

Substitutions in the *Escherichia coli* RNA polymerase σ^{70} factor that affect recognition of extended -10 elements at promoters

Andrew Sanderson, Jennie E. Mitchell, Stephen D. Minchin, Stephen J.W. Busby*

School of Biosciences, The University of Birmingham, Birmingham B15 2TT, UK

Received 7 March 2003; revised 5 May 2003; accepted 5 May 2003

First published online 14 May 2003

Edited by Ned Mantei

Abstract Previous work has shown that the base sequence of the DNA segment immediately upstream of the -10 hexamer at bacterial promoters (the extended -10 element) can make a significant contribution to promoter strength. Guided by recently published structural information, we used alanine scanning and suppression mutagenesis of Region 2.4 and Region 3.0 of the *Escherichia coli* RNA polymerase σ^{70} subunit to identify amino acid sidechains that play a role in recognition of this element. Our study shows that changes in these regions of the σ^{70} subunit can affect the recognition of different extended -10 element sequences.

© 2003 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Bacterial transcription initiation; RNA polymerase; σ^{70} subunit; Promoter recognition; Extended -10 promoters

1. Introduction

It is well established that the two principal DNA sequence elements for bacterial promoter activity are the -10 and -35 hexamer elements and that, during transcription initiation, these are recognised by the σ subunit of holo RNA polymerase (RNAP). Genetic studies identified different conserved regions in bacterial σ factors, and showed that promoter -10 elements are recognised by the conserved Region 2.4, whilst promoter -35 elements are recognised by the conserved Region 4.2 (reviewed in [1,2]). Recent crystallographic studies with σ factors have led to high resolution structures which show how the different domains of σ are organised with respect to the other RNAP subunits and how they contact different elements at a model promoter ([3–6]; reviewed in [7]). These studies reveal the structure of σ domains 2, 3 and 4 (Fig. 1A). The C-terminal Domain 4, including Region 4.2, which binds directly to the promoter -35 element, is joined to Domain 3 by an extended 30–35 residue linker that ‘threads’ through the RNA exit channel of RNAP. Domain 3 is joined by a short linker to Domain 2, which carries the determinants for -10 element recognition, located in an α helix (equivalent to Region 2.4) right at the C-terminal end of the domain. The structure of Domain 1 was not determined, although insights into its function were obtained from biophysical studies [8,9].

Research from many different laboratories has shown that the DNA segment immediately upstream of promoter -10 hexamers can provide a supplementary element that contributes significantly to promoter strength (reviewed in [10]). The DNA sequence 5' TGn 3', immediately upstream of a promoter -10 element, results in optimum promoter activity, and nearly 20% of *Escherichia coli* promoters contain this element [11]. Suppression genetics with *E. coli* σ^{70} have shown that residue E458 is involved in its recognition [12]. The location of E458 in the first helix of Domain 3 (Region 3.0; previously referred to as Region 2.5) pointed to an involvement of this domain in recognition of extended -10 elements. This was corroborated by studies in which an inorganic DNA cleavage reagent was attached at different positions in Region 3.0 of Domain 3 [13]. The recent crystallographic study of RNAP binding to DNA carrying promoter sequences [5] shows that the α helix at the end of σ Domain 2 (Region 2.4) and the α helix at the start of σ Domain 3 (Region 3.0) wrap around the DNA carrying the promoter -10 element and the extended -10 element. This is possible because of a sharp bend in the short linker between σ Domain 2 and Domain 3 (illustrated in Fig. 1B). Thus, the present evidence points to the involvement of σ Domain 3 in recognition of the extended -10 element, but the precise details of the interaction are unknown. Additionally, the crystallographic study [5] shows that involvement of Domain 2 would be possible. Because of this, we attempted a systematic genetic study of the interactions between *E. coli* σ^{70} and the 5' TGn 3' extended -10 element. To do this, we assayed the effects of alanine substitutions at different locations in Region 2.4 (in Domain 2) and Region 3.0 (in Domain 3) of σ^{70} on transcription initiation at a promoter known to be dependent on its extended -10 element. This led us to the conclusion that residues H455 and E458 in Region 3.0 are the principal residues where alanine substitution reduces recognition of an extended -10 promoter. Alteration of these residues, and assays with promoters in which the 5' TGn 3' motif was changed, confirmed that new promoter preferences could be created. Finally, suppression genetics identified other residues in Region 2.4 of σ^{70} Domain 2 that also contribute to recognition of extended -10 elements.

