DNA Deformability as a Recognition Feature in the RevErb Response Element‡

Michael L. Sierk,§,|| Qiang Zhao,§ and Fraydoon Rastinejad*,§,||,⊥

Department of Pharmacology, X-ray Diffraction Laboratory, Interdisciplinary Program in Biophysics, and Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia 22908-0735

Received May 29, 2001; Revised Manuscript Received August 24, 2001

ABSTRACT: Most nuclear receptors recognize the same consensus hexameric sequence, AGGTCA. An important question has been how the various members of this transcription factor family distinguish identity features in these closely related DNA sites. We determined structures from several crystal forms of the RevErb−DNA complex and analyzed the patterns of protein−DNA interactions and DNA distortions. We found a significant and consistent DNA distortion at a TA step directly preceding the first consensus 5′-AGGTCA-3′ recognition sequence. Importantly, while this base-pair sequence is associated with RevErb's high-affinity sites, there are no sequence-specific contacts formed with the protein. Our study shows that RevErb relies instead on the intrinsic geometry and flexibility of this TA site to make the required fit between the proteins' independent major groove and minor groove binding interactions, which occur on both sides of the TA step. Our findings extend the description of response element discrimination to include a role for sequence-dependent DNA deformations and suggest how other monomeric members of this superfamily, such as NGFI-B, SF-1, and ROR, could also recognize unique geometric features in their DNA targets.

RevErb is an orphan member of the nuclear hormone receptor superfamily, which includes the steroid receptors, the thyroid hormone receptor, retinoid receptors, and the vitamin D receptor and related proteins. These transcription factors target specific DNA binding sites, known as response elements, from which they regulate their target genes. Upon binding of ligand, these receptors are able to control specific genes responsible for differentiation, growth, and homeostasis in species ranging from insects to humans (*1*). The orphan receptors have been identified by sequence homology to known receptors, but their respective ligands are unknown, and in some cases the receptor may act in a ligand-independent manner (*2*). RevErb acts as transcriptional repressor, and is expressed from two genes: α and *β* (NR1D1 and NR1D2, respectively) (*3*). It is implicated in growth and differentiation in several tissues and species, including muscle and adipocyte differentiation in mice and nervous system development in mice and chickens (*4*−*6*), and has homologues in fruit flies (E75) (*7*) and worms (SEX-1) (*8*), which are involved in metamorphosis and sex determination, respectively. This receptor represents a subclass of orphan receptors that can bind DNA efficiently as monomers, but in the case of RevErb a homodimeric binding mode to DNA is also observed (*9, 10*).

One of the key questions about the superfamily has been how different receptors distinguish their respective response elements, since both the DNA binding domains (DBDs)[1] and the core DNA sequence elements are highly conserved. In this regard, a fundamental feature of all nuclear receptor DBDs is the use of a recognition alpha helix [residues 19−31 in Figure 1(a)] which always engages the major groove of AGGTCA half-sites found in nonsteroid receptor response elements, such as those of RXR, RAR, TR, RevErb, and NGFI-B, and AGAACA half-sites in steroid receptor response elements such as those of GR, PR, MR, and AR. Several recent DBD−DNA structures have better elucidated the factors that delineate target site discrimination in this family (*11*−*17*). These factors include monomeric vs dimeric binding, homodimeric vs heterodimeric binding, relative orientation of the two-half-sites (e.g., inverted, everted, direct repeat), and the size of the inter-half-site spacing (*18*−*20*).

In addition, the structures of RevErbα (*16*), nerve growth factor inducible B (NGFI-B) receptor (*12*), and thyroid hormone receptor (TR) DNA complexes (*14*) highlighted additional protein−DNA interactions involving the C-terminal extensions (CTEs) of the DBDs. These studies provided evidence that the CTE can form a second independent DNA binding surface that makes interactions along the minor groove 5′ to the half-sites. The sequence requirements in these cases were established previously through a site-selection assay starting with random sequences and/or a conventional DNA binding assay using mutated DNA elements in the case of RevErb (*9, 10*) NGFI-B, (*21*−*23*), TR (*24, 25*), and related receptors for which structural

‡ The coordinates for crystal forms II and III have been deposited in the PDB, accession numbers 1GA5 and 1HLZ, respectively. Crystal form I has accession number 1A6Y.

* Correspondence should be addressed to this author. Phone: 804-243-6722. Fax: 804-924-9057. E-mail: fr9c@virginia.edu.

§ Department of Pharmacology.

|| Interdisciplinary Program in Biophysics.

⊥ Department of Biochemistry and Molecular Genetics.

[1] Abbreviations: DBD, DNA binding domain; CTE, C-terminal extension to the DBD; RXR, 9-*cis*-retinoic acid receptor; RAR, *all-trans*-retinoic acid receptor; TR, thyroid hormone receptor; GR, glucocorticoid receptor; PR, progesterone receptor; MR, mineralocorticoid receptor; AR, androgen receptor; NGFI-B, nerve growth factor induced B receptor; SF-1, steroidogenic factor 1.

