

# SCIENTIFIC REPORTS



OPEN

## The first crystal structure of a DNA-free nuclear receptor DNA binding domain sheds light on DNA-driven allostery in the glucocorticoid receptor

Filipp Frank , C. Denise Okafor & Eric A. Ortlund

The glucocorticoid receptor (GR) is a steroid hormone receptor of the nuclear receptor family that regulates gene expression in response to glucocorticoid hormone signaling. Interaction with specific GR DNA binding sequences causes conformational changes in the GR DNA binding domain (DBD) that result in recruitment of specific sets of co-regulators that determine transcriptional outcomes. We have solved the crystal structure of GR DBD in its DNA-free state, the first such crystal structure from any nuclear receptor. In contrast to previous NMR structures, this crystal structure reveals that free GR DBD adopts a conformation very similar to DNA-bound states. The lever arm region is the most variable element in the free GR DBD. Molecular dynamics of the free GR DBD as well as GR DBD bound to activating and repressive DNA elements confirm lever arm flexibility in all functional states. Cluster analysis of lever arm conformations during simulations shows that DNA binding and dimerization cause a reduction in the number of conformations sampled by the lever arm. These results reveal that DNA binding and dimerization drive conformational selection in the GR DBD lever arm region and show how DNA allosterically controls GR structure and dynamics.

Glucocorticoids (GCs) are steroid hormones that are involved in various fundamental processes at the cellular and organismal level. Their anti-inflammatory and immune-suppressive functions have made them one of the most widely used class of drugs. However, their additional, undesired metabolic activities can cause severe side effects and limit their therapeutic use<sup>1,2</sup>. Separating beneficial anti-inflammatory from adverse side effects is the main challenge for improving GC drug performance.

In cells, GCs bind to the glucocorticoid receptor (GR), a steroid hormone nuclear receptor that regulates the expression of thousands of genes<sup>3</sup>. GR integrates hormone signals in a cell-type and gene-specific manner to enhance or repress transcriptional activity. GC binding to the GR ligand binding domain (LBD) in the cytoplasm causes it to translocate to the nucleus where its DNA binding domain (DBD) recognizes and binds to specific GR DNA binding sequences (GBSs) or inverted repeat GBSs (IR-GBSs). A GBS or IR-GBS can modulate expression of nearby genes by transactivation or transrepression, respectively<sup>4-6</sup>. The consensus GBS is ~15 base pairs in length and consists of two hexameric half-sites, AGAACA, separated by a spacer of three base pairs<sup>7</sup>. This spacer length is strictly required for dimerization of two GR DBDs in a head-to-head orientation with their dimerization loops (D-loops) facing and contacting each other<sup>8</sup>. This dimerization mode facilitates highly cooperative DNA binding<sup>9</sup>. The consensus IR-GBS, CTCC(N)<sub>0-2</sub>GGAGA, does not require a strict spacing between its two binding sites<sup>5</sup> and two GR molecules interact with IR-GBSs in a head-to-tail orientation on opposite sides of the DNA with their D-loops directed away from each other<sup>5</sup>. GR – IR-GBS interaction is characterized by negative cooperativity with one high affinity site and a second site of much lower affinity<sup>5</sup>.

Transactivation and transrepression require the selective recruitment of co-activators and co-repressors, respectively<sup>10-13</sup>. Furthermore, the level of transactivation itself is strongly affected by the exact sequence of the GBS<sup>4</sup>. These observations suggest that differences in DNA modulate GR structure, which is then sensed by its

Department of Biochemistry, Emory University School of Medicine, Atlanta, GA, 30322, USA. Correspondence and requests for materials should be addressed to E.A.O. (email: [eortlund@emory.edu](mailto:eortlund@emory.edu))

coregulators. In fact, crystal structures of GR DBD bound to various GBSs revealed a loop region in the DBD, the lever arm between the DNA recognition helix and the dimerization loop, whose conformation is sensitive to the bound DNA's sequence<sup>4</sup>. Different GBSs are found with different conformations of the lever arm region and mediate different transcriptional responses<sup>4</sup>. Similarly, DNA sequence also modulates the conformation of the dimerization loop in a head-to-head dimer<sup>14</sup>. Therefore, DNA may allosterically affect GR structure and, as a consequence, transcription<sup>1</sup>.

GR $\gamma$ , a GR isoform resulting from alternative splicing, differs from GR $\alpha$  only by a single arginine inserted into the lever arm. This difference results in distinct DNA binding properties, protein interaction partners, differential regulation of specific target genes and distinct functional roles at cellular level<sup>4,15–17</sup>. Together these studies suggest that the lever arm region is a flexible element whose specific conformational and dynamical state in a particular GR-DNA complex is responsible for the selective recruitment of co-factors to specific genomic loci.

While there are numerous crystal structures of GR DBD in complex with DNA of varying sequences, the only available structures of GR DBD in its free state were solved using NMR<sup>18–20</sup>. Comparison of free DBD NMR structures with crystal structures of the DNA-bound form suggested that DNA-binding is accompanied by specific conformational changes involving mostly the C-terminal half of the molecule<sup>18–20</sup>. In particular, no evidence was observed for the short distorted helix that forms part of the second zinc finger binding domain<sup>20</sup>. The authors reasoned that formation of this helix upon DNA binding may give rise to a conformational change in the adjacent D-loop, thus facilitating cooperative binding. Further, large conformational changes compared to the DNA-bound state were observed in the lever arm region<sup>19</sup>. It was suggested that rotation of the protein backbone around a point in the lever arm acts as a conformational relay through which small changes in the DNA recognition helix result in a reorientation of the D-loop.

