

Solution Structure of the Activator Contact Domain of the RNA Polymerase α Subunit

Young Ho Jeon, Tomofumi Negishi, Masahiro Shirakawa,
Toshio Yamazaki, Nobuyuki Fujita, Akira Ishihama,*
Yoshimasa Kyogoku*

The structure of the carboxyl-terminal domain of the *Escherichia coli* RNA polymerase α subunit (α CTD), which is regarded as the contact site for transcription activator proteins and for the promoter UP element, was determined by nuclear magnetic resonance spectroscopy. Its compact structure of four helices and two long arms enclosing its hydrophobic core shows a folding topology distinct from those of other DNA-binding proteins. The UP element binding site was found on the surface comprising helix 1, the amino-terminal end of helix 4, and the preceding loop. Mutation experiments indicated that the contact sites for transcription activator proteins are also on the same surface.

Activation of gene transcription in a prokaryote system is triggered by several kinds of transcription activators (1). In the *E. coli* RNA polymerase holoenzyme—composed of a core enzyme, α , β , β' , and one of several σ subunit species (2)—one of the regions responsible for transcription activation has been localized to the COOH-terminal third of the α subunit. Deletion of this region does not interfere with the assembly of the core or the holoenzyme, but reconstituted RNA polymerase containing COOH-terminal truncated α subunits cannot be activated by a group of transcription activator proteins (3). This group of proteins contains the class I transcription factors, and their contact sites have been placed at var-

ious positions in the COOH-terminal domain (4). In contrast, the so-called UP element in the *rrmB* P1 promoter is required for the transcription activation of the target ribosomal RNA gene. The isolated α subunit and its COOH-terminal domain protect the UP element region from deoxyribonuclease I (DNase I) digestion (5), which indicates that the COOH-terminal portion of the α subunit is responsible for the contact with *cis*-acting UP elements as well as with *trans*-acting transcription factors (6). Proteolytic cleavage experiments indicated that the COOH-terminal portion that is essential for the activation of polymerase forms independent structural domains (7). We isolated α CTD, a 98-amino acid COOH-terminal fragment of the α subunit (residues 233 to 329 plus methionine at the NH₂-terminus), and we determined its solution structure by multidimensional heteronuclear magnetic resonance spectroscopy (8). The secondary structure elements and the calculated three-dimensional (3D) structures are shown in Fig. 1.

The structure of α CTD is compactly folded and comprises four helices and two long loops at the terminals of the domain (Fig. 1B). Helix 1 (residues 264 to 273), helix 2 (residues 278 to 283), and helix 4 (residues 297 to 309) are almost perpendicular to each other. Helix 2 is short but is essential for the formation of the hydrophobic core; the H-D exchange rates of amide protons in this helix were found to be very slow. Helix 3 (residues 286 to 292) is roughly antiparallel to helix 4. Helix 4 is the longest helix and shows the most typical amphipathic character. The two loops that extend from helices 1 and 4 like arms enclose the hydrophobic core and meet each other through contact between Phe²⁴⁹ in the NH₂-terminal arm and Trp³²¹ and Ile³²⁶ in the COOH-terminal arm. The NH₂-terminal arm makes a turn at Pro²⁵⁶, and the four residues before Arg²⁵⁵ and the four residues after Pro²⁵⁶ form α -helical turns. The COOH-terminal region from Trp³²¹ to Ile³²⁶, which contains two proline residues, also shows a sharp turn in which Pro³²² takes on the *cis* configuration at the peptide bond. The root-mean-square deviation (RMSD) for the backbone heavy atoms of 50 structures from Phe²⁴⁹ to Ile³²⁶ (Fig. 1A) is 0.67 Å. Although this core region is rich in arms or loops, loop regions as well as helical regions are well determined (the RMSD for the region from Val²⁶⁴ to Ser³⁰⁹ is 0.57 Å). When the data were checked with Eisenberg's 3D profile analysis (9), this folding was found to be reasonable. Because the NH₂-terminal 16 residues preceding Phe²⁴⁹ and the COOH-terminal three residues did not show any long-range nuclear Overhauser effect (NOE), this region apparently does not form a single definite structure.

