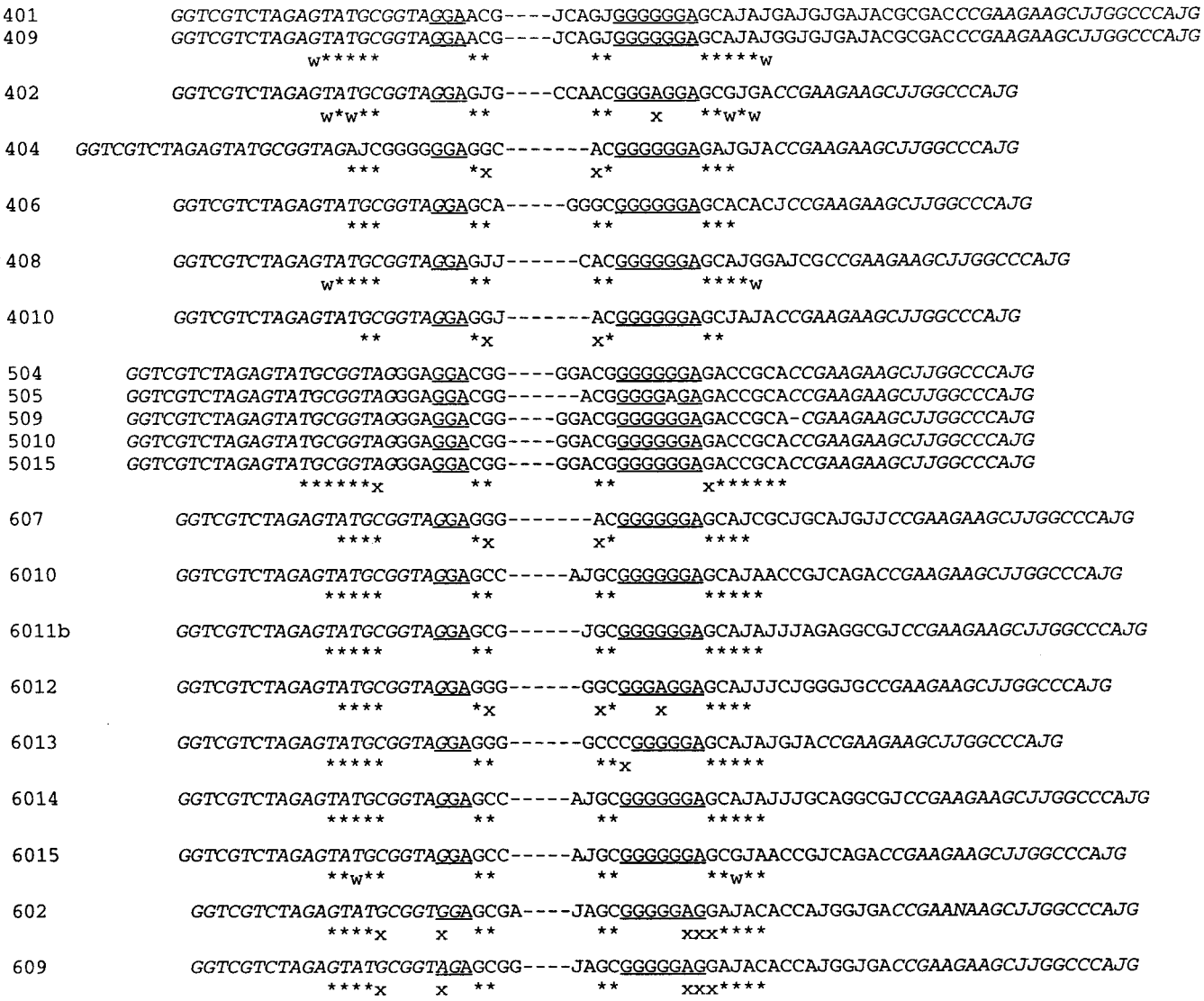

Figure 6. Most of the sequences arising from the dJ selection do not obviously fit the Lin−Patel−Huizenga−Szostak motif. They do, however, conform to a new motif, defined as having (i) a stretch of six G's followed by an A, (ii) a preceding GGA, (iii) the potential for Watson−Crick base pairs in inside stems just after the GGA and immediately before the GGGGGGA sequences, and (iv) the potential for Watson−Crick base pairs in outside stems four nucleotides before the GGA and immediately following the GGGGGGA sequence. Italicized regions are derived from primer or primer-binding regions. Segments having the potential for Watson−Crick complementarity are indicated by a * below the sequence. A G−T wobble is indicated by a w. Sequences presumed to be derived from a single ancestral sequence in the pool are grouped together.

