J. Mol. Biol. (1994) 240, 421–433

Genetic Studies of the *lac* Repressor

XIV†. Analysis of 4000 Altered *Escherichia coli lac* Repressors Reveals Essential and Non-essential Residues, as well as "Spacers" which do not Require a Specific Sequence

Peter Markiewicz, Lynn G. Kleina, Christina Cruz Susannah Ehret and Jeffrey H. Miller

Department of Microbiology and Molecular Genetics and the Molecular Biology Institute University of California, Los Angeles, CA 90024, U.S.A.

Amber mutations have been constructed at 328 positions, corresponding to residues 2 to 329 in the *E. coli lac* repressor protein. Synthetic and naturally occurring nonsense suppressors have been used to insert, in series, 12–13 amino acids at positions specified by an amber (UAG) codon in the *lacI* mRNA. The resulting set of over 4000 single amino acid replacements in the *lac* repressor protein allows a detailed analysis of its substitution tolerance along the linear array of residues, and reveals structure-function relationships in *lac* repressor and in proteins in general. (1) There are two main regions in the repressor which are extremely sensitive to amino acid replacements. One, the amino-terminal 59 residues, has been implicated in DNA and operator binding by a large body of work. The second, extending from approximately residues 239 to 289/292, forms the repressor core and shares the most homology with other repressor and DNA binding proteins. (2) Throughout the rest of the protein, segments of 6 to 14 amino acids, which are highly tolerant to single amino acid replacements, appear to act as "spacers" between one or several hydrophobic residues that are relatively intolerant to substitutions. (3) We have replaced the amino acids in these tolerant regions with spans of alanine residues, from 5 to 13 amino acids. In all five of the regions tested, alanine replacements, sometimes of up to 8 amino acids, still allowed functional repressor, while deletion of the same residues destroyed repressor function. This reinforces the view that many regions of a protein do not require a specific sequence to serve as spacers between more important residues. (4) A distinct pattern of substitutions leading to the I^S phenotype suggests the location of residues involved in inducer binding. (5) A number of general substitution patterns can be recognized. For instance, proline is not tolerated at over 40 sites which tolerate all the other amino acid replacements. Another set of sites tolerates only non-polar amino acids, whereas a third set tolerates a subset of the smallest amino acids, (serine, alanine, glycine and cysteine, and sometimes threonine and valine). (5) Overall, 93 of 328 sites (28%) tolerate all 13 amino acids tested, and 144 of 328 (44%) tolerate 12/13 or all 13 substitutions. We judge that 192 of 328 sites (59%) are generally tolerant to substitutions.

Keywords: *lac* repressor; essential residues; non-essential residues; spacers

1. Introduction

The study of protein structure-function relationship depends on the ability to create and analyze altered proteins. We have examined the *in vivo* effects of a large number of mutant *Escherichia coli lac* repressor proteins, many of which have been created by the suppression of nonsense mutations (Miller *et al.*, 1979; Kleina & Miller, 1990). This approach involves first obtaining a nonsense mutation, usually UAG (amber) at a specific site in a protein and then using characterized nonsense suppressor tRNAs to specifically insert different amino acids at the corresponding position in the protein. To increase the number of amino acids that can be inserted at UAG sites, we and our co-workers constructed a set of synthetic nonsense suppressor genes which were then introduced on plasmid vectors into *E. coli* strains (Normanly *et al.*, 1986, 1990; Kleina *et al.*, 1990). The resulting nonsense suppressors, together with those previously characterized, allow the insertion of 13 different

† Paper XIII in this series is Kleina & Miller (1990).

0022-2836/94/300421-13 $08.00/0

Table 1
Efficiency of suppression and specificity of insertion of tRNA suppressors

Suppressor	Gene	Amino acid inserted	Efficiency of suppression at UAG codons (%)
Su1	*supD*	Serine	6-54
Su2-89	*supE*	Glutamine	32–60
Su3	*supF*	Tyrosine	11–100
Su5	*supG*	Lysine	0·6–3; 5–29
Su6	*supP*	Leucine	30–100
$tRNA^{Ala}_{CUA}$	synthetic	Alanine	8–83
$tRNA^{Cys}_{CUA}$	synthetic	Cysteine	17–54
$tRNA^{GluA}_{CUA}$	synthetic	Glutamic Acid (80%) Glutamine (20%)	8–100
$tRNA^{Gly\ 1}_{CUA}$	synthetic	Glycine	24–100
$tRNA^{HisA}_{CUA}$	synthetic	Histidine	16–100
$tRNA^{Lys}_{CUA}$	synthetic	Lysine	9–29
$tRNA^{Phe}_{CUA}$	synthetic	Phenylalanine	48–100
$tRNA^{ProH}_{CUA}$	synthetic	Proline	9—60
FTOIRΔ26	synthetic	Arginine	4–28; 4–47

Data has been compiled from previous work on naturally occurring (Miller & Albertini, 1983) and synthetic (Kleina *et al.*, 1990) suppressors. The efficiency of suppression is expressed as a range of the highest and lowest values obtained in different contexts. The Su2-89 suppressor is from Bradley *et al.* (1981), and the FTOR1 26 suppressor from McClain & Foss (1988).

Table 2
Phenotypes resulting from different lacI *mutations*

Altered repressor function/property	Phenotypic symbol	Mutation dominant/recessive (d)	(r)
DNA binding	I^-	d	
Folding	I^-	r	
Aggregation	I^-	r	
Inducer binding	I^s	d	
Allosteric transition	I^s	d	
Tight binding to DNA	I^s, I^r		
Reversed allosteric transition	I^r, I^{rc}		
Stability increased by inducer binding	I^r		

Some of the altered repressors which have been characterized are shown. The I^- designation is for different defective repressors which can no longer block transcription of the *lac* genes. The I^s symbol denotes repressors that bind to operator but are not induced by IPTG. Several types of altered repressors display this characteristic, including those that bind operator more tightly. This latter class also displays a partial reverse induction profile (in which repression increases with increasing IPTG concentration), which is designated as I^r (Chamness & Wilson, 1970). I^r of I^{rc} (reverse curve; Myers & Sadler, 1971) repressors may also result from altered allosteric transition or from a stabilization of repressor by inducer binding. Temperature-sensitive derivatives of each type of repressor can also occur. For a more detailed description, see Kleina & Miller (1990), Mueller-Hill (1975) or Miller (1978).

amino acids at positions in a protein corresponding to a UAG site (Normanly *et al.*, 1990; Kleina *et al.*, 1990; McClain & Foss, 1988), and an additional amino acid at UGA sites. Table 1 depicts the amino acids which can be added by suppression of nonsense mutations at reasonable efficiencies.

The *E. coli lac* repressor is a tetrameric protein composed of four identical subunits, each consisting of 360 amino acids (Mueller-Hill, 1975; Farabaugh, 1978). The repressor consists of several domains. The amino-terminal 59 amino acids contain the DNA and operator binding sites, and the remaining portion (60 to 329) of the molecule contains the binding site for inducer and the dimer association sites (Mueller-Hill, 1975; Platt *et al.*, 1973; Ogata & Gilbert, 1978; Schmitz *et al.*, 1976; Miller & Schmeissner, 1979). The carboxyl-terminal 30–31 residues (330 to 360) are not required for the formation of active dimers, but are required for the dimer to tetramer transition (Platt *et al.*, 1970; Alberti *et al.*, 1991), and appear to contain a leucine zipper (Alberti *et al.*, 1991; Chakerian *et al.*, 1991). Numerous investigations have identified several classes of altered *lac* repressors (Mueller-Hill, 1975; Schmitz *et al.*, 1976; Miller & Schmeissner, 1979; Chamness & Willson, 1970; Miller, 1978) defective in one of these functions. Table 2 displays some of the phenotypes resulting from different *lacI* mutations.