α CTD is known to interact with the *rrmB* P1 promoter UP element in DNase I

Y. H. Jeon, T. Yamazaki, Y. Kyogoku, Institute for Protein Research, Osaka University, Suita, Osaka 565, Japan.
T. Negishi, N. Fujita, A. Ishihama, Department of Molecular Genetics, National Institute of Genetics, Mishima, Shizuoka 411, Japan.
M. Shirakawa, Graduate School of Bioscience, Nara Institute of Science and Technology, Takayama, Ikoma, Nara 630-01, Japan.

*To whom correspondence should be addressed.

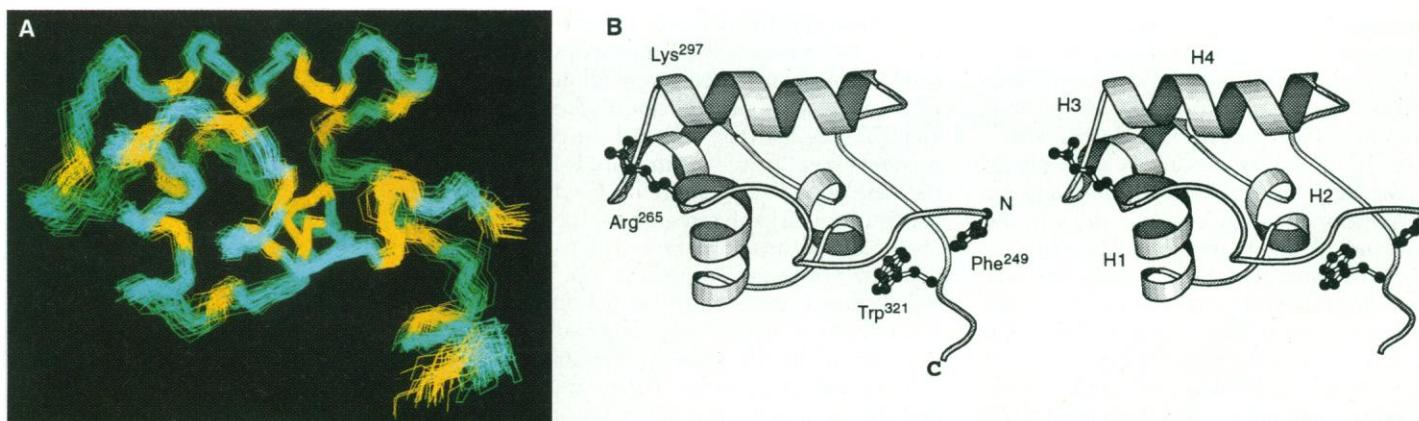


Fig. 1. (A) Peptide backbone traces (N, C α , C') of 50 simulated annealing structures of α CTD superimposed for residues Phe²⁴⁹ to Ile³²⁶. Hydrophobic residues (Phe, Trp, Val, Ile, and Leu) are shown in yellow. (B) Stereo view of a ribbon diagram representing the calculated mean structure folding of α CTD,

produced with the program Molscript (16). The side chains of Phe²⁴⁹, Arg²⁶⁵, and Trp³²¹ are shown; H, helix. The atomic coordinates of the mean structure were deposited at the Brookhaven Protein Data Bank (accession number 1COO).

