Asymmetric dimerization in a transcription factor superfamily is promoted by allosteric interactions with DNA

Abdul Kareem Mohideen Patel, Pierre Vilela, Tajith Baba Shaik, Alastair G Mcewen, Isabelle Hazemann, Karl Brillet, Eric Ennifar, Ali Hamiche, Gabriel V Markov, Vincent Laudet, et al.

https://doi.org/10.1093/nar/gkad632

Asymmetric dimerization in a transcription factor superfamily is promoted by allosteric interactions with DNA

Abdul Kareem Mohideen Patel [1,2,3,4,*], Pierre Vilela[1,2,3,4], Tajith Baba Shaik [1,2,3,4], Alastair G. McEwen[1,2,3,4], Isabelle Hazemann[1,2,3,4], Karl Brillet[5], Eric Ennifar[5], Ali Hamiche [1,2,3,4], Gabriel V. Markov[6], Vincent Laudet[7,8], Dino Moras[1,2,3,4], Bruno P. Klaholz [1,2,3,4] and Isabelle M.L. Billas [1,2,3,4,*]

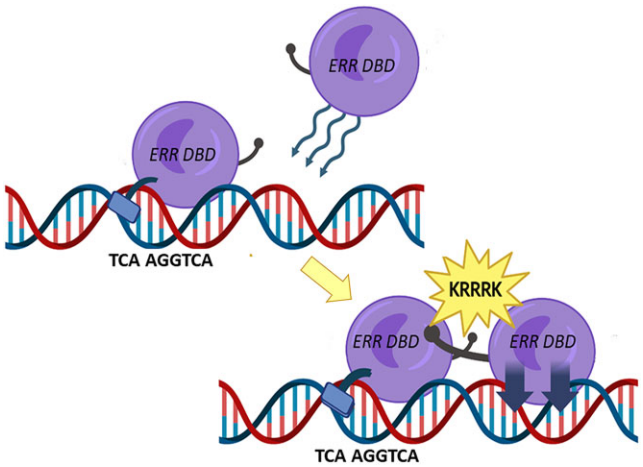
[1]IGBMC (Institute of Genetics and of Molecular and Cellular Biology), Centre for Integrative Biology (CBI), Illkirch, France, [2]Université de Strasbourg (Unistra), Strasbourg, France, [3]Institut National de la Santé et de la Recherche Médicale (INSERM) U1258, Illkirch, France, [4]Centre National de la Recherche Scientifique (CNRS) UMR 7104, Illkirch, France, [5]Architecture et Réactivité de L'ARN, CNRS UPR 9002, Institut de Biologie Moléculaire et Cellulaire, Université de Strasbourg, 67000, Strasbourg, France, [6]Sorbonne Université, CNRS, UMR 8227, Integrative Biology of Marine Models, (LBI2M, UMR8227), Station Biologique de Roscoff (SBR), 29680 Roscoff, France, [7]Marine Eco-Evo-Devo Unit. Okinawa Institute of Science and Technology. 1919-1 Tancha, Onna-son, 904-0495 Okinawa, Japan and [8]Marine Research Station, Institute of Cellular and Organismic Biology, Academia Sinica, 23-10, Dah-Uen Rd, Jiau Shi, I-Lan 262, Taiwan

Received March 12, 2023; Revised July 05, 2023; Editorial Decision July 10, 2023; Accepted July 21, 2023

ABSTRACT

Transcription factors, such as nuclear receptors achieve precise transcriptional regulation by means of a tight and reciprocal communication with DNA, where cooperativity gained by receptor dimerization is added to binding site sequence specificity to expand the range of DNA target gene sequences. To unravel the evolutionary steps in the emergence of DNA selection by steroid receptors (SRs) from monomeric to dimeric palindromic binding sites, we carried out crystallographic, biophysical and phylogenetic studies, focusing on the estrogen-related receptors (ERRs, NR3B) that represent closest relatives of SRs. Our results, showing the structure of the ERR DNA-binding domain bound to a palindromic response element (RE), unveil the molecular mechanisms of ERR dimerization which are imprinted in the protein itself with DNA acting as an allosteric driver by allowing the formation of a novel extended asymmetric dimerization region (KR-box). Phylogenetic analyses suggest that this dimerization asymmetry is an ancestral feature necessary for establishing a strong overall dimerization interface, which was progressively modified in other SRs in the course of evolution.

GRAPHICAL ABSTRACT



INTRODUCTION

Nuclear receptors (NRs) are ligand-dependent transcription factors essential for the proper regulation of the expression of genes involved in major developmental, physiological and metabolic pathways (1,2). Importantly, dysregulation in their mode of action is often linked to diseases including cancer, metabolic syndrome, reproductive failure etc., thus making NRs promising drug targets (3,4).

*To whom correspondence should be addressed. Tel: +33 388 653 203; Email: billas@igbmc.fr
Correspondence may also be addressed to Abdul Kareem Mohideen Patel. Email: mohideeen@igbmc.fr

NRs share a common modular structure composed of a highly conserved DNA-binding domain (DBD) and a less conserved ligand-binding domain (LBD), which give rise to the ligand-dependency of many, but not all NRs. These two domains are connected by a hinge region that contains the C-terminal extension (CTE), which is located right after the DBD and plays an important role in the recognition of DNA binding sites. The latter are specific genomic sequences, known as response elements (REs), located in the promoter-proximal locations or in enhancer-distal regions of target genes (1,2,5,6). These regulatory regions are usually composed of two half-sites of six base pairs each onto which NRs can bind as homodimers or as heterodimers. In fact, the dimerization of NRs is functionally advantageous, allowing the specific recognition of a larger diversity of binding sites in the genome.

The estrogen-related receptors (ERRs, NR3B) belong to the steroid receptor (SR) subfamily (NR3), which contains the vertebrate steroid receptors (ER and the oxosteroid receptors (oxoSRs) AR, PR, GR and MR) and play a crucial role in the transcriptional control of cellular energy metabolism. ERRs are orphan receptors, but they can bind synthetic ligands and react to various external factors, when high energy levels are needed by the organism (7). The ERRs are encoded by three paralogous genes (ERRα/ESRR1, ERRβ and ERRγ) in vertebrates and only one in the cephalochordate amphioxus and ascidians, which are the closest living relatives to vertebrates (8–10). Furthermore, the ERRs represent the most ancient subfamily (NR3B1) of the NR3 receptors and are present in all bilaterians including arthropods, mollusks and vertebrates (11,12). This led to the initial naming of early metazoan NR3 as ERR which is sometimes retained in the literature even if these NR3 members corresponds to early branches that are precursors of ERRs, but also of ERs and oxoSRs (13–15). In fact, the identification of hemichordate NR3s branching at the base of the NR3A (ER) and NR3C (oxoSRs) clade allowed to clearly establish that there was a late diversification of SRs among chordates (14). The SR ancestor AncSR1, which was predicted from phylogenetic tree analysis, was reconstructed and its DBD crystal structure was solved (16).

ERRα was identified based on the high similarity of its DBD sequence with that of the estrogen receptor (ERα, ESR1, NR3A1) (17). Like ERα, ERRα (as well as ERRβ and ERRγ) can bind *in vitro* to the estrogen response elements (EREs), which are composed of a palindromic PuGGTCA DNA motif with a three base-pair spacer (inverted repeat RE, IR3 RE) (18). However, in a cellular context, genome-wide studies revealed that ERRα and ERα recognize distinct DNA binding sites, ruling out the initial hypothesis of ERRα/ERα regulatory cross-talks, where ERRα would compete with ERα by binding to EREs. This suggests that ERRα and ERα differ with regards to transcriptional activity at the genomic, functional, and mechanistic levels (19). In fact, these studies and others (19–21) clearly established that ERRα preferentially recognizes DNA sequences referred to as ERR response elements (ERREs) that contain a single half-site of sequence PuGGTCA extended at its 5′-end by a 3 base pair TNA sequence, where 'C' is the most frequent nucleotide (N) (18,20,22,23). In addition, it was shown that a small, albeit functionally important, fraction of ERR binding sites contains 18-bp ERRE/ERE hybrid binding motifs composed of the extended ERRE half-site embedded into an inverted IR3 RE similar to the EREs (24). These composite ERRα/ERα binding sites, called embedded ERRE/IR3 (embERRE/IR3) REs, are found in less than ∼6% of the promoters/enhancers of target genes of ERRα/ERα. These embERRE/IR3 sites are significant to the biology of breast cancer including the genes located within the ERBB2 amplicon and GATA3 (19).

Due to the monomeric nature of ERRs binding sites, it was speculated that the ERRs can function as monomers (25,26), as do other NRs, such as Rev-Erba (NR1D1), NGFI-B (NR4A1) or LRH-1 (NR5A), by recognizing a single, extended half-site element of nine base pairs (27–30). However, functional studies have clearly demonstrated that ERR acts as a homodimer during cofactor recruitment and transcriptional activation (20,24,26,31–33). In addition, structural data show that the ERRα LBD homodimerizes in a similar fashion to that of the ERα LBD, with a stable dimerization interface (34,35). The apparent contradiction between the existence of ERRα dimerization and the monomeric nature of ERREs raises the fundamental questions of how the ERRα homodimer recognizes the DNA and the mechanisms that favor homodimerization of its DBD on DNA. These questions cannot be easily answered with the available single NMR structure that shows a monomer of the ERRβ DBD binding to its consensus ERRE (5′-TCAAGGTCA-3′) (36) and also with other NR DBD structures, in particular those of ERα (37,38), AncSR1, whose sequence was reconstituted from phylogenetic analyses (39) and monomeric NRs, such as SF1 and FTZ-F1/LRH-1 (28,40). Indeed, the molecular mechanism for the stabilization of an ERR DBD homodimer on DNA cannot be fully understood from simple comparison with the existing DBD structures. In particular, the NMR structure of ERRβ DBD is reported on a short 13 bp DNA fragment, much too short to permit dimerization.

Here, we report the first crystal structure of an ERR DBD homodimer bound to a long DNA fragment (26 bp) containing a natural embERRE/IR3 RE. The structure reveals an asymmetric arrangement and shows that the DNA sequence and its associated shape features are the key driving factors in the homodimerization of ERRα on DNA. Our structure provides a first-hand opportunity to gain insight into the molecular mechanisms of DNA recognition by ERRα which reveal a novel unexpected feature associated with DBD dimerization on DNA; in particular, the existence of an extended asymmetric dimerization interface encompassing not only the D box motif, but also the region upstream of helix H2 which includes a conserved KR-box motif. Our structural analyses of the ERRα DBD suggest that the DNA is the allosteric driver for the DBD dimer formation on DNA which brings key information about the oligomerization behavior of the ERR receptor. Thus, it sheds light on the evolutionary crossroad for a deeper understanding of SRs homodimerization.

MATERIALS AND METHODS

Cloning, protein expression, and purification

The mouse ERRα DBD (70–170 amino acids), whose DBD sequence is identical to that of human ERRα, was cloned into the in-house expression vector pnEAtH (41), which contains a His_6 affinity tag coding sequence. The vector was transformed into the *E. coli* BL21 (DE3) pRARE2 strain. Cultures were grown at 37°C and protein expression was induced at $OD_{600\,nm} = 0.6$ with 1 mM IPTG for 3 h at 25°C. The cell pellet was suspended in binding buffer (20 mM Tris pH 8.0, 400 mM NaCl, 10% glycerol, 4 mM CHAPS) and lysed by sonication. The crude extract was centrifuged at 45 000 g for 1 h at 4°C. The lysate was loaded on to a Ni affinity HisTrap FF crude column (GE Healthcare, Inc.) and the protein was eluted at a concentration of 250 mM imidazole. The hexa-histidine tag was cleaved overnight using thrombin protease. The ERRα DBD was further purified on a Heparin column using an increasing salt gradient, followed by size-exclusion chromatography (Superdex S75 16/60 column, GE Healthcare) using a SEC buffer (50 mM BisTris pH 7.0, 120 mM KCl, 0.5 mM CHAPS, 4 mM $MgCl_2$). Complexes were formed by mixing DNA to the purified ERRα DBD to achieve the desired DNA: protein molar ratios and incubated overnight at 4°C before performing the crystallization assays.

