# Structure and Function of the DNA-Binding Domain of the Glucocorticoid Receptor and Other Members of the Nuclear Receptor Supergene Family

TORLEIF HÄRD AND JAN-ÅKE GUSTAFSSON*

*Center for Structural Biochemistry and Department of Medical Nutrition, Karolinska Institute, Novum, S-141 57 Huddinge, Sweden*

*Received March 5, 1993*

## Background

The nuclear receptor superfamily includes receptors for steroid hormones, thyroid hormone, retinoic acids, and vitamin $D_3$, as well as receptors for several other, yet unidentified, ligands (orphan receptors). These receptors act as ligand-inducible transcription factors in that they are activated by ligand binding and then, in turn, associate with specific DNA sequences, where they act to regulate gene transcription. The various receptors within the superfamily have similar mechanisms of action. They are also similar in their functional architecture; i.e., the same regions or domains within the receptors are responsible for the various receptor functions, such as ligand binding, interaction with heat shock proteins (in the inactive state), nuclear translocation, DNA binding, and interactions with other proteins within the transcription regulation machinery. Several recent review articles have been devoted to the nuclear receptor superfamily.[1-7]

A few intense years of exciting research have resulted in the identification, characterization, and structure determination of nuclear receptor DNA-binding domains. The present article represents an account of these studies, which have been conducted within the classical areas of biophysics, chemistry, biochemistry, molecular biology, and biology. The discussion is focused on the glucocorticoid receptor DNA-binding domain (GR DBD), which is the domain on which most of our research has been focused.

## The Discovery of a DNA-Binding Domain

The first demonstration of the structural division of the GR into functional domains was performed by biochemical analysis of purified GR. It was shown that various functions, such as DNA and steroid binding, could be recovered in isolated protein fragments from a limited proteolysis of the intact GR molecule.[8,9] Subsequent cloning and sequencing of cDNA encoding glucocorticoid[10-13] and estrogen[14] receptors (ER) from various species, and comparison of these sequences with that of the oncogene v-*erbA*, revealed a highly conserved region, encompassing residues 440–505 of the rat GR, which was abundant in lysine, arginine, and cysteine residues (Figure 1). It was proposed that this region corresponded to the DNA-binding domain.[11,13,14] This notion was further tested in studies of various receptor mutants[15,16] and finally proven in an elegant experiment with a chimeric ER, where the DNA-binding domain had been replaced with that of the GR, which activated the expression of a glucocorticoid-inducible gene.[17]

Parallel studies had resulted in the identification of a variety of glucocorticoid-responsive genes. Gene-transfer studies, utilizing the mouse mammary tumor virus (MMTV) promoter and other promoters, were used to identify short DNA sequences conferring glucocorticoid sensitivity onto the gene in question. These sequences were shown to interact with purified GR[1] and constitute binding sites for nuclear receptors termed hormone response elements (HREs).[1,4,18] The 15 base pair consensus glucocorticoid response element

Torleif Härd was born in Göteborg, Sweden, in 1959. He obtained his M.Sc. in 1983 and Ph.D. in physical chemistry in 1986 from Chalmers University of Technology, Göteborg. Since 1990 he has been Associate Professor at the Center for Structural Biochemistry at the Karolinska Institute, where he is heading a group working in the general area of biophysical chemistry, mainly using NMR methods. The focus of his research is presently directed toward biophysical and structural aspects of steroid hormone receptors and related proteins.

Jan-Åke Gustafsson was born in Stockholm, Sweden, in 1943. He received a Ph.D. in 1968 and an M.D. in 1971 from the Karolinska Institute, Stockholm. Since 1979 he has been Chairman and Professor at the Department of Medical Nutrition, Karolinska Institute. He is also Director of the Center for Biotechnology, Center for Nutrition and Toxicology, and Center for Structural Biochemistry at Novum, Huddinge Hospital (South Campus of the Karolinska Institute). His major scientific achievements center around the glucocorticoid receptor, where he led early studies in purification of this protein to homogeneity and characterization of the three-domain structure of the receptor. His current research interests include the mechanism of action of nuclear receptors as well as mechanisms of regulation of cytochrome P-450.

