Biochemistry

pubs.acs.org/biochemistry Article

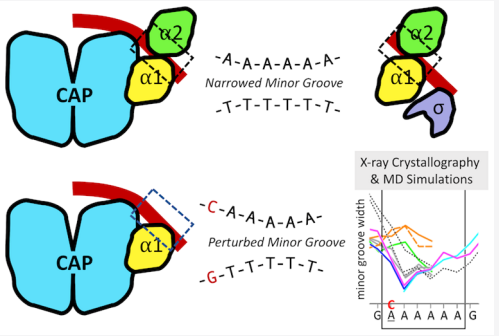
The RNA Polymerase α Subunit Recognizes the DNA Shape of the Upstream Promoter Element

Samuel Lara-Gonzalez, Ana Carolina Dantas Machado, Satyanarayan Rao, Andrew A. Napoli, Jens Birktoft, Rosa Di Felice, Remo Rohs,* and Catherine L. Lawson*

Cite This: https://dx.doi.org/10.1021/acs.biochem.0c00571 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We demonstrate here that the α subunit C-terminal domain of *Escherichia coli* RNA polymerase (αCTD) recognizes the upstream promoter (UP) DNA element via its characteristic minor groove shape and electrostatic potential. In two compositionally distinct crystallized assemblies, a pair of αCTD subunits bind in tandem to the UP element consensus A-tract that is 6 bp in length (A_6-tract), each with their arginine 265 guanidinium group inserted into the minor groove. The A_6-tract minor groove is significantly narrowed in these crystal structures, as well as in computationally predicted structures of free and bound DNA duplexes derived by Monte Carlo and molecular dynamics simulations, respectively. The negative electrostatic potential of free A_6-tract DNA is substantially enhanced compared to that of generic DNA. Shortening the A-tract by 1 bp is shown to "knock out" binding of the second αCTD through widening of the minor groove. Furthermore, in computationally derived structures with arginine 265 mutated to alanine in either αCTD, either with or without the "knockout" DNA mutation, contact with the DNA is perturbed, highlighting the importance of arginine 265 in achieving αCTD−DNA binding. These results demonstrate that the importance of the DNA shape in sequence-dependent recognition of DNA by RNA polymerase is comparable to that of certain transcription factors.



The *Escherichia coli* RNA polymerase (RNAP) is recruited to its promoter sites via interactions between its two flexibly tethered α-subunit C-terminal domains (αCTDs; UniProt P0A7Z4) and DNA and/or DNA-bound transcription activators such as the catabolite activator protein (CAP; UniProt P0ACJ8).[1] We demonstrate here that the RNAP αCTDs recognize their preferred DNA binding site, the upstream promoter (UP) element, via its unique DNA shape that is correlated with the electrostatic potential of the minor groove, proving the importance of a concerted shape readout mechanism for the recruitment of RNAP to DNA. The evidence complements our growing knowledge of the interplay between sequence and structure for protein−DNA recognition in a variety of genomic contexts.[2−4]

The UP element is an AT-rich DNA sequence motif found upstream of many promoters, the most extensively studied of which is *rrnB* P1.[5−7] UP elements stimulate both basal and factor-activated transcription through minor groove interactions with RNAP αCTDs.[5−7] A full UP element consists of two distinct successive upstream minor groove target sites on the same side of the DNA helix as the promoter −35 and −10 elements, with one binding site for each αCTD. The promoter-proximal UP element target site (consensus motif "AAAA-AARNR") can also function alone, stimulating transcription ≤170-fold.[5−7] Placement of an UP element target site within the *E. coli lac* operon promoter just upstream of the −35 element stimulates transcription and also facilitates formation of CAP, RNAP, and the promoter−DNA complex.[8]

We previously identified electrostatic focusing as the biophysical basis for sequence specificity of transcription factors in the minor groove.[9] We now revisit the question of how αCTD achieves DNA sequence specificity by examining two different crystal structures that share a common substructure: two αCTDs bound in tandem to the minor groove of an A_6-tract. Merging of these two complexes by superimposing their overlapping structures enabled construction of a near-atomic-resolution model of a CAP−RNAP−operator DNA assembly that is consistent with molecular shapes determined by electron microscopy.[10,11] Here, we compare the two structures and use them to analyze the DNA shape of the UP promoter element and derive the electrostatic potential along the minor groove using Poisson−Boltzmann calculations at physiological ionic strength. In addition, we describe a third crystal structure with the A-tract shortened by

Received: July 14, 2020
Revised: October 31, 2020

Table 1. Crystallographic Data Collection and Refinement Statistics

	(A) CAP-αCTD-DNA (CAD)	(B) CAP-αCTD-DNA[a]	(C) CAP-αCTD-DNA^{KO} (CAD-KO)	(D) αCTD-σR4-DNA (ASD)
X-ray source	NSLS-X29	NSLS-X25	NSLS-X25	NSLS-X25
wavelength (Å)	1.10	1.0	1.10	0.92
collection temperature (K)	100	100	100	100
space group	$P6_222$	$P6_222$	$P6_222$	$C222_1$
a, *b*, *c* (Å)	175.73, 175.73, 160.10	175.97, 175.97, 158.02	176.31, 176.31, 158.49	80.71, 86.41, 147.11
α, β, γ (deg)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 90
Matthews coefficient (Å^3 Da^{-1})	6.40	6.33	7.66	2.62
solvent content (%)	81	78	84	53
resolution range (Å)	41−3.0	20−3.1	50−5.0	25.4−3.25
no. of measured reflections	129653	332877	124215	24506
no. of unique reflections	29113	23331	6646	7461
completeness (%)	98.9	88.0	99.7	88.6
redundancy	4.5	14.3	18.7	3.3
mean(*I*)/sd(*I*)	11.5	13.7	10.1	11.1
R_{merge}	0.085	0.076	0.095	0.081
refinement resolution range (Å)	38.5−3.0 (3.1−3.0)	20−3.1 (3.21−3.1)	49.9−5.0 (5.40−5.0)	25.4−3.25 (3.72−3.25)
R_{work}	0.198 (0.356)	0.211 (0.448)	0.178 (0.231)	0.254 (0.376)
R_{free}	0.224 (0.401)	0.244 (0.475)	0.214 (0.315)	0.295 (0.374)
no. of working reflections	29042 (2039)	21027 (1705)	6615 (1145)	7422 (2235)
no. of free reflections	1481 (108)	2304 (182)	663 (127)	480 (150)
no. of non-hydrogen atoms	3737	3613	3099	2887
no. of water oxygen atoms	47	32	0	0
Wilson *B*-factor (Å^2)	100	82	207	99
RMSD for bond lengths (Å)	0.004	0.009	0.013	0.006
RMSD for bond angles (deg)	0.95	1.4	1.66	1.21
Molprobity Clashscore	9	39	13	21
Ramachandran outliers (%)	0	2.7	0	2.0
side-chain outliers (%)	5.6	9.8	2.9	8.2
PDB entry	3N4M	1LB2	5CIZ	3N97

[a]CAD structure determination statistics from Benoff et al.16 are provided for comparison.

1 bp. The "knockout" reduces minor groove narrowing, and the known correlation of minor groove width with electrostatic potential is likely responsible for loss of binding of one of the two αCTDs within the crystal. Molecular dynamics (MD) simulations confirm widening of the minor groove at the knocked out αCTD binding position and the importance of arginine 265 (Arg265), whose guanidinium group is inserted into the minor groove, in UP element recognition by αCTD.

