JOURNAL OF BACTERIOLOGY, Oct. 1995, p. 5818–5825 Vol. 177, No. 20
0021-9193/95/$04.00+0

PCR Mutagenesis Identifies a Polymerase-Binding Sequence of Sigma 54 That Includes a Sigma 70 Homology Region

YIN TINTUT AND JAY D. GRALLA*

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

Received 15 May 1995/Accepted 4 August 1995

Sigma 54 is a minor bacterial sigma factor that is not a member of the sigma 70 family of proteins but binds the same core RNA polymerase. Previously, we identified a region of sigma 54 that is important for binding core polymerase. In this work, PCR mutagenesis was used to identify specific amino acids important for this binding. The results show that important residues are clustered most closely in a short sequence that was previously speculated to be potentially homologous to a sequence in sigma 70. The mutagenesis also identifies important residues in the flanking hydrophobic-acidic region of sigma 54, which is absent in sigma 70. Overall, the data indicate that sigma 54 binds core polymerase through a sequence homologous to that of sigma 70 but in addition uses unique motifs to modify this interaction.

A number of different forms of RNA polymerase holoenzyme exist in *Escherichia coli* (5). These forms differ in that they contain different sigma factors, which determine the DNA-binding properties of the holoenzymes. This, in turn, determines the pattern of promoter usage in the cell. Most sigma factors are members of the sigma 70 family of proteins and contain variants of conserved motifs (12). The sole known exception is sigma 54, which is not a member of the sigma 70 family of proteins (12, 13). The amino acid sequence of sigma 54 has no obvious resemblance to that of sigma 70; however, it contains several regions that are largely conserved in sigma 54 proteins from a variety of bacteria (13, 14). Despite the lack of any significant sequence similarity, sigma 54 and sigma 70 bind the same core RNA polymerase.

When sigma 54 associates with core polymerase, it directs the polymerase to participate in a transcription mechanism that is different from that of sigma 70 (3, 8, 16, 18). This mechanism has a number of very unusual features, and they resemble features found in mammalian transcription mechanisms (4). Therefore, identification and characterization of the elements within sigma 54 that bind to core polymerase is of interest, because it would provide clues to how core polymerase can be bound by two quite different proteins and have its properties changed by the binding.

The sigma 70 family contains four conserved regions, all of which are absent in sigma 54 (5, 12). A core polymerase-binding region has been mapped to one of these sigma 70 regions, at the N-terminal side of conserved region 2 (10; see reference 12 for a review). Sigma 32, a sigma 70-related protein, also uses this analogous determinant to bind polymerase (9) and may have additional interactions near conserved region 3 (22). An attempt was also made to determine the core polymerase-binding region of sigma 54 but was thwarted by an inability to detect tight binding in the assay used (9).

Functional domains within *E. coli* sigma 54, including the one required for polymerase binding, have recently been identified by using a set of deletion mutants (20). The results showed that sigma 54 contains at least three functional domains: the C terminus, which binds DNA; the N terminus, which is necessary for transcriptional activation events; and the central region, which is necessary for binding core polymerase. Within this central region are features of potential interest: hydrophobic residues, including a subset arranged in a heptad repeat, and numerous acidic residues interdigitated with these hydrophobic residues.

Previously, we introduced and studied a small series of point mutations created within the core-binding region of sigma 54 by site-directed mutagenesis (19). The results showed that changes in multiple hydrophobic residues or multiple acidic residues could reduce the ability of sigma 54 to bind core polymerase. In that work, we noticed a small patch of residues (VEAVLK) between positions 175 and 180 that bear a slight resemblance to a patch (VEANLR) within the core-binding region of sigma 70 (10). Sequence comparisons (GAP alignment), however, do not indicate significant similarity between this or any region of sigma 54 and any region of sigma 70, including conserved region 2. Some acidic and hydrophobic mutations were within and directly flanking this small region of potential similarity.

In the present study, we attempted to assess the significance of amino acids in this general region by using random mutagenesis. Nonspecific mutations were introduced within this region via *Taq* polymerase in PCR, and the binding of the resulting sigma 54 mutants to polymerase was tested. The results confirmed the importance of this region for polymerase binding. Mutations causing loss of binding were found in all three features, hydrophobic, acidic, and sigma 70 similarity. However, the largest concentration of mutations was within or just adjacent to the potential sigma 70 similarity region. We infer that although the two sigmas have no statistically significant amino acid similarity, they use homologous regions as part of the mechanism of RNA polymerase binding. However, important determinants of sigma 54 binding also reside in the flanking regions, which are strikingly different from in the two sigmas.

MATERIALS AND METHODS

Strains and plasmids. Plasmid pYT2 was generated from the parent plasmid pTH7, which carries the *E. coli* sigma 54 gene (7). The sigma 54 gene is under the control of P_{tac}, an IPTG (isopropyl-β-D-thiogalactoside)-inducible promoter. However, in the absence of IPTG, there is still leaky expression of sigma 54, which is sufficient for growth on minimum media plates and for expression of the

* Corresponding author. Phone: (310) 825-1620. Fax: (310) 206-7286. Electronic mail address: Gralla@ewald.mbi.ucla.edu.

glnAp2 promoter (described below). A unique restriction site (*Bst*EII) was introduced at amino acid 153 by oligonucleotide-directed mutagenesis with the Sculptor in vitro mutagenesis system, RPn 1526, and the protocols recommended by Amersham, Arlington Heights, Ill. Two primers (20 bases long) flanking the region 154 to 193 were synthesized on a Gene Assembler Plus (Pharmacia-LKB Biotechnology, Inc., Piscataway, N.J.). *E. coli* YMC109^{tk} (*thi endA hsr* Δ*lacU169 rpoN*::Tn*10*/F′ *pro lacI*^{sq}ZU118 Tn5-102), lacking a wild-type chromosomal copy of the sigma 54 gene, was used as a host (6).

