Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 710–714, February 1986
Genetics

A complete library of point substitution mutations in the glucocorticoid response element of mouse mammary tumor virus

(oligonucleotide-directed mutagenesis/DNA sequence/enhancer)

CLYDE A. HUTCHISON III, STEVEN K. NORDEEN*, KENNETH VOGT, AND MARSHALL HALL EDGELL

Department of Microbiology and Immunology, Curriculum in Genetics, and Program in Molecular Biology and Biotechnology, The University of North Carolina, Chapel Hill, NC 27514

Communicated by Robert Sinsheimer, September 19, 1985

ABSTRACT **The glucocorticoid response element (GRE) of mouse mammary tumor virus (MMTV) was chemically synthesized as two complementary DNA strands bearing cohesive termini. During automated synthesis, random mutations were introduced into the DNA by "doping" each of the four nucleoside phosphoramidites (A, G, C, and T) with a low level of the other three. These preparations were annealed and cloned into an M13 phage vector to produce a library of GRE mutants. Mutations within the synthesized region were identified by sequencing phage isolates at random. All of the chemically distinct classes of transition and transversion mutations have been observed. Statistical considerations indicate that the library contains all of the possible 90 point substitution mutations within a 30-nucleotide mutagenic target. So far 88 of these substitutions have been isolated, 74 as single mutants. At least two of the three possible single mutants at each of the 30 positions have been identified.**

Several methods have been developed for introducing defined changes into DNA molecules of known sequence (for review, see ref. 1). The most precise of these is oligonucleotide-directed mutagenesis (2), which has been used to introduce single base substitutions at specific positions in DNA sequences as well as specific insertions and deletions. In fact, the major limitation of the method has been its high degree of specificity, because the strategies used have usually required the synthesis of a different mutagenic oligonucleotide for each desired mutation. In situations where it is necessary to produce many mutations throughout a sequenced region, other methods have been used, such as bisulfite mutagenesis (3) and the linker scanning method (4). The method of enzymatic nucleotide misincorporation (5) appears to produce a completely random set of substitution mutations. Recently, the potential of oligonucleotides as mutagens has been expanded through the use of mutagenic oligonucleotides synthesized to contain mixtures of bases at several positions within a target sequence (6–9). This allowed a number of different mutations to be introduced by using a single synthetic oligonucleotide preparation.

We present here a general method for producing and identifying every possible single base substitution mutation within a region. As described below, the method requires synthesis of each strand of the target sequence only once. To test the feasibility of the method, we have produced mutations within the glucocorticoid response element (GRE) of mouse mammary tumor virus (MMTV). The GRE resembles an enhancer that can stimulate expression of an adjacent gene in the presence of glucocorticoids (10, 11). Like enhancers, this activity exhibits some positional and orientational flexibility (12, 13). The GRE is specifically recognized by the glucocorticoid hormone receptor *in vitro* (14, 15), an interaction that is undoubtedly an integral part of the regulatory mechanism. Assays of various deletion mutants (12, 13, 16–19), combined with results of receptor–DNA footprint experiments (14, 15), indicate that a domain of ≈30 nucleotides (positions −160 to −190 relative to the start of transcription; see Fig. 1) contains a strong receptor binding site and is critical for activity of the element. We therefore expect that a combination of *in vitro* and *in vivo* assays of the mutants of this domain that we have isolated will yield insights into the sequence specificity of this regulatory element and the mechanism of steroid-mediated regulation of gene expression.

The method relies on automated synthesis of the target sequence in a way that yields a randomly mutagenized preparation. This is achieved by a slight modification of the usual synthetic procedure, in which chains are built stepwise from the 3′ end by the addition of nucleoside phosphoramidite monomeric units. Before synthesis begins, each of the four monomer reservoirs is "doped" with a small amount of each of the other three. Incorporation of a dopant molecule into the synthetic product results in a mutant sequence. Since the contaminating nucleotides are incorporated at random, this procedure results in a population of molecules containing 0, 1, 2, 3, or more mutations. It is possible to control the number of mutations per molecule by adjusting the composition of the phosphoramidite mixtures.