MATERIALS AND METHODS

Protein Expression and Purification. CAP was expressed and purified by cAMP affinity chromatography as described by Zhang et al.[12] and further purified on a Heparin 16/10 FF column (GE Healthcare). CAP was then concentrated and buffer exchanged with 0.3 mM dimer, as determined by a Bradford assay using a bovine serum albumin standard, in 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, and 0.1 mM EDTA. αCTD was expressed and purified as described by Lara-Gonzalez et al.[13] σR4 was prepared on the basis of the expression system for *Thermus aquaticus* σR4.[14] σR4 was purified with the heparin 16/10 FF column (GE Healthcare) and further purified by cation-exchange chromatography on a HiPrep SP FF 16/10 column (GE Healthcare). Fractions containing σR4 were then concentrated and buffer exchanged to 1.5−2.5 mM in 20 mM MES (pH 6), 200 mM NaCl, 0.1 mM EDTA, 0.2 mM DTT, and 0.02% NaN_3. The expressed σR4 fragment consisted of *T. aquaticus* σ residues 366−438 (UniProt Q9EZJ8), with mutation of the short helical region anticipated to interact with αCTD with corresponding residues from *E. coli* σR4 (UniProt P00579):[15] 424−427 "KYHE" was replaced with "RHP".

DNA Design, Synthesis, and Annealing. The fully 2-fold symmetric CAD and CAD-KO DNA duplexes were prepared as described by Benoff et al.[16] The ASD DNA duplex was assembled from complementary strands 5′-TGGAA-AAAAGTACTTGACATGG-3′ and 5′-CCATGTCAAG-TACTTTTTTCC-3′, yielding 22 bp with a 5′ T overhang at one end. C-18 dual reverse-phase high-performance liquid chromatography-purified DNA oligonucleotides were purchased from IDT. DNA duplexes were prepared in 5 mM sodium cacodylate buffer (pH 6.5), 10 mM $MgCl_2$, and 5 mM EDTA to yield final duplex concentrations of 0.5−1.0 mM. The duplexes were annealed by being heated at 90 °C for 5 min and then slowly cooled to 25 °C over a period of 12−15 h.

Determination of the Crystal Structure. CAD and CAD-KO crystals were grown and prepared for diffraction as described by Benoff et al.[16] ASD crystals were obtained by hanging drop vapor diffusion with a reservoir solution of 28% (w/v) PEG4000, 0.2 M ammonium acetate, 0.01 M sarcosine, and 0.1 M sodium citrate (pH 5.6). A single crystal was soaked in a cryoprotectant solution composed of the reservoir solution and 14% glycerol (see also ref 10).

X-ray diffraction data were collected at beamlines X25 and X29 of the National Synchroton Light Source. Initial phasing was performed via molecular replacement using Phaser,[17] using coordinates from previous crystal structures of Protein Data

https://dx.doi.org/10.1021/acs.biochem.0c00571
Biochemistry XXXX, XXX, XXX−XXX

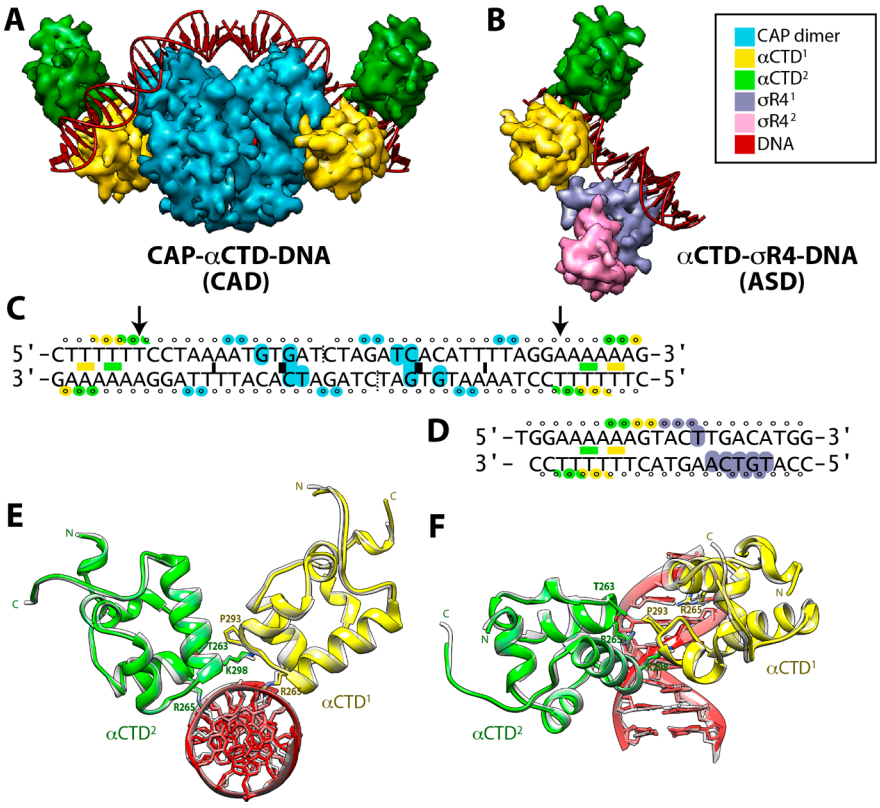


Figure 1. Comparison of two independent crystal structures with αCTDs tandemly bound to the UP element (A_6-tract) DNA. (A) CAD structure, with CAP (cyan) bound to its consensus DNA site, and αCTDs (yellow, green) bound to overlapping sites containing A-tract DNA on either side of CAP. Only αCTD^1 (yellow) makes protein−protein contacts with CAP. The 2-fold symmetric design was used to simplify crystallographic analysis (the asymmetric unit contains one-half of the full complex). (B) ASD structure, with one σR4 (purple) bound to a consensus −35 element DNA site, and αCTDs (yellow, green) bound to overlapping sites centered about adjacent A-tract DNA. Only αCTD^1 (yellow) makes protein−protein contacts with $\sigma R4^1$. A second σR4 (pink) forms an asymmetric dimer with DNA-bound σR4 (purple). In panels A and B, proteins are represented as surfaces; colors are as indicated in the legend at the right. (C and D) Corresponding protein−DNA contact footprints in the co-crystal structure. Nucleotide bases (letters) and backbone phosphates (○) are shaded according to the protein subunit making contact. Intercalation of the minor groove is indicated by shaded rectangles. In panel C, thick and thin black rectangles represent the positions of major and minor DNA kinks. Dotted vertical lines indicate designed breaks in the DNA backbone. Arrows indicate the base pair positions modified in the CAD "knockout" structure. (E and F) Superimposed αCTD/A-tract DNA regions (CAD, colors as in the legend; ASD, semitransparent gray). αCTD residues that participate in αCTD−αCTD contacts (Pro293 of αCTD^1 and Thr263 and Lys298 of αCTD^2) or that intercalate into the DNA minor groove (Arg265 of both subunits) are also displayed. The RMSD for 1484 equivalent protein and DNA atom positions in the two structures is 0.7 Å.