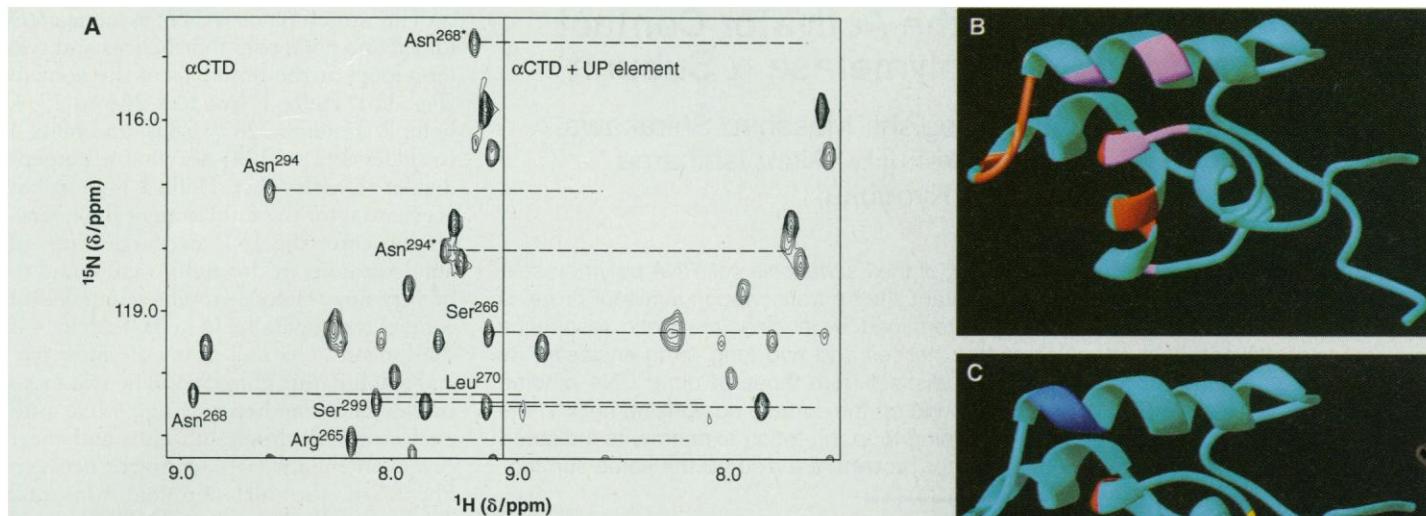


Fig. 2. (A) (^{15}N , ^1H) HSQC spectra of the ^{15}N -labeled αCTD and the complex with the UP-element duplex DNA (11), showing part of the main chain and side chain amide resonances. Left: Reference spectrum for 0.3 mM protein [in 0.5 M KCl and 20 mM phosphate buffer, $\text{H}_2\text{O}/\text{D}_2\text{O}$ ratio 90:10, pH 6.0, at 30°C] with a ^1H frequency of 500 MHz. The assignments of the amide resonances are indicated. Asterisks denote the amide side chain resonances. Right: Spectrum of the protein in the presence of 0.1 molar equivalent of the UP-element DNA. The solution conditions were the same as for the reference experiment. The contour levels were adjusted to the protein concentrations. Resonances that disappeared or were extremely broadened for the complex with DNA under high-salt conditions are indicated by dashed lines. (B) Mapping of the residues with a broadening effect in (^{15}N , ^1H) HSQC spectra. The residues whose resonances completely disappeared under high salt conditions are indicated in red (Arg 265 , Asn 268 , Leu 270 , Thr 292 , Asn 294 , Leu 295 , and Gly 296), and those with an intermediate broadening effect are colored magenta (Thr 263 , Val 264 , Ala 274 , Ser 299 , and Glu 302). (C) Mapping of the proposed contact sites for CRP and OxyR in the structure of αCTD , produced with the program Ribbons (17). The residues whose replacement renders RNA polymerase insensitive to activation by CRP are colored yellow and red (Leu 260 , Arg 265 , Asn 268 , Cys 269 , and Leu 270), and those in the case of OxyR are colored dark blue and red (Arg 265 , Asn 268 , Cys 269 , Lys 298 , and Ser 299); the residues in red are common to both factors. Mutations in the region from Pro 293 to Leu 300 also give weak reduction for CRP activation (15).

footprinting experiments. However, it is not known which residues of αCTD are involved in the interaction. To probe the binding site, we performed chemical shift perturbation experiments (10). Selective signal losses were observed in the (^{15}N , ^1H) heteronuclear single quantum correlation (HSQC) spectrum on mixing with a small amount of a 25-base pair (bp) DNA duplex with a sequence [d(TCAGAAAATTATTTTAAATTTTCCTC)] that corresponded to the *rrb* P1 UP element (from -61 to -37) (11) (Fig. 2A). Signal losses are observed when the exchange rate between the free and bound states is intermediate and when the chemical shift is largely perturbed by binding; in this case, the lost signals were attributable to amides of most of the residues from Glu 261 to Ile 275 and from Thr 292 to Ile 303 . These residues are located in helix 1, the NH_2 -terminal half of helix 4, and the loop region between helices 3 and 4 in the structure of αCTD (Fig. 2B). This observation indicates that helix 1 and the surrounding region are directly involved in the interaction with DNA. Together with the fact that the substitution of Arg 265 to any other amino acid residue, even to Lys, abolished the binding to the UP element (12), our results strongly support the idea that helix 1 recognizes the base sequence of the UP element. Asn 268 on the same side of helix 1 should also be involved in the rec-