PCR mutagenesis. PCR for the Mn^{2+} library was performed essentially as described previously (11). Briefly, plasmid pYT2 was linearized at the unique restriction site *Hin*dIII. PCR mixtures (100 μl) contained pYT2 template (1.0 ng); 1× magnesium-free *Taq* polymerase buffer (Promega); 0.5 mM $MnCl_2$; 6.1 mM $MgCl_2$; 10 mM β-mercaptoethanol; 10 μl of dimethyl sulfoxide; 1 mM (each) dGTP, dCTP, and dTTP; 0.2 mM dATP; two primers (1 μg); and 1 μl of *Taq* DNA polymerase.

The same PCR was performed for the Mg^{2+} library with the following modifications: 30 cycles of PCR in 1.5 mM $MgCl_2$ and 0.2 mM each deoxynucleoside triphosphate. $MnCl_2$, β-mercaptoethanol, and dimethyl sulfoxide were omitted from the reaction mixture.

The PCR products were purified on spin columns (QIAGEN) and digested with *Bst*EII and *Bgl*II. The digested amplification product was gel purified. The plasmid vector was prepared by digesting pYT2 with the same two enzymes and also gel purified. The ligation mixture of the PCR fragment and plasmid vector was transformed into YMC109^{tk} and plated on Luria broth (LB) plates containing tetracycline, kanamycin, and ampicillin. These colonies were tested by plating as described below.

Plating test. W-arg/X-Gal plates were used; these contain glucose minimal medium with arginine as the sole added nitrogen source (6) (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 [10.0-mg/ml] thiamine, 0.215 ml of 1 M $MgSO_4$, 2 ml of 25-mg/ml L-arginine, 1 ml of 20-mg/ml X-Gal [5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside], and appropriate antibiotics). The plating test was done in the absence of IPTG, taking advantage of the leaky expression from the tac promoter. The host strain, YMC109^{tk}, does not contain a chromosomal sigma 54 gene, and thus sigma 54-dependent *lacZ* expression relies on the plasmid. It also contains a glnAp2-*lacZ* fusion (1), which facilitates screening for sigma 54-dependent glnAp2 expression by blue color formation.

The fresh transformants from the PCR mutagenesis were transferred individually, via pipette tips, to the W-arg/X-Gal plates containing appropriate antibiotics. The plates were incubated at 37°C for 16 to 18 h. Two populations could be seen at this time. One set of colonies was blue, indicating significant levels of sigma 54-dependent expression, and these were taken to be colonies containing functional sigma 54. Other transformants gave either no growth or white pinpoints, and these were taken to contain nonfunctional sigma 54.

Gel mobility shift assay and sequencing. Extracts from the colonies of interest were prepared and subjected to gel mobility shift assay as described previously (19). The plasmids from the colonies were isolated, and the mutagenized region was sequenced with Sequenase version II and protocols recommended by United States Biochemical.

Statistical analysis. A statistical method was devised to analyze the results from the manganese library, in which each mutant sigma 54 contained multiple changes (see below). The analysis was designed to judge whether particular residues are preferentially mutated in non-core-binding forms of sigma 54. Each position was analyzed by using the distribution of changes among core-binding and non-core-binding forms of sigma 54. For each position, i, containing n_i mutations, the deviation (δ_i) of the observed number of mutations, x_i, in nonbinders from the expected number of such mutations (μ_i, as determined below) was calculated (15). This deviation was expressed in standard deviation units (s_i). The expected number (μ_i) was calculated from the probability (p) of seeing a nonbinding sigma 54 at each residue, where p is the fraction of mutations in the entire data set that are associated with nonfunctional forms of sigma 54. Residues with a δ value of 1 (68% confidence level) or greater were considered to be important positions where mutations are likely to damage polymerase binding.

Formulae used in this analysis are as follows:

$$\mu_i = n_i p \quad (1)$$

$$\delta_i = (x_i - \mu_i)/s_i \quad (2)$$

$$p \text{ (applies to the entire data set)} = \text{number of mutations in nonbinding forms of sigma 54/total number of mutations} \quad (3)$$

$$s_i = [n_i p (1 - p)]^{1/2} \quad (4)$$

From these equations, a sample calculation for residue V-178 (i = 178, where 9 of 10 changes are in nonfunctional colonies compared with 264 of 405 in the entire data set), is as follows: n_{178} = total number of mutations at residue 178 = 10; from equation 3, p = 264/405 = 0.65; from equation 1, $\mu_{178} = 10 \times 0.65 = 6.5$; from equation 4, $s_{178} = [10 \times 0.65 \times (1 - 0.65)]^{1/2} = 1.51$; x_{178} = number of mutations that are in nonfunctional colonies = 9; from equation 2, $\delta_{178} = [9 - 6.5)]/1.51 = 1.66$. This indicates that the observed frequency of mutations at position 178 appearing in nonbinding forms sigma 54 was greater than the expected number by 1.66 standard deviation units.