10.1021/bi011086r CCC: $20.00
Published on Web 10/06/2001

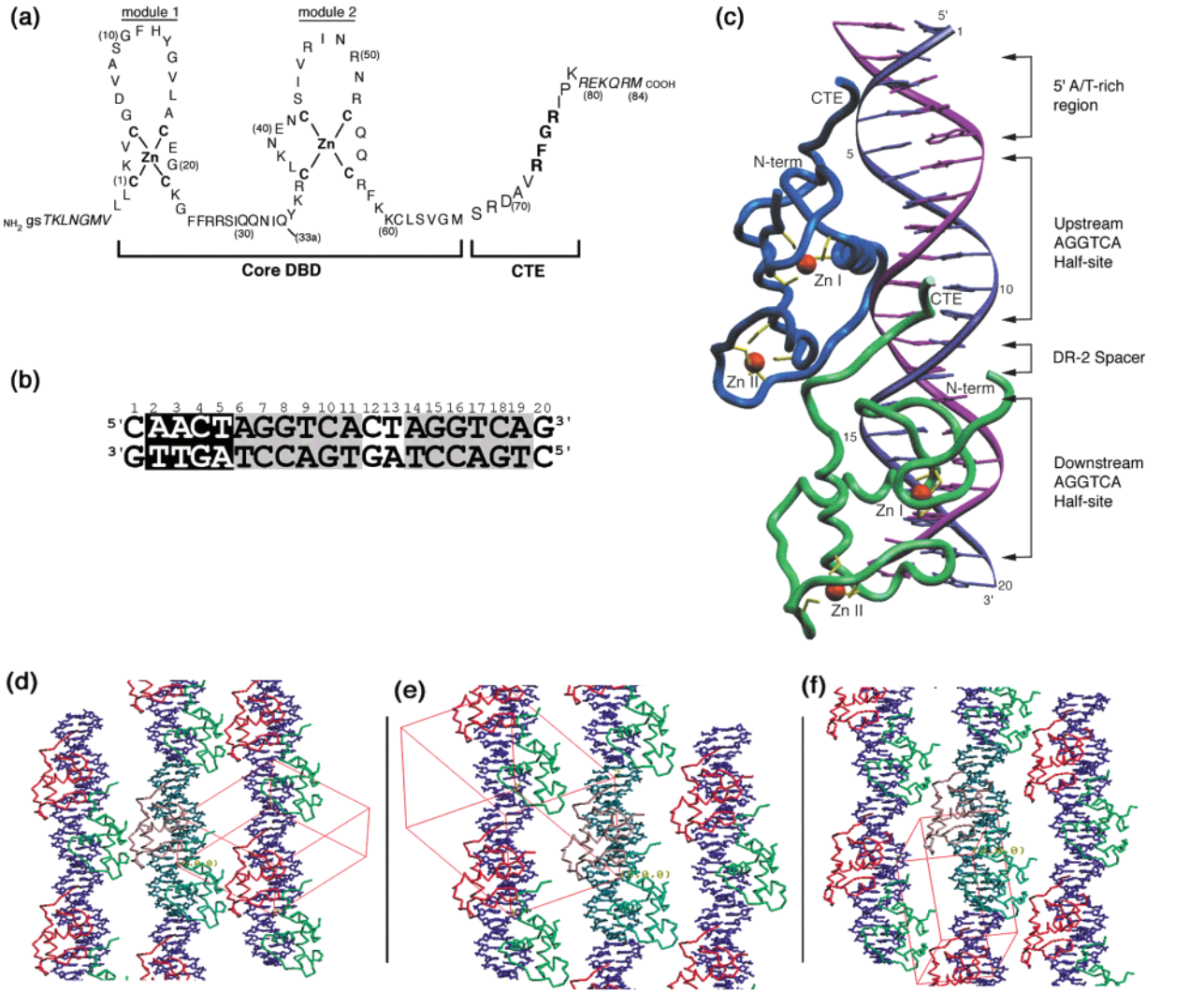

FIGURE 1: Sequence and structure of the RevErb−DNA binding domain (DBD) and its DNA response element (HRE). (a) Protein construct seen in the crystal structures. Numbering is according to the common numbering scheme for DBDs, in which the first coordinating cysteine in the first Zn module is labeled residue 1. (b) Sequence of the 20-mer DNA duplex in the structures. The AGGTCA half-sites are boxed in gray; the upstream (5′) flanking sequence required for tight binding are shown in reverse type. (c) Crystal structure of the overall complex, from crystal form I. Upstream monomer is shown in blue and downstream monomer in green. The 5′ and 3′ ends of the DNA are labeled, and zinc atoms are in red and their coordinating cysteine side chains are shown in yellow. (d−f) Diagrams showing the crystal packing in the different crystal forms: (d) crystal form I; (e) crystal form II; (f) crystal form III. The box indicates the unit cell, but in (e) the second complex has been left out for clarity.

information is still not available, such as SF-1 (*22*, *23*) and RORα (*26*).

Table 1 is a synopsis of these data for Reverb's monomeric binding, and the consensus sequences for the monomer and dimer were identical in this region. Each subunit of RevErb uses two distinct interfaces to bind its extended monomeric site, which consists of the conserved AGGTCA sequence as well as a specific 5′ flanking sequence [see Figure 1(b,c)]. At positions preceding the half-site (i.e., base-pairs 1−3), RevErb uses a so-called GRIP-box within its CTE to contact DNA at the minor groove. The GRIP-box was defined in the crystal structure as a second DNA binding interface, and is composed of Phe-73, Gly-74, Arg-75, and resides in the context of an extended peptide (VR{FGR}IP) that forms numerous contacts along the floor of the minor groove (*16*). This Grip-box does not form obvious sequence-discerning hydrogen bonds with the base-pairs, but the interactions with the minor groove are extensive and contribute favorably to the overall binding stability of the monomeric complex. The sequence preferences at base-pairs 1−3 are largely based on the steric properties at the minor groove edges of the base, which must allow the GRIP-box loop to penetrate deep in the groove in a way that leads to favorable van der Waals interactions. However, the initial crystal structure we determined could in no way explain the clear sequence preferences at positions 5 and 6 (see Table 1), which are located between the minor groove and the major groove binding sites. Therefore, we wanted to better assess how these base-pair sequences could contribute to RevErb's binding site selectivity.

Here we show how certain sequence requirements at the 5′ flanking sequences of the RevErb binding site contribute primarily to DNA deformability and not to contacts with the RevErb protein. To further investigate the structural determinants of these sequences in the RevErb binding site that were unexplained previously, we solved multiple crystal forms of the RevErb−DNA complex and looked at the degree of preservation in various DNA geometric features, including roll, twist, and slide. Multiple structure analysis

Table 1: Comparison of Biochemical Data and Crystallographic Data on the RevErb−DNA Interaction[a]

position	consensus sequence	no. of sequence representations in site selection assay	mutation/ effect on binding	location of contact
1	A/T	18/23		minor groove
2	A	20/23	C/−	minor groove
3	A/T	23/23	C/+	minor groove
4	N	−	A/+++++	no preference/ none
5	T	16/23	C/+++	none
6	A/G	23/23		none
7	G	22/23		major groove
8	G	22/23		major groove
9	T	22/23		major groove
10	C	23/23	G/++	major groove
11	A	23/23	C/++	major groove

[a] Data shown in columns 3 and 4 have been derived from Harding and Lazar, 1993 (*9*). Full binding is represented by five plus (+) symbols. Data in column 5 are from the analysis of the RevErb−DNA cocrystal structures reported here.

has provided a more precise description of a key DNA distortion at a TA step and suggests that it occurs irrespective of the crystal packing environment. The DNA distortion, composed of a large negative roll and large positive twist at this step, is seen in each of the four crystallographically independent structures precisely between the major and minor groove binding sites contacted by a RevErb protein. The distortion allows the DNA to undergo the necessary induced fit within the complex, allowing the recognition helix to contact the major groove, and the CTE to engage the minor groove. We propose that sequence-dependent DNA flexibility may act as an important discrimination feature for NGFI-B, SF-1, ROR, and other receptors which have similar protein organizations composed of coordinated major and minor groove binding interfaces and which bind to 5′ extended monomeric sites.

MATERIALS AND METHODS

Purification/Crystallization/Data Collection. The protein purification, DNA purification, complex formation, and crystallization are described elsewhere (*16*). Interestingly, all three crystal forms grew out of exactly the same conditions (*16*). The protein construct used is RevErbα. The previously described structure (crystal form I) has a 5-iodouridine in place of thymidine at position 13 on strand 1. The second structure (crystal form II), which has two copies of the complex in the unit cell (IIA and IIB), has a 5-iodouridine in place of thymidine at position 19 on strand 2. It is possible that the iodine atoms subtly perturb the local DNA geometry, but in neither case are they located near any protein−DNA contacts or critical DNA loci and thus are unlikely to affect the conclusions presented here.

Structure Determination and Refinement. The structure determination and refinement of model I are described elsewhere. Crystal form II was solved using molecular replacement in CNS (*27*), with model I used as a search model. The model consisted of both peptide chains (without the zinc atoms) and DNA residues 605−618 and 622−635. The cross-rotation was carried out with data from 8 to 4.0 Å. This gave a peak that was 3.5σ above the next highest peak, but all the solutions were fed into the translation search. The translation search was carried out from 15 to 4.0 Å, searching for one peak per cross-rotation solution. The highest peak from the cross-rotation search had a solution whose correlation coefficient was 0.29, which was 1.8σ above the next highest peak. This solution was fixed, and the second translation search carried out. The second solution gave a correlation coefficient of 0.65, which was 1.3σ above the next highest peak and had a high packing function score. This solution had an *R*-factor of 44.7%, and a free *R*-factor of 45.7%. After one round of rigid-body refinement from 15 to 2.5 Å in CNS, the *R*-factor and free *R*-factor were 41.3% and 41.6%, respectively. The initial $2F_o-F_c$ and F_o-F_c electron density maps showed strong density for the 8 Zinc atoms and 12 DNA base-pairs not in the search model, confirming that the molecular replacement solution was correct. Positional refinement was carried out in CNS using conjugate-gradient minimization against a maximum likelihood target using structure factor amplitudes, a bulk-solvent correction, and no data cutoffs. Model rebuilding was done in O (*28*) using SigmaA-weighted maps, with OOPS2 (*29*) used to help automate the process. DNA geometry and weak harmonic restraints were used initially and released in later stages of refinement. NCS restraints were used between the two independent complexes until the final rounds of refinement, which were carried out against all the data with no restraints. Initially, grouped *B*-factor refinements were carried out, with restrained individual *B*-factor refinement used in later stages. When the free *R*-factor had dropped to approximately 32%, CNS was used to place waters in appropriate F_o-F_c peaks, which were then checked in O for good hydrogen bonding environments. The refinement steadily converged to the values shown in Table 4. The model was checked with simulated-annealed omit maps, and there are no residues in disallowed regions of the Ramachandran plot as defined by the program PROCHECK (*30*).

Crystal form III was also solved by molecular replacement, using the same components of model I used for the crystal form II solution as a search model. The highest peak from the cross-rotation function was only 0.6σ above the next highest peak, but again all the solutions were fed into the translation function search. The translation function gave a peak 2.1σ above the next highest peak, with a good packing function score. This solution had an *R*-factor of 43.4% and a free *R*-factor of 47.2%. Rigid-body refinement from 15 to 3.0 Å in CNS gave an *R*-factor and a free *R*-factor of 41.2% and 45.6%, respectively. The same refinement strategy was used for model III as was used for model II, except that harmonic restraints were used, as well as NCS restraints between the upstream and downstream peptide core regions and AGGTCA sequences for the upstream and downstream half-sites. The DNA restraints were removed early in the refinement, but removing the NCS or harmonic restraints caused the free *R*-factor to increase, so they were included in the refinement until the final rounds (which were carried out against all the data with no restraints). Grouped *B*-factor refinement was carried out. Model III was also checked with omit maps, and there are no Ramachandran outliers, as defined by PROCHECK. The refinement statistics are shown in Table 4.