(1) Yamamoto, K. R. *Annu. Rev. Genet.* **1985**, 209–252 and references cited therein.
(2) Miesfeld, R. L. *Crit. Rev. Biochem. Mol. Biol.* **1989**, *24*, 101–117.
(3) Parker, M. G., Ed. *Nuclear Hormone Receptors*; Academic Press: London, 1991.
(4) Beato, M. *Cell* **1989**, *56*, 335–344 and references cited therein.
(5) Evans, R. M. *Science* **1988**, *240*, 889–895.
(6) Gustafsson, J.-A.; Carlstedt-Duke, J.; Poellinger, L.; Okret, S.; Wikström, A.-C. Brönnegård, M.; Gillner, M.; Dong, Y.; Fuxe, K.; Cintra, A.; Härfstrand, A.; Agnati, L. *Endocr. Rev.* **1987**, *8*, 185–234.
(7) Gronemeyer, H. *Annu. Rev. Genet.* **1991**, *25*, 89–123.
(8) Wrange, Ö.; Gustafsson, J.-A. *J. Biol. Chem.* **1978**, *253*, 856–865.
(9) Carlstedt-Duke, J.; Okret, S.; Wrange, Ö.; Gustafsson, J.-A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4260–4264.
(10) Miesfeld, R.; Okret, S.; Wikström, A.-C.; Wrange, Ö.; Gustafsson, J.-A.; Yamamoto, K. R. *Nature* **1984**, *312*, 779–781.
(11) Weinberger, C.; Hollenberg, S. M.; Rosenfeld, M. G.; Evans, R. M. *Nature* **1985**, *318*, 670–672.
(12) Hollenberg, S. M.; Weinberger, C.; Ong, E. S.; Cerelli, G.; Oro, A.; Lebo, R.; Thompson, E. B.; Rosenfeld, M. G.; Evans, R. M. *Nature* **1985**, *318*, 635–641.
(13) Miesfeld, R.; Rusconi, S.; Godowski, P. J.; Maler, B.; Okret, S.; Wikström, A.-C.; Gustafsson, J.-A.; Yamamoto, K. R. *Cell* **1986**, *46*, 389–399.
(14) Green, S.; Walter, P.; Kumar, V.; Krust, A.; Bornert, J.-M.; Argos, P.; Chambon, P. *Nature* **1986**, *320*, 134–139.
(15) Kumar, V.; Green, S.; Staub, A.; Chambon, P. *EMBO J.* **1986**, *5*, 2231–2236.
(16) Hollenberg, S. M.; Giguere, V.; Segui, P.; Evans, R. M. *Cell* **1987**, *49*, 39–46.
(17) Green, S.; Chambon, P. *Nature* **1987**, *325*, 75–78.
(18) Lucas, P. C.; Granner, D. K. *Annu. Rev. Biochem.* **1992**, *61*, 1131–1173.

0001-4842/93/0126-0644$04.00/0 

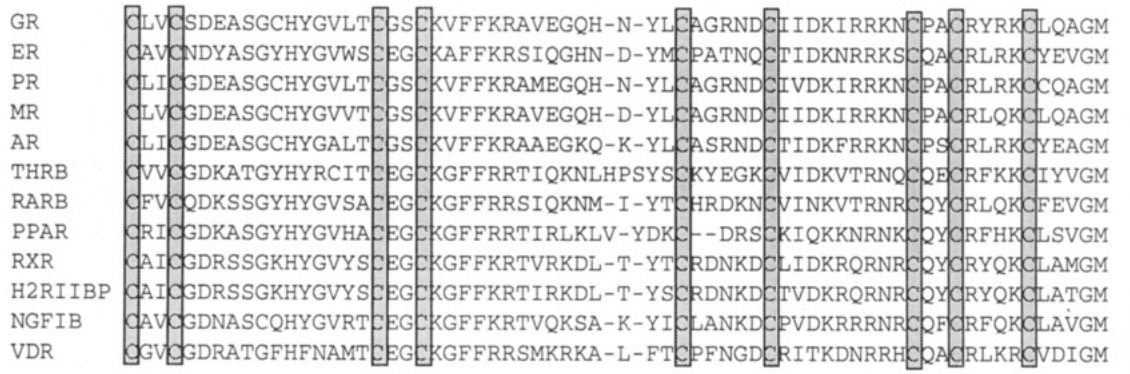


**Figure 1.** Aligned sequences of the core conserved regions of the DNA-binding domains of some of the members of the nuclear receptor supergene family with the nine conserved cysteine residues enclosed in shaded boxes. Alignments and receptor abbreviations have been adopted from Laudet *et al.*[69] Abbreviations used: GR, human (or rat) glucocorticoid receptor; ER, human estrogen receptor; PR, human progesterone receptor; MR, human mineralocorticoid receptor; AR, human androgen receptor; THRB, human thyroid hormone receptor $\beta$; RARB, human retinoic acid receptor $\beta$; PPAR, mouse peroxisome proliferator-activated receptor (orphan receptor); RXR, human retinoid X receptor; H2RIIBP, mouse H-2 region II binding protein (orphan receptor); NGFIB, (orphan receptor in rat); VDR, human vitamin D receptor.

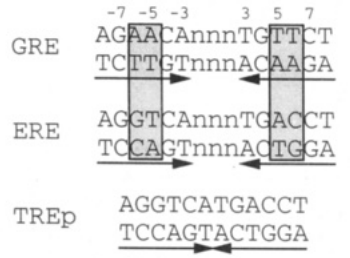


**Figure 2.** Sequences of palindromic hormone response elements of the GR (GRE), ER (ERE), and TR (TREp). The GRE base pair numbering is adopted from Luisi *et al.*[41] Arrows indicate inverted repeats of half-sites. Shaded boxes indicate base pairs that differ between the GRE and the ERE. It is important to note that the shown sequences are "idealized" binding sites and that not all naturally occurring response elements conform to these sequences.

(GRE) (Figure 2) contains two partially palindromic "half-site" hexamer sequences with three intervening nucleotides. The consensus sequence for the estrogen response element (ERE) differs from the GRE in only two base pairs within each hexanucleotide[19] (Figure 2). The three base pair intervening sequence constitutes a strict requirement for these two response elements.[20,21] The thyroid response elements (TREs), on the other hand, consist of hexanucleotide sequences that are similar to those in the ERE, but the consensus palindromic TRE does not contain any intervening nucleotides between the hexamer half-sites.[22]