We previously described the use of nonsense suppressors on a set of 141 nonsense mutations in the *lacI* gene (Kleina & Miller, 1990), which encodes the 360 amino acid *lac* repressor monomer. This work resulted in the generation of close to 1600 amino acid replacements. Using site-directed mutagenesis, we have constructed an additional 188 amber sites in the *lacI* gene. Here, we describe the effects of suppressing all of these mutations, spanning residues 2 to 329, with each of the characterized amber suppressors. (Amino-terminal fragments containing > 330 residues are partially or fully active *in vivo* and were not studied.) The replacement of 12–13 amino acids at each of the 328 amber sites results in over 4000 altered *lac* repressors, and allows a virtual genetic image reconstruction of the functional regions of the protein. We compare the results reported here from those of other large collections of altered proteins, including the pioneering studies of Perutz and co-workers on the hemoglobins and myoglobins (Perutz, 1965; Perutz *et al.*, 1965; Perutz & Lehmann, 1968), studies of HIV-1 protease (Loeb *et al.*, 1989), an N-terminal fragment of the lambda repressor (Bowie *et al.*, 1990), and T4 lysozyme (Rennell *et al.*, 1991). Additionally, we compare the substitutional tolerance of individual sites in the repressor with the evolutionary variability of those sites in homologous proteins.

2. Materials and Methods

(a) *Genetic methodology*

All genetic methods and assays were carried out as described by Miller (1972) and Kleina & Miller (1990). Mutants were constructed as described previously (Kleina & Miller, 1990), except that in many cases we crossed mutations directly from f1 onto the F′*lacpro* episome, without first cloning them onto a plasmid.

(b) *Correlation of tolerance and conservation*

The correlation coefficient, C_a (Schulz & Schirmer, 1979) is defined as:

$$C_a = \frac{(w)(x) - (y)(z)}{\sqrt{(x+y)(x+z)(w+y)(w+z)}}$$

where, as a percentage, or fraction: w is positive correct prediction, x is negative correction prediction, y is underprediction, z is overprediction, and where $w+x+y+z = 1$. Here, w is conserved, intolerant, x is not conserved, tolerant, y is predicted conserved, but not: i.e. nonconserved, intolerant, z is conserved, tolerant;

$$w = 97/328 = 0{\cdot}30; \qquad x = 163/328 = 0{\cdot}50;$$
$$y = 37/328 = 0{\cdot}11; \qquad z = 31/328 = 0{\cdot}09.$$

$$C_a = \frac{0{\cdot}15 - 0{\cdot}0099}{\sqrt{(0{\cdot}61)(0{\cdot}59)(0{\cdot}41)(0{\cdot}39)}}.$$

$C_a = 0{\cdot}58$, which is highly significant (Matthews, 1975).

(c) *Scoring for + and − among altered repressors*

In order for a replacement to score as +, it must result in a repressor with greater than 8 to 10% activity. Although the version of the Su2 glutamine-inserting suppressor (Su2–89) that we are using operates with high efficiency, the normal Su2 operates at a very low efficiency at certain amber sites. When coupled with the amber site corresponding to residue 117, glutamine is inserted at only 0·8 to 1·0% efficiency (Miller & Albertini, 1983; J. H. Miller, unpublished results). Since the repressor is overproduced 10-fold (Mueller-Hill *et al.*, 1968) in the strain employed, that means that glutamine is being inserted at 8 to 10% of the level of a typical wild-type strain. Even though glutamine is the wild-type amino acid at position 117, repression is not fully restored with the normal Su2 suppressor. Therefore, replacements must result in greater than 8 to 10% activity in order to be scored as a full +.

3. Results

(a) *Effects of amino acid replacements*

Figure 1 compiles all of the accumulated data for the 13 × 328 amino acid replacement matrix in the *lac* repressor. This Figure shows the fractional tolerance to substitution for each residue in the protein, without considering each substitution in detail. More quantitative data on the effects of specific replacement are given by Kleina *et al.* (1990), and additional quantitative data will be published elsewhere. On examining the linear array of sites in the protein, several features strike the eye. It is clear that large portions of the protein (60%) are tolerant to substitutions. However, two subregions sensitive to replacements are evident. The first, consisting of the amino-terminal 59 residues, has previously been shown to be involved in operator and DNA binding (Mueller-Hill, 1975; Platt *et al.*, 1973; Ogata & Gilbert, 1978), and contains the helix-turn-helix DNA binding motif seen in many regulatory proteins (McKay & Steitz, 1981; Kaptein *et al.*, 1985). Analysis of the properties of sites in this region shows that the sensitivity to amino acid exchanges can be correlated with the three-dimensional structure of this region as determined by NMR (Kaptein *et al.*, 1985), and depicted in Figure 2. Three helices are defined in the first 51 amino acids. Helix II is the recognition helix, including amino acids 16 to 23 (Lehming *et al.*, 1987, 1988). It can be seen that all amino acids in this helix are extremely sensitive to substitutions. It is particularly striking that for helix III, the residues whose side-chains point inwards making internal contacts (e.g. Val38, Ala41, Met42 and Tyr47) are sensitive to replacements, while those facing the exterior (e.g. Glu36, Gln39, Ala43, Glu44 and Asn46) are very tolerant of substitutions. A similar pattern is seen for helix I, with Val9, Ala10 and Ala13 intolerant, while other residues pointing outward are more substitutable. Thr5 is highly intolerant, but is believed to be in a close non-specific contact with the operator DNA (Kaptein *et al.*, 1985). The residues forming the loops between the helices are for the most part tolerant to substitutions.

The second region containing many sites intolerant to substitutions extends from amino acids 239 to 241 until amino acids 289 to 292. The experimental identification of this region as a crucial portion of the repressor has emerged from this study. We have carried out a computer search in GenBank and other data bases for proteins with significant homologies to *E. coli lac* repressor, and have found that the region of the repressor, aside from the amino-terminal DNA binding domain, which shares the greatest homology with other proteins is in fact in the region from residues 239 to 289 (see below, and Figures 3–4).

Figure 1 also reveals stretches of amino acids which appear to be almost completely tolerant to substitutions. For instance, very few substitutions between residues 100 to 112, 129 to 145, 151 to 160, 206 to 217 or 305 to 318 appear to be deleterious. Throughout the protein, it appears as if long stretches of tolerant residues separate one or small clusters of sites which are sensitive to substitutions. In most of these cases the sensitive residues form hydrophobic clusters which are probably buried or partly buried in the interior of the protein.

(b) *Inducer binding*

Replacements resulting in the I^S phenotype, inability to respond to inducer, are scored in the upper portion of Figure 1. Since the vast majority of sites in the repressor are not involved in inducer binding, only those sites where an effect occurs are shown. I^S mutations affect binding of the inducer molecule and/or the allosteric transition that reduces binding affinity to operator. The strongest effects are shown as darkened boxes (see legend to Figure 1). In addition to a large array of I^S sites from amino acids 66 to 99, and scattered sites over the next 90 residues, the remainder of the protein contains distinct clusters of I^S sites at regularly spaced intervals (see lower portion of Figure 1 and Miller *et al.*, 1979). Exactly how these sites are arranged in three dimensions will have to await the elucidation of the repressor structure.

