DNA-Binding Determinants of Sigma 54 as Deduced from Libraries of Mutations

YULI GUO AND JAY D. GRALLA*

Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California,
Los Angeles, Los Angeles, California 90095-1569

Received 12 September 1996/Accepted 3 December 1996

PCR mutagenesis was used to obtain libraries of mutations in the region between amino acids 300 and 400 in the DNA-binding domain of *Escherichia coli* sigma 54. Two hundred changes that did not alter function were identified. These were compared with a somewhat smaller number of changes that did alter function. Several important regions were identified. Single point mutations in two of these, near amino acids 363 and 383, destroyed the ability of sigma to bind DNA, as assayed by band shift analysis. A third segment from amino acids 327 to 347 is also a candidate for contributing to DNA binding. Comparison with data in the literature leads to testable proposals for the complex mode of DNA binding that is associated with sigma 54.

Recognition of *Escherichia coli* promoters by RNA polymerase is directed by the sigma subunit. Two types of sigma factors are known, the sigma 70 family of proteins (8) and sigma 54 (12, 14, 16, 22). In both cases, the promoters typically contain two primary recognition sequences separated by a spacer of defined length. These two sequences may be recognized by separate determinants on the sigma factors. Both genetic and biochemical evidence supports this view for members of the sigma 70 family of proteins (8), and genetic studies support this view for sigma 54 (19, 24).

The locations of the recognition determinants on sigma 54 remain uncertain. A series of deletions were constructed across the sigma 54 gene (19, 24). Deletions in the central region abolished the ability of sigma 54 to bind RNA polymerase, but sigma 54 could still bind DNA. Deletions in the N-terminal region abolished the ability to activate transcription, but DNA binding remained intact. Mutations in the C-terminal 40% of the protein could abolish DNA binding. A fragment from within this same region of sigma 54, containing amino acids 329 to the C terminus (at amino acid 477), was reported to bind DNA specifically, as assayed by footprinting (4). Protein-DNA cross-linking studies showed that the main site of DNA cross-linking to sigma 54 was within this same region, with the most likely site including residues between amino acids 332 and 345 (3). Sigma 54 with an internal deletion mutation that removed amino acids 293 to 332 could still bind DNA (24). Thus, the main DNA-binding determinant is clearly within the C-terminal one-third, perhaps just beyond amino acid 332.

It has also been proposed that the main DNA recognition determinant lies within this broad region but further beyond the site of cross-linking (5, 17). Site-directed mutations were constructed within a presumptive helix-turn-helix (HTH) motif. These studies led to the suggestion that a recognition helix near amino acid 380 was critically important for DNA binding. Thus, the primary determinants for DNA binding have been proposed to be somewhere between approximately amino acids 330 and 400.

Two other issues further complicate locating the DNA-bind-

ing determinants of sigma 54. First, mutations in various regions of sigma 54 can reduce DNA-binding affinity, although they do not abolish DNA binding, in contrast to the properties of the C-terminal region mutations (9, 19, 24). These mutations typically lead to a loss of protection over the -12 promoter element. Second, it is not known how many DNA-binding determinants exist. Because there are two known recognition elements at sigma 54 promoters, centered near positions -12 and -24, and because both are covered by sigma 54 (2, 14), it is presumed that there should be two DNA-binding determinants, but this has not been demonstrated directly. Moreover, the sequences in the promoter spacer that separates these two determinants also affect the affinity of sigma 54 for DNA (1), allowing for the possibility that there are additional DNA recognition determinants on sigma 54.

Thus far, genetic studies of DNA-binding determinants of sigma 54 have been restricted to the study of site-directed mutants targeted to regions of interest (5, 6, 9, 10, 17, 19, 24). These are highly selective in regard to which part of the region between amino acids 300 and 400 is being investigated. As discussed above, only large deletions and point mutations near amino acid 380 have been studied. To obtain a broad view of the role of this region in DNA binding, we created random libraries of mutations within it. These were screened for function and DNA binding. The results indicate that there are several determinants for DNA binding within this region and set the stage for future studies determining the role of each.

MATERIALS AND METHODS

Strains and plasmids. Plasmid pTH7 carries the *E. coli* sigma 54 gene, which is under the control of Ptac, an IPTG (isopropyl-β-D-thiogalactopyranoside)-inducible promoter (11). The leaky expression of sigma 54 in the absence of IPTG is sufficient to support growth on minimal medium plates. Two unique restriction sites, *Cl*I and *Av*rII, were introduced into pTH7 at amino acids 305 and 394, respectively, by use of the Amersham Sculptor in vitro mutagenesis kit. *E. coli* YMC109^{tk} (*thi* *endA* *hsr* Δ *lac*U169 *rpoN*::Tn10/F' *pro* *lac*I^qZU118 Tn5-102), lacking a wild-type chromosomal copy of the sigma 54 gene, was used as a host (10). The primers used for amplification were upstream from the *Cl*I site (primer 4; sequence CTGTTAGCCAATCAGAC) and downstream from the *Av*rII site (primer S5b; sequence GTACTTCGCAACAGTGC).

