Bioorganic & Medicinal Chemistry Letters, Vol.3, No.3, pp. 469-475, 1993
Printed in Great Britain

0960-894X/93 $6.00 + .00

PROTEIN EPITOPE TARGETING: OLIGONUCLEOTIDE DIVERSITY AND DRUG DISCOVERY

Michael I. Sherman*, Arthur H. Bertelsen and Alan F. Cook
PharmaGenics, Inc.
4 Pearl Court, Allendale, New Jersey 07401

Abstract. Oligonucleotides are ideal compounds to use for high-throughput screening. This article describes techniques by which oligonucleotides with potential therapeutic activity can be identified and chemically modified to serve as drugs.

Introduction

Although rational drug design is considered to be the wave of the future for drug discovery, past successes have often resulted from high-throughput random screening of compounds. In fact, with the availability of genetic engineering technologies to generate relatively large amounts of pure target molecules, usually proteins, the two approaches to drug discovery can usefully be employed in parallel: large numbers of compounds can be evaluated in an appropriate assay for ability to modulate the activity of a pure and highly-specific target molecule while efforts are made to determine the detailed structural characteristics of that target molecule. Thus, once an active compound is identified, it can be determined how it interacts physically with the target molecule and more potent adducts can be designed.

Success in high-throughput screening is not guaranteed. If the inventory tested does not include an active compound, the effort will not be successful, no matter how well-planned or diligent the program. Those efforts that have resulted in success have not been without considerable cost, in terms of time, labor and expense. In the past, only those research groups with large inventories of compounds or biological extracts (e.g., microbial broths) have been able to enter the high-throughput screening sweepstakes. Biological extracts hold the promise of greatly expanding the molecular diversity as compared to individual compound screening (since the former reflect Nature's ingenuity and breadth rather than that of a group of chemists). However, for biological extract screening to be successful, the active compound must be present in the extract in an adequate concentration to demonstrate activity, the compound must be extracted (often from among a group of many thousands of inactive compounds) and identified, and usually a synthetic approach must then be devised and implemented.

Recently, therapeutic discovery approaches involving peptides have become very popular. Although random peptides can be synthesized in an automated fashion, a limitation to their use is the rapidly increasing diversity that accompanies increases in the size of the peptide. For example, a random tripeptide library consisting of all combinations of the twenty common amino acids would have $20^3 = 8000$ different sequences, a manageable inventory; if tetrapeptides were generated, that would result in 160,000 sequences, a number that poses serious challenges in cataloguing and storage. A recent publication[1] describes how the generation of peptides on beads can address these problems. Another strategy is to generate peptides biosynthetically via phage particles, identify and expand the phage populations carrying active peptides and then deduce the active peptide sequences from the mRNA or DNA encoding them.[2,3]

Oligonucleotides are a versatile alternative to peptides for high-throughput screening. DNA and RNA oligonucleotides can be easily and efficiently generated with DNA synthesizers. Because relatively large oligomers can be made easily and inexpensively, the diversity of the mixtures can be extremely large. For example, a random 30-mer oligonucleotide will contain more than a trillion different sequences. Following incubation of target proteins with these mixtures, bound oligonucleotides can be isolated and identified. Oligonucleotides have an enormous advantage over peptides or any other class of compounds as screening materials because the bound oligonucleotide species can be amplified to facilitate identification. Active oligonucleotides can be inserted into vectors, cloned and sequenced or they can be amplified by the polymerase chain reaction (PCR)[4] and then sequenced directly. Because of this unique opportunity to amplify and rapidly identify one selected oligonucleotide from among a vast array, an entire oligonucleotide inventory can be efficiently and conveniently tested in a single assay sample. As an added bonus, since minute amounts of active oligonucleotide can be amplified and identified, relatively small amounts of target proteins are needed.

Proof of Principle

The concept of using fully degenerate, random-sequence oligonucleotides (randomers) to address biological questions was first applied by Oliphant and Struhl[5] to study the sequence requirements for E. coli promoter elements. In this early study a synthetic randomer was subjected to "mutually primed synthesis"[6] to generate a mixture that was amenable to cloning. This mixture was then used to replace functional elements in a prokaryotic promoter driving expression of a selectable marker. Under selection

pressure only those molecules able to provide the required promoter function could generate viable clones. In this way a vast number of independent sequences was rapidly evaluated for function and a consensus sequence of requirements determined.

