MASTER

UR-3490- 1581

Hybridization with Synthetic Oligonucleotides

Szostak, J.W.[†], Stiles, J.I.[‡], Tye, B.-K.[†], Sherman, F.[‡], and Wu, R.[†]

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

[†] Section of Biochemistry, Molecular and Cell Biology
Cornell University, Ithaca, NY 14853

[‡] Department of Radiation Biology and Biophysics
University of Rochester, School of Medicine, Rochester, NY 14642

EB

Recent advances in chemical¹ and enzymatic² synthesis of oligonucleotides have greatly increased the availability of these compounds. Oligonucleotides ten to twenty bases long are potentially useful as hybridization probes for the detection of unique genes in Southern blot filter hybridization experiments³ and for the screening of colony or bacteriophage banks for particular sequences. The amino acid sequence of many interesting proteins has been determined. From this information it is possible to deduce a partial nucleotide sequence for the corresponding mRNA or gene. The degeneracy of the genetic code results in ambiguity at the second base of some codons and at the third base of most codons. This effect can be minimized by selecting a region of the protein sequence consisting predominantly of unique codons (met, trp) and the other least ambiguous codons (asp, asn, cys, his, phe, tyr, glu, gln, lys). In these cases, the uncertainty is between A and G or T and C, and the effect of possible mismatches is minimized by selecting G for A/G ambiguity and T for T/C ambiguity in the oligonucleotide⁴. This results in either correct base pairing or a G=T mismatch. This type of mismatch is expected to cause less destabilization of the helix than any other mismatch⁵.

In this paper we describe procedures for the use of synthetic oligonucleotides for Southern blot experiments and gene bank screening, and demonstrate the effect of various mismatches on the efficiency of hybridization.

Sensitivity vs Specificity

To use synthetic oligonucleotide probes for hybridization they must first be end-labeled, and then annealed with single-stranded DNA bound to a nitrocellulose filter. The temperature should be 15-20° C below the

estimated T_m of the hybrid; in practice the conditions of the hybridization reaction must be carefully optimized in order to achieve high sensitivity and specificity. The specificity of the probe is determined by its length (and therefore the number of times its complementary sequence occurs in the DNA being probed), and by the stringency of the reaction conditions. If the conditions are insufficiently stringent, the probe will hybridize with many closely related sequences. However, the efficiency of the hybridization reaction declines as the reaction conditions are made more stringent. A balance must therefore be found between the opposing requirements of sensitivity and specificity.

Hybridization with a 12 nucleotide-long fragment (12-mer), with one mismatch, is sufficiently sensitive for the detection of correct binding to a restriction digest of λ DNA. We have tested the effects of several different mismatches and find that errors near the middle of the sequence are less critical than errors near either end.

An oligonucleotide of 13-15 nucleotides in length is sufficient for the detection of a unique gene in total yeast DNA by Southern blot analysis³, and a 15-mer can be used in the screening of a yeast DNA bank cloned in a λ vector. Conditions must be optimized and the specificity determined by Southern blot experiments before plaque hybridization is done. However, analysis of mammalian DNA by Southern blots may not be possible due to the greater complexity of this DNA, although plaques could still be detected. The use of single-stranded DNA phages such as certain M13 derivatives as cloning vectors may considerably enhance the usefulness of oligonucleotide hybridization in gene bank screening because of the greater sensitivity attainable with these vectors.

Materials

Oligonucleotides: d(A-G-C-A-C-C-T-T-T-C-T-T-A-G-C), complementary to bases 24-39 of the yeast iso-1 cytochrome c mRNA, was synthesized in our laboratory⁶ by the phosphotriester method.

d(G-A-G-C-G-G-A-T-A-A-C-A-A-T-T), and d(A-A-T-T-G-T-G-A-G-C-G-G), complementary to the E. coli lac operator, were a gift from Dr. S.A. Narang.

5' end-labeling was carried out with T4 polynucleotide kinase (New England Biolabs, Inc.) and [γ -³²P]-ATP (New England Nuclear, Inc.) at a specific activity of 1,000-3,000 Ci/mmol⁷.

DNA: Yeast DNA was prepared as described⁸ from the following strains:

D311-3A (a his1 trp2 lys2), used as a standard CYC1 (wild type) strain.

B-955 (a CYC1-91-A his1 trp2 lys2), has CAA in place of GAA^{13a}. and thus lacks the EcoRI site at bases 3-9.

B-2185 (a CYC1-183-AD his1 trp2 lys2), has a deletion encompassing amino acids 8 through 12¹⁴, the binding site of the 15-mer.

ϕ 80 dlac O^c mutants RV1, RV10, RV17, RV51, RV116 and RV120 were a gift of Dr. J. R. Sadler⁹.

Restriction enzymes were purchases from New England Biolabs, Inc.