2. Materials and methods

The *E. coli* K-12 strains used for cloning, manipulation and overexpression of the *rpoD* gene encoding σ^{70} were DH5 α (for routine cloning) and BL21(DE3) (for overexpression) as described by Rhodius and Busby [14]. All assays were performed in the *Alac* CAG20177 strain, whose chromosomal *rpoD* gene is repressed during growth in L broth [15]. Throughout this work, we used plasmid pKB σ carrying

*Corresponding author. Fax: (44)-121-414 7366.

E-mail address: s.j.w.busby@bham.ac.uk (S.J.W. Busby).

the *rpoD* gene as described by Barne et al. [12]. Derivatives of pKB σ encoding σ^{70} with alanine substitutions at different locations, or different substitutions at positions 455 or 458, were constructed by megaprimer PCR mutagenesis; full details are given in [16]. Mutant libraries of pKB σ , carrying random base changes in the segment of the *rpoD* gene bounded by the *Bam*HI and *Xho*I sites, encoding regions 2.4–3.1, were created by error-prone PCR, as before [12].

In this study, we used the model extended -10 KAB-TG promoter, illustrated in Fig. 2. To measure expression from this promoter, an *Eco*RI–*Hind*III fragment carrying KAB-TG was cloned into pRW50, a broad host range low copy number *lac* expression vector plasmid [17]. The resulting construct was transferred into CAG20177 and β -galactosidase expression was measured exactly as before [12] with σ^{70} supplied by different pKB σ derivatives. Each assay was performed independently at least six times. In some figures, absolute β -galactosidase activities are shown in standard Miller units whilst, in others, data are normalised with 100% taken as expression with wild type σ^{70} . Experiments were also performed with pRW50 carrying KAB-TG derivatives with different substitutions at positions -15 and -14 . To do this we exploited the set of 16 KAB-XY promoters described by Burr et al. [11] in which the T and G at positions -15 and -14 (see Fig. 2) are replaced by every possible combination. These pro-

motors are denoted KAB-XY, where XY describes the bases at positions -15 and -14 respectively.

In vitro experiments were performed using purified σ^{70} derivatives prepared using plasmid pET21 derivatives as described by Rhodius and Busby [14]. Segments of *rpoD* encoding mutant σ^{70} derivatives were transferred from pKB σ to pET21HisD [14] on *Sac*II–*Hind*III fragments. RNAP was reconstituted using core enzyme bought from Epicentre Technologies, and transcription experiments were performed with the KAB-TG and KAB-CC promoters cloned in plasmid pSR, as in Burr et al. [11], using protocols detailed by Rhodius et al. [18].

3. Results and discussion

3.1. Alanine scan mutagenesis across the Domain 2/Domain 3 boundary of σ^{70}

The starting point of this work was the plasmid pKB σ that encodes the wild type *E. coli* σ^{70} subunit controlled by a strong constitutive promoter [12]. Using standard recombinant DNA technology, 28 derivatives of pKB σ were made,

Table 1

The table shows measurements of β -galactosidase expression in strain CAG20177 carrying pRW50 derivatives encoding *lac* expressed from the KAB-TG promoter or derivatives carrying different bases at positions -15 and -14 (see Fig. 2)

