Determinants for *Escherichia coli* RNA Polymerase Assembly within the $\beta$ Subunit

Yi Wang$^1$, Konstantin Severinov$^1$, Nick Loizos$^1$, David Fenyo$^2$
Ewa Heyduk$^3$, Tomasz Heyduk$^3$, Brian T. Chait$^2$ and Seth A. Darst$^1$*

$^1$Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue New York, NY 10021, USA
$^2$Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, 1230 York Avenue New York, NY 10021, USA
$^3$Department of Biochemistry and Molecular Biology St. Louis University Medical School, 1402 S. Grand Boulevard, St. Louis MO 63104, USA

*Corresponding author

We used binding assays and other approaches to identify fragments of the *Escherichia coli* RNAP $\beta$ subunit involved in the obligatory interaction with the $\alpha$ subunit to form the stable assembly intermediate $\alpha_2\beta$ as well as in the interaction to recruit the $\beta'$ subunit into the $\alpha_2\beta_0$ sub-assembly. We show that two regions of evolutionarily conserved sequence near the C terminus of $\beta$ (conserved regions H and I) are central to the assembly of RNAP and likely make subunit-subunit contacts with both $\alpha$ and $\beta'$.

© 1997 Academic Press Limited

**Keywords:** E. coli RNA polymerase assembly; transcription; limited proteolysis; protein-protein footprinting

**Introduction**

The *Escherichia coli* DNA-dependent RNA polymerase (RNAP) comprises an essential catalytic core of two $\alpha$-subunits (each 36.5 kDa), one $\beta$ (150.6 kDa), and one $\beta'$ (155.2 kDa) subunit. The holoenzyme has an additional regulatory subunit, normally $\sigma^{70}$ (70.2 kDa). The RNAP assemblies according to the following pathway (reviewed by Zillig *et al.*, 1976; Ishihama, 1981):

$$2\alpha \rightarrow \alpha_2 \rightarrow \alpha_2\beta \rightarrow \alpha_2\beta\beta' \rightarrow \alpha_2\beta\beta'\sigma$$

The RNAP $\alpha$ subunit serves as the initiator for RNAP assembly. Igarashi & Ishihama (1991) showed that truncated $\alpha$ derivatives lacking about 100 C-terminal residues are completely competent for dimerization, RNAP assembly, and basal transcription. Five regions of the $\alpha$ subunit (designated A to E), all within the N-terminal assembly domain, are conserved in sequence among $\alpha$ homologs of prokaryotic, eukaryotic, archaeabacterial, and chloroplast RNAPs (reviewed by Ebright & Busby, 1995; Heyduk *et al.*, 1996). A number of mutagenesis studies have localized determinants for $\alpha$ dimerization and interaction with the other RNAP subunits to within these conserved regions (Igarashi *et al.*, 1990, 1991; Hayward *et al.*, 1991; Igarashi & Ishihama, 1991; Kimura *et al.*, 1994; Kimura & Ishihama, 1995a,b). Hydroxyl-radical protein footprinting was used to confirm these results and strengthen the conclusion that $\alpha$ conserved regions A and B are involved in direct interactions with the $\beta$ subunit while $\alpha$ conserved regions C and D are involved in direct interactions with the $\beta'$ subunit (Heyduk *et al.*, 1996). The sequence conservation within these regions in all organisms suggests that $\alpha$ homologs in archaeabacterial and eukaryotic RNAPs perform analogous functions (Kolodziej & Young, 1991; Azuma *et al.*, 1993; Lalo *et al.*, 1993; Pati, 1994; Ulmasov *et al.*, 1996).

In contrast to the $\alpha$ subunit, very little is known about regions of the $\beta$ and $\beta'$ subunits involved in RNAP assembly and subunit-subunit interactions. Sequence comparisons among eubacterial, archaeabacterial, and eukaryotic $\beta$ subunit homologs reveal regions of high sequence similarity separated by regions that are poorly conserved (Falkenburg
et al., 1987; Sweetser et al., 1987; Berghofer et al., 1988; Pühler et al., 1989; see Figure 4). We have recently shown that active E. coli RNAP can be assembled in vitro using β subunit fragments along with the other RNA polymerase subunits (Severinov et al., 1995a, 1996). The N and C-terminal boundaries of these fragments roughly delineate independent structural domains of the β subunit that do not have to be covalently linked for assembly and activity of the RNAP.

Because of the apparently analogous role of the α subunit and its regions of conserved primary structure in E. coli RNAP assembly to the role of α homologs in RNAP assembly in other organisms including man, it seems reasonable to assume that regions of conserved sequence in the β subunit are also involved in assembly. Which conserved region(s) of β? In this study, we use a combination of approaches to identify the conserved regions of the E. coli RNAP β subunit involved in the obligatory interaction with α to form the stable assembly intermediate αβ as well as in the interaction to recruit the β′ subunit into the αβ sub-assembly. We show that the two C-terminal most conserved regions of β (conserved regions H and I according to the notation of Sweetser et al., 1987) are central to the assembly of RNAP and likely make subunit-subunit contacts with both α and β′.

