Published in final edited form as:

Biochemistry. 2020 December 08; 59(48): 4523–4532. doi:10.1021/acs.biochem.0c00571.

RNA polymerase alpha subunit recognizes the DNA shape of the Upstream Promoter element

Samuel Lara-Gonzalez[1,§], **Ana Carolina Dantas Machado**[2,§], **Satyanarayan Rao**[2,§], **Andrew A. Napoli**[1], **Jens Birktoft**[1], **Rosa Di Felice**[2,3,4], **Remo Rohs**[2,3,5,6,*], **Catherine L. Lawson**[1,7,*]

[1]Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, 610 Taylor Road, Piscataway, NJ 08854, USA

[2]Quantitative and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA

[3]Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089, USA

[4]CNR-NANO Modena, Via Campi 213/A, 41125 Modena, Italy

[5]Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA

[6]Department of Computer Science, University of Southern California, Los Angeles, CA 90089, USA

[7]Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, 174 Frelinghuysen Road, Piscataway, NJ 08854, USA

Abstract

[*]**Corresponding Authors:** To whom correspondence should be addressed. Tel: 848-445-5494; Fax: 732-445-4320; cathy.lawson@rutgers.edu, Correspondence may also be addressed to Remo Rohs. Tel: 213-740-0552; Fax: 213-740-8631; rohs@usc.edu.

[§]**Present Addresses:** Samuel Lara-Gonzalez, IPICYT, Instituto Potosino de Investigación Científica y Tecnológica A.C., División de Biología Molecular, Camino a la Presa San José 2055, Lomas 4a. Sección 78216 San Luis Potosí, S.L.P. México

[§]**Present Addresses:** Ana Carolina Dantas Machado, Department of Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

[§]**Present Addresses:** Satyanarayan Rao, Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA

Author Contributions

The manuscript was written with contributions by all authors. All authors have given approval to the final version of the manuscript.

Conflict of Interest Statement. None declared.

Supplementary Methods and Materials: text with description of MD and MC protocols. Table S1: clustering analysis for trajectories. Table S2: Root mean square deviations (RMSD) between bound and unbound DNA derived from MD simulations Figure S1: CAD representative electron density. Figure S2: Residual Cα atom distances of DNA-bound vs. free αCTD coordinates following pairwise least-squares fits. Figure S3: Correlation between minor groove width and electrostatic potential in CAD and ASD co-crystal structures. Figure S4: DNA shape of bound versus unbound DNA from MD simulations. Figure S5: Shortening the A_6-tract widens the minor groove (MC trajectories). Figure S6: Convergence of MD simulations demonstrates stability of trajectories. Figure S7: Averaged minor groove width (MGW) profiles. Figure S8: Stability of the CAD_KO complex over 300 ns supports low-resolution crystal structure. Figure S9: Dimerization does not affect DNA shape in the MD simulation of the CAD complex. Figure S10: Residue-residue distance contact maps from which the difference contact map of Figure 4B (top panel) was derived. Supporting references.

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 accession codes: 3N4M (CAD), 5CIZ (CAD-KO), and 3N97 (ASD).

We demonstrate here that the alpha subunit C-terminal domain of *E. 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 of six base pairs 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 generic DNA. Shortening the A-tract by one base-pair 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 "knock-out" 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 DNA shape in sequence-dependent recognition of DNA by RNA polymerase is comparable to certain transcription factors.

Keywords

RNA polymerase; DNA binding; mechanisms; structure determination; molecular simulations

INTRODUCTION.

The *E. 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 increasing knowledge about 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 'AAAAAARNR') can also function alone, stimulating transcription up to 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 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 physiologic ionic strength. In addition, we describe a third crystal structure with the A-tract shortened by one base pair. The "knock-out" 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 in Zhang *et al.* (12) and further purified on a Heparin 16/10 FF column (GE Healthcare). CAP was then concentrated and buffer exchanged to 0.3 mM dimer, as determined by Bradford assay using a bovine serum albumin standard, in 10 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1 mM EDTA. αCTD was expressed and purified as described in Lara-Gonzalez *et al.* (13). σR4 was prepared based on the expression system for *T. 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% NaN3. 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 two-fold symmetric CAD and CAD-KO DNA duplexes were prepared as described in Benoff *et al.* (16). The ASD DNA duplex was assembled from complementary strands 5'-TGGAAAAAAGTACTTGACATGG-3' and 5'-CCATGTCAAGTACTTTTTTCC-3', yielding 22 base pairs with a 5' T overhang at one end. C-18 dual reverse-phase HPLC 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 concentrations of 0.5–1.0 mM duplex. The duplexes were annealed by heating at 90 °C for 5 minutes, and then slowly cooled to 25 °C over a period of 12 to 15 hours.