	GG	GA	GT	GC	AG	AA	AT	AC	TG	TA	TT	TC	CG	CA	CT	CC
H455	434	340	334	117	454	222	206	36	4449	267	45	43	272	134	76	76
HA455	626	479	179	117	580	336	258	37	1324	184	29	26	191	169	74	84
HC455	1086	734	348	151	1039	420	577	52	4258	352	53	53	320	206	126	153
HD455	390	260	200	97	533	316	265	56	4171	299	36	32	289	133	62	74
HE455	361	277	208	75	440	165	255	36	3614	246	34	27	172	162	125	82
HF455	433	245	234	44	435	190	139	29	3116	130	39	39	147	102	41	66
HG455	445	366	401	97	446	193	133	47	4739	223	40	43	183	137	50	78
HI455	1000	1152	461	192	1382	675	594	192	3640	471	39	38	549	372	278	241
HK455	600	484	419	133	437	223	125	41	4232	268	51	40	180	75	52	49
HL455	753	957	411	144	824	517	550	176	4502	488	41	35	463	217	220	148
HM455	450	495	171	87	403	198	163	35	4204	231	46	410	131	72	56	28
HN455	522	357	179	131	448	188	234	40	5065	261	42	51	154	58	71	54
HP455	554	423	300	129	542	267	323	39	4288	261	50	43	179	118	100	53
HQ455	374	317	141	57	384	209	285	35	4174	186	36	37	141	67	96	35
HR455	415	328	410	83	605	198	447	44	4358	209	36	36	233	145	150	196
HS455	2161	1855	470	424	2531	1547	1770	329	3902	1241	59	43	875	533	527	612
HT455	493	453	305	106	502	214	491	36	3805	231	56	40	156	116	96	43
HV455	437	309	264	110	401	207	215	28	3578	189	41	29	138	78	36	35
HW455	422	315	156	48	397	127	197	34	2732	223	38	36	135	76	55	65
HY455	431	398	424	66	589	211	497	38	2907	103	41	30	123	92	53	78
E458	434	340	334	117	454	222	206	36	4449	267	45	43	272	134	76	76
EA458	853	1044	1035	361	692	554	449	144	3599	1191	148	131	683	495	252	153
EC458	346	465	304	255	286	280	187	38	2372	281	64	82	192	157	92	82
ED458	513	329	292	97	292	199	120	58	3022	306	61	59	368	139	79	63
EF458	321	391	380	120	265	242	194	47	4051	347	66	68	230	178	46	88
EG458	256	393	519	118	467	346	396	36	3559	614	138	141	269	206	110	98
EH458	428	395	368	172	429	180	160	45	3853	332	68	69	376	120	104	76
EI458	311	292	366	138	409	220	188	36	3865	337	70	66	242	158	57	92
EK458	424	415	255	164	404	284	152	30	3828	284	60	70	170	165	74	67
EL458	381	734	711	248	425	369	344	58	2208	1168	124	126	449	356	87	83
EM458	281	602	692	239	441	341	236	30	2393	284	103	93	277	238	144	84
EN458	469	716	968	435	421	473	423	115	2590	1349	130	127	400	422	177	160
EP458	483	418	300	189	237	560	389	29	4434	479	75	105	249	92	19	42
EQ458	485	356	355	119	206	295	192	41	2362	469	112	126	226	115	29	48
ER458	459	395	317	194	267	253	184	38	4774	334	82	89	247	193	79	69
ES458	753	917	1167	414	630	438	271	120	2438	768	72	124	708	474	135	140
ET458	736	755	719	314	623	437	389	117	2404	656	58	107	477	342	160	61
EV458	585	821	1060	361	614	462	430	118	1779	616	79	87	703	441	220	110
EW458	408	431	396	162	690	209	283	22	3838	330	73	75	321	158	88	56
EY458	252	512	537	176	512	255	161	38	1798	211	75	106	373	168	85	65

The different derivatives are denoted by the bases as listed at the head of each column. Cells were grown in L broth in conditions where expression from the chromosomal *rpoD* gene was repressed, and σ^{70} was supplied from pKB σ encoding wild type σ^{70} or derivatives encoding σ^{70} with different substitutions of H455 (listed in the upper part of the table) or E458 (listed in the lower part of the table). Each data point (in Miller units) is the average of six experiments. For more than 90% of the data points, the standard deviation was less than 10%. The top line of each part shows activity with wild type σ^{70} with H455 and E458, and the central column in bold face shows activity with the starting KAB-TG promoter.