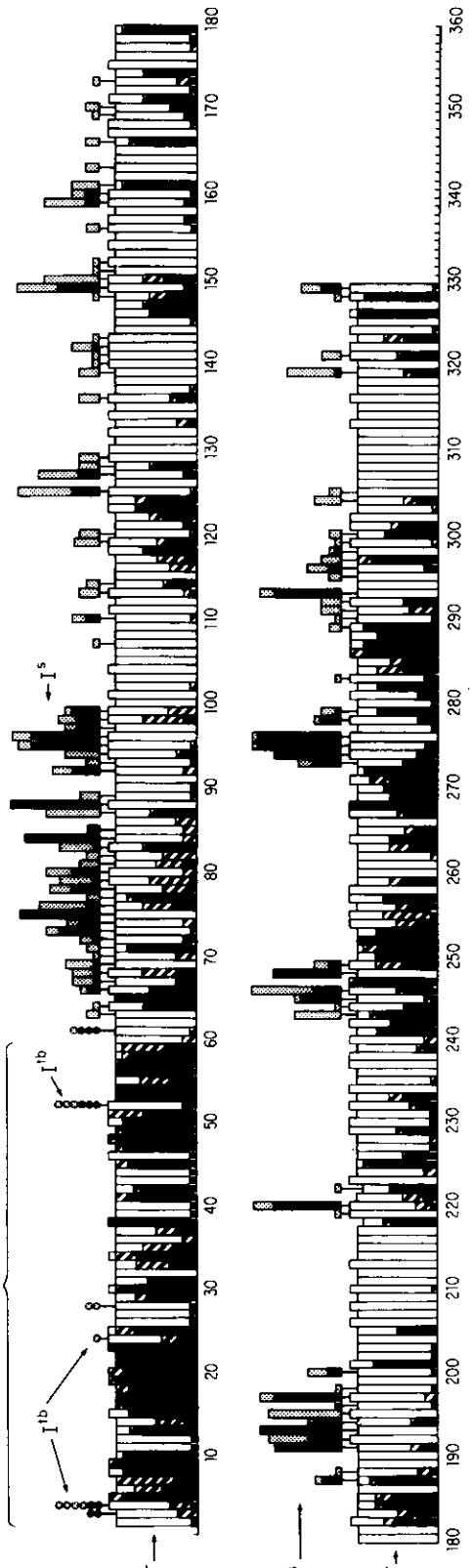

Figure 1. Effects of amino acid replacements in the *lac* repressor. The consequences of over 4000 single amino acid replacements are depicted here. The repressor protein is shown as an array of 360 sites (amino acids). An amber (UAG) mutation has been constructed at each position in the *lacI* gene, one mutation per construct, corresponding to residues 2 to 329 of the *lac* repressor (see Materials and Methods). At each position between residues 2 and 329, 12–13 amino acids can be exchanged through nonsense suppression and a bar has been drawn on the linear axis of the repressor polypeptide. The height of the bar represents the number of exchanges made (12 or 13). Each bar is divided into invisible segments, the number of segments corresponding to the number of replacements made. When a replacement generates the I^- phenotype, a corresponding portion of the bar is filled in. If the replacement does not alter the I^- phenotype, then the segment is left open or blank. Intermediate or temperature-sensitive effects are represented by a diagonal stripe. Thus, if all of the replacements destroy repressor function, then the whole bar is filled in (see position 22 or 38). If none of the replacements creates the I^- phenotype, then the whole bar is left open (see position 153 or 180). Above the 1st set of bars the replacements which create I^s repressors are shown. Here the bars were draewn only for those replacements causing this change in phenotype, the height of the bar indicating the number of replacements that generate the I^s repressor. The black segments are strong effects, and the open or dotted segments are weaker effects. The I^s phenotype can sometimes result from tighter binding to operator, as occurs for certain replacements in the amino-terminal portion (residues 1 to 61) of the repressor. These repressors are depicted as I^{1b} (tight-binding) and are represented by circles above the main set of bars. Filled-in circles indicate strong effects, and open circles weaker effects.

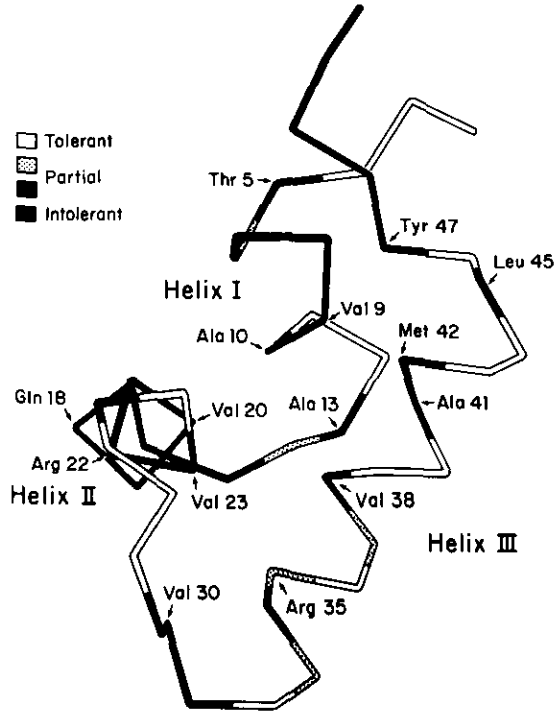

Figure 2. Three-dimensional configuration of *lac* repressor headpiece. Kaptein and co-workers have used NMR spectroscopy to determine the structure of the amino-terminal 51 amino acids of *lac* repressor (Kaptein *et al.*, 1985). Based on coordinates kindly provided by Dr R. Kaptein, we have used a Silicon Graphics Personal Iris workstation to generate this view of the repressor headpiece. The first 2 helices (I and II) are part of the helix-turn-helix motif seen in a number of regulatory proteins. The degree of tolerance to substitution, as derived from the data in Figure 1, is shown here. The residues most sensitive to substitution are shown in the darkest shading.

(c) *Homologies with related proteins*

One interesting application of the substitution data for *lac* repressor is to compare it to the observed evolutionary variation of equivalent sites of related proteins. To this end, we generated a multiple alignment of proteins homologous to *lac* repressor. FASTA (Lipman & Pearson, 1985) was used in the initial search against the PIR and translated GenBank (GenPept) protein data bases. FASTA recovered 16 proteins, all prokaryotic regulatory proteins except for the periplasmic d-ribose binding proteins of *E. coli* and *Salmonella typhimurium* (see Figures 3–4). Additional alignments of the 330 to 360 regions with eukaryotic proteins such as with *gag* polymerase and filamentous proteins were rejected after examining the pairwise alignments. Some of the matched proteins (*gal*, *cyt* repressor) have been aligned previously to *lac* repressor (von Wilcken-Bergman & Mueller-Hill, 1982). Additional searches using BLAZE and BLAST (Altschul & Lipman, 1990; Brutlag, D. L., Dautricourt, J.-P., Diaz, R., Fier, J. & Stamm, R., unpublished results) recovered 14 of the 16 proteins as highest scoring. It should be noted that other helix-turn-helix repressors such as lambda and cro do not appear in the matches. Though most of the matching proteins are bacterial repressors, only one of them, the *Klebsiella pneumoniae lac* repressor is identical in function to *E. coli lac* repressor.

Pairwise alignments between *lac* repressor and the database proteins revealed a number of conserved regions in the sequence, two of which were the most prominent. One, centered on the amino-terminal region of the proteins, comprises the helix-turn-helix DNA recognition motif. This match was found for all the proteins in the group which bind DNA. The *E. coli* periplasmic d-ribose binding protein (Groarke *et al.*, 1983) lacks this region, while the *Vibrio ScrR* protein (Blatch & Woods, 1991) consists only of this region. The second large region with extensive homologies begins at approximately residue 239 (*E. coli lac* repressor numbering) and extends at least to residue 289. It comprises a new conserved region for this family of proteins. Based on these data, a multiple sequence alignment was generated using either CLUSTALV (Higgins & Sharp, 1988) or the progressive similarity alignment of Feng & Doolittle (1990). The two programs gave very similar results. Various combinations of proteins and alignment parameters were used to ensure the robustness of the alignments. The maltose repressor (Reidl *et al.*, 1989) despite low homology in the 200 to 300 region, was aligned throughout its entire length, but the comparable region of the ribitol repressor (Wu *et al.*, 1985) could not be satisfactorily aligned. Therefore, only the first 60 amino acids of this protein was used in the alignment. No manual editing of the alignment was performed beyond these deletions. Figures 3 and 4 present the results of the alignment. Above the alignment we have placed a box whose shading indicates substitutional tolerance for that site in *E. coli lac* repressor, with intolerant residues black and tolerant ones white. This one-dimensional representation allows us to compare the conserved sites with the substitution data for *E. coli lac* repressor from Figure 1. An examination of Figures 2–3 shows a highly significant correlation between conserved sites in the alignment and sites which are intolerant to substitution in *E. coli lac* repressor, and conversely, between sites which are not conserved in evolution and sites which are tolerant to substitution. (We consider the implications of this correlation below.) For instance, there are 134 (41%) intolerant sites and 194 (59%) tolerant sites. Likewise, there are 128 (39%) conserved sites, and 200 (61%) non-conserved sites. If there were a random distribution of conserved and non-conserved sites, then among the intolerant sites we would expect 52 conserved sites, whereas we find 97, and we would expect 82 non-conserved sites, whereas we find 37. Similarly, among the tolerant residues, we would expect 76 conserved, whereas we find 31, and 118 non-conserved, whereas we find 163. These results are highly significant ($X^2 = 156$; $p \ll 0{\cdot}005$). Also, the correlation coeffi-