After synthesis, the mutagenized sequences are amplified by biological cloning to produce a mutant "library," a population containing a large number of single and multiple mutants of the original sequence. Every viable molecule in the library contains the synthetic sequence because of the particular cloning strategy used. Very rapid and convenient DNA sequencing procedures make it practical to identify every possible substitution mutation within the mutagenic target by simply sequencing random isolates from the library.

MATERIALS AND METHODS

Chemical Synthesis of DNA. Synthesis was performed by using an Applied Biosystems model 380A DNA synthesizer. We used the standard program cycle supplied with the machine (ABI001; 7/22/83), which was measured at 18.2 min. All synthetic reagents were also from Applied Biosystems (Foster City, CA). Standard operating procedures were used except for the preparation of the mutagenic nucleoside phosphoramidite mixtures. The contents of 0.5-g bottles of each of the four phosphoramidites were dissolved in the following amounts of dry acetonitrile injected through the septum to give 0.13 M solutions (A, 5.8 ml; G, 6.0 ml; C, 6.2 ml; T, 7.0 ml). After the phosphoramidites were com-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: GRE, glucocorticoid response element; MMTV, mouse mammary tumor virus.

*Present address: Department of Pathology, University of Colorado Health Sciences Center, Denver, CO 80262.

pletely in solution, all four were uncapped. The four solutions were cross-contaminated by transferring 100 μl from each one to each of the other three. This was done quickly with a Gilson Pipetman P200 without changing tips. The bottles were covered temporarily with Parafilm (American Can Company, Greenwich, CT) swirled briefly, and placed on the machine as quickly as possible by using the bottle changing routines. This resulted in roughly a 5% impurity in each phosphoramidite solution, or ≈1.5% of each of the three mutagenic species.

The oligonucleotides were detritylated and partially deprotected automatically by the synthesizer. The oligonucleotides were fully deprotected and purified by chromatography on Sephadex G-50 fine, followed by gel electrophoresis, using standard procedures (supplied by Applied Biosystems). A fairly broad segment was excised from the preparative gel (so that some of the species, one longer and one shorter than the desired length, were included) to be sure to include mutant sequences that might have altered mobility.

Cloning. The two complementary oligonucleotides were mixed at a concentration of 1 pmol/μl each in TM/16 (TM = 100 mM Tris·HCl, pH 8.5/50 mM $MgCl_2$) to give a total vol of 20 μl, in a 1.5-ml Eppendorf tube. This was floated on 1 liter of water at 100°C, and allowed to cool to room temperature over a period of several hours. Condensation was collected by brief centrifugation. This annealing reaction, and a 1:100 dilution in TM/8 (0.01 pmol/μl) were stored at −20°C. It should be noted that these oligonucleotides were not phosphorylated at their 5′ termini.

M13 mp11 replicative form DNA (20) was simultaneously digested with approximately a 10-fold excess of *Hin*dIII and *Sst* I in core buffer (BRL). The digest was phenol-extracted, ethanol-precipitated, rinsed with ethanol, dried, and dissolved in 10 mM Tris·HCl, pH 8.1/0.1 mM EDTA (TE) buffer to give a stock at 20 ng/μl.

One hundred nanograms (0.021 pmol) of digested M13 mp11 vector was mixed with 0.042 pmol of annealed oligonucleotide in 50 mM Tris·HCl, pH 7.5/10 mM $MgCl_2$/10 mM dithiothreitol/1 mM ATP/3 units of T4 DNA ligase (BRL), to give a total vol of 20 μl. This reaction mixture was incubated overnight at 4°C. The reaction mixture was phenol-extracted, ethanol-precipitated, rinsed with ethanol, dried, and redissolved in 50 μl of TE buffer. Six microliters of a solution of 100 mM Tris·HCl, pH 7.8/60 mM $MgCl_2$/1.5 M NaCl/2 mM EDTA/10 mM dithiothreitol was added, plus enough *Sal* I and *Bam*HI restriction enzymes to give a 50- to 100-fold overdigestion of the vector polylinker (58 μl total vol). Five microliters of this digest was used to transfect competent *Escherichia coli* strain JM107 (21) prepared by the method of Hanahan (22).