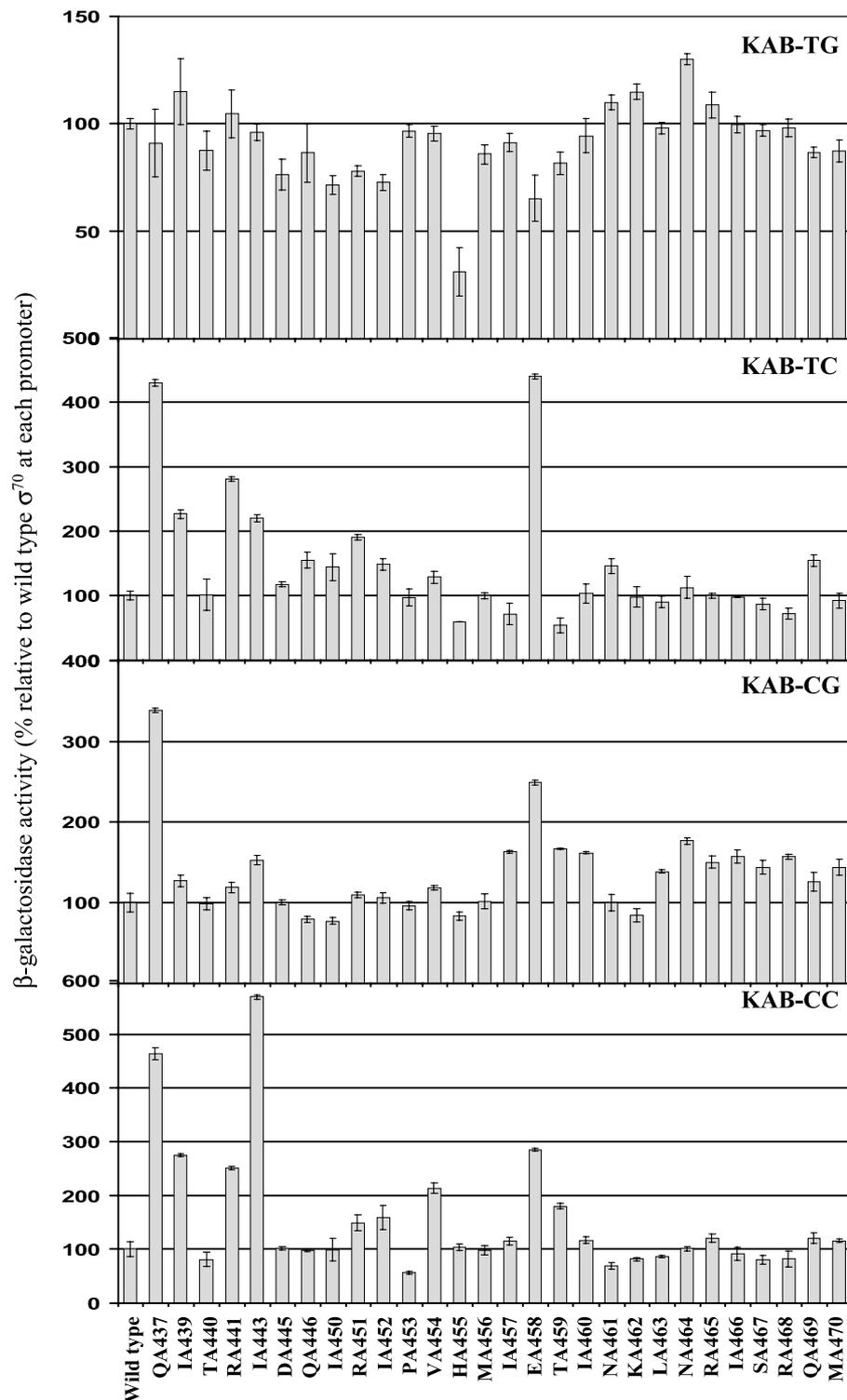


Fig. 3. Alanine scanning of *E. coli* σ^{70} . The bar charts describe measurements of β -galactosidase expression in strain CAG20177 carrying pRW50 derivatives with *lac* fused to the KAB-TG, KAB-TC, KAB-CG or KAB-CC promoters as indicated in each panel, and pKB σ , encoding σ^{70} with alanine substitutions at different locations, as indicated on the x-axis. Cells were grown in L broth in conditions where expression from the chromosomal *rpoD* gene was repressed and σ^{70} was supplied from the pKB σ derivative. Each bar represents the average of at least six independent measurements, with the standard deviation indicated. In each panel, data are normalised to 100%, which is taken as the activity with pKB σ encoding wild type σ^{70} . For the KAB-TG, KAB-TC, KAB-CG and KAB-CC promoters, this corresponds to 4449, 43, 272 and 76 Miller units respectively (Table 1: top data line).

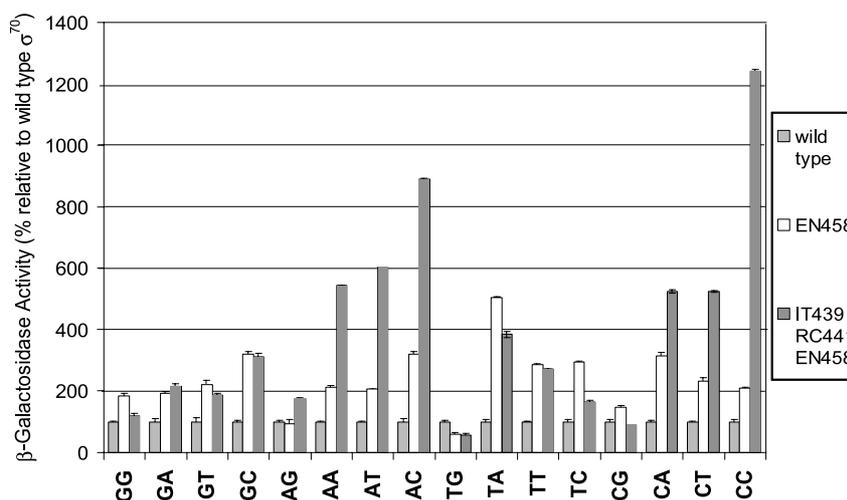


Fig. 4. Activity of a σ^{70} triple mutant. The bar chart represents measurements of β -galactosidase expression in strain CAG20177 carrying pRW50 derivatives encoding *lac* expressed from the KAB-TG promoter or derivatives carrying different bases at positions -15 and -14 (see Fig. 2). Each bar represents the average of at least six independent measurements, with the standard deviation indicated. The different derivatives are denoted by these bases as listed on the *x*-axis. Cells were grown in L broth in conditions where expression from the chromosomal *rpoD* gene was repressed, and σ^{70} was supplied from pKB σ encoding wild type σ^{70} or derivatives encoding σ^{70} with different substitutions as indicated in the inset.

each encoding σ^{70} with a single alanine substitution. These substitutions were located at positions from Q437 in Region 2.4 (towards the C-terminal end of Domain 2) to M470 in Region 3.0 (at the N-terminal end of Domain 3). The different substitutions are listed in Fig. 3 and their locations can be seen in Fig. 1A. To investigate the effects of the substitutions on promoter recognition, the different pKB σ derivatives were introduced into the *E. coli* strain CAG20177 carrying the synthetic KAB-TG promoter fused to the *lac* genes. The KAB-TG promoter (illustrated in Fig. 2) was chosen because its activity is dependent on an extended -10 'TG element' [11]. Strain CAG20177 was chosen as its chromosomal *rpoD* gene encoding σ^{70} can be repressed so that the sole source of σ^{70} is that provided by pKB σ [12].

