The Copper-Zinc Superoxide Dismutase Gene from *Drosophila melanogaster*: Attempts to clone the gene using two mixed sequence oligonucleotide probes

by

Nina O.L. Seto

B. Sc.(Hon) University of Toronto, 1984

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
Department of Biochemistry

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

February 1987
In presenting this thesis in partial fulfilment of the requirements for an advanced
degree at the University of British Columbia, I agree that the Library shall make it
freely available for reference and study. I further agree that permission for extensive
copying of this thesis for scholarly purposes may be granted by the head of my
department or by his or her representatives. It is understood that copying or
publication of this thesis for financial gain shall not be allowed without my written
permission.

Department of Biochemistry

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date March 18, 1987.
Superoxide dismutase is an enzyme which scavenges superoxide radicals and is thought to be a longevity determinant, as there exists a positive correlation between superoxide dismutase concentration and maximum life span potential. The cytosolic CuZn superoxide dismutase in *D. melanogaster* has been purified and sequenced, but the gene has not been cloned. However, when it is available the CuZn SOD gene may be reintroduced into the *Drosophila* genome via the P-element transformation system so its effects on the life span potential of *Drosophila* may be studied. This study describes attempts to clone the CuZn SOD gene from *D. melanogaster* using two mixed sequence oligonucleotide probes. The S1 probe corresponds to amino acids 43-48 of the protein sequence and contains 128 different oligonucleotide sequences representing all possible codon combinations predicted from the amino acid sequence. The GT3 probe is targeted to amino acids 90-95 of the protein. In this probe, deoxyguanosine was placed in positions where all four nucleotides may occur to decrease probe heterogeneity. The probes were used to screen *D. melanogaster* Canton-S and Oregon-R genomic lambda libraries. Three positive clones isolated from the Canton-S library had identical nucleotide sequence in the GT3 probe binding region, and sequencing of the probe binding site revealed that one member of the GT3 probe had formed a 15 bp duplex with the phage DNA. Screening of the
Oregon-R library produced four clones which hybridized with both GT3 and S1 probes. When these phage DNA were hybridized to polytene chromosomes by in situ hybridization, none mapped to 68AB on the third chromosome, the location of the CuZn SOD gene. These results suggest that modification of the classical strategy used in this study is necessary, and implications on probe design are discussed.
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>vii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>12</td>
</tr>
<tr>
<td>Materials</td>
<td>12</td>
</tr>
<tr>
<td>Bacterial strains</td>
<td>12</td>
</tr>
<tr>
<td>D. melanogaster genomic DNA preparation</td>
<td>13</td>
</tr>
<tr>
<td>Radioactive labelling of oligonucleotides</td>
<td>14</td>
</tr>
<tr>
<td>Screening genomic libraries by plaque hybridization</td>
<td>15</td>
</tr>
<tr>
<td>I. Primary screen</td>
<td>15</td>
</tr>
<tr>
<td>II. Secondary screen</td>
<td>16</td>
</tr>
<tr>
<td>Bacteriophage lambda DNA preparation</td>
<td>17</td>
</tr>
<tr>
<td>I. Large scale preparation</td>
<td>17</td>
</tr>
<tr>
<td>II. Lambda phage DNA isolation from primary screen</td>
<td>18</td>
</tr>
<tr>
<td>III. Lambda phage DNA isolation from secondary screen</td>
<td>18</td>
</tr>
<tr>
<td>Agarose gel electrophoresis and Southern hybridization</td>
<td>18</td>
</tr>
<tr>
<td>I. Restriction enzyme digests</td>
<td>18</td>
</tr>
<tr>
<td>II. Agarose gel electrophoresis</td>
<td>19</td>
</tr>
<tr>
<td>III. Southern transfer and hybridization</td>
<td>19</td>
</tr>
<tr>
<td>In situ hybridization to polytene chromosomes</td>
<td>20</td>
</tr>
<tr>
<td>I. Labelling the probe</td>
<td>20</td>
</tr>
</tbody>
</table>
II. In situ hybridization................................. 21
Subcloning phage DNA into plasmids....................... 21
I. Ligation reaction........................................ 21
II. Plasmid and M13 transformations....................... 21
III. Growth of transformants.............................. 22
DNA sequence determination.............................. 23
I. Template preparation................................ 23
II. DNA sequencing...................................... 23
RESULTS................................................. 26
The stringency of hybridization........................ 26
Primary screen of Canton-S lambda library............... 29
Secondary screen of Canton-S lambda library............ 29
DNA sequence analysis of GT3 binding site (CS-1).... 29
Primary screen of Oregon-R lambda library............. 37
Analysis of Oregon-R primary screen phages............ 37
Secondary screen of Oregon-R lambda library.......... 38
Mapping of phage DNA to polytene chromosomes.......... 43
The GT3 probe binding site of phage 109............... 43
DISCUSSION............................................ 50
I. Oligonucleotide probe heterogeneity................ 50
II. Oligonucleotide probe sequence..................... 51
III. Oligonucleotide probe length....................... 53
IV. Oligonucleotide probe design......................... 54
REFERENCES........................................... 57
LIST OF FIGURES

FIGURE 1. The amino acid sequence of Cu/Zn Superoxide dismutase for D. melanogaster .......................... 5

FIGURE 2. The mixed sequence oligonucleotide probes S1 and GT3.................................................. 8

FIGURE 3. Oregon-R genomic and lambda library DNA Southern analysis............................................. 27

FIGURE 4. Southern analysis of DNA from GT3 positive phage from the Canton-S genomic lambda library... 30

FIGURE 5. DNA sequencing strategy of the GT3 binding site of the Canton-S library GT3 positive phage... 33

FIGURE 6. The nucleotide sequence and translation of the GT3 binding site of phage CS-1, 2, and 3 from the Canton-S library................................................................. 35

FIGURE 7. Southern analysis of Oregon-R primary screen phages.......................................................... 39

FIGURE 8. Southern analysis of phage from the Oregon-R library that hybridizes both GT3 and S1 probes... 41

FIGURE 9. Determination of the chromosomal location of phage 109 and phage CS-2 by in situ hybridization to polytene chromosomes................................. 44

FIGURE 10. Nucleotide sequence and translation of the GT3 binding site of phage 109 from the Oregon-R library................................................................. 48
LIST OF ABBREVIATIONS

ATP adenosine-5' triphosphate
bp base pair(s)
BSA bovine serum albumin
cpm counts per minute
d deoxy
dd dideoxy
dNTP deoxynucleoside-5' triphosphate
dATP deoxyadenosine-5' triphosphate
dCTP deoxycytidine-5' triphosphate
dGTP deoxyguanosine-5' triphosphate
dTTP thymidine-5' triphosphate
DNA deoxyribonucleic acid
DNase deoxyribonuclease
DTT dithiothreitol
EDTA ethylenediaminetetraacetate
EtBr ethidium bromide
IPTG isopropyl-β-D-thiogalactopyranoside
kb kilobase pair(s)
kDa kilodalton(s)
LB Luria broth
LBMgT LB media with 5 μg/ml thymidine, 10 mM MgSO₄
mA milliamperes
mM millimolar
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PFU</td>
<td>plaque forming units</td>
</tr>
<tr>
<td>RF</td>
<td>replicative form</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>ss</td>
<td>single stranded</td>
</tr>
<tr>
<td>Southern</td>
<td>the Southern transfer procedure: the transfer of DNA from an agarose gel to a membrane</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>U</td>
<td>unit(s)</td>
</tr>
<tr>
<td>uCi</td>
<td>microcurie</td>
</tr>
<tr>
<td>ug</td>
<td>micrograms</td>
</tr>
<tr>
<td>uM</td>
<td>micromolar</td>
</tr>
<tr>
<td>uv</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>volts</td>
</tr>
<tr>
<td>Xgal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I wish to thank my supervisor Dr. Gordon Tener for encouragement, support and guidance. I am indebted to Dr. Shizu Hayashi for countless hours of discussion, help, and guidance. I thank Dr. Hayashi for all her contributions to this work and for her inestimable patience. I thank Jeff Leung and Craig Newton for sharing their knowledge and ideas with me. Last but not least, I thank Dr. Ian Gillam for his discourses - so I may be "educated".
INTRODUCTION

Superoxide dismutases are enzymes which catalytically scavenge superoxide radicals (O$_2^-$). They are thought to be an essential component of the biological defense against oxidative damage mediated by superoxide radicals, which are produced as a by-product of oxygen metabolism (1-4).