Model Evaluation/Comparison. LSQMAN (*29*) and O (*28*) were used to do the model superpositions. The protein−nucleic acid (*31*) and protein−protein (*32*) interaction servers

Table 2: Overall DNA Parameters[a]

	B-DNA	I	IIA	IIB	III	NGFI-B
major groove						
width (Å)	11.4	11.3(1.1)	11.2(0.8)	11.4(0.8)	11.4(0.7)	11.1(1.0)
depth (Å)	4.0	4.2(0.9)	3.9(1.3)	3.8(1.1)	4.1(0.9)	4.8(0.7)
minor groove						
width (Å)	5.9	6.0(0.6)	6.1(0.6)	6.0(0.6)	6.3(0.6)	6.6(1.2)
depth (Å)	5.5	5.3(0.5)	5.2(0.6)	5.2(0.5)	5.1(0.6)	5.0(1.5)
base-pair roll (deg)	1.7(0.1)	3.4(5.2)	3.5(4.5)	3.6(4.9)	2.6(5.3)	2.8(4.5)
base-pair slide (Å)	0.45(0.3)	0.33(0.9)	0.37(0.8)	0.35(0.7)	0.27(0.8)	0.00(1.0)
rise per base-pair (Å)	3.4(0.0)	3.3(0.2)	3.3(0.4)	3.3(0.3)	3.3(0.3)	3.3(0.1)
helical twist (deg)	35.9(0.9)	35.8(4.7)	35.5(5.3)	35.7(4.7)	35.4(5.1)	35.4(5.9)
helical D*x* (Å)	0.49(0.4)	−0.13(1.7)	−0.19(1.8)	−0.11(1.7)	−0.12(1.6)	−0.71(1.9)

[a] Values shown are averages from the four structures, with standard deviations indicated in parentheses. Groove widths and depths were calculated with Curves (*33*); all other values were calculated with 3DNA (*34*). NGFI-B data is from PDB entry 1CTT.

Table 3: DNA Deformation Energies[a]

	TA	CA	AT	AA
roll	1.84	3.85	2.26	1.84
twist	2.70	2.45	12.20	7.64
slide	3.58	1.93	30.45	10.85

[a] Shown are elastic deformation energies, in units of kT, where k is the Boltzmann constant and T is the absolute temperature in degrees kelvin, required for different base-step dimers to adopt the roll, twist, and slide conformations seen at the T5-A6 step in the RevErb−DR2 complex. The TA and CA dimers are clearly selected with respect to slide and twist, and the TA dimer is selected over the CA dimer with respect to roll. Values were calculated using the mean value reported in Olson et al. (*44*) for the given parameter and dimer, and the corresponding force constant derived from the inverse harmonic analysis listed at http://rutchem.rutgers.edu/~olson/force_cnst.html.

were used to calculate the buried surface areas. The DNA geometry was analyzed using the program Curves (*33*) for global helical axis and groove depth and width calculations, and the program 3DNA (*34*) for all other parameters. Figures were prepared with SETOR (*35*).

RESULTS

Overall Comparison of Four RevErb−DNA Complexes. The RevErb−DNA structure we reported earlier at 2.3 Å resolution is referred to here as model I (*16*). Here, we report three new structures derived from two additional crystal forms (see Table 4). The new structures in crystal forms II and III were solved by molecular replacement to resolutions of 2.4 and 2.8 Å, respectively. Crystal form II has two complete complexes in the asymmetric unit, referred to as model IIA and model IIB, respectively; models I and III each have one complex in the asymmetric unit. All crystal forms belong to the space group *P1*. Both the lateral lattice contacts and the angle between the axis of the DNA and the long axis of the unit cell are different in the three crystals [Figure 1(d−f)], indicating that the complexes are in different crystallographic environments and that consistent distortions in the DNA are not simply due to crystal lattice constraints.

Table 4 shows that crystal forms II and III have moderately high average crystallographic temperature (*B*) values and *R*-values. These *R*-values are in line with other protein−DNA complexes of similar *B*-values and diffraction limits. Several recently reported protein−DNA structures in the resolution range 2.4−3.1 Å show a range of *R*-values between 19.2 and 27.3%, and a range of free *R*-values between 27.3 and 32.0% (*12*, *36*−*39*). The larger atomic *B*-values associated with crystal forms II and III agreed well with those estimated for the data using Wilson plots. The models have good stereochemistry (see Table 4 and Materials and Methods) and good electron density in the core regions of the monomers and at the protein−DNA interfaces. Figure 2 shows the superposition of all the structures.

Table 4: Data Collection and Refinement Statistics

	crystal form		
	I	II	III
Data Collection Statistics			
wavelength (Å)	0.9054	0.9054	0.9207
resolution (Å)	19.4−2.30	19.6−2.35	18.6−2.70
R_{merge} (%)[a,b]	8.8 (26.1)	8.3 (32.3)	7.8 (25.8)
$\langle I/\sigma(I)\rangle$[b]	21.9 (2.8)	11.3 (2.9)	9.6 (3.18)
completeness (%)[b]	97.3 (94.0)	87.3 (82.8)	96.6 (93.3)
unique reflections[c]	13482 (666)	25370 (2476)	6854 (732)
average redundancy	3.5	2.5	1.7
unit cell parameters			
a,*b*,*c* (Å)	38.7,45.7,48.0	44.9,52.2,78.9	35.9,39.6,58.6
α,*β*,*γ* (deg)	75.2,80.6,85.6	85.8,76.6,74.5	84.1,72.1,70.2
space group (no. of complexes in asymmetric unit)	P1 (1)	P1 (2)	P1 (1)
solvent content (%)	51	58	56
Refinement Statistics			
resolution (Å)	19.4−2.3	19.6−2.4	18.6−2.8
R_{cryst}/R_{free} (%)[d]	19.5/28.0	25.3/29.9	26.1/31.3
total atoms (non-H)			
protein/DNA	1282/820	2367/1640	1163/820
zinc/solvent	4/234	8/279	4/9
RMS deviations			
bonds (Å)	0.011	0.007	0.009
angles (deg)	1.80	1.30	2.30
average *B*-factors (Å^2)			
protein/DNA	29.2/28.8	46.4/50.8	56.1/60.9
zinc/solvent	22.6/42.9	41.6/47.2	56.7/61.5

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_{i(hkl)} - \langle I_{hkl}\rangle|/\Sigma_{hkl}\Sigma_i\langle I_{i(hkl)}\rangle$, where I_i are individual measurements for any one reflection and $\langle I\rangle$ is the average intensity of the symmetry-equivalent reflections. [b] Values in parentheses are for the high-resolution bin: I(2.38−2.30 Å), II(2.43−2.35 Å), III(2.80−2.70 Å). [c] Values in parentheses are the number of reflections used in the test set for calculation of R_{free}. [d] $R_{cryst} = \Sigma|F_o - F_c|/\Sigma F_o$, where F_o and F_c are observed and calculated structure factors, respectively. R_{free} is calculated in the same way as R_{cryst} with a test set of reflections removed from the refinement.