PCR mutagenesis. Error-prone PCR (for the Mn²⁺ library) (20) was as follows: 50-μl reaction mixtures contained 1 ng of pTH7 template (cut with *Eco*RI), 1 μg of each primer, 1× *Taq* reaction buffer (Promega), 5 U of *Taq* polymerase, 6 mM MgCl₂, 0.64 mM MnCl₂, and 0.2 mM each dATP, dGTP, dCTP, and dTTP. After 30 cycles, reaction mixtures were extracted with phenol and chloroform and precipitated with ethanol. After cleavage with *Cl*I and *Av*rII, the digested fragment was purified by agarose gel electrophoresis. Full-length pTH7

* Corresponding author. Mailing address: Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, P.O. Box 951569, Los Angeles, California 90095. Phone: (310) 825-1620. Fax: (310) 206-7286. E-mail: Gralla@ewald.mbi.ucla.edu.

was also digested with the same two enzymes, and the large piece without the *Cla*I-*Avr*II fragment was purified. The ligation mixture of pTH7 large fragment and mutagenized insert was transformed into YMC109^{tk} and then plated on Luria broth plates containing ampicillin, kanamycin, and tetracycline.

The low-frequency-of-mutation PCR (Mg^{2+} library) used the same procedure with the following modifications: 100- μ l reaction mixtures contained 1 ng of template, 0.6 μ g of each primer, and 2 mM $MgCl_2$ as the only source of divalent cation.

Plating tests. W-arg plates (10) (500 ml of medium contains 5.25 g of K_2HPO_4 , 2.25 g of KH_2PO_4 , 7.5 g of agar, 10 ml of 20% glucose, 1.0 ml of thiamine [10.0 mg/ml], 0.215 ml of 1 M $MgSO_4$, 2 ml of arginine [25 mg/ml], and appropriate antibiotics) were used to screen transformants into two populations, functional mutants and nonfunctional mutants. Because host strain YMC109^{tk} lacks the sigma 54 gene, only those transformants with a functional sigma 54 gene can grow well on W-arg plates; arginine is the only source of nitrogen.

The fresh transformants from PCR mutagenesis were plated on Luria broth plates. These were transferred individually to W-arg plates containing appropriate antibiotics. The plates were incubated at 37°C for 36 to 40 h. Colonies that grew poorly compared with wild-type sigma 54 were identified, and these were presumed to be nonfunctional. For the lower-frequency-of-mutagenesis library, almost all of the initial candidates turned out to contain stop codons or frame-shifts, as identified by sequence analysis of the insert. We found that these colonies also failed to grow on W-gln plates, which have a higher tolerance for lack of sigma 54 (in this plate, the same amount of glutamine replaces arginine). To obtain some colonies from the low-frequency library that might yield expressed protein, this library was also screened on W-gln plates. Those colonies that grew poorly on W-arg plates but were able to form colonies on W-gln plates were analyzed further. As discussed below, a number of these colonies contained point substitutions in the insert of interest and were expressed as full-length sigma 54. All phenotypes were checked by replating.

As discussed in Results, all colonies from the low-frequency library were confirmed for sigma 54 protein expression as described previously (23). This involves adding radioactive amino acid precursors to cells just after IPTG induction and then identifying the radioactive sigma band by autoradiography of the protein gel. Selected colonies from the high-frequency library were assayed similarly, and all were found to produce detectable sigma 54 protein. To assay for DNA binding by sigma, cell extracts from the colonies of interest were prepared and subjected to a gel mobility shift assay as described previously (21, 24).

Statistical analysis. Two kinds of statistical analyses were used to analyze the data obtained from the high-frequency library. In one, described previously (20), the frequency of mutagenesis at each position is compared with the expected number of changes by standard deviation analysis. The changes associated with functional and nonfunctional libraries are analyzed separately. For each residue, the probability that the number of mutations observed deviates from the number expected based on random hits can be calculated. A large delta (in standard deviation units) represents a strong deviation from the expected distribution of mutations at that residue.