Figure 7. The dJ-poor sequences emerging from the selection with dJ libraries for receptors that bind to ADP. The three G-rich segments are underlined. Italicized regions are derived from primer or primer-binding regions. These oligonucleotides all appear to have arisen from a single ancestral sequence in the pool.

significant amounts of 1:1 complex were detected by affinity capillary electrophoresis.

Discussion

Although in vitro selection has been known as an experimental technique for more than a decade, few in vitro selection experiments have been used to quantitatively understand the relationship between structure and behavior in the large collections of molecules that are the starting point for these selection experiments. In vitro selection experiments may make their most important contribution by way of such analyses, as they concern questions at the center of both chemistry (How are behaviors distributed in a landscape of structure space?) and biology (How does Darwinian evolution identify functional behavior in a random search of a fitness-structure landscape?). These results show how the introduction of expanded genetic alphabets containing extra letters[6] and new chemical functional-

(20) A false perception of the affinity between a receptor and a ligand can be obtained if single data points are fitted to a model that makes an incorrect assumption about the molecularity of the complex; see Benner, S. A.; Burgstaller, P.; Battersby, T. R.; Jurczyk, S. in *The RNA World*, 2nd ed.; Cech, T. R., Atkins, J., Eds.; Cold Spring Harbor Press: Plainview, NY, 1999; 163−181.

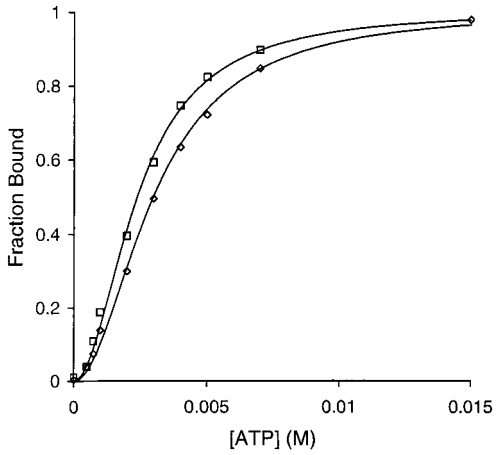

Figure 8. Equilibrium binding curves determined for Sequence A (diamonds) containing only standard nucleotides, and Sequence 409 (squares), combining the randomized region (in italics) with the 5′- and 3′-primer binding sites (GGTCGTCTAGAGTATGCGGTAG-*GAACGJCAGJG GGGGGAG CAJAJGGJGJGAJACGCGA*CCGAA-GAAGCJJGGCCCAJG), and containing the functionalized nucleotide dJ.

ity[7] empowers in vitro selection as a tool for addressing such general questions.

Several features of the results reported here are worth noting. First, our experiments with natural DNA libraries independently generated the same motif as generated by Huizenga and Szostak under analogous conditions. This convergence is not surprising, as the Huizenga−Szostak motif is short enough to be present in multiple copies in any random library of the size used in these experiments.

Less expected, however, was the convergence of molecular phenotype starting from two *types* of library, the natural library and the functionalized dJ library. Both libraries yielded receptors that bind *two* molecules of ATP cooperatively in a termolecular complex. We can only speculate why forming a ternary complex is optimal for the DNA and dJ-functionalized DNA systems. Almost certainly relevant is the fact that the selection recipe exploits an agarose affinity resin that is highly loaded with ATP (~3 mM effective concentration); most of the glucose units in the agarose strand had one ATP molecule covalently bound. For this reason, if a receptor binds one polymer-bound ATP molecule, it would almost always have the opportunity to bind a second. It appears that sequences exploiting this opportunity to grasp the solid support during the selection step are present in greater abundance in both libraries than sequences that have comparable affinity by binding just one agarose-bound ATP.

The independent searches of natural and unnatural sequence spaces also converged on similar affinity constants. The equilibrium constant for ATP binding to Sequence 409 is $6 \pm 1 \times 10^{-6}$ M^2 (note the units). The convergence of the two disassociation constants presumably reflects the fact that the effective concentration of adenosine derivative attached to the solid support was $\sim 3 \times 10^{-3}$ M. At these concentrations, the receptors were approximately half saturated with ligand. Thus, the receptor evidently evolved an affinity that is optimal for the selection conditions. The affinity was not maximized (that is, increased until the receptor could bind no tighter), presumably because no pressure selected for receptors with higher affinity, and higher-affinity receptors are more scarce in the sequence space than lower-affinity receptors.