A plate containing ≈5000 plaques was washed by overlaying with 4 ml of LB medium for 5 min. The medium was removed, centrifuged, and samples were stored at 4°C and −20°C. This is the GRE mutant library used in subsequent analysis.

DNA Sequencing. Sequencing was performed by the chain-termination method (23). Very rapid procedures for preparation of sequencing templates from 1.5-ml cultures, carrying out sequencing reactions in 96-well microtiter trays, and running 12 sets of reactions (48 samples) on a 20-cm wide gel were performed essentially as described to us by Alan Bankier (Medical Research Council Laboratory of Molecular Biology, Cambridge, England). These are slight modifications of the methods described by Bankier and Barrell (24), and a detailed protocol is available on request. These methods make it feasible to sequence 48 or more isolates in a day. One of us (C.A.H. III) sequenced 189 clones from the mutant library described in this paper in a 10-day period, but a graduate student should be able to proceed more rapidly. The ratio of dideoxynucleotides to normal deoxynucleotides was increased 15-fold above that used for routine sequencing to facilitate the sequencing of small inserts. A 15-nucleotide primer, 5′ AGTCACGACGTTGTA 3′ (Bethesda Research Laboratories), worked well for sequencing inserts into the primer proximal end of the M13 mp11 polylinker. The Klenow fragment of *E. coli* DNA polymerase I was obtained from Boehringer Mannheim, or from Bethesda Research Laboratories. The sequencing reactions were radioactively labeled with dATP [α-^{35}S] (25) obtained from New England Nuclear, and the products were analyzed on very thin (0.35 mm) 8% polyacrylamide gels (26).

RESULTS AND ANALYSIS

The strategy for synthesis and cloning of the synthetic GRE is shown in Fig. 1. Two strands were synthesized; one is 40 nucleotides long (GRE.1) with 5′ *Hin*dIII and 3′ *Sst* I cohesive ends. The other strand (GRE.2) is 32 nucleotides long and anneals with GRE.1 to provide the recessed termini of these two restriction sites. This sequence has 34 nucleotides of identity with the MMTV sequence (ref. 29; Fig. 1). Both strands were synthesized by using the same doped mixtures of monomers. The two synthetic products were labeled with ^{32}P at the 5′ end by T4 polynucleotide kinase and were analyzed by electrophoresis on a 12% polyacrylamide sequencing gel (Fig. 2*A*). A 17-mer synthesized under the same conditions, except that pure phosphoramidites were used, is shown for comparison. The mutagenized compounds show faint bands above the major band, possibly related to mutant sequences of altered electrophoretic mobility or to infrequent double addition events. Full-length products were purified by preparative gel electrophoresis, annealed, and ligated to M13 mp11 that had been digested with *Hin*dIII and *Sst* I. After ligation, the preparation was digested with a mixture of *Sal* I and *Bam*HI to select against any uncut or religated molecules of the M13 mp11 vector. This material was used to transfect competent *E. coli* by the Hanahan method, and an amplified library of phage particles was obtained from ≈5000 transfection events. An equivalent transfection using vector religated in the absence of the synthetic GRE gave only six plaques, indicating that selection against the parental vector was very effective.

Substitution of the wild-type synthetic GRE sequence for the *Hin*dIII/*Sst* I portion of the M13 mp11 polylinker results in a net deletion of 3 base pairs. The inserted sequence does not contain any termination codons in the *lacZ* reading frame. Consequently, we expected the wild-type and most GRE mutants to give blue plaques in the standard α-complementation assay (20). Approximately 90% of the phage in the amplified library gave blue plaques. When 10 white plaques were analyzed by sequencing, all contained substitution mutations that produced either a TAA or a TGA termination codon in the *lacZ* reading frame, or else a frameshift.

Plaques were picked at random from this library, phage DNA was prepared from small cultures, and the sequences of the inserted oligonucleotides were determined by the chain-termination method. A typical gel is shown in Fig. 2*B*. So far we have sequenced 546 isolates. All of these contain the synthetic GRE; 224 of these are wild-type in the sense that they contain the original MMTV sequence (Fig. 1). The remainder contain one or more mutations. These results are summarized in Fig. 3 and Table 1.