ognition. Because one helix and the NH_2 -terminal of another perpendicular helix make a binding surface to DNA, this feature may seem reminiscent of the helix-turn-helix (HTH) motif found in DNA-binding proteins. However, helices 1 and 4 of αCTD are distinct from the canonical or extended HTH (13); not only is there no direct connection between the two helices, but also the direction of one of the helices is opposite to that of HTH.

The COOH-terminal region of the α subunit is claimed to be the contact site for class I transcription factors (4). The mutations that largely reduced activation by CRP [adenosine 3',5'-monophosphate (cAMP) receptor protein] were all mapped to the narrow region between residues 260 and 270 (12, 14) (Fig. 2C). All the critical residues except Leu 260 are in helix 1. Essential residues for activation by OxyR (an activator for oxidative response genes) are also clustered in two narrow regions, residues 265 to 269 and 293 to 300 (15). These two regions correspond to helix 1 and the end of helix 4, respectively. The two separate regions in the sequence are spatially close to each other in the tertiary structure, and they overlap with the contact sites for the UP element of the *rrb* P1 promoter.

αCTD is one of the focal points in the control of transcription. Mutations in αCTD that affect the activation of many

transcription activator proteins have been reported, and these mutations have been mapped at various positions of the domain. Our results may lead to the determination of interaction surfaces for activator proteins through a better understanding of whether such mutations affect the structure or interaction of the domain.

REFERENCES AND NOTES

1. A. Ishihama, *Trends Genet.* **4**, 282 (1988).
2. R. R. Burgess, in *RNA Polymerase*, R. Losick and M. Chamberlin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1976), pp. 69-100; J. D. Hellman and M. J. Chamberlin, *Annu. Rev. Biochem.* **57**, 839 (1988).
3. K. Igarashi and A. Ishihama, *Cell* **65**, 1015 (1991); K. Igarashi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8958 (1991); A. Ishihama, *Mol. Microbiol.* **6**, 3283 (1992).
4. A. Ishihama, *J. Bacteriol.* **175**, 2483 (1993).
5. W. Foss *et al.*, *Science* **262**, 1407 (1993); E. E. Blatter *et al.*, *Cell* **78**, 889 (1994).
6. S. Busby and R. H. Ebright, *Cell* **79**, 743 (1994); R. H. Ebright and S. Busby, *Curr. Opin. Genet. Dev.* **5**, 197 (1995).
7. T. Negishi, N. Fujita, A. Ishihama, *J. Mol. Biol.* **248**, 723 (1995).
8. αCTD was overexpressed under the control of the T7 promoter. Uniformly ^{15}N , ^{13}C -doubly labeled αCTD was produced in M9 medium with [^{15}N]NH $_4$ Cl (0.5 g/liter) and [$^{13}\text{C}_6$]D-glucose (1 g/liter) as the nitrogen and carbon sources. αCTD was collected from the supernatant fraction of the sonicated cell suspension. After salting out and dialysis, the supernatant was applied to a DEAE-Sepharose column (Pharmacia). The peak fractions of αCTD were applied to a gel-filtration column and then to a Protein-pak G-DEAE column (Waters); purity was >99%. Sequence analysis of the NH_2 -terminal end showed