To assess the importance of small regions, rather than individual residues, a second-order moving-average analysis (18) was done. The entire sequence of the insert was considered to be associated with five amino acid blocks. For each amino acid residue, N , the total number of functional mutations within a block of 5 amino acid residues centered at N was calculated (first-order moving total for N). For example, there were 12 functional changes from residue 348 to 352; 12 is the first-order moving total for residue 350. As the 12 mutations accumulated in a 5-residue block, the first-order moving average around amino acid 350 was 12/5 or 2.4. The data were smoothed by obtaining a second-order moving average, in which the procedure was repeated with data from the first-order moving average as input. For example, for the second-order moving average about amino acid 350, the input data would include the value of 2.4 associated with amino acid 350 and the values determined as described above for each of the residues 348, 349, 351, and 352.

RESULTS

Creation of libraries and analysis of functional mutants.

The method for creating a library of mutations in the primary DNA-binding region is outlined in Fig. 1. The procedure is based on that developed previously to study a much smaller region of sigma 54 that is involved in binding to RNA polymerase (20). It involves amplifying the targeted region by error-prone PCR. Amplification relied on two unique restriction enzyme sites that were introduced by site-directed mutagenesis. The *Cla*I site is N-terminal to the cross-linking region, and the *Avr*II site is C-terminal to the presumptive HTH motif (see the introduction). The collection of amplified PCR products was cleaved with the two enzymes and cloned back into a vector from which the same region had been removed. This

FIG. 1. Outline of the PCR mutagenesis procedure for the region between the introduced restriction sites at amino acids 307 and 392.

creates a library of sigma 54 genes containing the amplified inserts.

The PCR amplification was done under error-prone conditions, based on procedures used previously on a small central segment of the sigma 54 gene (20). Two different conditions were used. In one, manganese was included in the PCR, leading to a higher error frequency (13). In the other, manganese was not used, leading to a lower error frequency. As described below, the low-frequency library resulted in fewer than 1% of the colonies containing clones of interest; almost all of the colonies were either functional or contained trivial down-mutants of the frameshift and stop codon types. Thus, the information content was sparse, although the few mutants obtained were informative. By contrast, the high-frequency library yielded a great deal of information and will be described first.

The library created by manganese PCR was transformed into strain YMC109^{tk}, which lacks a chromosomal copy of the sigma 54 gene. Individual transformants containing plasmids were isolated on nonselective rich medium plates containing appropriate antibiotic markers to obtain colonies for testing. These colonies were transferred to W-arg plates. W-arg plates do not support significant growth of YMC109 unless the bacteria are transformed with plasmids containing functional sigma 54. Thus, colonies were identified as being one of two types. Colonies that grow should contain functional sigma 54 proteins. Those that form very small colonies (or do not grow) should not contain functional sigma 54. Both types of colonies were analyzed further (Fig. 2).

The purpose of analyzing both types of colonies is to compare mutations that lead to loss of sigma 54 function with mutations that can be tolerated. Approximately one-third of the colonies grew on W-arg plates and thus were judged to contain functional sigma 54 genes. The inserts of approximately 100 of these were sequenced (discussed below) and

FIG. 2. Outline of procedure to separate the higher-frequency-of-mutagenesis library into two populations.

found to contain 2.1 mutations per insert on average. This provides a collection of more than 200 mutational changes that do not destroy sigma 54 function.

We tested approximately one-quarter of these mutant forms of sigma 54 (mutant sigmas) for DNA binding to confirm their categorization by the W-arg plate test. This test involves making crude cell extracts from each colony and doing a band shift assay with labeled NifH promoter probe. This assay has been used previously and shown to measure specific promoter recognition (21). The position of a sigma 54 holoenzyme-DNA complex, which is the most sensitive indicator of the ability of sigma 54 to bind DNA, is shown in Fig. 3a (labeled holo). The band just below that band represents a nonspecific interaction of the probe with core polymerase. The broad band at the bottom of Fig. 3a represents the binding of any excess sigma 54 to DNA. Band shift analysis for five clones that grew on W-arg plates and thus were judged to be functional is shown in Fig. 3a. A holoenzyme-DNA band is present in all lanes (lanes 3 to 7), confirming that the mutant sigmas can still bind DNA (compare with the positive wild-type control in lane 2; see also Fig. 3b for patterns of nonfunctional mutants found not to bind DNA, as discussed below). In three cases, this band is very strong (lanes 3, 6, and 7), and in two cases, it is weaker (lanes 4 and 5). The weaker band may indicate some lowering of affinity for DNA but not enough to prevent sigma 54 from functioning. Alternatively, the weaker band pattern may represent a somewhat lower stability of the mutant protein; when expression levels (see below) of the wild type are deliberately lowered (lane 1), this same lower-affinity pattern appears. In any case, for all colonies tested, growth on W-arg plates was accompanied by the ability to bind DNA.