Superoxide dismutases (SOD) are metalloproteins with either copper and zinc, manganese, or iron as ligands. Prokaryotes possess the two closely related FeSOD and MnSOD, whereas eukaryotes have a tetrameric MnSOD (MW 80-90kDa) in the mitochondrial matrix and an independently evolved dimeric (MW 31-33kDa) CuZn SOD in the cytosol (5,6).

Superoxide dismutase catalyzes the dismutation of two superoxide radicals into molecular oxygen and hydrogen peroxide:

\[
2O_2^- + 2H^+ \rightarrow O_2 + H_2O_2
\]

Catalase and/or peroxidase removes hydrogen peroxide by changing it to water:

\[
2H_2O_2 \rightarrow 2H_2O + O_2
\]

Any existing superoxide and hydrogen peroxide form the very reactive hydroxyl radical (•OH) via the iron-catalyzed Haber-Weiss reaction:

\[
O_2^- + H_2O_2 \rightarrow •OH + O_2 + OH^-
\]

The hydroxyl radical may initiate free radical reactions which result in lipid, protein and DNA damage (7). Despite the
unique nature of the superoxide substrate, a multiplicity of assays for measurement of superoxide dismutase activity has been devised (8-13).

The biological function of superoxide dismutase is as an antioxidant and its role as a possible longevity determinant has been proposed (14,15). Non-genetic theories of aging attribute damage of structural and cellular components of the cell to daily "wear and tear". The free radical theory of aging postulates that free radical damage increases with age and contributes to the biological changes observed with aging (16-19). Free radicals produced as transient intermediates in normal cellular metabolism may attack otherwise stable molecules and thus contribute to the observed cellular damage. It has been proposed that the aging process begins with dysdifferentiation or the loss of the proper differentiated state of the cells. The process of aging represents the response of the organism to these changes. There is evidence that antioxidants such as SOD protect against oxygen free radical damage which can contribute to dysdifferentiation and aging (14,15).

There exists a relationship between metabolic rate and aging in both poikilotherms and homeotherms. Mammals have stable metabolic rates and species-specific life spans. Life spans of poikilotherms are generally variable and determined by environmental conditions. Factors which reduce metabolic rate extend the life span of poikilotherms. Species with higher
metabolic rates (and increased oxygen utilization) have higher intracellular concentrations of superoxide radicals. A small fraction of superoxide radicals escape quenching despite elaborate enzymatic and non-enzymatic defenses against them. Therefore, antioxidants may be important in determining lifespan.

In an attempt to determine if age-related changes occur in antioxidant efficiency, superoxide dismutase concentration as a function of age has been determined for various species and results have been varied (20-26). In *D. melanogaster* mitochondrial MnSOD declined 21% between 5 and 58 days of age whereas cytosolic CuZn SOD remained relatively constant (27-29). However, a positive correlation exists between the tissue concentration of SOD per specific metabolic rate (SMR) and life span potential (LSP) for the 12 primate and 2 rodent species studied (30,31). Therefore, since SOD/SMR=k(LSP), LSP is directly related to the level of SOD in the tissue. The ratio of SOD per amount of oxygen consumed is constant for each species, and longer lived species must have a higher concentration of SOD.

There are many advantages to using *Drosophila* as an experimental animal in aging studies. *Drosophila* grow vigorously in the laboratory, are "old" at 40 days, and more is known about its genetics than any other higher eukaryote.

Elucidating the role of SOD in aging was the impetus behind cloning the CuZn and Mn SOD genes from *D. melanogaster*. The
cloned SOD genes may be reintroduced into the *Drosophila* genome via the P-element transformation system (32-34). P-transposable DNA injected into the early embryo integrates into the genome of germ line cells of the individual. P-element vectors carry the neomycin resistance gene which allows selection of transformants on the antibiotic G418 (35).

In order to clone the CuZn SOD gene a probe was needed and to synthesize the probe it was necessary to know the amino acid sequence of the protein. The eukaryotic CuZn SOD proteins sequenced are: human, bovine, porcine, horse, yeast, swordfish, and *Drosophila* (36-38). Comparison of these eukaryotic sequences shows conserved regions. The *Drosophila* CuZn SOD (32 kDa) constitutes 0.4% of the total soluble protein and has a specific activity 1.5 times higher than any other purified SOD (39-41). The protein consists of 151 amino acids and is 57% homologous to the bovine protein (Figure 1).

The strategy used to clone CuZn SOD was to use mixtures of synthetic oligonucleotides representing all possible codon combinations predicted from a 6 amino acid segment of the protein as a probe to identify the cloned DNA. Complex probe mixtures are undesirable but often inevitable due to the redundancy of the genetic code. Ideally, the amino acid sequence chosen should use the least number of codons possible. Methionine and tryptophan have unique codons, but are two of the rarest amino acids in proteins. All other amino acids are coded for by 2, 4 or even 6
FIGURE 1. The amino acid sequence of copper/zinc superoxide dismutase from *D. melanogaster* (36). The enzyme exists as a dimer, consisting of two identical subunits. Each subunit consists of 151 amino acids and the amino terminus of the protein is acetylated. Two electrophoretically distinguishable forms of SOD differ at position 96. The SOD(fast) variant (shown here) has Asp-96, whereas the SOD(slow) variant has Lys-96. (A=Ala, R=Arg, N=Asn, D=Asp, B=Asx, C=Cys, Q=Gln, E=Glu, Z=Glx, G=Gly, H=His, I=Ile, L=Leu, K=Lys, M=Met, F=Phe, P=Pro, S=Ser, T=Thr, W=Trp, Y=Tyr, V=Val).
<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>VVKAVCVING</td>
<td>DAKGVFFEQ</td>
<td>ESSGTPVKVS</td>
<td></td>
</tr>
<tr>
<td>GEVCGLAKGL</td>
<td>HGPHVHEFGD</td>
<td>NTNGCMSGP</td>
<td></td>
</tr>
<tr>
<td>HFNPGKEHG</td>
<td>APVDENRHLG</td>
<td>DLGNIEATGD</td>
<td></td>
</tr>
<tr>
<td>CPTKVNITDS</td>
<td>KITLFGADSI</td>
<td>IGRTVVHAD</td>
<td></td>
</tr>
<tr>
<td>ADDLGQGGHE</td>
<td>LSKSTGNAGA</td>
<td>RIGCGVIGIA</td>
<td></td>
</tr>
</tbody>
</table>

151

K
In principle, each sequence in a mixed oligonucleotide probe may elicit a positive signal. Therefore, if two probes targeted to different regions of the protein are constructed, then the probability that only one DNA fragment will be common to both oligonucleotide families is high.

The first probe named S1 is targeted to amino acids number 43-48 of the CuZn SOD sequence (Figure 2). The S1 probe was designed before the Cu/Zn SOD sequence from Drosophila was available. Sequence data from human, cow, horse and yeast show amino acids 42-48 are highly conserved, except for amino acid 47 which is either glutamic acid or glutamine (Figure 2). Therefore, both glutamic acid and glutamine codons had to be included in the S1 probe. Later, the D. melanogaster sequence was found to have a Glu-47, which is consistent with what was predicted for the S1 probe. Inclusion of every possible codon for amino acids 42-48 resulted in 128 different 17 base oligonucleotides, which together form the S1 probe.

The second probe GT3 was designed with the D. melanogaster sequence available, and is targeted to amino acids 93-95, a segment unique to the Drosophila Cu/Zn SOD. To include every possible codon, this segment would also require a family of 128 different oligonucleotides. Therefore, to decrease the degeneracy of the probe, deoxyguanosine(dG) was placed in positions 9 and 12 where all four nucleotides dA, dT, dC, dG are
FIGURE 2. The mixed sequence oligonucleotide probes S1 and GT3. The S1 probe corresponds to amino acids 43-48 of the CuZn superoxide dismutase protein. This region was conserved with the exception of position 47 where Glu-47 was found in yeast, horse and D. melanogaster and Gln-47 was found for human and bovine sequences. The S1 probe was designed to include both Glu and Gln at position 47, as the probe was designed before the D. melanogaster sequence was available. The S1 probe consists of 128 different sequences (17 nucleotides in length) and is the complement of the coding strand. The GT3 probe corresponds to residues 90-95, a variable segment unique to the D. melanogaster protein. The probe heterogeneity was decreased by placing dG at positions 9 and 12 (G*) where all four nucleotides may occur. The GT3 probe consists of 8 variants (17 nucleotides in length) and represents the coding strand.
**S1 PROBE**

<table>
<thead>
<tr>
<th>43</th>
<th>44</th>
<th>45</th>
<th>46</th>
<th>47</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHE</td>
<td>HIS</td>
<td>VAL</td>
<td>HIS</td>
<td>GLU</td>
<td>PHE</td>
</tr>
</tbody>
</table>