The DNA bases deviate from standard Watson−Crick geometry at positions 5 and 6. An estimation of the coordinate errors based on Luzzati plots (*27*) indicates that the expected coordinate errors are smaller than the observed shifts in the coordinates of base-pairs 5 and 6 from their standard (Watson−Crick) positions. Model I has a coordinate

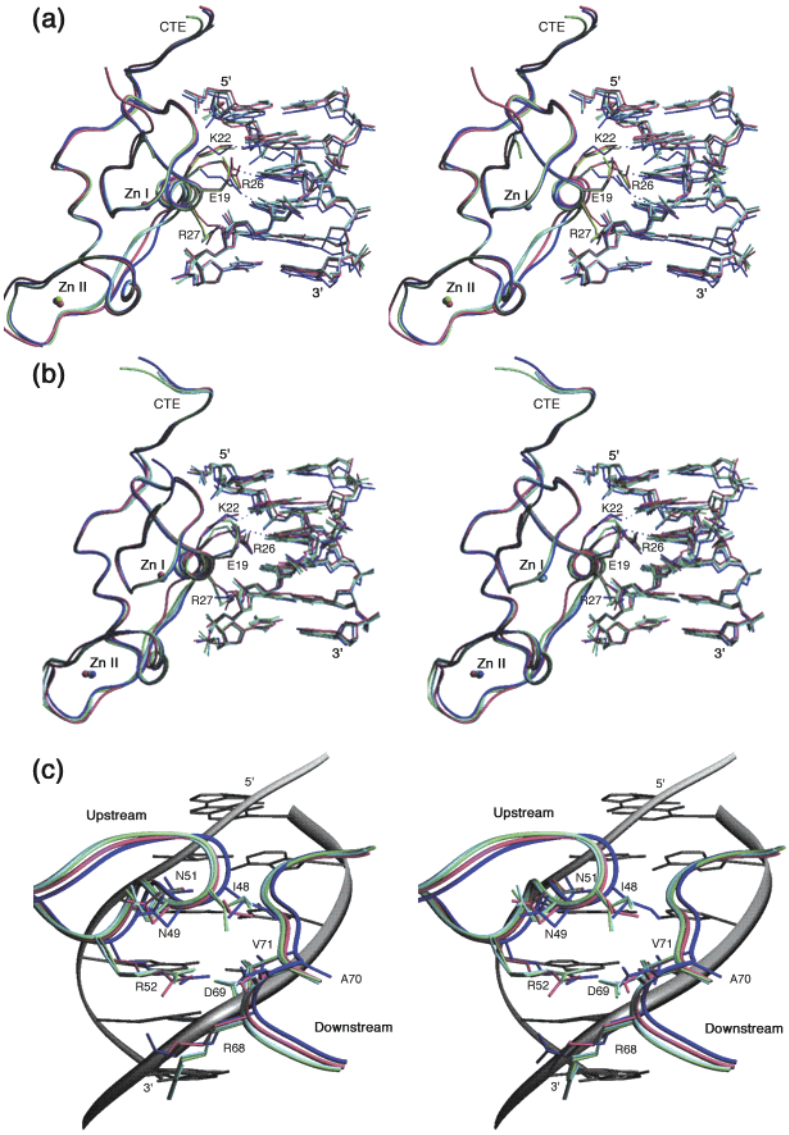

FIGURE 2: Stereopairs showing comparisons of the four different structures of the RevErb−DNA complex. (a) Upstream monomer/AGGTCA half-site. (b) Downstream monomer/AGGTCA half-site. (c) Dimer interface. In each case, model 1 is in dark blue, model IIA is in green, model IIB is in light blue, and model III is in magenta. In (c), only the DNA from model I was used, shown in gray. Figures prepared using SETOR (*35*).

error of 0.34 Å, but the T5-A6 atoms are shifted 1.42 Å (RMSD) from their standard positions. Models IIA and IIB have coordinate errors of 0.42 Å, and the shift in T5-A6 from their standard positions is 1.46 and 1.51 Å, respectively. Model III has a coordinate error of 0.5 Å, but the T5-A6 atoms are shifted 1.56 Å from their standard positions. Figure 3 shows that canonical B-DNA geometry would not fit the observed density in this region.

The gross features of the two independent DNA binding interfaces of RevErb are preserved in the four independent complexes we examined. However, there are subtle differences in protein conformation, DNA geometry, and water molecule positioning that illustrate the degree of plasticity possible in such structures. Pairwise superpositions of the four crystallographically independent complexes show that overall the models have an RMSD of 0.8 Å for all Cα and phosphorus atoms. Difference-distance matrixes that have been corrected for the uncertainty of the coordinates (*40*) (not shown) indicate that the only significant differences in protein main chain position between the structures occur in the loop regions immediately surrounding the Zn II region (residues 31−42), which are flexible in most nuclear receptor DBD structures. The protein buries an average of 2035 Å^2 (SD 90 Å^2) of solvent-exposed surface on the DNA in the four complexes, with the downstream half-sites burying an average of 150 Å^2 more surface area than the upstream half-site (*32*). This difference in buried surface area between the two half-complexes is due to the fact that the upstream subunit stabilizes the downstream subunit's CTE through the protein dimer interface, which buries an average of 143 Å^2 (SD 16 Å^2) of surface area (*32*).

The upstream and downstream subunits have CTEs which are in different local environments, owing to the participation of the downstream CTE in protein−protein dimerization [see Figure 1(c)]. As the upstream CTE is not stabilized by any subunit interactions, it is slightly more disordered and marked by somewhat higher atomic *B*-factors. In fact, while none of the four upstream CTEs exhibit density beyond Arg75, all of the downstream CTEs have good density at least through Pro77. Also, the upstream and downstream CTEs

face different DNA sequences: CAACT for the upstream CTE and TCACT for the downstream CTE.

Overall, the backbone atoms of the downstream CTE (residues 67−75) align well with respect to one another (RMSD < 1.0 Å). The side chains of residues Ser67, Asp69, Ala70, Val71, and Phe73 also align closely, while the side chains of Arg68, Arg72, and Arg75 show more variability. The side chains of Arg68 in models I and III make hydrogen bond contacts with both upstream Arg52 and the phosphate oxygen atoms of G15. Those from models IIA and B contact the G15 phosphate oxygens and the carbonyl oxygen from Ile28. In models I and III, the Arg72 guanidino group forms water-mediated hydrogen bonds to the carbonyl oxygens of Arg72 and Gly74, while in models IIA and B, the guanidino group functionally replaces the water molecule, forming direct hydrogen bonds with the two carbonyl oxygens (see Figure 4c).

RevErb's Readout of DNA Geometry and Distortion. The RevErb−DNA is in the B-form, with an average rise per base-pair of 3.35 Å (Table 2). Alignment of the RevErb−DNA with a canonical B-DNA model of the oligomer (average RMSD 1.5 Å for the four models) (*34*) shows that the response element DNA, when viewed over its entire length, appears slightly skewed toward the upstream bound protein [Figures 1(c) and 6(a)]. Our examination of these various crystallographic models shows several unique DNA distortions that are well preserved in all of the structures. These consistent structural features include the major groove width, which has an average minimum of 9.8 Å in the region of G7 and G15 in the AGGTCA half-sites (compared to a width of 11.4 Å for fiber B-DNA), and a broad maximum of between 12.5 and 13.8 Å in the spacer region. The major groove depth has sharp minima ranging between 2 and 3 Å at the T-A step at the 5′ end of both half-sites (T5-A6/T13-A14), compared with 4.0 Å in B-DNA. Upstream of the first AGGTCA half-site, the minor groove narrows and deepens slightly, to ∼5.3 Å and ∼5.9 Å, respectively (compared to a width of 5.9 Å and depth of 5.5 Å for canonical B-DNA). This is in contrast to the same region in NGFI-B, which widens to over 7 Å (*12*).

To detect other types of distortions in the DNA, we show in Figure 3(a) the parameters sensitive to deformation (base-pair roll, twist, and slide) as a function of base-pair sequence. We identified a significant and consistent distortion at the T5-A6 step immediately 5′ to the upstream AGGTCA half-site, where there is significant negative roll (base-pair edges rotating away from each other on the major groove side), increased twist, and positive slide (relative translation along the long axis of the bases). There is also positive *x*-displacement (translation toward the major groove with respect to the locally defined helical axis) at this juncture in all four complexes (not shown). Figure 3(b) specifically shows that the region of the electron density map around base-pairs 5 and 6 of the DNA is well-defined and correctly represented in the refined models, and not compatible with idealized Watson−Crick base-pair geometry. The distortions result in an overall local increased winding about the helical axis, with concomitant compression of the base-pairs in the minor groove and movement of the A6 base-pair toward the major groove.