Whether it occurs in biology or in nonliving chemical systems, convergence of phenotype suggests that the phenotype converged upon is *optimal* for a particular environment. This is, we believe, the first time that such convergence has been observed in a "non-living" system. It is, of course, commonplace in living systems. Convergence at the level of sequence may also occur from time to time.[21] Thus, certain features of bat wings and bird wings are strictly analogous (weight−power ratios, for example), even though the two physiological structures are not homologous. The analogous phenotypes suggest that the features of the phenotype are optimal; such a conclusion would not be possible in the absence of convergence. Thus, the interplay between environmental conditions demanding a function of a biopolymer and the ability of the biopolymer to deliver that function is strictly analogous to that observed during natural selection, illustrating the nature of life as a self-sustaining chemical system capable of Darwinian evolution.

Is there any reason to believe that structure space delivered by the functionalized biopolymer containing dJ offered a richer density of functionality than the natural biopolymer? More work needs to be done to say for certain. However, it is interesting to note that the sequences selected to bind ATP were longer and, consequently, had a higher number of dJs than the sequences arising from selection to bind the less negatively charged adenosine analogues ADP and AMP.

It is clear, however, that the addition of functionality did not change dramatically the phenotype-structure landscape. The first indicator of this is the fact that the dJ library generated a few oligonucleotides that almost certainly fold in the same way that the Lin−Patel−Huizenga−Szostak motif folds and presumably bind adenosine derivatives in an analogous way. Further, we prepared one of the dJ-containing receptors, Sequence 409, with T replacing dJ. This receptor was found to bind ATP, although with a 2-fold weaker affinity ($13 \pm 4 \times 10^{-6}$ M^2) than the analogous sequence containing dJ. The fact that the Lin−Patel−Huizenga−Szostak fold evidently appears only infrequently in the dJ-selection suggests that the selection procedure is demanding enough to exclude receptors with 2-fold weaker affinity. Together, these results suggest that the functionalization contributed by dJ influences modestly, rather than dramatically, the affinity-structure landscape, at least for these receptors selected under rather nondemanding conditions. When called upon to provide catalytic activity, however, DNA makes greater use of functionality (Ang et al., in preparation).

Experimental Section

Reagents were purchased from Aldrich or Fisher, unless otherwise indicated. NMR spectra were recorded on a Varian XL 300 spectrometer at 300 MHz (^{1}H) referenced to TMS, at 75.4 MHz (^{13}C) referenced to solvent, and at 121.4 MHz (^{31}P) with H_3PO_4 as standard in the solvents as given. UV spectra were measured on a Varian Cary 1 Bio spectrophotometer. Mass spectrometry was performed by Spectroscopy Services of the University of Florida Chemistry Department.

Chromatography. Column chromatography was performed with silica gel (230−425 mesh, Fisher) and the indicated eluting solvents. Silica gel plates with 254 nm fluorescence (Whatman) were used for TLC. TLC separations were also visualized by staining with a Ce/Mo reagent (2.5% phosphormolybdic acid, 1% $Ce(IV)(SO_4)_2 \cdot 4H_2O$, 6% H_2SO_4 in H_2O) and heating. AG 1-X8 resin (Bio-Rad) was obtained as the Cl^- form and converted to the HCO_3^- form by washing with 16 volumes of 1 M NH_4HCO_3 solution, followed by deionized water and finally with 0.5 M NH_4HCO_3 solution; following this treatment, no eluting Cl^- was detected. Ion-exchange chromatography was done with DEAE Sephadex (Sigma) equilibrated in 0.2 M $(Et_3NH)HCO_3$ (pH 7.0). HPLC was performed using a Waters PrepLC 4000 System with a 486 tunable absorbance detector. Reversed-phase HPLC separation was done

(21) Stewart, C. B.; Schilling, J. W.; Wilson, A. C. *Nature* **1987**, *330*, 401−404.

using a Waters PrepLC 25 mm module containing a single PrepPak cartridge (Prep Nova-Pak HR C_{18} 6 μm 60 Å, 25 × 100 mm).