## Characterization of the DNA-Binding Domain

The key event for the functional characterization and later structure determination of the GR DBD was the expression of large amounts of recombinant protein fragments in bacteria.[23,24] Freedman *et al.* initially studied a 150-residue GR fragment including the DBD, which was found to specifically recognize and protect the same GRE sequences as the full length GR.[23] The purified protein fragment was shown to reversibly ligate two zinc ions. These could be removed by dialysis at low pH in the presence of chelating agents to yield an apo-protein that failed to bind DNA. However, DNA binding could be reconstituted by preincubating the protein with $Zn^{2+}$ or $Cd^{2+}$ ions. (The affinity for $Cd^{2+}$ is about 100 times higher than for $Zn^{2+}$.) The metal ion coordination was investigated using EXAFS spectroscopy, and comparison of spectra with those of a reference model compound (Zinc(II) tetrathiophenolate) enabled Freedman *et al.* to deduce a tetrahedral sulfur coordination for both zinc ions. Thus, eight of 10 cysteine residues in the studied protein fragment (and within the DBD) are involved in zinc coordination. The remaining two cysteines could subsequently also be accounted for by the detection of two reactive sulfhydryl groups. Although convincing, the results of Freedman *et al.* were not completely unambiguous in terms of cysteine sulfur coordination by zinc ions, because all members of the receptor superfamily contain *nine* completely conserved cysteine residues. Logically, one of these should be dispensable for correct protein folding, given the EXAFS data. This issue was resolved using site-directed mutagenesis on the intact GR, which unambiguously showed that Cys500 can be replaced by a Ser or an Ala residue without loss of receptor activity[25] and DNA binding by the DBD.[26] The zinc coordination scheme shown in Figure 3 was later also confirmed in the structure determination of the GR DBD in complex with a GRE.

Independent studies in our (J.-Å.G.) laboratory had also shown that a 108-residue recombinant protein fragment containing the GR DBD binds two zinc ions that are essential for proper folding. However, the DNA-binding affinities of recombinant DNA fragments to various GRE sequences were found to be significantly lower than those of the intact receptor, and the protection patterns obtained in DNase I footprinting assays were also different from those of intact GR.[24] A suspicion that two recombinant DBD fragments bound with a weak cooperativity to the two GRE half-sites,

(19) Klock, G.; Strähle, U.; Schütz, G. *Nature* **1987**, *329*, 734–736.
(20) Dahlman-Wright, K.; Siltala-Roos, H.; Carlstedt-Duke, J.; Gustafsson, J.-A. *J. Biol. Chem.* **1990**, *265*, 14030–14035.
(21) Nordeen, S. K.; Suh, B. J.; Kühnel, B.; Hutchison, C. A. *Mol. Endocrinol.* **1990**, *4*, 1866–1873.
(22) Glass, C. K.; Holloway, J. M.; Devary, O. V.; Rosenfeld, M. G. *Cell* **1988**, *54*, 313–323.
(23) Freedman, L. P.; Luisi, B. F.; Korszun, Z. R.; Basavappa, R.; Sigler, P. B.; Yamamoto, K. R. *Nature* **1988**, *334*, 543–546.
(24) Dahlman, K.; Strömstedt, P.-E.; Rae, C.; Jörnvall, H.; Flock, J.-I.; Carlstedt-Duke, J.; Gustafsson, J.-A. *J. Biol. Chem.* **1989**, *264*, 804–809.
(25) Severne, Y.; Wieland, S.; Schaffner, W.; Rusconi, S. *EMBO J.* **1988**, *7*, 2503–2508.
(26) Zilliacus, J.; Dahlman-Wright, K.; Carlstedt-Duke, J.; Gustafsson, J.-A. *J. Steroid Biochem. Mol. Biol.* **1992**, *42*, 131–139.


**Figure 3.** Sequence of the GR DBD fragment for which the structure has been determined showing the zinc ion coordination by eight conserved cysteine residues. The residue numbering used here and in the text refers to the rat GR DBD, but the amino acid sequence within the DBD is identical in the human GR. Helical segments are enclosed in boxes. Boxed amino acids correspond to the "P-box", i.e., amino acids that are essential for the specific binding to a GRE versus an ERE, the residues covered with stripes are those that make specific contacts to GRE bases in a sequence specific DBD–DNA complex.[41] Residue in circles correspond to the "D-box", which was initially identified as crucial for cooperative DNA binding, and shaded residues are those found to be involved in protein–protein interactions in the dimeric DBD–DNA complex.

whereas the intact GR had a higher affinity due to dimeric or very strongly cooperative binding, was substantiated by the results from gel retardation assays, methylation protection experiments,[27] and equilibrium titrations.[28] In these studies it was found that a recombinant DBD molecule initially associates with the 5′TGTTCT3′ half-site of the GRE, and that a second DBD molecule binds cooperatively to the adjacent GRE half-site. The cooperative binding of GR DBD to the two GRE half-sites was also found to be strongly dependent on the intervening sequence of three nucleotides and relative orientation of the hexamer half-sites, but independent of the integrity of the DNA backbone.[20] A segment of five amino acid residues within the DBD that might be essential for dimerization had been identified by Umesuno and Evans and named the "D-box" (Figure 3).[29] Studies on recombinant DBD fragments showed that these residues indeed are required for cooperative binding of GR DBD to GRE.[30]

However, the cooperativity observed in the binding of recombinant GR DBD fragments to DNA might be artifactual due to the isolation of the DBD from the rest of the GR, which can form a dimer also in the free state[31] due to additional protein–protein interactions outside the DBD.[32] It is still not completely clear if the intact GR exists in the monomeric state and binds cooperatively to DNA, or if it exists and binds as a dimer, although recent investigations into this matter seem to suggest that the GR binds to GRE as a preformed dimer.[31]