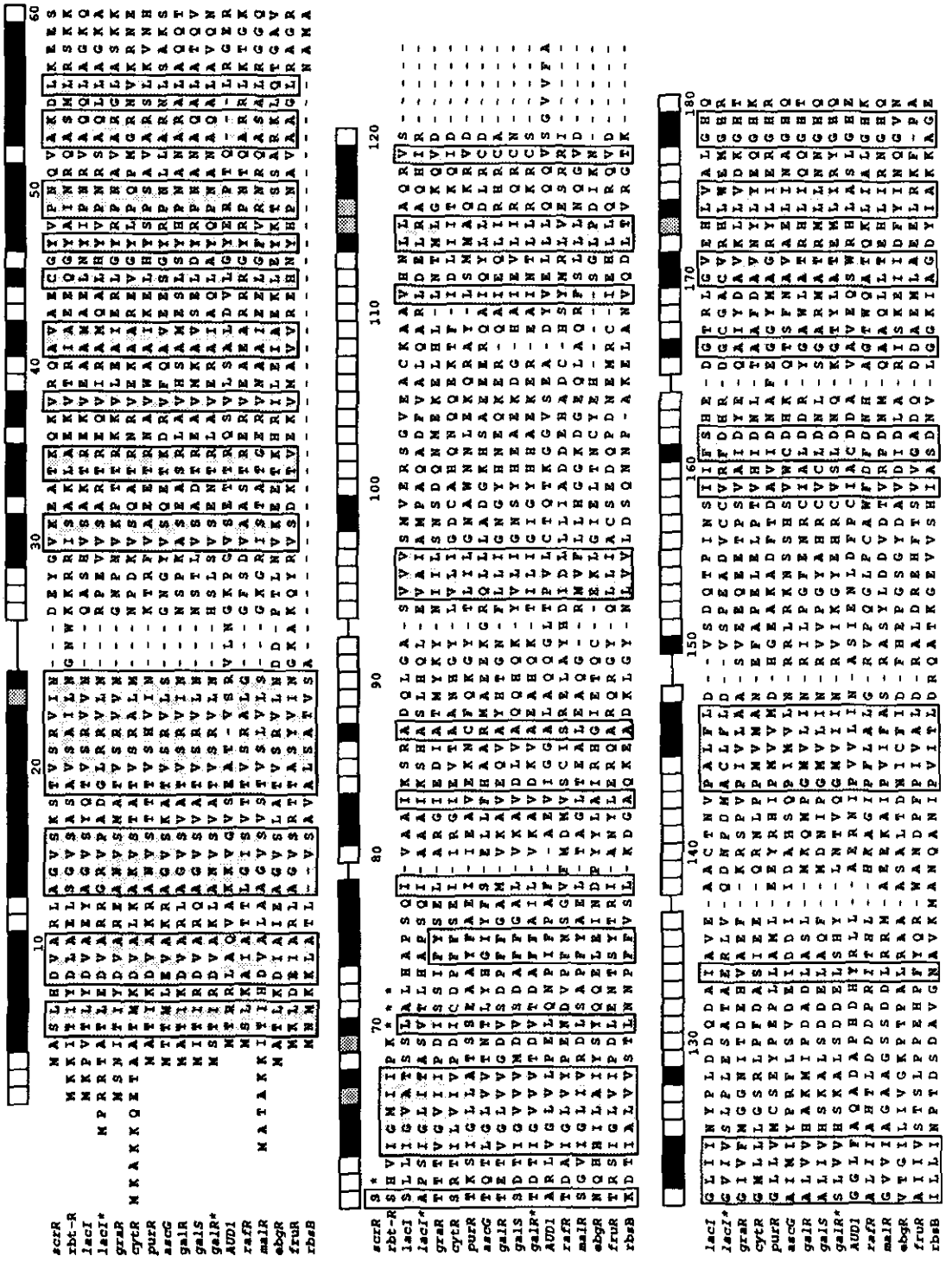

Fig. 3.

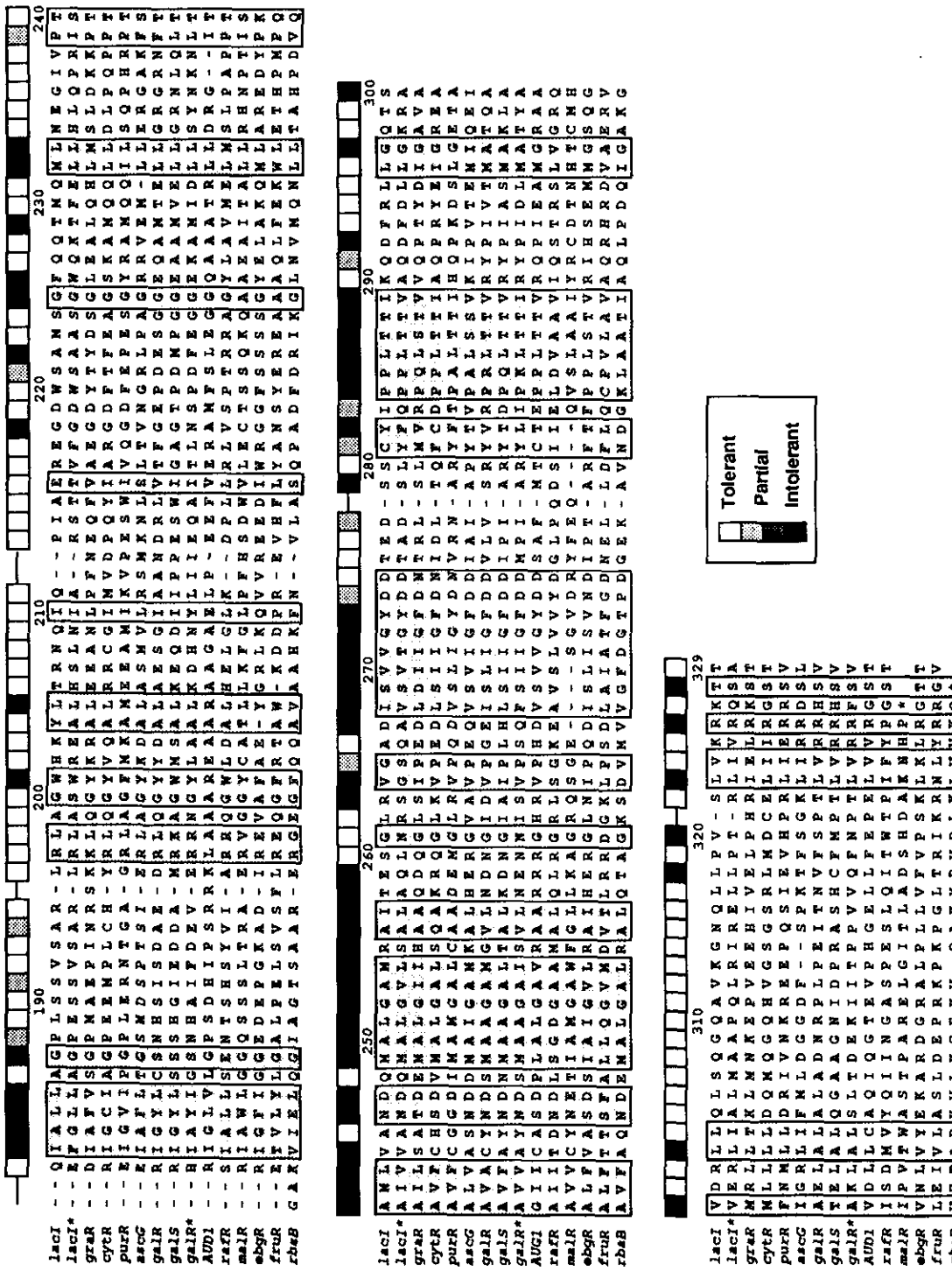

Fig. 4.