Here, we have solved the crystal structure of the GR DBD in its DNA-free state. Its structure is similar to the DNA-bound form with differences in the lever arm and D-loop that are not as pronounced as previously observed in NMR structures. Molecular dynamics simulations of GR-DBD in complex with a GBS, an IR-GBS, and the free state provide evidence that DNA binding and dimerization reduce conformational sampling of the lever arm region in a sequence-specific manner. These observations offer a mechanistic explanation for allosteric control of GR structure and its transcriptional output by specific DNA sequences.

## Results

**General description of the crystal structure.** The GR DBD crystallized in its free form in space group P1 with eight molecules in the asymmetric unit (Table 1). The eight molecules contain four dimers that are linked by a disulfide bridge via Cys431 (Fig. 1A). While this dimer appears to be advantageous for crystallization (no reducing agents were present during crystallization) it is not likely to form under physiological conditions in the intracellular reducing environment. Purified GR DBD is monomeric in solution<sup>21</sup>.

Within each dimer, one molecule shows clear density within the C-terminal helix (H3) following the core DBD, whereas this stretch is mostly disordered in the other molecule (Fig. 1A,B). H3 had been resolved to varying extents in previous DNA-bound crystal structures. It lines up over the minor groove, makes contact with the DNA backbone 3 bp upstream of the GBS, and contains five lysine residues for potential further DNA interactions<sup>4</sup>. In the full-length receptor, the DBD is followed by a short linker connecting directly to the LBD. The orientation of H3 along the minor groove may thus serve to constrain the LBD to a conformation competent for dimerization. In the free state observed here, H3 is rotated by approximately 60 degrees so that it would lie across the minor groove of a bound DNA (Fig. 1C). This conformation is likely influenced by crystallization and the formation of the disulfide-bridged dimer. However, the current structure suggests that the orientation of H3 relative to the core DBD is not fixed in the absence of DNA and adopts its active conformation only upon GR binding to a recognition sequence.

The core DBD (residues 421–491) of the individual molecules within the unit cell are very similar in structure (Fig. 2A; average pairwise RMSD = 0.49 Å). The largest differences are located within the lever arm region (Fig. 2A,B) and are similar in extent to the variations observed in GR DBD bound to different GBSs (Fig. 2C)<sup>3,4</sup>. His453 is known to either pack into the protein's core or flip outwards depending on the sequence of bound DNA<sup>4</sup>. In the free crystal structure, His453 is flipped out in all molecules. However, close inspection of electron density shows evidence for an alternative, packed conformation in four molecules (Fig. 3). These observations suggest that the lever arm, including His453, is mobile in the free state and samples multiple conformations resembling the various structures in the presence of different DNA sequences.

**Significant differences with previous free DBD structures solved by NMR.** Previous structures of free GR DBD, solved by NMR, reveal significant conformational changes compared to the DNA-bound form found in crystal structures<sup>18–20</sup>. Hydrogen bonding interactions defining the short antiparallel beta-sheet in the N-terminal Zn finger were not present in NMR structures<sup>8,21</sup>. The short, distorted helix in the second zinc finger (residues 469–473) had not been unambiguously observed in NMR experiments<sup>18,20</sup>. The largest change in conformation was described in the adjacent D-loop (residues 457–465), which mediates dimerization of GR DBD on DNA<sup>18</sup>. Finally, NMR studies identified the lever arm region (residues 450–456) as a hinge around which small differences within the preceding DNA reading helix are amplified to cause large conformational changes in the succeeding D-loop<sup>19</sup>. Although some of the conformational differences in NMR structures may have been due to a lack of well-defined NMR constraints<sup>18</sup>, it had been accepted that DNA binding is accompanied by significant conformational changes in the DBD.

Comparison of the current structure of free GR DBD with NMR structures as well as DNA-bound crystal structures shows that the free GR DBD crystal structure described here is more similar to the DNA-bound form (Fig. 4A and B; RMSDs 0.72–1.08 Å; Table 2) than to the solution structures of free GR DBD (Fig. 4C; RMSDs 1.83–2.77 Å; Table 2). In particular, the N-terminal beta-sheet as well as the short, distorted helix are formed

Wavelength	1.000
Resolution range	36.04–2.5 (2.589–2.5)
Space group	P 1
Unit cell	65.8, 65.4, 72.7; 71.3, 85.0, 68.4
Unique reflections	33141 (2095)
Multiplicity	2.4 (1.4)
Completeness (%)	89.47 (56.85)
Mean I/sigma(I)	14.5 (1.3)
Wilson B-factor	53.58
R <sub>meas</sub>	0.18 (0.65)
R <sub>rim</sub>	0.10 (0.46)
CC <sub>1/2</sub>	(0.56)
Reflections used in refinement	33024 (2092)
Reflections used for R-free	1991 (126)
R-work	0.22 (0.33)
R-free	0.22 (0.35)
Number of non-hydrogen atoms	4707
macromolecules	4664
Zn <sup>2+</sup>	16
solvent	27
Protein residues	619
RMS(bonds)	0.006
RMS(angles)	1.00
Ramachandran favored (%)	97.1
Ramachandran allowed (%)	2.6
Ramachandran outliers (%)	0.3
Rotamer outliers (%)	0.59
Clashscore	10
Average B-factor	76.78
macromolecules	76.96
ligands	54.02
solvent	58.69