Although the cohesive termini of the synthetic GRE were mutagenized at the same level as the remainder of the molecule, we expected that mutations in these sequences would not be efficiently cloned. It is apparent from inspection of Fig. 3 that there is a low yield of mutations that alter the *Hin*dIII and *Sst* I restriction sites. In the following analysis, we will consider the central 30 nucleotides, excluding these sites, as the "mutagenic target." This target corresponds to

```
                RECEPTOR PROTECTION
 -197         -------------------------               -1Ø2
   :       ------------------------                     :
5' TAAGTAAGTTTTTGGTTACAAACTGTTCTTAAAACGAGGAT-5Øbp-AGCTCTG-promoter-MMTV  WILD-TYPE GRE
                     *      *  *                                         CONTEXT

pBR322-TAAGTTTTTGGTTACAAACTGTTCTTAAAACGAGGAT-5Øbp-AGCTCTG-promoter-HSVTK FUNCTIONAL
                                                                         GRE
   TAAGTAAGTTTTTGGTTACAAACTGTTCTTAAAACGAGG.(deletion).CTG-promoter-β-gal CONSTRUCTS

  Hind III                            Sst I
  ----           GRE.1 (4Ø)              --
5' AGCTTAAGTTTTTGGTTACAAACTGTTCTTAAAACGAGCT 3'    SYNTHETIC MUTAGENIZED GRE
3'     ATTCAAAAACCAATGTTTGACAAGAATTTTGC     5'
       :         GRE.2 (32)            :
     -193                             -16Ø
```

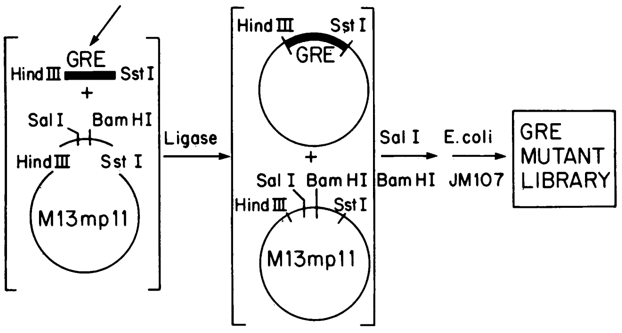

FIG. 1. Design and cloning of the synthetic GRE. The coding strand of the wild-type MMTV sequence (27) is shown. Numbering of the sequence is relative to the start of transcription, according to ref. 27. Dashes above the sequence indicate residues protected from DNase I, by binding of receptor, for the coding (upper dashes) and noncoding (lower dashes) strands (14, 15). Asterisks represent the positions of guanine residues protected from methylation by receptor binding and, conversely, whose methylation abolishes receptor binding (28). The wild-type MMTV sequence is aligned with two functional GRE constructs: (*i*) a deletion of sequences 5′ to position −193 with the MMTV promoter fused to the thymidine kinase gene of herpes simplex virus (*HSVTK*) (17) and (*ii*) a deletion of 56 base pairs (−104 to −160) fused to the β-galactosidase coding region (19). Both strands, GRE.1 and GRE.2, of the synthetic mutagenized GRE sequence are shown. The cloning strategy is diagrammed.

nucleotides 6–35 of Fig. 3, or −163 to −192 of the MMTV sequence (27).

We have passed several landmarks in the analysis of GRE mutations from the library. First, mutations at all 30 positions were identified. Next, single substitution mutations were identified at all 30 positions. At least two of the three possible single mutants have now been identified at every position within the target. Altogether we have observed 88 of the 90 possible substitution mutations. Of these, 74 have been recovered as single mutants. The 16 mutations that have not yet been recovered as single mutants are indicated in Fig. 3. It is interesting that the two mutations we have not yet recovered, even in multiple mutants, lie at one end of the target (C→G at position 33 and T→G at position 34; numbering refers to Fig. 3). Consequently, we have recovered all possible substitution mutations in the 27-nucleotide stretch numbered 6–32 in Fig. 3.