All possible codons

<table>
<thead>
<tr>
<th>UUU</th>
<th>CAU</th>
<th>GUG</th>
<th>CAU</th>
<th>CAA</th>
<th>UUU</th>
</tr>
</thead>
</table>

Coding strands

5' TTT CAT GTG CAT CAA TT 3'

S1 Probe (Complementary Strand)

5' TTT CAT GTG CAT CAA TT 3'

**GT3 PROBE**

<table>
<thead>
<tr>
<th>90</th>
<th>91</th>
<th>92</th>
<th>93</th>
<th>94</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASP</td>
<td>LYS</td>
<td>PRO</td>
<td>THR</td>
<td>LYS</td>
<td>VAL</td>
</tr>
</tbody>
</table>

All possible codons

<table>
<thead>
<tr>
<th>GAU</th>
<th>UGU</th>
<th>CCG</th>
<th>ACG</th>
<th>AAA</th>
<th>GUG</th>
</tr>
</thead>
</table>

GT3 Probe

5' GAT TGT CCG* ACG* AAA GT 3'
possible (Figure 2). The substitution of dG into these positions rests on the assumption that dG can base pair with dC, and that pairing with dT and dA is not destabilizing. With this modification, the degeneracy of the GT3 probe was reduced to 8 different 17 base oligonucleotides.

These oligonucleotides (GT3 and S1) may be radioactively labelled and used as direct hybridization probes to screen genomic DNA which has been cloned into lambda phage libraries (42,43). In principle, selective hybridization of the probes to their cognate targets is possible as internal mismatches exhibit a much lower melting temperature than perfect duplexes (44,45). These conditions have to be empirically determined as the dissociation temperature ($T_d$) of the oligonucleotide duplexes are highly sequence dependent and empirically determined equations such as $T_d(°C) = 4(G+C) + 2(A+T)$ can only be used as an estimate of the actual dissociation temperature (46,47).

The DNA of putative positives may be mapped to polytene chromosomes to determine whether the sequence originates from the same chromosomal location as the CuZn SOD gene. The presence of polymorphic variants of SOD in several natural populations of Drosophila led to the genetic mapping of the cytosolic SOD gene to 32.5 on the third chromosome (48). Production of cytosolic SOD defective mutants mapped the gene to 68AB on 3L (the left arm of the third chromosome) on the polytene chromosomes. Cytosolic SOD mutant lethality correlated with increased oxygen metabolism by
Drosophila at the time of eclosion (49).

Attempts to clone the Drosophila CuZn SOD gene by using the two mixed oligonucleotide probes (GT3 and S1) to screen genomic lambda libraries is the subject of this thesis.
MATERIALS AND METHODS

MATERIALS

*E. coli* JM101, LE392, and Q358 were provided by J. Leung. The *Drosophila melanogaster* Oregon-R genomic library was constructed in lambda EMBL3 by J. Leung in 1984. The Canton-S library was constructed in lambda Charon 4 and obtained from Maniatis (43). M13mpl8 and mpl9 were amplified from original P.L. Biochemical stocks. M13mp8 and pUC13 were provided by S. Hayashi. Oligonucleotides GT3, S1, and M13 universal primers were synthesized by T. Atkinson (M. Smith Laboratory, UBC). γ-[\(^{32}\)P]ATP (3000 Ci/mmol, 10 uCi/ul) and α-[\(^{32}\)P]dNTP (3000 Ci/mmol) were from Amersham Corp., and deoxy and dideoxyNTPs from P.L. Biochemicals. Restriction enzymes, T4 ligase, *E. coli* DNA polymerase I, *E. coli* DNA polymerase (Klenow fragment) were purchased from New England Biolabs, Bethesda Research Laboratories, Boehringer Mannheim, or Promega Biotech. and used interchangeably unless specified. *E. coli* RNA polymerase was from P.L. Biochemicals. Calf intestinal phosphatase was from Boehringer Mannheim, BSA and lysozyme from Sigma, agarose (Ultrapure) from BRL, acrylamide from Eastman Kodak Co., N,N'-Methylene bisacrylamide from Matheson Coleman and Bell, and TEMED from BioRad. Hybond-N (nylon) membranes in rolls and 82 mm discs were from Amersham Corp. The X-ray film (Curix RP1) was from Agfa-Gevaert.

BACTERIAL STRAINS
The hosts for *D. melanogaster* Canton-S and Oregon-R genomic libraries were LE392 and Q358, respectively. Both stocks were kept frozen in 15% glycerol. Q358 (in NZYM (56)) and LE392 (in LBMgT) were grown to stationary phase with 0.2% maltose to induce the lambda receptor (56). The cells were pelleted by spinning at 4000g (5 min.). The pellet was resuspended in 0.5 volumes of 0.01M MgSO$_4$ and stored at 4°C (viable for 1 week).

**D. MELANOGASTER GENOMIC DNA PREPARATION**

*D. melanogaster* high molecular weight genomic DNA was prepared essentially as described with the following modifications (42). One gram of frozen *D. melanogaster* (Oregon R stock from G. Tener) was placed in the bottom of a Dounce homogenizer and ground with 5 ml of sterile ice cold solution (10 mM Tris-HCl (pH 7.5), 60 mM NaCl, 10 mM EDTA, 0.15 mM spermine, 0.15 mM spermidine) and filtered through fine gauze or 2 layers of Miracloth (Calbiochem-Behring Corp.) into a sterile centrifuge tube. The nuclei were pelleted by centrifugation (12,000g, 10 minutes, 4°C) and the pellet and tube rinsed with 0.1 M NaCl, 20 mM Tris-HCl (pH 8), 10 mM EDTA and resuspended in 5 ml of the same buffer. Proteinase K (500 ug) was added and all the nuclei were lysed simultaneously by injecting 1 ml of 10% SDS into the solution. The solution was incubated for 3 hours at 50°C and 4 ml of the same buffer was added before phenol extraction. After lysis, mechanical shearing (by mixing, shaking, etc.) was kept to a minimum and solutions contained no more than 1 mg DNA/20 mls
of nuclear lysate during phenol extractions. The DNA solution was extracted with equal volumes of phenol/CHCl$_3$(1:1) gently by rolling on a wheel (15 min., room temperature). The tubes were centrifuged (10,000g, 5 minutes) to separate the phases. The extraction was repeated with phenol/CHCl$_3$ and then once with CHCl$_3$. The DNA was precipitated by adding cold 95% EtOH to the final aqueous phase and the DNA was removed with a glass rod. The DNA was air dried briefly and redissolved in 500 ul of TE 8 (10 mM Tris-HCl (pH 8), 1 mM EDTA). When necessary, heating at 65°C was used to redissolve the entire pellet.

**RADIOACTIVE LABELLING OF OLIGONUCLEOTIDE PROBES**

The 5' OH of the oligodeoxyribonucleotides were labelled with [${}^{32}$P] by polynucleotide kinase which transfers the radioactive phosphorus from $\gamma$-[${}^{32}$P]ATP. One hundred picomoles of GT3 or S1 were reacted with 250 uCi (80 pmoles) of $\gamma$-[${}^{32}$P]ATP and 25 units of polynucleotide kinase in 0.1 M Tris-HCl (pH 7.5), 20 mM MgCl$_2$, 0.2 mM EDTA, 0.2 mM spermidine, and 10 mM DTT for 45 minutes at 37°C. The reaction was terminated by heating at 65°C for 10 minutes. *E. coli* tRNA (250ug) was added as a carrier and the sample diluted with 200 ul of TE 8. The reaction mixture was loaded onto a 0.5 ml column of DEAE-cellulose and washed with 2 ml of TE 8. The unincorporated label was eluted with 8 ml of 0.2 M NaCl/TE 8 and the oligonucleotides were eluted in 3.0 ml of 1.0 M NaCl/TE 8. Total incorporated counts were routinely $2 \times 10^8$ cpm (30-40% incorporation). The probe was
hybridized to filters at 40 ul/cm² and no less than 1 x 10⁷
cpm/ml in 6 x SSC (1 x SSC = 0.15 M NaCl, 0.015 M NaCitrate (pH 7.2)), 50 mM sodium phosphate (pH 6.8), 5 x Denhardt's reagent
(1xDenhardt's reagent = 0.02% Ficoll, polyvinyl pyrrolidone, and
BSA), 0.5% SDS, and 20 ug/ml E. coli tRNA.