Bank (PDB) entries 1LB2,[16] 1KU7,[14] and 3K4G.[13] Final coordinates were obtained by iterating between automated refinement with Phenix[18] and manual adjustment to sigmaA-weighted model phased density maps with Coot.[19] For each structure, strong noncrystallographic symmetry restraints were imposed between domain components present in duplicate within the asymmetric unit (αCTD, σR4) and overall translation-libration-screw (TLS) parameters were refined (single TLS group) to improve the fit of the model to strongly anisotropic diffraction data. For ASD, additional pseudobond restraints were applied between canonical base pair hydrogen-bonded atoms. CAD-KO atom positions were additionally restrained using reference dihedral angles generated from the CAD structure. Initial refinement of CAD-KO included both αCTD copies, but αCTD^2 was dropped in final refinement rounds because the supporting density was clearly lacking. For CAD and ASD, grouped atomic displacement parameters (one per residue) were also refined. Data collection, processing, and final model statistics are listed in Table 1.

Relative αCTD occupancies were estimated using group occupancy refinement with CNS version 1.3.[20] Starting with the final refined model atom positions for each structure, all *B*-factors were reset to 50 $Å^2$, and grouped occupancies were refined to convergence against observed diffraction data in the resolution range of 50−5.0 Å for each αCTD position, while all other atom occupancies were kept fixed at 1.0. Bulk solvent correction was turned off.

Visualization and Analysis of Structure. Structure superpositions were carried out using PyMol[21] and UCSF Chimera.[22] Interfaces were analyzed using PISA.[23]

DNA Shape Analysis and Electrostatic Potential Calculations. Curves 5.3[24] was used to calculate minor groove width as a function of nucleotide sequence. The electrostatic potential was calculated in reference points at the center of the minor groove in the approximate plane of each base pair.[9] DelPhi[25] was used for nonlinear Poisson−Boltzmann calculations[26] of the electrostatic potential at the physiological ionic strength of 0.145 M. DelPhi is a numerical approach to computing the electrostatic potential of a complex molecular system in solution, given the charge distribution that is determined by the chemical nature and the structural arrangement of atoms. We used the nonlinear Poisson−

https://dx.doi.org/10.1021/acs.biochem.0c00571
Biochemistry XXXX, XXX, XXX−XXX

Boltzmann equation and a previously described numerical protocol.[9] Minor groove width and electrostatic potential were plotted as a function of nucleotide sequences and indicated the Arg265 contacts. Molecular surface representations of shape and electrostatic potential were generated with GRASP2.[27]

Computational Predictions of DNA Structure. Bound and unbound DNA structures were predicted using all-atom MD simulations with Gromacs 5.0.4.[28] The Amber99sb force field[29] was used for the protein and DNA, while explicit water was described using the TIP3P model.[30] Amber99sb is a version of the Amber force field (developed on the basis of Amber94) with refined dihedral parameters to reduce structural transitions in DNA oligomers to noncanonical conformations. Each system was simulated for 300 ns, after a standard minimization−equilibration protocol that is described in the Supporting Information. The investigated systems included the crystal structures of CAD, CAD_KO, and ASD (abbreviations defined in Table 1), as well as systems derived from the crystal structures: CAD+ASD (merged structure), CAD_R265Aα1 (Arg265 mutation in αCTD[1]), CAD_R265Aα2 (Arg265 mutation in αCTD[2]), CAD_KO_R265Aα1, and CAD_KO_R265Aα1_noCAP (CAP coordinates removed) complexes. For each complex, except for the latter two, the unbound DNA duplex was also simulated for the same amount of time. Monte Carlo (MC) simulations of CAD and CAD_KO unbound DNA were also performed (Supporting Information, Supplementary Methods and Materials) as an independent validation of MD data.[31]

DNA shape was computed with Curves 5.3[24] for regular MD snapshots every 10 ps, using a computational toolkit, Trj2Shape, that can process the entire trajectory in a high-throughput manner (Supporting Information, Supplementary Methods and Materials) and complies with shape definitions published elsewhere.[32,33] Instructions and scripts for Trj2Shape are available through GitHub at https://github.com/satyanarayan-rao/Trj2Shape. Contact maps were obtained with Gromacs tools. Interface contacts were analyzed using DNAproDB[34,35] using default settings. Details are provided in the Supporting Information.

Structure Deposition. The crystal structures are available in the Protein Data Bank as entries 3N4M (CAD), 3N97 (ASD), and 5CIZ (CAD-KO).

RESULTS

Two Complexes with αCTDs Bound to A-Tract DNA. In this study, we obtained co-crystal structures for two distinct multiprotein−DNA assemblies that share a common feature, a pair of αCTDs bound to A_6-tract DNA (Figure 1).

In the 2-fold symmetric "CAP-αCTD-DNA" complex (CAD), αCTD is bound to A_6-tracts flanking each side of the CAP binding site (Figure 1A,C). Here, we re-refined the original structure reported by Benoff et al.[16] using new undulator source synchrotron data (Table 1, column A) and improved source coordinates for αCTD.[13] The updated structure is a close variant of the original with modestly improved model statistics but a similar diffraction limit (in Table 1, compare columns A and B; extensive efforts to obtain higher-resolution diffraction data for CAD were unsuccessful, likely because of the high solvent content). The updated CAD structure confirms that both αCTDs interact with DNA via the A-tract DNA minor groove (Figure 1A,C and Figure S1A,B) and that DNA-bound CAP and adjacent DNA-bound αCTD[1] form a small (~350 Å^2) but well-defined interface. At the "primary kink" position where the 44 bp DNA bends toward CAP (Figure 1C, thick rectangles[36]), the new structure has a roll angle (52°) somewhat larger than that in the original determination (44°) or in other CAP-DNA complexes (on average 43°).[37] The larger roll and resulting kink are accompanied by an unusual electron density feature consistent with a purine base intercalated into the minor groove side of the CpA base pair step (Figure S1C). Its most likely source is an adenine from cyclic AMP. The increased roll angle and intercalated density are the only major differences that we observe between the original structure and this new determination; all protein−protein and protein−DNA interfaces described by Benoff[16] are preserved in the new CAD structure. Given the high solvent content (81%), cryocooling alone may have introduced this structural difference.

In the asymmetric "αCTD-σR4-DNA" complex (ASD), αCTD is bound to an A_6-tract adjacent to a −35 element with σ subunit region 4 (σR4) bound [Figure 1B,D; because *E. coli* σR4 was refractory to crystallization, σR4 in this structure is a chimera of *T. aquaticus* σR4 with a short *E. coli* segment inserted near the expected position of interaction with αCTD[15] (see Materials and Methods for details)]. A preliminary description of this structure was given by Hudson et al.;[10] here we report the fully refined 3.25 Å structure (Table 1, column D). The region of ASD that encompasses σR4 and the −35 element DNA is essentially equivalent to the co-crystal structure of the *T. aquaticus* σR4/−35 element DNA complex determined by Campbell et al.[14] In both structures, two σR4 domain copies form an extended asymmetric dimer (the RMSD for all equivalent Cα atoms in each σR4 dimer is 0.6 Å). Only one copy forms an interface with the −35 element DNA (σR4[1]; purple footprint in Figure 1D). The fortuitous dimerization supports the folded state of the isolated 72-residue σR4 domain fragment in the crystal. Within the RNAP holoenzyme, σR4 is supported by other structural elements, notably the "β-flap".[38,39] The ASD structure confirms the proximity of an acidic surface on αCTD (residues 257−261) and a basic surface on σR4 (residues 593−604, *E. coli* numbering) upon DNA binding, as has been predicted for RNAP bound to both CAP class I and UP element promoters.[40,41] In the crystal, there is only one direct contact between αCTD and σR4 consisting of a weak (3.5 Å) hydrogen-bonded salt bridge. However, a modest bend in the connecting DNA could readily bring the predicted regions into more direct contact, as is observed in the CAP−RNAP−operator DNA complex structure recently determined at 3.9 Å using cryo-electron microscopy.[11] We note that comparative analysis beyond the coarse level of domain placement is not warranted given limited resolution and concern that chimeric σR4 in the ASD structure may have introduced structural artifacts.