Results

The C-terminal portion of the β subunit interacts with α

We have shown previously that the β subunit can be split nearly in half, into two separate peptide fragments, β residues 1 to 643 and 643 to 1342 (denoted β_{ABCD} and β_{EFGH}, respectively, to reflect the locations of the conserved regions; see Table 1 for a description of all the β subunit fragments used in this study), but still assemble in vitro with the other RNA polymerase subunits into active RNAP (Severinov et al., 1996). We used RNAP α subunit engineered with an N-terminal hexahistidine tag (His_{α}; Tang et al., 1995) in a Ni²⁺-NTA agarose co-immobilization assay to investigating the binding of these β subunit fragments to α (Figure 1). RNAP immobilized by virtue of His_{α} subunits was transcriptionally active in the immobilized state (Tang et al., 1995), indicating that the N terminus of at least one of the α subunits must be solvent exposed since immobilization of the RNAP in this way did not interfere with any necessary subunit-subunit interactions. Since overexpressed β and its fragments were generally found in inclusion bodies, the binding experiments were done by mixing His_{α} and β (or β fragments) together in the denatured state (in buffer containing 6 M guanidine-HCl) and then removing the denaturant by dialysis (see Materials and Methods). When a mixture of His_{α} and β was reconstituted and then loaded onto Ni²⁺-NTA agarose beads (Figure 1a, lane 1), the unbound material removed (Figure 1a, lane 2), the beads washed with 25 mM imidazole buffer (Figure 1a, lane 3), then eluted with 100 mM imidazole buffer, β was found in the eluted fraction along with His_{α}-α (Figure 1a, lane 4). Since rena-uterized β in the absence of His_{α}-α did not interact with the beads (Figure 1a, lane 8), we conclude that β was retained in the presence of His_{α}-α by virtue of a direct interaction with immobilized His_{α}-α, as expected. When a mixture of His_{α}-α and β_{ABCD} was loaded onto Ni²⁺-NTA agarose beads (Figure 1a, lane 9), all of the β_{ABCD} appeared in the unbound fraction (Figure 1a, lane 10), whereas an analogous experiment with His_{α}-α and β_{EFGH} indicated an apparently stoichiometric interaction of β_{EFGH} with the immobilized His_{α}-α. In the absence of His_{α}-α, β_{EFGH} did not interact with the Ni²⁺-NTA agarose beads (Figure 1a, lane 20) and so we conclude that β_{EFGH} by itself, directly interacts with α subunit whereas the remainder of the β polypeptide, β_{ABCD}, does not. Moreover, when a mixture of His_{α}-α, β_{ABCD}, and β_{EFGH} were loaded together onto Ni²⁺-NTA agarose beads, only β_{EFGH} was retained with His_{α}-α in the eluted fraction (Figure 1b). Thus, β_{ABCD} also does not interact with the complex between α and β_{EFGH} and so all of the determinants for the interaction between α and β appear to be contained within β_{EFGH}.

Identification of β subunit domains in complex with α by limited proteolysis

Limited proteolysis has often been used to define the domain organization of proteins (Wilson, 1991). When β alone was treated with limiting amounts of a variety of proteases, the C-terminal half was rapidly degraded (K.S., D.F., E. Severino-va, B.T.C., and S.A.D., unpublished results). We designed an assay to identify, if present, the proteolytically-resistant core of β_{EFGH} while in a complex with α. A battery of five proteases was used; trypsin, chymotrypsin, proteinase K, Pronase, and endoproteinase Glu-C. Rather than full length α, which itself is susceptible to proteolysis (Blatter et al., 1994), we used the assembly-competent N-
terminal domain (NTD) of α engineered with an N-terminal hexahistidine tag (His6-α[1-235], hereafter referred to as His6-αNTD; Tang et al., 1995). Control experiments showed that His6-αNTD, when immobilized on Ni²⁺-NTA agarose, was completely resistant to cleavage even by the highest concentrations of each of the five proteases used (data not shown). The beads were then washed with buffer containing 25 mM imidazole (W), then the bound proteins were eluted with buffer containing 100 mM imidazole (E). The beads were then analyzed by SDS-PAGE on an 8% gel (NOVEX). His6-α was omitted from the experiments shown in lanes 5 to 8 and lanes 17 to 20 as a control for non-specific binding of β or its fragments to the Ni²⁺-NTA-agarose beads. b, Ni²⁺-co-immobilization assay with His6-α renatured with molar excesses of βABCD and βEFGH [C10T988] simultaneously. Since βABCD and βEFGH are nearly the same size and co-migrate on the SDS-PAGE system used, βEFGH[C10T988], with a 127 residue insertion after amino acid 988 (Kashlev et al., 1989; Borukhov et al., 1991), and thus with a lower mobility, was used. In this assay, βEFGH[C10T988] was added to the renaturation mix from a crude inclusion body preparation, explaining the presence of numerous contaminating bands in the load and flow-through fractions.

Figure 1. Binding of full length β and βEFGH, but not βABCD, to immobilized His6-α by the Ni²⁺-co-immobilization assay. a, His6-α was renatured with a molar excess of full length β (lanes 1 to 4), βABCD (lanes 9 to 12), or βEFGH (lanes 13 to 16), loaded onto Ni²⁺-NTA-agarose beads (L) and the unbound protein (or flow-through) removed (F). The beads were then washed with buffer containing 25 mM imidazole (W), then the bound proteins were eluted with buffer containing 100 mM imidazole (E). The protein fractions were then analyzed by SDS-PAGE on an 8% gel (NOVEX). His6-α was omitted from the experiments shown in lanes 5 to 8 and lanes 17 to 20 as a control for non-specific binding of β or its fragments to the Ni²⁺-NTA-agarose beads. b, Ni²⁺-co-immobilization assay with His6-α renatured with molar excesses of βABCD and βEFGH [C10T988] simultaneously. Since βABCD and βEFGH are nearly the same size and co-migrate on the SDS-PAGE system used, βEFGH[C10T988], with a 127 residue insertion after amino acid 988 (Kashlev et al., 1989; Borukhov et al., 1991), and thus with a lower mobility, was used. In this assay, βEFGH[C10T988] was added to the renaturation mix from a crude inclusion body preparation, explaining the presence of numerous contaminating bands in the load and flow-through fractions.
Figure 2. Limited proteolysis of Ni\textsuperscript{2+}-NTA-agarose-immobilized (His\textsubscript{6}-\(\alpha\)NTD)\(\beta\)EFGH\(I\) analyzed by SDS-PAGE on 4% to 20% gradient gels (Novex). In each gel, the lane labeled U is untreated with protease. The ratios across the top of the other lanes indicate the mass ratio of protease to (His\textsubscript{6}-\(\alpha\)NTD)\(\beta\)EFGH\(I\). The positions of molecular mass markers (Sigma) are indicated to the left of each gel. 