The method described here should, in principle, yield a completely random collection of point mutations. We have been interested in assessing how closely this expectation is met. Since the level of impurity in the phosphoramidites used in the GRE synthesis is 1/20 (5%), we expected an average of 1.5 substitutions per oligonucleotide in the 30-nucleotide target region. If each mismatch is either repaired randomly *in vivo* or resolved to a homoduplex by DNA replication, then the expected number of substitution mutations per clone is the same (1.5). The observed number, 0.83 (453 substitutions per 546 clones), is somewhat lower. The distribution of these mutations among the clones fits a random distribution, as estimated from the Poisson approximation, within statistical sampling error (Table 1). We do not know the reason for the lower than expected yield of mutations, and we do not yet know whether it is a general phenomenon or just a property of this particular GRE mutant library. Possible explanations include (*i*) selection for wild-type sequences at the annealing step, and (*ii*) lethal effects of mismatch repair. However, these mechanisms would not necessarily yield a population of mutants that so closely fits the Poisson distribution. A different type of explanation involves inactivation of the small amounts of phosphoramidites used as dopant, by moisture at the time of mixing.

Twelve chemically distinct types of misincorporation events could occur during synthesis of the GRE (3 different misincorporations for each of the 4 bases). Since we cannot determine which synthetic strand contributed a particular mutation, we are only able to distinguish six classes of transition and transversion mutations (Table 2). All six classes are observed. One class, the A·T→G·C transition, was recovered somewhat less frequently than expected if all substitutions are equally likely.

In addition to the expected substitution mutations, 14 single base insertions, and 6 single base deletions have been observed (Fig. 3). We do not know whether these were induced during chemical synthesis or *in vivo*. It should be noted, however, that they may be simply explained as errors during chemical synthesis. Each of the observed insertions could arise by double addition during a single synthetic cycle. If the phosphoramidite preparations contained a small amount of material unprotected by a trityl group at the 5′ terminus, then such events could occur. The deletion events could be explained in two ways. If detritylation is incomplete then molecules bearing a 5′ trityl group would be unable to couple in one synthetic cycle, but they could participate in subsequent cycles. Alternatively, deletions could be explained as a result of inefficiency of the capping reaction. This reaction, which acetylates unreacted 5′ hydroxyls, is intended to prevent unreacted molecules from participating in subsequent cycles of synthesis.

DISCUSSION

In spite of the few unexpected findings described above, it is clear that the mutagenesis procedure works almost exactly as planned. More than 95% of the mutations are substitutions, and we have not noticed any feature that differs significantly from random expectation by more than 2-fold.

The Statistics of Complete Mutational Libraries. We were interested in calculating the number of mutations that must be