5-(3″-Trifluoroacetamidopropynyl)-2′-deoxyuridine. 5-Iodo-2′-deoxyuridine (Sigma; 1.41 mmol, 500 mg) was dissolved in dry DMF (12 mL). Ar was passed through this solution for 10 min. $(Ph_3P)_4Pd$ (0.1 equiv, 0.141 mmol, 163 mg) was added, and Ar was passed through the solution for another 5 min. Triethylamine (2.0 equiv, 2.83 mmol, 285 mg, 0.393 mL) was added via syringe, followed by addition of *N*-propynyltrifluoroacetamide (2.5 equiv, 3.53 mmol, 533 mg) and CuI (0.2 equiv, 0.282 mmol, 53.7 mg). The mixture was stirred at 40 °C for 5 h, the solvent was evaporated, and the residue was dissolved in MeOH/methylene chloride 1:1 (10 mL). Ion-exchange resin (AG1-X8 HCO_3^- form, 1.5 g) was added to remove the Et_3N·HI byproduct, and the mixture was stirred at room temperature for 30 min. The mixture was filtered through Celite, the solid was washed with MeOH/methylene chloride 1:1 (10 mL), and the combined filtrates were evaporated. The residue was purified by silica column chromatography (chloroform/MeOH 8.25:1.75) to yield product as a yellow foam containing some DMF. R_f: 0.42 (chloroform/MeOH 8.25:1.75). ^{1}H NMR (DMSO-d_6): 2.13 (m, 2H, 2′), 3.60 (m, 2H, 5′), 3.81 (m, 1H, 4′), 4.24 (m, 3H, H-9, 3′), 5.12 (t, 1H, 5′-OH), 5.27 (d, 1H, 3′-OH), 6.11 (t, 1H, 1′), 8.22 (s, 1H, H-6), 10.09 (t, 1H, NH chain), 11.67 (s, 1H, NH cycl.). ^{13}C NMR (DMSO-*d*6): 29.5 (C-9), 40.7 (2′), 61.0 (5′), 70.2 (3′), 75.4 (C-8), 84.8 (1′), 87.5, 87.7 (4′, C-7), 97.7 (C-5), 113.9, 117.8 (q, CF_3, J = 287.95 Hz), 144.2 (C-6), 149.5 (C-2), 155.9, 156.4 (q, *CO*CF_3, J = 37.25 Hz), 161.7 (C-4).

5-(3″-Trifluoroacetamidopropynyl)-5′-*O*-dimethoxytrityl-2′-deoxyuridine. The product from above (1.34 mmol, 506 mg) was coevaporated with pyridine, then dissolved in dry pyridine (10 mL) and cooled to 0 °C. Et_3N (2 equiv, 2.68 mmol, 271 mg, 0.373 mL), DMAP (0.25 equiv, 0.336 mol, 41 mg), and 4,4′-dimethoxytrityl chloride (1.2 equiv, 1.61 mmol, 545 mg) were added, and the mixture was stirred at 0 °C for 5 min and then at room temperature for 4 h. TLC (chloroform/10% MeOH, R_f = 0.47) did not show any starting material. MeOH (2 mL) was added and the mixture evaporated. The residue was extracted (ethyl acetate/aqueous $NaHCO_3$ solution), the combined organic layers were washed with water and dried (Na_2SO4), and the solvent was evaporated. The residue was purified by silica column chromatography (chloroform/10% MeOH) to give 902 mg (99%) of product as a yellow foam. ^{1}H NMR ($CDCl_3$): 2.44−2.61 (m, 2H, 2′), 3.34 (m, 2H, 5′), 3.73 (s, 6H, MeO), 3.89 (m, 1H, 4′), 4.14 (m, 2H, H-9), 4.59 (m, 1H, 3′), 6.34 (t, 1H, 1′), 6.80 (m, 4H, DMT), 7.14−7.33, 7.61−7.70 (m, 9H, DMT), 8.21 (s, 1H, H-6). ^{13}C NMR ($CDCl_3$): 30.3 (C-9), 41.6 (2′), 55.2 (MeO), 63.5 (5′), 72.0 (3′), 75.3 (C-8), 86.0 (1′), 86.9 (Cq trityl), 87.0, (C-7), 87.3 (4′), 99.9 (C-5), 113.3 (DMT), 113.8, 117.6 (q, CF_3, J = 286.52 Hz), 126.9, 127.8, 128.0, 129.9, 135.4 (all DMT), 143.6 (C-6), 144.5 (DMT), 149.4 (C-2), 156.4, 156.9 (q, *CO*CF_3, J = 37.70 Hz), 158.5 (DMT), 162.6 (C-4).