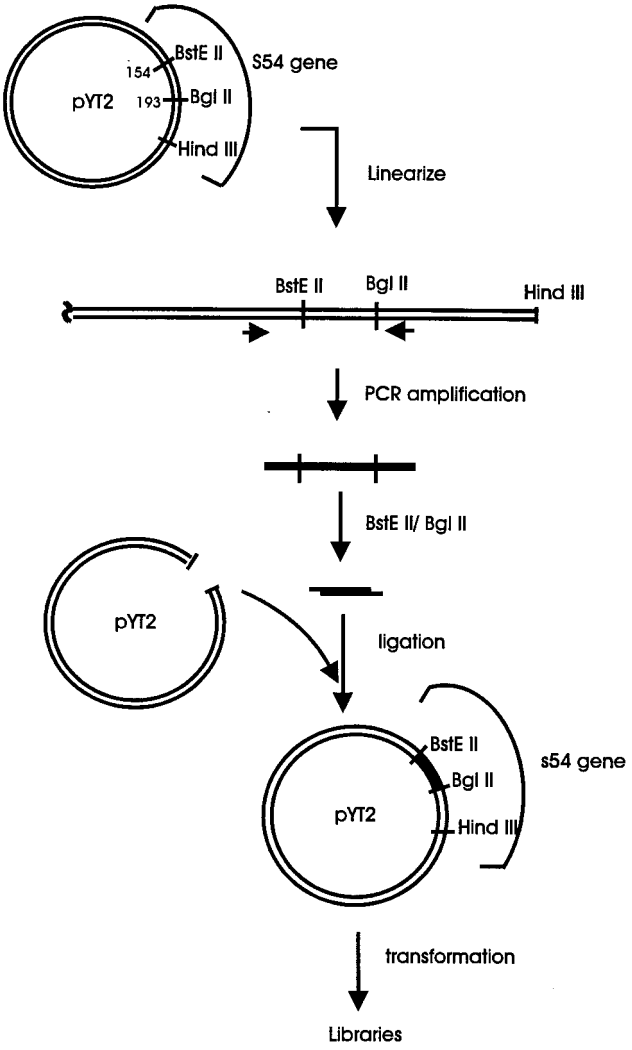

FIG. 1. Outline of the PCR mutagenesis protocol.

RESULTS

The region containing features of current interest for polymerase binding lies approximately between amino acids 160 and 190 of sigma 54 (19). To introduce "random" mutations within this region by PCR procedures, a unique restriction site was introduced in the codon corresponding to amino acid 153. This was paired, for PCR mutagenesis, with a naturally occurring unique site at amino acid 193. The procedure uses plasmid pYT2, which carries the entire *E. coli* sigma 54 gene. To generate libraries, the region containing amino acids 154 to 193 was amplified by PCR, subcloned back into pYT2, and transformed into YMC109^{tk}, which lacks a chromosomal copy of the sigma 54 gene. Colonies were screened for blue color on W-arg/X-Gal plates to indicate sigma 54 function. This plating procedure was verified by using preexisting known mutants of sigma 54 (Fig. 1) (see Materials and Methods for details).

PCR amplification was done under two conditions, one of which enhances error-prone copying to enhance the mutation frequency. Amplification by *Taq* polymerase with 1.5 mM

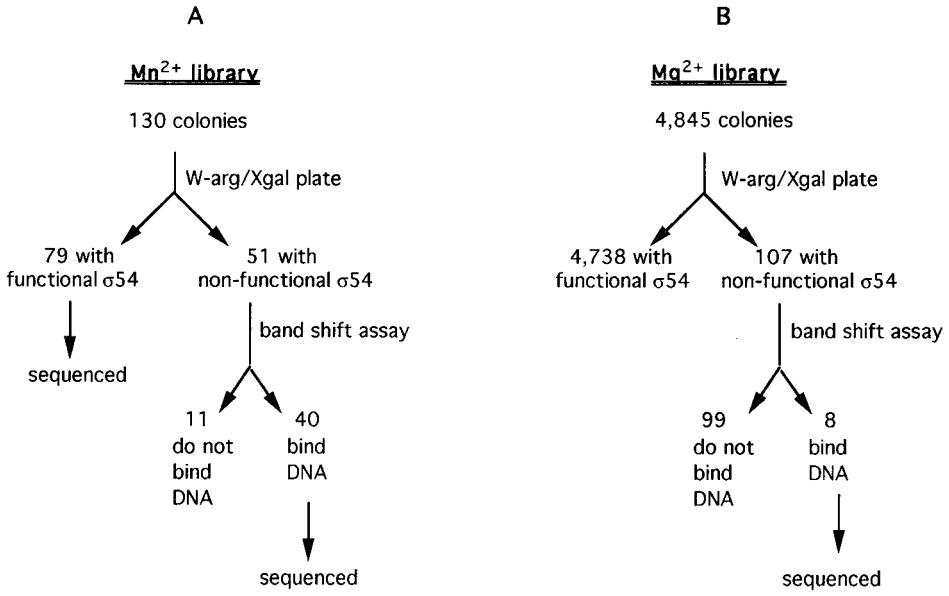

FIG. 2. Flow chart of the procedure leading to the generation of sequence information from mutated forms of sigma 54.

Mg^{2+} is known to introduce mutations at a low frequency (11, 21). The point of creating this library is to identify amino acids that are critical for sigma 54 function. However, because the mutation frequency is low, the information derived from this library will be limited. To expand the information content, a second amplification procedure was used in the presence of 0.5 mM Mn^{2+}. Copying is more error prone under these conditions (11), and multiple mutations might be expected. In this case, there are more data but data analysis is more complicated than that for single changes.

Mn^{2+} library. The outline of the procedure used to screen the Mn^{2+} library is shown in Fig. 2A. An initial transformation yielded 130 colonies on LB plates. When these were tested for function on W-arg/X-Gal plates, 79 colonies were found to be functional and 51 colonies were found to be nonfunctional. The nonfunctional colonies were subjected to further screening.