Comparing the CAD and ASD structures, we find that the regions encompassing the two αCTDs bound to A_6-tract DNA are effectively identical (Figure 1E,F; the RMSD for all equivalent modeled protein and DNA atom positions is 0.7 Å). Each αCTD binds in an identical way to the minor groove side of a 5 bp DNA target (yellow and green footprints in Figure 1C,D), with a modest protein−DNA interface area of 400 Å^2 (compare with a value of 2500 Å^2 for the full CAP dimer−DNA interface). In each structure, the two αCTDs bind to the A-tract DNA with the same polarity and precisely 2 bp apart from each other (approximate rotation of 70° and rise of 6.8 Å with respect to the DNA helix axis). The small protein−

https://dx.doi.org/10.1021/acs.biochem.0c00571
Biochemistry XXXX, XXX, XXX−XXX

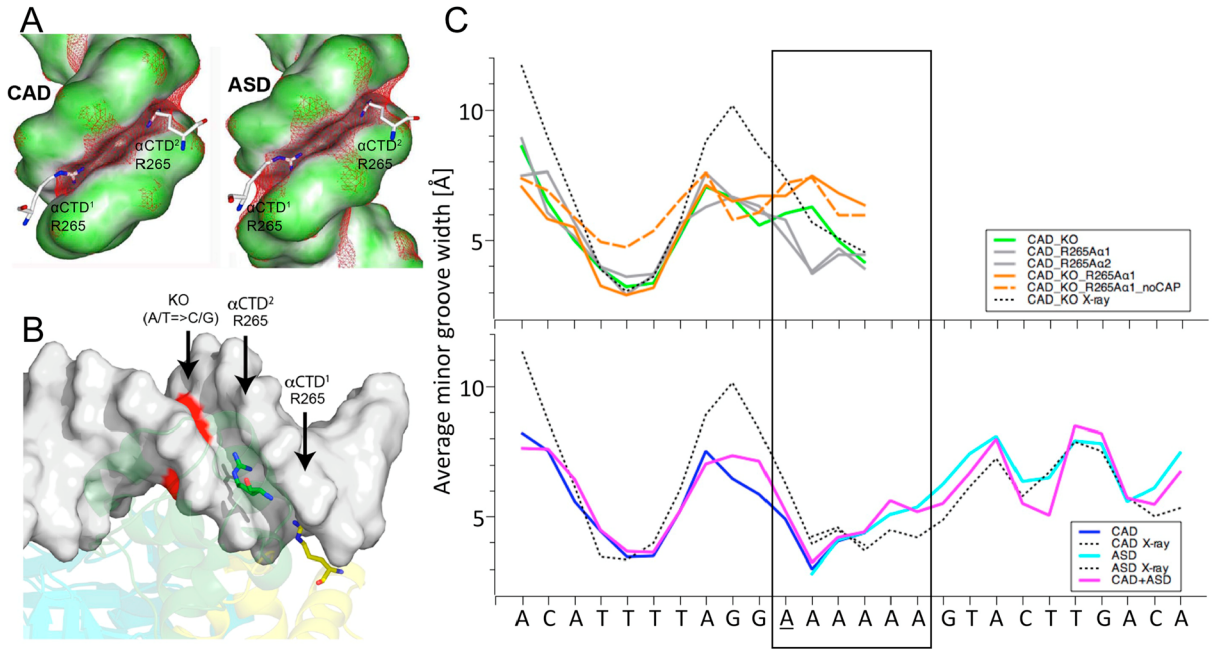


Figure 2. Minor groove width of the A_6-tract that is very narrow in CAD and ASD crystal structures, with a consistent minimum of enhanced negative electrostatic potential. (A) Molecular surface of the A_6-tract for CAD (left) and ASD (right) color-coded by shape (convex surfaces colored green, concave surfaces colored dark gray). Darker gray shading corresponds to a narrower minor groove. The Arg265 residues from each αCTD that insert into the minor groove are shown as sticks. The isopotential surface at −5 kT/e (red mesh) is shown, superimposed on the molecular surfaces. (B) CAD A_6-tract region with αCTD Arg265 residue positions. The surface of the base pair that is mutated in the CAD-KO knockout complexes is colored red. (C) DNA minor groove width profiles from crystal structures and MD trajectories, as a function of nucleotide position. On the horizontal axis, CAD and ASD DNA sequences are merged via overlap in the A_6-tract. The underlined position marks the location at which adenine is mutated to cytosine in CAD-KO, CAD_KO_R265Aα1, and CAD_KO_R265Aα1_noCAP. The region corresponding to the A_6-tract is highlighted with a black rectangle (the left-most narrow minor groove region is part of the CAP binding site). In the short (3 bp) region where the CAD X-ray- and ASD X-ray-derived curves overlap, the two can be distinguished by the continuity of connected points.

protein interface between αCTD^1 and αCTD^2 (160 Å^2) involves just 12 residues and three hydrogen bonds (compare with 350 Å^2 and 20 residues, respectively, for the αCTD^1–CAP protein–protein interface). The hydrogen bonds at this interface are coordinated through a single ε-amino nitrogen on αCTD^2 from Lys298, which contacts main-chain carbonyl oxygens of αCTD^1 Leu290, Thr292, and Leu295 and the DNA backbone phosphate oxygen. The equivalent site at αCTD^2 is coordinated by water. The observation of identical shape complementarity, hydrogen bonding, and charge neutralization in both crystal structures suggests that this organization may constitute a preferred polarity/orientation for multiple αCTDs cooperatively bound to DNA.

The αCTDs in CAD and ASD were also compared with the previously determined structure of free αCTD in the absence of DNA.[13] Given that the free αCTD coordinates were used as starting models in CAD and ASD crystallographic refinement, bound versus free αCTD conformations were expected to be similar. Overall, RMSDs are ∼0.7 Å for pairwise C_α superpositions. A modest and remarkably consistent bound versus free deformation is observed, however, with the largest shifts at residues involved in minor groove recognition: V264, R265, and N294 (Figure S2). Relative to the free state, bound residues are ∼1.5 Å further away from the bulk domain and closer to the DNA minor groove (Figure S2, inset). This deformation enables the side chain of R265 to insert slightly further into the minor groove.

αCTD-Bound A-Tracts: DNA Shape and Electrostatic Potential. In each of the four αCTD–DNA interfaces in these two crystal structures, the side-chain guanidinium groups of αCTD Arg265 are inserted into the A-tract minor groove. The minor groove width of the A-tract is in each case decreased to ∼3 Å, compared to 5.8 Å for canonical B-DNA (Figure 2 and Figure S3). This minor groove narrowing generates a region with an enhanced negative electrostatic potential of approximately −9 kT/e, which in turn attracts the Arg265 residues that reside at the center of the DNA binding surface of αCTD (Figure 2A). The decreased groove width and enhanced negative electrostatic potential are nearly identical for each αCTD binding site (Figure S3). The fact that this readout pattern is detected for all independent observations suggests that this specific shape is an intrinsic feature of the DNA sequence and thus provides a structural basis for UP element recognition. Indeed, the comparison between minor groove width profiles of bound and unbound DNA from our MD trajectories (Figure S4) supports this hypothesis. Crystal structures of unbound DNA containing A_6-tracts have similarly narrowed minor grooves[42,43] (Figure S5).