DNA sequence analysis of these clones containing functional forms of sigma 54 led to the identification of more than 200 mutations in the targeted region of sigma 54. Figure 4 shows the sequence of the region from amino acids 307 to 392 (residues are indicated in column II by a bold single-letter code). All mutational changes associated with functional forms of sigma 54 are shown to the left of the residue (a number indicates the number of repeated occurrences). These represent approximately 200 changes that apparently can be tolerated in that sigma 54 can still function and still bind DNA. Although the screen would require many thousands more mutations to approach saturation, it is apparent from this collection that at least 80% of the amino acids can be mutated

FIG. 3. Band shift assay of crude cell extract for the presence of DNA-binding sigma 54. DNA binding is indicated by the band labeled holo. (a) Lanes: 1, wild-type sigma 54 with expression not further induced with IPTG; 2, wild-type sigma 54 with expression induced to the highest level; 3 to 7, different functional mutants. (b) Lanes: 1, wild-type sigma 54 with expression not further induced; 2, wild-type sigma 54 with expression induced to the highest level; 3 to 10, various nonfunctional mutants.

without loss of function (69 of 86). The 17 amino acids that are not mutated in these functional forms of sigma 54 are not contiguous; therefore, it is not obvious that any one region is the most intolerant to mutation.

In an attempt to identify contiguous stretches of amino acids that tolerate mutations poorly, an analysis of the data was done. The approximately 200 changes are derived from proteins that typically have two linked changes; even the two together do not cause a loss of function. For each amino acid, the frequency with which these functional changes occur in the 5-amino-acid region surrounding it was calculated. The idea is that certain regions may be selectively associated with fewer of these harmless mutations because changes there would destroy function. This was repeated twice to obtain a second-order moving average (see Materials and Methods), which represents the tendency of each amino acid to be within a region that tolerates mutation in that it still retains function. This is illustrated in Fig. 5, with the residues that are associated with a lower-than-average number of mutations represented by filled symbols. These presumably are less tolerant of functional mutation since they have mutational frequencies that are below the median moving average of 2.23 for the insert as a whole. The residues are clearly not located randomly, and their five clusters are shaded. The much smaller double-shaded regions below the broken line include the residues in which the frequency of mutations observed is less than the expected random frequency by more than 1 standard deviation. Four of the five clusters have residues that meet this more-stringent criterion for evidence of inability to sustain mutation and retain function.

The three largest regions, which also have small segments

I	II	III	I	II	III
307	F R G S3 C	-1.78	350	Q L	1.35
	S D4K Y2 N	-2.09		L Q Q	-0.74
	S T Y D2 N	-1.66	H L2R3 Q L H2R		0.30
310	V P T2C E A	-1.87		T S A T	-0.07
	S H2C2 R C H L	0.12		L3 F S	-0.44
	I N I Y	1.13	355	L2Y F Y L S2	1.20
	Y H G E N2 D G V	-0.62		A E K V2	1.66
	R G	-0.74	L3R2K Q L R		-0.62
315	A G2N V D	-1.66		G D	1.35
	C G2 S R	-0.44		G2 E D	-0.07
	Q L R	1.91	360	D E A	0.43
	F S L	1.91		Y N K C	2.34
	V I	-0.74	T3L I M K		-1.02
320	L2 R	-1.04		K I E Q	2.34
	I C G S	-1.28		Q S P A	-0.07
	T S N D	-0.07	365	K L M V2	0.61
	V R L M P	0.61		A V	-0.74
A V G L R2E2	Q L2	-1.02		R L Q	0.43
325	N2 D N	-0.07		D	
	S A D	0.43	N G2Y2	A	-1.66
	N2I E2 K E2I	0.12		V I N	0.43
	W R	1.35		T D A	-1.04
	L S2	1.91		R L2 Q R L2P	1.20
330	T2V2 I L	-0.72		A T	1.35
	Q E K R2	0.61	G I A V		-1.28
	C R2 S R	-0.44	375	G K V E G D	0.22
	L P3	2.34	T K2V2I M I T2V		0.30
	D G V2 E G D	-0.16		R Q	-1.04
335	C2I2 S G2	-0.16		E G	1.35
	H2 R H3	1.15		S P	1.35
	S4 N	-1.48	380	M S3 T P	-0.72
	N4 D	-1.48		V2 I L V	0.61
	M T	-0.74		F S P	0.43
340	L			R	
	Q3 L P	-0.44	G M A V G		-0.44
H C S L R H		-0.72		I T I	0.43
	A V M	0.43	A S2 T A		-0.44
	C3 S G4	1.20	R H2 Q		-1.28
345	H2 R L	-0.07		K E3R N	3.02
	S C W S	1.13		Y	
	I T2F	2.34	390	N M2 L P	-0.44
	I2 V	-1.04		L Y R3 H G	-1.02
G2R E		-1.28		C S C2I	1.66