Crystal structure determination

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 reservoir solution 28% (w/v) PEG4000, 0.2 M ammonium acetate, 0.01 M sarcosine, 0.1 M sodium citrate, pH 5.6. A single crystal was soaked in cryoprotectant solution composed of reservoir solution plus 14% glycerol (see also Hudson *et al.* (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), with coordinates from previous crystal structures with PDB IDs 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 non-crystallographic 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 model fit to strongly anisotropic diffraction data. For ASD, additional pseudo-bond 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 since 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 shown in Table 1.

Relative αCTD occupancies were estimated using group occupancy refinement with CNS v.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 against observed diffraction data in the resolution range 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.

Structure visualization and analysis

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

We used Curves 5.3 (24) to calculate minor groove width as a function of nucleotide sequence. We calculated the electrostatic potential in reference points at the center of the minor groove in the approximate plane of each base pair (9). We used DelPhi (25) for non-linear Poisson-Boltzmann calculations (26) of the electrostatic potential at the physiologic ionic strength of 0.145 M. DelPhi is a numerical approach to compute 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 non-linear Poisson-Boltzmann equation and a previously described numerical protocol (9). We plotted minor groove width and electrostatic potential 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

We predicted bound and unbound DNA structures 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 based on Amber94) with refined dihedral parameters to reduce structural transitions in DNA oligomers to non-canonical 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 three crystal structures: 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, Methods and Materials) as an independent validation of MD data (31).

We computed the DNA shape with Curves 5.3 (24) for regular MD snapshots every 10 ps, using a computational toolkit, Trj2Shape, that can process the entire trajectory in high-throughput manner (Supporting Information, 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 under accession codes 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 multi-protein–DNA assemblies that share a common feature: a pair of αCTDs bound to A_6-tract DNA (Figure 1).

In the two-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 similar diffraction limit (Table 1, compare columns A and B; extensive efforts to obtain higher resolution diffraction data for CAD were unsuccessful, likely owing to 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, 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 towards CAP (Fig. 1C, thick rectangles(36)), the new structure has a somewhat larger roll angle (52°) than 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 is the only major difference that we observe between the original structure and this new determination; all protein–protein and protein–DNA interfaces described by Benoff 2002 (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 the sigma subunit region 4 (σR4) bound (Figure 1B,D; since *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 (RMSD for all equivalent Cα atoms in each σR4 dimer is 0.6 Å) were observed. Only one copy forms an interface with the −35 element DNA (σR4[1]; purple footprint in Fig. 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 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; 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 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 two base pairs apart from each other (approximate rotation of 70° and rise of 6.8 Å with respect to the DNA helix axis). The small protein–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 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, Leu295, and DNA backbone phosphate oxygen. The equivalent site at αCTD^2 is coordinated by water. 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 vs. free αCTD conformations were expected to be closely similar. Overall, RMSDs are ~0.7 Å for pairwise C_α superpositions. A modest and remarkably consistent bound vs. 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 sidechain 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 about 3 Å, compared to 5.8 Å for canonical B-DNA (Figure 2 and Figure S3). This minor groove narrowing generates a region with enhanced negative electrostatic potential of about −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 highlight in Figure 2B, underlined in Figure 2C) with C/G, reducing the A-tract length to five base pairs. At an UP element proximal site, this substitution reduces transcriptional activity, but only by about 30% (5). A structure for this ‘knockout’ complex (‘CAD-KO’) was obtained based on crystal diffraction data 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). Reduced 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, as well as the corresponding unbound DNA duplexes, and compared the result with the known crystal structures (Figure 2C, Figures S4 and S5). The KO mutation is two base pairs distant 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 solution state (Figure 2C, blue solid line), as well as in two previously determined unbound DNA crystal structures with sequences CGCGAAAAAACG (PDB ID 1D89) and CGCAAAAAAGCG (PDB ID 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 knock-out 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 very similar characteristics as the two separate complexes CAD and ASD; (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 vs. maxima (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).

Since 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 knock-out 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, 45). 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 of partial residue–residue contact maps between two structures: CAD_R265Aα1 versus CAD in the top panel, 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 the 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 knock-out 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 (46, 47). However, in the minor groove, direct contacts cannot distinguish A/T versus T/A base pairs, nor G/C versus C/G base pairs, because of degeneracy in the pattern of functional groups (48). Shape readout of sequence-dependent flexibility and/or structural features can also be used in sequence recognition (46, 47, 49).

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 A-rich sequence. We previously reported similar observations for transcription factors (3, 9, 50–53). 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 (54). 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 guanidium 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 (55). In particular, Gaal and coworkers showed that the R265A mutation reduced UP-element dependent transcription levels 15-fold (56), 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 sidechain to more deeply insert into the minor groove. Comparison of MD simulations from CAD with native αCTD and R265A αCTD confirm that contact between αCTD and the minor groove is weakened by the R265A mutation.