It is important to note that a similarly configured CTAG sequence also appears at the downstream half-site (T13-A14),

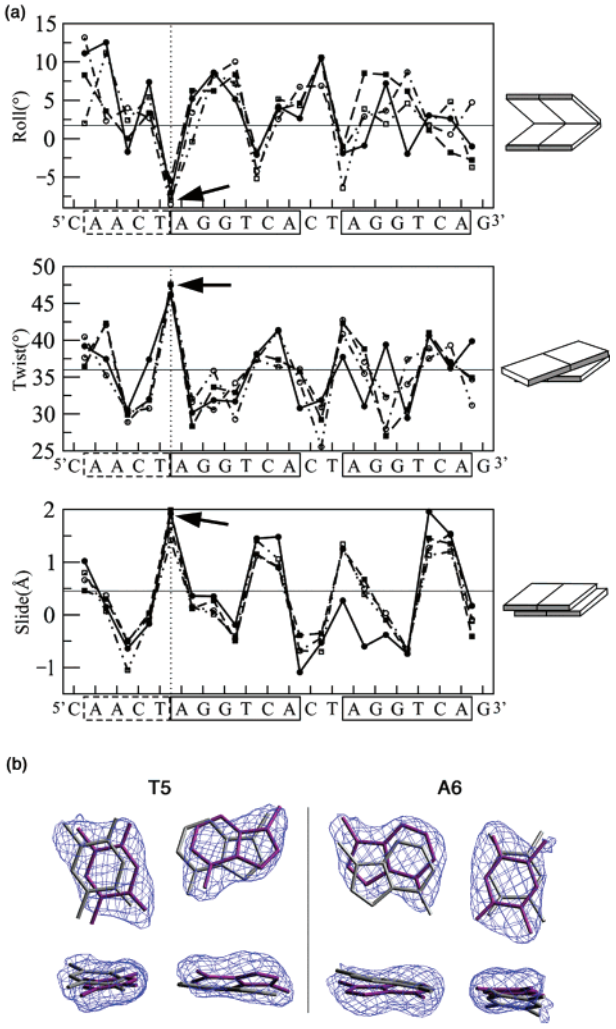

FIGURE 3: (a) Comparison of DNA parameters (roll, twist, and slide) for the four RevErb models. Model I, filled circles; model IIA, filled squares; model IIB, open circles; model III, open squares. Thin solid lines indicate ideal B-DNA geometries. The vertical dotted lines denote the loci of consistent distortions at the T5-A6 step, and arrows highlight the consistent distortion. AGGTCA half-sites are boxed with solid lines; the upstream AACT sequence is boxed with dashes. In the pictograms to the right, the minor groove side of the base-pair is shaded. (b) Electron density of the model IIA T5 and A6 bases, contoured at 1.3σ, demonstrating that the deformation is seen clearly in the electron density map. The magenta models are from model IIA. For comparison, base-pairs in gray indicate that base-pairs with standard (B-DNA) parameters would not fit the density equally well.

but that this TA step has not adopted the same structure as T5-A6. This region of the DNA, where the downstream GRIP-box is interacting with the DNA, also has protein contacts in the major groove from the upstream recognition helix, which does not allow the geometry to be determined solely by the GRIP-box interaction. The differences in geometry seen for these two dinucleotide steps (T5-A6 versus T13-A14) suggest that the TA step has an intrinsic flexibility that may be an important recognition feature for the RevErb protein. Interestingly, mutating T13 to a C modestly reduces dimeric binding (*10*), and as there are no base-specific protein contacts at this base-pair, this result suggests that the local geometry induced in T13-A14 is also important, though different from that of T5-A6.

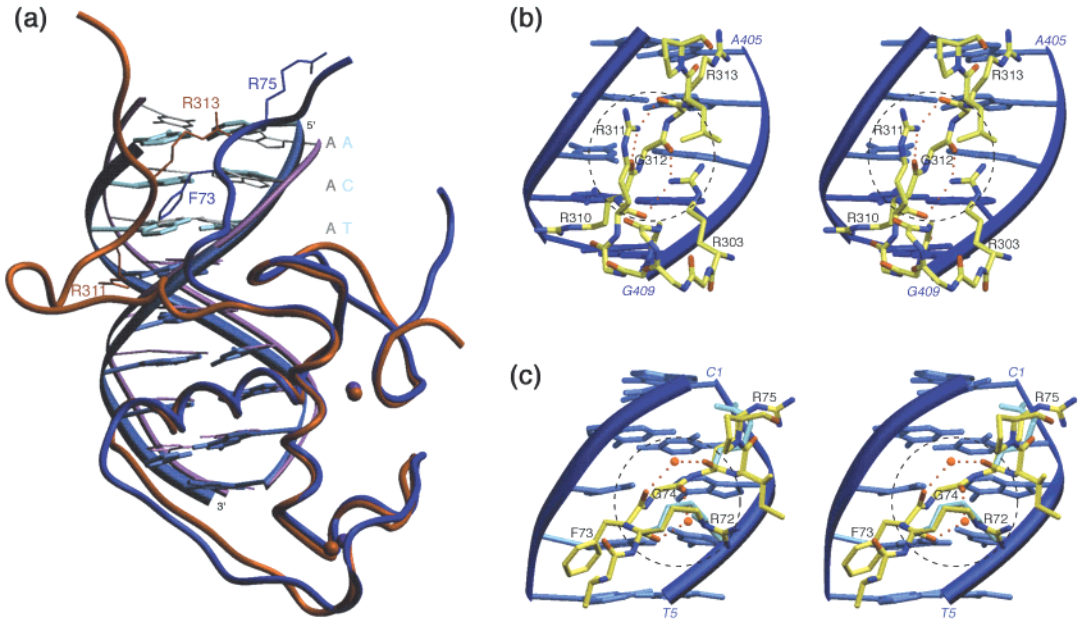

FIGURE 4: Comparison of RevErb and NGFI-B DNA binding complexes. (a) Superposition of the downstream RevErb monomer and its associated DNA with the monomeric NGFI-B−DNA complex. RevErb monomer is shown in dark blue; RevErb−DNA backbone and AGGTCA half-site bases are in blue; RevErb upstream (flanking) base-pairs are in light blue; NGFI-B protein is in red; NGFI-B DNA backbone and AGGTCA half-site bases are in magenta; NGFI-B upstream bases are in gray. The side chains from residues F73, R75 from RevErb, and R311, R313 from NGFI-B are shown. (b) NGFI-B−CTE interactions at the minor groove, showing the hydrogen bond network between the carbonyl oxygens of (RevErb equivalents in parentheses) R310(72), R311(73), G312(74), and R313(75). R303 has no counterpart in RevErb. (c) Downstream RevErb CTE, showing the hydrogen bond network between the carbonyl oxygens of R72, F73, G74, and R75. Alternate side chain conformations of R72 and R75 from model IIA are shown in cyan. Amino acids in the peptides are colored by atoms, waters are shown as red spheres, and hydrogen bonds are red dotted lines. The dashed circles highlight the conserved backbone configuration in RevErb and NGFI-B.

A roll/twist combination of approximately −8° and 48°, respectively, is adopted at the T5-A6 step (Figure 3) and is the major DNA distortion in the 20 base-pair sequence; this is true for each of the four crystallographically independent structures that we have studied. DNA site selection experiments show that the base at position 5 must be a thymidine for ideal monomer or dimer binding, but that the base at position 6 can be either an adenine or a guanine, both of which are purines (Table 1) (*9*). The greater flexibility of pyrimidine-purine base-pair steps over all other combinations has been described previously from gel mobility assays, computational studies, and surveys of protein−DNA structures (*41*−*43*). Olson et al. (*44*) recently proposed an empirically derived dimer model, based on harmonic potentials for the three rotational (roll, twist, and tilt) and three translational (rise, slide, and shift) base-pair degrees of freedom, that gives elastic force constants impeding these motions for a given base-pair step. This analysis showed that the T-A step is also the most flexible of the pyrimidine-purine steps. Using the mean value of a parameter for a given step, along with the derived force constant (http://rutchem.rutgers.edu/~olson/force_cnst.html), one can calculate, in units of kT, the deformation energy required for several types of base-pair steps to adopt the precise conformation seen at the T5-A6 site in our structures. Table 3 shows a comparison of energies for selected base steps in the conformation seen in the four RevErb−DNA complexes at the T5-A6 step. It is clear that the TA and CA steps would be able to adopt the conformation seen in the crystal structure with the lowest energetic costs.

Therefore, the energies shown in Table 3 provide a reasonable explanation for the biochemical data suggesting the association of the T5-A6 step with the optimal binding sites of RevErb. Position 6 can be either purine (A or G), and position 5 is preferentially a T, although mutating it to a C, which is still a pyrimidine, reduces monomeric RevErb binding by only 40% (see Table 1), whereas other changes have a more dramatic effect on binding (*9*). Purine-purine and purine-pyrimidine steps were not represented at this location in any of the RevErb binding sequences derived from the DNA site selection assay. As Table 3 shows, these combinations (represented by AA and AT, respectively) require much more energy to deform into the observed conformation. In comparing the TA and CA steps, the observed roll conformation strongly favors the TA step. Thus, the requirement for the TA step at the 5′ end of the half-site is not a result of any specific contacts, but stems from a requirement for a specific DNA geometry that can be best accommodated by this pyrimidine-purine step and therefore lead to the best fit between the proteins' major and minor groove binding elements and their cognate sequences. Our analysis of the crystallographic *B*-values further indicates that the DNA at the T5-A6 step may be pushed into a locally stabilized conformation, as the atomic *B*-factors in that region of the DNA are 5−10 Å^2 less than those of their neighboring bases in each structure.