3′-*O*-Acetyl-5-(3″-trifluoroacetamidopropynyl)-5′-*O*-dimethoxytrityl-2′-deoxyuridine. The product from above (1.33 mmol, 902 mg) was coevaporated with pyridine and then dissolved in pyridine (10 mL), and DMAP (0.25 equiv, 0.332 mmol, 40.5 mg), Et_3N (2.5 equiv, 3.32 mmol, 335 mg, 0.462 mL), and Ac_2O (1.2 equiv, 1.60 mmol, 163 mg, 0.15 mL) were added. It was stirred at room temperature for 2.5 h (TLC: chloroform/10% MeOH, R_f = 0.75); then MeOH (5 mL) was added to stop the reaction and it was evaporated to dryness. The residue was extracted (water/ethyl acetate), the organic layer was dried (Na_2SO_4), and the solvent was evaporated. The residue was purified by silica column chromatography (chloroform/10% MeOH) to give 937 mg (98%) of product as a yellow foam. ^{1}H NMR ($CDCl_3$): 2.08 (s, 3H, Ac), 2.38−2.64 (m, 2H, 2′), 3.42 (m, 2H, 5′), 3.77 (s, 6H, MeO), 3.96 (m, 1H, 4′), 4.18 (m, 2H, H-9), 5.45 (m, 1H, 3′), 6.34 (t, 1H, 1′), 6.85 (m, 4H, DMT), 7.23−7.48, 7.63−7.71 (m, 9H, DMT), 8.20 (s, 1H, H-6).^{13}C NMR ($CDCl_3$): 20.8 (Ac), 30.2 (C-9), 38.7 (2′), 55.1 (MeO), 63.5 (5′), 74.9 (3′), 75.1 (C-8), 84.4 (1′), 85.3 (4′), 87.1 (Cq trityl), 87.2 (C-7), 99.4 (C-5), 113.3 (DMT), 113.6, 117.4 (q, CF_3, J = 287.42 Hz), 126.9, 127.7, 128.0, 129.9, 135.2 (all DMT), 143.1 (C-6), 144.3 (DMT), 149.4 (C-2), 156.3, 156.8 (q, *CO*CF_3, J = 37.78 Hz), 158.6 (DMT), 162.0 (C-4), 170.3 (Ac).

3′-*O*-Acetyl-5-(3′-trifluoroacetamidopropynyl)-2′-deoxyuridine. The compound obtained above (1.30 mmol, 937 mg) was dissolved in dry MeOH (1 mL). A solution of anhydrous HCl in anhydrous MeOH (10%, 1 mL) was added, and the reaction was complete after 3 min (TLC, chloroform/10% MeOH, R_f = 0.46). It was cooled to 0 °C, and aqueous $NaHCO_3$ solution was added to adjust pH 7. It was extracted (ethyl acetate) and dried (Na_2SO_4) and the solvent evaporated. The residue was purified by column chromatography (chloroform/10% MeOH) to give 371 mg (68%) of a slightly yellow foam. ^{1}H NMR (CD_3OD): 2.03 (s, 3H, Ac), 2.18−2.40 (m, 2H, 2′), 3.71 (m, 2H, 5′), 4.02 (m, 1H, 4′), 4.19 (d, 2H, H-9), 5.21 (m, 1H, 3′), 6.14 (t, 1H, 1′), 8.25 (s, 1H, H-6). ^{13}C NMR (CD_3OD): 20.9 (Ac), 30.7 (C-9), 39.0 (2′), 62.8 (5′), 76.0 (3′), 76.4 (C-8), 86.9 (1′), 87.1 (C-7), 88.5 (4′), 99.9 (C-5), 111.6, 115.4, 119.2, 123.0 (q, CF_3, J = 285.99 Hz), 145.5 (C-6), 151.1 (C-2), 158.2, 158.7 (q, $COCF_3$, J = 37.78 Hz), 164.4 (C-4), 172.2 (Ac). HRMS (FAB^+): calculated, 420.1019; found, 420.1026.