Murakami and coworkers 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 (57). 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 (57) that not only positive charge, but 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 compared to arginine due to the different Born radii of their sidechains' 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 owing to the low diffraction limit (16). Owing to 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, 58, 59), and either of the two αCTDs of promoter-bound RNAP may function in this role (5). In addition, only the αCTD^1 position observed in the CAD/ASD crystal structures is accessible to promoter-bound RNAP, while the αCTD^2 orientation/position is stereochemically inaccessible, owing to 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 ID 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, private 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 ribosomal RNA 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

We thank Brian Hudson for assistance with sample preparation and data collection, Seth Darst and Richard Ebright for providing the σR4 expression plasmid, 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.

Funding Sources

This work was supported by the National Institutes of Health [R01GM021589 to C.L.L.; R01GM106056 and R35GM130376 to R.R.]. R.D.F. was supported by the USC WiSE program.

ABBREVIATIONS

CAD	CAP-αCTD-DNA complex
CAD-KO	CAP-αCTD-DNA^{KO} complex (with the A-tract shortened by one base-pair)
ASD	CAP-sR4-DNA complex
CAD+ASD	structure obtained by merging CAD and ASD
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

REFERENCES

[1]. Ebright RH; Busby S (1995) The Escherichia coli RNA polymerase alpha subunit: structure and function, Current opinion in genetics & development 5, 197–203. [PubMed: 7613089]

[2]. Abe N; Dror I; Yang L; Slattery M; Zhou T; Bussemaker HJ; Rohs R; Mann RS (2015) Deconvolving the recognition of DNA shape from sequence, Cell 161, 307–318. [PubMed: 25843630]

[3]. Slattery M; Zhou T; Yang L; Dantas Machado AC; Gordan R; Rohs R (2014) Absence of a simple code: how transcription factors read the genome, Trends in biochemical sciences 39, 381–399. [PubMed: 25129887]

[4]. Mathelier A; Xin B; Chiu TP; Yang L; Rohs R; Wasserman WW (2016) DNA Shape Features Improve Transcription Factor Binding Site Predictions In Vivo, Cell Syst 3, 278–286 e274. [PubMed: 27546793]

[5]. Estrem ST; Ross W; Gaal T; Chen ZW; Niu W; Ebright RH; Gourse RL (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit, Genes & development 13, 2134–2147. [PubMed: 10465790]

[6]. Ross W; Gosink KK; Salomon J; Igarashi K; Zou C; Ishihama A; Severinov K; Gourse RL (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase, Science 262, 1407–1413. [PubMed: 8248780]

[7]. Gourse RL; Ross W; Gaal T (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition, Molecular microbiology 37, 687–695. [PubMed: 10972792]

[8]. Busby S; Ebright RH (1999) Transcription activation by catabolite activator protein (CAP), Journal of molecular biology 293, 199–213. [PubMed: 10550204]

[9]. Rohs R; West SM; Sosinsky A; Liu P; Mann RS; Honig B (2009) The role of DNA shape in protein-DNA recognition, Nature 461, 1248–1253. [PubMed: 19865164]

[10]. Hudson BP; Quispe J; Lara-Gonzalez S; Kim Y; Berman HM; Arnold E; Ebright RH; Lawson CL (2009) Three-dimensional EM structure of an intact activator-dependent transcription initiation complex, Proceedings of the National Academy of Sciences of the United States of America 106, 19830–19835. [PubMed: 19903881]

[11]. Liu B; Hong C; Huang RK; Yu Z; Steitz TA (2017) Structural basis of bacterial transcription activation, Science 358, 947–951. [PubMed: 29146813]

[12]. Zhang XP; Gunasekera A; Ebright YW; Ebright RH (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, Journal of biomolecular structure & dynamics 9, 463–473. [PubMed: 1667734]

[13]. Lara-Gonzalez S; Birktoft JJ; Lawson CL (2010) Structure of the Escherichia coli RNA polymerase alpha subunit C-terminal domain, Acta crystallographica. Section D, Biological crystallography 66, 806–812. [PubMed: 20606261]

[14]. Campbell EA; Muzzin O; Chlenov M; Sun JL; Olson CA; Weinman O; Trester-Zedlitz ML; Darst SA (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit, Molecular cell 9, 527–539. [PubMed: 11931761]

[15]. Lawson CL; Swigon D; Murakami KS; Darst SA; Berman HM; Ebright RH (2004) Catabolite activator protein: DNA binding and transcription activation, Current opinion in structural biology 14, 10–20. [PubMed: 15102444]

[16]. Benoff B; Yang H; Lawson CL; Parkinson G; Liu J; Blatter E; Ebright YW; Berman HM; Ebright RH (2002) Structural basis of transcription activation: the CAP-alpha CTD-DNA complex, Science 297, 1562–1566. [PubMed: 12202833]