To further investigate the observed binding mechanisms, we analyzed a single DNA base pair substitution in the CAD complex that we predicted would disrupt binding of αCTD^2 but would still permit crystallization (within the CAD crystal lattice, αCTD^2 is highly solvent-exposed and does not participate in crystal contacts). The substitution replaces the first A/T base pair of the A_6-tract (arrows in Figure 1B, red in Figure 2B, underlined in Figure 2C) with C/G, reducing the A-tract length to 5 bp. At an UP element-proximal site, this substitution reduces transcriptional activity, but by only ∼30%.[5] A structure for this "knockout" complex ("CAD-KO") was obtained on the basis of crystal diffraction data

https://dx.doi.org/10.1021/acs.biochem.0c00571
Biochemistry XXXX, XXX, XXX−XXX

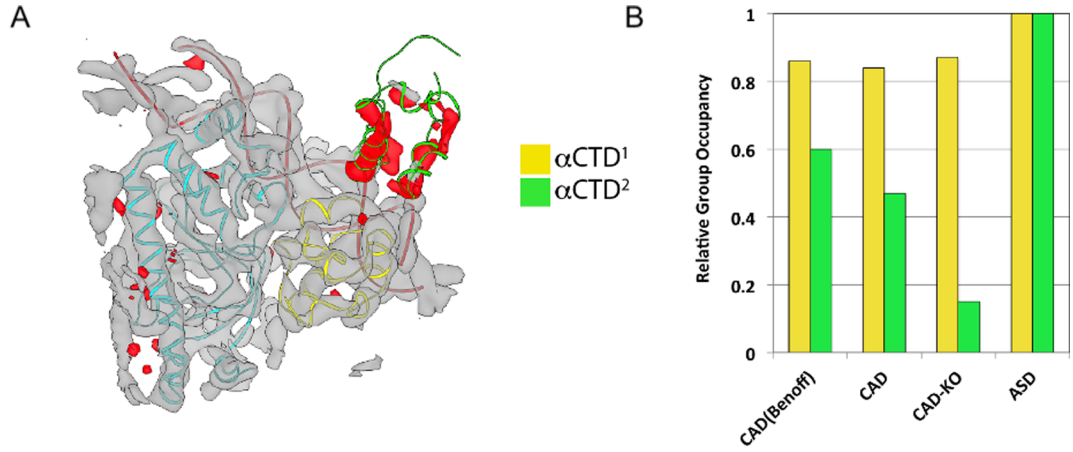


Figure 3. Shortening of the A-tract impairs binding of the second αCTD. (A) Model map for CAD-KO ($2mF_o - DF_c$, gray contour at 2.0σ) and data vs the model difference map ($mF_o - DF_c$, red contour at -3.25σ) obtained after atomic position refinement with the full (two αCTDs) CAD model. The orientation and color scheme correspond to those of Figure 1A (right half). (B) Relative αCTD occupancies in CAD, CAD-KO, and ASD crystals obtained by grouped refinement against the measured diffraction data sets (see Materials and Methods).

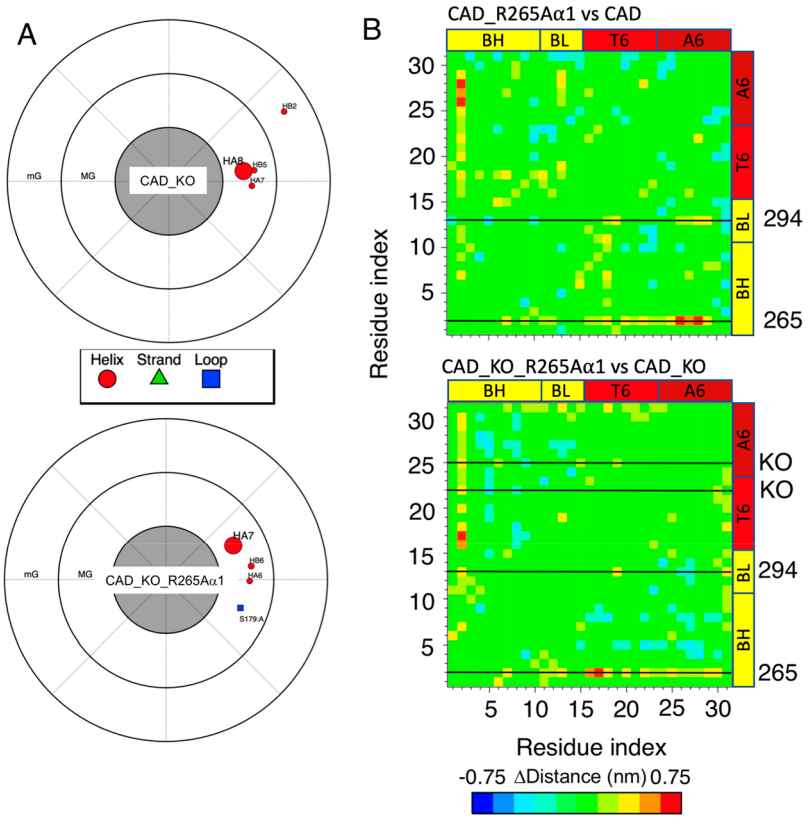


Figure 4. Contact between αCTD^1 and the DNA minor groove weakened upon mutation of Arg265 to Ala. (A) DNAproDB polar contact maps for CAD_KO (top) and CAD_KO_R265Aα1 (bottom), obtained for the most representative structure that covers the final span of the 300 ns MD trajectory [65% cluster population for CAD_KO and 17% cluster population for CAD_KO_R265Aα1 (see Table S1)]; the contact between αCTD^1 and the DNA is represented by the HB2 red dot in the minor groove annulus (labeled as "mG"). (B) Residue−residue difference distance contact map, obtained with Gromacs; the map is symmetric with respect to the diagonal. A positive value of Δdistance indicates that a residue−residue distance increases as an effect of the R265A protein mutation, in the absence (top) and presence (bottom) of the knockout DNA mutation. Residue indices: binding helix BH (1−10), binding loop BL (11−15), T_6+-tract (16−23), and A_6+-tract (24−31). The positions of αCTD^1 protein residues Arg265 (in BH) and Asn294 (in BL), as well as the positions of the DNA knockout mutation when present (bottom), are marked by horizontal lines.

obtained to a resolution limit of 5 Å (Table 1, column C), following a conservative structure refinement strategy with geometrical constraints derived from unsubstituted CAD. This structure confirmed that binding of αCTD^2 is indeed disrupted by the shortening of the A-tract (Figure 3). Weakened binding is indicated by strongly negative difference density (Figure 3A) and negligible refined occupancy for αCTD^2 relative to αCTD^1 (Figure 3B).