Figure 5. Deletions in the *lacI* gene. Different lengths of the *lacI* gene were deleted by site-directed mutagenesis (see Materials and Methods). The amino acids missing from the resulting repressors are shown here. Thus, delta 100–112 means that amino acids 100 to 112 are missing from the repressor. Dark bars indicate that the repressor has no activity, and an open bar (as in the case of the single amino acid deletions at positions 208 and 313) indicates that the repressor has activity (see also Table 3). The information in Figure 1 has been summarized on the main line here to indicate the degree of tolerance to substitutions of different regions of repressor. The darker the shading, the more intolerant to substitution the region is.

cient C_a (Schulz & Schirmer, 1979) is 0·58 (see Materials and Methods), which is highly significant (Matthews, 1975).

(d) *Regions of repressor that are tolerant to substitutions*

The substitution profile shows that segments of the protein are virtually insensitive to single amino replacements. What function do these stretches of up to 14 amino acids serve? To test whether these stretches could be eliminated, we deleted the residues comprising the substitution-tolerant regions using oligonucleotide-directed site-specific mutagenesis. Figure 5 shows the regions we deleted. In each case, the resulting repressor lost function *in vivo*. Namely, it could no longer repress the *lac* operon. We also constructed mutants in which we replaced the amino acid stretches with runs of alanines, as depicted in Figure 6. We used alanine because it was the most common functional replacement in our data set. In several cases replacements of up to eight amino acids with alanine runs still resulted in repressors with apparently normal function, since the *lac* genes were still IPTG inducible (see Table 3). Because some stretches can be replaced with a completely different amino acid sequence, but not deleted, it may be that these stretches serve to correctly space crucial residues in the protein, and that a unique sequence is not necessary for this function. Comparison to the three-dimensional structure, when it becomes avail-

Figures 3–4. Multiple sequence alignment of proteins with sequence homology to *lac* repressor. *scrR*, *Vibrio alginolyticus* sucrose uptake repressor (Blatch & Woods, 1991); *rbt-R*, ribitol operon repressor from *Klebsiella* (Wu *et al.*, 1985); *lacI*, *E. coli lac* repressor (Farabaugh, 1978); *lacI** (*klbR*), *Klebsiella pneumoniae lac* repressor (Buvinger & Riley, 1985); *cytR*, *E. coli cyt* repressor (Valentin-Hansen *et al.*, 1986); *purR*, *E. coli* purine biosynthesis operon repressor (Rolfes & Zalkin, 1988); *galR*, *E. coli* galactose repressor (von Wilcken-Bergman *et al.*, 1982); *ebgR*, evolved beta-galactosidase operon repressor (Stokes & Hall, 1985); *rafR*, *E. coli* raffinose operon repressor (Aslandis & Schmitt, 1990); *fruR*, fructose phosphotransferase system repressor (Jahreis *et al.*, 1991; unpublished GenBank submission X55457); *malR*, *E. coli* maltose repressor (Altshul *et al.*, 1990); *rbsB*, d-ribose periplasmic binding protein (Groarke *et al.*, 1983); *graR*, catabolite repression protein for alpha-amylase gene expression in *B. Subtilis* (Henkin, T. M., Grundy, F. J., Nicholson, W. L. & Champliss, G. H., unpublished GenBank submission M85182); *ascG*, *asc* operon regulatory protein in *E. coli* (Hall & Xu, 1992); *galS*, isorepressor of the *gal* regulon in *E. coli* (Weickert & Adhya, 1992); *galR**, *H. influenzae gal* repressor (Maskell, D. J., Szabo, M. J., Deadman, M. & Moxon, E. R., unpublished GenBank submission X65934); *AUD1*, amplification element of *S. lividans* (Piendl, W., Eichenseer, C., Viel, P., Altenbuchner, J. & Cullum, J., unpublished GenBank submission X65465). The alignments were generated using the algorithm of Feng & Doolittle (1990). Only the first 72 amino acids of the ribitol repressor were used in the alignment, since the program and CLUSTAL V (Higgins & Sharp, 1988) generated excessive gaps in the remainder of the sequence during several alignment runs. Regions of homology were boxed using a specific algorithm we developed to avoid "curve-fitting". First, for each column of amino acids at a particular site, scores from a normalized log-odds matrix (as in Feng & Doolittle, 1990) were averaged. Resulting scores ranged between 0·1 to 0·9, and columns with scores between 0·1 to 0·4 were boxed. Second, the number of the most common amino acid at any particular site was derived. The second score was used to correct for columns with log-odds scores > 0·40 that consisted of many identical amino acids (such as lysine) which have very low conservation scores in the log-odds matrix. Columns with log-odds scores between 0·4 and 0·5 for which the number of the most common amino acids was > / = 5 were boxed. Columns with log-odds scores between 0·3 and 0·4 for which the number of the most common amino acid was < 5 were unboxed. Correlation coefficients were taken from the resulting conservations and substitutional tolerance.

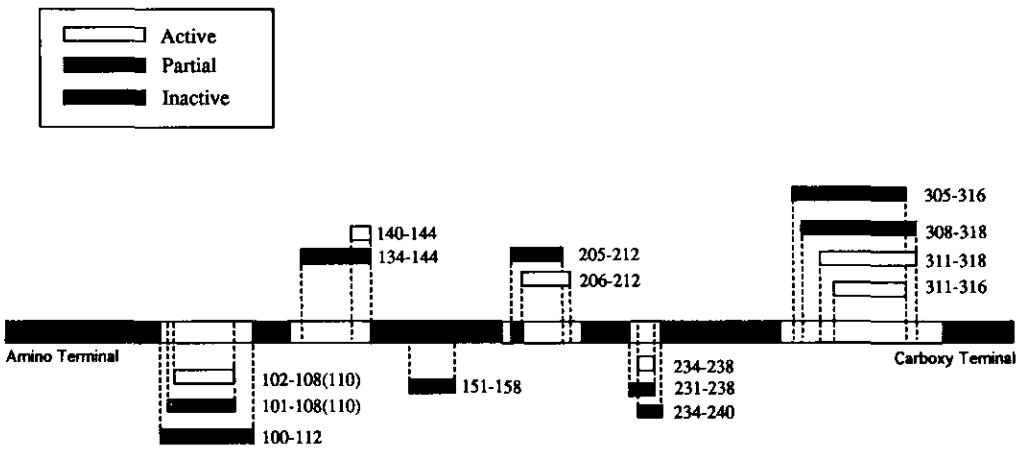

Figure 6. Alanine replacements in the *lac* repressor. Site-directed mutagenesis was used to replace segments of the *lac* repressor with continuous runs of alanines. The numbers indicate the extent of the replacements. Dark bars indicate inactive repressor, open bars indicate active repressors, and shaded bars indicate partially active repressors. See also Table 3 and the legend to Figure 5.

able, will allow the determination of whether the tolerant stretches are contained in specific structural elements, such as surface loops or helices.

(e) *Substitution patterns*

Several substitution patterns are evident. (A complete tabular representation of the substitution data will be published elsewhere). It is clear that *proline is not tolerated at many otherwise tolerant* positions. At many of the sites shown in Figure 1, only a single substitution partially or fully destroys protein function. In all cases, these represent the replacement of the wild-type residue by proline. Out of the 328 sites examined, 144 (44%) are extremely tolerant to substitutions, in that (excluding proline for the moment) they accept all 12 of the amino acids inserted by the nonsense suppressors. However, 51 (34%) of these otherwise tolerant sites *do not tolerate proline.*

Another set of sites tolerates only hydrophobic amino acids, as shown by the examples depicted in Table 4. At a number of sites, only certain small amino acids (glycine, alanine, serine and cysteine, and sometimes threonine and valine) are tolerated, as summarized in Table 5.