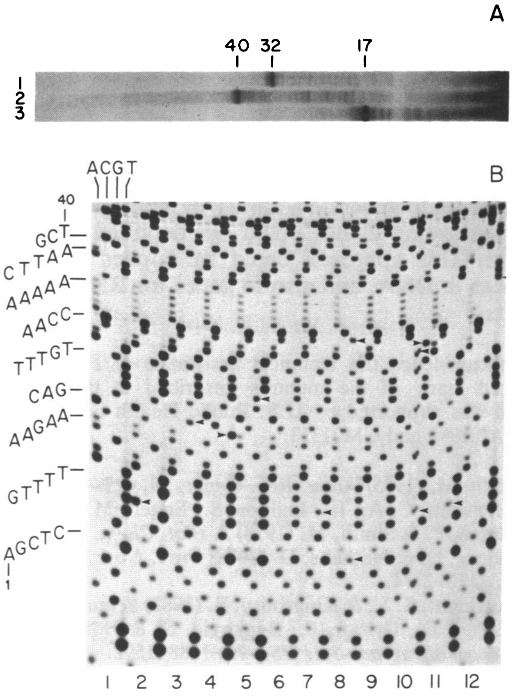

FIG. 2. (*A*) Gel analysis of mutagenized synthetic oligonucleotides. Mutagenized oligonucleotides (0.1 pmol) GRE.2 (lane 1, length 32) and GRE.1 (lane 2, length 40) were ^{32}P-labeled, using T4 polynucleotide kinase, and fractionated on a 12% polyacrylamide gel. An unmutagenized oligonucleotide of length 17 was also included for comparison (lane 3). (*B*) Sequence analysis of GRE mutations. A sequencing gel showing 12 isolates (isolate numbers g620–g631) from the GRE mutant library is shown. The wild-type GRE sequence is indicated to the left of the autoradiogram. Numbering is the same as that used in Fig. 3. Each set of sequence lanes is arranged ACGT, left to right. Arrowheads mark those bands corresponding to mutations. Three sequences are wild type (lanes 1, 6, and 9), seven are single mutants [lanes 2 (T→A at position 8), 3 (C→G at position 16), 4 (A→T at position 15), 5 (T→G at position 19), 7 (T→G at position 8), 11 (T→G at position 8), and 12 (C→T at position 33)], one is a double mutant [lane 8 (T→G at position 4 and C→G at position 26)], and one is a triple mutant [lane 10 (T→G at position 8, A→G at position 24, and A→T at position 25)].

analyzed to obtain a complete set of all possible substitution mutations. If the number of possible independent mutations is *s*, then the probability *P*(1), that one particular mutant will be present among *N* analyzed is given by an expression equivalent to that described for "representative" genomic libraries by Clarke and Carbon (30):

$$P(1) = 1 - [1 - (1/s)]^N. \quad [1]$$

By an extension of this reasoning, the probability *P*, that all *s* independent mutations occur within a set of *N* selected at random may be calculated as a product of *s* terms:

$$P = \prod_{i=N}^{i=N-s+1} [1 - (1 - 1/s)^i]. \quad [2]$$

Since calculation of this expression is time consuming for large values of *s*, an approximation is useful. For values of *N* that are large compared to *s*, *P* may be approximated as:

$$P = [1 - (1 - 1/s)^{N-s/2}]^s. \quad [3]$$

This expression may be rearranged to give an estimate of *N* in terms of *P* and *s*:

$$N = \frac{\ln(1 - p^{1/s})}{\ln(1 - 1/s)} + s/2 \quad [4]$$

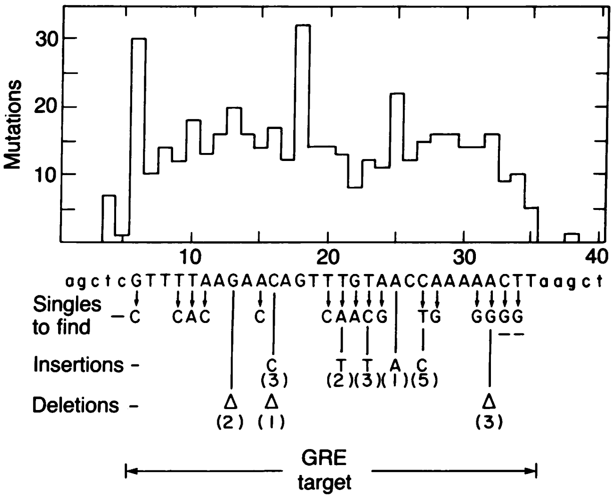

FIG. 3. Distribution of mutations in the synthetic GRE. The polarity of the sequence shown is that read from the sequencing gel (Fig. 2*B*), and it is the bottom strand (GRE.2) of the synthetic GRE diagrammed in Fig. 1. The total number of substitution mutations at each position is plotted. Substitution mutations that have not been recovered as single mutants are listed below the sequence (Singles to find). The two substitution mutations that have not been observed within the 30-nucleotide target region (see text) are underlined (C→G at position 33 and T→G at position 34). Single bases may be inserted to the left of the positions indicated to give the observed insertion mutant sequences. The number of isolates of each insertion and deletion mutation are indicated below the mutant base in parentheses.