The 51 colonies could be nonfunctional as a result of lack of sigma 54 binding to polymerase or a variety of other reasons, including both translational truncation of the protein by mutation and general lack of expression. Previously, we showed that expressed sigma 54 is still capable of binding to a *Rhizobium meliloti nifH* promoter fragment (2) even if it lacks the ability to bind core RNA polymerase (19). Therefore, we used DNA binding as a screen. Nonexpressed proteins obviously will not bind DNA. In addition, we determined that truncated proteins do not bind DNA, because translation of the C-terminal domain of sigma 54 is required for DNA binding (20).

At the same time, we screened directly for ability of sigma 54 to bind polymerase. Both screens involve a single band shift assay described previously (19). Crude cell extracts were prepared from each of the 51 nonfunctional colonies. The forms of sigma 54 present in each colony were then assayed with a labeled *Rm nifH* promoter fragment as the probe.

The results of the screening as applied to 4 of these 51 colonies are shown in Fig. 3. Lane 1 shows the result of the assay in which an extract from colonies containing wild-type sigma 54 was used. As described previously (19, 20), DNA binding by sigma 54 alone is indicated by a broad band in the region marked by a bracket in the figure. Polymerase binding by sigma 54 is indicated by a band in the holoenzyme position as indicated by an arrow in the figure (see reference 19 for the identification of these bands).

The four colonies containing nonfunctional mutants assayed here fall into two distinct classes. Lanes 3 and 5 show no indication of any ability to bind the DNA probe. Such mutants were not characterized further, because they are not specifically defective in the ability to bind polymerase. By contrast, lanes 2 and 4 show mutants that retain the ability to bind DNA probe, as shown by the appearance of a band in the bracketed region. However, neither lane shows a band in the holoenzyme position (compare with the positive control pattern in lane 1). The results show that the forms of sigma 54 present in lanes 2 and 4 can bind DNA and are specifically defective in the ability

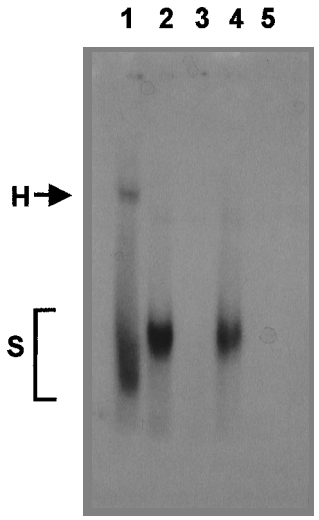

FIG. 3. Band shift assay in crude cell extracts. Lanes: 1, wild-type sigma 54; 2 to 5, sigma 54 mutants. The arrow indicates the mobility of holoenzyme (H) (core polymerase and sigma 54) position, and the bracket indicates the mobility of sigma 54 (S)-alone position (19).

TABLE 1. Distribution of mutations in core-binding and non-core-binding forms of sigma 54

No. of mutations in sigma 54	No. of colonies with:	
	Core-binding sigma 54	Non-core-binding sigma 54
0	19	
1	20	
2	19	
3	6	1
4	6	8
5	9	4
6		7
7		7
8		7
9		0
10		5
11		1

to bring core RNA polymerase into the complex. Of the 51 colonies, 40 contained forms of sigma 54 that had this desired phenotype (Fig. 2A).

The region of the sigma 54 gene that was subjected to PCR mutagenesis was sequenced from the 79 functional forms of sigma 54 as well as the 40 non-core-binding forms of sigma 54 (Fig. 2A). The relationship between the functionality of sigma 54 and the number of mutations present is shown in Table 1. The data suggest that multiple mutations are associated with nonfunctional forms of sigma 54 that prevent polymerase binding. Nearly all of the nonbinding forms of sigma 54 contain at least four altered amino acids. No sigma 54 with six or more mutations was able to bind polymerase. If 6 of 40 residues have to be mutated to destroy the polymerase binding, there may be roughly five times as many unimportant residues as important residues for binding polymerase in this region. In any case, the data indicate that there are many residues within this region that can be mutated but do not lead to a loss of ability to bind polymerase.

Overall, the data reveal some individual changes that can be tolerated by sigma 54 and other changes that are associated with forms of sigma 54 that cannot bind polymerase. Defective core binding by sigma 54 does not result from degradation of unfolded forms, because all mutants retain the ability to bind DNA, although local structural changes are possible. In addition, no doubt there are many potential changes not represented in the data set. This is due to the need for multiple changes in codons to generate certain substitutions, to some bias in the creation of mutations by the *Taq* polymerase, and to the analysis of the limited number of colonies that are available.

The interpretation of changes that are tolerated for core binding by sigma 54 is straightforward. However, interpretation of individual changes within the set of multiple changes that cause sigma 54 to be nonbinding is more difficult. To assess the potential importance of individual amino acids in these cases, we analyzed the data by two independent methods: one involves the number of changes, and the other involves the type of mutations.

Figure 4A shows the distribution of changes resulting from sequencing the sigma 54 mutants from the Mn^{2+} library. The amino acid sequence of the wild-type sigma 54 is shown, with the residue positions at the top of the figure. The pair of rows just below shows the number of colonies in which each residue is mutated in either nonbinding or binding forms of sigma 54. For example, at residue 155 (threonine), there were eight colonies with nonbinding sigma 54 in which this residue was changed. The same threonine was changed in nine different functional colonies (the types of changes are considered in the second screen below).

To determine the statistical significance of the distribution of changes in each of the residues within this region, the δ value for each residue was calculated (Fig. 4B; see Materials and Methods for a sample calculation). The δ value measures the deviation of the observed number of nonfunctional mutations from the expected number, with the latter determined from the statistics of the entire data set. Large positive values of δ (in standard deviation units) indicate that changes in the residue are associated with statistically significant loss of function. The extreme example of this would be leucine 154, which has a δ value of greater than 3 standard deviation units because it is changed 18 times in nonfunctional sigma 54 and never appears in a functional sigma 54.