[17]. McCoy AJ; Grosse-Kunstleve RW; Adams PD; Winn MD; Storoni LC; Read RJ (2007) Phaser crystallographic software, Journal of applied crystallography 40, 658–674. [PubMed: 19461840]

[18]. Adams PD; Afonine PV; Bunkoczi G; Chen VB; Davis IW; Echols N; Headd JJ; Hung LW; Kapral GJ; Grosse-Kunstleve RW; McCoy AJ; Moriarty NW; Oeffner R; Read RJ; Richardson DC; Richardson JS; Terwilliger TC; Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta crystallographica. Section D, Biological crystallography 66, 213–221. [PubMed: 20124702]

[19]. Emsley P; Cowtan K (2004) Coot: model-building tools for molecular graphics, Acta crystallographica. Section D, Biological crystallography 60, 2126–2132. [PubMed: 15572765]

[20]. Brunger AT; Adams PD; Clore GM; DeLano WL; Gros P; Grosse-Kunstleve RW; Jiang JS; Kuszewski J; Nilges M; Pannu NS; Read RJ; Rice LM; Simonson T; Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, Acta crystallographica. Section D, Biological crystallography 54, 905–921. [PubMed: 9757107]

[21]. Delano WL (2002) The Pymol Molecular Graphics System, Delano Scientific, Palo Alto.

[22]. Pettersen EF; Goddard TD; Huang CC; Couch GS; Greenblatt DM; Meng EC; Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis, J Comput Chem 25, 1605–1612. [PubMed: 15264254]

[23]. Krissinel E; Henrick K (2007) Inference of macromolecular assemblies from crystalline state, Journal of molecular biology 372, 774–797. [PubMed: 17681537]

[24]. Lavery R; Sklenar H (1989) Defining the structure of irregular nucleic acids: conventions and principles, Journal of biomolecular structure & dynamics 6, 655–667. [PubMed: 2619933]

[25]. Rocchia W; Sridharan S; Nicholls A; Alexov E; Chiabrera A; 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, Journal of computational chemistry 23, 128–137. [PubMed: 11913378]

[26]. Honig B; Nicholls A (1995) Classical electrostatics in biology and chemistry, Science 268, 1144–1149. [PubMed: 7761829]

[27]. Petrey D; Honig B (2003) GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences, Methods in enzymology 374, 492–509. [PubMed: 14696386]

[28]. Hess B; Kutzner C; van der Spoel D; Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, J Chem Theory Comput 4, 435–447. [PubMed: 26620784]

[29]. Hornak V; Abel R; Okur A; Strockbine B; Roitberg A; Simmerling C (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters, Proteins 65, 712–725. [PubMed: 16981200]

[30]. Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water, J Chem Phys 79, 926–935.

[31]. Rohs R; Sklenar H; 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. [PubMed: 16216581]

[32]. Zhou T; Yang L; Lu Y; Dror I; Dantas Machado AC; Ghane T; Di Felice R; Rohs R (2013) DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale, Nucleic acids research 41, W56–62. [PubMed: 23703209]

[33]. Zhou T; Shen N; Yang L; Abe N; Horton J; Mann RS; Bussemaker HJ; Gordan R; Rohs R (2015) Quantitative Modeling of Transcription Factor Binding Specificities Using DNA Shape, Proceedings of the National Academy of Sciences of the United States of America 112, 4654–4659. [PubMed: 25775564]

[34]. Sagendorf JM; Berman HM; Rohs R (2017) DNAproDB: an interactive tool for structural analysis of DNA-protein complexes, Nucleic acids research 45, W89–W97. [PubMed: 28431131]

[35]. Sagendorf JM; Markarian N; Berman HM; Rohs R (2020) DNAproDB: an expanded database and web-based tool for structural analysis of DNA–protein complexes Nucleic acids research 48, D277–D287. [PubMed: 31612957]

[36]. Parkinson G; Wilson C; Gunasekera A; Ebright YW; Ebright RH; Berman HM (1996) Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface, Journal of molecular biology 260, 395–408. [PubMed: 8757802]

[37]. Napoli AA; Lawson CL; Ebright RH; Berman HM (2006) Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: Recognition of pyrimidine-purine and purine-purine steps, Journal of molecular biology 357, 173–183. [PubMed: 16427082]

[38]. Geszvain K; Gruber TM; Mooney RA; Gross CA; Landick R (2004) A hydrophobic patch on the flap-tip helix of E.coli RNA polymerase mediates sigma(70) region 4 function, Journal of molecular biology 343, 569–587. [PubMed: 15465046]

[39]. Murakami KS; Masuda S; Campbell EA; Muzzin O; Darst SA (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex, Science 296, 1285–1290. [PubMed: 12016307]

[40]. Chen H; Tang H; Ebright RH (2003) Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription, Molecular cell 11, 1621–1633. [PubMed: 12820974]