To reinforce this finding, we carried out MD simulations to predict the structure of bound CAD and CAD-KO complexes,

https://dx.doi.org/10.1021/acs.biochem.0c00571
Biochemistry XXXX, XXX, XXX−XXX

as well as the corresponding unbound DNA duplexes, and compared the result with the known crystal structures (Figure 2C and Figures S4 and S5). The KO mutation is 2 bp from the binding position of αCTD[2] R265 (Figure 2B). For the CAD DNA duplex, the locations and extent of minor groove narrowing observed for the A_6-tract in the protein-bound co-crystal structure (Figure 2C, dotted black line, left part from base pair 1 to 14) were also present in the predicted CAD complex structure in the solution state (Figure 2C, blue solid line), as well as in two previously determined unbound DNA crystal structures with sequences of CGCGAAAAAACG (PDB entry 1D89) and CGCAAAAAAGCG (PDB entry 1D98) (Figure S5). We find that the A/T to C/G substitution in CAD-KO widens the minor groove in the region contacted by αCTD[2], both in the computationally predicted bound structure (Figure 2C, green solid line) and in the co-crystal structure (Figure S5). Thus, our results suggest that arginine residues are used to specifically recognize the UP element through means of electrostatic potential and narrow minor groove width, a recognition mechanism known as DNA shape readout.[44]

The observations from MD simulations of the protein–DNA complexes are summarized in Figure 2C in terms of the minor groove width profile for all structures. In addition to the minor groove widening effect of the knockout DNA mutation noted above, we also note that (i) the A_6-tract in the ASD complex has the same narrow minor groove width as in the CAD complex, (ii) the combined CAD+ASD complex has characteristics very similar to those of the two separate complexes CAD and ASD, and (iii) profiles from predicted structures (Figure 2C, solid blue, cyan, and magenta lines) closely match those from crystal structures (Figure 2C, dotted lines). Minor groove width profiles between bound and unbound DNA all follow the same trends in terms of minima versus maxima in groove width (Figure S4): the DNA intrinsically determines the shape that accommodates the protein. Convergence of the simulations is documented in Figures S6–S8. The CAD complex reported in Figure 2C is monomeric (half of the biological assembly), but we have verified in MD simulations that the dimeric form displays the same minor groove width profile (Figure S9).

Because the intrinsic DNA shape of the AT-rich sequence was observed to be important for the electrostatic potential-dependent recognition of αCTD by R265, we next asked whether mutation of R265 to alanine would affect this recognition mode. The R265A mutation does not induce major changes in the minor groove width profile by itself but noticeably increases the minor groove width at the CAD binding site when combined with the knockout DNA mutation (Figure 2C). In Figure 4, we show that structural changes upon protein mutation mainly involve residue 265 alone, with negligible effects on the rest of the αCTD DNA binding domain. The polar contact maps in Figure 4A (CAD_KO, top; CAD_KO_R265Aα1, bottom) were obtained with the DNAproDB tool.[34,35] The contact-labeled HB2, between an α-helix in αCTD[1] and the DNA, visible at the 2-o'clock position in the minor groove on the polar contact map of CAD_KO (Figure 4A, top panel), is lost in CAD_KO_R265Aα1 (Figure 4A, bottom panel). Contact maps reveal that this loss of contact is primarily due to the loss of the arginine side chain alone. Figure 4B illustrates the difference in partial residue–residue contact maps between two structures: CAD_R265Aα1 versus CAD in the top panel and CAD_KO_R265Aα1 versus CAD_KO in the bottom panel (absolute distance contact maps of CAD_R265Aα1 and CAD are reported for reference in Figure S10). These maps are focused on the binding domain of the complexes: the binding helix (BH), the binding loop (BL), the T_6-tract plus one base at each edge (T_6+), and the complementary A_6-tract plus one base at each edge (A_6+). The maps report the minimum inter-residue distance over all atom–atom distances for each residue pair, averaged (details in Table S1 and related text). Residue 265 is more distant from the DNA in the complex with the mutated protein than in the complex with the native protein, which reflects a loss of binding strength, both with and without the knockout DNA mutation. All other residues in the binding domain are unaffected.

DISCUSSION

In the major groove, sequence recognition is dominated by base readout: the four possible base pairs can be distinguished by their chemical patterns of functional groups in the major groove.[45,46] However, in the minor groove, direct contacts cannot distinguish A/T versus T/A base pairs, or G/C versus C/G base pairs, because of degeneracy in the pattern of functional groups.[47] Shape readout of sequence-dependent flexibility and/or structural features can also be used in sequence recognition.[45,46,48]

This study confirms that the αCTD domains of RNAP recognize the specific minor groove geometry generated by the A-tract sequence of the UP element resulting in a DNA shape-dependent electrostatic potential. The observed narrowing of the minor groove correlates with the enhanced negative electrostatic potential (see Figure S3). These biophysical effects are sequence-dependent; thus, the shape readout mechanism that exploits the negative electrostatic potential is intrinsic to the AT-rich sequence. We previously reported similar observations for transcription factors.[3,9,49−52] The importance of DNA shape readout has been demonstrated in quantitative predictions of DNA binding specificities for diverse families of transcription factors.[2,33] However, observations of shape readout for other DNA binding proteins have been sparse or lacking.[53] Here, we can generalize the use of DNA shape readout to related bacterial RNA polymerases for which the biophysical recognition mechanism was previously unknown. Our findings reveal shape readout as a recognition mechanism that plays an important role in the DNA binding by *E. coli* RNAP.

While multiple residues of αCTD form direct contacts to backbone phosphates at the edge of the narrow minor groove, just one residue, Arg265, specifically recognizes the shape-dependent minor groove negative electrostatic potential of the A_6-tract, through insertion of the planar, positively charged guanidinium group directly into the groove. This residue is known to be crucial to RNAP function: the single amino acid substitution Arg265 to Ala eliminates UP element binding by RNAP to the same extent as deleting the entire αCTD.[54] In particular, Gaal and co-workers showed that the R265A mutation reduced UP element-dependent transcription levels 15-fold,[55] suggesting that this substitution results in a complete loss of UP element dependence. Here, we have shown that αCTD undergoes a slight deformation that permits the Arg265 side chain to more deeply insert into the minor groove. Comparison of MD simulations from CAD with native αCTD and R265A αCTD confirms that contact between αCTD and the minor groove is weakened by the R265A mutation.

https://dx.doi.org/10.1021/acs.biochem.0c00571
Biochemistry XXXX, XXX, XXX−XXX

Murakami and co-workers showed that the only permissible substitution at position 265 supporting CAP-dependent transcription activation is another positively charged residue (Lys), yet no substitution, not even R265K, is tolerated in UP element-dependent transcription.[56] Given that αCTD's overall position relative to DNA is fixed through multiple interactions with the DNA backbone, a Lys amino nitrogen at position 265 would not be able to reach as deeply into the minor groove as an Arg guanidinium group, even with the modest deformation we observed between bound and free αCTD. The partial electrostatic complementarity offered by Lys is apparently insufficient to attract αCTD to transcription activation sites lacking additional determinants (such as CAP). Our results provide further support for the authors' conclusion[56] that not only positive charge but also side-chain structure plays a critical role in the αCTD UP element interaction. An additional consideration is the desolvation cost that is higher for lysine than for arginine due to the different Born radii of their side chains' head groups where the positive charge resides. This difference in desolvation makes it energetically more favorable for arginine to insert deeply into the minor groove.[9]

The narrow minor groove of A-tract DNA results from negative propeller twist angles of A/T base pairs that is stabilized by inter-base pair hydrogen bonds in the major groove, as well as a characteristic spine of hydration in the minor groove.[42,43] The original 3.1 Å CAD structure reported several water molecules positioned near the αCTD R265 guanidinium and with placement consistent with a spine of hydration in the narrow minor groove; however, confidence in water placement was limited because of the low diffraction limit.[16] Because of an apparent hard diffraction limit of ∼3 Å for the CAD crystal system, we are unfortunately unable to provide further insight into the role of water in A-tract DNA recognition, although a few water molecules conservatively placed in the updated structure have positions consistent with a spine of hydration.