Hybridization: DNA fragments were transferred to nitrocellulose paper (Schleicher and Schuell, Inc.) by the blotting procedure of Southern³ or the plaque transfer procedure of Benton and Davis¹⁰. All hybridizations were carried out in 2xSSC, with 0.2% polyvinylpyrrolidone, Ficoll and BSA (Sigma, Inc.) in sealed plastic bags.

Hybridization of a 12-mer to the λ endolysin gene

The amino acid sequence of the λ endolysin gene is known¹¹ and contains a region with low nucleotide sequence ambiguity. We¹² synthesized a 12-mer complementary to this sequence, with four possible ambiguities

of the G=T type. Hybridization to exonuclease III-digested λ DNA was observed and primer-extension and DNA sequencing work showed that hybridization to the correct site had occurred. It was not possible, however, to determine how many mismatched base pairs, if any, were actually present.

Hybridization of oligonucleotide probes to the *E. coli* *lac* operator

The effect of different types of mismatches in different positions on the efficiency of oligonucleotide hybridization was tested with a series of O^c mutants of the *E. coli* *lac* operator. G=A, A=C and G=T mismatches near the middle or ends of the complementary sequence were studied. The $\phi 80$ d *lac* O^c mutants were obtained from Dr. J.R. Sadler. Each of the *lac* O^c mutants is denoted by its strain number and the position and base change of each mutation¹³ is noted below and in Figure 1. A synthetic 15-mer was hybridized with the wild type and two mutants with the following base changes:

5' G-A-G-G-G-A-T-A-A-C-A-A-T-T 3'	: 15 mer
-- C-T-C-G-C-C-T-A-T-T-G-T-T-A-A ---	: λ chromosome
RV10 RV17	: mutation
G A	
A C	: mismatch

DNA from the $\phi 80$ d*lac* O^c mutants was isolated and digested with the restriction enzyme *Eco* R1. The resulting fragments were separated by agarose gel electrophoresis and transferred to nitrocellulose filters by the blotting technique of Southern³. The synthetic probe was labelled with [γ -³²P]-ATP using T4 polynucleotide kinase, and then hybridized

to the DNA fragment containing the lac operator sequence.

Depending on the position and nature of the mutation, the stability of the hybrid varied. A G=A mismatch at the terminal nucleotide of the 15-mer (RV10, above) led to a large decrease in hybridization efficiency relative to the wild type (no mismatch) control at 37⁰, and to a lesser decrease at 31⁰. An A=C mismatch at the penultimate nucleotide (RV17, above) results in a similar large decrease in efficiency at 37⁰, while at 31⁰ the efficiency is comparable to that of the wild type strain. This suggests that the destabilizing effect of a mismatch is enhanced at higher temperatures when the stability of the helix is marginal even without mismatches, i.e. a decrease in the T_m of the hybrid was observed with the introduction of base-pairing errors.

The fact that hybridization was detectable at 37⁰ with a mismatch in the penultimate base of a 15-mer (perhaps equivalent to the use of a 13-mer) suggested that smaller oligonucleotides might be suitable as hybridization probes for the Southern blot experiments. A synthetic dodecanucleotide, d(A-A-T-T-G-T-G-A-G-C-G-G) was hybridized with 1-2 µg of Eco R1 digested φ80 dlac DNA (wild type and mutants), at 30⁰ C in 2xSSC. The results of this experiment are shown in Fig. 1. A G=T or A=C base pair mismatched at the fifth or eighth position did not affect the hybridization of the probe under these conditions. However, a G=A mismatch at position seven rendered the double helix very unstable, giving a barely detectable signal for hybridization to the lac operator. Likewise, a G=T or A=C mismatch at positions 10 or 11 disrupted the helix formation, and a G=A mismatch at position 9 was also not tolerated.

These data suggest that an internal G=T or A=C mismatch has a

less disruptive effect on the conformation of the double helix than a similar mismatch near the end of the sequence, or an internal G=A mismatch. A decline in the specificity of the hybridization reaction with the 12-mer relative to the 15-mer is also evident. Some hybridization of the 12-mer with another fragment of DNA is seen with all of the strains, whereas the 15-mer was entirely specific to the lac operator fragment. Even with an internal mismatch, a 12-mer binds strongly enough to its complementary sequence to be detected by filter hybridization.

Hybridization of synthetic probes to the CYC1 gene of yeast

The sequence of the first 44 nucleotides of the yeast cycl gene was determined by Stewart and Sherman¹⁴ by amino acid sequence analysis of frameshift mutants. A 15-mer complementary to part of this sequence was synthesized for use as a hybridization probe by Szostak et al.,⁶. This 15-mer has been used in two ways to detect the cycl gene in Southern blot experiments. First, the 15-mer was used to make cDNA complementary to the cycl mRNA as previously described⁶. The labeled cDNA is a highly specific and sensitive probe, but is difficult to prepare and available in only small amounts. Second, end labeled 15-mer was used directly as a hybridization probe. While this probe is easily prepared and is sufficiently specific, the efficiency of the hybridization is low and a large excess of labeled probe must be used.