[41]. Ross W; Schneider DA; Paul BJ; Mertens A; Gourse RL (2003) An intersubunit contact stimulating transcription initiation by E coli RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4, Genes & development 17, 1293–1307. [PubMed: 12756230]

[42]. Nelson HC; Finch JT; Luisi BF; Klug A (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications, Nature 330, 221–226. [PubMed: 3670410]

[43]. DiGabriele AD; Steitz TA (1993) A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend, Journal of molecular biology 231, 1024–1039. [PubMed: 8515463]

[44]. Dantas Machado AC; Saleebyan SB; Holmes BT; Karelina M; Tam J; Kim SY; Kim KH; Dror I; Hodis E; Martz E; Compeau PA; Rohs R (2012) Proteopedia: 3D visualization and annotation of transcription factor-DNA readout modes, Biochem. Mol. Biol. Educ 40, 400–401. [PubMed: 23166030]

[45]. Sagendorf JM; Markarian N; Berman HM; Rohs R (2020) DNAproDB: an expanded database and web-based tool for structural analysis of DNA-protein complexes, Nucleic acids research 48, D277–D287. [PubMed: 31612957]

[46]. Lawson CL; Berman HM (2008) Indirect Readout of DNA Sequence by Proteins, In Protein-Nucleic Acid Interactions (Rice PA, and Correll CC, Eds.), pp 66–90, RSC Press, Cambridge, UK.

[47]. Rohs R; Jin X; West SM; Joshi R; Honig B; Mann RS (2010) Origins of specificity in protein-DNA recognition, Annu Rev Biochem 79, 233–269. [PubMed: 20334529]

[48]. Seeman NC; Rosenberg JM; Rich A (1976) Sequence-specific recognition of double helical nucleic acids by proteins, Proceedings of the National Academy of Sciences of the United States of America 73, 804–808. [PubMed: 1062791]

[49]. Locasale JW; Napoli AA; Chen S; Berman HM; Lawson CL (2009) Signatures of protein-DNA recognition in free DNA binding sites, Journal of molecular biology 386, 1054–1065. [PubMed: 19244617]

[50]. Joshi R; Passner JM; Rohs R; Jain R; Sosinsky A; Crickmore MA; Jacob V; Aggarwal AK; Honig B; Mann RS (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure, Cell 131, 530–543. [PubMed: 17981120]

[51]. Kitayner M; Rozenberg H; Rohs R; Suad O; Rabinovich D; Honig B; Shakked Z (2010) Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs, Nat Struct Mol Biol 17, 423–429. [PubMed: 20364130]

[52]. Yang L; Zhou T; Dror I; Mathelier A; Wasserman WW; Gordan R; Rohs R (2014) TFBSshape: a motif database for DNA shape features of transcription factor binding sites, Nucleic acids research 42, D148–155. [PubMed: 24214955]

[53]. Yang L; Orenstein Y; Jolma A; Yin Y; Taipale J; Shamir R; Rohs R (2017) Transcription factor family-specific DNA shape readout revealed by quantitative specificity models, Mol Syst Biol 13, 910. [PubMed: 28167566]

[54]. Chang YP; Xu M; Machado AC; Yu XJ; Rohs R; Chen XS (2013) Mechanism of origin DNA recognition and assembly of an initiator-helicase complex by SV40 large tumor antigen, Cell reports 3, 1117–1127. [PubMed: 23545501]

[55]. Ross W; Gourse RL (2005) Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate Escherichia coli RNA polymerase-lacUV5 promoter association, Proceedings of the National Academy of Sciences of the United States of America 102, 291–296. [PubMed: 15626760]

[56]. Gaal T; Ross W; Blatter EE; Tang H; Jia X; Krishnan VV; Assa-Munt N; Ebright RH; Gourse RL (1996) DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture, Genes & development 10, 16–26. [PubMed: 8557191]

[57]. Murakami K; Fujita N; 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. [PubMed: 8861963]

[58]. Murakami K; Kimura M; Owens JT; Meares CF; Ishihama A (1997) The two alpha subunits of Escherichia coli RNA polymerase are asymmetrically arranged and contact different halves of the

DNA upstream element, Proceedings of the National Academy of Sciences of the United States of America 94, 1709–1714. [PubMed: 9050843]

[59]. Murakami K; Owens JT; Belyaeva TA; Meares CF; Busby SJ; Ishihama A (1997) Positioning of two alpha subunit carboxy-terminal domains of RNA polymerase at promoters by two transcription factors, Proceedings of the National Academy of Sciences of the United States of America 94, 11274–11278. [PubMed: 9326599]