We report here the observation of the same tandem complex of two αCTDs bound to A_6-tract DNA in two completely different crystal contexts. Until recently, we were unaware of any biochemical, genetic, or structural evidence supporting a biological role for cooperatively bound αCTDs within a single A_6-tract. In the context of either a proximal UP element or *lac* (class I CAP-dependent) promoter, only one αCTD is required for activity,[5,57,58] and either of the two αCTDs of promoter-bound RNAP may function in this role.[5] In addition, in the class I CAP-dependent promoter, only the αCTD^1 position observed in the CAD and ASD crystal structures is accessible to promoter-bound RNAP, while the αCTD^2 orientation or position is stereochemically inaccessible, because of the short (16-residue) flexible linker between the α-subunit N- and C-terminal domains.[10,15] A reviewer kindly alerted us to a new 4.1 Å cryo-EM structure of *E. coli* RNAP bound to the *rrnB* P1 promoter in a closed complex (PDB entry 6WR6; EMD-21879). In this structure, the two RNAP αCTDs are observed to bind in tandem to the promoter UP element using the same geometry as in the CAD and ASD crystal structures (K. S. Murakami, personal communication). This structure shows that tandem αCTD binding can occur within an intact transcription complex and therefore may have a biological role in activation of rRNA transcription. Tandem αCTD binding to DNA might also occur under certain other situations such as initial recruitment of RNAP prior to promoter binding or recruitment of a second RNAP following promoter binding.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.0c00571.

Supplementary Methods and Materials (text with description of MD and MC protocols), clustering analysis for trajectories (Table S1), RMSD between bound and unbound DNA derived from MD simulations (Table S2), CAD representative electron density (Figure S1), residual Cα atom distances of DNA-bound versus free αCTD coordinates following pairwise least-squares fits (Figure S2), correlation between minor groove width and electrostatic potential in CAD and ASD co-crystal structures (Figure S3), DNA shape of bound versus unbound DNA from MD simulations (Figure S4), shortening the A_6-tract that widens the minor groove (MC trajectories) (Figure S5), convergence of MD simulations that demonstrates the stability of trajectories (Figure S6), averaged minor groove width (MGW) profiles (Figure S7), stability of the CAD_KO complex over 300 ns supports low-resolution crystal structure (Figure S8), dimerization that does not affect DNA shape in the MD simulation of the CAD complex (Figure S9), residue−residue distance contact maps from which the difference contact map of Figure 4B (top panel) was derived (Figure S10), and supporting references (PDF)

Accession Codes

UniProt accession codes P0A7Z4 (RNA polymerase α subunit C-terminal domain), P0ACJ8 (catabolite activator protein), and Q9EZJ8 (RNA polymerase σ subunit region 4). PDB entries 3N4M (CAD), 5CIZ (CAD-KO), and 3N97 (ASD).

■ AUTHOR INFORMATION

Corresponding Authors

Catherine L. Lawson − *Department of Chemistry and Chemical Biology and Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-3261-7035; Phone: 848-445-5494; Email: cathy.lawson@rutgers.edu; Fax: 732-445-4320

Remo Rohs − *Quantitative and Computational Biology, Department of Biological Sciences, Department of Physics and Astronomy, Department of Chemistry, and Department of Computer Science, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0003-1752-1884; Phone: 213-740-0552; Email: rohs@usc.edu; Fax: 213-740-8631

Authors

Samuel Lara-Gonzalez − *Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-3313-0165

Ana Carolina Dantas Machado − *Quantitative and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, United States*

Satyanarayan Rao − *Quantitative and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0001-7678-3505

Andrew A. Napoli − *Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States*

Jens Birktoft − *Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States*

Rosa Di Felice − *Quantitative and Computational Biology, Department of Biological Sciences and Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, United States; CNR-NANO Modena, 41125 Modena, Italy;* orcid.org/0000-0002-7772-6550

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.0c00571

Funding

This work was supported by the National Institutes of Health (Grant R01GM021589 to C.L.L. and Grants R01GM106056 and R35GM130376 to R.R.). R.D.F. was supported by the University of Southern California WiSE program.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Brian Hudson for assistance with sample preparation and data collection, Seth Darst and Richard Ebright for providing the the σR4 expression plasmid, and Helen Berman, Richard Ebright, and Barry Honig for discussions and helpful suggestions. X-ray diffraction data were collected at Brookhaven National Laboratory National Synchrotron Light Source beamlines X25 and X29.

■ ABBREVIATIONS

CAD, CAP−αCTD−DNA complex; CAD-KO, CAP−αCTD−DNA^{KO} complex (with the A-tract shortened by 1 bp); ASD, CAP−sR4−DNA complex; CAD+ASD, complex obtained by merging CAD and ASD coordinates; CAD_R265Aα1, CAD complex with mutation R265A in αCTD^{1}; CAD_R265Aα2, CAD complex with mutation R265A in αCTD^{2}; CAD-KO_R265Aα1, CAP−αCTD−DNA^{KO} complex with the R265A mutation in αCTD^{1}; CAD-KO_R265Aα1_noCAP, CAP−αCTD−DNA^{KO} complex with the R265A mutation in αCTD^{1} and CAP coordinates removed; RMSD, root-mean-square deviation.

■ REFERENCES

(1) Ebright, R. H., and Busby, S. (1995) The *Escherichia coli* RNA polymerase alpha subunit: structure and function. *Curr. Opin. Genet. Dev. 5*, 197−203.

(2) Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H. J., Rohs, R., and Mann, R. S. (2015) Deconvolving the recognition of DNA shape from sequence. *Cell 161*, 307−318.

(3) Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordan, R., and Rohs, R. (2014) Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci. 39*, 381−399.

(4) Mathelier, A., Xin, B., Chiu, T. P., Yang, L., Rohs, R., and Wasserman, W. W. (2016) DNA shape features improve transcription factor binding site predictions *in vivo*. *Cell Syst. 3*, 278.

(5) Estrem, S. T., Ross, W., Gaal, T., Chen, Z. W., Niu, W., Ebright, R. H., and Gourse, R. L. (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev. 13*, 2134−2147.

(6) Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science 262*, 1407−1413.

(7) Gourse, R. L., Ross, W., and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol. Microbiol. 37*, 687−695.

(8) Busby, S., and Ebright, R. H. (1999) Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol. 293*, 199−213.

(9) Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. (2009) The role of DNA shape in protein-DNA recognition. *Nature 461*, 1248−1253.

(10) Hudson, B. P., Quispe, J., Lara-Gonzalez, S., Kim, Y., Berman, H. M., Arnold, E., Ebright, R. H., and Lawson, C. L. (2009) Three-dimensional EM structure of an intact activator-dependent transcription initiation complex. *Proc. Natl. Acad. Sci. U. S. A. 106*, 19830−19835.

(11) Liu, B., Hong, C., Huang, R. K., Yu, Z., and Steitz, T. A. (2017) Structural basis of bacterial transcription activation. *Science 358*, 947−951.

(12) Zhang, X. P., Gunasekera, A., Ebright, Y. W., and Ebright, R. H. (1991) Derivatives of CAP having no solvent-accessible cysteine residues, or having a unique solvent-accessible cysteine residue at amino acid 2 of the helix-turn-helix motif. *J. Biomol. Struct. Dyn. 9*, 463−473.