Site-directed mutagenesis on GR and ER conducted in several laboratories resulted in the identification of particular amino acid residues in the DBD that are crucial for the discrimination between GREs and EREs by the receptors. First, Green *et al.* used chimeric ER constructs to localize the sequence-specific DNA-binding function to the N-terminal (or CI) region of the DBD containing the first zinc-coordinating subdomain.[33] The authors also (correctly) postulated that the C-terminal (CII) region, containing the second zinc-coordinating subdomain, might be involved in dimerization interactions in the bound state. Second, three groups subsequently and independently identified amino acids that are crucial for the sequence specificity. The residues Gly458, Ser459, and Val462 (the "P-box"[29]) in the GR DBD can be replaced with the corresponding residues in the ER DBD (Glu, Gly, and Ala) to convert the receptor specificity for a GRE to specificity for an ERE[29,34] (Figure 3). The corresponding effect in the context of the ER was also demonstrated.[35] These observations in combination with a secondary structure prediction led Berg to (correctly) suggest that the region Lys461–Glu469 could form an amphipathic $\alpha$-helix, which, perhaps initiated and/or stabilized by Cys457–Cys460, might form a "recognition helix", analogous to those found in other DNA-binding proteins.[36]

Thus, two short years of studies had resulted in a detailed characterization of the GR DBD at a biochemical level, and many of the important functions and interactions had been mapped onto the primary structure (Figure 3). The stage was now (in the spring of 1990) set for a complete determination of the three-dimensional structure of GR DBD.

## The Three-Dimensional Structure of the DNA-Binding Domain

The structures of the GR[37,38] and ER[39] DBDs, determined in solution using nuclear magnetic resonance (NMR), were both solved and published in 1990. The two proteins turned out to be very similar, as expected from the high degree of sequence identity and similar biochemical and DNA-binding properties. In both cases it was also possible to use the existing genetic, biochemical, and biophysical data to predict models for the dimeric binding to GREs and EREs, respectively. In the case of GR DBD the model was confirmed by an NMR study of a GR DBD in complex with a DNA fragment containing a GRE half-site.[40] These initial papers were soon to be followed by structure determinations of two different complexes between DNA and two GR DBD molecules using X-ray crystallography.[41]

(27) Tsai, S. Y.; Carlstedt-Duke, J.; Weigel, N. L.; Dahlman, K.; Gustafsson, J.-Å.; Tsai, M.-J.; O'Malley, B. W. *Cell* **1988**, *55*, 361–369.
(28) Härd, T.; Dahlman, K.; Carlstedt-Duke, J.; Gustafsson, J.-A.; Rigler, R. *Biochemistry* **1990**, *29*, 5358–5364.
(29) Umesuno, K.; Evans, R. M. *Cell* **1989**, *57*, 1139–1146.
(30) Dahlman-Wright, K.; Wright, A.; Gustafsson, J.-A.; Carlstedt-Duke, J. *J. Biol. Chem.* **1991**, *266*, 3107–3112.
(31) Drouin, J.; Sun, Y. L.; Tremblay, S.; Lavender, P.; Schmidt, T. J.; de Léan, A.; Nemer, M. *Mol. Endocrinol.* **1992**, *6*, 1299–1309.
(32) Dahlman-Wright, K.; Wright, A. P. H.; Gustafsson, J.-A. *Biochemistry* **1992**, *31*, 9040–9044.
(33) Green, S.; Kumar, V.; Theulatz, I.; Wahli, W.; Chambon, P. *EMBO J.* **1988**, *7*, 3037–3044.
(34) Danielsen, M.; Hinck, L.; Ringold, G. M. *Cell* **1989**, *57*, 1131–1138.
(35) Mader, S.; Kumar, V.; de Verneuil, H.; Chambon, P. *Nature* **1989**, *338*, 271–274.
(36) Berg, J. M. *Cell* **1989**, *57*, 1065–1068.
(37) Härd, T.; Kellenbach, E.; Boelens, R.; Maler, B. A.; Dahlman, K.; Freedman, L. P.; Carlstedt-Duke, J.; Yamamoto, K. R.; Gustafsson, J.-A.; Kaptein, R. *Science* **1990**, *249*, 157–160.
(38) Härd, T.; Kellenbach, E.; Boelens, R.; Kaptein, R.; Dahlman, K.; Carlstedt-Duke, J.; Freedman, L. P.; Maler, B. A.; Hyde, E. I.; Gustafsson, J.-A.; Yamamoto, K. R. *Biochemistry* **1990**, *29*, 9015–9023.
(39) Schwabe, J. W. R.; Neuhaus, D.; Rhodes, D. *Nature* **1990**, *348*, 458–461.
(40) Remerowski, M. L.; Kellenbach, E.; Boelens, R.; van der Marel, G. A.; van Boom, J. H.; Maler, B. A.; Yamamoto, K. R.; Kaptein, R. *Biochemistry* **1991**, *30*, 11620–11624.
(41) Luisi, B. F.; Xu, W. X.; Otwinowski, Z. Freedman, L. P.; Yamamoto, K. R.; Sigler, P. B. *Nature* **1991**, *352*, 497–505.

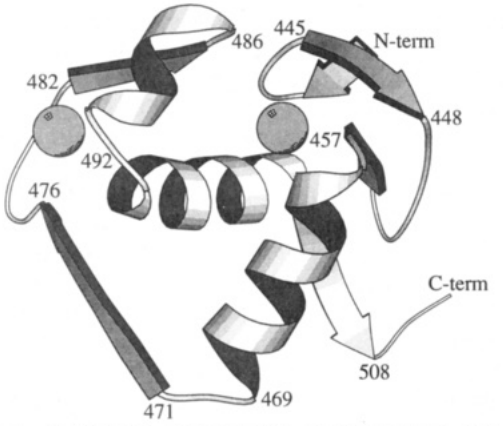


**Figure 4.** Schematic (Molscript[70]) drawing of the GR DBD structure in solution determined using NMR.[46] Helices and regions of extended peptide conformation (arrows) are indicated. Solid spheres represent bound zinc ions.