Figure 4B shows the δ value in standard deviation units plotted against the positions of residues. The distribution has its strongest significant cluster in the region 177 to 182. These changes are either in or just adjacent to the region (between residues 175 and 180) previously identified as being potentially homologous to residues within the sigma 70 core-binding region (19). The result provides an initial indication that this region is important for polymerase binding.

Of 40 amino acids, 11 displayed δ values of +1 or greater, indicating that they are preferentially mutated in nonbinding sigma 54. In view of the suggestion (above) that one in five residues may be important, we decided to consider these amino acids to be initial candidates for important residues. These 11 amino acids were L-154, E-163, I-165, I-170, A-177, V-178, L-179, R-181, I-182, G-189, and V-190 (Fig. 4A). We note that only 2 of these 11 residues are hydrophilic. On the other hand, seven positions show δ values of −1 or less, and these consist exclusively of hydrophilic residues: T-155, E-159, S-164, E-168, D-173, Q-183, and K-193. The large negative δ values at these positions may indicate that changes here actually assist polymerase binding, raising the possibility that the wild-type hydrophilic residues function to reduce polymerase binding to levels appropriate for the minor sigma factor.

We also devised a second screen for residues that reduce binding. The criteria used in this screen are based on the types of mutations observed at each position (Fig. 4C). The premise of this second screen is that a residue is more likely to be important if it is associated with changes that occur only in nonbinding forms of sigma 54. The last row of Fig. 4C shows every mutation that was found only in nonbinding forms of sigma 54. We required that a residue be associated with at least two such "unique" changes (that is, to be changed to two residues that never appear in that position in functional sigma 54) to pass this second screen. To increase the stringency, very conservative changes were not considered significant. That is, we considered the following amino acids to be equivalent: D = E, L = I, S = T, and K = R. For example, residue V-178 passes the screen because it is changed to alanine or to aspartate only in nonbinding forms of sigma 54. By contrast, E-169 would not pass this screen because, although it is changed to glycine in nonbinding forms of sigma 54, the same change also appears in forms of sigma 54 that retain core binding.

This second analysis, independent of the first screen, selected 15 residues which were changed uniquely in at least two nonfunctional colonies. These 15 unique residues were L-154, E-159, E-163, I-165, I-170, D-171, I-172, A-177, V-178, L-179, K-180, R-181, I-182, F-185, and G-189 (shaded residues in Fig. 4C). Yet again, the residues between positions 177 and 182, which are within the potential sigma 70 homology, are sug-