(13) Lara-Gonzalez, S., Birktoft, J. J., and Lawson, C. L. (2010) Structure of the *Escherichia coli* RNA polymerase alpha subunit C-terminal domain. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 806−812.

(14) Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L., and Darst, S. A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell 9*, 527−539.

(15) Lawson, C. L., Swigon, D., Murakami, K. S., Darst, S. A., Berman, H. M., and Ebright, R. H. (2004) Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol. 14*, 10−20.

(16) Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y. W., Berman, H. M., and Ebright, R. H. (2002) Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science 297*, 1562−1566.

(17) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658−674.

(18) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 213−221.

(19) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126−2132.

(20) Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.

(21) Delano, W. L. (2002) *The Pymol Molecular Graphics System*, DeLano Scientific, Palo Alto, CA.

(22) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF

Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605−1612.
(23) Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol. 372*, 774−797.
(24) Lavery, R., and Sklenar, H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn. 6*, 655−667.
(25) Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., and Honig, B. (2002) Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem. 23*, 128−137.
(26) Honig, B., and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science 268*, 1144−1149.
(27) Petrey, D., and Honig, B. (2003) GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. *Methods Enzymol. 374*, 492−509.
(28) Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput. 4*, 435−447.
(29) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet. 65*, 712−725.
(30) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys. 79*, 926−935.
(31) Rohs, R., Sklenar, H., and Shakked, Z. (2005) Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of papillomavirus E2-DNA binding sites. *Structure 13*, 1499−1509.
(32) Zhou, T., Yang, L., Lu, Y., Dror, I., Dantas Machado, A. C., Ghane, T., Di Felice, R., and Rohs, R. (2013) DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res. 41*, W56−62.
(33) Zhou, T., Shen, N., Yang, L., Abe, N., Horton, J., Mann, R. S., Bussemaker, H. J., Gordan, R., and Rohs, R. (2015) Quantitative modeling of transcription factor binding specificities using DNA shape. *Proc. Natl. Acad. Sci. U. S. A. 112*, 4654−4659.
(34) Sagendorf, J. M., Berman, H. M., and Rohs, R. (2017) DNAproDB: an interactive tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res. 45*, W89−W97.
(35) Sagendorf, J. M., Markarian, N., Berman, H. M., and Rohs, R. (2020) DNAproDB: an expanded database and web-based tool for structural analysis of DNA−protein complexes. *Nucleic Acids Res. 48*, D277−D287.
(36) Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. H., and Berman, H. M. (1996) Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. *J. Mol. Biol. 260*, 395−408.
(37) Napoli, A. A., Lawson, C. L., Ebright, R. H., and Berman, H. M. (2006) Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: Recognition of pyrimidine-purine and purine-purine steps. *J. Mol. Biol. 357*, 173−183.
(38) Geszvain, K., Gruber, T. M., Mooney, R. A., Gross, C. A., and Landick, R. (2004) A hydrophobic patch on the flap-tip helix of *E. coli* RNA polymerase mediates sigma(70) region 4 function. *J. Mol. Biol. 343*, 569−587.
(39) Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science 296*, 1285−1290.
(40) Chen, H., Tang, H., and Ebright, R. H. (2003) Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription. *Mol. Cell 11*, 1621−1633.
(41) Ross, W., Schneider, D. A., Paul, B. J., Mertens, A., and Gourse, R. L. (2003) An intersubunit contact stimulating transcription initiation by E coli RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4. *Genes Dev. 17*, 1293−1307.
(42) Nelson, H. C., Finch, J. T., Luisi, B. F., and Klug, A. (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature 330*, 221−226.
(43) DiGabriele, A. D., and Steitz, T. A. (1993) A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol. 231*, 1024−1039.
(44) Dantas Machado, A. C., Saleebyan, S. B., Holmes, B. T., Karelina, M., Tam, J., Kim, S. Y., Kim, K. H., Dror, I., Hodis, E., Martz, E., Compeau, P. A., and Rohs, R. (2012) Proteopedia: 3D visualization and annotation of transcription factor-DNA readout modes. *Biochem. Mol. Biol. Educ. 40*, 400−401.
(45) Lawson, C. L., and Berman, H. M. (2008) Indirect Readout of DNA Sequence by Proteins. In *Protein-Nucleic Acid Interactions* (Rice, P. A., and Correll, C. C., Eds.) pp 66−90, RSC Press, Cambridge, U.K.
(46) Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., and Mann, R. S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem. 79*, 233−269.
(47) Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. U. S. A. 73*, 804−808.
(48) Locasale, J. W., Napoli, A. A., Chen, S., Berman, H. M., and Lawson, C. L. (2009) Signatures of protein-DNA recognition in free DNA binding sites. *J. Mol. Biol. 386*, 1054−1065.
(49) Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., Jacob, V., Aggarwal, A. K., Honig, B., and Mann, R. S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell 131*, 530−543.
(50) Kitayner, M., Rozenberg, H., Rohs, R., Suad, O., Rabinovich, D., Honig, B., and Shakked, Z. (2010) Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol. 17*, 423−429.
(51) Yang, L., Zhou, T., Dror, I., Mathelier, A., Wasserman, W. W., Gordan, R., and Rohs, R. (2014) TFBSshape: a motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Res. 42*, D148−155.
(52) Yang, L., Orenstein, Y., Jolma, A., Yin, Y., Taipale, J., Shamir, R., and Rohs, R. (2017) Transcription factor family-specific DNA shape readout revealed by quantitative specificity models. *Mol. Syst. Biol. 13*, 910.
(53) Chang, Y. P., Xu, M., Dantas Machado, A. C., Yu, X. J., Rohs, R., and Chen, X. S. (2013) Mechanism of origin DNA recognition and assembly of an initiator-helicase complex by SV40 large tumor antigen. *Cell Rep. 3*, 1117−1127.
(54) Ross, W., and Gourse, R. L. (2005) Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate *Escherichia coli* RNA polymerase-lacUV5 promoter association. *Proc. Natl. Acad. Sci. U. S. A. 102*, 291−296.
(55) Gaal, T., Ross, W., Blatter, E. E., Tang, H., Jia, X., Krishnan, V. V., Assa-Munt, N., Ebright, R. H., and Gourse, R. L. (1996) DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture. *Genes Dev. 10*, 16−26.
(56) Murakami, K., Fujita, N., and Ishihama, A. (1996) Transcription factor recognition surface on the RNA polymerase alpha subunit is involved in contact with the DNA enhancer element. *EMBO J. 15*, 4358−4367.
(57) Murakami, K., Kimura, M., Owens, J. T., Meares, C. F., and Ishihama, A. (1997) The two alpha subunits of *Escherichia coli* RNA polymerase are asymmetrically arranged and contact different halves of the DNA upstream element. *Proc. Natl. Acad. Sci. U. S. A. 94*, 1709−1714.
(58) Murakami, K., Owens, J. T., Belyaeva, T. A., Meares, C. F., Busby, S. J., and Ishihama, A. (1997) Positioning of two alpha subunit carboxy-terminal domains of RNA polymerase at promoters by two transcription factors. *Proc. Natl. Acad. Sci. U. S. A. 94*, 11274−11278.

https://dx.doi.org/10.1021/acs.biochem.0c00571
Biochemistry XXXX, XXX, XXX−XXX