a, Limited proteolysis using chymotrypsin. b, Limited proteolysis using proteinase K. c, Limited proteolysis using trypsin. d, Summary of proteolysis results and mapping of \(\beta\) fragments using N-terminal sequencing and MALDI-MS. The horizontal black bar represents the primary sequence of \(\beta\)EFGH\(I\) with amino acid numbering corresponding to full length \(\beta\) shown above the bar. Evolutionarily conserved regions are shaded grey and labeled E to I below the bar according to Sweetser \textit{et al.} (1987). The open box represents a region containing large deletions found in several chloroplast RNAP \(\beta\) homologs (Hudson \textit{et al.}, 1988). Deletions encompassing dispensable region II (DRII; Borukhov \textit{et al.}, 1991) are represented as bars above the primary structure. The locations of Lys\textsubscript{1065} and His\textsubscript{1237}, which cross-link to initiating nucleotide analogs (Mustaev \textit{et al.}, 1991), are indicated (*). Below the primary structure are the results of mapping the N and C termini of the major protease-resistant fragments of \(\beta\) resulting from limited proteolysis of (His\textsubscript{6}-\(\alpha\)NTD)\(\beta\)EFGH\(I\). Cleavage sites that were mapped to within one residue are indicated by vertical arrows, sites that were mapped to within a range of residues are indicated by a line spanning that range. The labeled horizontal lines below illustrate the identified fragments in the context of the primary structure.
chymotrypsin yielded a band with a mobility corresponding to about 60 kDa (Figure 2a, ct-I). As the chymotrypsin concentration was increased, a second stable product accumulated with a mobility corresponding to about 35 kDa (Figure 2a, ct-II). N-terminal sequencing of these two products revealed that each contained a major and a minor component. The band labeled ct-I contained proteins with N termini at β residue 700 (major component) and 681 (minor). The band labeled ctII contained proteins with N termini corresponding to 907 (major) and 903 (minor). All of these N termini are consistent with the cleavage specificity of chymotrypsin. Next, the following experiment was performed. (1) (His₆-NTD)₅PEFGH complex, immobilized on Ni²⁺-NTA agarose, was subjected to chymotrypsin cleavage at a ratio of chymotrypsin:complex of 1:300 (w/w). At this concentration of chymotrypsin, significant amounts of both ct-I and ct-II were generated (Figure 2a). (2) The beads were washed to remove unbound proteins. (3) The beads were then suspended in buffer containing 6 M guanidine-HCl to denature and elute the remaining β fragments from the immobilized His₆-NTD. (4) The eluted β fragments were then subjected to matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS; Hillenkamp et al., 1991) to accurately measure their masses.

From the N-terminal sequences, the measured masses, and with consideration of the cleavage specificity of chymotrypsin, the C termini of the proteolytic fragments could be roughly inferred. The N-terminal sequencing and MALDI-MS data that led to the identification of the β fragments are tabulated in Table 2.

Limited proteolysis with trypsin, and analysis of the bound peptides by SDS-PAGE, yielded one major, stable product from PEFGHI (t-I, Figure 2c) with a mobility corresponding to about 37 kDa. N-terminal sequencing of the product labeled t-I revealed that it contained one component with N terminus at β residue 904, consistent with the cleavage specificity of trypsin and also with an earlier study of RNAP trypsinolysis (Borukhov et al., 1991). The products (predominantly t-I) resulting from limited proteolysis at a ratio of trypsin:complex of 1:5000 (w/w) were analyzed by MALDI-MS, allowing the C terminus to be determined (Table 2). The results of this analysis of the domain organization of PEFGHI in complex with α are schematically summarized in Figure 2d.

### Table 2. Identification of β subunit fragments generated by limited proteolysis

<table>
<thead>
<tr>
<th>Protein</th>
<th>Conserved regions</th>
<th>N-terminal sequence</th>
<th>Observed mass (kDa)</th>
<th>N</th>
<th>β Residues</th>
<th>C</th>
<th>Calculated mass (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>β₁₂₅ᵦ</td>
<td>FGH₆₅</td>
<td>XXANMQRQAVP</td>
<td>61.5</td>
<td>681</td>
<td>1234-1241</td>
<td>61.2-61.9</td>
<td></td>
</tr>
<tr>
<td>β₁₂₅ᵦ</td>
<td>FGH₆₅</td>
<td>VGTGMRRAV</td>
<td>61.5</td>
<td>700</td>
<td>1253-1261</td>
<td>61.6-61.9</td>
<td></td>
</tr>
<tr>
<td>β₁₂₅ᵦ</td>
<td>FGH</td>
<td>VAVDSGVTVAV</td>
<td>44.7</td>
<td>708</td>
<td>1110-1116</td>
<td>44.3-45.0</td>
<td></td>
</tr>
<tr>
<td>β₁₂₅ᵦ</td>
<td>FGH</td>
<td>DSGVTAYAKR</td>
<td>44.7</td>
<td>711</td>
<td>1113-1119</td>
<td>44.3-45.0</td>
<td></td>
</tr>
<tr>
<td>β₁₂₅ᵦ</td>
<td>H₁₅</td>
<td>XIAGFKES</td>
<td>38.9</td>
<td>903</td>
<td>1245-1252</td>
<td>38.5-39.3</td>
<td></td>
</tr>
<tr>
<td>β₁₂₅ᵦ</td>
<td>H₁₅</td>
<td>AIFKGKADY</td>
<td>38.8</td>
<td>904</td>
<td>1246</td>
<td>38.5</td>
<td></td>
</tr>
<tr>
<td>β₁₂₅ᵦ</td>
<td>H₁₅</td>
<td>XEKA SDVKDD</td>
<td>38.9</td>
<td>907</td>
<td>1249-1256</td>
<td>38.5-39.2</td>
<td></td>
</tr>
</tbody>
</table>
After overexpression, about 50% of $\beta_{FGHIN}$ was found in the cytosol, the other 50% in inclusion bodies. All of the other fragments were found almost totally in inclusion bodies and so all of the reconstitutions were done from inclusion body preparations. The binding of each fragment to $\alpha$ was investigated using Ni$^{2+}$-co-immobilization assays with His$_{6}$-$\alpha$NTD, and in side-by-side reactions with $\alpha$NTD as a control for non-specific binding of the $\beta$ fragment to the Ni$^{2+}$-NTA-agarose beads. Load (L), flow-through (F), wash (W), and elution (E) fractions were analyzed by SDS-PAGE on 8 to 25% gradient PhastGels (Pharmacia).