FIG. 4. Mutations from the Mn^{2+} library and their analysis. Columns I include the *E. coli* sigma 54 residue numbers. Columns II show the wild-type residue in bold single-letter code. At the left of the residues are shown changes associated with functional mutants, followed by the number of times that changes appeared in the whole population. Those without a number are changes that appeared only once. To the right of the residues are shown changes found in nonfunctional mutants along with their frequency of appearance. Columns III show the delta values, which represent statistical measures of how frequently a residue is mutated in functional sigmas (see Materials and Methods). High positive values indicate that the residue is mutated much less frequently than expected in forms of sigma that retain function. For example, a delta value of 1.0 indicates that the mutation frequency is 1 standard deviation unit different from that expected based on random hits.

that meet this more-stringent criterion, are located downstream from amino acid 332, where the main DNA-binding determinants are thought to reside (see the introduction). One is centered over the presumptive recognition helix of the HTH motif (minimum at position 378); one is centered near position 364, just upstream from the beginning of the presumptive HTH motif (5); and one is centered near position 347 at the downstream edge of the cross-linking region (3). Each patch has a maximum of approximately 15 amino acids that are associated with below-average numbers of functional mutations, and each has a few amino acids with mutational frequencies at least 1 standard deviation lower than expected based on random mutational frequencies. Two additional patches are smaller and are located prior to the major DNA-binding region as determined in prior studies. The analysis suggests that there are at least three, and perhaps up to five, patches of amino acids that are intolerant to functional mutation.

Multiple mutations associated with nonfunctional sigma 54. To complement these data, clones from colonies that did not grow on W-arg plates, indicating a loss of sigma 54 function, were studied (Fig. 2). The inserts were first subjected to single-lane DNA sequencing. This preliminary analysis screened out frameshift mutations that appeared in one-third to one-half of the clones that did not grow on W-arg plates. The remaining

clones were analyzed further. The band shift analysis described above confirmed that these mutant forms of sigma 54 were unable to bind DNA (Fig. 3b; compare the patterns with those of the binding mutants shown in Fig. 3a and with those of the nonbinding point mutants shown below). Protein gels (as shown below for nonfunctional single point mutations) confirmed that these mutant sigmas were expressed (data not shown).

The inserts from these nonfunctional clones were fully sequenced. The average number of mutations per insert in this collection was 4.9. This frequency of mutation is higher than the 2.1 associated with functional clones, consistent with the inactivation of sigma 54 causing the inability to grow on W-arg plates. This also suggests that in many cases, it takes multiple mutations to destroy DNA binding. Approximately 100 changes associated with nonfunctional forms of sigma 54 were identified. These are shown to the right of column II in Fig. 4.

Because on average each clone had five of these changes, the existence of any individual change is not a direct indication that it has led to loss of function. To address this issue, a statistical analysis that was used previously in a similar study of a smaller (polymerase binding) region of sigma 54 (20) was applied. This analysis takes each amino acid and compares the frequency with which it is mutated in nonfunctional sigmas with that in functional sigmas. A standard deviation from the mean is calculated for each amino acid. This gives a probability that the distribution of nonfunctional changes deviates from the mean. These delta values are shown in column III of Fig. 4. High positive delta values are by definition associated with amino acids that are changed frequently in nonfunctional sigmas and rarely in functional sigmas. Such high positive values suggest that the residues are important for function (20).

The eight amino acids with the highest values (delta values above 1.9, which indicates a confidence level higher than 94% [7]) are shown associated with vertical bars in Fig. 5. Each amino acid has a probability greater than 94% that it is mutated in nonfunctional sigmas more frequently than expected. For example, the extreme would be lysine 388, which is mutated five times in nonfunctional sigmas but never in functional sigmas. This compares to the typical amino acid which is mutated once in nonfunctional sigmas and twice in functional sigmas. At the lower end, but still highly significant, would be amino acids 317, 318, and 329, each of which is mutated twice in nonfunctional sigmas and never in functional sigmas.

Seven of these eight amino acids are found within the five patches identified as being intolerant to mutation in the moving-average analysis of Fig. 5. The exception is Leu-333, which is just outside one of these patches. Since these five patches include only one-half of the total number of amino acids within the region mutated, the occurrence of seven of these eight residues within them supports the importance of these patches.