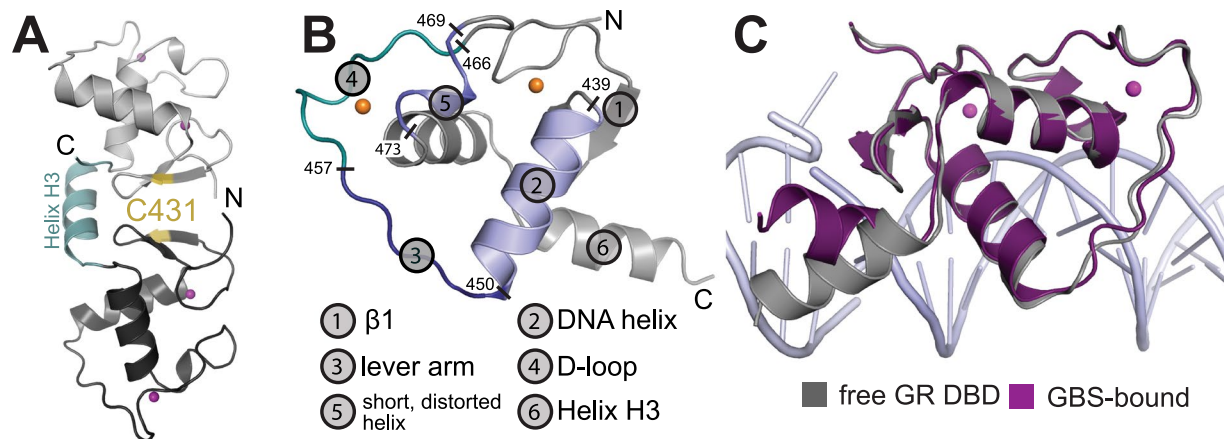
**Table 1.** X-ray diffraction data. Statistics for the highest-resolution shell are shown in parentheses.

and identical in structure to the DNA-bound state (Fig. 5A,B) and the D-loop conformation is similar to the one found in DNA-bound structures (Fig. 4A,B). The conformation around the lever arm region is characterized by a certain degree of flexibility in the current structures as evidenced by multiple different conformations in the eight monomers of the unit cell as well as electron density indicative of multiple distinct conformations in His453 within the monomers (Fig. 3). However, it is very different from the conformations found in NMR structures and more closely resembles the DNA-bound state (Fig. 4).

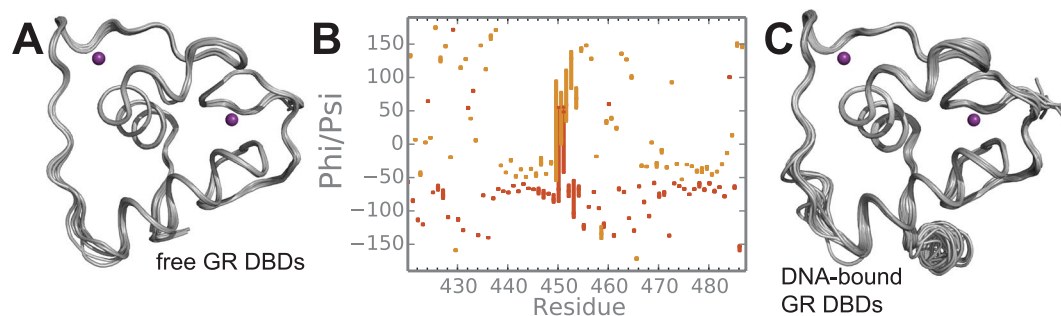
Changes between the DNA-free and DNA-bound crystal structures are minor. Differences are located around the mobile lever arm region and extend into the D-loop (Fig. 4A,B). Interestingly, the peptide backbone conformation of the DNA-binding helix, including the residues in direct contact with DNA, is essentially unchanged between the free and GBS DNA-bound forms (Fig. 5C). However, the side chains of these residues are solvent-exposed in the free form and as a result they are highly dynamic, and their positions are not resolved. Thus, the side chains appear to become ordered into their active conformations only upon contact with DNA.

An important advantage of NMR spectroscopy over crystallography is that it probes macromolecular structure in solution, rather than in a solid, crystalline state. Under these conditions dynamic protein elements fluctuate between different conformations as they would in their physiological environment in the cell. The fact that no single local conformation for such elements exists in solution may result in poorly defined regions. In a crystalline environment, on the other hand, the protein is generally stabilized in a single, low energy state. In the asymmetric portion of the unit cell, we observe eight independent DNA-free GR DBD molecules, which closely resembles the DNA-bound state observed in other crystal structures. This suggests that the DNA binding conformation is accessible to the GR DBD in solution