Yeast DNA from a wild type strain has an Eco RI site at bases 3 to 9 of the cycl gene (amino acids 2 to 3 of the iso-1 cytochrome c), whereas DNA from a strain carrying the CYC1-9¹-A mutation does not have this Eco RI site. Since the strains are isogenic, we expect that the cycl Eco RI fragment will be larger in the mutant than the wild type,

while all other Eco R1 fragments will be of the same size. DNA from the two strains was digested with excess Eco R1, electrophoresed on a 1% agarose gel and transferred to nitrocellulose paper. The filter was annealed with cycl cDNA (50-150 bases long) at 65°, washed and autoradiographed (see Fig. 2). Hybridization was observed with DNA fragments of 4.45 Md from the wild type strain, and 6.2 Md in the mutant. The size shift conclusively identifies these fragments as carrying the cycl gene. The 6.2 Md fragment is the fusion product of the 4.45 and 1.75 Md fragments from the wild type strain.

In a separate experiment, hybridization of end-labeled 15-mer alone was observed to the 1.75 Md fragment. This fragment did not appear with DNA from a strain carrying the CYC1-183-AD deletion which exactly covers the 15-mer binding site. The 1.75 Md fragment must be on the 3' end of the 6.2 Md fragment, and is the fragment carrying the structural gene for iso-1 cytochrome c.

Hybridization of labeled 15-mer to Eco R1, Bam HI and Eco R1 plus Bam HI-digested wild type yeast DNA is illustrated in Fig. 3. In each case there is a major and a minor band. The major band in the Eco R1 digest is the 1.75 Md fragment carrying the cycl gene. The minor band is probably due to a DNA fragment carrying a closely related sequence that also binds the 15-mer.

The hybridization was repeated at several temperatures between 25° and 50° C. The optimum for sensitivity and specificity was in the 42° - 45° C range. The efficiency of the hybridization declined rapidly above 45° C; at lower temperatures many additional bands were seen. Only one minor band remained at 44° C.

9

We have used the labeled 15-mer to screen a λ phage bank of Eco R1 fragments of total yeast DNA. With the conditions optimized as described above, and using the plaque transfer procedure¹⁰, we observed a clear pattern of hybridization to individual plaques, however the sensitivity was greatly enhanced by spotting the plaques to a grid before proceeding with the plaque transfer. In addition to the regular pattern, plaques prepared in this manner gave stronger signals and were larger in size making them more distinct from the background (see Fig. 4). Analysis of these clones has shown both the cycl gene, and the cloned fragment responsible for the minor band seen in Fig. 3 (manuscript in preparation).

Hybridization of a synthetic 13-mer to restriction enzyme digests of total yeast DNA has been reported by Montgomery et al.,¹⁵. They found hybridization to seven different fragments, including the cycl gene. The correct band was identified with the Eco R1 site mutant described above. A cycl clone was also identified by hybridization with a 13-mer. DNA of the correct size was isolated from a gel and used to prepare a phage λ bank. The plaques were transferred to a grid and screened by hybridization with labeled 13-mer. Hybridization was detected to a clone carrying the cycl gene.

Problems with filter hybridizations

A common technical problem encountered in filter hybridization is a heavy background of dark spots on the film that can obscure the desired bands. This problem becomes severe with the very high levels of radioactive probe used for oligonucleotide hybridizations. The background can be minimized by using the lowest possible concentration of probe (usually $10^5 - 10^6$ cpm/ml is sufficient), and by doing the hybridization in 0.2% polyvinylpyrrolidone, Ficoll and bovine serum albumin. It is also important that the filter be thoroughly wetted before the labeled probe is added. Nitrocellulose filters wet poorly in high salt; the best procedure is to wet the filters in distilled water. They are then

sealed in plastic bags and incubated briefly in hybridization buffer at 65° after which the labeled probe is added (R. Rothstein, personal communication).

Cloning eucaryotic genes

The yeast genome contains 2×10^7 base pairs of DNA. A sequence of 12 nucleotides is likely to be unique. Since neither a 13-mer (7 bands) nor a 15-mer (2 bands) hybridized to a single site, the hybridization conditions were probably not sufficiently stringent to eliminate binding to closely related sequences. The complexity of mammalian DNA requires a 15-16 nucleotide-long sequence to be unique, and probably an 18-20 nucleotide sequence is more ^edesirable to avoid unwanted hybridization to related sequences. This would be difficult to derive from the amino acid sequence of a protein. An alternative may be to screen a phage bank with two shorter probes of 11-14 nucleotides in length, and to search for clones to which both short probes bind.