The GR DBD (Figure 4) was found to consist of two "modules", approximately corresponding to the two exons coding for the DBD, with the general composition [zinc subdomain–helix–$\beta$-strand], which act together to adopt an almost globular folded protein. Two helices in these subdomains are oriented approximately perpendicular to each other, and the (conserved) hydrophobic side chains of these and the following extended regions form a hydrophobic protein core. This core is nearly spherical with a diameter of about 20 Å. The two zinc ion coordination sites are located outside the core at a distance of about 13 Å from each other. The first zinc-coordinating subdomain is folded onto the core and makes several contacts with the two helices and the C-terminal extended region. This domain does not adopt much regular secondary structure, with the exception of a very short stretch of antiparallel $\beta$-sheet involving residues Leu439–Leu441 and Leu455–Cys457. The second zinc-coordinating subdomain, which protrudes out from the protein core with a zinc ion at the base, forms a loop involving residues Cys476–Cys482, followed by a $\beta$-strand encompassing Asp481–Asp485 and a short $\alpha$-helix formed by Lys486–Asn491. The two zinc-coordinating subdomains are stabilized by several "interdomain" contacts involving residues Cys443, Asp445, and Ile484–Lys486.

It is important to note that the two zinc-coordinating subdomains of the steroid receptor DBDs are structurally (and functionally) distinct from zinc-coordinating subdomains found in other transcription factors and DNA-binding proteins, e.g., the TFIIIA-type "zinc fingers" found in several proteins in a variety of species,[42] the *gag* protein p55 from human immunodeficiency virus,[43] and the DNA-binding domain of the yeast transcription factor GAL4.[44] All these zinc-coordinating "motifs" differ from each other and from the steroid hormone receptor DBDs in the nature of zinc coordination as well as in secondary structure composition and overall folding.

Comparison of the initial NMR structure of the GR DBD and the X-ray structure of the same DBD fragment in complex with DNA revealed some apparent differences within the second zinc-coordinating subdomain, and it was discussed whether or not these differences might be due to conformational flexibility within this region in the uncomplexed protein and/or a DNA-induced structural change in the DBD upon binding to DNA. This unresolved issue first prompted us to investigate the extent of rapid internal motions along the protein backbone using $^{15}$N NMR relaxation measurements.[45] However, the extent of rapid motions was found to be uniform and limited along the entire DBD backbone, as can be expected for a folded protein with a well-defined structure. A subsequent refinement of the NMR structure of the free protein carried out in our laboratory revealed that the actual differences between free and DNA-bound molecules are small,[46] although there are some minor structural changes in the DBD upon binding to DNA, which are discussed in more detail below.

Structure determinations have also been carried out on the uncomplexed forms of the DBDs from the human retinoic acid receptor $\beta$ (RAR-$\beta$)[47,48] and the human retinoid X receptor $\alpha$ (RXR-$\alpha$),[49] using NMR. These proteins and the ER DBD[39] have secondary structure compositions and globular folds that are similar to those of the GR DBD, with possible exceptions for the folding of the second zinc-coordinating subdomain. Rapid amide proton exchange are observed in this region of the ER DBD, indicating a clear difference in conformational flexibility compared to the GR DBD (D. Rhodes, personal communication). Another notable difference between the four DBDs for which structures have been determined is that the C-terminal residues of the hRXR-$\alpha$ DBD form an additional $\alpha$-helix which is not observed in the other DBDs.[49]

## Structure of the DBD–GRE Complex

A key paper on the GR DBD is the account of the crystal structures of two DBD–DNA complexes by Luisi *et al.*[41] The first of these is a structure of a dimeric DBD bound to a pseudosymmetric GRE fragment with two palindromic half-sites with the correct three base pair spacing. Unfortunately, this structure could only be refined to a resolution of about 4 Å. The crystallization of a complex containing two DBD molecules bound to a completely symmetric DNA fragment containing two GRE half-sites with a four base pair spacing was more successful in that the structure (initially) could be refined to 2.9 Å. However, the two DBD molecules in this complex did not both bind to their specific GRE half-sites. Instead they preferred a nonsymmetric configuration, in which there is one DBD bound in the correct specific site and a second DBD bound cooperatively to the first DBD in a site, which does not seem to be sequence specific, but which has the correct three base pair displacement to a specific

(42) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809–817 and references cited therein.
(43) Summers, M. F.; South, T. L.; Kim, B.; Hare, D. *Biochemistry* **1990**, *29*, 329–340 and references cited therein.
(44) Marmorstein, R.; Carey, M.; Ptashne, M.; Harrison, S. C. *Nature* **1992**, *356*, 408–414 and references cited therein.
(45) Berglund, H.; Kovács, H.; Dahlman-Wright, K.; Gustafsson, J.-A.; Härd, T. *Biochemistry* **1992**, *31*, 12001–12011.
(46) Baumann, H.; Paulsen, K.; Kovács, H.; Berglund, H.; Wright, A. P. H.; Gustafsson, J.-A.; Härd, T. *Biochemistry*, accepted for publication.
(47) Katahira, M.; Knegtel, R. M. A.; Boelens, R.; Eib, D.; Schilthuis, J. G.; van der Saag, P. T.; Kaptein, R. *Biochemistry* **1992**, *31*, 6474–6480.
(48) Knegtel, R. M. A.; Katahira, M.; Schilthuis, J. G.; Bonvin, A. M. J. J.; Boelens, R.; Eib, D.; van der Saag, P. T.; Kaptein, R. *J. Biomol. NMR* **1993**, *3*, 1–17.
(49) Lee, M. S.; Kliewer, S. A.; Provencal, J.; Wright, P. E.; Evans, R. M. *Science* **1993**, *260*, 1117–1121.