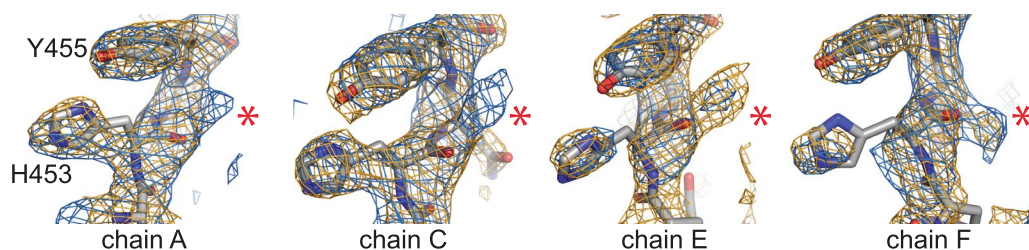
**The lever arm is flexible in MD simulations.** Lever arm conformation is sensitive to the presence or absence of DNA, the sequence of bound DNA in a GR-DNA complex, and even to external factors such as buffer conditions during crystallization<sup>4</sup>. One flexible part of the lever arm is residue His453, which can adopt two distinct conformations: packed in the core of the protein (“in”) or flipped outwards (“out”). Together, these observations suggest that dynamics in this region play an important role in GR DBD function. To explore this further, we carried out two replicates each of 1.0 μs molecular dynamics simulations containing GR DBD molecules in



**Figure 1.** Crystal structure of the DNA-free GR DBD. **(A)** Dimerization via formation of a disulfide between adjacent C431 residues. **(B)** Overall structure of the free GR-DBD with important structural elements highlighted. **(C)** The orientation of helix H3 differs between the free DBD and the DNA-bound state. In the DNA-bound state H3 lies across the major groove and makes interactions with the DNA phosphate backbone. The conformation of H3 found in the free state is rotated by  $\sim 60$  degrees so that the helix lines up perpendicular to the major groove.



**Figure 2.** Lever arm flexibility. **(A)** Structural alignment of the eight monomers found in the asymmetric unit of the crystal. **(B)** Box plot of the Phi and Psi backbone dihedral angles observed in the eight GR DBD monomers. Each box represents the interquartile range. The highest variability is observed in the lever arm region. **(C)** Structural alignment of crystal structures of the GR DBD bound to different GBSs.



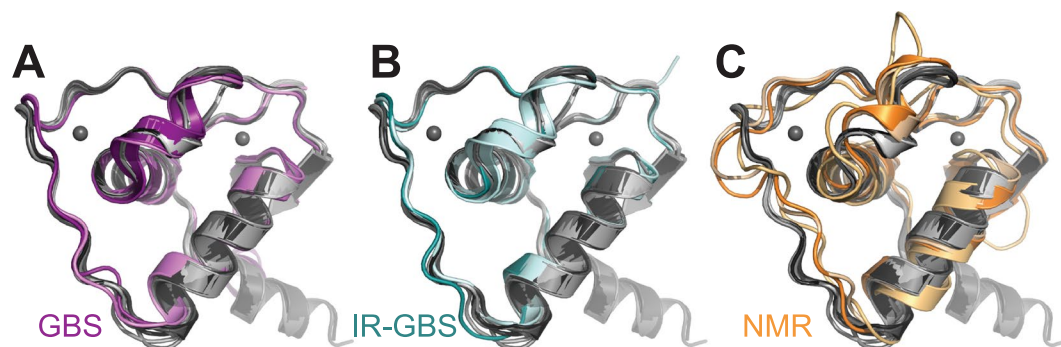
**Figure 3.** Electron density indicative of multiple conformations for His453. Shown in orange is a  $2F_o - F_c$  map at  $1.5\sigma$  and in blue is a composite omit map contoured at  $2\sigma$ . The red asterisks mark the alternate location of the His453 side chain in the “in” conformation.

different functional states: (i) the free GR DBD reported here (molecule C of the crystal structure), (ii) a GR DBD dimer bound to a representative GBS containing a perfect palindromic hexameric repeat with two high-affinity binding sites (GBS1 and GBS2; PDB 3G99, chains A and B, respectively)<sup>4</sup>, and a GR DBD dimer bound to a palindromic IR-GBS (PDB 4HN5) containing a high-affinity (IR-GBS1) and a low-affinity (IR-GBS2) binding site<sup>5</sup>. The two GR binding sites within the IR-GBS are characterized by strong negative cooperativity of binding<sup>5</sup> and IR-GBS2 is likely not fully occupied *in vivo*. Its presence in the structure resulted from stabilizing contacts that facilitated crystallization.

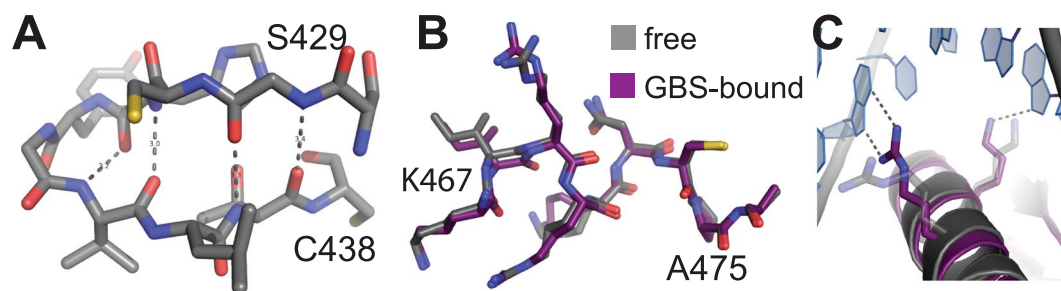


RMSD [Å]	1GDC	1RGD	2GDA	3G99_A	3G99_B	4HN5_A	4HN5_B
Free GR DBD	2.03	2.77	1.83	0.976	1.08	0.92	0.73
1GDC		2.69	0.759	1.95	1.91	1.94	1.99
1RGD			2.63	2.79	2.92	2.76	2.73
2GDA				1.80	1.80	1.80	1.78
3G99_A					0.614	0.67	0.54
3G99_B						0.67	0.62
4HN5_A							0.70

**Table 2.** Pairwise RMSD between different GR DBD structures. For the free GR DBD, molecule C of the crystal structure described here was used.