SCREENING GENOMIC LIBRARIES BY PLAQUE HYBRIDIZATION

I. Primary Screen

The titer of the Canton-S and Oregon-R lambda libraries was
determined by plating a series of 10-fold dilutions of the
original phage stock. The phage was added to 100 ul of 10 mM
MgCl₂/CaCl₂ and 100 ul of resuspended Q358. The phage and
bacteria were incubated for 15 minutes at 37°C, and plated with
3 ml of soft top agarose (0.7% agarose in NZYM at 55°C) on NZYM
plates. After the agarose hardened the plates were inverted and
incubated at 37°C for 10-14 hours.

Both libraries were screened by the in situ plaque
hybridization method developed by Benton and Davis (50). Each
library was plated on six plates with 10⁴ plaque forming
units (PFU) per 82mm plate and incubated until the plaques were
just touching but not confluent. The plates were chilled at
4°C for 1 hour before blotting onto nylon filters (Hybond-N).
The phage and DNA was adsorbed by placing two filters on each
plate sequentially, the first for 3 minutes and the second for 5
minutes. The filters were orientated by asymmetric ink spots,
and air dried. The phage and DNA were denatured and fixed by
placing the filters DNA side up on blotting paper soaked in 1.5 M NaCl/0.5 M NaOH then neutralized with 1.5 M NaCl/1.0 M Tris-HCl (pH 7.5) for 5 minutes each. After air drying the filters, the DNA was covalently fixed to the nylon membrane by irradiation with uv light (254nm, 4 minutes). The filters were washed in 6xSSC by gently rubbing with a gloved finger and then prehybridized (6xSSC, 10xDenhardt's reagent, 0.2% SDS) for at least 3 hours before hybridization with the GT3 probe for at least 12 hours at 37°C. The filters were washed in 6xSSC (2 changes of buffer for 30 minutes) at room temperature and then washed at higher stringency (52°C for Canton-S and 50°C for Oregon-R screens) twice for 15 minutes each. Filters were exposed to X-ray film for 4 days at -70°C with an intensifying screen. Agar plugs containing plaques from the region corresponding to a positive on the autoradiogram were picked (with the wide end of a sterile pasteur pipette) into 0.5 ml of SM buffer with 5 ul of CHCl₃.

II. Secondary Screen

In order to isolate a single positive phage from the positive agar plug picked in the high density screen, the plaques must be plated to a density of 200 to 400 PFU per plate and re-screened as described for the primary screens. Plaques were allowed to grow bigger for the secondary screens and the autoradiogram exposure time was decreased to 2 days. Positive spots aligned with a single plaque. This was picked into 200 ul of SM buffer with 5 ul of CHCl₃ (using the narrow end of a
pasteur pipette).

**BACTERIOPHAGE LAMBDA DNA PREPARATION**

**I. Large scale preparation**

To isolate DNA from lambda phage clones, 1.25 ml of resuspended Q358 was infected with about $1 \times 10^8$ PFU of phage diluted in 1.0 ml of 10 mM MgCl$_2$/CaCl$_2$. The phage was adsorbed for 15 minutes at 37°C and then used to inoculate 250 ml of prewarmed NZYM. The culture was shaken vigorously at 37°C until it lysed clear (3-8 hours, depending on titer). CHCl$_3$ (1 ml) was added and swirled slowly for 5 minutes before pelleting cell debris by centrifugation for 15 minutes at 8000 g. The supernatant was mixed with 0.15 volumes 5 M NaCl/0.3 volumes 50% PEG6000 and the phage precipitated for at least 12 hours at 4°C (51). The PEG-precipitated phage was pelleted by centrifugation (8000 g, 15 min., 4°C). The supernatant was poured off and the tubes were spun for another minute and the supernatant removed as cleanly as possible. The phage pellet was resuspended in 5 ml of DNaseI buffer (50 mM Tris-HCl (pH 7.5), 5 mM MgCl$_2$, 0.5 mM CaCl$_2$) and incubated at 37°C with 50 ul of DNaseI (1 mg/ml in 0.15 M NaCl, 50% glycerol) and 50 ul RNase A (10 mg/ml). After incubation for 3 hours the debris was pelleted by centrifugation at 12,000 g for 5 minutes and the phage supernatant was treated with 75 ul Proteinase K (10 mg/ml), 0.5 ml 10% SDS, 100 ul of 250 mM EDTA (pH 7.5) and incubated at 68°C for 2 hours. The DNA was purified by two phenol/CHCl$_3$(1:1)
extractions and one CHCl₃ extraction. The DNA was concentrated by EtOH precipitation and resuspended in 250 ul TE 8 (Yield 200-250 ug).

II. Lambda phage DNA isolation from primary screen phages

Preparation of phage DNA from the primary screen plugs was exactly as described above except 100 ul of host Q358 was infected with 50 ul of the primary plug phage solution and used to inoculate a 20 ml NZYM culture. The precipitated DNA was redissolved in 50 ul of TE 8 and the concentration checked on an agarose gel.

III. Lambda phage DNA isolation from secondary screen phages

Individual phages isolated from the secondary screen were also grown in 20 ml NZYM cultures. These cultures were inoculated with 140 ul of phage suspension and 100 ul of Q358. The DNA was prepared as described above.

AGAROSE GEL ELECTROPHORESIS AND SOUTHERN HYBRIDIZATION

I. Restriction enzyme digests

Restriction digests were performed using 10-50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 1 mM DTT, 100 ug/ml BSA, and 0, 50, or 100 mM NaCl as required by the enzyme. The DNA was digested with 2-fold excess of enzyme for 3 hours in a 100 ul volume and 5 ul was loaded onto a mini agarose gel to check for completeness of digestion before the rest of the digest was precipitated by EtOH, resuspended in TE 8, and loaded onto a large gel for transfer onto membranes.
II. Agarose gel electrophoresis

Agarose gel electrophoresis was in 90 mM TBE (90mM Tris, 90 mM boric acid, 1 mM EDTA, pH 8.3) and gels varied from 0.5-1.2% agarose (with 1 ug/ml EtBr). The gels were usually run at 2.5V/cm for 12 to 14 hours. Digested genomic DNA (40 ug) was resuspended in 40 ul TE 8 and heated at 65°C for 5 minutes before loading onto a gel. Primary screen phage DNA (10 ug) and individual phage DNA (5 ug) were redissolved in 15 ul TE 8. Gel loading buffer (40% sucrose, 0.1% bromphenol blue, 0.05% xylene cyanol, 0.1% SDS, 1 mM EDTA) was added before loading (5 ul).

III. Southern transfer and hybridization

The agarose gels were denatured 2x10 minutes in 1.5 M NaCl/0.5 M NaOH and neutralized 2x10 minutes in 1.5 M NaCl/1.0 M Tris-HCl (pH 7.5) at room temperature with shaking (52). The DNA was then transferred to nylon membranes (Hybond-N) with 20xSSC overnight. The membrane was air dried before covalent linkage of the DNA by uv irradiation (254 nm, 4 minutes) and washed in 6xSSC by gently rubbing with a gloved finger before prehybridization (6xSSC, 10xDenhardt's reagent, 0.2% SDS) in heat sealed bags for at least 3 hours. The DNA on the filters was hybridized with either GT3 or S1 oligonucleotides (6xSSC, 50 mM sodium phosphate (pH 6.8), 5 x Denhardt's reagent, 0.5% SDS, 20 ug/ml E. coli tRNA) at 37°C. Membranes with DNA from individual phage were hybridized for at least 12 hours, whereas membranes with genomic DNA were hybridized for at least 40 hours. The filters were
washed at room temperature in 6xSSC for 30 minutes and then a higher stringency wash was performed at either 50 or 52°C (GT3) and 47°C (SI) in 6xSSC. Filters were exposed for 2-7 days, depending on the counts retained.

IN SITU HYBRIDIZATION TO DROSOPHILA POLYTENE CHROMOSOMES

I. Labelling the probe

The DNA probe was labelled with $[^{125}\text{I}]$-CTP ($[^{125}\text{I}]$-5-iodo CTP). This was done by in vitro transcription of the phage DNA by E. coli RNA polymerase in the presence of $[^{125}\text{I}]$-CTP (53-55). The phage DNA was treated with RNase A (1 ug/ul) and phenol extracted twice and precipitated with 0.6 volumes of 20% PEG/2.5 M NaCl and resuspended in 20 ul TE 8. The transcription reaction consisted of 5 ug of phage DNA and 80 uCi $[^{125}\text{I}]$-CTP in 40 mM Tris-HCl (pH 7.5), 10 mM DTT, 10 mM MgCl$_2$, 20 uM ATP, GTP, UTP, and 5 U of E. coli RNA polymerase. This was incubated at 37°C for 2 hours followed by DNase I treatment (10 ug DNase I, 40 mM Tris-HCl (pH 7.8), 100 ug E. coli tRNA) for 20 minutes at room temperature. This mixture was extracted with phenol/CHCl$_3$ (once) and with CHCl$_3$ (three times) before loading the aqueous phase on a Sephadex G-25 column (0.196cm$^2$x 20 cm). The first radioactive peak was collected (in 0.3 M NaOAC (pH 7.2), 50 uM EDTA, 0.01% SDS) and precipitated by 2.5 volumes of 95% EtOH and redissolved in an appropriate volume of 70% formamide, 0.06 M KH$_2$PO$_4$, 0.06 M K$_2$HPO$_4$, 5 mM EDTA, 4mM KOH, and 0.5 M KCl.
II. In situ hybridization

The probe (prepared above) was placed onto a slide containing salivary gland chromosomes at $10^6$ cpm per slide and hybridized at 45°C for 2 days. The slides were washed, coated with emulsion, and developed essentially as described (55).