Comparison of DNA Deformations in the RevErb and NGFI-B Complexes. Figure 7 shows how several orphan nuclear receptors have sequence similarity in their CTEs including related GRIP-box elements, and it is also known that these receptors require specific sequences 5′ to their AGGTCA sites for optimal binding (*10*, *22*, *45*). The CTE is sufficient to determine which upstream DNA sequence a receptor will bind to; chimeric constructs with either the CTE

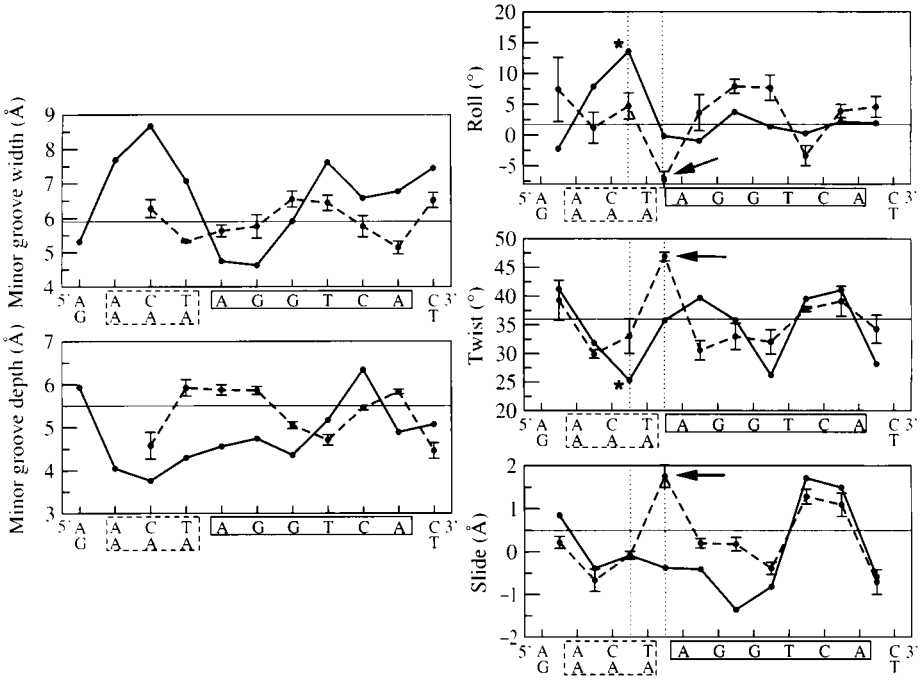

FIGURE 5: Comparison of NGFI-B and RevErb−DNA geometries. NGFI-B parameters are shown using solid lines, and RevErb parameters are shown using dashed lines. For RevErb, data shown are the average value for base-pairs 2−12 from the four models, with one standard deviation shown as error bars. For NGFI-B, data are shown for the similar DNA site composed of base-pairs 3−13 (−4 to 7 in the authors original numbering scheme). The RevErb sequence is shown along the *x*-axis, with the NGFI-B sequence shown below where the two differ. The prominent distortions in the RevErb and NGFI-B response element DNA occur 1 base-pair away from each other (indicated with arrows and asterisks, respectively) and are different in terms of roll, twist, and slide parameters.

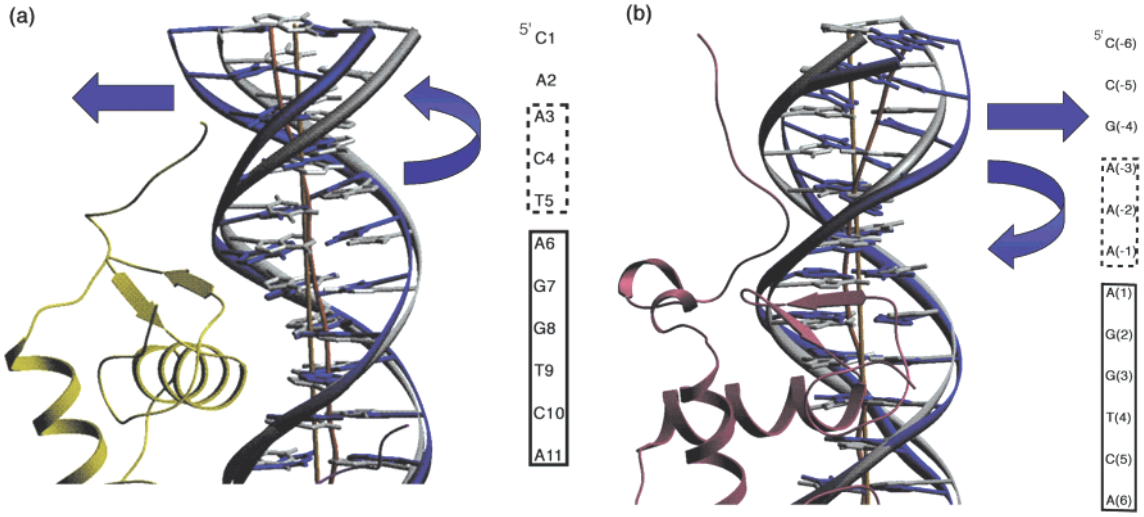

FIGURE 6: Helical curvatures of RevErb and NGFI-B DNA compared with standard B-DNA. Arrows indicate direction of distortion. (a) RevErb−DNA from crystal form I shown in dark blue with its global helical axis from curves shown in red. The standard B-DNA structure for this sequence is shown in gray. The DNA sequence is listed to the right and is boxed as in Figures 3 and 5. (b) NGFI-B−DNA structure. The NGFI-B−DNA is in dark blue, with its helical axis shown in red. B-DNA is represented in gray.

of SF-1 attached to the core DBD of NGFI-B or the CTE of NGFI-B attached to the core DBD of RORα confer specificity for the SF-1 and NGFI-B response elements, respectively (*46*). The comparison of RevErb and NGFI-B structures shown in Figure 6 demonstrates how DNA site selection could be accomplished through recognition of DNA geometry, despite the fact that both receptors use highly related recognition alpha helices, a consensus AGGTCA site, and a superficially related GRIP-box that interacts with the minor groove of 5′ flanking sequences. One critical difference to note is the size of the connecting loop linking the GRIP-box to the core DBD, with NGFI-B having seven residues inserted with respect to RevErb (Figure 7). Figure 4(a) shows that the core DBDs of RevErb and NGFI-B superimpose quite well, while their CTEs (residues past Met66 in RevErb and Met297 in NGFI-B) follow different trajectories in approaching the DNA minor groove, stemming from the different sizes of their CTE preceding the GRIP-box. The NGFI-B CTE buries nearly 100 Å^2 more protein surface area on the DNA than does the RevErb CTE (*31*), with more overall and base-specific contacts to the DNA. However, close inspection of the GRIP-boxes (the sequence RFGR in RevErb and RRGR in NGFI-B) reveals that they have a conserved architecture, even though they are displaced from

Receptor	CTE		Monomer response element
	66	Position: 4321	
RevErb	GMSRDAV-------**RFGR**IPKR		AACT**AGGTCA**
NGFI-B	GMVKEVVRTDSLKG**RRGR**LPSK		AAA**AGGTCA**
ROR	GMSRDAV-------**KFGR**MSKK		WWCW**AGGTCA**(α1)/AACT**AGGTCA**(α2)
SF-1	GMRLEAVRADRM--**RGGR**NKFG		TCA**AGGTCA**

FIGURE 7: Alignment of the CTEs and the DNA binding sites for RevErb, NGFI-B, RORα, and SF-1. The amino acid sequences shown start with the last residue of the conserved core DBD (Gly-66) and include the CTEs. Dashes indicate absence of any corresponding sequence, and the GRIP-box residues are in boldface type. The preferred DNA binding sites for these monomeric response elements are also shown, with the common hexameric sequence shown in boldface type and their distinct 5′ sequences indicated. ROR has two isoforms which differ primarily in sequences preceding their core DBD, and both binding sites are shown (α1 and α2). W refers to an A or T base.

one another when the core DBDs are aligned. As shown in Figure 4(b,c), the carbonyl oxygen atoms from the first and third residues (Arg72/Gly74 in RevErb, Arg310/Gly312 in NGFI-B) and the second and fourth residues (Phe73/Arg75, Arg311/Arg313) interact through either bridging nitrogen atoms from arginine side chains or water molecules. This required conformation explains part of the 5′ A/T-rich region's sequence specificity in both receptors, as various studies have shown that a G at position 3 for RevErb or at the position immediately upstream of the AGGTCA for NGFI-B abolishes binding, due to steric exclusion of the conserved glycine which is fixed by the interlocking hydrogen bonds (*9*, *10*, *22*).