Subsequent work by Oliphant et al.[7] refined the concept so that it could be applied to situations where genetic selection was not readily available. The approach, involving in vitro enrichment of the randomer mixture by repeated affinity selection prior to cloning, was utilized to refine the definition of consensus binding site requirements for the yeast transcriptional activator GCN4. At about the same time, Kinzler and Vogelstein[8] were attempting to isolate specific DNA sequences from human genomic DNA capable of binding the transcription factor TFIIIA from Xenopus.

In their approach, Kinzler and Vogelstein[8] generated a library of sheared human genomic fragments and ligated these fragments to "catch linkers" which could serve both as cloning sites and priming sites for the polymerase chain reaction. The "catch-linked" library was then mixed with Xenopus TFIIIA and the protein was immunoprecipitated. Nucleic acid sequences that were bound to the transcription factor were co-precipitated. Because the inherent selection efficiency of the specific binding sequence(s) was expected to be low, repeated rounds of binding and isolation were used to enrich for the desired sequences. To increase the quantity of bound DNA following each selection cycle, the catch-linked sequences were used to prime a PCR reaction with the co-precipitated DNA. The newly synthesized enriched library of fragments was then recycled via binding and immunoprecipitation, thereby generating a reiterative procedure of enormous power. With multiple cycles even a marginal difference in binding affinity could result in the isolation of the most avidly binding fragments from the remainder of the population. Kinzler and Vogelstein[8] reported that after four cycles of selective enrichment the complexity of the population of human genomic sequences was adequately reduced to permit identification of known TFIIIA protein binding sequences by cloning and sequence analysis.

A limitation of the work by Kinzler and Vogelstein is that the complexity that can be examined is limited by the starting complexity of the fragmented genomic DNA. In the case of the human genome, owing to repetitiveness the complexity is considerably less than the 6×10^9 base pairs of the diploid cell. However, this level of complexity can be achieved with oligonucleotides simply by synthesizing a fully degenerate 17-base randomer. Increasing the size of the randomer can result in a virtually limitless number of DNA (or RNA) sequences. In fact, the major limitation to sampling sequences becomes the mass of DNA that must be added to the target to fully represent the randomer complexity.

As anticipated and demonstrated by Oliphant and Struhl[7] and by Kinzler and Vogelstein,[8] these methods, with minor variations, have proven to be valuable research tools. In subsequent studies, Tuerk and Gold[9] have used this methodology to identify RNA sequences bound by T4 DNA Polymerase, whereas Ellington and Szostak[10] have used single-stranded RNA generated from DNA randomers to identify nucleotide sequence requirements for binding to a variety of dye molecules. This basic approach has also been used to identify or confirm the specific DNA sequences recognized by numerous nucleic acid binding proteins including GCN4,[11] SP1,[12] myoD and E2A,[13] WTL protein,[14] U1snRNP-A,[15] BmFTZ-F1,[16] c-myb[17] and p53.[18]

Protein Epitope Targeting

The foregoing studies illustrate the utility of random oligonucleotide screening for basic research purposes. These studies have provided the impetus to search for oligonucleotides that might have therapeutic utility. For example, it is generally accepted that some oncogenes are DNA-binding proteins.[17,19] Introduction into cancer cells of oligonucleotide sequences to which these oncogene proteins bind could sequester them and thereby prevent them from eliciting their undesirable activity. However, even proteins not normally known to bind nucleic acids might interact with considerable affinity with one or more among the trillion or so 30-mer oligonucelotides present in a random mixture. A demonstration of this assumption, involving the protein thrombin, has recently been published.[20]

If the premise is accepted that for many or most proteins, one or more oligonucleotide sequences will be identifiable which bind(s) with relatively high affinity, pharmacologically-relevant oligonucleotides might be identified which bind specifically to active epitopes of those target proteins. We call this approach Protein Epitope Targeting. The strategy, then, is to attempt to identify oligonucleotides which will block the physiological action of a target protein by interacting with its active site. This approach is particularly attractive for identifying inhibitors of the interaction between protein ligands and their receptors. There are numerous ligand-receptor interactions which it would be desirable to interrupt in various pathological states (e.g. growth factors, adhesion molecules), and scientists have been searching for appropriate inhibitory compounds. Although antibodies and modified receptors have been used for this purpose, there has been little success in identifying low molecular weight

inhibitors of ligand-receptor interactions. By evaluating trillions of oligonucleotides, the Protein Epitope Targeting approach might be productive where other screening efforts, involving several orders of magnitude fewer compounds, have failed.