**Figure 3.** Binding of $\beta$ fragments to $\alpha$NTD. Ni$^{2+}$-co-immobilization assays of $\beta$ fragment binding to either His$_{6}$-$\alpha$NTD (lanes 1 to 4, 9 to 12, 17 to 20, 25 to 28, and 33 to 36) or $\alpha$NTD (lanes 5 to 8, 13 to 16, 21 to 24, 29 to 32, and 37 to 40), as a control for non-specific binding of the $\beta$ fragment to the Ni$^{2+}$-NTA-agarose beads. Load (L), Flow-through (F), wash (W), and elution (E) were analyzed by SDS-PAGE on 8 to 25% gradient PhastGels (Pharmacia).

RNA Polymerase Assembly

<table>
<thead>
<tr>
<th>$\beta$ fragment</th>
<th>$\alpha$ His-tag</th>
<th>$\beta_{FGHI}$</th>
<th>$\beta_{FGHI-N}$</th>
<th>$\beta_{FGH}$</th>
<th>$\beta_{FGH-N}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>fraction</td>
<td></td>
<td>L F W E L F W E</td>
<td>L F W E L F W E</td>
<td>L F W E L F W E</td>
<td>L F W E L F W E</td>
</tr>
<tr>
<td>$\beta_{FGHI}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{FGHI-N}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$(NTD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{FGH}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{FGH-N}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$(NTD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta$ fragment</th>
<th>$\alpha$ His-tag</th>
<th>$\beta_{HI}$</th>
<th>$\beta_{HN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>fraction</td>
<td></td>
<td>L F W E L F W E</td>
<td>L F W E L F W E</td>
</tr>
<tr>
<td>$\beta_{HI}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{HN}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$(NTD)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Binding of $\beta$ fragments to $\alpha$NTD. Ni$^{2+}$-co-immobilization assays of $\beta$ fragment binding to either His$_{6}$-$\alpha$NTD (lanes 1 to 4, 9 to 12, 17 to 20, 25 to 28, and 33 to 36) or $\alpha$NTD (lanes 5 to 8, 13 to 16, 21 to 24, 29 to 32, and 37 to 40), as a control for non-specific binding of the $\beta$ fragment to the Ni$^{2+}$-NTA-agarose beads. Load (L), Flow-through (F), wash (W), and elution (E) were analyzed by SDS-PAGE on 8 to 25% gradient PhastGels (Pharmacia).
To address whether β conserved region H is required for α interaction, we investigated the binding of two additional β fragments, β_{EFGH} and β_I (Table 1 and Figure 4). Neither fragment bound α in the Ni^{2+}-co-immobilization assay (data not shown). Thus, we conclude that amino acid residues in both β conserved regions H and I_N are required for α binding.

**Specificity of β fragment interactions with α**

We used three additional tests to investigate the specificity and authenticity of the observed inter-
actions between the α subunit and β fragments; protein footprinting, binding to an α point mutant defective in β binding, and competition between full length β and its fragments.

Protein footprinting

Hydroxyl-radical protein footprinting was recently used to show that residues within α conserved regions A and B (specifically α residues 30 to 55 and 65 to 75, respectively) are involved in direct interactions with the β subunit in the αβ sub-assembly, while residues within α conserved regions C and D (specifically α residues 175 to 185 and 195 to 210, respectively) are involved in direct interactions with the β subunit (Heyduk et al., 1996). We utilized the same footprinting method to ask if the interactions observed between α and β fragments containing conserved regions H and Iα were identical to the interactions between α and full length β.

For this work, we used an α derivative with an N-terminal heart muscle protein kinase recognition site (Li et al., 1989). With the purified, 35P-end-labelled α derivative (α*), complexes were formed with excess β or β fragments and purified to homogeneity. For each complex, we then performed hydroxyl-radical-mediated cleavage and compared the cleavage pattern quantitatively (Heyduk et al., 1996) with α* alone (Figure 5). Separate, independent cleavage experiments for α* alone were compared to determine the level of background noise (Figure 5a). In addition to the analysis of Heyduk et al. (1996), we tested the statistical significance of the differences between the (α*+α*) control experiments and the (α*β-α*), (α*βGH-α*), and (α*βGHIL-α*) experiments using a Student’s t-test for small samples with unequal variances (Mendenhall & Sincich, 1988). Regions where statistically significant differences from the (α*+α*) control were observed at a confidence level of 0.99999% are shown as black bars just above the x-axis in each plot of Figure 5. There were no statistically significant differences between two independent data sets of (α*+α*).

Figure 5b compares the cleavage pattern of the α*β complex to that of α* alone. Negative values of the curve represent protection of α from cleavage by the binding of β. The characteristics of the difference plot resemble qualitatively the difference plot of Heyduk et al. (1996), demonstrating the reproducibility of the method. Figure 5c and d compares cleavage patterns of the α*βGHIL and α*βGHIL complexes, respectively, to that of α* alone. All three plots (Figure 5 to c) show qualitatively and quantitatively (in terms of statistically significant differences) the same protection pattern. For all three experiments, statistically significant differences occurred within two broad peaks of protection centered around α amino acids 43 to 52 and 61 to 76, corresponding with the earlier work of Heyduk et al. (1996). In addition, weaker but statistically significant protection was also observed in a region around residues 159 to 176. These results demonstrate that all of the contacts made by the β subunit on α are also made identically by βGHIL and βGHIL.