**Figure 4.** Structural alignment of DNA-free GR DBD with previous structures. (A–C) Structural alignments of the 8 monomers found in the crystal structures of free GR-DBD (in grey) overlaid with structures of DNA-bound GR DBD from crystal structures 3G99 (A) and 4HN5 (B) and DNA-free GR DBD from NMR structures 1GDC, 1RGD, and 2GDA (C).

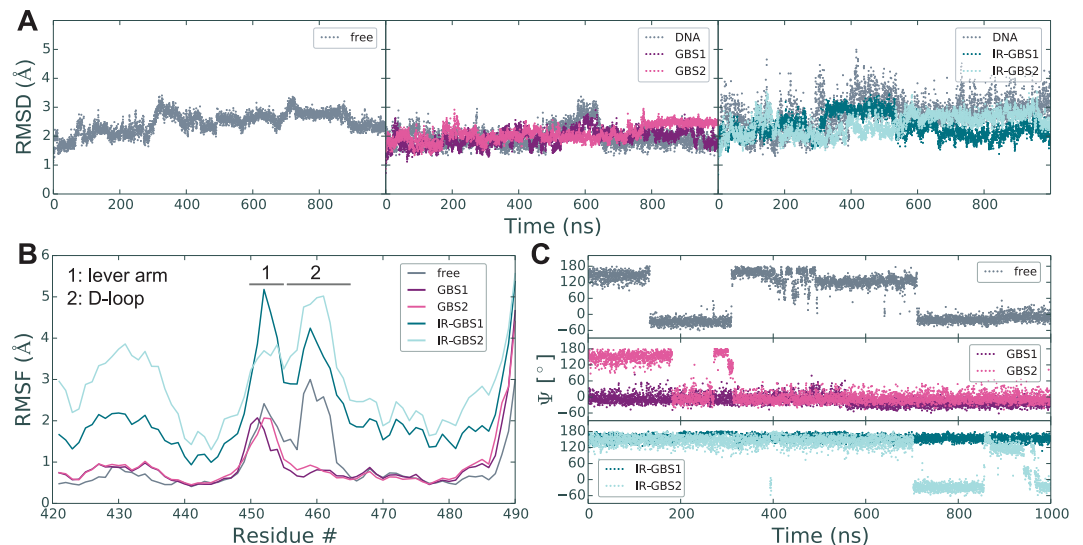


**Figure 5.** Structural elements of GR DBD in the DNA-free state. (A) Hydrogen-bonds forming the antiparallel beta-sheet in the N-terminal Zn finger in the DNA-free GR DBD structure. (B) The short, distorted helix in the C-terminal Zn finger in DNA-bound (PDB ID: 3G99) and free GR DBD adopt the same conformation. (C) Conformation of the DNA reading helix. Shown is a structural alignment of a representative molecule of DNA-free GR DBD with DNA-bound GR-DBD (3G99). The DNA reading helix aligns well between the two proteins and the position of two residues that make direct contact with the DNA in 3G99 is unchanged in the DNA-free structure.

To further validate results from this initial set of simulations, two alternative systems for the DNA-free state and the GBS bound state were simulated, also in duplicate: (i) molecule A of the crystal structure reported here and (ii) PDB ID 3G9J<sup>4</sup>, a structure similar to 3G99, but with 18 nucleotide DNA strands instead of 16.

All GR DBD molecules were characterized by stable root mean square displacements (RMSDs) over the course of the entire simulation (Fig. 6A and Supplementary Figs 1 and 2). Root mean square fluctuations (RMSFs), a measure of flexibility, reveal the lever arm and D-loop as the most mobile regions in the proteins (Fig. 6B and Supplementary Figs 3 and 4). As expected, the head-to-head dimerization between GBS1 and GBS2 stabilized the D-loop of those molecules, which resulted in lower RMSFs in that region compared to the free state or the IR-GBS1 and IR-GBS2 states. Due to their spatial proximity, stabilization of the D-loop may have direct effects on the dynamical behavior of the lever arm region.