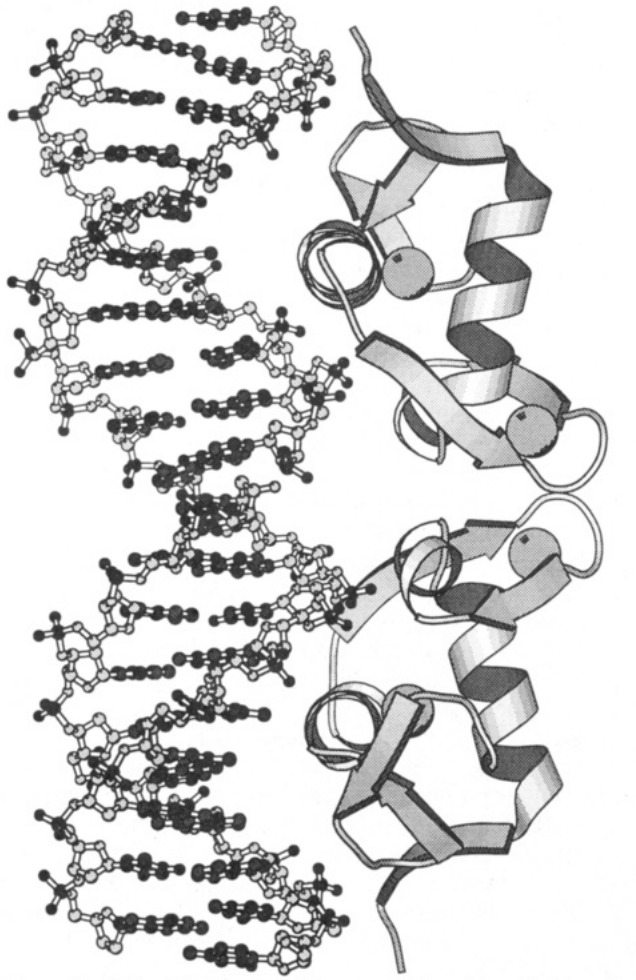

**Figure 5.** Molscript drawing of the crystal structure of a dimeric DBD–DNA complex[41] (Protein Data Bank entry PDB1GRE, 2.9-Å resolution). The upper DBD molecule is specifically bound to a GRE half-site. The view is adjusted so that the recognition helices, encompassing residues Gly458–Glu469, are viewed from the ends. The lower DBD molecule is bound cooperatively in a more nonspecific manner (see the text for details).

half-site. In other words, the penalty paid by binding nonspecifically was preferred over the penalty paid by binding noncooperatively, at the conditions under which the complex could be crystallized. This structure is perhaps not completely unexpected, in view of what is now known about the relative strengths of cooperativity versus specific/nonspecific binding.[28,50] Still, the complex refined to 2.9 Å now instead provides detailed information, not only about sequence specific DBD–GRE interactions but also on protein–protein interactions and nonspecific DBD–DNA interactions. These interactions are described in detail by Luisi *et al.*[41] and are therefore only reviewed briefly here.

In the DBD–GRE complex two DBD molecules bind with the N-terminal helices inserted into two successive major grooves of the DNA (Figure 5), in a manner that is consistent with the models mentioned above and that accounts for results from methylation protection experiments[27,51] and "missing base" contact and ethylation interference experiments.[52] There are no DNA–protein interactions involving DNA bases within the three base pair intervening sequence between the two DNA half-sites. The DNA major groove is widened by about 2 Å to accommodate the recognition helix in the specific site, but there are no bends or other major structural distortions induced in the DNA upon DBD binding.

It is interesting to compare the refined NMR structure of the free DBD[46] with the crystal structure of the DBD in complex with DNA. The structure of the protein core, the relative positions of helices, etc. remain unchanged upon binding to DNA whereas minor structural changes occur in a few loop regions that are involved in DNA–protein and protein–protein interactions in the complex. Most notable is the change of the orientation of the loop containing the D-box, i.e., residues Ala477–Asp481, which has been "reoriented" upon formation of the dimeric DBD–DNA complex, most likely as a result of protein–protein interactions involving residues in this region. Minor structural changes also occur within the segment at the "tip" of the first zinc-coordinating subdomain and the $\beta$-strand following the first (recognition) helix, which both contain residues that interact with DNA phosphates. The network of hydrogen bonds supporting various side chains to form appropriate contacts with DNA phosphates, as discussed by Luisi *et al.*,[41] seems to be absent in the uncomplexed protein.