5-(3″-Aminopropynyl)-2′-deoxyuridine 5′-triphosphate. The compound from the previous step (1.21 mmol, 505 mg) was coevaporated with pyridine and dissolved in 1.75 mL of pyridine/5.25 mL of dioxane (anhydrous), and 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one (1.1 equiv, 1.33 mmol, 269 mg), dissolved in 1.4 mL dioxane, was added. A white precipitate formed immediately, and the mixture was stirred at room temperature for 10 min. Tributylammonium pyrophosphate (819 mg), dissolved in 4.55 mL of DMF/1.3 mL of tributylamine, was added, and the precipitate was dissolved within a few seconds. The solution was stirred for another 10 min, iodine (1.1 equiv, 1.33 mmol, 337 mg), dissolved in 30 mL of pyridine/0.6 mL of water was added, and the resulting solution was stirred for 15 min. The reaction was quenched by addition of 5% aqueous Na_2SO_3. The solvent was evaporated at room temperature, and the residue was dissolved in water (25 mL) and stirred at room temperature for 30 min. Concentrated aqueous ammonia (100 mL) was added, and it was stirred for another 5 h. The solvent was again evaporated at room temperature and the residue purified by ion-exchange chromatography on DEAE Sephadex using a gradient of 0.1M-0.5M NEt_3HCO_3. The UV absorbing fractions were evaporated at 20 °C, and the product was stored at −20 °C. Reversed-phase HPLC purification of the triphosphate was done at a flow rate of 5.1 mL/min using 25 mM (Et_3NH)OAc (pH 7.0) (solvent A) and 20% (v/v) MeCN in solvent A (solvent B) in the following linear binary gradient: 100% A (1 min); 100% A to 90% A (39 min); retention time 33 min. The concentration of dJTP was estimated using an extinction coefficient (ϵ_{290} = 9000 $M^{-1}cm^{-1}$) consistent with measured values for analogously substituted compounds.[22] A high-resolution mass spectrum was obtained showing an exact mass of 519.9885, corresponding to a calculated exact mass of 519.9923. ^{31}P NMR (D_2O): −1.97, −2.13 (d, J = 19.18 Hz), −6.33, −6.48 (d, J = 19.18 Hz), −16.80 (dt). ^{1}H NMR (D_2O): 1.24 (t, 12H, Et_3N), 2.40 (m, 2H, 2′), 3.18 (q, 8H, Et_3N), 4.06 (m, 2H, 5′), 4.19 (m, 2H, H-9), 4.25 (m, 1H, 4′), 4.61 (m, 1H, 3′), 6.21 (t, 1H, 1′), 8.31 (s, 1H, H-6). ^{13}C NMR (D_2O): 6.8 (Et_3N), 28.2 (C-9), 37.9 (2′), 45.0 (Et_3N), 63.4 (5′), 68.1 (3′), 76.4 (C-8), 83.9, 84.1, 84.4 (1′, 4′, C-7), 96.5 (C-5), 144.4 (C-6), 148.7 (C-2), 162.5 (C-4). UV (H_2O, pH 7) λ_{max} 230, 290 nm. MS (ESI^+) 522 $[M + H]^+$, 544 $[M + Na]^+$.

DNA Reagents and Materials. ThermoPol 1× buffer and Vent DNA polymerase were purchased from New England Biolabs, standard dNTPs and Taq DNA polymerase from Promega, Microspin S-300 HR columns from Pharmacia, streptavidin-agarose from Fluka, and AMP-, ADP-, and ATP-agarose (C-8 attachment) from Sigma. The random-sequence DNA library was synthesized by Microsynth (Balgach, Switzerland); all primers, templates, and 3′-fluorescein-tagged Sequences A and B were from Integrated DNA Technologies (Coralville, IA). Oligonucleotides were purified by polyacrylamide gel electrophoresis (PAGE).

DNA Pools. A library of DNA containing the four natural bases in equal proportion and having a 70 base random region flanked by two defined primer binding sites of 22 and 20 bases (Figure 3) was synthesized (200 nmol scale). The library was purified (denaturing 8% PAGE), yielding 3.2 nmol of oligonucleotide. Approximately 10% of the DNA was able to be PCR amplified, giving ∼2 × 10^{14} different molecules. All 120 μg of DNA was PCR amplified in a reaction volume of 50 mL (10 mM KCl; 10 mM $(NH_4)_2SO_4$; 20 mM Tris·HCl, pH 8.8; 2 mM $MgSO_4$; 0.1% Triton X-100 (1× ThermoPol reaction buffer); 0.2 mM dNTPs; 0.5 μM primers; 20 units/mL Vent DNA polymerase)

(22) Robins, M. J.; Barr, P. J. *J. Org. Chem.* **1983**, *48*, 1854−1862.

for six cycles (temperature cycle conditions: 94 °C, 4 min; 55 °C, 7 min; 72 °C, 7 min). The PCR product was extracted with phenol/$CHCl_3$, ethanol precipitated, redissolved in 1 mL of TE (10 mM Tris·HCl, pH 7.6; 1 mM EDTA), and purified using Microspin S-300 HR columns.