SUBCLONING PHAGE DNA INTO PLASMIDS

I. Ligation Reaction

The plasmid (pUC13, M13mpl8 or mpl9 RF) was digested (2 ug in 20 ul volume) with the desired restriction enzymes and heated at 65°C for 10 minutes. Calf intestinal phosphatase was added (CIP 1 unit in 50 mM Tris-HCl pH 9.0, 1 mM MgCl$_2$, 0.1 mM ZnSO$_4$, 1 mM spermidine—total volume 50 ul) and reacted for 30 minutes at 50°C. To terminate the reaction 40 ul H$_2$O, 5 ul 10% SDS, and 10 ul 10xSTE buffer (10xSTE= 0.1 M Tris-HCl pH 8.0, 1 M NaCl, 10 mM EDTA) was added and heated at 68°C for 15 minutes. The mixture was phenol extracted twice (phenol/CHCl$_3$ 1:1) and chloroform extracted once before EtOH precipitation of DNA.

The phage DNA to be subcloned was digested and mixed in varying molar ratios with 100 ng of linearized and dephosphorylated plasmid or RF DNA and ligated with 0.1 U T4 DNA ligase/10 ul in 50 mM Tris-HCl pH 7.5, 10 mM MgCl$_2$, 1 mM DTT, and 0.8 mM ATP at 15°C for 12-18 hours (56).

II. Plasmid and M13 Transformations

Plasmid and M13 DNA was introduced into JM101 that were made competent by the CaCl$_2$ method (56). A colony of JM101 from a
minimal glucose plate was grown in 2YT until OD$_{600}$ was 0.5-0.6. The cells were pelleted (5 min. at 4000 g) and resuspended in half the original culture volume of sterile 50 mM CaCl$_2$. After incubation on ice for 20 minutes the cells were pelleted and resuspended in 1/10 the original culture volume. Aliquots (300 ul) of competent JM101 were incubated on ice with the aliquots of the ligation mixture for 40 minutes. For pUC plasmids, the cells were heat shocked for 3 minutes at 42°C and plunged on ice; then 0.7 ml LB was added and the cells shaken for 1 hour at 37°C. Aliquots (0.1-0.3 ml) were plated on LB Amp (100 ug/ml) plates with 50 ul Xgal(2%) and 10 ul IPTG (100mM). For M13 transformations, the cells were returned to room temperature after heat shock. Fifty ul Xgal(2%), 10 ul IPTG (100mM), 200 ul fresh exponential JM101, and 3 ml of soft agar (0.7% in 2YT at 55°C) were added and the cells plated on 2YT plates (57). After incubation overnight at 37°C, white transformants (plaques or colonies) were picked individually or screened by transfer onto Hybond-N.

III. Growth of transformants

Transformants containing pUC plasmids were picked into 2 mls of LB Amp (100 ug/ml) and grown to saturation. M13 transformants were grown in 2 ml of 2YT containing 20 ul of JM101 (grown overnight in minimal glucose) at 37°C for 5 hours. Plasmid and RF preparations were by the alkaline lysis method (56). Part of the cultures were frozen in 15% glycerol as stock and the supernatant from M13 cultures were stored as high titer phage
stocks or used for ssDNA template preparations.

**DNA SEQUENCE DETERMINATION**

I. **Template preparation**

Single stranded DNA template from M13 transformants was prepared by spinning 1.2 ml of the M13 culture for 5 minutes in a microfuge. The supernatant was carefully removed and 0.3 ml of 20% PEG6000/2.5 M NaCl was added and mixed. The phage were precipitated at room temperature for 15 minutes and pelleted by centrifugation for 15 minutes in the microfuge. The supernatant was discarded and the pellet was resuspended in 0.2 ml of TE 8 and extracted twice with an equal volume of phenol/CHCl₃ (1:1) being careful to leave all of the interface. The aqueous layer was EtOH precipitated twice and redissolved in 50 ul of TE 8.

Double stranded DNA templates (either plasmids or M13 RFs) were prepared by taking 5 ug of plasmid and digesting with RNase A (1 ug/ul), phenol extracting once, and making the solution up to 0.5 M NaCl/TE 8 (100 ul total). The RNA was removed by passing the DNA through a spun column (desktop centrifuge, 1400rpm, approx. 5 min., calibrated to recover 100 ul) of crosslinked Sepharose 4B. The recovered DNA was precipitated and rehydrated in 40 ul of 0.2 M NaOH and denatured at room temperature for 5 minutes. The solution was neutralized with 1/10 vol of 2 M NH₄OAc (pH 4.5) and precipitated with 2.5 volumes of 95% EtOH. The DNA was dissolved in 10 ul TE 8 just before sequencing (58).

II. **DNA Sequencing**
DNA sequence was determined by the dideoxynucleotide chain termination method (59). The ssDNA template (5 ul) was mixed with 1 ul of M13 universal forward or reverse primer (4 pmoles) and 2 ul of annealing buffer (100 mM NaCl, 100 mM Tris-HCl (pH 7.5)). The mixture was heated at 65°C for 10 minutes and cooled slowly to room temperature. For double stranded sequencing, 1-2 ug of template was annealed for 15 minutes at 37°C (58). In both cases, 1 ul 15 uM dATP and 1.5 ul α-[32P]-dATP 3000 Ci/mmmole) were added to the template/primer mix and 2.2 ul of the final mixture was distributed to tubes containing 1.5 ul of each of the A,T,C,G deoxy/dideoxy mixes. The following nucleotide mixes were used (60):

- dG/ddG: 89 uM ddGTP, 7.9 uM dGTP, 158 uM dTTP, 158 uM dCTP
- dA/ddA: 116 uM ddATP, 111 uM dGTP, 111 uM dTTP, 111 uM dCTP
- dT/ddT: 547 uM ddTTP, 158 uM dGTP, 7.9 uM dTTP, 158 uM dCTP
- dC/ddC: 547 uM ddCTP, 158 uM dGTP, 158 uM dTTP, 10.5 uM dCTP

The reaction was initiated by the addition of 1 ul of DNA polymerase I Klenow fragment (0.5U/ul in 80 mM potassium phosphate (pH 7.5), 0.8 mg/ml BSA, 40% glycerol, 10 mM DTT). The reaction mixtures were incubated at 37°C for 15 minutes and then chased by the addition of 1 ul dNTP (0.5 mM) to each tube with another 15 minute incubation. The reactions were terminated by adding 5 ul of formamide/dye mix (98% deionized formamide, 10 mM EDTA, 0.2% bromophenol blue and xylene cyanol). The samples were heated for 3 minutes at 90-100°C and placed on ice.
The samples (1 ul) were electrophoresed in 6% or 8% polyacrylamide gels (acrylamide : methylenebisacrylamide (19:1), 8 M urea, 0.06% ammonium persulfate, 20 ul TEMED, 50 mM TBE) at 1600-1800 V such that the current did not exceed 25mA. The gel was run until the bromophenol blue reached the bottom (40cmx18cmx0.35cm gels) or 1.5 hr after the last loading with a maximum of 3 loadings (0, 1.5, 3 hr). The gels were dried onto Whatman 3MM paper using a vacuum gel drier (1.5 hr, 80°C) and were exposed to X-ray film at room temperature for hours or days depending on radioactivity incorporated.
RESULTS