Interestingly, the IR-GBS complex exhibited the greatest RMSFs amongst all the simulations, in both replicates. This was surprising since the free structure and the GBS complex behaved similarly. It appears that binding to an IR-GBS



**Figure 6.** Overview of MD simulations. **(A)**  $C_{\alpha}$  root mean square displacements (RMSDs) of replicate #1 simulations. **(B)**  $C_{\alpha}$  root mean square fluctuations (RMSFs) of replicate #1 simulations. **(C)** Dynamics of His453 flipping between ‘in’ and ‘out’ conformations. Shown are the trajectories of the His453  $\psi$  dihedral angle, which reflects how the residue flips between its ‘in’ ( $\sim -28^{\circ}$ ) and ‘out’ ( $\sim 143^{\circ}$ ) conformation.

sequence induces significant flexibility in the DBD compared to the free or GBS-bound state. This is likely due to the strong negative cooperativity which disfavors simultaneous binding of two GR molecules at this DNA sequence. Previous studies showed that strong allosteric communication flows through the DNA to mediate this effect.

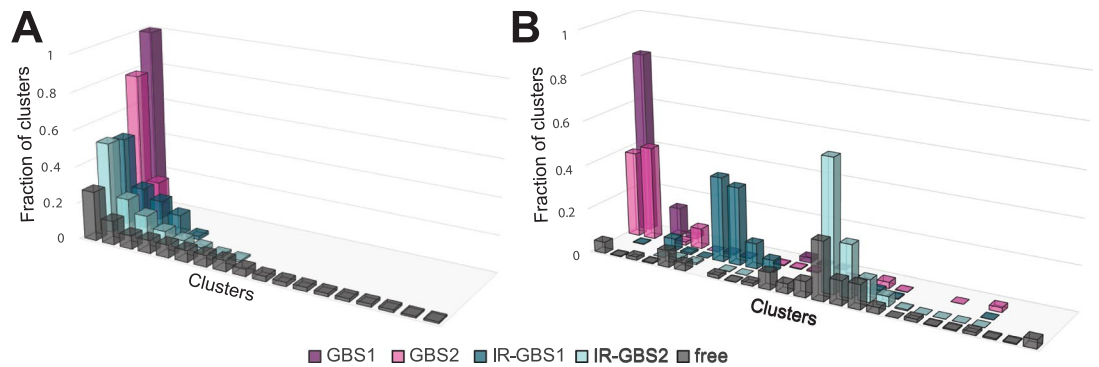
His453 in the lever arm was observed to flip between its two main conformations in a number of cases (Fig. 6C). The conformation of His453 is found in two main states represented by  $\psi$  backbone dihedral angles of  $143^{\circ} \pm 20^{\circ}$  (out) and  $-28^{\circ} \pm 11^{\circ}$  (in), respectively. Switching between these states is rapid with a potential on-pathway intermediate at  $119^{\circ} \pm 13^{\circ}$  in the out-to-in transition. These data underscore the dynamic behavior of this residue and are consistent with the multiple conformations detected in crystal structures.

### Molecular dynamics uncover distinct dynamic properties for different functional states of the GR DBD.

Previous MD simulations over 50–100 ns of GR DBD in different states revealed small changes in lever arm RMSF values that correlated with transcriptional output of the simulated complexes<sup>22,23</sup>. Considering all of the 1 $\mu$ s simulations performed in this study, we did not observe significant, systematic differences in lever arm RMSFs between the DNA-free and the DNA-bound state. To more closely examine the relationship between lever arm dynamical behavior and GR DBD functional state, we performed a conformational clustering analysis. For each simulation, 50,000 frames, equally distributed in time over the course of the trajectories, were clustered based on their lever arm conformations. Clustering frames from each trajectory independently showed a trend in the number of conformations sampled by each monomer. GBS1 molecules maintained a single conformation over the entire simulation, whereas GBS2 is distributed over two and three separate clusters in the duplicate simulations (Fig. 7A and Supplementary Fig. 5). IR-GBS1 and IR-GBS2 conformations distributed into an intermediate number of clusters (IR-GBS1: 5 and 6; IR-GBS2: 8 and 3). Strikingly, the greatest number of clusters was identified for simulations of the DNA-free state: 12 and 18 for molecule C, and 10 and 8 for molecule A (Fig. 7A and Supplementary Fig. 5). The increase in the number of clusters in the free structure was not a consequence of a general increase in protein mobility as RMSF values were comparable to simulations of GBS1 and GBS2. It more likely results from the lack of stabilizing contacts in the adjacent D-loop that are present in the GBS structures. The resulting increase in D-loop flexibility may amplify the conformational space accessible to the lever arm. D-loop flexibility alone, however, does not explain the entire effect on lever arm clustering since IR-GBS structures would be expected to sample more distinct conformations. The smaller number of clusters for IR-GBS1 and IR-GBS2 may be explained by stabilization through neighboring contacts with DNA. Thus, the clustering analysis suggests that, while flexible to a similar degree in all states, lever arm conformational sampling is progressively reduced with the addition of nearby contacts through DNA binding via the DNA-helix and dimerization via the D-loop.

The clustering analysis was also carried out on the two alternative systems that were simulated for the DNA-free state and the GBS bound state. Again, the DNA-free state was characterized by a greater number of clusters compared to the DNA-bound states, confirming that these observations are robust in different structural contexts (Supplementary Fig. 6).