DNA sequences and annealing

HPLC purified single-stranded DNAs were ordered from Sigma-Aldrich, Inc and Euromedex, France.

The oligonucleotide DNA sequences for crystallization and ITC experiments are:

- ***embERRE/IR3*** (sense): 5′-ATG***TCA***AGGTCA CCGTGACCTTTACG-3′
- and (anti-sense): 5′-TCG***TAA***AGGTCACGGTGACCT TGACA-3′;
- ***tff1 IR3/ERE*** (sense): 5′-GGATTAAGGTCA GGTTGGAGG AGACT-3′
- and (anti-sense): 5′-CAGTCTCCTCCAACCTGACCT TAATC;
- ***traERRE*** (sense): 5′-TTTG***TCA***AGGTCACA CAGTGAGTTAC-3′ and (anti-sense): 5′-GTAACTCACTGTGTGACCT***TGA***CAAA

The oligonucleotide DNA sequences for EMSA experiments:

- BE24Cons-ERREC: 5′-TT**TCAAGGTCA**GGTTGGAG GAGAC and (anti-sense): 5′-GTCTCCTCCAACCT GACCTTGAAA
- BE24Cons-ERRET: 5′-TT**TTAAGGTCA**GGTTGGAG GAGAC and (anti-sense): 5′-GTCTCCTCCAACCT GACCTTAAAA
- BE24Cons-ERREG: 5′-TT**TGAAGGTCA**GGTTGG AGGAGAC and (anti-sense): 5′-GTCTCCTCCAAC CTGACCTTCAAA
- BE24Cons-ERREA: 5′-TT**TAAAGGTCA**GGTTGG AGGAGAC and (anti-sense): 5′-GTCTCCTCCAAC CTGACCTTTAAA
- BE24-ARE (negative control): 5′-ATGAGAGAACTggcTGACCATGGG and (anti-sense): 5′-CCCATGGTCAgccAGTTCTCTCAT

The single-stranded DNA oligonucleotide and its respective anti-sense DNA fragment were mixed in a ratio of 1:1 and heated in the DNA annealing buffer (10 mM Tris pH 8.0, 100 mM NaCl, 1% DMSO, 0.1 mM EDTA) at 95°C and gradually cooled down to 4°C using a BIO-RAD PCR thermocycler. Finally, 4 mM $MgCl_2$ was added to the annealed DNA sample.

Crystallization and data collection

The ERRα DBD protein and the embedded DNA were mixed in a ratio of 2:1 and the final concentration of the complex (5 μg/μl) was adjusted with SEC buffer. ERRα DBD/DNA complex crystals were obtained by mixing the complex with the reservoir solution in a ratio of 1:1 at 20°C using the screen Morpheus I screen from Molecular Dimensions (https://www.moleculardimensions.com/products/morpheus). Crystals appeared within 3 days. Crystals were cryo-cooled through flash-freezing into liquid nitrogen directly from the reservoir solution, which is already a cryo-protectant. Data were collected at the Proxima 2A microfocus beamline of the SOLEIL synchrotron (Gif-sur-Yvette, France), using an Eiger X 9M detector (Dectris).

Structure determination and analysis

The data showed anisotropic diffraction and were therefore processed with the autoPROC suite (42). The space group of the crystal is $P2_12_12$ ($a = 168.11$ Å, $b = 70.82$ Å, $c = 94.84$ Å, α, β, γ = 90°), with two copies of the complex in the asymmetric unit and an approximate solvent content of 57.7%. The structure was determined by molecular replacement using the ERα DBD/IR3 structure (PDB code 1HCQ) (38), as an initial model using the Phaser software (43) in the Phenix suite (44). A primary search was carried out using the entire DNA/protein model. The structure was refined iteratively using Phenix software through model building with COOT (45) and refined with Buster (46), which allows considering all reflections for refinement regardless of the crystal orientation related to the anisotropic diffraction. Overall map features were improved using the feature-enhance map (FEM) routine of the Phenix software (47). In addition, a few key residues were refined individually with the Polder map from Phenix suite (48), where the Polder map represents an omit map that excludes the bulk solvent around the omitted region and hence allows to visualize weak densities close to the bulk solvent. FEM and Polder map analysis procedures are part of the Phenix suite (44). The DNA geometric parameters were calculated using 3DNA (49). All the structural figures were prepared using PyMol (PyMOL Molecular Graphics System, Version v 2.5.4, Schrödinger, LLC).

Electromobility shift assay

Native polyacrylamide gels were prepared to a final concentration of 5% using an Acrylamide:Bisacrylmide solution

with 10 mM HEPES-KOH (pH 7.5) and several concentrations of $MgCl_2$. Gels were pre-run at a constant voltage of 120 V for 45–60 min. The different DNA oligonucleotide-protein complexes were reconstituted and incubated for at least 20 min on ice before running at a constant voltage of 150 V for 2–3 h, followed by staining with ethidium bromide.

Isothermal titration calorimetry

The purified DBD and the annealed DNA fragment were dialyzed separately overnight against the SEC buffer using a dialysis unit (Thermo Fisher Scientific) with a 10 kDa cut-off membrane. Samples were diluted to the working concentration and degassed prior to the experiments. ITC data were obtained on a Malvern Panalytical PEAQ-ITC microcalorimeter. Measurements were performed by titrating the protein (cell) with DNA (syringe). The heat of dilution for the protein and DNA were obtained by titrating them separately against the buffer. Experiments were performed at 10, 15, 20, 25 and 30°C for the embERRE/IR3 DNA in order to evaluate the temperature-dependency of the protein/DNA interaction. Data were processed with the MicroCal PEAQ-ITC Analysis software and with the AFFINImeter v 1.2.3 software using a two binding site model (50). ITC data were also collected for the binding of the ERRα DBD to IR3/ERE DNA and to the ERRE extended half-site DNA (at 15 and 20°C, respectively, in order to optimize the signal/noise ratio for these DNA sequence) and to the ERRE extended half-site DNA (at 20°C).

Sequence and evolutionary analysis

The ERR DBD sequence was used to perform a blast search against the protein data bank. Multiple sequence alignment was carried out using the MAFFT program. Sequence alignments were checked and manually edited using Jalview. Final figures were prepared using ESPript. Collected NR sequences were aligned using Clustal Omega (51). Alignments were checked manually and edited with Seaview (52). Phylogenetic trees were built using PHYML (53) using the LG model (54) with a gamma law. The reliability of the nodes was assessed by the likelihood-ratio test (55). Ancestral character reconstruction and stochastic mapping (56) were performed under R version 4.1.2 (57) using the make.simmap function as implemented in the phytools package version 1.0-1 (58). Character evolution was inferred using a model of symmetrical transition rates between the character states (SYM). 10000-character histories were sampled to allow the incorporation of the uncertainty associated with the transition between different states. Inferred state frequencies for ancestral nodes were plotted using the describe.simmap function.

RESULTS

Overall architecture of ERRα DBD/DNA crystal structure

To gain insight into the molecular mechanisms of DNA recognition by the ERRα DBD homodimer, we reconstituted, characterized, and crystallized the complex between the protein (mmERRa-70-170, Figure 1A and B) and a 26 base-pair DNA containing an embERRE/IR3 RE (Figure 1C) which had been shown to stabilize the homodimeric form of the ERRα DBD (59). The asymmetric unit contains two equivalent DNA-protein complexes (RMSD 0.83 Å) arranged opposite to each other, with the subunits DBD1 and DBD2, forming a head-to-head homodimer on the DNA. An additional subunit (DBD3) is seen on the opposite side of the DNA due to crystal packing, but this does not affect the homodimer (Figure S1A and B and Table S1). The structure of one of the two DNA complexes found in the asymmetric unit is shown in Figure 1D, and the main protein-DNA interactions depicted in Supplementary Figure S1C. DBD1 is bound to the TCA 5′-extended half-site and DBD2 is bound to the second 5′-TAA extended half-site on the complementary DNA strand (i.e. located at the 3′-end of the RE). These two different sequences offer the possibility to compare the role of different 5′-extensions in the same complex (Figure 1C).

Asymmetry of the dimerization interface

The DNA sequence is key for the stabilization of the ERRα DBD homodimer. The embERRE/IR3 RE is composed of two palindromic half-sites which are flanked by two different 5′-extensions (Figure 1C). However, since two half-sites are symmetric, we asked whether the protein dimer is arranged in a symmetric or asymmetric manner. At first look, the ERRα subunits DBD1 and DBD2 are organized in a similar manner to the homodimers of SRs bound to their respective IR3 REs (ER, GR, AR, PR and MR) (37,38,60–64). The two DBDs contact each other through their D-box motif (Pro116 to Glu120 in Figure 1B) *via* a classical 'hand-shake' mode of interaction (Figure 2A). The dimerization interface comprising the two D-boxes is formed by hydrophobic packing between the methyl groups of Ala117 in each DBD and from the van der Waals contacts of the Pro116 pyrrole ring which underlies the Ala117 methyl group (Figure 2A). Besides this, polar contacts are also formed with the Pro116 backbone carbonyl group of DBD1 establishing a hydrogen bond with the amide group of Glu122 of DBD2 and *vice versa*, with relatively symmetric bonding schemes and equivalent interaction distances (3.1 and 2.9 Å, Figure 2A).

However, a clear conformational asymmetry is seen at the interface of the two DBDs, in the region following the D-box and prior to helix H2 (Glu122-Ala130 in Figure 1B). In DBD1, this region comprises a loop between Glu122-Thr124 and a 3_{10} helix formed by residues of the KRRRK motif from Lys125-Lys129 (hereafter referred to as 'KR-box'; Figures 1B, D and 2B) with a topology similar to other DBD structures, *e.g.* the NMR structure of ERRβ/ERRE (PDB ID: 1LO1 (36), and the crystal structures of the ERα DBD (PDB ID : 1HCQ (38), and AncSR1 DBD (PDB ID: 4OLN (16,39). In stark contrast, in DBD2 a conformational difference is observed in the KR-box region, which folds as a loop, thereby extending the dimerization region. The DBD arrangement therefore becomes asymmetric (Figure 2B). The KR-box residues are strictly conserved for all the three ERR subtypes highlighting the functional importance of this structure (Figure 2C and Supplementary Figure S2A).