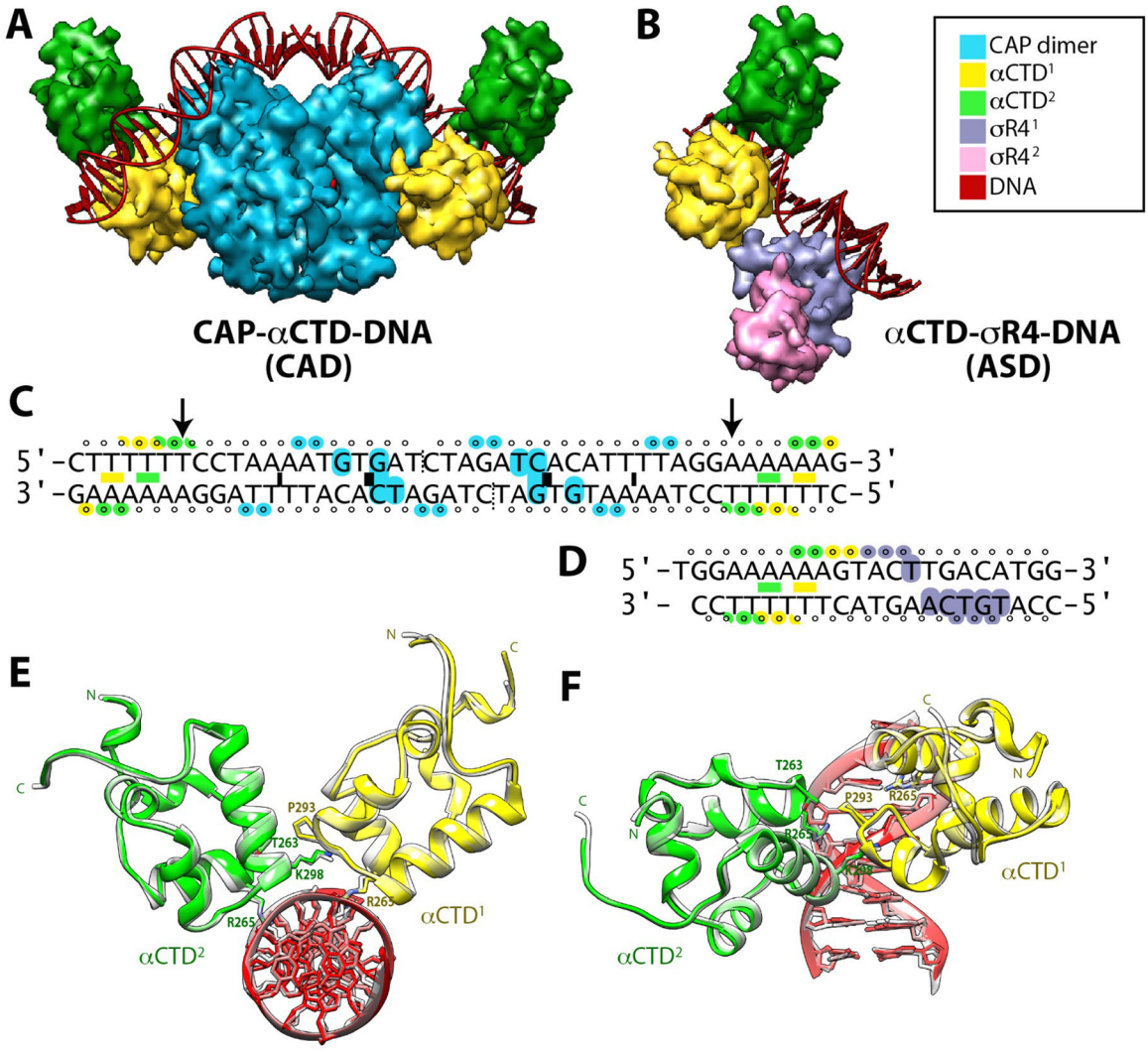

Figure 1.

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 two-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 (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 co-crystal structure. Nucleotide bases (letters) and backbone phosphates (open circles) are shaded according to the protein subunit making contact. Intercalation of the minor groove is indicated by shaded rectangles. In (C), thick and thin black rectangles represent 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: semi-transparent grey). αCTD residues that participate in αCTD-αCTD contacts (Pro293 of αCTD^1, Thr263 and Lys298 of αCTD^2) or that intercalate into the DNA minor groove (Arg265, both subunits) are also displayed. The RMSD for 1484 equivalent protein + DNA atom positions in the two structures is 0.7 Å.