Point mutations that interfere with DNA binding. Next, we turned to construction and analysis of a low-frequency-of-mutagenesis library. In preliminary studies, we found that approximately 10% of colonies grew poorly or not at all on the W-arg plates described above, indicating loss of function. However, preliminary sequence analysis indicated that almost all of these were either frameshift or stop codon mutants, which did not express sigma 54 protein. Because this prevented detection of the point mutations with nonfunctioning but expressed sigma, we sought an alternative plate test to screen out these nonexpressed mutants. We found that replacing arginine with the alternative nitrogen source glutamine in the plates accomplished this purpose. On these W-gln plates, only the rare point down-mutations grew whereas the nonexpressed mutants grew poorly. Thus, we revised the screen to exclude all colonies that

FIG. 5. Statistical analysis of mutations pointing to important residues (see Materials and Methods). The curve at the top of the graph indicates the second-order moving average (shown at left) that was calculated from the distribution of mutations in functional sigma across the region. Regions that exhibit lower-than-average mutation frequencies in functional sigma are shown shaded below the solid line. Residues within this region have closed symbols. The double shaded areas below the broken line include only regions which are mutated so rarely in functional sigmas that they are more than 1 standard deviation lower than expected. The amino acid numbers are shown below the plot. Regions that may cross-link to DNA or form a presumptive HTH motif are indicated. The vertical bars represent individual residues that have a greater-than-94% probability of showing fewer mutations than expected (delta values shown at right). The filled circles below the amino acid numbers show the positions of five single point mutations that eliminate DNA binding. They are F354S, E360G, S379L, R383H, and R383L.

grew poorly on W-gln plates. That is, colonies that showed some growth on W-gln plates and none or very slow growth on W-arg plates were candidates for nonbinding but expressed sigmas. This revised screen produced more colonies with expressed point mutations but of course is associated with the possibility that very strong down-mutations might not be detected.

Twenty colonies were obtained by this screening procedure. After sequence analysis, six of these turned out to have unique single point mutations. The others either contained double mutations or stop codons (all were amber) or were duplicates of the single mutations. Only three of these did not grow at all on W-arg plates, and these are single point mutations at arginine 383 (to either histidine or leucine) and at serine 379 (to leucine). Three changes, glutamate 360 to glycine, phenylalanine 354 to serine, and phenylalanine 318 to valine, were associated with abnormally small colonies on W-arg plates.

These six sigmas were subjected to band shift analysis (as described above) to test for their ability to bind NifH DNA. Figure 6 shows that five have completely lost the ability to bind DNA. In these cases (lanes 4 to 8), there is no band in either the holoenzyme or sigma positions. For comparison, various patterns are shown: no sigma (lane 1), small amounts of sigma (lane 2), large amounts of sigma (lane 3), and sigma with a lower affinity for polymerase (lane 10). The exception is sigma with the F-to-V mutation at position 318 (F318V), which still binds DNA (compare lane 9 with lane 3, the wild-type positive control, and lane 10, the polymerase-binding mutant). This result is consistent with prior studies showing that deletion of the region containing F-318 did not prevent DNA binding. Apparently, the small patch associated with residue 318 has

some other important function. This could conceivably be related to surface structure since protease digestion studies suggest that the region is highly exposed (3).

The five nonbinding forms of sigma, containing single point mutations in four amino acids, are represented by filled circles at the bottom of Fig. 5. All are expressed, as tested by our prior procedure for producing and detecting radiolabeled sigma 54 (Fig. 7; see mutants in lanes 4 to 8 compared with the positive control in lane 3 and the negative controls in lanes 1 and 2). R-383 and S-379 are within the most-C-terminal functional patch of important amino acids identified above from study of the manganese library; this corresponds to the recognition

FIG. 6. Band shift assay for DNA binding of sigma 54 with single point mutations. Lanes: 1, cells without sigma 54; 2, wild-type sigma 54 before full induction; 3, wild-type sigma 54 after full induction; 4, S379L; 5, R383H; 6, F354S; 7, E360G; 8, R383L; 9, F318V; 10, G189R.

FIG. 7. Protein expression of nonfunctional single point mutants. Lanes: 1, cells without sigma 54; 2, cells with sigma 54 but no induction; 3, wild type after full induction; 4, S379L; 5, R383H; 6, F354S; 7, E360G; 8, R383L.

helix of the presumptive HTH motif. E-360 is within one of the other two large patches, and F-354 is nearby. These are also adjacent to 2 of the 8 residues (amino acids 361 and 363) that are rarely mutated in functional forms of sigma. Two patches, separated by approximately 20 amino acids, can eliminate DNA binding.