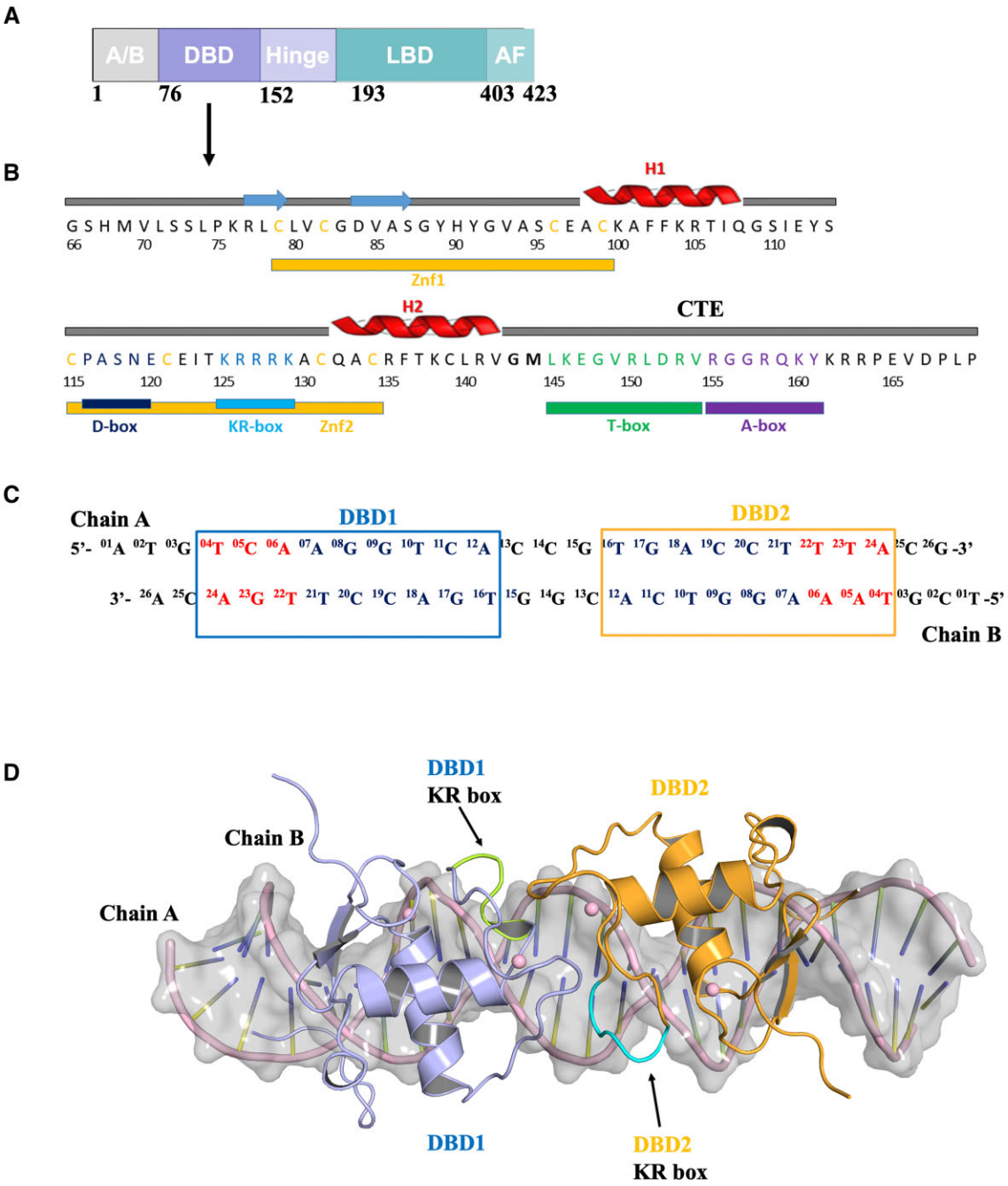


Figure 1. (**A**) Domain organization of ERRα. (**B**) ERRα DBD protein sequence and schematic secondary structure representation. (**C**) Sequence of the 26 bp DNA fragment, embERRE/IR3, used for crystallization where the two binding sites for DBD1 and DBD1 are outlined with blue and orange boxes, respectively. (**D**) Cartoon representation of the homodimer of ERRα DBD bound to the embERRE/IR3 response element, with DBD1 in light blue and DBD2 in orange color. The KR-box indicated in (B) is highlighted on the structure, in lemon green for DBD1 and teal for DBD2.

The asymmetric dimerization interface of the ERRα DBD results in an increased dimerization area which arises from the interactions between residues of the KR box motif of DBD2 and DBD1 (ASA: 486 Å^2 versus 468 Å^2, for the KR-box of DBD1 and DBD2 respectively). The side chains of Arg128 and Lys129 of DBD1 interact with the DNA phosphate backbone and Arg127 makes a contact with Ser114 of DBD2 (Figure 2D and F). In DBD2, the Arg128 side chain participates in protein-protein interactions with DBD1 by forming an extended-loop with the help of Lys129 that in turn makes an intra-hydrogen bond with Ala130 (DBD2), a conserved residue unique to the ERRα sequences. This interaction is not possible for Ala130 of DBD1, due to the dissimilar conformations of the KR boxes. Thus, it contributes to the asymmetry of the interface and to the asymmetric distribution of electrostatic potential in the 'KR-box' regions. Furthermore, Arg127 of DBD2 makes additional contacts with Ser114 and Pro116

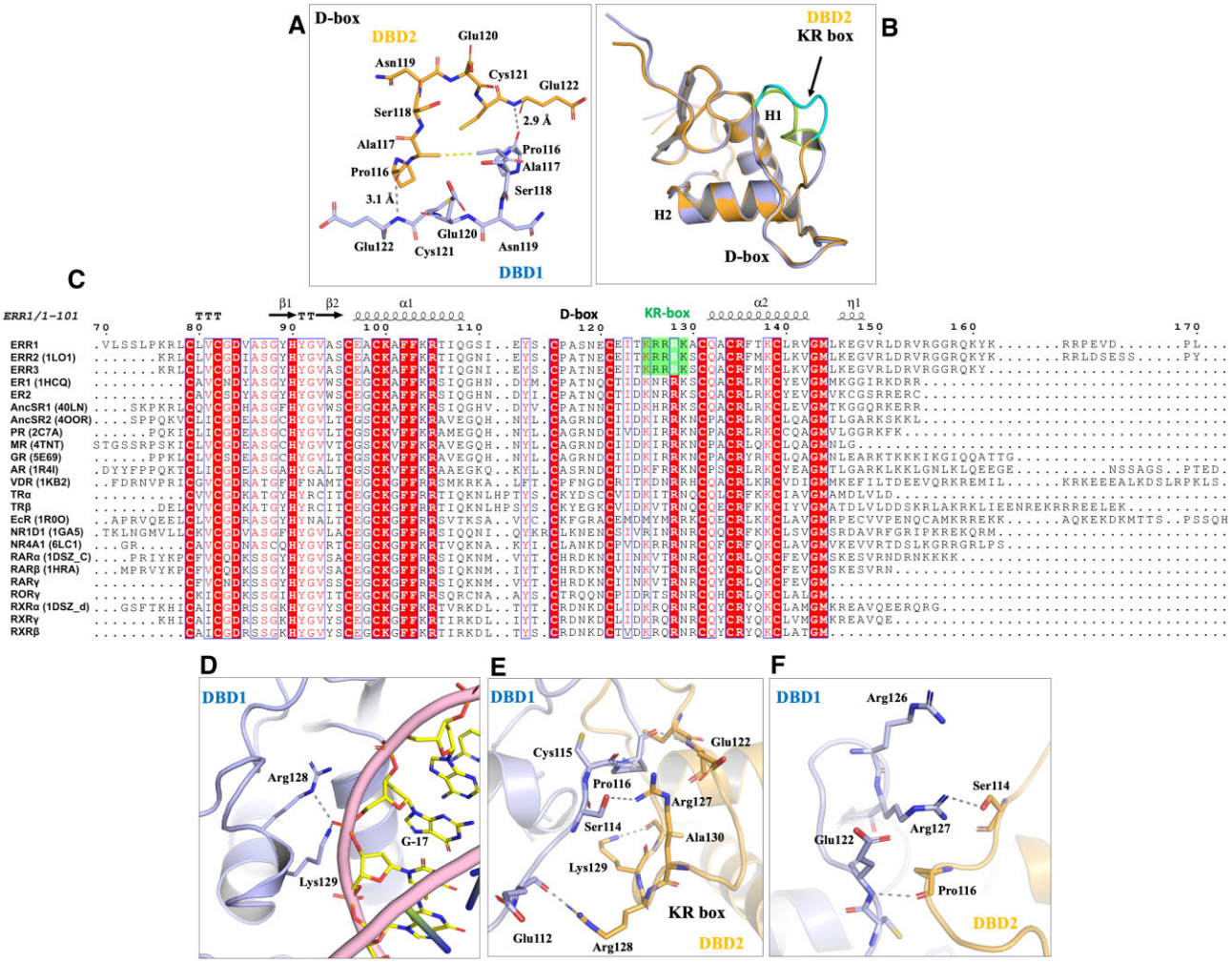


Figure 2. The asymmetry of the dimerization interface involves a conserved KR-box motif. (**A**) Enlarged view of the residues belonging to two D-boxes (DBD1 and DBD2) that form the canonical dimerization interface. (**B**) Cartoon representation of the KR-box region (lemon green for DBD1 and teal for DBD2) for the subunits superimposed, with DBD1 in light blue and DBD2 in orange color. (**C**) Multiple sequence alignment of the human ERR DBD sequences with those of the other steroid NRs (NR3 subfamily (AncSR1, ER, AncSR2, AR, GR, MR, PR) and NRs of the subgroups NR1 (TR, RAR, PPAR, ROR, VDR), NR2 (RXR and HNF4), NR4 (NGF-IB, SF1), showing KR box conservation. The sequence numbering corresponds to the human ERRα sequence and the colors relate to the percentage of sequence identity. The top panel represents the secondary structure elements of ERRα. The KR box region of the ERRs is highlighted in green. (**D**) Stick representation of the interactions between the KR box residues Arg128 and Lys129 of DBD1 with the DNA phosphate backbone of G17 of the complementary strand (chain B). (E, F) Enlarged view of the dimerization interface interactions between DBD1 and DBD2 looking at the region of the KR-box of (**E**) DBD2 and (**F**) DBD1. Residues are shown with a stick representation, with carbon atom in light blue for DBD1 and orange for DBD2, red for oxygen, blue for nitrogen and yellow for sulfur atom.

(Figure 2E and Supplementary Figure S2B–D). Altogether, the structure of the ERRα DBD homodimer shows that the conformational adaptations of the DBD2 D-box and KR-box regions lead to an asymmetric arrangement on the almost symmetric DNA resulting in stabilization of the dimer on its cognate DNA RE.

Evolution of the KR box in the NR3 subfamily

To better understand the evolutionary forces acting on the KR box, we performed a phylogenetic analysis within the NR3 family. We categorized 8 main types of possible KR boxes with the upstream T124 residue which are shown in Supplementary Figure S2E-F, allowing substitutions only with amino acid residues having equivalent physicochemical properties: (i) the relatively weakly constrained motif that contains two positions able to vary and which is found as DKQRRK in the NR3F sequences of the early diverging eumetazoan *Trichoplax*, as DRQ/NRRK in some annelid NR3Ds, and as DK/RNRRK motif in chordate ERs; (ii) the relaxed DK/RMSRK motif found in the cnidarian specific NR3Es (13); (iii) a NKMSRK variant specific to the NR3E from Hydra; (iv) the highly conserved ERR KR box with the motif KRRRK that is present in all ERRs of almost all bilaterians; (v) a SKQRRK variant specific to ERR from *Priapulus caudatus*; (vi) the DKHRRK motif, which is the most abundant in bilateria before the divergence of chordates at the AncSR1 node and (v) the DKI/FRRK motif from vertebrate SRs and; (vi) the unique DRRLKK motif found in the amphioxus SR.

We then mapped the occurrence of these 8 types on a simplified version of the updated phylogeny as shown in Supplementary Figure S2F, which is fully consistent in its topology with previous publications based on a similar dataset (14). The ancestral state reconstruction for every node of the phylogeny illustrates successive changes in constraints with a basal DKQRRK motif found in the early diverging eumetazoan Trichoplax, followed by an episode of high divergence in Cnidarians, leading to loss of the ancestral NR3 in some species (anthozoans) and retention of a diverging DK/RMSRK motif in medusozoan cnidarian NR3Es, where the otherwise almost universally conserved R in the third position is switched to an S, and a further derived motif in Hydra, where the D at first position in the motif switched to N. From the observations that the ancestral DK/RxRRK motif (in blue in Supplementary Figure S2F) is present in two distant parts of the tree (SRs/NR3Ds, and Trichoplax NR3F), we conclude that it must have been the one present at the basal nodes, including AncSR1, but we cannot formally exclude a convergent appearance of the DK/RxRRK in those two branches. However, even removing the basal *Trichoplax* sequences does not change the inference of the ancestral state, given the long branch in cnidarian NR3Es. The ERR type (T/NKRRRK) appears then as a specific subtype of the ancestral DK/RxRRK motif that has remained remarkably well constrained during bilaterian evolution. Overall, our evolutionary analysis highlights the structural and functional importance of the ERR KR box and suggests that it is a constrained subtype of the ancestral motif DK/RxRRK present in vertebrate steroid receptors.