We attempted to determine footprints of βHI and βHI on α but were unable to obtain reproducible difference plots. We believe this is due to the relatively weak binding of these shorter β derivatives to α (Figure 3). After the formation of complexes in the presence of a large excess of βHI several steps of purification were required to obtain homogeneous complexes free of excess βHIS and its fragments, and free α*H*, resulting in disassociation and very low yields of complex.

The effect of an α point mutant defective in β binding

Kimura & Ishihama (1995a,b) described a mutant α subunit with a single amino acid substitution, Arg45 to Ala (αR45A), that was properly folded and dimerized but was defective in binding β to form the assembly intermediate αββ. If the interactions of the β fragments with α are authentic, the β fragment interactions with the αR45A mutant should also be negatively affected. This was tested using Ni2+-co-immobilization assays with Hisα-αR45A-NTD and β or β fragments containing conserved regions H and Iα (Figure 6).

Full length β bound to Hisα-αR45A-NTD in the Ni2+-co-immobilization assay (Figure 6, lane 2), but the binding was weak compared to wild-type (wt) αNTD (Figure 1, lane 4). Estimating from the Coomassie stained bands, there was about a fivefold reduction in the amount of β retained (relative to α) when Hisα-αR45A-NTD was used rather than wt Hisα-αNTD.

The binding of the β fragments βEFGHI, βGHIL, and βGHIL to the αR45A mutant was undetectable by the Ni2+-co-immobilization assay (Figure 6, lanes 4, 6, and 8). In contrast, the shorter β fragments βHI (Figure 6, lane 10), β950-1342 (Figure 6, lane 12), and βHI (not shown) were co-immobilized similarly by Hisα-αR45A-NTD and wt Hisα-αNTD. We attempted to titrate down the molar ratio of β fragment:α in the renaturation mixture (normally 4:1) but this did not reveal any obvious difference in binding affinity between the smallest β fragments and the αR45A or wt α.

Competition between full length β and its fragments for α binding

Only one β subunit interacts with an α dimer in the RNAP assembly. We tested the ability of full length β to bind competitively with its fragments to α. In all cases tested, β fragments containing conserved regions H and Iα could not bind α simultaneously with full length β. For example, equimolar amounts of Hisα-αNTD and βHI were renatured in the presence of increasing amounts of either full length β or βGHIL, and binding to the Hisα-αNTD was then tested by the Ni2+-co-immobil-
Figure 5. Difference plots showing hydroxyl-radical protein footprinting results. Normalized difference intensity, \((I - I_{2*})/I\), is plotted against residue number, where \(I\) is the corrected intensity for the complex under study, and \(I_{2*}\) is the corrected intensity for \(2^*\). Statistically significant differences according to a Student's t-test (confidence level of 0.99999) are denoted by the thick black bars just above the x-axis. a, \(\alpha_2^* - \alpha_2^*\) (averaged data for 4 lanes); b, \(\alpha_2^* - \alpha_2^*\) (averaged data for 9 lanes); c, \(\alpha_2^* - \alpha_2^*\) (averaged data for 17 lanes); d, \(\alpha_2^* - \alpha_2^*\) (averaged data for 11 lanes).
Recruitment of $\beta'$ to complexes between $\beta$ fragments and $\alpha$

Mutational analysis and hydroxyl-radical protein footprinting have shown that $\alpha$ conserved regions C and D make direct interactions with the $\beta'$ subunit that are critical for core RNAP assembly (Kimura et al., 1994; Kimura & Ishihama, 1995a,b; Heyduk et al., 1996). Nevertheless, in the absence of $\beta$, $\beta'$ does not interact with $\alpha$, as shown by the Ni$^{2+}$-co-immobilization assay (Figure 8, lane 4). Thus, both the $\alpha$ and $\beta$ subunits make critical interactions with $\beta'$ to form core RNAP (Figure 8, lane 8). Interestingly, we found that the complex of His$_{8}\alpha$NTD with $\beta_{\text{FGHI}}$ was sufficient to recruit $\beta'$ (Figure 8, lane 12). We used the Ni$^{2+}$-co-immobilization assay with His$_{8}\alpha$NTD and the various $\beta$ fragments to determine the conserved regions of $\beta$ required for recruitment of $\beta'$ into their complex with $\alpha$. Since the formation of the $\alpha$-$\beta$ complex is obligatory, we only investigated the $\beta$ fragments containing, at a minimum, conserved regions H and I$_{3}$ since, as determined above, only these $\beta$ fragments could interact with $\alpha$. While all of the $\beta'$ binding results are summarized schematically in Figure 4, the most revealing experiments are shown in Figure 8, lanes 13 to 20. When a mixture of His$_{8}\alpha$NTD, $\beta'$, and $\beta_{\text{FGHI}}$ was renatured and subjected to the Ni$^{2+}$-co-immobilization assay (Figure 5, lanes 13 to 16), the complex between His$_{8}\alpha$NTD and $\beta_{\text{FGHI}}$ was formed, as expected, but this complex was unable to recruit $\beta'$ (Figure 5, lane 16). However, renaturation of a mixture of His$_{8}$NTD, $\beta'$, and $\beta_{\text{FGHI}}$ resulted in the formation of the ternary complex (Figure 5, lane 20), indicating that the C-terminal half of $\beta$ (EFGHI) alone interacts with $\alpha$ while the N-terminal half (EFGHI) does not (Figure 1), suggesting that this assay could be conveniently used to parse the $\beta$ subunit into functional domains that interact with $\alpha$ to initiate RNAP assembly. To further divide $\beta$ into smaller fragments that were functionally relevant, we did not want to rely completely on the locations of the evolutionarily-conserved regions since structural domains may not necessarily correspond to the regions of conserved sequence (Severinova et al., 1996). Therefore, we used limited
proteolysis to define domains of $\beta_{\text{FGHI}}$ complexed with $\alpha$. Of most significance from the proteolysis studies was the high degree of clustering of all but one of the N and C-terminal cleavage sites (Figure 2d and Table 2) for three different proteases, indicating that the locations of these sites contain information about the domain architecture of $\beta$ rather than the properties of the individual proteases. Two clusters of N-terminal cleavage sites were observed. The first N-terminal cluster contained four cleavage sites (681, 700, 708, and 711) from two different proteases (chymotrypsin and proteinase K), all within 30 residues of each other, just C-terminal of $\beta$ conserved region E. The second N-terminal cluster contained three cleavage sites (903, 904, and 907) from two different proteases (chymotrypsin and trypsin), all within four residues of each other, just C-terminal of conserved region G. A cluster of C-terminal cleavage sites occurred with two proteases (chymotrypsin and trypsin) within a region of 28 residues (1234-1261) in the middle of conserved region I, again indicating that structural domains as defined by limited proteolysis may not necessarily correspond to domains defined by sequence analysis.