The upper panel of Fig. 3 shows the measured effects of the different alanine substitutions on expression from the KAB-TG promoter. Many of the substitutions have little or no effect and the biggest reductions are seen with the HA455 and EA458 substitutions. As controls, the experiment was repeated with three derivatives, KAB-TC, KAB-CG and KAB-CC, in which the extended -10 'TG' motif at positions -15 and -14 (see Fig. 2) was replaced by TC, CG or CC. The results in the three lower panels of Fig. 3 show that the HA455 substitution causes lesser reductions from these promoters whilst, the EA458 substitution, as previously reported [12], causes increased expression. Interestingly, alanine substitutions at Q437, I439, R441 and I443 also cause increased expression and the pattern of the increases is promoter specific. We suppose that most of these increases result from the different substitutions, either directly or indirectly, removing clashes.

The apparent importance of σ^{70} residues H455 and E458 prompted us to construct derivatives of pKB σ encoding other changes at position 455 and 458 of the *rpoD* gene. This resulted in two sets of 20 derivatives encoding σ^{70} with every possible residue at codon 455 and at codon 458. Results in Table 1 (central column in bold) list the effects of the different substitutions on expression from the KAB-TG promoter. The

data show that the only change at position 455 to reduce expression by more than 40% is HA455. This argues that H455 is not essential for promoter recognition and that the effects of the HA substitution are, most likely, indirect. In contrast, nine of the changes at position 458 cause a greater than 40% reduction. This underscores the importance of E458 in promoter recognition. Surprisingly, the decrease is not seen with some substitutions, e.g. EP458 or ER458. Thus, there may be more than one mechanism by which the KAB-TG promoter can be recognised.

In our previous study, Burr et al. [11] described a set of 15 derivatives of the KAB-TG promoter in which the sequence at positions -15 and -14 was replaced by every possible combination. We therefore measured promoter activity with each of these derivatives in the presence of every member of the two sets of 20 pKB σ derivatives, encoding σ^{70} with every possible residue at codon 455 and at codon 458. The data from these assays are summarised in Table 1. The results show that, with every mutant σ^{70} , the KAB-TG promoter is 'preferred' over all of the other promoters carrying different bases at positions -15 and -14 . However, in many cases, substitutions at positions 455 and 458 allow increased expression from promoters carrying certain bases at positions -15 and -14 . The best example is the HS455 substitution, which permits improved recognition of several different promoters. We conclude that amino acid sidechains at both positions 455 and 458 of σ^{70} can contribute to the specificity of recognition of bases at positions -15 and -14 .

3.2. Suppression genetics

Determinants other than H455 and E458 of σ^{70} must also play some role in recognition of the base sequence at positions -15 and -14 of promoters. To attempt to identify these, we searched for substitutions in σ^{70} that would permit better recognition of the KAB-CC promoter (carrying 'CC' at positions -15 and -14). To facilitate our search we started with the pKB σ derivative encoding σ^{70} carrying the EN458 substitution that we had already found slightly to improve recog-