Differential interactions of DBD1 and DBD2 on ERRE half sites

A strong asymmetry in the extended DBD dimerization region is observed even though the sequence of the two half-sites is identical. Thus, we question whether the interactions seen between the core of the DBDs and the DNA major grooves composing the half-sites are similar. For this, we superimposed the DBD2-ERRE on the DBD1-ERRE and compared these also to other NR DBD structures. The recognition helix (H1) of ERRα interacts with the DNA major groove half-site via base-specific interactions with its residues Glu97, Lys100, Lys104 and Arg105 (Figure 3A). These interactions are conserved in ERα and other PuGGTCA-binding NRs, such as RXR, RAR or VDR (6). In particular, Glu97 has a key role in discriminating between AGGTCA from AGAACA half-site sequences which are also recognized by oxoSRs (1,5). A similar binding pattern is observed in the two DBD subunits, but with some important qualitative differences in the electron density maps of the residues interacting with the bases in the DNA major grooves. For DBD1, clear H-bond interactions are observed between the side-chains of Glu97, Lys100, Lys104, and Arg105 and DNA bases of the major groove (Figure 3A and Supplementary Figure S3A). Instead, in DBD2, the side chains of Glu97 and Arg105 (clearly defined in the electron density map) are oriented towards C19 and G17 in chain A (Figure 3B), but Lys100 and Lys104 could not be defined from the weak electron density map and appear disordered (Supplementary Figure S3B). This shows that the stabilization of DBD2 on the DNA is not strong compared to DBD1. In accordance with this notion, the experimental temperature factor of these residues, which is a measure of the disorder of each atom position, is higher in DBD2 than in DBD1, demonstrating a slightly more flexible conformation for DBD2. Likewise, the core of the ERRα DBD is inherently stabilized by a patch of hydrophobic residues from helix H1 and H2 and this unique feature is absent in ERα (Figure 3C).

Taken together, the analysis of the interactions between ERR-specific residues and the DNA major grooves are conserved, but DBD1 and DBD2, have stronger interface involvement to their cognate half-site. It further emphasizes a unique asymmetry not only at the level of the dimer interface, but also for the mode of interaction of the two individual subunits with the DNA major groove of each respective half-site.

The 5′ DNA extension as a key determinant of the A-box conformation

Our structure offers a unique opportunity to study in atomic details the interactions between the DBD and a characteristic 5′ TCA (for DBD1 and DBD3) or a less frequent 5′ TAA (for DBD2), extended half-site. The CTE region contains the T- and A-boxes (Figure 1B), which are visible in the electron density until the main chain of Arg158 for DBD1 and DBD2. The T-box consists of one-turn of a 3_{10} helix (Lys146-Gly148) with Val149 packing against Phe103 and Ile107 of helix H1 before crossing the phosphate backbone and making a β-turn formed by residues Arg150 to Arg153 (Figure 4A), similar to the NMR structure of the ERRβ DBD monomer on ERRE (36). The A-box region (Arg155–Lys162) comprises an AT-hook (Arg155, Gly156, Gly157 and Arg158) and represents the main motif for the interaction with the 5′ DNA extension by insertion of this loop into the minor groove. The interaction pattern of the AT-hook motif is similar to that seen in the monomeric ERRβ DBD, with two glycine residues actively participating in the interactions with the 5′ TCA extension (Figure 4B). Since the electron density of the side chain of Arg158 was rather weak, we performed a polder (omit) map analysis to improve the local electron density, which excludes bulk solvent around the omitted region (48). This analysis suggests that Arg158 is inserted into the minor groove which forms an H-bond with the O4 atom of the ribose of C25 on the complementary strand (Supplementary Figure S4A and B). On the other hand, compared to the ERRβ DBD solution structure, additional specific interactions that stabilize the AT-hook motif are seen for DBD1. In particular, salt bridge interactions between the side chains of two residues of the T-box, Arg150 and Asp152 help maintain a stable conformation of the AT-hook which interacts with the DNA backbone (Figure 4A). These observations contrast with those seen for DBD2 and DBD3, where the side chain of Arg150 interacts with the carbonyl moieties of Val154, Gly156 (A-box) and Gly157 (Figure 4C). In stark contrast, in DBD2, the AT hook glycine residues Gly156 and Gly157 do not interact with the minor groove of the 5′ TAA extension, leading to loss of specific interactions with

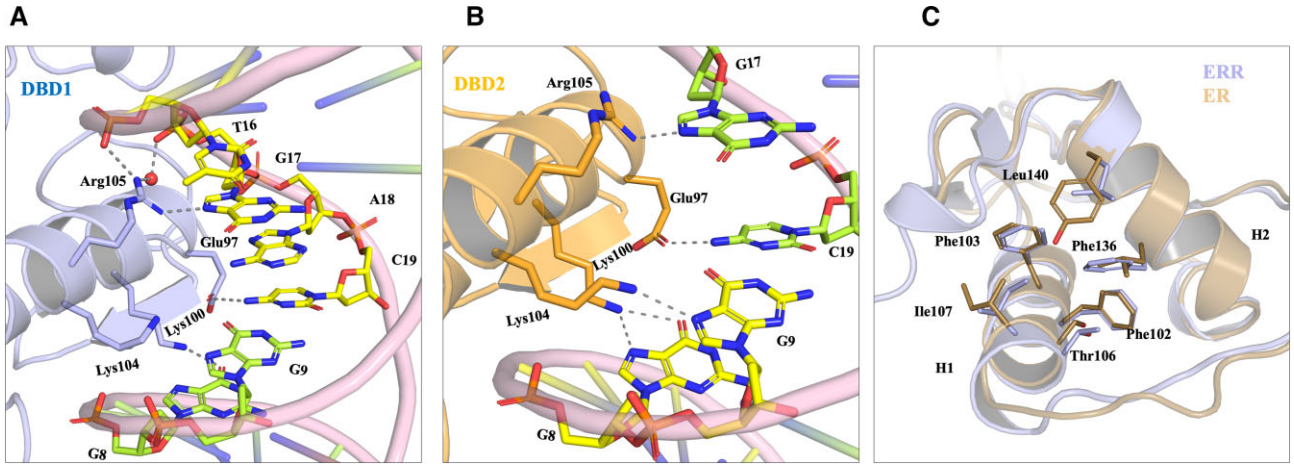


Figure 3. Major groove interactions in DBD1 and DBD2. (A, B) The residues Glu97, Lys100, Lys104 and Arg105 in H1 interact with the major groove half-site via base-specific interactions, in a similar manner for DBD1 (**A**) and DBD2 (**B**), but the interaction network is much stronger in DBD1 than in DBD2, as indicated by the structural water molecules seen in the electron density of DBD1, but not of DBD2. In addition, the side chains of residues Lys100 and Lys104 are only well defined in DBD1, but not in DBD2 (see Supplementary Figure S3A and B). (**C**) Enlarged view of the superimposition of one subunit of ERRα DBD (DBD1 in light blue) with the corresponding ERα DBD in the region encompassing H1 and H2. The core of ERRα DBD is stabilized by a pocket of hydrophobic residues. In ERα, Phe136, Leu140 and Thr106 are replaced by Ile, Tyr, and Ser, respectively, weakening the contribution of hydrophobic contacts. The Phe136 residue in ERRα, replaced by Leu in ERα, further strengthens the core interactions through an intricate network of π-π stacking interactions.

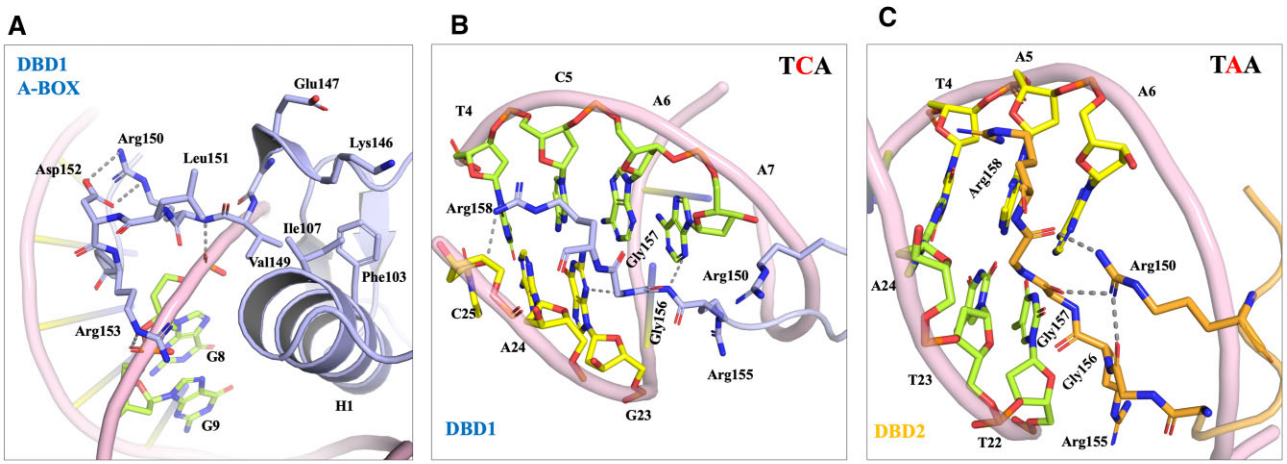
the second base (Figure 4C). Improvement of the electron density with the polder map analysis for Arg158 suggests an H-bond with the N3 position of A5 in chain B (Supplementary Figure S4B). Consistently, the strong H-bonds seen in the TCA 5′-extension are lost in the TAA 5′ extension. These observations unravel for the first time the molecular details of the role played by the 5′ extension in the stabilization of the first DBD on DNA, highlighting the importance of the central base C/G in the TNA 5′-extension. Lastly, our structure clarifies the functional role of the Tyr161 residue of the A-box. In fact, this residue does not directly stabilize the core of the protein as reported in the solution structure (36), but, instead, plays a key role in the stabilization of the CTE through interaction with the phosphate backbone of the 5′-extension (Supplementary Figure S4D).

Our structural observations of the TCA/TAA differential interactions are supported by electrophoretic migration shift assays (EMSA) of the full-length ERRα bound to ERRE response elements containing different types of 5′ extensions, including TCA, TTA, TGA and TAA (Figure 4D). In EMSA, the concentration of $MgCl_2$ was progressively increased in order to determine the conditions for the best binding specificity. Simultaneously, we examined the binding of ERRα to a negative control DNA sequence for increasing $MgCl_2$ concentration. Above 50 μM $MgCl_2$, we lose the non-specific binding of the protein to the negative control sequence. At higher concentration of $MgCl_2$ (100 μM $MgCl_2$ (gel III) and 250 μM $MgCl_2$ (gel IV)), the binding of ERRα to the TTA (lane 4) and TAA (lane 6) containing ERREs is progressively lost, in contrast to the TCA (lane 3) and TGA (lane 5) containing ERREs. This suggests that the binding to the TCA/TGA 5′extensions is stronger than to a TTA/TAA extension. Together with the structural analysis, our data unveil the molecular mechanisms for the preferential binding of the ERRα CTE to TCA and TGA 5′ extended half-sites.

Asymmetry in the positioning of the ERRα DBD homodimer on DNA as compared to related NRs

To gain a global understanding of the binding mode of the ERRα DBD homodimer on DNA, we compared the structure of ERRα DBD with ERα DBD (37,38) and the DBD of the reconstructed ancestral steroid receptor AncSR1 (16). The superimposition of the ERRα DBD homodimer with that of the ERα DBD (DBD1 subunit as a reference) shows that the corresponding second subunit (DBD2) is not placed in the same manner (Figure 5A). In fact, the ERRα DBD2 is rotated counter-clockwise when looking down the DNA axis from the 3′-end compared to the symmetrical ERα DBD (Figure 5B). This movement impacts the position of the two major helices, H1 and H2, and affects the positioning of the D-box loops and the extended dimerization region including the KR-box. When the dimer interface is visualized along the pseudo-symmetry axis, the entire dimerization region of the second subunit is strongly displaced (Figure 5B).