Table 3
Assays of β-galactosidase in strains with altered repressors

	Beta-galactosidase activity	
lacI region on plasmid	No IPTG	IPTG (10^{-3} M)
None	5600	5500
Wild-type	4·1	840
Multiple alanine replacements		
102–108	5·9	660
101–108	22	270
140–144	3·3	2700
151–158	180	850
206–212	2·4	770
234–238	4·5	1500
311–316	1·6	760
311–318	1·4	700
305–316	810	1700
Deletions		
100–112	5000	5500
206–212	8000	8000
311–318	2500	3160

Plasmids carrying each of the mutated *lacI* genes were constructed as previously described (Kleina & Miller, 1990), and put into a strain deleted for *lac* and *pro*, but carrying the *lac* region on an F′ *lacpro* episome. The episomal *lac* region carries a frameshift mutation, *1378* (Calos & Miller, 1981), in the *lacI* gene. Therefore, beta-galactosidase is synthesized constitutively from the episomal *lac* promoter, unless it is repressed by the *I* gene product synthesized from the plasmid. Beta-galactosidase was assayed after growth at 37 °C. Assay conditions and units are as previously described (Miller, 1972).

(f) *Reliability of data*

The efficiency of suppression is rarely 100%, and the efficiency of suppression varies from suppressor to suppressor and from codon to codon (Miller & Albertini, 1983; Bossi, 1983), However, by using the $lacI^{Q}$ allele, which results in a tenfold overexpression of the repressor (Mueller-Hill *et al.*, 1968), we can compensate for the lack of complete suppression in almost all cases. For the majority of suppressed proteins, the amount of repressor being produced varies within a threefold range, near to the normal amounts of wild-type repressor, since we are using a single copy F′ carrying the *lacI* gene. Thus, we are not creating false positives by greatly overproducing the repressor. Also, the assays we are using (Kleina & Miller, 1990) to determine repressor *function in vivo can recognize, as partially defective,* repressors with as much as 8 to 10% activity (see Materials and Method).

Leakiness of the amber mutation is not a factor in the experiments reported here. All amber mutations allow some level of transmission even in strains lacking known suppressors. Studies of fusion strains estimate this level to vary between 0·01% and 2%

Table 4

Substitution patterns of selected sites in the repressor

	Amino acid appearing at position														
Wild-type	Ile	Leu	Phe	Val	Pro	Cys	Tyr	Ala	Gly	Ser	His	Gln	Arg	Lys	Glu
Residue															
Ile64	(+)	+	+		+	+	+	+	−	−	+	−	−	−	−
Val66		+	+, s	(+)	−	+	+, s	+	−	−, ts	+	−	−	−	−
Leu71		(+)	+			−	+	−	−	−		−, ts	−	−	−
Ile123	(+)	+	+	+	−	+	−	+	−	−	−	−	−	−	−
Ile124	(+)	+	+	+	−	+	+, ts	±	−	−	+	−	−	−	−
Phe147		+	(+)		−	±	+	+	−	−	+	−, ts	−	−	−
Phe161		+, ts	(+)		−, ts	−		−	−	−, ts	+	−	−	−	−
Leu174		(+)	+		−	+	+	+	−	−, ts	−	−, ts	−		−
Ile182	(+)	+	+	+	+	+	+	+	−	−	−	−	−	−	−
Leu243		(+)	+		−	+	−	±	−	−	−	−	−	−	−
Val270		+	+	(+)	−	+	−	+	−	−	−	−	−	−	−
Leu286		(+)	−	+	−	±	−	±, ts	−	−	−	−	−	−	−
Ile289	(+)	+		+	−	+	−	−	−	−	−	−	−	−	−

All amino acid replacements were made by using the amber suppressors indicated in Table 1. See also Kleina & Miller (1990). Different designations are used for the I phenotype. In general, + refers to no significant detected alteration (greater than 200-fold repression of beta-galactosidase in cases where measured); ± indicates altered repression, but retention of the ability to repress beta-galactosidase synthesis 20 to 200-fold; − usually designates less than 4-fold repression. These designations are only a rough guide. Substitutions resulting in a temperature-sensitive phenotype are indicated by ts. Substitutions resulting in a loss of response to inducer are indicated as s for I^s repressors, or ws for a weaker I^s phenotype.

(Miller & Albertini, 1983). However, even with the overproducing I^Q allele operating, none of the amber mutations in the *lacI* gene (from codons 2 to 329) exhibit any measurable repressor activity in the absence of a suppressor under any experimental conditions we have employed.

(g) *Implications*

The work of Perutz and co-workers on the hemoglobins and myoglobins (Perutz *et al.*, 1965; Perutz & Lehmann, 1968) established that in globular proteins the non-polar residues in the interior of the protein are not replaceable with polar amnino acids, but in many cases are replaceable with certain other non-polar amino acids. On the other hand, residues on the surface of the protein are usually freely exchangeable between non-polar and polar amino acids, unless they are part of the substrate or ligand binding site, or part of intersubunit contacts. Among similar proteins in different species, residues at interior positions are more highly conserved than residues on the surface (Pertuz *et al.*, 1965). Work with extensive amino acid replacements in other proteins has reinforced these conclusions. Sauer and co-workers (Bowie *et al.*, 1990) have found, using combinatorial cassettee mutagenesis, that many residues in the phage lambda repressor amino-terminal domain can be freely substituted, but buried residues are more refractory to amino acid

Table 5

Effects of amino acid replacements at selected sites in the repressor

	Amino acid appearing at position													
Wild-type	Leu	Phe	Pro	Cys	Tyr	Ala	Gly	Ser	Thr	His	Gln	Arg	Lys	Glu
Residue														
Ala57	−	−	−		−	(+)	±			−	−	−	−	−
Gly65	−	−	−	−	−	−	(+)	−		−	−	−	−	−
Gly166	−	−	−	±, ts	−	+	(+)	+		−	−	−	−	−
Gly218	±, ts	−	−	+	−	+	(+)	+		−	−	−	−	−
Gly225	−	−	−	−	−	+	(+)	±, ts		−	−	−	−	−
Ala241	−, ts	−, ts	−	+	−	(+)	+	+		−	−	−	−	−
Ala250	−	−	−	+	−	(+)	+	+		−	−	−	−	−
Gly252	−	−	−	−	−	−	(+)	+		−	−	−	−	−
Gly272	−	−	−	−	−	+	(+)	−		−	−	−	−	−
Ser279	−	−	−	+, ws	−	+, ws	+, s	(+)		−	−	−	−	−
Thr287	−	−	−	+	−	+	−	+	(+)	−	−	−	−	−
Thr288	−	−	−	+	−	+	+	+	(+)	−	−	−	−	−
Gly297	−, ws	−	−	−	−, ws	±	(+)	±		−	−	−	−	−
Thr328	−	−	−	−	−	±	+	+	(+)	−	−	−	−	−

For details, see legend to Table 4.

exchanges. Only positions on the surface could tolerate replacements by polar residues. A recent study of the systematic replacement of amino acids at 163 of the 164 positions in bacteriophage T4 lysozyme (Rennell *et al.*, 1991), using our amber suppressor system, has shown a very strong correlation between degree of substitutional tolerance and the degree of solvent accessibility. Residues which are exposed to solvent (and on the surface) are freely substitutable, whereas those which are not accessible (and buried) are intolerant to substitution. In light of these and our previous studies (Miller *et al.*, 1979; Kleina & Miller, 1990), the finding that a large number of sites in the repressor can be freely substituted is not surprising. In the repressor work reported here, 28% of the 328 sites tolerated all substitutions, 44% all substitutions except proline, and 59% most substitutions. The T4 lysozyme study revealed that 55% of the sites tolerated all of the substitutions. The somewhat higher percentage of sites tolerating every substitution can be partly attributed to the different sensitivities of the assays used (Rennell *et al.*, 1991; see Materials and Methods). The form of the results in the two data sets is strikingly similar. For example, in both the repressor and lysozyme study, proline is frequently not tolerated at otherwise tolerant sites. Extensive substitutions have also been used to identify important functional regions in other proteins. For examples, random mutagenesis and screening was used to detect missense mutations in each of the 99 coding positions in the HIV-1 protease, which revealed three important functional regions of the protein (Loeb *et al.*, 1989).