We next compared the DNA geometry in the RevErb and NGFI-B binding sites to see if we could more precisely account for their different DNA structure, and found that there are significant differences in geometry associated with their flanking A/T-rich regions (Figure 5), resulting in different DNA curvatures when each is compared with canonical B-DNA (Figure 6). Figure 5 shows that the RevErb target's minor groove is noticeably narrower and deeper, while the NGFI-B target's minor groove becomes wider and shallower than normal B-DNA. The NGFI-B DNA exhibits typical "A-tract bending" (*47*) with positive roll and reduced twist at the A406-A407 step two steps upstream of the AGGTCA site, resulting in a slight bending (∼15°) of the helix toward the major groove [Figure 6(b)](*12*). This contrasts with the RevErb−DNA geometry discussed above, which has negative roll and increased twist at the T5-A6, one step upstream of the AGGTCA site, resulting in slight bending toward the minor groove (∼5°). NGFI-B prefers a more convex surface to accommodate simultaneous binding of its CTE and core DBD. Due to the presence of seven fewer residues in RevErb, this receptor needs a slightly concave surface at its DNA element to accommodate optimal binding. Therefore, it is clear that in both RevErb and NGFI-B response elements, the DNA duplexes must be predisposed or induced into specific conformations to form the complementary binding surface for their proteins.

DISCUSSION

We have shown that a necessary thymidine-adenine base-pair step in the 5′ flanking sequence of the RevErb binding site adopts a unique and accommodating distortion when bound to RevErb in all of the crystal forms of the complex that we have solved. Importantly, a similar TA sequence (T13-A14) in the same DNA duplex does not adopt the same distortion, in part because the two subunits must form their dimer interface effectively at this site. This observation argues that the T5-A6 step is highly flexible, and is useful in terms of the induced fit required to form a stable complex with RevErb. Many proteins that bind DNA take advantage of the increased flexibility at pyrimidine-purine steps to create a better binding surface (*41*−*44*). Both Dickerson (*41*, *42*) and Olson (*44*) found in studies of high-resolution protein−DNA structures that the thymidine-adenine step had the greatest flexibility of all the pyrimidine-purine steps. In RevErb's case, a roll/twist combination of approximately −8° and 48°, respectively, is needed at the T5-A6 step. This TA sequence appears to be critical based on DNA site selection experiments, and our analysis suggests that this is because it accommodates these roll and twist values with the least amount of input energy for geometric readjustment, rather than because of specific sequence recognition by the protein side chains.

Importance of DNA Flexibility in Protein−DNA Recognition. It has been known for some time, based on computational studies and gel mobility assays, that the DNA structure can adopt various types of distortion, and that certain sequences are more flexible than others. In addition, several early protein−DNA crystal structures, such as the *trp* repressor−operator complex (*48*), showed that proteins could bind to specific sequences without direct contacts to the DNA bases, implying that there must be some form of indirect readout of the DNA sequence. More recent studies have investigated subtle DNA distortions, such as those reported here, that can modulate the formation of protein−DNA complexes. Garvie and Phillips (*49*) showed that there is a flexible TA step in the Met repressor−operator complex that is critical for indirect readout of the correct operator sequence. Multiple crystal structures of the interferon regulatory factor (IRF) DBD homodimer bound to DNA showed a conserved deformation of the core GAAA sequence (*50*), which had no base-specific contacts from the protein. This protein binds cooperatively to the DNA, but there are no protein−protein contacts in the crystal structure, implying that the DNA conformation is the primary cause of the cooperativity. Several studies of the Papillomavirus E2 protein bound to different DNA sequences by Hegde et al. (*51*−*53*) have shown that the intrinsic flexibility of the target DNA is crucial for determining binding behavior. Several groups have studied the sequence requirements for binding of the TATA-box by TATA binding protein (TBP) (*54*−*56*). This is another case where TA steps are crucial for protein binding, although the deformation is much more severe, as phenylalanine side chains are inserted between consecutive base-pairs.

There is evidence that DNA flexibility is used as a discrimination feature by other nuclear receptors. The crystal structure of the RXR DBD homodimer bound to a DR1 response element showed that the global helical DNA axis was bent by 12°, while in the RXR−RAR DR1 crystal structure, the axis was bent by only 6° (*15*). This result is particularly interesting, since the DNA used in both complexes contains the same DR1 architecture with two AGGTCA sites, implying that the DNA can be induced to distort differentially in these two complexes, with the locus of the DNA distortion the central spacer element separating the two half-complexes. This suggests that differential distortability at the spacer base-pair or flanking sequences may allow the RXR homodimer and the RXR−RAR heterodimer to select their cognate response elements preferentially, even though the core binding sites are DR1 in both cases. In addition, a phasing analysis of DNA bending by the RXR−TR heterodimer on a DR4 response element indicated that the DNA was bent by 10°, which is consistent with the crystal structure (*57*). Differential distortability may play prominently in site selection by RXR−TR and RXR−LXR, both of which must discriminate their binding sites despite their related architecture. In this way, sequences outside the hexameric sites may be induced to adopt different distortions that allow each of these heterodimers to discern their distinct gene regulatory sites.

In a number of cases, local deformations in the DNA that do not propagate obviously along the entire duplex have not been carefully considered in terms of their importance to the induced fit between a protein and its DNA site. This is likely due in part to the fact that there are few cases where there are multiple crystal forms of the complex that improve the signal-to-noise ratio of subtle geometric anomalies enough to allow sufficient confidence in the precision of these irregularities.

Site Discrimination by Other Monomeric Receptors. Since specific 5′ flanking sequences are also required for optimal binding of other monomeric receptors, we have considered how receptors with CTE sequences related to RevErb and NGFI-B could discriminate their sites. Importantly, our study would predict that DNA distortability could be an important factor for these other receptors, with bipartite DNA binding motifs requiring both a major groove and a minor groove interface. Moreover, NGFI-B, RevErb, and several other orphan receptors all bind to monomeric sites containing the same AGGTCA site but extended with different 5′ sequences. Figure 7 shows that the orphan receptor steroidogenic factor 1 (SF-1 in humans, FTZ-F1 in fruit flies, NR5A1) (*3*), like RevErb and NGFI-B, also contains an Arg-Gly motif within its CTE, and this receptor has also been shown to bind monomerically to an extended AGGTCA site (*22*). However, SF-1 has a peptide link between its GRIP-box (sequence RGGR) and core DBD that is two residues shorter than that of NGFI-B (see Figure 7) (*22*). DNA site selection and DNA binding experiments have shown that SF-1 binds to sequences in which TCA appears directly upstream of the AGGTCA half-site (*22*, *23*). In contrast to NGFI-B and RevErb, SF-1 appears to require a guanine amine group in the minor groove (*22*). However, a homology model of the SF-1 DBD−DNA complex based on the NGFI-B crystal structure does not allow us to explain these sequence requirements (not shown). The different protein and DNA sequences in the SF-1−DNA complex indicate that both the protein and the DNA will have distinct conformations from those seen in the RevErb and NGFI-B complexes.

The orphan receptor RORα has a more closely related CTE to that of RevErb, including a very similar GRIP-box (KFGR in ROR versus RFGR in RevErb) and the same size connecting loop as RevErb, and has been shown to bend DNA (Figure 7) (*58*). However, RORα has two isoforms (1 and 2) that have differences in their amino-terminal domains and which have been shown to bind slightly different flanking sequences (Figure 7) (*45*). RORα2 has an identical consensus target sequence and binding pattern to mutant oligonucleotides as those of RevErb, but RORα1 has a consensus sequence of A/T-T/A-C-A/T-AGGTCA and binds with 65% of wild-type affinity to a probe with a G at position 5 (*45*). These results would indicate that the region N-terminal to the core DBD may interact with the DNA in RORα, and could involve yet another type of DNA distortion.