We further noticed that the AncSR1 DBD homodimer bound to IR3/ERE also presents an asymmetric conformation that has not been originally highlighted (Figure 5C and D). The KR-box of one AncSR1 DBD subunit is folded as a 3_{10} helix with the last two residues interacting non-specifically with the DNA, while the KR-box of the other subunit unfolds and adopts a loop conformation and interacts with the neighboring subunit. However, the loop of AncSR1 is kinked inward and forms a tiny dimerization interface. In stark contrast, in the ERRα DBD dimer, the KR box loop extends outwards forming a larger interface which contributes to an increased dimerization surface (Figure 5C and D). The similarity between ERRα and AncSR1 suggests that the asymmetry of the dimerization interface is an ancestral feature of receptor homodimerization which is necessary for establishing strong contacts between the two



Lane No.	Sample
1	Marker
2	Free DNA
3	TCA
4	TAA
5	TGC
6	TTA
7	ARE control

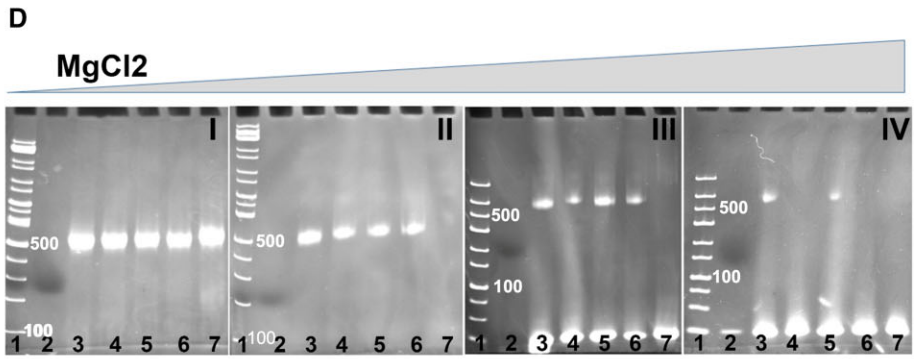


Figure 4. TCA and TAA 5′ extensions are key in the conformation of the A-box. (**A**) In DBD1, the conformation of the T-box (residues 145–154) adjacent to the A-box (residues 155–161) is well defined and helps the A-box loop to penetrate into the minor groove of the 5′ extension. (**B**) Interaction pattern of the CTE residues of DBD1 with the TCA 5′extension. (**C**) Interaction pattern of the CTE residues of DBD2 with the TAA 5′extension. Only a few H-bond interactions are observed, demonstrating the weak interaction at TAA 5′extension. (**D**) EMSA analyses show the significant role of 5′- extension TNA nucleotides. EMSA gel for different concentration of $MgCl_2$ 25, 50 100, 250 μM (gels I to IV); lane 1: marker, lane 2; free DNA control; lane 3 TCA ERRE; lane 4: TTA ERRE; lane 5: TGA ERRE; lane 6: TAA ERRE and lane 7: ARE (Androgen receptor RE) DNA as a control DNA sequence (see Materials and Methods for sequences).

D-box regions and the extended dimerization region and thus increasing the overall dimerization interface.

DNA is the allosteric driver for cooperative ERRα DBD dimerization

Isothermal titration calorimetry (ITC) was used to gain biophysical insights into the dimerization mechanism of the ERRα DBD on an embERRE/IR3. Our previous work showed that the ERRα DBD could behave as a monomer on different types of DNA REs (such as the *tra* ERRE and the *tff1* IR3/ERE, but not on embERRE/IR3) (59). The analysis of the ITC data indicates that a dimer of the ERRα DBD binds to the embERRE/IR3 DNA (stoichiometry $n = 0.5$) in the nanomolar range ($K_d = 4$–25 nM) (Supplementary Figure S5A–E and Table S2). The event is strongly enthalpy driven in the 10–30°C range with a large unfavorable binding entropy. Temperature-dependence analysis shows that the heat capacity change ΔCp for embERRE DNA binding is highly negative (ΔCp = –0.473 kcal/mol/K, Supplementary Figure S5F), which is indicative of a significant local folding coupled to site-specific DNA binding (65). An initial non-specific binding event is observed at lower molar ratios, when the protein is in excess over DNA, with unfavorable binding enthalpy and positive entropic contribution. These un-specific electrostatic interactions can be attributed to the avidity of the DBD for DNA at high protein concentration (66–68). In contrast, ITC data show that only one monomer of the ERRα DBD binds specifically to the *tra* ERRE (Supplementary Figure S5G) and to the *tff1* IR3/ERE (Supplementary Figure S5H) in agreement with our previous biophysical results (59). However, the ITC data analysis indicates significantly different thermodynamic profiles for these two types of REs. Regarding the *tff1* IR3/ERE, one monomer binds specifically (stoichiometry $n = 1.0$) in the low nanomolar range ($K_d = 28.2$ nM) and this interaction is highly enthalpy-driven ($\Delta H_1 = -11.7$ kcal/mol) with an unfavorable entropic component ($-T\Delta S_1 = 1.7$ kcal/mol) (Supplementary Figure S5H and Table S2). A second unspecific binding event, likely of a second monomer ($n = 0.7$), is observed on this DNA

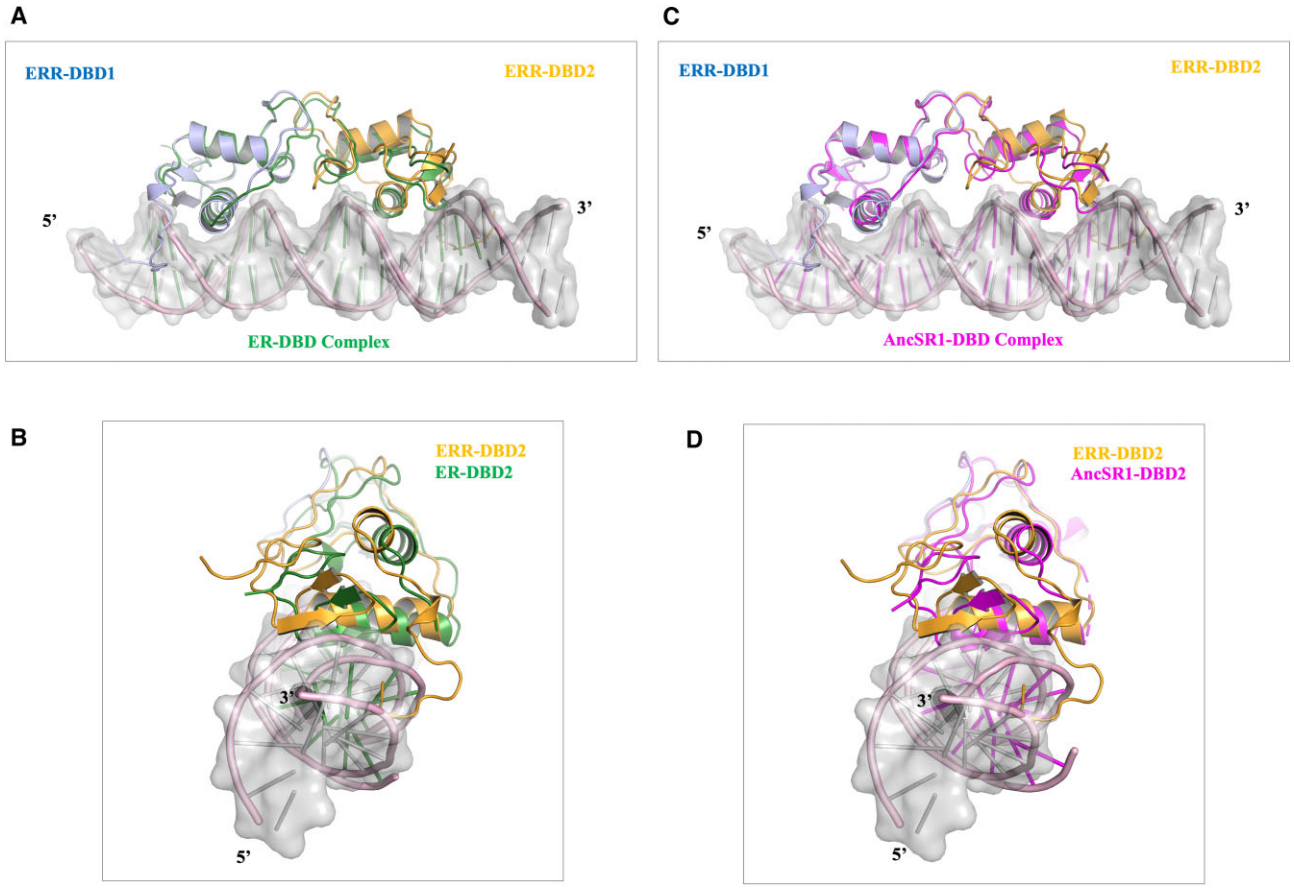


Figure 5. The KR-box motif is an ancestral feature crucial for the asymmetric organization of ERRα DBD on DNA compared to ERα DBD and AncSR1 DBD. (**A, B**) Superimposition of ERRα DBD and ERα DBD on DNA (2 orientations, with a 90° rotation between them). (C, D) Superimposition of ERRα DBD and AncSR1 DBD on DNA, with (**C**) a global view and (**D**) an enlarged view in the KR-box region of DBD2. ERRα DBD1 is colored in light blue, ERRα DBD2 in orange, ERα DBDs in green, and AncSR1 in magenta. The views of the superimposition along the DNA axis (in B and in D, for ERα and AncSR1, respectively) emphasize the shift of ERRα DBD2 positioning on DNA.

sequence and results from a strongly entropy-driven interaction (ΔH_2 = 0.9 kcal/mol, $-T\Delta S_2$ = –12.3 kcal/mol), which is a common feature of unspecific binding events (Privalov *et al.*, 2010). Regarding the binding of the ERRα DBD to the *tra* ERRE extended half-site, a monomer binds in the low nanomolar range (stoichiometry $n = 0.9$, $K_d = 26.8$ nM) with an enthalpy-driven interaction (ΔH_1 = –17.5 kcal/mol, $-T\Delta S_1$ = 7.3 kcal/mol), as well as a second, likely unspecific binding of dimer (n = 0.53), with a rather similar binding profile (n = 1.3, K_d = 29.7 nM, ΔH_2 = –16.1 kcal/mol, $-T\Delta S_2$ = 6.0 kcal/mol). The ITC data are furthermore consistent with native gel electrophoresis analysis of the protein/DNA complexes, where a first monomer and then a dimer are seen for the embERRE/IR3 RE (Supplementary Figure S5I), but only a monomer is observed for the *tff1* IR3/ERE DNA (Supplementary Figure S5J).