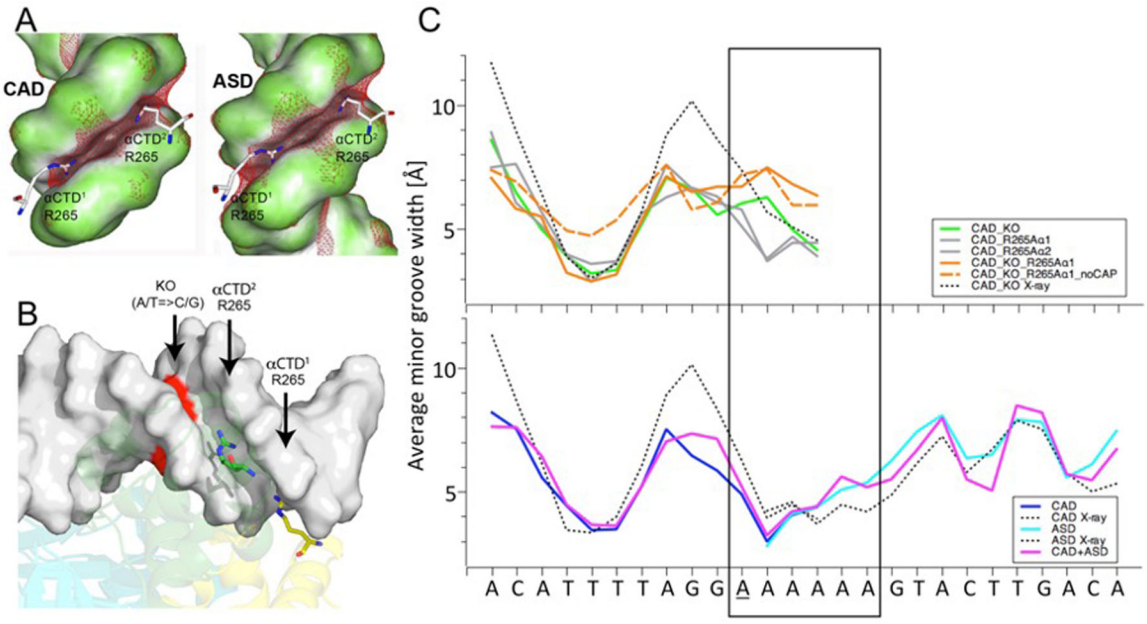

Figure 2.

Minor groove width of the A_6-tract 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) is shown color-coded by shape (convex surfaces in green, concave surfaces in dark gray). Darker gray shading corresponds to narrower minor groove. The Arg265 residues from each αCTD that insert into the minor groove are shown in stick model. The isopotential surface at −5kT/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 highlighted in 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 base-pair) region where the CAD X-ray and ASD X-ray derived curves overlap, the two can be distinguished by continuity of connected points.

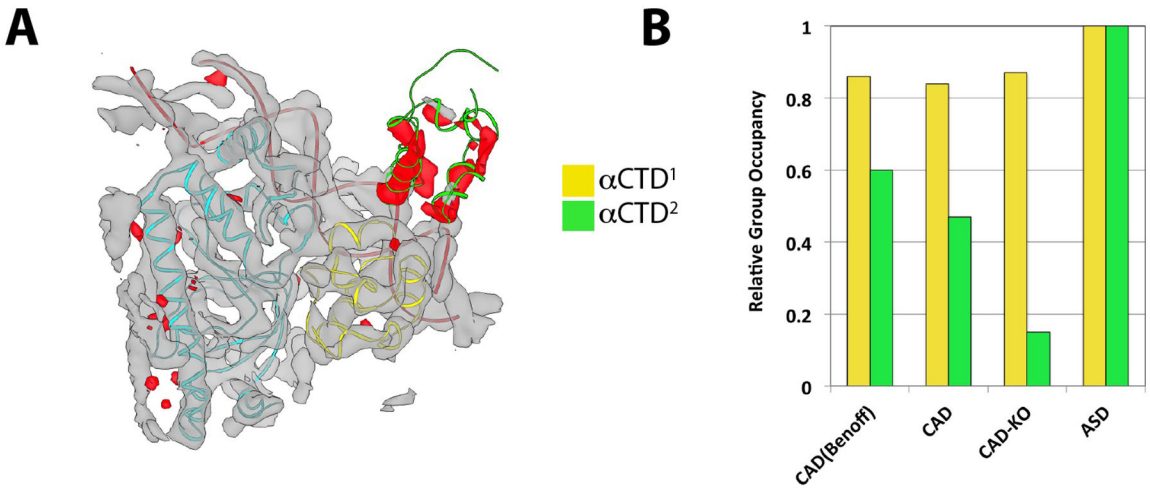

Figure 3.