From Oligonucleotide to Drug

Often in a screening procedure, active compounds are not directly useful as drugs but serve as lead structures which are further modified to generate the final pharmaceutical product. After a Protein Epitope Targeting sequence has been identified by in vitro binding assays, it will then be desirable to improve the pharmacokinetic and pharmacodynamic properties of the oligonucleotide so that adequate concentrations of the therapeutic agent can be maintained in vivo. This can potentially be achieved in any number of ways.

One of the challenges in the design of oligonucleotide analogs is to develop compounds which have improved stability as compared with their natural counterparts. Natural oligonucleotides are rapidly degraded in vivo primarily via cleavage of the phosphodiester backbone[21] and this instability has stimulated efforts to make oligonucleotides more resistant to degradation. A widely studied approach is the replacement of the negatively charged oxygen with a nonionic methyl group to give methyl phosphonates[22] which are more stable towards degradation but less soluble in aqueous solvents. Other nonionic backbone modifications include phosphotriesters[23] and phosphoramidates.[24] Another tactic is to replace the negatively charged oxygen with a sulfur atom to give phosphorothioates, which provide increased stability while retaining the desirable solubility characteristics of natural oligos.[25] Such substitutions would be logical for Protein Epitope Targeting-selected oligonucleotides providing that the presence of the negatively charged backbone oxygen is not critical for binding to the target molecule.

Drug delivery is a key issue for intracellular targets. Although natural oligonucleotides can penetrate the cell membrane via a receptor-mediated process,[26] this process is relatively inefficient. One approach to the improved delivery of oligonucleotides is to attach a conjugate group. Examples of such groups include peptides,[27,28] polylysine[29] and phospholipids.[30] Such groups can also serve to reduce degradation if attached at sites which prevent nuclease activity.

An alternative approach to improvement of delivery to the cell is to increase the lipophilic character of the oligonucleotide so that it will be taken up more effectively by

passive diffusion. This can be achieved by elimination of the negative charge as with methyl phosphonates,[22] by attachment of lipophilic or amphiphilic groups such as cholesterol,[31] fatty acids or polyethylene glycol,[32] or by attachment of a more lipophilic base.

If it is desirable to permanently disable the target protein, such as, for example, an oncogene-derived target, attachment of a cross-linking adduct to the oligonucleotide would provide a viable approach. The potential advantage of being able to permanently disable the target protein would be partially offset by the disadvantage that cross-linking agents are generally electrophilic in nature so that side reactions with other cellular components might contribute to unwanted toxicity. Oligonucleotides conjugated with alkylating agents, such as psoralen,[33] platinum complexes,[34] proflavine[35] and ethyleneimine[36] have the potential to permanently inactivate their RNA targets. This strategy could also be exploited in the Protein Epitope Targeting approach.

The design of Protein Epitope Targeting-selected oligonucleotide analogs does not suffer from the need to preserve duplex formation via hydrogen bonding of complementary bases as is required for antisense oligonucleotides, and it is possible that with Protein Epitope Targeting, hydrogen bonding via specific positions of the purine and pyrimidine bases may not be the primary mode of interaction with target. Thus, at positions where the nature of the base or sugar is not critical, modifications can be introduced to provide for increased stability, lipophilicity or both. Alternatively, the base or sugar can be completely removed and replaced by a substitute which serves primarily as a linker group to bridge the residues required for binding.

Conclusions

The molecular diversity that can easily be achieved with oligonucleotides, together with the ease with which active oligonucleotide species can be amplified and identified, gives this class of compounds unique advantages in high-throughput screening. It is too early to predict how widespread the utility of oligonucleotides as drugs will be. However, Protein Epitope Targeting describes a generic approach to drug discovery in virtually every therapeutic area and the kinds of chemical modification described above could be universally useful in improving the pharmacokinetic properties of active oligonucleotides.