The ITC analyses indicate the dimerization of the ERRα DBD on the embERRE/IR3 is driven by the DNA. It is well established that in addition to the base-specific recognition mechanism, the DNA shape plays a key role in the specific recognition of DNA by proteins. The DNA shape parameters refer to the geometrical and physical properties of DNA, such as minor and major groove widths, twist and roll parameters, and can be predicted based on the DNA sequence {DNAshape software (69,70)}. When applied to the various ERR binding sequences, the analysis suggests that the sequence-dependent DNA shape is a key structural component promoting dimer formation (59). Our crystal structure thus gave us the possibility to calculate the experimental DNA structural parameters and to discuss them in light of our initial hypotheses (49) (Figure 6A–D). The minor groove width (MGW) in the experimental structure shows two strong minima in the middle of each half-site (Figure 6A) as predicted from the sequence, confirming our hypothesis that the second DBD requires a narrow minor groove for dimerization (59). However, the minima of the experimental and predicted MGW profiles are not identical. In the prediction, the two minima are located at position ±6 (corresponding to the second nucleotide of the half-site). This is true for the first minimum of the experimental MGW profile (Figure 6A), but not for the second one, which is found at position +5, and thus displaced by one nucleotide with respect to the predicted symmetrical value. Hence, the experimental MGW profile is asymmetric despite an equivalent half-site composition. This suggests that the binding of the dimer is conditioned by a narrow DNA MGW, but it affects reciprocally the DNA shape in

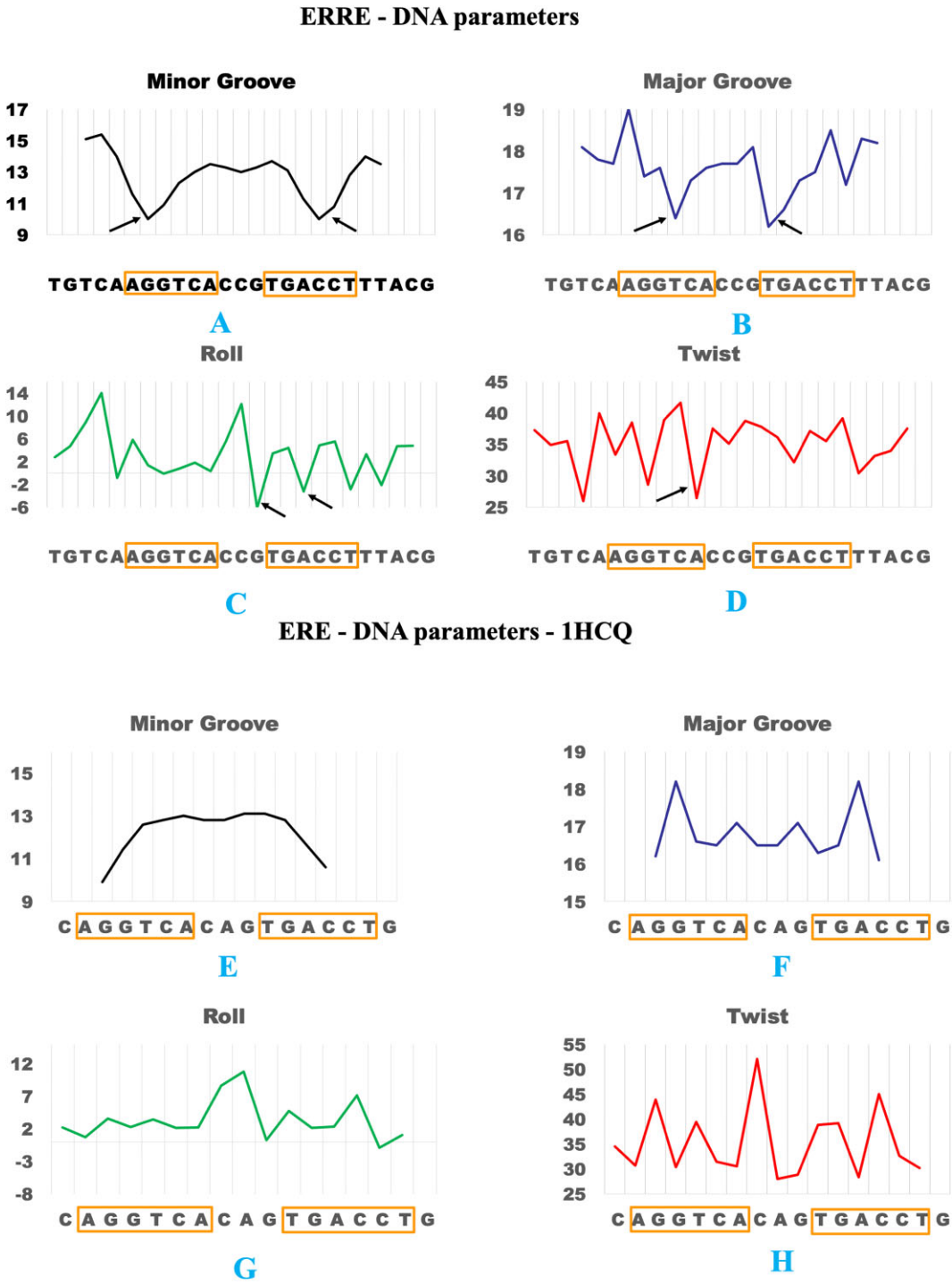


Figure 6. (A–D) DNA geometrical parameters of the embERRE/IR3 DNA bound by ERRα, showing the (**A**) minor groove width, (**B**) major groove width, (**C**) roll, and (**D**) twist parameters. (E–H) DNA geometrical parameters of the ERE-IR3 bound by ERα DBD homodimer (PDB code 1HCQ, (38)), showing the (**E**) minor groove width, (**F**) major groove width, (**G**) roll and (**H**) twist parameters.

the second half-site. An asymmetry can also be seen for the experimental major groove width and roll parameter (Figure 6B and C). In particular, the width of the major groove at position +2 (T/A) in the second half-site adopts a strong minimum, displaced compared to its position in the first half-site, leading to a strong asymmetry. This observation can be rationalized in terms of local base-pair deformation, in particular visible from the alternation of positive and negative roll values along with the spacer and the second half-site sequence. The detailed analysis suggests that Lys100 in H1 can interact with two consecutive G nucleotides (at positions +5 and +6), leading to local DNA deformation. These observations strongly contrast with the experimental DNA shape parameters of the ERE/IR3 as calculated from the structure of the ERα DBD on this cognate sequence (37,38) (Figure 6E–H). In the latter case, the two half-sites are perfectly symmetrical with respect to the central base pair (position 0), both in composition and DNA shape. The formation of these molecular interactions may be triggered by differences in the dimerization interface of the DBD domain, consistent with the observed asymmetric DBD interface. This is not observed for the roll values of ERα binding to ERE/IR3 which are always positive. Similarly, in the structure of AncSR1 bound to ERE/IR3, no major variations of the MGW, roll and twist parameters are observed along the DNA sequence (Supplementary Figure S6A–D).

Altogether, our observations suggest that the binding of the second ERRα DBD subunit to its cognate DNA is favored by a narrow minor groove. However, the binding of the protein leads concomitantly to significant deformation of the DNA shape for adequate positioning of the two DBD subunits. Hence, the observed conformation of the ERRα DBD dimer differs from that of the ERα DBD or the AncSR1 bound to palindromic ERE/IR3 sequences, highlighting the singularity of the molecular mechanisms that lead to the asymmetric homodimerization of the ERRα DBD on DNA.

DISCUSSION

Cooperative homo- or heterodimerization is a mechanism used by NRs to strengthen their binding to their respective REs and also increase their DNA binding site repertoire in the genome. However, the ERRs clearly contrast with this general view, because most of their DNA targets are single extended half-site elements, suggesting at first that the ERRs act as monomers (22). However, this view is challenged by a large set of functional data which proved that the ERRs are obligate homodimers for the binding of coactivators and transcriptional activation (20,24,26,31,32). To resolve this apparent contradiction, we focused on embERRE/IR3 RE that are natural ERRα binding sites found in the human genome and shown to carry important biological functions in the etiology of breast cancer (20). Our previous analysis of the isolated ERRα DBD bound to embERRE/IR3 RE revealed that the ERRα DBD forms a stable dimer (59). Based on this, we hypothesized that the sequence-dependent DNA shape of the ERRα binding sites could be pivotal for the cooperative binding of the ERRα homodimer to DNA. To address this in detail, we now determined the crystal structure of the ERRα DBD dimer on a natural embERRE/IR3 RE and unveiled the molecular determinants for the stabilization of the ERRα DBD homodimer which is asymmetrically organized on DNA.

The ERRα DBD homodimer is globally organized in a similar manner to other steroids NRs in which a strong cooperative homodimerization is observed on their cognate IR3 REs (5,6,37,38,60,61,71). Furthermore, the recognition of the core half-site sequence by each DBD subunit is identical to that which is observed for the ERα DBD and the AncSR1 DBD. In fact, the base-specific interactions between residues of the recognition helix H1 and the AGGTCA half-site nucleotides of the major groove are preserved (Figure 4A), suggesting evolutionary conserved recognition mechanisms. In the case of the oxoSRs, the molecular determinants of DNA recognition evolved through permissive substitutions into a different type of binding, both in terms of protein residues and DNA half-site sequence (AGAACA), which remained identical for the ER and ERR subfamilies (72).

The ERRα DBD homodimer differs from the DBD structures of SR and related receptors, considering that it forms an asymmetric assembly. The asymmetric dimerization region covers the interface formed by the D-box of each subunit and a unique extended dimerization interface formed between the KR-box of DBD2 and the N-terminal part of the D-box of DBD1 (Figure 2E). In fact, the KR box of DBD1 features a 3_{10} helix that interacts with the DNA (Figure 2D), whereas the KR-box of DBD2 forms an extended loop and makes additional interactions with the adjacent DBD1, leading to an increased dimerization surface area and eventually the stabilization of the dimer (Figure 2B and E). This asymmetric dimerization interface in a homodimer complex was unexpected considering the symmetry of the DNA REs and the prevalent symmetry of ER homodimer and other steroid NRs. It highlights the flexibility and adaptability of the ERR DBD architecture. The crucial role of the KR-box in the dimerization process may explain previous biochemical data reporting the role played by Thr124 on the ERR dimerization (25). In fact, mutating this residue, that is located right before the KR-box motif, to alanine, was shown to disrupt the ERR dimer (25). In our structure, weak electrostatic interactions are observed between Thr124 and the neighbouring residues of the KR-box. Mutating Thr124 to alanine may indeed destabilize the overall KR box, and consequently affect the dimerization capacity of ERR.

Our evolutionary analysis suggests that the role of the KR-box in the dimerization process has been overlooked for other SRs. The residues of the KR-box (KRRRK) are highly conserved in all ERR sequences which represents a constrained version of the box found in ERs (KNRRK) and oxoSRs (KF/IRRK) as well as in the reconstituted ancestral receptor AncSR1 (KHRRK) (Figure 2C). In fact, a single amino acid change separates the ERR KR box from other the oxoSRs box (K/RxRRK) motifs (Figure 2C). The ERα DBD bound to its corresponding IR3 REs shows no significant change in the conformation of its KR box, despite a slight asymmetry of the distances involved in the interaction network formed by the KR-box residues (Fig-

ure 5). In contrast, the DBD of the ancestral steroid receptor AncSR1, binds to an AGGTCA IR3 RE in an asymmetric manner. However, since the sequence of AncSR1 was predicted from phylogenetical analysis (16), it cannot be considered as a real receptor isolated from a living organism. Besides, the conformational change of the KR box motif to a loop for the second subunit leads to no additional interactions, a negligible increase of the dimerization interface and consequently no further stabilization of the dimer. Along with these observations is a minor deformation seen for the IR3 DNA fragment bound by the AncSR1 DBD dimer (Supplementary Figure S6A). In contrast, in the case of the ERRα DBD, the DNA shape of the embERRE/IR3 (characterized by physical properties such as minor and major groove widths, twist and roll parameters) suggests that the binding of the protein leads to the deformation of the DNA to adequately position the ERRα DBD dimer. This DNA deformation reinforces the overall interaction network and the stabilization of the dimeric complex. Hence, the DNA sequence is the allosteric driver for the formation of the asymmetric dimerization of the ERRα DBD. The observed asymmetric binding of the ERRα DBD homodimer represents a remarkable case of adaptation in the binding of two, sequence identical, protein subunits to a relatively symmetrical IR3 binding site.