CONCLUSIONS

By examining multiple crystal forms of the orphan receptor RevErb−DNA binding complex, we identified a consistent, packing-independent DNA distortion that explains sequence requirements not involved in protein contacts with RevErb. By thoroughly examining a set of structures, we have identified subtle DNA distortions that would not be discernible from studying any one crystal structure alone. In particular, our examination highlighted an unusual DNA configuration that is associated with the high-affinity binding of RevErb, and which appears to assist the induced fit between the protein and its DNA target. Sequence-dependent DNA distortability is likely to play an important part in the DNA discrimination and site selection of various other nuclear receptors, most of which use the same general hexameric binding sequence (AGGTCA), but with unique flanking sites and dimeric arrangements. In a larger sense, the analysis of the RevErb−DNA binding complex adds to the evidence that DNA can be recognized in terms of its local structures and not just through sequence-specific contacts.

ACKNOWLEDGMENT

We thank Mitch Lazar for the RevErbα expression construct, Sepideh Khorasanizadeh for assistance with crystallography, and Robert Kretsinger and Christine Wright for helpful discussions. We are grateful to the staff at beamline X31, DORIS storage ring, DESY, Hamburg, Germany, and the staff at beamline X4A, NSLS, Brookhaven National Lab, for assistance with data collection.

REFERENCES

1. Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., et al. (1995) *Cell 83*, 835−839.
2. Giguere, V. (1999) *Endocr. Rev. 20*, 689−725.
3. Nuclear Receptor Nomenclature Committee (1999) *Cell 97*, 161−163.
4. Bonnelye, E., Vanacker, J. M., Desbiens, X., Begue, A., Stehelin, D., and Laudet, V. (1994) *Cell Growth Differ. 5*, 1357−1365.
5. Chawla, A., and Lazar, M. A. (1993) *J. Biol. Chem. 268*, 16265−16269.

6. Downes, M., Carozzi, A. J., and Muscat, G. E. (1995) *Mol. Endocrinol. 9*, 1666−1678.
7. Segraves, W. A., and Hogness, D. S. (1990) *Genes Dev. 4*, 204−219.
8. Carmi, I., Kopczynski, J. B., and Meyer, B. J. (1998) *Nature 396*, 168−173.
9. Harding, H. P., and Lazar, M. A. (1993) *Mol. Cell. Biol. 13*, 3113-3121.
10. Harding, H. P., and Lazar, M. A. (1995) *Mol. Cell. Biol. 15*, 4791−4802.
11. Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., and Sigler, P. B. (1991) *Nature 352*, 497−505.
12. Meinke, G., and Sigler, P. B. (1999) *Nat. Struct. Biol. 6*, 471−477.
13. Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993) *Cell 75*, 567−578.
14. Rastinejad, F., Perlmann, T., Evans, R. M., and Sigler, P. B. (1995) *Nature 375*, 203−211.
15. Rastinejad, F., Wagner, T., Zhao, Q., and Khorasanizadeh, S. (2000) *EMBO J. 19*, 1045−1054.
16. Zhao, Q., Khorasanizadeh, S., Miyoshi, Y., Lazar, M. A., and Rastinejad, F. (1998) *Mol. Cell 1*, 849−861.
17. Zhao, Q., Chasse, S. A., Devarakonda, S., Sierk, M. L., Ahvazi, B., and Rastinejad, F. (2000) *J. Mol. Biol. 296*, 509−520.
18. Khorasanizadeh, S., and Rastinejad, F. (1999) *Curr. Biol. 9*, R456−458.
19. Khorasanizadeh, S., and Rastinejad, F. (2001) *Trends Biochem. Sci. 26*, 384−390.
20. Rastinejad, F. (2001) *Curr. Opin. Struct. Biol. 11*, 33−38.
21. Wilson, T. E., Fahrner, T. J., Johnston, M., and Milbrandt, J. (1991) *Science 252*, 1296−1300.
22. Wilson, T. E., Fahrner, T. J., and Milbrandt, J. (1993) *Mol. Cell. Biol. 13*, 5794−5804.
23. Wilson, T. E., Paulsen, R. E., Padgett, K. A., and Milbrandt, J. (1992) *Science 256*, 107−110.
24. Olson, D. P., and Koenig, R. J. (1997) *J. Biol. Chem. 272*, 9907−9914.
25. Katz, R. W., and Koenig, R. J. (1993) *J. Biol. Chem. 268*, 19392−19397.
26. Medvedev, A., Yan, Z. H., Hirose, T., Giguere, V., and Jetten, A. M. (1996) *Gene 181*, 199−206.
27. 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) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.
28. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
29. Kleywegt, G., and Jones, T. A. (1996) *Acta Crystallogr., Sect. D 52*, 829−832.
30. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
31. Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999) *J. Mol. Biol. 287*, 877−896.
32. Luscombe, N. M., Laskowski, R. A., Westhead, D. R., Milburn, D., Jones, S., Karmirantzou, M., and Thornton, J. M. (1998) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 1132−1138.
33. Lavery, R., and Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.
34. Lu, X. J., Shakked, Z., and Olson, W. K. (2000) *J. Mol. Biol. 300*, 819−840.
35. Evans, S. V. (1993) *J. Mol. Graphics 11*, 134−138, 127−128.
36. Chen, L., Glover, J. N., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) *Nature 392*, 42−48.
37. Galburt, E. A., Chadsey, M. S., Jurica, M. S., Chevalier, B. S., Erho, D., Tang, W., Monnat, R. J., Jr., and Stoddard, B. L. (2000) *J. Mol. Biol. 300*, 877−887.
38. Najmudin, S., Cote, M. L., Sun, D., Yohannan, S., Montano, S. P., Gu, J., and Georgiadis, M. M. (2000) *J. Mol. Biol. 296*, 613−632.
39. Yokoyama, H., Mizutani, R., Satow, Y., Komatsu, Y., Ohtsuka, E., and Nikaido, O. (2000) *J. Mol. Biol. 299*, 711−723.
40. Schneider, T. R. (2000) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 56*, 714−721.
41. Dickerson, R. E., and Chiu, T. K. (1997) *Biopolymers 44*, 361−403.
42. Dickerson, R. E. (1998) *Nucleic Acids Res. 26*, 1906−1926.
43. Suzuki, M., Amano, N., Kakinuma, J., and Tateno, M. (1997) *J. Mol. Biol. 274*, 421−435.
44. Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., and Zhurkin, V. B. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 11163−11168.
45. Giguere, V., Tini, M., Flock, G., Ong, E., Evans, R. M., and Otulakowski, G. (1994) *Genes Dev. 8*, 538−553.
46. Giguere, V., McBroom, L. D., and Flock, G. (1995) *Mol. Cell. Biol. 15*, 2517−2526.
47. Goodsell, D. S., and Dickerson, R. E. (1994) *Nucleic Acids Res. 22*, 5497−5503.
48. Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) *Nature 335*, 321−329.
49. Garvie, C. W., and Phillips, S. E. (2000) *Struct. Fold. Des. 8*, 905−914.
50. Fujii, Y., Shimizu, T., Kusumoto, M., Kyogoku, Y., Taniguchi, T., and Hakoshima, T. (1999) *EMBO J. 18*, 5028−5041.
51. Hines, C. S., Meghoo, C., Shetty, S., Biburger, M., Brenowitz, M., and Hegde, R. S. (1998) *J. Mol. Biol. 276*, 809−818.
52. Kim, S. S., Tam, J. K., Wang, A. F., and Hegde, R. S. (2000) *J. Biol. Chem. 275*, 31245−31254.
53. Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R. S., and Shakked, Z. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 15194−15199.
54. Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J., and Dickerson, R. E. (1996) *J. Mol. Biol. 261*, 239−254.
55. Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993) *Nature 365*, 520−527.
56. Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) *Nature 365*, 512−520.
57. Shulemovich, K., Dimaculangan, D. D., Katz, D., and Lazar, M. A. (1995) *Nucleic Acids Res. 23*, 811−818.
58. McBroom, L. D., Flock, G., and Giguere, V. (1995) *Mol. Cell. Biol. 15*, 796−808.

BI011086R