### $\beta$ Conserved regions H and I$_N$ are required for the interaction with $\alpha$

Using domains of $\beta$ defined by the proteolysis study as well as additional $\beta$ fragments (Table 1), we used the Ni$^{2+}$-co-immobilization assay to investigate $\alpha$ binding. The results are summarized in Figure 4. The different $\beta$ fragments grouped qualitatively into three classes; those that bound $\alpha$ approximately stoichiometrically (based on the Coomassie stained gels, denoted ++ in Figure 4), those that bound $\alpha$ but obviously less tightly (+), and those that did not detectably bind $\alpha$ (−). Only $\beta$ fragments containing both conserved regions H and the N-terminal half of region I (I$_N$) interacted with $\alpha$ and the results in total allow us to conclude that $\beta$ conserved regions H and I$_N$ are required for the interaction of $\beta$ with $\alpha$. This is the main conclusion from this study. While conserved regions F and G are not required for $\alpha$ binding, they appear to play some role in stabilizing the interaction since the only difference between the strong binding $\beta$ fragments (++) and those that bound weakly (+) was the presence of regions F and G.

While we have shown that active RNAP can be reconstituted from peptides comprising the $\beta$ subunit split into at least four separate fragments (Severinov et al., 1995a, 1996), the $\beta$ subunit split into two separate peptides at a site between conserved regions H and I will not support the assembly of functional enzyme (K. S., A. Vasenko, I. Bass, S. A. D., unpublished results). The finding that both regions H and I$_N$ of $\beta$ are required for $\alpha$ interaction explains the inability of the $\beta$ polypeptide split between the two regions to assemble into functional enzyme since the assembly process would not be able to initiate.

Because almost all of the $\beta$ fragments used in this study were relatively insoluble and tended to aggregate and precipitate on their own in solution, the $\alpha$-$\beta$ fragment complexes were renatured by dialysis from the denatured state, leading to the possibility of artifacts due to non-specific aggregation effects. Therefore, we performed a number of experiments, in addition to Ni$^{2+}$-co-immobilization, to test the specificity and authenticity of the observed interactions between $\beta$ fragments containing conserved regions H and I$_N$ and $\alpha$. The most stringent test of specific and authentic binding was protein footprinting, which was used to show that the fragments $\beta_{\text{FGH}}$ and $\beta_{\text{FGHI}}$ bound to $\alpha$ and protected the same $\alpha$ residues from hydroxyl-radical mediated cleavage as full length $\beta$ (Figure 5). This unequivocally verifies that the negative results obtained in the Ni$^{2+}$-co-immobilization assay for some of the $\beta$ fragments are not artifacts of the assay and demonstrate that these regions of $\beta$ do not interact with $\alpha$, at least not in a way that is detectable by two very different assays, the Ni$^{2+}$-co-immobilization or by hydroxyl-radical protein footprinting. In addition, these two $\beta$ fragments ($\beta_{\text{FGH}}$ and $\beta_{\text{FGHI}}$): (1) did not bind an $\alpha$ mutant with a point substitution (R45A) known to be specifically compromised in its ability to bind $\beta$ (Figure 6), and (2) bound $\alpha$ competitively with full

![Figure 8](image-url)
length $\beta$. Thus, we conclude that the fragments $\beta_{\text{pGBH}}$ and $\beta_{\text{pGBH}}$ bound $\alpha$ in a highly specific manner and using the same interactions as full length $\beta$ in forming the assembly intermediate $\alpha_2\beta$.

We were unable to observe a reproducible protein footprint on $\alpha$ from the binding of the shorter $\beta$ fragments $\beta_{\text{HI}}$ and $\beta_{\text{HIL}}$, probably due to their weaker binding affinity. Also, the shorter $\beta$ fragments bound the $\alpha_{\text{RGSH}}$ mutant similarly to wt $\alpha$. It should be noted, however, that in the Ni$^{2+}$-co-immobilization assay, the $\alpha_{\text{RGSH}}$ mutant is not completely defective in binding full length $\beta$ (Figure 6), and under conditions that drive in vitro RNAP assembly (Tang et al., 1995), the $\alpha_{\text{RGSH}}$ mutant is equally effective as wt $\alpha$ in reconstituting active RNAP as measured by an affinity-labeling assay (Grachev et al., 1987; Severinov et al., 1995a; Y.W., K.S., A. Mustaev, & S.A.D., unpublished results), suggesting that the $\alpha_{\text{RGSH}}$ mutant may not be an appropriate test of $\alpha$-$\beta$ interactions in vitro. These results are not inconsistent with the data presented by Kimura & Ishihama (1995a,b). In favor of the specificity of $\beta_{\text{HI}}$ binding to $\alpha$, $\beta_{\text{HI}}$ bound $\alpha$ competitively with full length $\beta$, indicating that they interact with the same site on $\alpha$. In addition, the $\beta$ subunit, when split into two separate peptides between conserved regions G and H, remains competent to assemble active RNAP (Severinov et al., 1995a), indicating that $\beta$ fragments containing only regions H and I are competent for initiating RNAP assembly.