DISCUSSION

PCR methods were used to create libraries of mutations in an approximately 90-amino-acid region of sigma 54 thought to contain its main DNA binding determinants (3–5, 19, 24). Three different populations of mutations were analyzed to determine the most likely locations for these determinants. First, approximately 200 changes that did not interfere with sigma 54 function were isolated. Analysis of these led to the identification of four or five patches of amino acids that were intolerant to mutation in that they were rarely mutated in functional sigmas. Second, approximately 100 changes associated with inactive forms of sigma 54 were identified and analyzed. Seven of the eight amino acids that were identified as being most important by this independent analysis were within the patches identified by the first analysis. Finally, five single point mutations that prevented DNA binding were found. These were mainly clustered in the two most-C-terminal of these patches. The locations of the patches are shown in Fig. 5, and these are compared with two potential markers of DNA binding, a region that cross-links to DNA (3) and a hypothesized HTH motif (5).

The mutational data are most compelling for two regions centered roughly near amino acids 383 and 363. Both regions are rarely mutated in functional sigmas (Fig. 5, shaded regions), and both contain single point mutations that abolish DNA binding (Fig. 5, filled circles) and contain residues that are frequently mutated in nonfunctional sigmas (Fig. 5, vertical bars). The most-C-terminal patch is clearly coincident with the recognition helix of the hypothetical HTH motif and begins near amino acid 378. Residues within this region have previously been subjected to site-directed changes and have led to inactivation of sigma 54, which was suggested to occur by defective promoter recognition (5). Our data have shown directly that point mutations at residues serine 379 and arginine 383 within this region inactivate DNA binding. This confirms the importance of this region and these residues, showing that it is indeed necessary for binding DNA.

The second region required for DNA binding lies approximately 20 amino acids upstream. The amino acids lysine 363, tyrosine 361, glutamate 360, and phenylalanine 354 are implicated by analysis of nonfunctional sigmas (Fig. 5). None of these amino acids is within the presumptive positioning helix of the suggested HTH motif, which begins at amino acid 367 (5). Thus the data indicate that two separated patches of amino

acids are involved in DNA binding but do not advance knowledge of the structure of the DNA recognition determinants.

The data also suggest that there are additional determinants for DNA binding, but the evidence is not as compelling since no single-point mutations were obtained. The regions involved are centered near amino acids 329 and 347 (Fig. 5). These two regions mark the approximate boundaries of the part of sigma 54 that is believed to cross-link most readily to DNA, which lies approximately between amino acids 332 and 345 (3). It has been suggested that this region could potentially be alpha-helical (3). We note that an alpha-helical wheel analysis of amino acids 327 to 350 places the seven residues that are least tolerant to functional mutation (see Fig. 4 for residues with delta values greater than 1, implying a confidence level of 1 standard deviation) on a single face (data not shown). These seven residues lie mostly at the two edges of the region, raising the possibility of DNA interactions with the edges of a long alpha-helix. The segment also contains five positively charged amino acids, which have the potential to bind the DNA backbone and contribute to the affinity of sigma 54 for DNA.

Sigma 54 is known to bind to two distinct promoter sequence elements, located near positions –12 and –24 (14). The regions near amino acids 383 and 363 are candidates for recognition of these elements, although direct biochemical evidence is lacking. Recognition of the –12 region is believed to be complex. Mutations in several separated parts of sigma 54 each interfere with recognition at position –12, suggesting that its recognition may involve a complex structure. The structure has been proposed to involve alpha-helical bundles, including parts of the N-terminal and the polymerase-binding domains (9, 19, 24), and separate experiments have implicated the region near residue 383 (17). One possibility is that –12 region recognition involves a four-helix bundle analogous to that observed for sigma 70 (15) with the region near residue 340 as a candidate for the fourth helix. Recognition of region –24 appears to be simpler and to be dominant in terms of affinity (19, 24). The spacer sequence between the –12 and –24 elements is known to influence the affinity of sigma 54 for DNA (1).

Any of the identified regions are of course candidates to influence any of these interactions. The current identification of these candidates for the various DNA-binding determinants allows one to begin a biochemical analysis of selected forms of sigma 54 to learn how it binds DNA by use of the interplay among these protein domains and DNA elements.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant GM35754.

We thank Yi Song for her excellent technical assistance.