Fig. 4 depicts the probability of completeness as a function of the size (*N*) of a collection of mutants. The parameter *s* may refer to the size of any set of independent equally probable classes of mutants. This analysis may be applied to the case of the 30-nucleotide GRE target as follows. About 200 mutations must be sequenced to give a probability of 95% that mutations at all 30 positions within the GRE target sequence would be observed. Since there are 0.83 substitution mutations per clone, this would require sequencing ≈240 clones. Similarly, 200 single mutants would give a 95% probability of including single mutants at all 30 positions. Since 40% of the clones are single mutants (Table 1), 500 clone sequences would meet this criterion. The same expression may be used to show that ≈700 mutations (850 clones) must be analyzed to give a 95% probability of identifying all 90 possible substitution mutations. Analysis of 500 mutations (600 clones) gives about an even chance (56%) of identifying all 90 possible substitutions.

The reasoning outlined above also applies to multiple mutations. There are 3915 possible double mutations in the GRE target [(90 × 87)/2]. To achieve a 90% chance of a complete set of doubles, a library would need to contain

Table 1. Mutation frequency in the synthetic GRE

Substitution mutations in clone	Number of clones	
	Observed	Predicted*
0	224 (41%)	238 (44%)
1	218 (40%)	198 (36%)
2	81 (15%)	82 (15%)
3	19 (3%)	23 (4%)
4	4 (0.7%)	3 (0.9%)

This table is based on the sequences of 546 clones, containing 453 substitution mutations (0.83 mutations per clone).

*Predicted by Poisson analysis.

Table 2. Summary of 453 substitution mutations isolated from the synthetic GRE library

	To				
From	C·G	G·C	A·T	T·A	Total
(8) C·G	—	38	57	50	145
(22) A·T	111	73	—	124	308

≈44,000 double mutants. Since our library contains about 15% double mutants (Table 1), a library constructed from the same mutagenic synthesis would require a minimum of roughly 300,000 independent clones to meet this criterion. This should be readily achievable.

Future Prospects. The complete collection of single substitution mutations in the GRE of MMTV described here, will be useful for both *in vitro* and *in vivo* analysis of GRE function. One approach is to assay the mutant sequences for *in vitro* glucocorticoid receptor binding activity. Preliminary evidence (P. Scheible and J. Cidlowski, personal communication) indicates that the synthetic wild-type sequence cloned in M13 binds receptor as efficiently as when embedded in the normal MMTV context. We have constructed an M13-based vector to assay biological activity of the mutants *in vivo*. The synthetic GRE can be conveniently inserted into this vector upstream from the MMTV promoter fused to the coding sequences of the chloramphenicol acetyl transferase (*CAT*) gene.

The methodology used in this work could be directly applied to analyze a wide variety of regulatory sequences that are small enough to be easily synthesized. The most convenient cases would be those in which the mutants could be functionally assayed directly in a (perhaps modified) M13 vector. In the case of longer sequences, such as large protein coding genes, the mutagenic target could be synthesized as a series of "cassettes." Each would be synthesized in both a wild-type and a mutagenized form, so that they could be assembled in any desired combination. This would reduce the amount of sequencing necessary to catalog a mutant collection of the desired level of completeness.

Although we were interested in obtaining single substitution mutations in the present study, for many purposes multiple mutations may be of interest. This could, of course, be readily achieved simply by adjusting the composition of the mutagenic phosphoramidite mixtures. One such application would be to generate a library of multiple mutants from which interesting clones could be identified by biological selection. This approach could be applied to regulatory sequences in DNA, and also to genes for RNA and protein products.

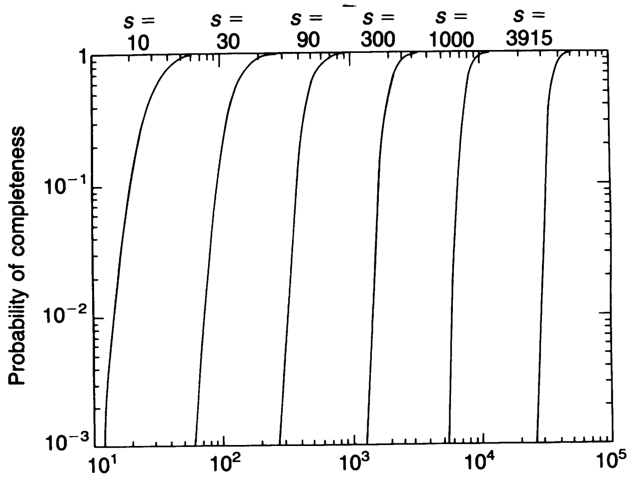

FIG. 4. Probability of completeness of a collection of random mutations. The probability (P) of completeness of a collection of N mutations was calculated from Eq. 2 (or Eq. 3 in cases in which the difference is less than the resolution of this figure). Curves are plotted for several different values of s, the size of a complete set of equally probable mutations.