THE STRINGENCY OF HYBRIDIZATION

The stringency of hybridization must be decided upon before embarking on screening of the library. Hybridization 5°C below the dissociation temperature ($T_d$) allows perfect duplexes to be detected (61). The hybridization range for GT3 and S1 was determined to be 45-53°C and 37-47°C, respectively (an estimate from $T_d = 4(G+C) + 2(A+T)$ in 6xSSC (44,46)). Reactions should be performed under the most stringent conditions to eliminate hybridization to related sequences but with mixed sequence oligonucleotides the stringency must be a compromise i.e. the stringency must be high enough for signals to be discernible over background but still allow the hybridization of possibly all sequences in the mix. The hybridization temperature was determined by probing genomic and library DNA and looking for the presence of distinct bands above background at the chosen temperature (Figure 3). Hybridization of GT3 to digested Oregon-R genomic and library DNA at 50°C showed multiple bands of varying intensities. Due to its high degeneracy (128 variants) and low melting, bands were not visible when the S1 probe was used to probe membrane-bound genomic DNA at or below 47°C. Subsequent hybridization of GT3 was at 52°C (high stringency) and 50°C (lower stringency) whereas S1 was at 47°C. Lower temperatures were not feasible due to background interferences so it is possible that lower melting sequences in the probe were
FIGURE 3. Oregon-R genomic and lambda library DNA Southern analysis. Oregon-R genomic DNA (I) and lambda library DNA (II) were digested with (a) EcoRI (b) EcoRI/SalI (c) HindIII (d) HindIII/SalI and transferred onto nylon membranes (Hybond-N). The membranes were hybridized with $[^{32}\text{P}]$ labelled GT3 at 37°C and washed at increasingly higher temperatures until positive bands were discernible above background. The membranes with (III) genomic DNA and (IV) lambda library DNA were washed at 50°C, in 6xSSC, and exposed to X-ray film for 4 days at -70°C. The autoradiograms show GT3 hybridizes to different fragments at varying intensities. The size marker (m) was lambda DNA digested with HindIII.
excluded. Direct correlation of bands in the Oregon-R genomic and library DNA was not possible since the library was constructed as MboI partial digests ligated into the BamHI site of the EMBL3 vector (42).

PRIMARY SCREEN OF CANTON-S LAMBDA LIBRARY

The GT3 probe was used to screen the D. melanogaster Canton-S lambda library (43). The primary screen of $6 \times 10^4$ PFU (equivalent to 6 genomes) by GT3 at $52^\circ$C resulted in 5 positive signals clearly discernible above background. The SI probe could not be used for primary screens as it showed non-specific hybridization to lambda arms or some component of the lysed bacteria.

SECONDARY SCREEN OF CANTON-S LAMBDA LIBRARY

The individual phages responsible for the positive signals were isolated and the DNA inserts were released from the Charon 4 vector by digestion with EcoRI (which produces 19 and 11 Kb vector arms). Analysis of the EcoRI fragments showed 3 different phages (CS-1, CS-2, and CS-3) with common as well as unique size fragments. Phage CS-3 had a 3.7 Kb EcoRI band which had been cleaved to smaller fragments in CS-2 (Figure 4). Therefore, these three phages may have overlapping genomic inserts with polymorphic EcoRI sites. Hybridization with GT3 showed that the 4.7 Kb (from CS-1) and 7.0 Kb (from CS-2, CS-3) EcoRI fragments hybridized GT3 (Figure 4).

DNA SEQUENCE ANALYSIS OF GT3 BINDING SITE (CANTON-S LIBRARY)
FIGURE 4. Southern analysis of DNA from GT3 positive phage from the Canton-S genomic lambda library. The genomic insert was released from the phage vector arms (Charon 4) by digestion with EcoRI. The two highest molecular weight bands (19 and 11 Kbp) on the agarose gel (I) are the vector arms. The EcoRI digestion patterns reveal three different phage: CS-1 (a,b), CS-2 (c,d), and CS-3 (e,f,g,h,i). The DNA separated on the agarose gel was transferred to Hybond-N and hybridized to GT3. The filter was washed at 52°C in 6XSSC and the resulting autoradiogram is shown (II). Phage CS-1 (a,b) has a 4.7 Kb positive EcoRI fragment whereas phage CS-2 and CS-3 (c-i) have a 7.0 Kb GT3 binding EcoRI fragment.
The 4.7 Kb (from CS-1) and 7.0 Kb (from CS-2 and 3) GT3 positive EcoRI fragments were cloned into pUC 13 and pEMBL8-, respectively. The plasmid with the 4.7 Kb insert (CS-1) was digested with PstI to produce a 500 bp GT3 positive fragment, which was subcloned into M13mp8 for Sanger sequencing. Sequencing the single stranded template for 200 bp (from one end of the insert) did not reveal any homology to CuZn SOD when translated in all 3 reading frames. T-track analysis of the 10 other clones containing the 500 bp PstI insert showed that the insert was cloned in the same orientation in all 10 clones. Sequencing from the other end of the insert by double stranded sequencing of the M13 RF showed the GT3 hybridizing site starting one nucleotide after the PstI recognition site (Figure 5). The GT3 hybridizing site of the 7.0 Kb fragment from CS-2 and CS-3 was sequenced on both strands by the method of Maxam and Gilbert (S. Hayashi). The three positive phage (CS-1, 2 and 3) had identical nucleotide sequence in the GT3 probe binding region and identical GT3 binding sites.

The GT3 binding region was recognized by translation of the DNA sequence resulting in the amino acid sequence Asp-Cys-Pro-Thr-Lys-Lys whereas GT3 was originally targeted to Asp-Cys-Pro-Thr-Lys-Val in CuZn SOD (Figure 6). The one GT3 sequence responsible for binding this phage is 5'-GAT TGT CCG ACG AAG GT-3' which forms 15 perfect Watson-Crick base pairs and one dG:dT pair with the phage DNA (Figure 6). This member of the GT3
FIGURE 5. DNA sequencing strategy for the GT3 binding site of the Canton-S library GT3 positive phage. The three phage from the Canton-S library (CS-1, CS-2, CS-3) had identical nucleotide sequence in the GT3 probe hybridizing region. The GT3 binding site (open box) was sequenced by the Sanger chain termination method (solid arrows) on one strand of CS-1 and on both strands of CS-2 and CS-3 by the Maxam and Gilbert chemical method (dotted arrows).
PstI   TaqI   PstI   TaqI

230   130   200

Sanger Sequencing
Maxam and Gilbert Sequencing
Oligonucleotide binding site

100 bp
FIGURE 6. The nucleotide sequence and translation of the GT3 binding site of phage CS-1, 2, and 3 from the Canton-S library. Translation of the nucleotide sequence around the GT3 binding site produced a sequence of 38 amino acids. More nucleotide sequence upstream of the GT3 binding site is required to determine the number of amino acids that is encoded by this sequence. There exists a polymorphic site six nucleotides 5' to the start of the GT3 binding site (G*). The CS-2 and 3 sequence has GG at this site. The GT3 probe binding site was recognized by the amino acid sequence Asp-Cys-Pro-Thr-Lys-Lys as opposed to the desired SOD sequence Asp-Cys-Pro-Thr-Lys-Val. All nucleotide sequences were translated in all three phases and no other homology to the SOD protein was found. The phage sequence 3'-CTA ACA GGC TGC TTC TT-5' hybridized the GT3 sequence 5'-GAT TGT CCG ACG AAG GT-3'. 
<table>
<thead>
<tr>
<th>ATA</th>
<th>GCC</th>
<th>AAG</th>
<th>GCG</th>
<th>AAG</th>
<th>ATA</th>
<th>GCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILE</td>
<td>ALA</td>
<td>LYS</td>
<td>ALA</td>
<td>LYS</td>
<td>ILE</td>
<td>ALA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AAA</th>
<th>TAC</th>
<th>TTG</th>
<th>GCG</th>
<th>CGG</th>
<th>GCT</th>
<th>GAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>LYS</td>
<td>TYR</td>
<td>LEU</td>
<td>ALA</td>
<td>ARG</td>
<td>ALA</td>
<td>GLU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GAG</th>
<th>ATT</th>
<th>CAC</th>
<th>GCC</th>
<th>AAT</th>
<th>TTC</th>
<th>CTG*</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLU</td>
<td>ILE</td>
<td>HIS</td>
<td>ALA</td>
<td>ASN</td>
<td>PHE</td>
<td>LEU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GCC</th>
<th>AAC</th>
<th>GAT</th>
<th>TGT</th>
<th>CCG</th>
<th>ACG</th>
<th>AAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALA</td>
<td>ASN</td>
<td>ASP</td>
<td>CYS</td>
<td>PRO</td>
<td>THR</td>
<td>LYS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AAG</th>
<th>CTG</th>
<th>CAG</th>
<th>TTC</th>
<th>CAG</th>
<th>CTC</th>
<th>TCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>LYS</td>
<td>LEU</td>
<td>GLN</td>
<td>PHE</td>
<td>GLN</td>
<td>LEU</td>
<td>SER</td>
</tr>
</tbody>
</table>

| CTG | LEU |
mixed oligonucleotide family appears to be the highest melting member to have hybridized to a genomic sequence from the Canton-S library.