References and Notes

1. Lam, K.; Salmon, S.E.; Hersh, E.M.; Hruby, V.J.; Kazmierski, W.M.; Knapp, R.J. *Nature* **1991**, *354*, 82.
2. Scott, J. K.: Smith, G.P. *Science* **1990**, *249*, 386.
3. Devlin, J.J.; Panganiban, L.C.; Devlin, P.E. *Science* **1990**, *249*, 404.
4. Mullis, K.B.; Faloona, F.A. *Methods Enzymol.* **1987**, *155*, 335.
5. Oliphant, A.R.; Nussbaum, A.L.; Struhl, K. *Gene* **1986**, *44*, 177.
6. Oliphant, A.R.; Struhl, K. *Nucl. Acid Res.* **1988**,*16*, 7673.
7. Oliphant, A.R.; Brandl, C.J.; Struhl, K. *Mol. Cell Biol.* **1989**, *9*, 2944.
8. Kinzler, K.W.; Vogelstein, B. *Nucl. Acid Res.* **1989**, *17*, 3645.
9. Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
10. Ellington, A.D.; Szostak, J.W. *Nature* **1990**, *346*, 818.
11. Mavrothalassitis, G.; Beal, G.; Papas, T.S. DNA Cell. Biol. **1990**, *9*, 783.
12. Thiesen, H.-J.; Bach, C. *Nucl. Acid Res.* **1990**, *18*, 3203.
13. Blackwell, T.K.; Weintraub, H. *Science* **1990**, *250*, 1104.
14. Rauscher, F.J.; Morris, J.F.; Tournay, O.E.; Cook, D.M.; Curran, T. *Science* **1990**, *250*, 1259.
15. Tsai, D.E.; Harper, D.S.; Keene, J.D. *Nucl. Acid Res.* **1991**, *19*, 4931.
16. Ueda, H.; Hirose, S. *Nucl. Acid Res.* **1991**, *19*, 3689.
17. Howe, K.M.; Watson, R.J. *Nucl. Acid Res.* **1991**, *19*, 3913.
18. Kern, S.E.; Kinzler, K.W.; Bruskin, A.; Jarosz, D.; Freidman, P.; Prives, C.; Vogelstein, B. *Science* **1991**, *252*, 1708.
19. Blackwell, T.K.; Kretzner, L.; Blackwood, E.M.; Eisenman, R.N.; Weintraub, H. *Science* **1990**, *250*, 1149.
20. Bock, L.C.; Griffin, L.C.; Latham, J.A.; Vermaas, E.H.; Toole, J.J. *Nature* **1992**, *355*, 564.
21. Shaw, J-P.; Kent, K.; Bird, J.; Fishback, J.; Froehler, B. *Nucl. Acid Res.* **1991**, *19*, 747.
22. Miller, P. S.; Ts'o, P. O. P. *Anti-Cancer Drug Design* **1987**, *2*, 117.
23. Miller, P. S.; Braiterman L. T.; Ts'o, P. O. P. *Biochemistry* **1977**, *16*, 1988.
24. Froehler, B.; Ng, P.; Matteucci, M. D. *Nucl. Acid Res.* **1988**, *16*, 4831.
25. Matsukura, M.; Zon, G.; Egan, W.; Shinozuka, K.; Stein, C. A.; Mitsuya, H.; Cohen, J. S.; Broder, S. *Gene*, **1988**, *72*, 343.
26. Loke, S. L.; Stein, C. A.; Zhang, X. H.; Mori, K.; Nakanishi, M.; Subasinghe, C.; Cohen, J. S.; Neckers, L. M. *Proc. Natl. Acad. Sci. USA* **1989**, *86*,3474.
27. Eritja, R.; Pons, A.; Escarceller, M.; Giralt, E.; Albericio, F. *Tetrahedron* **1991**, *47*, 4113.
28. Haralambidis, J.; Duncan, L.; Tregear, G. W. *Tetrahedron Lett.* **1987**, *28*, 5199.
29. Lemaitre, M.; Bayard, B.; Lebleu, B. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 648.
30. Ramirez, F.; Mandal, S. B.; Marecek, J. F. *J. Amer. Chem. Soc.* **1982**, *104*, 5483.
31. Stein, C. A.; Pal, R.; DeVico, A. L.; Hoke, G.; Mumbauer, S.; Kinstler, O.; Sarngadharan, M. G.; Letsinger, R. L. *Biochemistry* **1991**, *30*, 2439.
32. Tullis, R. H. **1990**, *U.S. Patent* 4,904,582.
33. Kean, J. M.; Murakami, A.; Blake, K. R.; Cushman, C.; Miller, P. S. *Biochemistry* **1988**, *27*, 9113.
34. Chu, B. C. F.; Orgel, L. E. *Nucl. Acid Res.* **1989**, *17*, 4783.
35. Praseuth, D.; Le Doan, T.; Chassignol, M.; Decout, J-L.; Habhoub, N.; Lhomme, J.; Thuong, N. T.; Helene, C. *Biochemistry* **1988**, *27*, 3031.
36. Webb, T. R.; Matteucci, M. D. *Nucl. Acid Res.* **1986**, *14*, 7661.