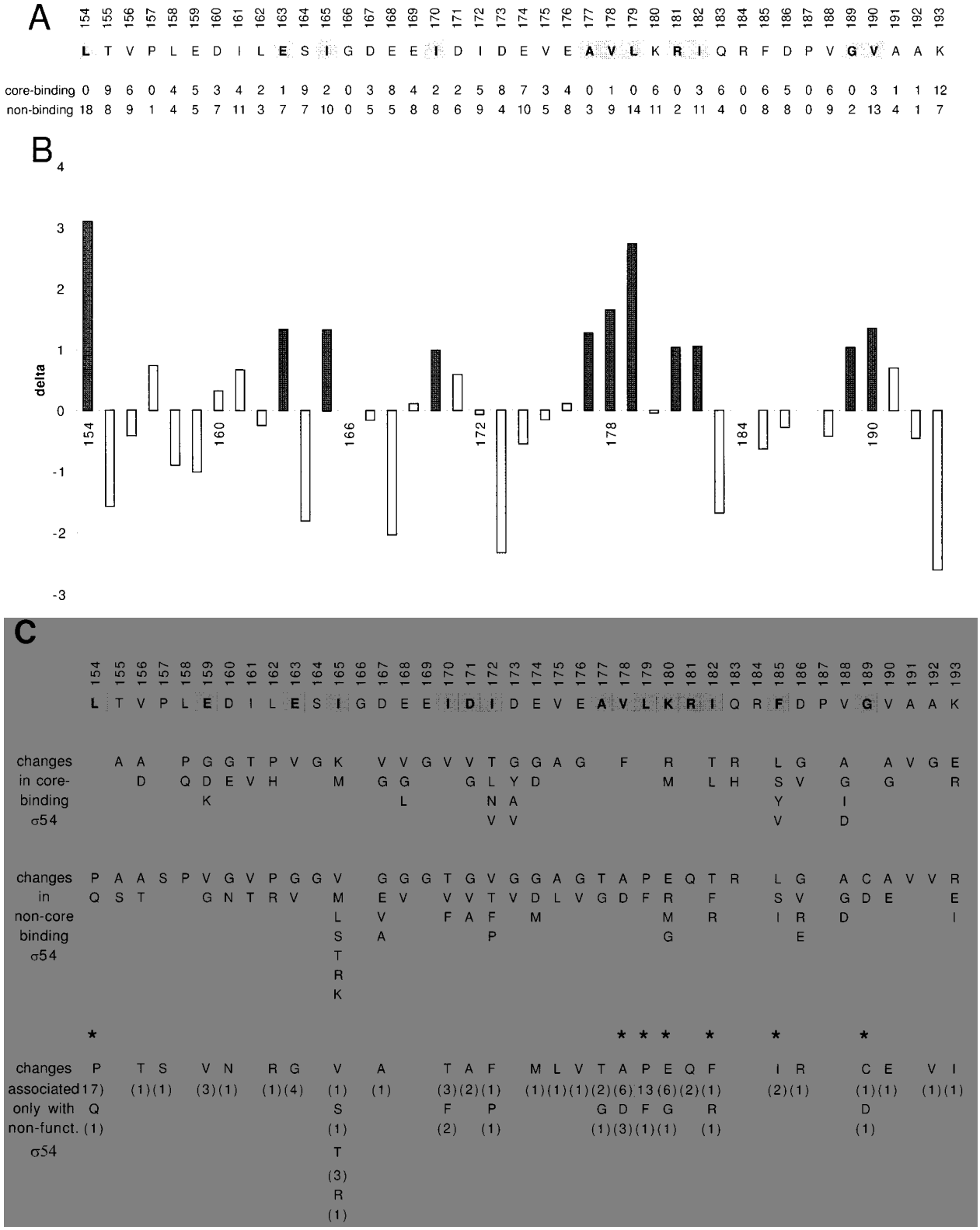

FIG. 4. Analysis of mutations from the Mn^{2+} library performed by addressing the frequency and type of mutations. (A) The number of colonies carrying mutations at a residue is indicated. Numbers are compiled separately for core-binding and non-core-binding forms of sigma 54. The shaded residues are suggested to be important ($\delta \geq 1$). (B) The δ value, a measure of the statistical significance of the residue being important (see Materials and Methods), was plotted against the position of each residue. High bars above the axis indicate residues that are more likely to be statistically significant. (C) All the individual mutations are shown. Using E-159 as an example, E was found changed to G, D, and K in core-binding forms of sigma 54 and to V and G in non-core-binding forms of sigma 54. The change to V was observed in three different nonbinding forms of sigma 54. The shaded residues are associated with at least two nonbinding forms of sigma 54 in which that residue was altered in a way that never occurred in a functional sigma. *, positions that are mutated in nonbinding forms of sigma 54 that do not occur in combination with other mutations listed in the last row.

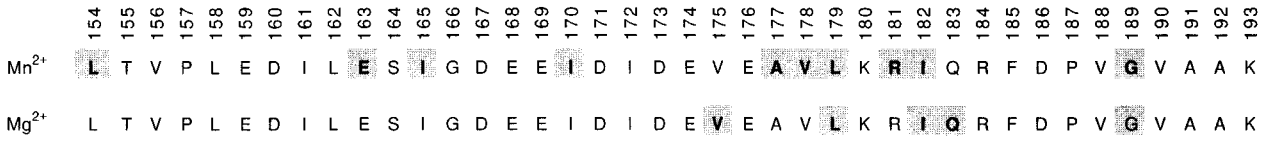

FIG. 5. Important residues (shaded) are as shown separately from analysis of the Mn^{2+} and Mg^{2+} libraries.

gested to be important. Of the 40 nonfunctional forms of sigma 54, 10 contained only one unique change, and these were in one of the seven residues marked non-binding shown in the last row of Fig. 4C. These again center within the potential sigma homology region. Of the 40 forms of sigma 54, 28 contained multiple unique changes and only the 2 remaining forms contained no unique residues among the six or seven mutations observed in each.

Of 11 residues identified by the statistical analysis of Fig. 4A, 10 also passed the second screen (V-190 is the exception) and are thus prime candidates for amino acids that are important for core polymerase binding. These residues are L-154, E-163, I-165, I-170, A-177, V-178, L-179, R-181, I-182, and G-189 (first row of Fig. 5). All have a high probability of being changed preferentially in non-core-binding forms of sigma 54 (Fig. 4B), and each is mutated in at least two ways that occur only in nonbinding forms (Fig. 4C).

Mg^{2+} library. As discussed above, a library with DNA obtained from amplification reactions containing magnesium rather than manganese was also created. These transformants were subjected to the identical screen for sigma 54 function involving plating first on LB and then on W-arg/X-Gal media. Most colonies survived this screen, as expected from the low frequency of introduction of errors in this PCR amplification protocol. Of 4,845 colonies, only 107 proved to be nonfunctional by this plate test (Fig. 2B). This 2% yield is obviously far less that the high yield obtained from the manganese library, which was created by error-prone amplification.

The 107 nonfunctional colonies were further screened by band shift assay as described above (see Fig. 3 for an example). Of these 107, only 8 retained the ability to bind DNA probe while losing the ability to bind core polymerase (Fig. 2B). The remaining colonies were presumably largely frameshift or translation stop mutations. Thus, the yield of information from this analysis is quite low; only 8 colonies of nearly 5,000 gave the desired phenotype. This low yield limits the extensive use of this procedure, because each nonfunctional colony (107 in this case) must be analyzed by band shift assays.

Upon DNA sequence analysis, seven of the eight colonies proved to have a single change in the amplified region and one colony had a double mutation. Each colony represented a different change. All changes are between residues 175 and 189: V175D, L179P, I182N, Q183P, G189V, and G189R (Mg^{2+} row in Fig. 5). We observed changes of L-179 to proline in two different colonies and also changes of V-175 to aspartate in two colonies, one with a single change and the other paired with V-156 to I. No analysis is required to interpret these results. Single changes in each of these five amino acids disrupt the ability of sigma 54 to bind core polymerase. As indicated independently in each of the two previous analyses, the highest concentration of mutations is near the potential sigma 70 homology region centered at residue 179.

Three of these five residues were also identified by analysis of the manganese library (L-179, I-182, and G-189 [Fig. 5]). Possibly, the other changes identified in the manganese library are not so damaging as to be able to fully disrupt polymerase binding without the assistance of other changes. Alternatively, it is obvious that the magnesium screen is nowhere near saturated from the isolation of eight positive colonies. Because 5,000 colonies were screened on plates and over 100 colonies were screened by band shift, it might require thousands of proteins to be assayed by band shift assay to ensure saturation. Nonetheless, the two screens together identified 12 of the 40 residues in this region as being important for the binding of core polymerase (Fig. 6). Below, we discuss the relationship between the residues identified by these two PCR protocols and the motifs present in this region of sigma 54.