Functions for $\beta$ conserved regions H and I$_N$ are not known but of note are two sites, Lys1065 within region H, and His1237 within I$_N$, that cross-link to initiating nucleotide analogs (Mustaev et al., 1991), placing these residues within about 3 Å of the initiating nucleotide $\alpha$-phosphate. Furthermore, protein-DNA cross-linking has implicated a region of $\beta$ between Met1230 and Met1273 within region I as participating in critical protein-DNA interactions near the transcript 3′-end (Nudler et al., 1996). Thus, residues within $\beta$ conserved regions H and I$_N$ are positioned very close to, and appear to contribute essential components to the enzymes active center. Since there is no evidence that $\alpha$NTD is located near the DNA template or RNA transcript, nor that it plays a role in the RNAP catalytic function, it is surprising to find that RNAP assembly involves interactions between $\alpha$NTD and $\beta$ conserved regions H and I$_N$.

**$\beta$ Conserved region I$_C$ appears to be involved in recruiting $\beta'$ into the $\alpha$-$\beta$ complex**

An additional observation that $\beta_{\text{EFGHI}}$ was sufficient to recruit the $\beta'$ subunit into its complex with $\alpha$ (Figure 8) led us to take the Ni$^{2+}$-co-immobilization assay one step further to investigate which $\beta$ fragments, once bound to $\alpha$, were able to recruit $\beta'$. The results with some of the fragments clearly indicated that, in addition to the $\beta$ regions previously identified as being required for $\alpha$ binding, the C-terminal half of $\beta$ conserved region I ($I_C$) was required to bind $\beta'$. So, for instance, $\beta_{\text{EFGH}}$ and $\beta_{\text{GBH}}$, both bound $\alpha$ quite strongly, but only $\beta_{\text{EFGH}}$ was able to recruit $\beta'$ into the complex (Figure 8). However, $\beta_{\text{HI}}$ and $\beta_{\text{HIL}}$, both bound $\alpha$ and both appeared able to recruit $\beta'$ into the complex, clouding this interpretation. Indirect support for the conclusion that a direct interaction occurs between $\beta$ conserved region I$_C$ and $\beta'$ comes from genetic suppressor studies. Experiments designed to find interacting regions within the $\beta$ subunit failed to find intragenic suppressor mutations of three different alleles within conserved region I$_C$ (substitutions at $\beta$ positions 1249, 1266, and 1272; Tavormina et al., 1996). In S. cerevisiae RNA polymerase II, a conditional mutation resulting from substitution of a highly conserved Gly residue (corresponding to position 1282 of E. coli $\beta$) within conserved region I$_C$ of the $\beta$ homolog was suppressed by mutations in the $\beta'$ homolog (Martin et al., 1990; Scafe et al., 1990). These results together indirectly support the possibility that region I$_C$ of $\beta$ may interact with regions of $\beta'$.

An emerging view of RNA polymerase assembly

The results of this study, combined with results from many other studies (Igarashi et al., 1990; Hayward et al., 1991; Igarashi et al., 1991; Igarashi & Ishihama, 1991; Blatter et al., 1994; Kimura et al., 1994; Kimura & Ishihama, 1995a,b; Severinov et al., 1995a,b; Heyduk et al., 1996) provide support for the following sequence of events along the RNAP assembly pathway (Zillig et al., 1976; Ishihama, 1981). (1) Assembly is initiated by the dimerization of the $\alpha$ subunits, which is mediated by the $\beta$NTD. (2) The assembly intermediate $\alpha_2\beta$ is formed by interactions between $\alpha$ conserved regions A and B and $\beta$ conserved regions H and I$_N$. The complex is stabilized by regions F and G of $\beta$, probably through interactions with regions H and I. (3) $\beta'$ is recruited into the $\alpha_2\beta$ complex through interactions with conserved regions in both $\alpha$ (regions C and D) and $\beta$ (including region I$_C$). (4) The condensation of further interactions between conserved regions in $\beta$ and $\beta'$ (or between intrasubunit interactions), which are not currently known, results in the formation of core RNAP. (5) RNAP holoenzyme subsequently forms by the addition of the $\sigma$ subunit through unknown interactions but which do not involve the $\alpha$ subunit.

The $\beta$ subunit homolog in archaeabacteria is encoded by two separate genes (Berghofer et al., 1988; Pühler et al., 1989). Conserved regions A to I are arranged co-linearly with respect to E. coli $\beta$ except the split site (corresponding to approximately codon 650 of E. coli rpoB) separates conserved regions A to D ($\beta_N$) from E to I ($\beta_C$) on two separate peptides. Our results also suggest that in archaeabacteria, assembly of RNAP involves the initial interaction of $\beta_C$ with $\alpha_2$ prior to the recruitment of $\beta'$. Subsequently, $\beta_N$ would enter the complex to form the core RNAP.
Further studies investigating the interactions of other large subunit domains (Severinov et al., 1995a,b, 1996) with each other and with sub-assemblies of other RNAP subunit domains will help elucidate additional subunit-subunit interactions that occur in the RNAP assembly.

Materials and Methods

Proteins

α Derivatives

Plasmids expressing His₆-α, αNTD, His₆-αNTD (Tang et al., 1995), and His₆-αNTD with an N-terminal calf heart protein kinase site (Heyduk et al., 1996) were generous gifts from R. Ebright (Waksman Institute, Rutgers University). The proteins were expressed and purified as described (Tang et al., 1995).

β Derivatives

Full length β and β' were obtained from expression plasmids pMKSe2 (Severinov et al., 1993) and T7β' (Zalenskaya et al., 1990), respectively. Construction of expression plasmids for β₁₄, βCDEFG, and β₁₉₅₀-₁₃₄₂ was described (Severinov et al., 1995a), as were expression plasmids for β₁₄CD and β₁₄FGH (Severinov et al., 1996). β₁₄FGH[ω786] was cloned from rpoB containing the ω786 insertion (Kashlev et al., 1989; Borukhov et al., 1991) into the plasmid expressing β₁₄CD. The β fragments β₁₄CD, β₁₄FGH, β₁₄FGH, β₁₄CD, and β₁₄FGH were obtained from expression plasmids constructed by PCR cloning from pMKSe2 into NcoI and BamHI sites of the T7-based pET15b vector (Novagen). Use of the ω₃-NcoI site results in loss of the vector-based N-terminal His₆-tag; the resulting protein products contain only native β sequences. β was cloned from pMKSe2. β₁₄FGH(A[910-982]) and β₁₄FGH(A[946-1062]) were cloned from rpoB mutants harboring the deletions (constructed as described by Borukhov et al., 1991) into the plasmid expressing β₁₄FGH.