To more directly compare lever arm conformations across the different simulations, we performed a combined cluster analysis using frames equally distributed over the duplicate simulations of the initial set of structures (GBS1 and GBS2, IR-GBS1, and IR-GBS2, and DNA free; Fig. 7B). The analysis identified a total of 25 unique clusters among the trajectories. The results of the two clustering analyses are in overall agreement in that lever arm conformations identified for GBS1 and GBS2 were distributed over a smaller number of clusters compared to IR-GBS1 and IR-GBS2. The free state was most heterogeneous. Examination of the degree of conformational overlap between the



**Figure 7.** Cluster analysis for lever arm conformations. **(A)** Clustering of individual MD trajectories shows trends in the conformational space sampled in different states of GR. For each molecule clusters are ordered by decreasing frequency. **(B)** Combined clustering of the duplicate trajectories for DNA-free GR DBD (molecule C), 4HN5, and 3G99 shows which clusters are shared between the different molecules. The results show that the two monomers bound to a GBS spend ~40% of their time in the same conformation. Further significant overlap is observed between free GR and the low affinity binding site within the IR-GBS complex (IR-GBS2). Molecule IR-GBS1 does not have significant overlap with any other molecules.

different DBD monomers reveals that the greatest overlap existed between GBS1 and GBS2 (~40%; Fig. 7B). Another significant overlap occurred between the free and IR-GBS2 states, which share three clusters covering at least 10% of the time each. Interestingly, IR-GBS1 is somewhat separate from the other states having only minor overlap with other molecules and spending most of its time in conformations exclusive to this state (Fig. 7B).

The results of the clustering analyses allow several conclusions. First, the large number of clusters observed in the free state shows that the lever arm samples multiple distinctive conformations in the absence of DNA. Conformational sampling is restricted upon binding to DNA and is further reduced by dimerization via the D-loop as seen in complexes with a GBS. This suggests that DNA binding and dimerization allosterically bias the conformation of the lever arm region in a specific DBD-DNA complex. Importantly, this mechanism is consistent with changes in GR DBD chemical shifts observed by  $^1\text{H}$ - $^{15}\text{N}$ -HSQC experiments in the presence and absence of DNA<sup>14</sup>. Second, while conformational sampling of the lever arm is reduced in DNA-bound complexes, it retains enough flexibility to sample a small number of distinct conformations. These may provide the ability to interact with different co-regulators while bound to a particular DNA sequence and the recruitment of a specific binding partner may depend on further signals. The different regulatory outcomes achieved by GBS and IR-GBS sequences may be explained by a combination of distinct lever arm conformations and increased lever arm conformational sampling in GR bound to an IR-GBS.

## Conclusions

We have solved the first crystal structure of a free steroid receptor DNA binding domain. The structure contains eight independent molecules in the asymmetric unit and superposition of these molecules reveals that the free GR DBD is very similar to its DNA-bound form. Differences are minor and located mostly within the lever arm and the adjacent D-loop. This is in contrast to NMR studies of the free GR DBD, which had identified a number of protein segments with major conformational differences compared to DNA-bound crystal structures<sup>18–20</sup>. The absence of any large conformational changes between the DNA-free and DNA-bound crystal structures suggests that minor structural variations found within the lever arm and the D-loop are responsible for their differential functional properties. This view is consistent with previous studies showing that the lever arm and D-loop are the main sites of variability in structures of GR DBD bound to DNA of different sequences<sup>4,14,23</sup>.

In addition to variability between different GR DBD crystal structures, the lever arm is also the most flexible element in molecular dynamics simulations, consistent with previous reports<sup>22,23</sup>. We reasoned that differences in its intrinsic dynamics are responsible for functional differences. However, lever arm RMSF values did not display significant differences between the DNA-free and DNA-bound states. Instead, a clustering analysis of the MD trajectories based on lever arm conformation revealed differences between the different functional states: GBS-bound molecules spent most of their trajectories in the same conformational cluster, IR-GBS-bound molecules were characterized by an intermediate number of clusters, and the lever arm of DNA-free GR DBD sampled the most diverse conformations (Fig. 7A). A combined cluster analysis further uncovered that the two GBS-bound molecules populated a shared cluster, while there was only minimal overlap with clusters populated in any of the other simulated molecules. Thus, clustering analysis of MD trajectories is able to distinguish the different functional states of the GR DBD.

Together, the data presented here show that large-scale conformational changes in the GR DBD may not be required for DNA binding. It is rather more likely that the lever arm undergoes conformational selection from a highly heterogeneous collection of clusters in the free state to a small number of unique clusters in the functional, DNA-bound state. Furthermore, lever arm clusters of DNA-bound GR DBD correlated with the direction of transcriptional output mediated by the particular DNA segments. These insights contribute to our understanding of how DNA sequence inputs control the conformational landscape of the GR DBD, which may have consequences for the transcriptional regulation of target genes.

## Methods

**Protein expression and purification.** The GR DBD (residues 418–517) was cloned into pSmt3 vector for expression with an N-terminal His<sub>6</sub>-Sumo-tag<sup>24</sup> using BamHI and NotI restriction sites. The protein was expressed in E coli BL21 DE3. Cells were grown to an OD<sub>600</sub> of 0.8–1.0 and overexpression was induced with 0.5 mM IPTG for 4 hours at 37 °C. Protein was purified by nickel affinity chromatography, followed by cleavage of the His<sub>6</sub>-Sumo-tag using Ulp-1 enzyme. After removal of the His<sub>6</sub>-Sumo-tag by a second round of nickel-affinity chromatography, the protein was concentrated and purified by size exclusion chromatography in 25 mM Tris pH7.5, 100 mM NaCl, and 5% glycerol. Purified protein was flash frozen and stored at a concentration of ~5 µg/µL.