Recently the phage M13 has been modified so as to be a useful cloning vector¹⁶. This phage produces plaques containing up to 10^9 phage particles. Plaque hybridization with this system is approximately 100 times more efficient than with λ plaques (R. Rothstein, personal communication). This property may greatly enhance the usefulness of synthetic oligonucleotides in screening phage banks. Probes short enough to be derived from amino acid sequences may be sufficiently long to have the sensitivity required for M13 plaque hybridization.

Acknowledgements

This work was supported in part by NSF research grant 77-20313 awarded to RW, NSF grant PCM78-02341 and NIH grant AI 14980-01 awarded to BKT, NIH grant GM12702 awarded to FS, and NIH Postdoctoral Fellowship GM05441-02 to JIS and in part by the U.S. Department of Energy at the University of Rochester, Department of Radiation Biology. This paper has been designated Report No. UR-3490-1581.

Figure Legends

Figure 1. Hybridization of a dodecanucleotide to the E. coli lac operator region in $\phi 80$ dlac. Hybridization (at arrow) is visible with the wild type sequence, and with an internal A=C mismatch, but not with an internal G=A mismatch, or any mismatch near the end of the probe. Insert shows a photograph of the ethidium bromide stained gel.

Figure 2. Hybridization of polymerase-extended 15-mer (cycl cDNA) to yeast DNA from wild type (A) and from a mutant lacking the E. coli RI site (B). Loss of the Eco RI site results in the fusion of two adjacent fragments to give one larger fragment (6.2 Md) seen in the mutant.

Figure 3. Hybridization of labeled 15-mer to restriction enzyme digests of yeast DNA. The lower band in the Eco RI digest is at 1.75 Md and is due to hybridization to the cycl gene.

Figure 4. Hybridization of labeled 15-mer to a λ phage bank of Eco RI fragments of total yeast DNA. Plaques were transferred to a grid such that there are approximately 200 plaques per plate. The three common spots on all filters are a previously isolated clone containing the minor band seen in the Eco RI track of Fig. 3. The solid arrows indicate positive plaques, the open arrow nonspecific binding.

References

1. R. Wu, C.P. Bahl, and S.A. Narang (1978) in Progress in Nucleic Acids Research and Molecular Biology, Vol. 21, p. 101-104, ed: Academic Press, N.Y.
2. S. Gillam, R. Jahnke, and M. Smith (1978) J. Biol. Chem. 253, 2532-2539.
3. E. Southern (1975) J. Mol. Biol. 98, 503-517.
4. R. Wu (1972) Nature New Biology 236, 198-200.
5. O.C. Uhlenbeck, F.H. Martin, and P. Doty (1971) J. Mol. Biol. 57, 217-229.
6. J.W. Szostak, J.I. Stiles, C.P. Bahl, and R. Wu (1977) Nature 265, 61-63.
7. A. Maxam, and W. Gilbert (1977) Proc. Natl. Acad. Sci. USA
8. D.R. Cryer, R. Eccleshall, and J. Marmur (1975) in "Methods in Cell Biology", Vol., XII ed. D.M. Prescott, p. 39, Academic Press, N.Y.
9. A. Jobe, J.R. Sadler, and S. Bourgeois (1974) J. Mol. Biol. 85, 231-248.
10. W.D. Benton, and R.N. Davis (1977) Science 196, 205-208.
11. M. Imada, and A. Tsugita (1971) Nat. N. Biol. 233, 230.
12. R. Wu, C.D. Tu, and R. Padmanabhan (1973) Bioch. Biophys. Res. Comm. 55, 1092-1099.
13. W. Gilbert, J. Gralla, J. Majors, and A. Maxam (1975) Protein-Ligand Interactions (Ed., by H. Sund, and G. Blauer) Walter de Gruyter, Berlin, N.Y.
14. J.W. Stewart, and F. Sherman (1974) in Molecular and Environmental Aspects of Mutagenesis (ed. by L. Pradash, F. Sherman, M.W. Miller, C.W. Lawrence, H.W. Taber) 102-127 (Charles C. Thomas, 1974).
15. D.L. Montgomery, B.D. Hall, S. Gillam, and M. Smith (1978) Cell 14, 673-680.
16. B. Gronenborg, and J. Messing (1978) Nature 272, 375-377.
- 13a. F. Sherman and J.W. Stewart (1973) in "The Biochemistry of Gene Expression in Higher Organisms" ed. J.K. Bollak and J.W. Lee, p. 56-86 Australian and New Zealand Book Co. PTY., LTD., Sydney Australia.

Fig. 1

D311-3A B955

-6.0 Md

-4.45Md

Fig. 2

Fig. 3