For all of the β fragments, overexpression was induced with 1 mM isopropyl-β-D-thiogalactoside at 37°C for three hours. The proteins in inclusion bodies were then prepared according to Tang et al. (1995).

Reconstitution of α-β complexes

Purified α or αNTD derivatives were mixed with β or its fragments (from washed inclusion bodies) at a molar ratio of 2:1 and a total protein concentration of not more than 0.5 mg/ml in reconstitution buffer (40 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM EDTA, 20 mM MgCl₂, 10 μM ZnCl₂, 20% (v/v) glycerol, 7 mM β-mercaptoethanol, 6 M guanidine-HCl) and dialyzed overnight at 4°C against 2 x 2 l of the same buffer without guanidine-HCl.

Limited proteolytic digestion

Complexes of His₆-αNTD with β fragments (about 10 μg total protein) were immobilized on Ni²⁺-NTAagarose beads (Qiagen) by incubation in 50 μl of buffer A (20 mM Tris-HCl, pH 7.9, 100 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol, 5% (v/v) glycerol) + 2.5 mM imidazole and incubated for 30 minutes at 4°C with gentle mixing. Protease was then added and digestion reactions were carried out at 25°C with gentle mixing. After 30 minutes, phenylmethylsulfonyl fluoride was added to a final concentration of 1 mM, the beads were pelleted by centrifugation, washed four times with 500 μl of buffer A, and eluted with 10 μl of buffer A + 100 mM imidazole. The eluted products were then analyzed by SDS-PAGE using 4 to 20% gradient gels (Novex).

Matrix-assisted laser desorption mass spectrometry

Matrix-assisted laser desorption mass spectra (Hillenkamp et al., 1991) of the proteolytic fragments of β were collected using a time-of-flight mass spectrometer constructed at The Rockefeller University (Beavis & Chait, 1989). The fragments were mixed with z-cyano-4-hydroxycinnamic acid (10 g/l in formic acid/water/isopropanol (1:3:2, by vol.) to obtain a final protein concentration of about 2 pmol/μl. An aliquot (0.5 μl) was placed on the mass spectrometer probe tip and air dried. The sample was irradiated with 10 ns duration laser pulses (355 nm wavelength) from a Nd(YAG) laser. The resulting ions were accelerated in an electrostatic field and their time-of-flight was measured with a LeCroy 9350 M oscilloscope. The observed masses are listed in Table 2.

Ni²⁺-NTA agarose co-immobilization binding assays

Protein complexes (about 10 μg) were mixed with pre-equilibrated Ni²⁺-NTA-agarose beads (Qiagen) in 50 μl buffer A and incubated for 30 minutes at 4°C with gentle mixing. The beads were pelleted by centrifugation and washed twice with 500 μl of buffer A, 2.5 mM imidazole, then twice with 500 μl of buffer A, 25 mM imidazole. The protein samples were then eluted from the beads with buffer A, 100 mM imidazole and analyzed by SDS-PAGE using 8 to 25% gradient PhastGels (Pharmacia).

Hydroxy-radical protein footprinting

Purified α derivative was ³²P-end-labelled in a 230 μl reaction containing 90 μM α derivative, 90 units bovine heart protein kinase (Sigma), 0.1 μM [γ³²P]ATP (New England Nuclear, 7-110 TBq/mmol), 10 mM Tris-HCl (pH 7.9), 50 mM NaCl, 0.5 mM EDTA, 25% (v/v) glycerol, and 30 mM DTT for one hour at 37°C. Complexes (αβ, αβFGH, and αβFGH) were prepared essentially as described (Tang et al., 1995). Briefly, labelled α derivative and excess β (or β fragment) was mixed under denaturing conditions. Following dialysis into non-denaturing conditions, complexes were purified by batch-mode metal ion-affinity chromatography, then gel-filtration chromatography on a Superose-6 FPLC column (Pharmacia) as in Borukhov & Goldfarb (1993) except that DTT and glycerol were omitted from the running buffer. Footprint reactions were performed at room temperature in a buffer containing 10 mM Mops-NaOH (pH 7.2), 100 mM KCl, 10 mM MgCl₂, 10 μM z5 or complex, 1 mM (NH₄)Fe(SO₄)₂, 2 mM EDTA, 20 mM ascorbate and 1 mM H₂O₂ for 30 minutes. Reactions were initiated and terminated as described (Heyduk et al., 1996) and the products were analyzed by tricine SDS-PAGE (Schagger & von Jagow, 1987; Heyduk & Heyduk, 1994) followed by phosphorimager analysis (Molecular Dynamics Storm). The gels were quantified and the difference plots were generated essentially as described (Heyduk et al., 1996), as was the calibration of amino acid positions.
using standards generated by CNBr, endoproteinase Lys-C, or endoproteinase Glu-C cleavage.

Acknowledgments

We thank R. Ebright (Waksman Institute, Rutgers University) for plasmids expressing εNTD derivatives and for helpful discussions, G. Meredith (Stanford University) for assistance with the t-test analysis of the protein footprinting results, and E. Campbell for helpful discussions. K.S. is a recipient of The Burroughs Wellcome Fund. Y. Azuma, M. Yamagishi, and A. Ishihama (1993). Subunits of Escherichia coli RNA polymerase II: enzyme purification and structure of the subunit 3 gene. Nucl. Acids Res. 21, 3739–3754.

References


Edited by R. Enright

(Rceived 25 February 1997; received in revised form 25 April 1997; accepted 25 April 1997)