REFERENCES

1. Buck, M., and W. Cannon. 1992. Specific binding of the transcription factor σ^{54} to promoter DNA. *Nature* **358**:422–424.
2. Cannon, W., S. Austin, M. Moore, and M. Buck. 1995. Identification of close contacts between the sigma N (sigma 54) protein and promoter DNA in closed promoter complexes. *Nucleic Acids Res.* **23**:351–356.
3. Cannon, W., F. Claverie-Martin, S. Austin, and M. Buck. 1994. Identification of a DNA-contacting surface in the transcription factor σ^{54} . *Mol. Microbiol.* **11**:227–236.
4. Cannon, W., S. Missailidis, S. Smith, A. Cottier, S. Austin, M. Moore, and M. Buck. 1995. Core RNA polymerase and promoter DNA interactions of purified domains of σ^N : bipartite functions. *J. Mol. Biol.* **248**:781–803.
5. Coppard, J. R., and M. J. Merrick. 1991. Cassette mutagenesis implicates a helix-turn-helix motif in promoter recognition by the novel RNA polymerase sigma factor σ^{54} . *Mol. Microbiol.* **5**:1309–1317.
6. Cullen, P. J., D. Foster-Hartnett, K. K. Gabbert, and R. G. Kranz. 1994. Structure and expression of the alternative sigma factor, RpoN, in *Rhodobacter capsulatus*; physiological relevance of an autoactivated nifU2-rpoN superoperon. *Mol. Microbiol.* **11**:51–65.

7. **Dryer, R. L., and G. F. Lata.** 1989. Experimental biochemistry, p. 298–331. Oxford University Press, Oxford, United Kingdom.
8. **Gross, C. A., M. Lonetto, and R. Losick.** 1992. Bacterial sigma factors, p. 129–176. In S. L. McKnight and K. R. Yamamoto (ed.), *Transcriptional regulation*, vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
9. **Hsieh, M., and J. D. Gralla.** 1994. Analysis of the N-terminal leucine heptad and hexad repeats of sigma 54. *J. Mol. Biol.* **239**:15–24.
10. **Hsieh, M., Y. Tintut, Y. Jiang, and J. D. Gralla.** 1994. Functional roles for the glutamines within the glutamine-rich region of the transcription factor sigma 54. *J. Biol. Chem.* **269**:373–378.
11. **Hunt, T. P., and B. Magasanik.** 1985. Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerase and the products of *glnF*, *glnG*, and *glnL*. *Proc. Natl. Acad. Sci. USA* **82**:8453–8457.
12. **Kustu, S., E. Santero, J. Keener, D. Popham, and D. Weiss.** 1989. Expression of sigma 54 (*ntrA*)-dependent genes is probably united by a common mechanism. *Microbiol. Rev.* **53**:367–376.
13. **Leung, D. W., E. Chen, and D. V. Goeddel.** 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**:11–15.
14. **Magasanik, B.** 1996. Regulation of nitrogen utilization, p. 1344–1356. In F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, Jr., B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 1. ASM Press, Washington, D.C.
15. **Malhotra, A., E. Severinova, and S. Darst.** 1996. Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. *Cell* **87**:127–136.
16. **Merrick, M. J.** 1993. In a class of its own- the RNA polymerase sigma factor σ^{54} (sigma N). *Mol. Microbiol.* **10**:903–909.
17. **Merrick, M. J., and S. Chambers.** 1992. The helix-turn-helix motif of σ^{54} is involved in recognition of the –13 promoter element. *J. Bacteriol.* **174**:7221–7226.
18. **Porkess, R.** 1991. The HarperCollins dictionary of statistics, p. 139–141. HarperCollins Publisher, Inc., New York, N.Y.
19. **Sasse-Dwight, S., and J. D. Gralla.** 1990. Role of eukaryotic-type domains found in the prokaryotic enhancer receptor factor σ^{54} . *Cell* **62**:945–950.
20. **Tintut, Y., and J. D. Gralla.** 1995. PCR mutagenesis identifies a polymerase-binding sequence of sigma 54 that includes a sigma 70 homology region. *J. Bacteriol.* **177**:5818–5825.
21. **Tintut, Y., C. Wong, Y. Jiang, M. Hsieh, and J. D. Gralla.** 1994. RNA polymerase binding using a strongly acidic hydrophobic-repeat region of sigma 54. *Proc. Natl. Acad. Sci. USA* **91**:2120–2124.
22. **Wang, J. T., A. Syed, M. Hsieh, and J. D. Gralla.** 1995. Converting *Escherichia coli* RNA polymerase into an enhancer-responsive enzyme—role of an NH_2 -terminal leucine patch in sigma 54. *Science* **270**:992–994.
23. **Wong, C., and J. D. Gralla.** 1992. A role for the acidic trimer repeat region of transcription factor σ^{54} in setting the rate of temperature dependence of promoter melting *in vivo*. *J. Biol. Chem.* **267**:24762–24768.
24. **Wong, C., Y. Tintut, and J. D. Gralla.** 1994. The domain structure of sigma 54 as determined by analysis of a set of deletion mutants. *J. Mol. Biol.* **236**:81–90.