## DNA–Protein Interactions

Three amino acid side chains in the DBD molecule bound in the specific GRE half-site make specific contacts with DNA bases. Most importantly, the methyl groups of Val462 are in direct van der Waals contact with the methyl of T(+5) (see Figure 2 for GRE numbering), thereby contributing to the preferential binding of GR DBD to GREs compared to EREs. The other contacts are made by Arg466 and Lys461, which are conserved throughout the receptor superfamily, to two G(+4) and G(−7), respectively, which are conserved among the known response elements.[53] Lys461 also provides two hydrogen bonds, mediated by a bound water molecule, to G(−7) and T(+6). These specific interactions are backed up by an array of what might be less specific interactions with DNA phosphates, including side chain as well as main chain groups in His451, Tyr452, Tyr474, Arg489, and Lys490. Not all of these latter interactions are present in the "nonspecific" complex formed by the second, cooperatively bound, DBD molecule, suggesting that some of these interactions might instead be of a more specific nature, as discussed by Luisi *et al.*[41]

The identified specific contacts account for several of the conserved DNA bases within the different classes of hormone response elements. The substitution of three residues in the ER with those of the GR (residues Glu, Gly, and Ala of the ER are replaced by Gly, Ser, and Val of the GR at positions 458, 459, and 462, respectively) respecifies the binding site from an ERE to a GRE, as mentioned above. This effect is not very easily comprehended as only one of the three residues (Val462) is involved in direct interactions with DNA. Considerable insight into the origin of the ERE/GRE discrimination has been provided by extensive mu-

(50) Lundbäck, T.; Cairns, C.; Gustafsson, J.-A.; Carlstedt-Duke, J.; Härd, T. *Biochemistry* **1993**, *32*, 5074–5082.

(51) Scheidereit, C.; Westphal, H. M.; Carlsson, C.; Bosshard, H.; Beato, M. *DNA* **1986**, *5*, 383–391.

(52) Cairns, C.; Gustafsson, J.-A.; Carlstedt-Duke, J. *Mol. Endocrinol.* **1991**, *5*, 598–604.

(53) Forman, B. M.; Samuels, H. H. *Mol. Endocrinol.* **1990**, *4*, 1293–1301.

tagenesis analyses[54,55] and a recent crystal structure of a dimeric ER DBD–ERE complex, determined by Schwabe *et al.*[56]

Zilliacus *et al.* have made a detailed analysis of a series of GR DBD mutants, containing all possible combinations of ERE and GRE specific residues at positions 458, 459, and 462 of the GR, with regard to their affinity for various GRE and ERE binding sites.[54,55] Statistical analysis of the results revealed that several strategies appear to be utilized in accomplishing discrimination. For instance, Val462 in the GR and Glu439 (the amino acid residue of the ER corresponding to Gly439 in the GR) seem to form both positive contacts with specific base pairs in the cognate binding site and negative contacts in the noncognate binding site. Ser459, on the other hand, has a negative effect in all sites, but the negative effect is stronger for the ERE than for the GRE. Zilliacus *et al.* also noted that combinations of amino acid residues appear to act synergistically, most often with the effect of an attenuated binding affinity for noncognate sites.

The sequence-discriminating role of the Glu residue replacing the GR Gly439 is very clear in the ER DBD–ERE complex,[56] where this Glu is found to make both direct and water-mediated hydrogen bonds with the adenine and cytosine in the ERE that replaces the two thymines of the GRE. On the other hand, it seems that this substitution effect is not the only cause of the sequence discrimination, because some of the other DNA base-interacting side chains have different arrangements in the ER DBD–ERE complex compared to the GR DBD–GRE complex and a more intricate network of ordered water molecules is also found in the former complex. For instance, a conserved lysine (Lys465 of the GR), which makes no specific interactions in the GR DBD–GRE complex, is used in the recognition of the ERE, as discussed by Schwabe *et al.*[56] It is also notable that the DNA conformations seem to differ in the two crystal structures.

Thus, extensive structural and mutagenesis data have given a detailed structural and functional picture of the origin of sequence specificity. However, we believe that these studies also need to be complemented by thermodynamic analyses in order to allow conclusions about the relative weight of observed hydrogen bonds and the relative importance of entropic effects to affinity and specificity (Lundbäck et al., unpublished results).

## Protein–Protein Interactions

Recombinant GR DBD protein fragments bind cooperatively to DNA[27,28,41] due to interactions involving residues on one surface of the DBD: Ala477, Arg479, and Asp481 within the D-box, and Ile483 and Ile487, as shown in the crystallographic analysis by Luisi *et al.*[41] The dimeric ER DBD–ERE complex seems to involve somewhat fewer protein–protein interactions than the GR DBD–GRE complex. This crystal also contains a second complex is which the second zinc-coordinating domain seems partially unfolded,[56] an observation that might be related to the lower stability of this region in the ER DBD compared to the GR DBD, mentioned above.

The homodimerization of the GR and ER DBDs seems to be a property of the steroid hormone receptors, for which the classical response elements are inverted repeats of two half-sites with the strict spatial requirement of three intervening base pairs. This is in contrast to, e.g., the thyroid hormone receptor, retinoic acid, and vitamin D receptor, which bind to a variety of response elements including direct repeats to which they bind as heterodimers with the retinoid X receptors, RXR.[57] The possibility that the D-box within the GR DBD could act as a constitutive dimerization motif that might function in the context of other receptors was recently demonstrated.[58] It was, for instance, found that the substitution of the 5-residue segment of the TR DBD with the corresponding D-box of the GR DBD resulted in cooperative binding of recombinant TR DBD fragments to a modified TRE with a three base pair intervening sequence between the half-sites. The mechanism for "head-to-tail" heterodimerization and the structural role of the spacing between direct repeat half-sites in the discrimination of binding partners in heterodimers involving several non-steroid receptors in complexes with RXR still remain to be elucidated, as discussed below.