Biotinylated primer to initiate complementary strand synthesis and functionalized nucleoside dJ were introduced to create a pool of dJ-containing biotinylated DNA (dJ pool) by PCR amplifying 22 μg of DNA (0.3 nmol) in a volume of 30 mL (1× ThermoPol reaction buffer; 0.2 mM dATP, dCTP, and dGTP; 0.2 mM dJTP; 0.5 μM biotinylated primer and standard primer; 20 units/mL Vent DNA polymerase) over 10 cycles (temperature cycle conditions as above). A biotinylated pool made of the four naturally occurring nucleosides (standard pool) was amplified with the biotinylated primer and the four standard dNTPs over six cycles. After phenol/$CHCl_3$ extraction, each pool was redissolved in TE (200 μL). Analogous PCR amplifications were done for each pool (0.5 mL) with 20 μCi α-^{32}P-dATP and combined with the larger amplification, to have a small portion of radiolabeled DNA.

A streptavidin-agarose column (0.5 mL), which had been equilibrated in wash buffer (50 mM NaCl; 10 mM Tris·HCl, pH 7.5; 1 mM EDTA), was used to make the DNA single-stranded. The DNA was applied to the column and incubated (5 min, rt), and then the column was washed with 10 volumes of wash buffer. The streptavidin-agarose was suspended in 0.5 M NaOH (1.5 mL), removed from column, and incubated (37 °C, 15 min). The suspension was returned to the column, and the NaOH solution containing the single-stranded DNA was eluted and collected. The DNA was ethanol precipitated and redissolved in water (150 μL).

Selection. The pools were preselected to minimize enrichment of DNA molecules interacting with column material.[23] A preselection agarose gel was prepared from 6-aminohexanoic acid *N*-hydroxysuccinimide ester-sepharose 4B. The gel was washed with 1 mM HCl (200 mL/g) for 1 h and then incubated (5 min, rt) in 0.5 M NaCl; 0.1 M Tris·HCl, pH 8.0 solution. The gel was washed with several volumes of 0.1M NaOAc, pH 4.0; 0.5 M Na Cl and then with 0.1 M Tris·HCl, pH 8.0; 0.5 M Na Cl. The resulting preselection agarose was stored in 1 mM EDTA, pH 8.0 at 4 °C.

All permutations of selections against AMP-, ADP-, or ATP-agarose using the standard pool or the dJ pool were done (totaling six separate selection experiments). The adenosine phosphate immobilized on agarose was washed with several volumes of water and stored in 1 mM EDTA, pH 8.0, at 4 °C. Preselection agarose (0.2 mL) and adenosine phosphate agarose (0.5 mL) were equilibrated with 20 volumes of selection buffer (250 mM Na Cl; 50 mM Tris·HCl, pH 7.6; 5 mM $MgCl_2$). Single-stranded DNA (33 μg, ~1 nmol), from the combination of the 30 mL PCR product with the analogous radiolabeled 0.5 mL PCR product, was denatured in selection buffer (95 °C, 3 min) and allowed to cool (15−20 min). The DNA was applied to the preselection column, eluted with 2.5 volumes of selection buffer, and applied directly to the selection (immobilized adenosine phosphate) column. The selection column was washed with 5 volumes of selection buffer. Interacting molecules were eluted with 3 volumes of selection buffer containing the appropriate adenosine phosphate at 1.5× the concentration indicated by the manufacturer of the AMP-, ADP-, or ATP-agarose. The amount of adenosine phosphate on the agarose varied between batches (1.3−4.5 mM).

The eluted DNA was ethanol precipitated in the presence of 100 μg of glycogen and redissolved in water (50 μL). The precipitated DNA was PCR amplified (0.5 mL reaction volume) in the presence of 20 μCi α-^{32}P-dATP, as described above, to generate a biotinylated pool. Generally, 10−14 cycles of PCR were necessary to get a sufficient amount of DNA. The PCR product was extracted with phenol/$CHCl_3$, ethanol precipitated, and made single-stranded with streptavidin-agarose as above. Half of this single-stranded DNA was stored at −20 °C, and the remaining half was used for the next round of selection. The iterative process was repeated with another selection column for 8 further rounds.

Cloning and Sequencing. After the ninth round of selection, the eluted DNA was amplified in a 0.3 mL reaction volume (1× Taq polymerase buffer; 0.2 mM dNTPs; 1.0 μM primers; 6 U Taq DNA polymerase) for 10 cycles. The resulting DNA with T replacing dJ was cloned into the vector pCR 2.1-TOPO using the TOPO TA Cloning Kit (Invitrogen). For the standard pool, 10 clones were sequenced per selection experiment, and for the dJ pool, 15 clones were sequenced. The randomized region of the dJ-containing sequences was shortened during the selection. This result suggests biases against longer oligonucleotide receptors and dJ, although these results do not permit a conclusion as to whether these biases arise in the selection or the amplification steps.