Notably, the experimental values of the DNA geometrical parameters obtained from our crystal structure are fully consistent with the ones we previously predicted (59) based on the DNA sequence (69,70). The analysis demonstrates that the sequence-dependent DNA shape is crucial in the initial recognition process by the receptor, especially a narrow minor groove in the flanking region of the second half-site. In fact, the DNA shape favors the positioning of the dimer on the DNA, even in the case when the second DBD does not make the most appropriate interactions with DNA. This observation is supported by the weak electron density of some protein side chains of DBD2 that make base-specific interactions and by the corresponding higher *B*-factors. But this appears to be largely compensated by the conformational asymmetry of the protein interface, which is a key element in the stabilization of the dimer on DNA.

Our structure sheds further important light on crucial molecular determinants for the binding of the ERRα DBD to two different 5′-extended half-sites. DBD1 binds to a TCA extended half-site, whereas DBD2 binds to a TAA extended half-site, the latter is a less favorable extension in terms of stability and transcriptional activation (24,25). Our structure emphasizes the strong intricate H-bond network formed between the DBD CTE and the minor groove TCA extension, whereas interactions are much fewer for the TAA extension. These observations agree with our EMSA assays and functional data. We further clarified the structural role of the functionally important residue Tyr161, the last residue of the A-box, in the stabilization of the ERR-DNA complex, through interactions with the DNA backbone (Supplementary Figure S4D), rather than with the hydrophobic protein core as suggested by the solution structure (36)

In our work, we considered the special case of the embERRE/IR3 binding site, whereas most of the ERR binding sites encompass a 3′ DNA sequence with no particular consensus for the binding of the second ERR subunit, given that the ERRs function as dimers. However, all ERR binding sites feature a 5′ extended half-site of the type TNAAGGTCA (19,20,22). Our structure therefore highlights structural principles allowing the binding of the ERR homodimer to DNA (Figure 7). Specifically, it suggests that the DNA acts as an allosteric driver for the formation of a functional ERR homodimer. This is consistent with previous data showed that the sequence of the 5′ TNA extension in the ERRE conveys information that is transferred to the bound ERR, altering its conformation and establishing whether the coactivator PGC-1α will form a productive interaction with the ERR (25). Our work provides further remarkable insight into the ERR dimerization mechanisms on DNA that involve the extended asymmetric dimerization interface made of different KR-box conformations. We propose that given the imprinted asymmetric behavior, the resulting ERR homodimer will be in an appropriate conformation for the binding of coregulators, resulting in its stabilization onto DNA and the formation of a complex fully functional for gene regulation processes.

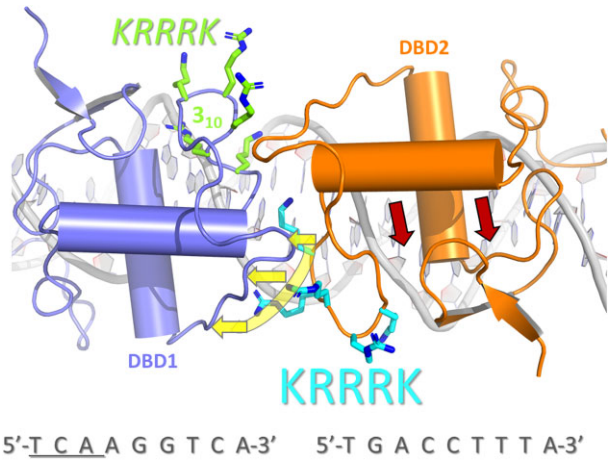


Figure 7. A representation of the ERRα DBD/ERRE complex, summarizing our main structural observations. The dimerization interface of ERRα DBD homodimer is made by interactions between the D box of each subunit. Furthermore, it is reinforced by the presence of an extended interaction area arising from the KR-box region of DBD2 that unfolds as a loop and interacts with the dimerization region of DBD1, as indicated by yellow arrows. The dimerization interface is asymmetric with different KR-box conformations for the two subunits (highlighted by the lemon green and teal colours) and it is required for proper stabilization of the DBD homodimer on DNA (as suggested by the red arrows), suggesting the importance of the DNA, which acts as allosteric driver in accommodating the receptor in a fully functional state suitable for gene expression regulation. The subunit bound to the extended half-site, DBD1, is shown in blue colour and DBD2 is shown in orange colour.

Finally, our results provide insights and new perspectives into the evolution of DNA recognition and binding specificity by the steroid nuclear receptors. The ERRs represent one of the most ancient and extant groups of SRs with a remarkable sequence conservation throughout evolution (11,12). We hypothesize that such a sequence conservation is linked to the preservation of the biological functions of ERRs and may be related to the crucial and most ancestral processes in the regulation of energy metabolism. Indeed, in Drosophila it was shown that ERR is instrumental in

the metabolic program that coordinates the conversion of nutrients into biomass to support larval growth by upregulating the transcription of metabolic genes involved in glycolysis, the pentose phosphate pathway and lactate production (12,73). Similarly, in mammals, when viewed in the light of tumor growth, ERRα directs the transcriptional switch that supports oxidative metabolic pathways, matching what is observed in Drosophila (18). This suggests that the coordination between growth and maturation, a feature often neglected, but central for the metazoan life cycle, may be the ancestral function of ERRs.

This ancestral function may be a critical functional element for the evolutionary conserved DNA recognition mechanisms by the ERRs. Our crystal structure of the ERRα DBD suggests that asymmetry is used for function, allowing a better stabilization of a functional dimer on the cognate DNA sequence in the presence of coactivators. Given the ancestral character of the ERRs, the ancestor of the ER/ERR/SR clade may function in a similar manner. Evolutionary diversification in bilaterians further led to the emergence of the common ancestor of the ERRs and the common ancestor of the bilaterian NR3 group containing the steroid receptors, which may have been an AncSR1-like molecule. The comparative analysis of our structure with AncSR1 DBD suggests that the asymmetry still remained in AncSR1 as a structural remnant of the ancestral asymmetrical and functional feature seen in the ERRα DBD, but it is likely not used for the homodimer stabilization of SRs and hence for its function. Therefore, our crystal structure could serve as a useful model for the understanding of the structural evolution in the DNA binding recognition by the steroid NR family that are of utmost medical and therapeutic importance.

DATA AVAILABILITY

Protein Data Bank accession number: The crystal structure of the ERRα DBD/embERRE/IR3 complex has been deposited to the Protein data Bank (www.pdb.org) under the accession number **PDB ID 8CEF**.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the Integrated Structural Biology platform of IGBMC, in particular P. Poussin-Courmontagne for his assistance in protein crystallization. We thank Dr. G. Travé for comments and critical reading of the manuscript. We acknowledge SOLEIL for provision of synchrotron radiation facilities and we would like to thank William Shepard for assistance in using beamline Proxima 2A.

Authors contributions: I.M.L.B., A.K.M.P. conceptualized the project and designed the experiments; I.H and A.K.M.P crystallized the complex; A.G.M. screened the crystals; A.K.M.P and I.M.L.B. solved and refined the structure; A.K.M.P designed and performed the EMSA experiment; P.V., K.B. and E.E. performed the ITC experiment; G.V.M. and T.B.S. performed phylogenetic analysis; I.M.L.B., A.K.M.P., D.M., B.P.K., A.H., V.L. contributed to the methodologies; I.M.L.B., A.K.M.P., P.V., T.B.S., A.G.M., E.E., G.V.M., V.L. analyzed the results; I.M.L.B., A.K.M.P., E.E., A.H., V.L., D.M., B.P.K. supervised the experiments and the analyses; I.M.L.B., A.K.M.P. wrote the original draft of the paper. All authors contributed to the critical revision of the manuscript and approved the final version.

FUNDING

CNRS, INSERM, University of Strasbourg and l'Alsace contre le Cancer; French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05, Instruct-ERIC]; Interdisciplinary Thematic Institute IMCBio, as part of the ITI 2021–2028 program of the University of Strasbourg, CNRS and Inserm, was supported by IdEx Unistra [ANR-10-IDEX-0002]; SFRI-STRAT'US project [ANR 20-SFRI-0012]; EUR IMCBio [ANR-17-EURE-0023] under the framework of the French Investments for the Future Program; B.P.K. acknowledges support by Fondation pour la Recherche Médicale (FRM); P.V. acknowledges the 'Ligue Contre le Cancer' and the 'Association pour la Recherche sur le Cancer' (ARC). Funding for open access charge: ANR.

Conflict of interest statement. None declared.

REFERENCES

1. Gronemeyer,H., Gustafsson,J.A. and Laudet,V. (2004) Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug. Discov.*, **3**, 950–964.
2. Billas,I.M.L. and Moras,D. (2013) Allosteric controls of nuclear receptor function in the regulation of transcription. *J. Mol. Biol.*, **425**, 2317–2329.
3. Lazar,M.A. (2017) Maturing of the nuclear receptor family. *J. Clin. Invest.*, **127**, 1123–1125.
4. Zhao,L., Zhou,S. and Gustafsson,J.-Å. (2019) Nuclear receptors: recent drug discovery for cancer therapies. *Endocr. Rev.*, **40**, 1207–1249.
5. Helsen,C., Kerkhofs,S., Clinckemalie,L., Spans,L., Laurent,M., Boonen,S., Vanderschueren,D. and Claessens,F. (2011) Structural basis for nuclear hormone receptor DNA binding. *Mol. Cell. Endocrinol.*, **348**, 411–417.
6. Helsen,C. and Claessens,F. (2014) Looking at nuclear receptors from a new angle. *Mol. Cell. Endocrinol.*, **382**, 97–106.
7. Audet-Walsh,E. and Giguere,V. (2015) The multiple universes of estrogen-related receptor [alpha] and [gamma] in metabolic control and related diseases. *Acta Pharmacol. Sin.*, **36**, 51–61.
8. Bardet,P.L., Laudet,V. and Vanacker,J.M. (2006) Studying non-mammalian models? Not a fool's ERRand!. *Trends Endocrinol. Metab.*, **17**, 166–171.
9. Bertrand,S., Belgacem,M.R. and Escriva,H. (2010) Nuclear hormone receptors in chordates. *Mol. Cell. Endocrinol.*, **334**, 67–75.
10. Gomes,I.D.L., Gazo,I., Nabi,D., Besnardeau,L., Hebras,C., McDougall,A. and Dumollard,R. (2019) Bisphenols disrupt differentiation of the pigmented cells during larval brain formation in the ascidian. *Aquat. Toxicol.*, **216**, 105314.
11. Bannister,R., Beresford,N., Granger,D.W., Pounds,N.A., Rand-Weaver,M., White,R., Jobling,S. and Routledge,E.J. (2013) No substantial changes in estrogen receptor and estrogen-related receptor orthologue gene transcription in Marisa cornuarietis exposed to estrogenic chemicals. *Aquat. Toxicol.*, **140-141**, 19–26.
12. Tennessen,J.M., Baker,K.D., Lam,G., Evans,J. and Thummel,C.S. (2011) The drosophila estrogen-related receptor directs a metabolic switch that supports developmental growth. *Cell Metab.*, **13**, 139–148.
13. Khalturin,K., Billas,I.M.L., Chebaro,Y., Reitzel,A.M., Tarrant,A.M., Laudet,V. and Markov,G.V. (2018) NR3E receptors in