DISCUSSION

Sigma 54 and sigma 70 direct polymerase to participate in quite different types of transcription mechanisms (4). The two sigmas are not related by sequence but nonetheless bind the same core RNA polymerase (12, 13). Prior work identified a region of sigma 54 that is important for its binding to core RNA polymerase (20). This region contained three motifs of potential interest: hydrophobic, acidic, and a very short stretch that bears a slight resemblance to a region of sigma 70 involved

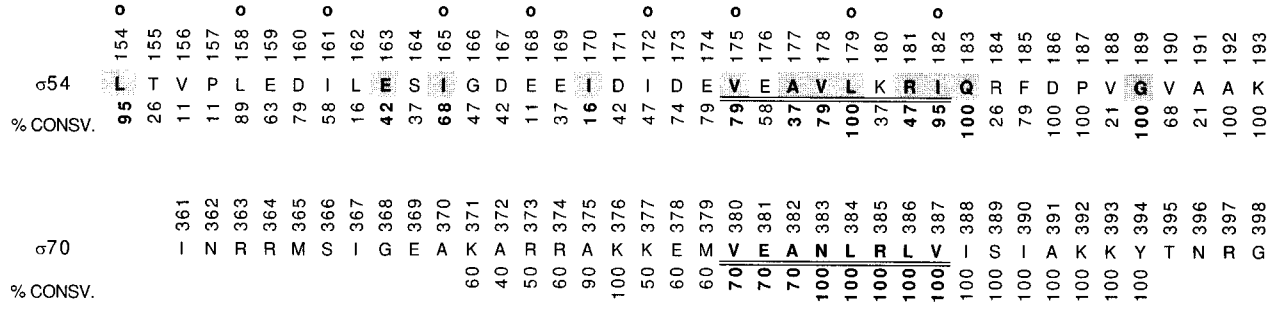

FIG. 6. Important residues in sigma 54 and comparison with sigma 70 (12). The shaded residues in sigma 54 are the important residues from analyses of both libraries. The homologous patch of residues between the two sigmas is underlined. The alignment of 19 different types of sigma 54 (14) by MegAlign by LASERGENE Navigator introduced a gap between residues 165 and 171. The open circles above the sigma 54 sequence indicate residues that are in the first and fourth positions of a potential heptad repeat. The sequences and the percent conservation (% consv.) within each family are shown for each sigma factor. The following amino acids were considered to be equivalent in calculating the percent conservation: D = E, L = I, S = T, and K = R.

in polymerase binding (19). Although this latter resemblance is too slight to be apparent in statistics-based programs, the results of the current mutagenesis study strongly support the importance of this small region.

Previously, we pointed to the short sigma 54 sequence VEA VLK as possibly similar to the sigma 70 sequence VEANLR, but we could not reach a conclusion concerning its functional significance (19). Of the 12 amino acids identified as important in the present work, 4 are in this short stretch of 6 and 3 are directly adjacent. That is, 7 important residues are now found to be clustered in this 9-amino-acid region, whereas the other 5 important residues are scattered in the 31 flanking amino acids (Fig. 6). Moreover, the residues within this region are identified as important independent of the details of the analysis. That is, four come from single changes in sigma 54 (Fig. 5), five have high delta values in the analysis of Fig. 4B, and six are identified by the analysis of Fig. 4C. Therefore, we infer that this sigma 70 similarity region is indeed a critical determinant of polymerase binding.

Figure 6 displays the important residues and shows the extent to which they are conserved among sigma factors. Within the 9-amino-acid region just identified (amino acids 175 to 183), the 7 residues deemed important in assisting polymerase binding by mutagenesis are conserved between 37 and 100% in a compilation of sigma 54 sequences from 19 different bacteria (14) (% consv.; row 2). The comparable region is conserved between 70 and 100% among sigma 70 proteins from 10 different bacteria (% consv.; row 4) (group I sigma 70 [12]). This high degree of conservation supports the general importance of this region. We note, however, that some of the very conserved residues in sigma 54 are not found to be important by the analysis. Because the analysis is based on a limited number of colonies, it is likely that not all the important residues were identified. Moreover, some conserved residues may have related functions such as preventing polymerase from being bound so tightly by sigma 54 as to interfere with the association of polymerase with sigma 70 (for example, residues with large negative delta values). Less highly conserved residues that are found to be important may vary among organisms because of differences in their polymerases or differences in the affinity for polymerase that is appropriate to the physiology of the different organisms.

The polymerase-binding determinants of sigma 54 and sigma 70 are different in three identifiable ways. First, they differ within the homology region. For example, the fourth amino acid (residue 178) is valine in sigma 54 and is 79% conserved among 19 sigma 54 proteins. However, the fourth residue (residue 383) for sigma 70 proteins is asparagine, and this is 100% conserved among 10 sigma 70 proteins. We infer that the two sigmas use a similar determinant in binding polymerase but that important differences exist within it.

Second, the results support the importance of sigma 54 hydrophobic residues in the flanking region (19), which are lacking in sigma 70. Seven hydrophobic residues are found to be important in the region between amino acids 154 and 182. These include three hydrophobic residues that were previously selected for site-directed mutagenesis (I-165, V-175 and L-179 [19]). This region, now known to be important for polymerase binding, was proposed to have the potential to form an amphipathic alpha-helix on the basis of the presence of hydrophobic residues in a heptad repeat (17). Site-directed mutagenesis showed that some hydrophobic residues on the first and fourth positions of the repeat (Fig. 6) were important for binding core polymerase (19). However, this does not settle whether the hydrophobic repeat arrangement is important; of the maximum of eight such residues, five are identified as important (L-154, I-165, V-175, L-179, and I-182) and three are not (L-158, I-161, and I-172), and two hydrophobic positions (V-178 and I-170) shown to be important are not on this potential repeat. In any case, the very much lower frequency of hydrophobic residues in the sigma 70 core-binding region supports the view that these residues are related to a sigma 54 specific function.