Fluorescently Labeled Sequence 409. Sequence 409 with T replacing each dJ was used as a template to produce Sequence 409 in parallel PCR amplifications (20 × 0.1 mL reaction volumes) with dJTP and a biotinylated primer. Each 0.1 mL PCR reaction (1× ThermoPol reaction buffer; 0.2mM dATP, dCTP, and dGTP; 0.2 mM dJTP; 0.5 μM biotinylated primer and standard primer; 20 units/mL Vent DNA polymerase; 1 nM template) was amplified over 12 cycles (temperature cycle conditions: 94 °C, 1 min; 55 °C, 2 min; 72 °C, 2 min). The PCR products were pooled, phenol/$CHCl_3$ extracted twice, the aqueous phase was ethanol precipitated, and the resulting pellet was redissolved in TE (250 μL). The redissolved PCR product was made single-stranded using streptavidin-agarose, as described for the DNA pools. The oligonucleotide was ethanol precipitated with 3 μg of glycogen as a carrier.

Sequence 409 was fluorescently labeled at the 3′ end with terminal transferase (Boehringer Mannheim) and fluorescein-12−2′,3′-dideoxyuridine 5′-triphoshate (Boehringer Mannheim). Sequence 409 (0.2 nmol) was dissolved in 45 μL of labeling buffer (200 mM potassium cacodylate; 25 mM Tris·HCl, pH 6.6; 0.25 mg/mL bovine serum albumin) containing cobalt chloride (5 mM), fluorescein-12-ddUTP (40 μmol), and terminal transferase (0.6 units/μL). The reaction mixture was incubated (1 h, 37 °C) and phenol/$CHCl_3$ extracted twice, and the aqueous phase was ethanol precipitated with glycogen as a carrier.

***K_d* Determinations.** Affinity capillary electrophoresis was performed using a Beckman P/ACE 2200 automated capillary electrophoresis system with laser-induced fluorescence (LIF) detection (Beckman-Coulter, Fullerton, CA). The 488 nm line (3 mW) of the Ar^+ laser was used for excitation, and fluorescence was collected at 520 nm. Separations were performed using 50 μm i.d. capillaries (Polymicro Technologies, Phoenix, AZ). All capillaries were coated with polyacrylamide using a method similar to that of Dolník et al.[24] The acrylamide solution used in preparing the capillary was 3 mL of aqueous acrylamide (10% w/v; Sigma), 7 mL of water, 100 μL of 10% aqueous ammonium persulfate (Sigma), and 10 μL of tetramethylethylenediamine (Sigma). Capillaries were stored in tris buffer (50 mM Tris·HCl, pH 7.6; 50 mM sodium chloride; 50 mM magnesium chloride) when not in use.

DNA samples (~200 nM) were prepared daily in tris buffer. Each DNA sample was heated at 88−90 °C for 2.5 min and then cooled to room temperature before beginning experiments. A 30−50 mM ATP (Sigma) stock was made in tris buffer, and the pH was adjusted to 6.5 with 1 M sodium hydroxide (Fisher). Electrophoresis buffers were made from this stock in concentrations ranging from 500 μM to 10 mM ATP. The capillary was rinsed with the electrophoresis buffer for 3 min before each 5 s hydrodynamic injection. The separation potential was 10 kV in a 37 cm capillary. The length to the detector was 30 cm. Data were collected with Beckman P/ACE software version 2.0 using a 386 PC computer. Binding constants were determined by nonlinear least-squares regression using the variance-covariance method.

Acknowledgment. The authors acknowledge support from Office of Naval Research (N00025-96-1-0362 to S.A.B.), the National Institutes of Health (GM 54048 to S.A.B.), and the National Science Foundation (CHE 9531428).

Supporting Information Available: Additional experimental procedures and data (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA9816436

(23) Famulok, M. *J. Am. Chem. Soc.* **1994**, *116*, 1698−1706.

(24) Dolník, V.; Xu, D.; Yadav, A.; Bashkin, J.; Marsh, M.; Tu, O.; Mansfield, E.; Vainer, M.; Madabhushi, R.; Barker, D.; Harris, D. *J. Microcolumn Sep.* **1998**, *10*, 175−184.