It is interesting to note that the GR DBD itself contains some activity with regard to regulation of transcription,[16] indicating that it might be involved in protein–protein interactions other than the dimerization interactions, although it should be noted that transactivation primarily seems to involve regions in other parts of the receptor.[59,60] Random mutagenesis of GR followed by screening for defective receptors in yeast allowed Schena *et al.*[61] to identify several point mutations within the GR DBD that do not affect DNA binding, but still affect activation of transcription. These mutations involve residues within the short helix in the second zinc-coordinating subdomain (Arg488, Arg489, and Asn491) and nearby residues in the following helix (Pro493 and Ala494). Thus, this region might be involved in interactions within the GR, or it might interact directly with other proteins of the transcription activation machinery. Furthermore, Luisi *et al.* noted that the nearby hydrophobic side chains of Ile483 and Ile487 are exposed to the solvent, which is unusual in proteins, implying that these residues also might be involved in contacts with other domains of the receptor or other proteins.[41]

## Issues That Remain To Be Resolved

With regard to the GR DBD, most of its various functions are now fairly well understood at a molecular level, although it would be desirable to obtain more detailed information about the nature of the interactions between the DBD and other parts of the receptor.[62]

(54) Zilliacus, J.; Dahlman-Wright, K.; Wright, A.; Gustafsson, J.-A.; Carlstedt-Duke, J. *J. Biol. Chem.* **1991**, *266*, 3101–3106.
(55) Zilliacus, J.; Wright, A. P. H.; Norinder, U.; Gustafsson, J.-A.; Carlstedt-Duke, J. *J. Biol. Chem.* **1992**, *267*, 24941–24947.
(56) Schwabe, J. W. R.; Chapman, L.; Finch, J. T.; Rhodes, D. *Cell*, in press.
(57) Kliewer, S. A.; Umesuno, K.; Mangelsdorf, D. J.; Evans, R. M. *Nature* **1992**, *355*, 446–449.
(58) Dahlman-Wright, K.; Grandien, K.; Nilsson, S.; Gustafsson, J.-A.; Carlstedt-Duke, J. *J. Steroid Biochem. Mol. Biol.* **1993**, *45*, 239–250.
(59) Hollenberg, S. M.; Evans, R. M. *Cell* **1988**, *55*, 899–906.
(60) Freedman, L. P.; Yoshinaga, S. K.; Vanderbilt, J. N.; Yamamoto, K. R. *Science* **1989**, *245*, 298–301.
(61) Schena, M.; Freedman, L. P.; Yamamoto, K. R. *Genes Dev.* **1989**, *3*, 1590–1601.
(62) Zandi, E.; Galli, I.; Döbbeling, U.; Rusconi, S. *J. Mol. Biol.* **1993**, *230*, 124–136.

It is also somewhat unsatisfactory that no role as yet has been found for the "ninth" conserved cysteine residue (Cys500).

The fact that the GR DBD is well characterized in terms of structure, dynamics, and biochemical behavior makes it an interesting model system for detailed biophysical studies of a more general nature, for instance concerning the thermodynamic basis for sequence specific DNA–protein interactions, or as a test system for molecular dynamics simulations of free energy differences in complexes with various response elements.

Although there is a very large degree of sequence identity within the DBDs of the various members of the hormone receptor superfamily, there are some remaining functions which cannot be addressed using the GR DBD as a model system. Probably the most important unresolved aspect is the heterodimerization of the TR, VDR, RAR, and the peroxisome proliferator-activated receptor (PPAR) with RXR at direct repeat response elements with different half-site spacings (refs 18, 57, and 63–66 and work cited therein).

(63) Green, S. *Nature* **1993**, *361*, 590–591.
(64) Perlmann, T.; Rangarajan, P. N.; Umesuno, K.; Evans, R. M. *Genes Dev.* **1993**, *7*, 1411–1422.
(65) Towers, T. L.; Luisi, B. F.; Asianov, A.; Freedman, L. P. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6310–6314.
(66) Gearing, K. L.; Göttlicher, M.; Teboul, M.; Widmark, E.; Gustafsson, J.-A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1440–1444.
(67) Wilson, T. E.; Paulsen, R. E.; Padgett, K. A.; Milbrandt, J. *Science* **1992**, *256*, 107–110.
(68) Ueda, H.; Sun, G.-C.; Murata, T.; Hirose, S. *Mol. Cell. Biol.* **1992**, *12*, 5667–5672.
(69) Laudet, V.; Hänni, C.; Coll, J.; Catzeflis, F.; Stéhelin, D. *EMBO J.* **1992**, *11*, 1003–1013.
(70) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946–950.

An additional class of DBDs within the hormone receptor superfamily, that binds to single response elements without forming homo- or heterodimers, is represented by the NGFI-B orphan receptor. The DBD of this receptor binds to a single-site response element containing eight DNA base pairs, and a short segment (the "A-box") at the C-terminus, outside the consensus DBD, is required for identification of the two "extra" DNA base pairs.[67] Similar properties have been observed in at least one more orphan receptor DBD.[68] Additional structure determinations are clearly required to elucidate the binding modes of the heterodimerizing and monomeric binding DBDs.

## Concluding Remarks

We have in this Account described the studies resulting in the identification, characterization, and structure determination of DNA-binding domains of nuclear hormone receptors, especially the GR DBD. The studies have yielded an understanding of the molecular basis for DNA binding by steroid hormone receptors. The GR DBD has, within a few years, become one of the most well-characterized eucaryotic DNA-binding proteins which may be used as a model system for addressing the function of related sequences as well as more general questions concerning the biophysical chemistry of sequence specific DNA–protein interactions.

*Our studies referred to in this Account have been supported by grants from the Swedish Medical Research Council (No. 13X-2819) and the Swedish Natural Sciences Research Council.*