that methionine was retained. A sedimentation equilibrium experiment on a 0.1 mM protein solution revealed a molecular weight of 11,800, which indicates that most of the protein molecules exist as monomers. Size exclusion chromatography also showed a monomeric molecular size. Nuclear magnetic resonance measurements of 2 mM solutions of the protein [in 30 mM KCl, 20 mM phosphate buffer, and 1 mM dithiothreitol (pH 6.0) at 30°C] were performed with Bruker AMX500 and DMX500 spectrometers. Most of the observable ^1H , ^{15}N , and ^{13}C nuclei were given resonance assignments by means of the following 3D experiments: ^{15}N -edited total correlation spectroscopy (TOCSY)-HSQC, CBCA(CO)NH, CBCANH, HBHA(CBCACO)NH, HBHA(CB-CA)NH, HNCO, and HN(CA)CO for the backbone nuclei [A. Bax and S. Grzesiek, in *NMR of Proteins*, G. M. Clore and A. M. Gronenborn, Eds. (Macmillan, London, 1993), pp. 33–52], and HEHOHEHAHA [A. Majumdar, H. Wang, R. C. Morshauer, E. R. P. Zuiderweg, *J. Biomol. NMR* **3**, 387 (1993)] for the side chain nuclei. Distance information was collected using two-dimensional homonuclear nuclear Overhauser effect spectroscopy (NOESY) with a 100-ms mixing time, and 3D (^1H , ^{13}C) and (^1H , ^{15}N) NOESY-HSQC spectra with a 100-ms mixing time. For torsion angle constraints, the backbone vicinal coupling constants ($^3J_{\text{HN,H}\alpha}$) were estimated from a two-dimensional heteronuclear multiple quantum coherence J (HMJQC-J) spectrum [L. E. Kay and A. Bax, *J. Magn. Reson.* **86**, 110 (1990)] of uniformly ^{15}N -labeled αCTD . The NOE connectivities from strong, medium, and weak cross-peaks were assumed to correspond to the upper limits for proton-proton distances of 3.0, 4.0, and 5.0 Å, respectively. Because stereospecific assignments for the methyl and methylene protons have not yet been performed, appropriate corrections were added for constraints including the pseudoatoms. The hydrogen bond constraints were added for slowly exchanging amides as upper and lower limits of 3.0 and 2.7 Å for N–O, respectively, and 2.0 and 1.8 Å for H–O, respectively. For torsion angle constraints, residues that gave $^3J_{\text{HN,H}\alpha}$ values of <5.5 Hz, 8.0 to 10 Hz; and >10 Hz were estimated to have torsion angles ϕ of -90° to -40° , -160° to -80° , and -140° to -100° , respectively, except for the cases of very strong intraresidue HN–H α NOEs. In total, 822 NOE constraints, 52 dihedral angle constraints, and 19 sets of hydrogen bond constraints were used. Simulated annealing calculations [M. Nilges, G. M. Clore, A. M. Gronenborn, *FEBS Lett.* **229**, 317 (1988)] were performed on Silicon Graphics workstations with X-PLOR [A. T. Brünger, version 3.1 (Yale Univ. Press, New Haven, CT, 1992)]. Three hundred structures were calculated, and 50 structures with distance violations of <0.3 Å and angle violations of $<5^\circ$ were obtained.

9. J. U. Bowie, R. Lüthy, D. Eisenberg, *Science* **253**, 164 (1991); Y. Matsuo and K. Nishikawa, *Protein Sci.* **3**, 2055 (1994).
10. N. Dekker *et al.*, *Nature* **362**, 852 (1993); Y. Kyogoku *et al.*, *Methods Enzymol.* **261**, 524 (1995).
11. A DNA titration experiment was done by adding 0.05, 0.1, 1.0, and 1.2 molar equivalents of DNA to the protein at 30 mM KCl. However, signal broadening was so severe that most peaks were barely observed. We extended the experiment by increasing the KCl concentration to 800 mM in the presence of 0.1 molar equivalent of DNA. Almost every peak reappeared at the 800 mM salt condition. No precipitation occurred throughout these experiments. The solution condition of 500 mM KCl for Fig. 2A was chosen as the best for observation of strongly perturbed peaks separately from others. To determine whether the signal losses were specifically caused by the UP-element DNA, we performed a similar experiment in the presence of a 22-bp DNA duplex [d(TAATGTGAGT TAACTCACATTA)] that corresponded to the consensus binding sequence of CRP. We observed only slight shifts of peak positions for two or three residues, but we could not see any disappearance of peaks because of broadening. Thus, the drastic change in the spectrum in the presence of the UP element was clearly distinguished from the control experiment.
12. K. Murakami, N. Fujita, A. Ishihama, unpublished data.
13. S. C. Harrison and A. K. Aggarwal, *Annu. Rev. Biochem.* **59**, 933 (1990); N. Assa-Munt, R. J. Mortish-