One striking feature of Figure 1 is how it reveals the regions which are sentitive to substitution. Two sizeable regions, from residues 1 to 59 and from residues 241 to 289, are very sensitive to substitutions, creating the I^- phenotype. The amino-terminal 59 amino acids form a separate domain which includes the DNA and operator binding sites (Mueller-Hill, 1975; Platt *et al.*, 1973; Ogata & Gilbert, 1978), which explains its intolerance to many amino acid substitutions. The second, 241 to 289 region, has an unknown function. It is precisely this segment which shown the strongest conservation among proteins related to the repressor (Figures 2–3). Note that this region also includes several clusters of I^S sites, presumably defining part of the inducer binding site.

A second striking feature of Figure 1 is the presence of segments which are almost completely tolerant of single amino acid substitutions. This includes residues 100 to 112, 129 to 145, 151 to 160, 206 to 217 and 305 to 318. These "open" regions are interdispersed between single residues or clusters of residues that are sensitive to substitution. The finding that the residues in the open regions cannot be deleted (Figure 5) but can be replaced by poly-alanine stretches (Figure 6) indicates that stretches of the sequence serve as spacers for the hydrophobic residues in the core. In fact, in the vast majority of cases the sensitive residues separated by the spacers are hydrophobic amino acids, and the suggestion is that they form part of the hydrophobic core of the protein. Although the three-dimensional structure of the *lac* repressor is not yet known, X-ray crystallographic data on the structure is forthcoming (Pace *et al.*, 1990). The data in Figures 1, 5 and 6 predict that the open regions which are replaceable by poly-alanine stretches may represent surface residues, since they can accept replacements of up to eight amino acids at a time. It should be noted that a method called "clustered charged-charged-to-alanine scanning mutagenesis" (Bennett *et al.*, 1991; Bass *et al.*, 1991, see also Wertman *et al.*, 1992) surveys the surface of the protein by substituting alanines for all the charged residues within a five amino acid stretch. Our replacements complement the recent study of Matthews and co-workers (Heinz *et al.*, 1992), who replaced ten consecutive alanine residues in bacteriophage T4 lysozyme, in a segment known to form an alpha-helix on the exterior of the protein. The lysozyme retained normal function and stability. Our work differs in making the replacements using functional, rather than structural, criteria, and may therefore demonstrate other types of structure tolerant to substitutions.

It is interesting to examine the nature of substitutionally intolerant residues in the repressor. The amino-terminal 59 amino acids are particularly sensitive to amino acid replacements. However, if we look at the remainder of the repressor (actually from residues 60 to 329), it is evident that virtually all sensitive sites are either hydrophobic residues that can only be replaced by certain other hydrophobic residues, or else small amino acids that can only be substituted by other small amino acids. Table 6 shows that only 6 of 73 hydrophilic residues in this region are sensitive to substitution, whereas 42 of 97 hydrophobic residues and 28 of 93 small residues are sensitive to substitution.

There is a strong, although not complete, correlation (see Materials and Methods) between the evolutionarily conserved regions of the repressor and the residues that are sensitive to substitution, as can be seen from Figures 2–3. In others words, the substitution map of the protein is strongly correlated with the "evolution map" of the *lac* repressor family. Considering that all but one of the proteins being compared are not *lac* repressors but other regulatory proteins, this correlation is remarkable. One possible implication of this correlation is that the sequence divergence of these proteins resulted from near neutral mutations which occurred relatively independently, with multiple compensating mutations being rare. In other words, the pattern of allowed amino acids for a particular site in a protein may remain constant even when up to 80% of the amino acids in the protein have changed during evolutionary divergence. A similar study of suppressor-generated changes in phage T4 lysozyme showed that whereas 74 of 163 residues tested (45%) are sensitive to at least one substitution, all 14 residues that are fully conserved among five phage-encoded lysozymes are sensitive to

substitutions (Poteete *et al.*, 1992). It should be noted that although the variability of a site generally correlates with the evolutionary variability, this is not always the case. For instance, Tyr17 and Gln18 in the amino-terminal region of the repressor are intolerant to substitution yet vary considerably in the aligned proteins. These sites have both been implicated in controlling the specificity of DNA binding (Lehming *et al.*, 1987, 1988). Since operator sequences vary among DNA binding proteins, one expects similar variation of amino acids conferring specificity. With respect to the above, it is interesting to note that the I^S sites (residues presumably involved in inducer binding) are not well conserved. One might rationalize this result by arguing that the specific ligands being bound by each regulatory protein are different, so that parts of the binding site involved in specificity should be variable in the *lac* repressor family, but intolerant to substitutions in any given protein from the set, for the ligand binding property.

There are also a number of residues, such as Gly91, which are highly conserved in the aligned set but can be freely substituted in the repressor. In some cases these residues represent sites which simply do not conform to the general pattern, and in other cases may be due to specious homologies or unrecognized subtle functions of the repressor protein.

We thank Drs Ponzy Lu, Mitch Lewis, David Eisenberg, and Brian Matthews for valuable discussions. We are indebted to Dr Robert Kaptein for supplying the coordinates of the *lac* repressor headpiece shown in Figure 2. This work was supported by grants from the National Institute of Health (USHHS RO1 GM 43827-02) and from the Department of Defense (DAALO3-92-G-0173).

References

Alberti, S., Oehler, S., Wilcken-Bergmann, B. v., Kraemer, H. & Mueller-Hill, B. (1991). Dimer-to-tetramer assembly of *lac* repressor involves a leucine heptad repeat. *New Biol.* **3**, 57–62.

Altshul, S. F. & Lipman, D. J. (1990). Protein database searches for multiple alignments. *Proc. Nat. Acad. Sci., U.S.A.* **87**, 5509–5513.

Aslanidis, C. & Schmitt, R. (1990). Regulatory elements of the raffinose operon: nucleotide sequences of operator and repressor genes. *J. Bacteriol.* **172**, 2178–2180.

Bass, S. H., Mulkerrin, M. G. & Wells, J. A. (1991). A systematic mutational analysis of hormone-binding determinants in the human growth hormone receptor. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 4498–4502.

Bennett, W. F., Paoni, N. F., Keyt, B. A., Botstein, D., Jones, A. J., Presta, L., Wurm, F. M. & Zoller, M. J. (1991). High resolution analysis of functional determinants on human tissue-type plasminogen activator. *J. Biol. Chem.* **266**, 5191–5201.

Blatch, G. L. & Woods, D. R. (1991). Nucleotide sequence and analysis of the *Vibrio alginolyticus scr* repressor-encoding gene (*scrR*). *Gene*, **101**, 45–50.

Bossi, L. (1983). Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *J. Mol. Biol.* **164**, 73–87.

Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990). Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science*, **247**, 1306–1310.

Bradley, D., Park, J. V. & Soll, L. (1981). $tRNA_2^{Gln}Su^{+2}$ mutantrs that increase amber suppression. *J. Bacteriol.* **145**, 704–712.

Buvinger, W. E. & Riley, M. (1985). Regulatory region of the divergent *Klebsiella pneumoniae lac* operon. *J. Bacteriol.* **163**, 858–862.

Calos, M. P. & Miller, J. H. (1981). Genetic and sequence analysis of frameshift mutations induced by ICR-191. *J. Mol. Biol.* **153**, 39–66.

Chakerian, A. E., Tesmer, V. M. Manly, S. P., Brackett, J. K., Lynch, M. J., Hoh, J. T. & Matthews, K. S. (1991). Evidence for Leucine zipper motif in lactose repressor protein. *J. Biol. Chem.* **266**, 1371–1374.

Chamness, G. C. & Willson, C. D. (1970). An unusual *lac* repressor mutant. *J. Mol. Biol.* **53**, 561–565.

Farabaugh, P. J. (1978). Sequence of the *LacI* gene. *Nature (London)*, **274**, 765–769.

Feng, D. F. & Doolittle, R. F. (1990). Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol.* **183**, 375–387.

Groarke, J. M., Mahoney, W. C., Hope, J. N., Furlong, C. E., Robb, F. T., Zalkin, H. & Hermodson, M. A. (1983). The amino acid sequence of D-ribose-binding protein from *Escherichia coli* K12. *J. Biol. Chem.* **258**, 12952–12956.

Hall, B. G. & Xu, L. (1992). Nucleotide sequence, function, activation, and evolution of the cryptic *asc* operon of *Escherichia coli* K12. *Mol. Biol. Evol.* **9**, 688–706.