**PRIMARY SCREEN OF OREGON-R GENOMIC LAMBDA LIBRARY**

Screening with GT3 at a lower stringency (to include lower melting members of GT3) was performed at 50°C in a newly constructed Oregon-R genomic lambda library. The Oregon-R library has never been amplified (in contrast to the much older Canton-S library) and all clonable sequences should be present. Screening $6 \times 10^4$ PFU (6 genomes) of the Oregon-R library resulted in 97 primary screen plugs corresponding to potential positive signals.

**ANALYSIS OF OREGON-R PRIMARY SCREEN PHAGES**

The DNA from the primary screen phage plugs (containing approximately 100 individual phage species) was purified in order to eliminate false positives due to non-specific hybridization of GT3 to bacterial DNA and debris. An initial strategy was to amplify and prepare DNA from the 97 primary plugs in pools of ten and to determine which pool had both GT3 and S1 hybridizing inserts. The results suggested that five of the ten pools did not contain inserts which hybridized both probes. However, subsequent analysis of DNA from individual primary plugs within the pool showed this was not the case. This may be due to over or under representation of certain phage during competitive growth. Therefore, DNA from all 97 primary screen plugs had to be
purified. Hybridization of both probes to primary plug DNAs revealed that: 75 hybridized S1 probe with 31 of these also hybridizing GT3, 3 were GT3 positive only, and 19 did not bind either probe (Figure 7).

SECONDARY SCREEN OF OREGON-R GENOMIC LAMBDA LIBRARY

Low density screening of all 31 GT3/S1 positive primary plugs resulted in the isolation of single phage species capable of binding GT3. Digestion with SalI released the entire insert from the EMBL3 vector, but this analysis was difficult since SalI cuts the genomic insert infrequently. Therefore, each phage was digested with HindIII and hybridized with both probes. All phages showed different restriction fragment patterns (although some may be overlapping phages) with the exception of 5 phage which were identical. Four phages (109, 224, 253, 292) hybridized both probes (Figure 8). The gels were deliberately overloaded to increase the intensity of the signal. Different HindIII fragments hybridized GT3 and S1 in phage 109, 224, and 253, whereas in phage 292 both probes hybridized the same fragment. Hybridization of GT3 and S1 on different HindIII fragments is plausible since the probe binding sites may be separated by introns in the gene. By analogy to the human CuZn SOD sequence (62), there are two introns between the sequences where the two probes should hybridize. Phage 253 appears to have a strong (2.9 Kb) and weak (1.3 Kb) S1 hybridizing fragment. This may be due to two variants of S1 binding different HindIII fragments of the
FIGURE 7. Southern analysis of Oregon-R primary screen phages. The phage DNA from each primary screen plug was digested with HindIII and the fragments separated on an agarose gel (I). There are many HindIII fragments per lane, as each primary plug contains approximately 100 phage species. The most prominent band is the 4.4 Kb HindIII fragment which is the EMBL3 vector arm. The DNA fragments were transferred onto Hybond-N and hybridized with GT3 (50°C) or S1 (47°C). The autoradiogram of the GT3 probed filter (II) shows that a HindIII fragment from primary plugs 1-3 (41-43) and 6 (46) strongly hybridizes to the GT3 probe whereas the rest of the phage weakly or questionably hybridize the GT3 probe. The autoradiogram of the S1 probed filter (III) shows that multiple HindIII fragments from each primary screen pool hybridize S1 sequences. Therefore, all the primary plugs (41-49) examined may contain GT3 and S1 probe binding sequences and all were analyzed further. The size marker (m) was lambda DNA digested with HindIII.
FIGURE 8. Southern analysis of phage from the Oregon-R library that hybridize both GT3 and S1 probes. The result of screening the Oregon-R genomic library was four phages (a)=109, (b)=224, (c)=253, (d)=292) that hybridize both S1 and GT3 probes. (I) The phage DNA was digested with HindIII and analyzed on an 0.7% agarose gel. All four phage show different restriction patterns and a common 4.4 Kb EMBL3 vector arm. The DNA was transferred onto Hybond-N and hybridized with GT3 and S1 probes. The autoradiograms of the filters probed with GT3 (II) and S1 (III) show that different HindIII fragments hybridized the two probes, except in phage 292 where both probes hybridize the same 7 Kb HindIII fragment. Phage 253 has two S1 positive fragments which may be due to hybridization of different S1 sequences. All four phage were mapped on polytene chromosomes to determine whether they originate from the same chromosomal location as the SOD gene. The size marker (m) was lambda DNA digested with HindIII.
insert. Phage 253 is also the weakest GT3 binding phage as hybridization of GT3 does not occur at 52°C. Phage 109, 224, and 292 still bind GT3 at 54°C.

**MAPPING OF PHAGE DNA TO DROSOPHILA POLYTENE CHROMOSOMES**

The chromosomal location of CuZn SOD has been determined (48,49) to be at 68AB on the third chromosome and a phage which maps to the same site would be a good candidate for being the CuZn SOD clone. *In situ* hybridization of phage 109, 224, 253, and 292 showed that none of these phage hybridized to 68AB. Phage 109, 224, and 292 mapped to 3A3-4 (X), 87D5-7 (3R), and 77B1-2 (3L), respectively (Figure 9a). The GT3 positive phage from the Canton-S library (CS-2) mapped to 90EF (3R) (Figure 9b). Phage 253 mapped to many chromosomal locations.

**THE GT3 PROBE BINDING SITE OF PHAGE 109**

The GT3 probe binding region of phage 109 was sequenced to determine whether the probe was hybridizing to related sequences (and possibly pseudogenes) or via mismatched duplexes. A 2.6 Kb GT3 positive BamHI fragment was subcloned into pUC13 to facilitate subcloning smaller fragments into M13 vectors. The BamHI clone was digested with Sau3A and all the fragments were ligated into M13mpl9 and screened with GT3. A 190 bp GT3 positive Sau3A insert was sequenced on both strands. The GT3 probe binding site was found to be 140 bp from one end of the insert and was recognized by translation in one reading frame resulting in Asp-Cys-Pro-Thr, four of the six amino acids
FIGURE 9. Determination of the chromosomal location of phage 109 and phage CS-2 by in situ hybridization to polytene chromosomes. In vitro transcription of (a) phage CS-2 (the 7.0 Kb EcoRI fragment cloned into pEMBL8-) and (b) phage 109 in the presence of \[^{125}\text{I}]\) -CTP produced labelled RNA which was hybridized to D. melanogaster polytene chromosomes. Phage CS-2 mapped to 90EF (3R), and phage 109 mapped to 3A3-4 (X). Similarly phage 224, 253 and 292 were also mapped on the polytene chromosomes (figures not shown), none of which hybridized to 68AB (on 3L) where the CuZn SOD gene is located. This figure is courtesy of Dr. Shizu Hayashi.
specified by GT3 (Asp-Cys-Pro- Thr-Lys-Val). The phage 109 sequence 3'-CTG ACG GGC TGC TAT AC-5' forms 13 perfect Watson-Crick base pairs with GT3 (Figure 10). The duplex ends at position 14 with a dA:dA mismatch (reading 5' --> 3' on GT3). This phage 109 sequence may hybridize to 5'-GAC TGC CCG ACG AAA GT-3' and 5'-GAC TGC CCG ACG AAG GT-3' of the GT3 probe. These GT3 sequences form a 13 base pair duplex with phage 109 which is stable even at 54°C, probably due to the high GC content (9/13) of the first 13 nucleotides of the sequence. Translation of the entire 190 bp insert in all 3 phases did not show any other amino acid homology to the CuZn SOD sequence.
FIGURE 10. Nucleotide sequence and translation of the GT3 probe binding site of phage 109 from the Oregon-R library. Translation of the nucleotide sequence produced the sequence Asp-Cys-Pro-Thr, four of the six amino acids specified by GT3 (Asp-Cys-Pro-Thr-Lys-Val). The phage 109 sequence 3'-CTG ACG GGC TGC TAT AC-5' forms 13 perfect Watson-Crick base pairs with the corresponding GT3 sequences 5'-GAC TGC CCG ACG AAA GT-3' and 5'-GAC TGC CCG ACG AAG GT-3'.