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

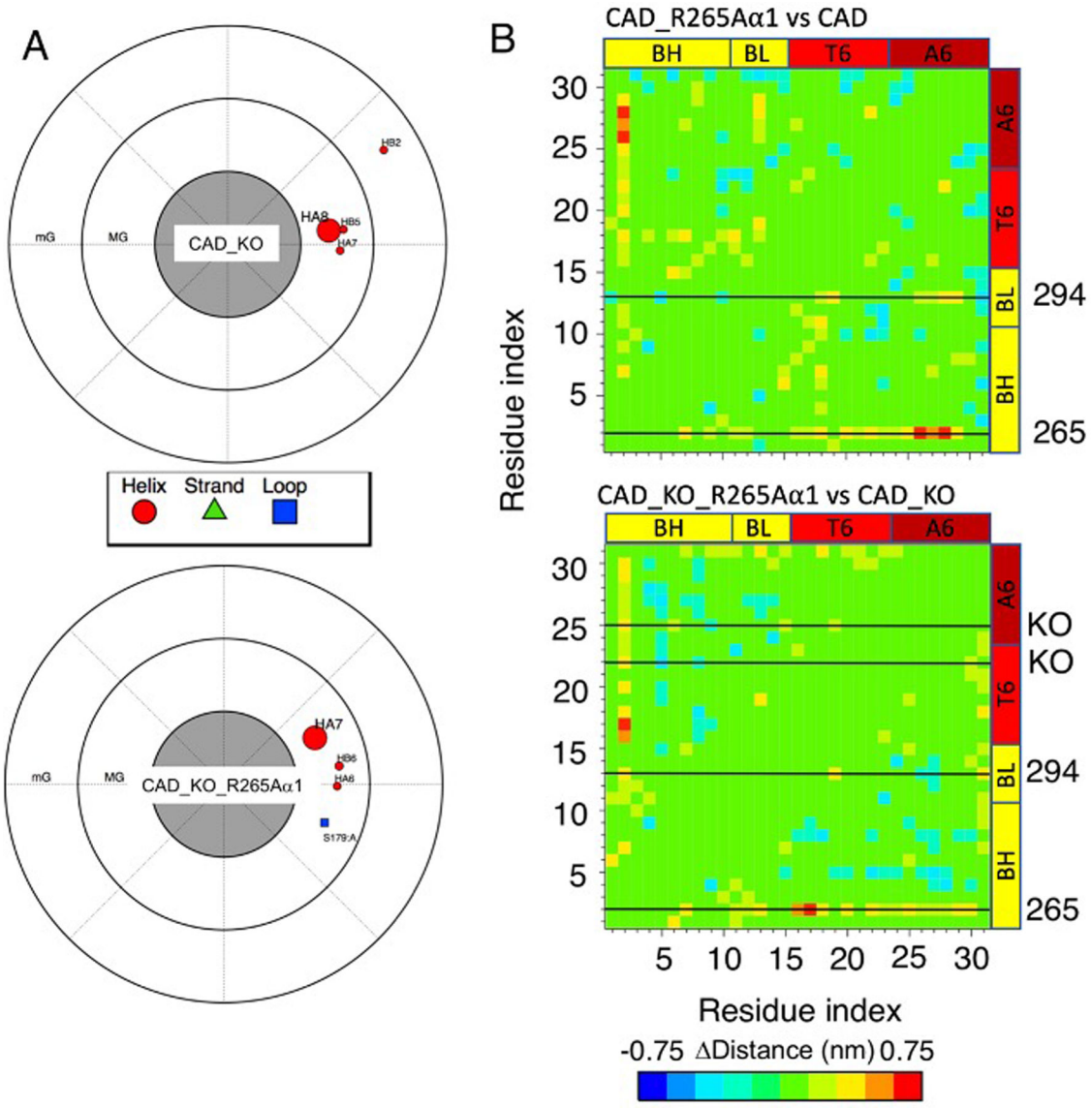

Figure 4.

Contact between αCTD[1] and the DNA minor groove is weakened upon mutation of Arg265 to Ala. (A) DNAproDB polar contact maps for CAD_KO (top) and for 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, 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 a residue–residue distance increases as an effect of the R265A protein mutation, in the absence (top) and presence (bottom) of the knock-out DNA mutation. Residue indices: binding helix BH (1–10), binding loop BL (11–15), T_6+-tract (16–23), 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 knock-out mutation when present (bottom), are marked by horizontal lines.

Table 1.

Crystallographic Data, Collection and Refinement Statistics.

STRUCTURE	A. CAP-αCTD-DNA (CAD)	B. CAP-αCTD-DNA (16)*	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	$P6_222$	$P6_222$	$P6_222$	$C222_1$
Unit cell *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
Unit cell α β γ (°)	90 90 120	90 90 120	90 90 120	90 90 90
Matthews coefficient A^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. measured reflections	129653	332877	124215	24506
No. 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}	.085	.076	.095	.081
Refinement resol 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)
Reflections working	29042 (2039)	21027 (1705)	6615 (1145)	7422 (2235)
Reflections free	1481 (108)	2304 (182)	663 (127)	480 (150)
non-hydrogen atoms	3737	3613	3099	2887
water oxygen atoms	47	32	0	0
Wilson B-value ($Å^2$)	100	82	207	99
Bond lengths (Å)	0.004	0.009	0.013	0.006
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 id	3N4M	1LB2	5CIZ	3N97

*Benoff et al. 2002 CAD structure determination statistics are provided for comparison