The third striking difference is the net charge in the amino-terminal sides of this region in the two sigmas; sigma 54 is strongly acidic and sigma 70 is strongly basic, as discussed previously (19). The mutagenesis identifies only a single acidic residue within this region. We cannot extend our prior discussion in this regard because, as speculated previously, a greater change in net charge may be required to alter function, which would require numerous changes.

We conclude that despite the lack of statistically significant sequence similarity, sigma 54 and sigma 70 use a homologous sequence to help bind core RNA polymerase. Both sigmas also use accessory interactions, which in the case of sigma 54 are concentrated significantly in the flanking acidic-hydrophobic region. Residues within this flanking region are not as well conserved as those within the sigma 70 homology region; in addition, a gap appears in the flanking region near the border of the sigma 70 homology region when all members of the sigma 54 family are aligned. This may imply that the flanking and homology regions bind two different sites on polymerase. That is, each sigma may bind multiple determinants on polymerase, only one of which is common. Because each sigma confers a distinct transcription mechanism on core polymerase, these difference may be important in determining the transcriptional properties of the bound polymerase. Thus, the two sigmas may use a related motif to stabilize bound polymerase and then use different secondary contacts to allow the polymerase to undergo the different changes required at sigma 54 and sigma 70 type promoters.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Training Grant GM07185 to Y.T. and grant GM35754 to J.D.G.

We thank Jim Bowie for his suggestions in the statistical analysis.

REFERENCES

1. **Backman, K., Y. M. Chen, and B. Magasanik.** 1981. Physical and genetic characterization of the gln A-gln G region of the Escherichia coli chromosome. Proc. Natl. Acad. Sci. USA **78:**3743–3747.
2. **Buck, M., and W. Cannon.** 1992. Specific binding of the transcription factor sigma 54 to promoter DNA. Nature (London) **358:**422–424.
3. **Collado-Vides, J., B. Magasanik, and J. D. Gralla.** 1991. Control site location and transcriptional regulation in *Escherichia coli*. Microbiol. Rev. **55:** 371–394.
4. **Gralla, J. D.** 1991. Transcriptional control—lessons from an *E. coli* promoter data base. Cell **66:**415–418.
5. **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.
6. **Hsieh, M., Y. Tintut, 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.
7. **Hunt, T. P., and B. Magasanik.** 1985. Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerase and the products of *glnF*, *glnF* and *glnL*. Proc. Natl. Acad. Sci. USA **82:**8453–8457.
8. **Kustu, S., E. Santero, J. Keener, D. Popham, and D. Weiss.** 1989. Expression of σ^{54} (*ntrA*)-dependent gene is probably united by a common mechanism. Microbiol. Rev. **53:**367–376.
9. **Lesley, S. A., M. A. Brow, and R. R. Burgess.** 1991. Use of in vitro protein synthesis from polymerase chain reaction-generated templates to study interaction of Escherichia coli transcription factors with core RNA polymerase and for epitope mapping of monoclonal antibodies. J. Biol. Chem. **266:**2632–2638.
10. **Lesley, S. A., and R. R. Burgess.** 1989. Characterization of the *Escherichia*

coli transcription factor sigma 70: localization of a region involved in the interaction with core RNA polymerase. Biochemistry **28:**7728–7734.
11. **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.
12. **Lonetto, M., M. Gribskov, and C. A. Gross.** 1992. The sigma 70 family: sequence conservation and evolutionary relationship. J. Bacteriol. **174:**3843–3849.
13. **Merrick, M. J., and J. R. Gibbins.** 1985. The nucleotide sequence of the nitrogen regulation gene *ntrA* of *Klebsiella pneumoniae* and comparison with conserved features in bacterial sigma factors. Nucleic Acids Res. **13:**7607–7620.
14. **Powell, B. S., D. L. Court, T. Inada, Y. Nakamura, V. Michotey, X. Cui, A. Reizer, M. H. Saier, Jr., and J. Reizer.** 1995. Novel proteins of the phosphotransferase system encoded within the rpoN operon of Escherichia coli. J. Biol. Chem. **270:**4822–4839.
15. **Rees, D. G.** 1989. Essential statistics, 2nd ed., p. 65–76. Chapman & Hall, New York.
16. **Reitzer, L., and B. Magasanik.** 1986. Transcription of *glnA* in *E. coli* is stimulated by activator bound to sites far from the promoter. Cell **45:**785–792.
17. **Sasse-Dwight, S., and J. D. Gralla.** 1990. Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor sigma 54. Cell **62:**945–954.
18. **Tintut, Y., J. T. Wang, and J. D. Gralla.** A novel bacterial transcription cycle involving sigma 54. Genes Dev., in press.
19. **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.
20. **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.
21. **Zhou, Y., X. Zhang, and R. H. Ebright.** 1991. Random mutagenesis of gene-sized DNA molecules by use of PCR with Taq DNA polymerase. Nucleic Acids Res. **19:**6052.
22. **Zhou, Y. N., W. A. Walter, and C. A. Gross.** 1992. A mutant σ^{32} with a small deletion in conserved region 3 of σ has reduced affinity for core RNA polymerase. J. Bacteriol. **174:**5005–5012.