- ire-Smith, R. Aurora, W. Herr, P. E. Wright, *Cell* **73**, 193 (1993).
14. C. Zou, N. Fujita, K. Igarashi, A. Ishihama, *Mol. Microbiol.* **6**, 2599 (1992); M. Tang *et al.*, *Genes Dev.* **8**, 3058 (1994).
15. K. Tao, C. Zou, N. Fujita, A. Ishihama, *J. Bacteriol.*, in press.
16. P. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
17. M. Carson, *ibid.*, p. 958.
18. We thank B. J. Lee of Seoul National University for

CRP binding site DNA, and K. Nishikawa of the Protein Engineering Research Institute and K. Shimizu of the Nara Institute of Science and Technology for the use of the 3D-1D profile program COMPASS. Supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science and Culture of Japan, and by Human Frontier Science Program grant RG-401/95M.

4 August 1995; accepted 19 October 1995

A Population Genetic Test of Selection at the Molecular Level

M. F. J. Taylor,* Y. Shen, M. E. Kreitman

The role of natural selection in molecular evolution has been inferred primarily by rejection of null hypotheses based on neutral theory, rather than by acceptance of specific predictions based on selection. In this report, a population genetic test of a specific prediction for selection on DNA polymorphism is presented. Pyrethroid insecticide use constitutes an experiment for which form of selection and molecular target (voltage-gated sodium channels) are both known. As predicted, differential pyrethroid selection on tobacco budworm populations generated significant geographic heterogeneity in sodium channel marker allele frequencies, compared with arbitrary loci.

Many studies have sought to test the null hypothesis of neutral evolution of DNA sequence variation in populations. However, the lack of specific prior predictions based on selection has weakened such tests. Even for the most well-studied case of the high-to-low latitudinal clines in the *Adh* fast-slow polymorphism in *Drosophila melanogaster*, the form of selection presumed to maintain this polymorphism remains elusive, although patterns of polymorphism are consistent with a model of balancing selection (1).

If a specific form of selection is expected at a particular locus, then a prediction may be developed on the basis of selection rather than neutrality, which may be tested with the use of appropriate statistics. Differentiation in allele frequencies among populations can be estimated by F_{st} , the between population component of standardized genetic variance. Population structure and historical contingencies (range expansions or contractions) have genome-wide effects on gene frequency variation between populations, whereas selection affects variation only at target loci. Differentiation in the form of clines of allele frequency for some allozyme loci, and not for others, has been taken as evidence that spatially variable selection has been acting at such loci,

whereas gene flow has homogenized frequencies at other unselected loci (2). However, clinal patterns for some loci may result from historical patterns of colonization and subsequent admixture, whereas nonclinal loci are homogenized by selection, acting uniformly in space and causing convergence of frequencies over a wide area (3). The study of nucleotide variation has permitted some resolution of this dilemma for the *Adh* cline in *D. melanogaster* (4). Such studies highlight the difficulties of testing for selection at the molecular level in natural populations, in the absence of prior expectations about the strength, duration, or loci of action of selection.

The widespread use of insecticides in agriculture has resulted in the rapid evolution of resistance for many pest species. Such natural "experiments" can enable explicit prediction of the outcome of selection at the molecular level. Pyrethroid insecticides act on voltage-gated sodium channels in nerve membranes (5). Resistance to the pyrethroid permethrin in a field-derived strain of the tobacco budworm (*Heliothis virescens*), a major cotton pest, is linked to a DNA marker for a sodium channel locus *hscp*, homologous to the *para* locus of *D. melanogaster* (6). Levels of resistance are known to vary considerably among populations, and resistance has arisen in response to selection in just the last 10 years (7). Population genetic surveys in North America, ranging from Texas to Georgia for 13 allozyme loci (8), and from Sonora, Mexico, to Georgia for mitochondrial DNA markers (9), have shown that tobacco budworm populations are little differentiated, which

M. F. J. Taylor and Y. Shen, Department of Entomology, Center for Insect Science, University of Arizona, Tucson, AZ 85721, USA.

M. E. Kreitman, Department of Ecology and Evolution, University of Chicago, 1103 East 57th Street, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: Taylor@biosci.arizona.edu