Heinz, D. W., Baase, W. A. & Matthews, B. W. (1992). Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence. *Proc. Nat. Acad. Sci., U.S.A.* **89**, 3751–3755.

Higgins, D. G. & Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, **73**, 237–244.

Jahreis, K., Postma, P. W. & Lengeler, J. W. (1991). Nucleotide sequence of the ilv H-fruR gene of *Escherichia coli* K-1 and *Salmonella typhimurium* LT2. *Mol. Gen. Genet.* **226**(1-2), 332–336.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & Van Gunsteren, W. F. (1985). A protein structure from nuclear magnetic resonance data. *J. Mol. Biol.* **182**, 179–182.

Kleina, L. G. & Miller, J. H. (1990). Genetic studies of the *lac* repressor. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. *J. Mol. Biol.* **212**, 295–318.

Kleina, L. G., Masson, J.-M., Normanly, J., Miller, J. H. & Abelson, J. (1990). Construction of *Escherichia coli* amber suppressor tRNA Genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. *J. Mol. Biol.* **213**, 705–717.

Lehming, N., Sartorius, J., Oehler, S., Wilcken-Bergman, B. v. & Mueller-Hill, B. (1987). The interaction of the recogition helix of *lac* repressor with *lac* operator. *EMBO J.* **6**, 3145–3153.

Lehming, N., Sartorius, J., Oehler, S., Wilcken-Bergman, B. v. & Mueller-Hill, B. (1988). Recognition helices of *lac* and lambda repressor are oriented in opposite

directions and recognize similar DNA sequences. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 7947–7951.

Lipman, D. J. & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science*, **227**, 1435–1441.

Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E. & Hutchinson, C. A., III (1989). Complete mutagenesis of the HIV-1 protease. *Nature (London)*, **340**, 397–400.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. (The Netherlands)*, **405**, 442–451.

McClain, W. H. & Foss, K. (1988). Changing the acceptor identity of a transfer RNA by altering nucleotides in a variable pocket. *Science*, **241**, 1804–1807.

Mckay, D. B. & Steitz, T. A. (1981). Structure of catabolite gene activator protein at 2·9 Å resolution suggests binding to left-handed *B*-DNA. *Nature (London)*, **290**, 744–749.

Miller, J. H. (1992). *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Miller, J. H. (1978). A review. *The Operon* (Miller, J. H. & Reznikoff, W. S., eds), pp. 31–88, Cold spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Miller, J. H. & Albertini, A. M. (1983). Effects of surrounding sequence on the suppression of nonsense codons. *J. Mol. Biol.* **164**, 59–71.

Miller, J. H. & Schmeissner, U. (1979). Analysis of missense mutations in the *lacI* gene. *J. Mol. Biol.* **131**, 223–248.

Miller, J. H., Coulondre, C., Hofer, M., Schmeissner, U., Sommer, H., Schmitz, A. & Lu, P. (1979). Generation of altered proteins by the suppression of nonsense mutations. *J. Mol. Biol.* **131**, 191–222.

Mueller-Hill, B. (1975). A review. *lac* repressor and *lac* operator. *Prog. Biophys. Mol. Biol.* **30**, 227–252.

Mueller-Hill, B., Crapo, L. & Gilbert, W. (1968). Mutants that make more *lac* repressor. *Proc. Nat. Acad. Sci., U.S.A.* **59**, 1259–1264.

Myers, G. L. & Sadler, J. R. (1971). Mutational inversion of control of the lactose operon of *E. coli*. *J. Mol. Biol.* **58**, 1–28.

Normanly, J., Kleina, L. G., Masson, J.-M., Abelson, J. & Miller, J. H. (1990). Construction of *Escherichia coli* amber suppressor tRNA *genes*. III. Determination of tRNA specificity. *J. Mol. Biol.* **213**, 719–726.

Normanly, J., Masson, J.-M., Kleina, L. G., Abelson, J. & Miller, J. H. (1986). Construction of two *Escherichia coli* amber suppressor genes: tRNA Phe/Cua and tRNA Cys/Cua. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 6548–6552.

Ogata, R. & Gilbert, W. (1978). An amino-terminal fragment of *lac* repressor binds specifically to *lac* operator. *Proc. Nat. Acad. Sci., U.S.A.* **75**, 5851–5854.

Pace, H. C., Lu, P. & Lewis, M. (1990). *lac* repressor: crystallization of intact tetramer and its complexes with inducer and operator DNA. *Proc. Nat. Acad. Sci., U.S.A.* **87**, 1870–1873.

Perutz, M. F. (1965). Structure and function of haemoglobin. I. tentative atomic model of horse oxyhaemoglovin. *J. Mol. Biol.* **13**, 646–668.

Perutz, M. F. & Lehmann, H. (1968). Molecular pathology of human haemoglobin. *Nature (London)*, **219**, 902–909.

Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965). Structure and function of haemoglobin. II. Some relations between polypeptide chain configuration and amino acid sequence. *J. Mol. Biol.* **13**, 669–678.

Platt, T., Files, J. G. & Weber, K. (1973). Specific proteolytic destruction of the NH_2-terminal region and loss of the deoxyribonucleic acid-binding activity. *J. Biol. Chem.* **248**, 110–121.

Platt, T., Miller, J. H. & Weber, K. (1970). *In vivo* degradation of mutant *lac* repressor. *Nature (London)*, **228**, 1154–1156.

Poteete, A. R., Rennell, D. & Bouvier, S. E. (1992). Functional significance of conserved amino acid residues. *Proteins*, **13**, 38–40.

Reidl, J., Roemisch, K., Ehrmann, M. & Boos, W. (1989). Mal I, a novel protein involved in regulation of the maltose system for *Escherichia coli*, is highly homologous to thge repressor proteins GalR, CytR and *Lac*I. *J. Bacteriol.* **171**, 4888–4899.

Rennell, D., Bouvier, S. E., Hardy, L. W. & Poteete, A. R. (1991). Systematic mutation of bacteriophage T4 lysozyme. *J. Mol. Biol.* **222**, 67–87.

Rolfes, R. J. & Zalkin, H. (1988). *Escherichia coli* gene purR encoding a repressor protein for purine nucleotide synthesis: cloning, nucleotide sequence, and interaction with the purF operator. *J. Biol. Chem.* **263**, 19653–19661.

Schmitz, A., Schmeissner, U., Miller, J. H., & Lu, P. (1976). Mutations affecting the quaternary structure of the *lac* repressor. *J. Biol. Chem.* **251**, 3359–3366.

Schulz, G. E. & Schirmer, R. H. (1979). *Principles of Protein Structure*, pp. 124–125, Springer-Verlag, New York.

Stokes, H. W. & Hall, B. G. (1985). Sequence of the ebgR gene of *Escherichia coli*: evidence that the EBG and LAC operons are descended from a common ancestor. *Mol. Biol. Evol.* **2**, 478–483.

Valentin-Hansen, P., Larsen, J. E. L., Hojrup, P., Short, S. A. & Barbier, C. S. (1986). Nucleotide sequence of the CytR regulatory gene of *E. coli* K-12. *Nucl. Acids Res.* **14**, 2215–2228.

Weickert, M. J. & Adhya S. (1992). Isorepressor of the gal regulon in *Escherichia coli*. *J. Mol. Biol.* **226**, 69–83.

Wertman, K. F., Drubin, D. G. & Botstein, D. (1992). Systematic mutational analysis of the yeast *ACT1* gene. *Genetics*, **132**, 337–350.

Wilcken-Bergman, B. v. & Mueller-Hill, B. (1982). Sequence of *galR* gene indicates a common evolutionary origin of *lac* and *gal* repressor in *Escherichia coli*. *Proc. Nat. Acad. Sci., U.S.A.* **79**, 2427–2431.

Wu, J., Anderson-Loviny, T., Smith, C. A. & Hartley, B. S. (1985). Structure of wild-type and mutant repressors and of the control region of the *rbt* operon of *Klebsiella aerogenes*. *EMBO J.* **4**, 1339–1344.

Edited by B. W. Matthews

(Received 30 April 1993; accepted 18 April 1994)