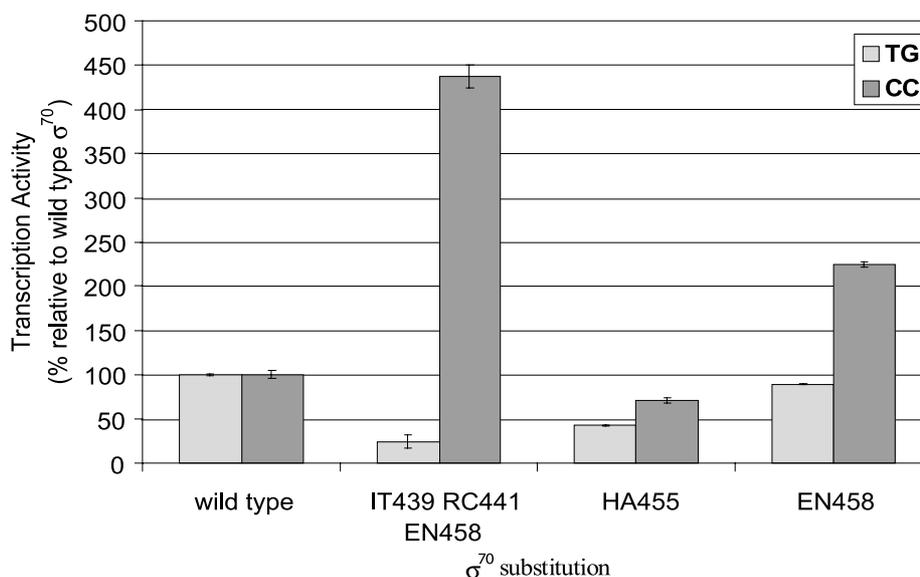


Fig. 5. In vitro transcription with RNAP containing mutant σ^{70} . The bar chart represents quantification of transcripts initiating at the KAB-TG (lighter grey) or KAB-CC (darker grey) promoters with purified RNAP reconstituted with wild type or mutant His-tagged σ^{70} carrying different substitutions as indicated on the x-axis. To perform this experiment the KAB-TG or KAB-CC promoters were cloned into plasmid pSR such that transcripts run to a strong terminator [18]. After incubation of DNA with RNAP and ^{32}P -labelled NTPs, electrophoresis and autoradiography were used to quantify transcripts starting at the KAB-TG and KAB-CC promoters (as in [11]). Quantification was performed using Imagequant software (Molecular Dynamics) and data are represented in arbitrary units with the plasmid-encoded RNA-I taken as an internal standard. The bar chart shows the production of transcripts at each promoter as a percentage of the amount of transcript with wild type σ^{70} .

nition of the KAB-CC promoter (Table 1). PCR was used to generate mutant pKB σ libraries that were screened in *E. coli* strain CAG20177 carrying the synthetic KAB-CC promoter fused to the *lac* genes. After making 30 independent libraries and screening over 250 000 colonies, we isolated a single colony that exhibited increased Lac expression due to substitutions in pKB σ . This mutant derivative encoded σ^{70} carrying the EN458 substitution in Region 3.0 and the IT439 and RC441 substitutions in Region 2.4. The ability of this mutant to drive expression from the KAB-TG derivatives carrying different base sequences at positions -15 and -14 was then measured. The data, illustrated in Fig. 4, show that the triple IT439 RC441 EN458 mutant gives increased expression with the KAB-CC promoter as expected, but also shows increased expression with other promoters (e.g. KAB-AC). For comparison, data is also shown with σ^{70} carrying the EN458 alone. From this, it is clear that the IT439 and RC441 substitutions cause supplementary effects, and we conclude that residues 439 and 441 of σ^{70} contribute to recognition of base pairs at position -15 and -14 . To confirm this, we purified σ^{70} carrying the IT439 RC441 and EN458 substitutions and, after reconstitution into RNAP, we measured its ability to recognise the KAB-TG and KAB-CC promoters in an in vitro transcription assay. For comparison, RNAP was also reconstituted with wild type σ^{70} and with σ^{70} carrying solely the EN458 or HA455 substitutions. The results, summarised in Fig. 5, indicate that RNAP with the triple IT439 RC441 EN458 mutant σ^{70} gives increased transcription with the KAB-CC promoter and reduced transcription with KAB-TG. Comparison with the single EN458 mutant confirms that the IT439 and RC441 substitutions contribute to this alteration in promoter preference.