(*** = amber codon)
80  
CTC  CGC  CAC  TAG  TCC  ACT  AGT  
LEU  ARG  HIS  ***  SER  THR  SER  

100  
TAG  TTG  CCT  CCT  CTG  CGA  GCC  
***  LEU  PRO  PRO  LEU  ARG  ALA  

110  
ATC  ACA  CCT  CAA  TAC  TGT  TCA  
ILE  THR  PRO  GLN  TYR  CYS  SER  

120  
GAC  TGC  CCG  ACG  ATA  TGT  TGT  
ASP  CYS  PRO  THR  ILE  CYS  CYS  

130  
CCG  TGT  GCC  TGC  CCA  GGC  GCC  
PRO  CYS  GLY  CYS  PRO  GLY  ALA  

140  
TCC  TCT  T  
SER  SER  

150  

160  

170  

180  

190  

200  

DISCUSSION

To date, mixed sequence oligodeoxyribonucleotide probes 13-17 bases in length have been successfully used to isolate protein genes of unknown DNA sequence from libraries of low to moderate complexity (61,63). The mixed sequence probes often lack the specificity required for probing sequences as complex as those in a Drosophila or mammalian genomic library. Successful use of mixed sequence oligonucleotide probes rests largely on probe design i.e. heterogeneity, base sequence, and length of the probe. Each of these factors and their implications on probe design will be discussed in light of the results presented in this study.

I. OLIGONUCLEOTIDE PROBE HETEROGENEITY

The redundancy of the genetic code often leads to complex mixtures of oligonucleotide probes. Genes have been isolated using mixtures of great complexity (>32 different sequences) but only in cDNA libraries (64). In mixtures of greater than 16 sequences, each individual sequence comprises less than 6% of the total mixture, resulting in unfavourable signal to noise ratios in hybridization experiments. For example, genomic DNA Southern hybridized with S1 did not produce discernible bands as the background was high. Therefore, the great sequence complexity of the S1 probe (128 variants) restricted its use to hybridization with pure cloned DNA only.

The greater the heterogeneity of the probe, the larger the
range of dissociation temperatures for oligonucleotide-DNA duplexes will be. The dissociation temperature of each oligonucleotide differs, making the selection of a suitably stringent and selective hybridization condition for all members difficult or impossible. Therefore, a large number of positives are picked up in screening complex libraries. The greater the complexity of the probe, the greater the probability of binding unrelated sequences will be. This was observed in the primary screen of the Oregon-R library. In principle, every sequence in the SI probe may elicit a positive signal and most of the phage (75/97) did hybridize SI sequences. The four phages (109, 224, 253, 292) which hybridized to both GT3 and SI probes all mapped to different chromosomal locations. Therefore, it is very important to keep the number of sequences in a probe to a minimum. Low degeneracy probes are possible if an amino acid sequence contains methionine and tryptophan, both having only one codon. A single codon may also be chosen for an amino acid by examining codon preference data in that organism, as there exists a bias in the usage of the several degenerate codons for an amino acid (65,66). However, this is only possible if extensive codon usage data for the organism are available and there exists the possibility that the preferred codon may not be used at the sequence where the oligonucleotide probe is targeted.

II. OLIGONUCLEOTIDE PROBE SEQUENCE

The oligonucleotide sequence determines the stability of the
oligonucleotide-DNA complex. The stability depends not only on the nucleotide composition, but also on the nucleotide sequence of the probe (67-69). It has been noted that in a 17 nucleotide long mixed sequence probe, hybridization of the lowest melting probe (usually lowest GC content) will also allow hybridization of 12 or 13 bp regions (70). The GT3 binding site of phage 109 shows that at the hybridization stringency chosen to include lower melting sequences, a 13 bp duplex was responsible for eliciting the GT3 positive signal.

The degeneracy of the 17 base GT3 probe was decreased from 128 to 8 variants by the placement of dG in positions 9 and 12, where all four nucleotides may occur. Pairing with dC is normal, but pairing with dG is considered unfavorable. It was assumed that dG forms a favorable wobble base pair with dT and that pairing to dA is weakly stabilizing. This assumption was based on observations that the ribonucleotides rG:rU and rG:rT form stable base pairs. However, it has been argued that the assumption is false since dG:dT is less stable than rG:rT (68,71). Recent studies on the stability of mismatches in oligonucleotide duplexes reveal that the stability of dG:dT, dG:dG and dG:dA pairs are comparable and stabilizing (69).

The placement of dG as the third base of the proline and threonine codons may not agree with codon preferences in Drosophila. Although extensive codon usage data are not available for Drosophila, the sequence analysis of 8 genes
reveals that for proline the codon CCG is the least used and that CCC is the most used. Similarly for threonine, ACG is rarely used and ACC is the most frequently used (66). In both cases, the most favored codon would require a ΔG:ΔG pairing with the GT3 sequence. Although ΔG:ΔG interactions are not destabilizing, it is important to remember that the difference in stability between ΔC:ΔG and the mismatched ΔG:ΔT, ΔG, ΔA is large (69). Therefore, the placement of ΔG in these positions in GT3 would increase the tendency of the probe to bind related but incorrect sequences containing ΔC at the corresponding position on the target DNA. Both GT3 binding sites sequenced in this study show ΔG:ΔC pairing at positions 9 and 12 of the probe, and both strongly hybridize unrelated sequences.

III. OLIGONUCLEOTIDE PROBE LENGTH

The length of the oligonucleotide probe required to predict an unique sequence may be statistically calculated. For a Drosophila genome (1.65x10^8 bp), a probe 15 nucleotides long should predict a unique sequence whereas for the human genome (3x10^9 bp) a probe 17 nucleotides in length is required (71). Each member of the mixed sequence probe may hybridize a unique sequence specifying the same amino acids. The clone isolated from the Canton-S library had a 15 bp sequence capable of coding for five of the amino acids in CuZn SOD. This nucleotide sequence should be unique, and the sequence in the SOD gene should be
specified by another member of GT3. The GT3 binding sequences in the two clones studied are probably intergenic sequences since the potential open reading frames are not very long. These intergenic sequences would be eliminated in a cDNA library. Screening lower complexity libraries (such as cDNA libraries) has not always been successful using two short oligonucleotide probes (72,73). This has prompted the use of longer unique probes (74).

In long probes (>50 bp), the uncertainty at each codon is largely ignored, as the probe length is used to confer probe specificity. Theoretically, the specificity of a two probe combination is less than that of a probe of length equal to the combined length of the shorter probes (68).

Isolation of genomic clones using short mixed oligonucleotide probes has never been reported. An unique 52 base long probe was used to isolate a genomic clone containing the human factor IX gene (74). The two other cases of genomic clone isolation used even longer probes, although they may have been longer than necessary (75,76). Therefore, longer unique probes may be used to overcome the problems encountered using short mixed probes.

IV. OLIGONUCLEOTIDE PROBE DESIGN

Theoretical and practical considerations of probe design have recently been described in great detail (68). The problems encountered in searching for genes in DNA banks of greater complexity has required modification of the classical strategy
used in this study.

Longer probes have been used to increase probe specificity as short stretches of perfect homology occurring by chance have produced false positives. One approach used to obtain longer probes has been to "snap together" 10 overlapping short oligonucleotides and filling in the gaps to form a long double stranded probe (76).

Another modification of probe design uses deoxyinosine (dI) in positions of ambiguity where all 4 nucleotides are possible. Inosine is a guanosine analog and occurs naturally in the wobble position of the anticodon of some transfer RNAs where it appears to pair with rC, rU and rA. The thermal stability of oligonucleotide duplexes containing deoxyinosine has been studied. Independent of sequence effects, the order of stability is dI:dC > dI:dA > dI:dT, dI:dG (77).

Deoxyinosine is superior to dG opposite dA/dC ambiguities since dI pairs less strongly with dC and more strongly with dA than dG. For probing dG/dT ambiguities, both dG and dI are non-selective. Use of dI in these cases is also advantageous, as dI does not increase the tendency of the probe to bind unrelated sequences by forming strong dG:dC pairs. With these factors in mind, one must decide whether the best sequence for the probe corresponds to the coding or non-coding strand.

In light of the results of this study, an improved probe might be a modified GT3 which uses dI in positions of base
ambiguity:

\[5'- \text{ACC TTI GTI GGA CAA TCI CCI GTI GC } -3'\]

The revised GT3 probe is 26 nucleotides in length and is derived from the non-coding strand as opposed to GT3, which was from the coding strand. The new probe would also be 9 nucleotides longer and specify 3 additional amino acids.

Despite the guidelines available for oligonucleotide probe design, decisions concerning probe length, heterogeneity, and sequence are still a matter of educated guesswork and whether any probe will find the desired gene sequence can only be determined by experimentation.
REFERENCES