cnidarians: a new family of steroid receptor relatives extends the possible mechanisms for ligand binding. *J. Steroid Biochem. Mol. Biol.*, **184**, 11–19.
14. Hochberg,G.K.A., Liu,Y., Marklund,E.G., Metzger,B.P.H., Laganowsky,A. and Thornton,J.W. (2020) A hydrophobic ratchet entrenches molecular complexes. *Nature*, **588**, 503–508.
15. Baker,M.E. (2008) Trichoplax, the simplest known animal, contains an estrogen-related receptor but no estrogen receptor: implications for estrogen receptor evolution. *Biochem. Biophys. Res. Commun.*, **375**, 623–627.
16. McKeown,A.N., Bridgham,J.T., Anderson,D.W., Murphy,M.N., Ortlund,E.A. and Thornton,J.W. (2014) Evolution of DNA specificity in a transcription factor Family produced a new gene regulatory module. *Cell*, **159**, 58–68.
17. Giguère,V., Yang,N., Segui,P. and Evans,R.M. (1988) Identification of a new class of steroid hormone receptors. *Nature*, **331**, 91–94.
18. Tremblay,A.M. and Giguère,V. (2007) The NR3B subgroup: an ovERRview. *Nucl. Recept. Signal*, **5**, e009.
19. Deblois,G., Hall,J.A., Perry,M.C., Laganiere,J., Ghahremani,M., Park,M., Hallett,M. and Giguère,V. (2009) Genome-wide identification of direct target genes implicates estrogen-related receptor alpha as a determinant of breast cancer heterogeneity. *Cancer Res.*, **69**, 6149–6157.
20. Dufour,C.R., Wilson,B.J., Huss,J.M., Kelly,D.P., Alaynick,W.A., Downes,M., Evans,R.M., Blanchette,M. and Giguère,V. (2007) Genome-wide orchestration of cardiac functions by the orphan nuclear receptors ERR[alpha] and [gamma]. *Cell Metab.*, **5**, 345–356.
21. Tremblay,A.M., Dufour,C.R., Ghahremani,M., Reudelhuber,T.L. and Giguere,V. (2010) Physiological genomics identifies estrogen-related receptor alpha as a regulator of renal sodium and potassium homeostasis and the renin-angiotensin pathway. *Mol. Endocrinol.*, **24**, 22–32.
22. Sladek,R., Bader,J.A. and Giguère,V. (1997) The orphan nuclear receptor estrogen-related receptor alpha is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene. *Mol. Cell. Biol.*, **17**, 5400–5409.
23. Xia,H., Dufour,C.R. and Giguère,V. (2019) ERRα as a bridge between transcription and function: role in liver metabolism and disease. *Front. Endocrinol.*, **10**, 206.
24. Vanacker,J.-M., Bonnelye,E., Chopin-Delannoy,S., Delmarre,C., Cavaillès,V. and Laudet,V. (1999) Transcriptional activities of the orphan nuclear receptor errα (Estrogen Receptor-Related Receptor-α). *Mol. Endocrinol.*, **13**, 764–773.
25. Barry,J.B., Laganiere,J. and Giguere,V. (2006) A single nucleotide in an estrogen-related receptor {alpha} site can dictate mode of binding and peroxisome proliferator-activated receptor {gamma} coactivator 1{alpha} activation of target promoters. *Mol. Endocrinol.*, **20**, 302–310.
26. Barry,J.B. and Giguere,V. (2005) Epidermal growth factor-induced signaling in breast cancer cells results in selective target gene activation by orphan nuclear receptor estrogen-related receptor {alpha}. *Cancer Res.*, **65**, 6120–6129.
27. Seacrist,C.D., Kuenze,G., Hoffmann,R.M., Moeller,B.E., Burke,J.E., Meiler,J. and Blind,R.D. (2020) Integrated structural modeling of full-length LRH-1 reveals inter-domain interactions contribute to receptor structure and function. *Structure*, **28**, 830–846.
28. Solomon,I.H., Hager,J.M., Safi,R., McDonnell,D.P., Redinbo,M.R. and Ortlund,E.A. (2005) Crystal structure of the Human LRH-1 DBD-DNA complex reveals ftz-F1 domain positioning is required for receptor activity. *J. Mol. Biol.*, **354**, 1091–1102.
29. Meinke,G. and Sigler,P.B. (1999) DNA-binding mechanism of the monomeric orphan nuclear receptor NGFI-B. *Nat. Struct. Biol.*, **6**, 471–477.
30. Sierk,M.L., Zhao,Q. and Rastinejad,F. (2001) DNA deformability as a recognition feature in the reverb response element. *Biochemistry*, **40**, 12833–12843.
31. Horard,B., Castet,A., Bardet,P.L., Laudet,V., Cavailles,V. and Vanacker,J.M. (2004) Dimerization is required for transactivation by estrogen-receptor-related (ERR) orphan receptors: evidence from amphioxus ERR. *J. Mol. Endocrinol.*, **33**, 493–509.
32. Huppunen,J. and Aarnisalo,P. (2004) Dimerization modulates the activity of the orphan nuclear receptor errγ. *Biochem. Biophys. Res. Commun.*, **314**, 964–970.
33. Zhang,Z. and Teng,C.T. (2000) Estrogen receptor-related receptor alpha1 interacts with coactivator and constitutively activates the Estrogen response elements of the Human lactoferrin gene. *J. Biol. Chem.*, **275**, 20837–20846.
34. Kallen,J., Lattmann,R., Beerli,R., Blechschmidt,A., Blommers,M.J.J., Geiser,M., Ottl,J., Schlaeppi,J.M., Strauss,A. and Fournier,B. (2007) Crystal structure of Human estrogen-related receptor {alpha} in complex with a synthetic inverse agonist reveals its novel molecular mechanism. *J. Biol. Chem.*, **282**, 23231–23239.
35. Greschik,H., Althage,M., Flaig,R., Sato,Y., Chavant,V., Peluso-Iltis,C., Choulier,L., Cronet,P., Rochel,N., Schule,R. *et al.* (2008) Communication between the erralpha homodimer interface and the PGC-1alpha binding surface via the helix 8-9 loop. *J. Biol. Chem.*, **283**, 20220–20230.
36. Gearhart,M.D., Holmbeck,S.M., Evans,R.M., Dyson,H.J. and Wright,P.E. (2003) Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition. *J. Mol. Biol.*, **327**, 819–832.
37. Schwabe,J.W., Chapman,L., Finch,J.T., Rhodes,D. and Neuhaus,D. (1993) DNA recognition by the oestrogen receptor: from solution to the crystal. *Structure*, **1**, 187–204.
38. Schwabe,J.W.R., Chapman,L., Finch,J.T. and Rhodes,D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*, **75**, 567–578.
39. Thornton,J.W., Need,E. and Crews,D. (2003) Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science*, **301**, 1714–1717.
40. Little,T.H., Zhang,Y., Matulis,C.K., Weck,J., Zhang,Z., Ramachandran,A., Mayo,K.E. and Radhakrishnan,I. (2006) Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. *Mol. Endocrinol.*, **20**, 831–843.
41. Diebold,M.-L., Fribourg,S., Koch,M., Metzger,T. and Romier,C. (2011) Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J. Struct. Biol.*, **175**, 178–188.
42. Vonrhein,C., Flensburg,C., Keller,P., Sharff,A., Smart,O., Paciorek,W., Womack,T. and Bricogne,G. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr.*, **67**, 293–302.
43. McCoy,A. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr.*, **63**, 32–41.
44. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.*, **66**, 213–221.
45. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr.*, **66**, 486–501.
46. Bricogne,G., Blanc,E., Brandl,M., Flensburg,C., Keller,P., Paciorek,W., Roversi,P., Sharff,A., Smart,O.S., Vonrhein,C. *et al.* (2017) In: *BUSTER Version 2.10.3*. Global Phasing Ltd., Cambridge, United Kingdom.
47. Afonine,P.V., Moriarty,N.W., Mustyakimov,M., Sobolev,O.V., Terwilliger,T.C., Turk,D., Urzhumtsev,A. and Adams,P.D. (2015) FEM: feature-enhanced map. *Acta. Crystallogr. D Biol. Crystallogr.*, **71**, 646–666.
48. Liebschner,D., Afonine,P.V., Moriarty,N.W., Poon,B.K., Sobolev,O.V., Terwilliger,T.C. and Adams,P.D. (2017) Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr.*, **73**, 148–157.
49. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
50. Piñeiro,Á., Muñoz,E., Sabín,J., Costas,M., Bastos,M., Velázquez-Campoy,A., Garrido,P.F., Dumas,P., Ennifar,E., García-Río,L. *et al.* (2019) AFFINImeter: a software to analyze molecular recognition processes from experimental data. *Anal. Biochem.*, **577**, 117–134.
51. Sievers,F. and Higgins,D.G. (2014) In: Russell,D.J. (ed.) *Multiple Sequence Alignment Methods*. Humana Press, Totowa, NJ, pp. 105–116.

52. Gouy,M., Guindon,S. and Gascuel,O. (2009) SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.*, **27**, 221–224.
53. Guindon,S. and Gascuel,O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, **52**, 696–704.
54. Le,S.Q. and Gascuel,O. (2010) Accounting for solvent accessibility and secondary structure in protein phylogenetics is clearly beneficial. *Syst. Biol.*, **59**, 277–287.
55. Anisimova,M. and Gascuel,O. (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.*, **55**, 539–552.
56. Huelsenbeck,J.P., Nielsen,R. and Bollback,J.P. (2003) Stochastic mapping of morphological characters. *Syst. Biol.*, **52**, 131–158.
57. R-Core-Team (2021) In: *R Foundation for Statistical Computing*. Vienna, Austria.
58. Revell,L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.*, **3**, 217–223.
59. Mohideen-Abdul,K., Tazibt,K., Bourguet,M., Hazemann,I., Lebars,I., Takacs,M., Cianférani,S., Klaholz,B.P., Moras,D. and Billas,I.M.L. (2017) Importance of the sequence-directed DNA shape for specific binding site recognition by the estrogen-related receptor. *Front. Endocrinol.*, **8**, 140.
60. Luisi,B.F., Xu,W.X., Otwinowski,Z., Freedman,L.P., Yamamoto,K.R. and Sigler,P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*, **352**, 497–505.
61. Meijsing,S.H., Pufall,M.A., So,A.Y., Bates,D.L., Chen,L. and Yamamoto,K.R. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science*, **324**, 407–410.
62. Shaffer,P.L., Jivan,A., Dollins,D.E., Claessens,F. and Gewirth,D.T. (2004) Structural basis of androgen receptor binding to selective androgen response elements. *Proc. Natl Acad. Sci. U.S.A.*, **101**, 4758–4763.
63. Roemer,S.C., Donham,D.C., Sherman,L., Pon,V.H., Edwards,D.P. and Churchill,M.E.A. (2006) Structure of the progesterone receptor-deoxyribonucleic acid complex: novel interactions required for binding to half-site response elements. *Mol. Endocrinol.*, **20**, 3042–3052.
64. Hudson,W.H., Youn,C. and Ortlund,E.A. (2014) Crystal structure of the mineralocorticoid receptor DNA binding domain in complex with DNA. *PLoS One*, **9**, e107000.
65. Spolar,R. and Record,M. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
66. Peters,W.B., Edmondson,S.P. and Shriver,J.W. (2004) Thermodynamics of DNA binding and distortion by the hyperthermophile chromatin protein Sac7d. *J. Mol. Biol.*, **343**, 339–360.
67. Oda,M., Furukawa,K., Ogata,K., Sarai,A. and Nakamura,H. (1998) Thermodynamics of specific and non-specific DNA binding by the c-myb DNA-binding domain11Edited by P. E. Wright. *J. Mol. Biol.*, **276**, 571–590.
68. Privalov,P.L., Dragan,A.I. and Crane-Robinson,C. (2010) Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. *Nucleic Acids Res.*, **39**, 2483–2491.
69. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
70. Zhou,T., Yang,L., Lu,Y., Dror,I., Dantas Machado,A.C., Ghane,T., Di Felice,R. and Rohs,R. (2013) DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.*, **41**, W56–W62.
71. Denayer,S., Helsen,C., Thorrez,L., Haelens,A. and Claessens,F. (2010) The rules of DNA recognition by the androgen receptor. *Mol. Endocrinol.*, **24**, 898–913.
72. Ortlund,E.A., Bridgham,J.T., Redinbo,M.R. and Thornton,J.W. (2007) Crystal structure of an ancient protein: evolution by conformational epistasis. *Science*, **317**, 1544–1548.
73. Tennessen,J.M. and Thummel,C.S. (2011) Coordinating growth and maturation — Insights from Drosophila. *Curr. Biol.*, **21**, R750–R757.