3.3. Conclusions

The precise molecular mechanism by which RNAP recog-

nises specific promoter sequences is a long standing puzzle that has received much attention [1,2,19,20]. In this study, we took a genetic approach to understand how promoter extended -10 elements are contacted. Our results are easy to interpret in the light of the recent structures of σ factors either alone or within RNAP [3–7]. It is clear that the extended -10 element can make contact with sidechains in both Region 2.4 and Region 3.0, presumably because these two α helices form a pincer that grips the promoter DNA. Our data argue for involvement of residues 439, 441, 455 and 458, and it is clear that different specificities can be generated by different side-chain combinations. The precise details remain to be determined. Interestingly, residues 440 and 441 have been implicated in -10 element recognition [1,2]: thus, a degree of flexibility is likely.

Acknowledgements: This work was supported by a Wellcome Trust project grant and by a University of Birmingham Medical School Ph.D. studentship. We thank Seth Darst for many helpful discussions and for sharing results prior to publication. We are grateful to the University of Birmingham Genomics Facility for automated DNA sequencing.

References

- [1] Gross, C.A., Lonetto, M. and Losick, R. (1992) in: Transcriptional Regulation (McKnight, S.L. and Yamamoto, K.R., Eds.), pp. 129–176, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- [2] Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J. and Young, B. (1998) Cold Spring Harbor Symp. Quant. Biol. 63, 141–155.
- [3] Campbell, E., Muzzin, O., Chlenov, M., Sun, J., Olson, C., Weinman, O., Trester-Zedlitz, M. and Darst, S. (2002) Mol. Cell 9, 527–539.
- [4] Murakami, K., Masuda, S. and Darst, S. (2002) Science 296, 1280–1284.

- [5] Murakami, K., Masuda, S., Campbell, E., Muzzin, O. and Darst, S. (2002) *Science* 296, 1285–1290.
- [6] Vassilyev, D., Sekine, S., Laptenko, O., Lee, J., Vassilyeva, M., Borukhov, S. and Yokoyama, S. (2002) *Nature* 417, 712–719.
- [7] Young, B., Gruber, T. and Gross, C. (2002) *Cell* 109, 417–420.
- [8] Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A., Niu, W., Ebright, Y., Levy, R. and Ebright, R.H. (2002) *Cell* 108, 1–20.
- [9] Camerero, J., Shehktman, A., Campbell, E., Chlenov, M., Gruber, T., Bryant, D., Darst, S., Cowburn, D. and Muir, T. (2002) *Proc. Natl. Acad. Sci. USA* 99, 8536–8541.
- [10] Bown, J., Barne, K., Minchin, S. and Busby, S. (1997) in: *Nucleic Acids and Molecular Biology* (Lilley, D. and Eckstein, F., Eds.), Vol. 11, pp. 41–52, Springer.
- [11] Burr, T., Mitchell, J., Kolb, A., Minchin, S. and Busby, S. (2000) *Nucleic Acids Res.* 28, 1864–1870.
- [12] Barne, K., Bown, J., Busby, S. and Minchin, S. (1997) *EMBO J.* 16, 4034–4040.
- [13] Bown, J., Owens, J., Meares, C., Fujita, N., Ishihama, A., Busby, S. and Minchin, S. (1999) *J. Biol. Chem.* 274, 2263–2270.
- [14] Rhodius, V. and Busby, S. (2000) *J. Mol. Biol.* 299, 311–324.
- [15] Lonetto, M., Rhodius, V., Lamberg, K., Kiley, P., Busby, S. and Gross, C. (1998) *J. Mol. Biol.* 284, 1353–1365.
- [16] Sanderson, A. (2003) Ph.D. thesis, The University of Birmingham, UK.
- [17] El-Robh, M. and Busby, S. (2002) *Biochem. J.* 368, 835–844.
- [18] Rhodius, V., Savery, N., Kolb, A. and Busby, S. (2001) in: *Methods in Molecular Biology: Protein-DNA Interaction Protocols* (Moss, T., Ed.), chapter 31, pp. 451–464, Humana Press.
- [19] Busby, S. (1995) *J. Biol. Educ.* 29, 95–96.
- [20] Busby, S. and Ebright, R. (1994) *Cell* 79, 743–746.