We acknowledge the technical assistance of Pearl Cole, who sequenced many of the mutants described in this paper. This research was supported by U.S. Public Health Service Research Grants AI08998 and GM21313.

1. Smith, M. (1985) *Annu. Rev. Genet.* **19,** 423–463.
2. Hutchison, C. A., III, Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P. & Smith, M. (1978) *J. Biol. Chem.* **253,** 6551–6560.
3. Shortle, D. & Nathans, D. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2170–2174.
4. McKnight, S. L. & Kingsbury, R. (1982) *Science* **217,** 316–324.
5. Shortle, D., Grisafi, P., Benkovic, S. & Botstein, D. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1588–1592.
6. Hui, A., Hayflick, J., Dinkelspiel, K. & de Boer, H. A. (1984) *EMBO J.* **3,** 623–629.
7. Matteucci, M. D. & Heyneker, H. L. (1983) *Nucleic Acids Res.* **11,** 3113–3121.
8. Murphy, M. H. & Baralle, F. E. (1983) *Nucleic Acids Res.* **11,** 7695–7700.
9. Wells, J. A., Vasser, M. & Powers, D. B. (1985) *Gene* **34,** 315–323.
10. Lee, F., Mulligan, R., Berg, P. & Ringold, G. M. (1981) *Nature (London)* **294,** 228–232.
11. Huang, A. L., Ostrowski, M. C., Berard, D. & Hager, G. L. (1981) *Cell* **27,** 245–255.
12. Chandler, V. L., Maler, B. A. & Yamamoto, K. R. (1983) *Cell* **33,** 489–499.
13. Ponta, H., Kennedy, N., Skroch, P., Hynes, N. E. & Groner, B. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1020–1024.
14. Scheidereit, C., Geisse, S., Westphal, H. M. & Beato, M. (1983) *Nature (London)* **304,** 749–752.
15. Payvar, F., DeFranco, D., Firestone, G. L., Edgar, B., Wrange, O., Okret, S., Gustafsson, J. A. & Yamamoto, K. R. (1983) *Cell* **35,** 381–392.
16. Hynes, N., Van Ooge, A. J. J., Kennedy, A., Herrlich, P., Ponta, H. & Groner, B. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3637–3641.
17. Majors, J. & Varmus, H. E. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5866–5870.
18. Buetti, E. & Diggelmann, H. (1983) *EMBO J.* **2,** 1423–1429.
19. Lee, F., Hall, C. V., Ringold, G. M., Dobson, D. E., Luh, J. & Jacob, P. E. (1984) *Nucleic Acids Res.* **12,** 4191–4206.
20. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
21. Yanisch-Perron, C., Vieria, J. & Messing, J. (1985) *Gene* **33,** 103–119.
22. Hanahan, D. (1983) *J. Mol. Biol.* **166,** 557–580.
23. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
24. Bankier, A. T. & Barrell, B. G. (1983) in *Techniques in Nucleic Acid Biochemistry*, ed. Flavell, R. A. (Elsevier, Limerick, Ireland), Vol. B5, pp. 1–34.
25. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.
26. Sanger, F. & Coulson, A. (1978) *FEBS Lett.* **87,** 107–110.
27. Majors, J. & Varmus, H. E. (1983) *J. Virol.* **47,** 495–504.
28. Scheidereit, C. & Beato, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3029–3033.
29. Ucker, D. S., Firestone, G. L. & Yamamoto, K. R. (1983) *Mol. Cell. Biol.* **3,** 551–561.
30. Clarke, L. & Carbon, J. (1976) *Cell* **9,** 91–99.