**Protein crystallization and data collection.** In an attempt to crystallize GR DBD in complex with an RNA GBS mimic present in the lncRNA Gas5<sup>25</sup> protein was mixed with a 35nt short RNA hairpin (sequence: GCTCCCAGTGGTCTTTGTAGACTGCCTGATGGAGC) at a 1.1:1 ratio of protein to RNA at a concentration of ~4 µg/µL. Protein was crystallized by hanging drop vapor diffusion at 18 °C in solutions containing 25% PEG3350, 50 mM Tris pH8.5, and 200 mM di-sodium tartrate. Crystals were cryo-protected by addition of well solution containing 20% glycerol and flash frozen in liquid nitrogen. Data were collected remotely from the South East Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source 22ID beamline (Argonne National Laboratories, Chicago). Data were processed and scaled using *HKL2000* (ref Ottwinowski) and phased by molecular replacement using Phaser-MR in *PHENIX*<sup>26</sup>. PDB 4HN5 chain A was used as the search model. Model building and refinement were carried out using *Coot*<sup>27</sup> and *PHENIX*<sup>26</sup>. No electron density indicating the presence of nucleic acid was found anywhere in the unit cell.

**Model construction for molecular dynamics simulations.** Three GR DBD complexes were prepared for molecular dynamics simulations: 1) Free GR DBD (PDB 6CFN) (GR monomer), 2) GR-GBS (PDB 3G99) (dimer), and 3) GR-IR-GBS (PDB 4HN5) (two monomers). For consistency, all GR sequences were truncated to contain residues 421–490 (human numbering).

**Molecular dynamics simulations.** The complexes were solvated in an octahedral box of TIP3P water with a 10 Å buffer around the protein/protein-DNA complexes. Na<sup>+</sup> and Cl<sup>-</sup> ions were added to neutralize the protein and achieve physiological conditions. All systems were set up using xleap in *AmberTools* (version 17) and force fields contained in *Amber16*<sup>28</sup>. The *ff14SB* force field was used for protein and OL15 for DNA. Zinc metal centers were modeled with the zinc Amber force field (ZAFF)<sup>29</sup>. All minimizations and simulations were performed with *Amber16*. Systems were minimized with 5000 steps of steepest descent followed by 5000 steps of conjugate gradient minimization with 500-kcal/mol-Å<sup>2</sup> restraints on all non-solvent atoms. The minimization was repeated sequentially with the following conditions: i) 500-kcal/mol-Å<sup>2</sup> restraints on protein and Zn atoms only (for GR-DNA complexes), ii) 500-kcal/mol-Å<sup>2</sup> restraints on Zn<sup>2+</sup> atoms and coordinating cysteines only, iii) 500-kcal/mol-Å<sup>2</sup> restraints on Zn atoms only, and iv) no restraints. The systems were heated from 0 to 300 K using a 100-ps run with constant volume periodic boundaries and 10-kcal/mol-Å<sup>2</sup> restraints on all non-solvent atoms. The following stepwise equilibration was performed using the NPT ensemble: i) 1 ns with 10-kcal/mol-Å<sup>2</sup> on all non-solvent atoms; ii) 1 ns with 10-kcal/mol-Å<sup>2</sup> on protein and Zn atoms only (GR-DNA complexes), iii) 1 ns with 10-kcal/mol-Å<sup>2</sup> on Zn<sup>2+</sup> atoms and coordinating cysteines only; and iv) 1 ns with 10-kcal/mol-Å<sup>2</sup> on Zn atoms only. All restraints were removed and 1 µs production simulations were performed for each system in the NPT ensemble. A 2-fs time step was used and all bonds between heavy atoms and hydrogens were fixed with the SHAKE algorithm<sup>30</sup>. A cutoff distance of 10 Å was used to evaluate long-range electrostatics with particle mesh Ewald and for van der Waals forces. For analysis, 50,000 evenly spaced frames were obtained from each simulation.

**Analysis.** Structural averaging and analysis were performed with the *CPPTRAJ* module<sup>31</sup> of *AmberTools*. The MMTSB toolset<sup>32</sup> was used to perform a cluster analysis of GR monomer conformations. The *k-means* clustering algorithm (kclust) was used with a 2.3 Å RMSD cutoff radius. This radius was selected as the optimal radius to reflect the conformational variability of the selected fragment (i.e. the lever arm) across all GR monomers. 50,000 evenly spaced snapshots from each simulation were RMSD fitted to the lever arm and clustered. To identify sampled conformations that overlap between the free and DNA-bound GR monomers in the two replicate simulations, 5,000 evenly spaced snapshots from each monomer (50,000 total) were combined, RMSD fitted and clustered.

## References

- Vandewalle, J., Luypaert, A., De Bosscher, K. & Libert, C. Therapeutic Mechanisms of Glucocorticoids. *Trends Endocrinol. Metab.* **29**, 42–54 (2018).
- Whitehouse, M. W. Anti-inflammatory glucocorticoid drugs: reflections after 60 years. *Inflammopharmacol* **19**, 1–19 (2010).
- Weikum, E. R., Knuesel, M. T., Ortlund, E. A. & Yamamoto, K. R. Glucocorticoid receptor control of transcription: precision and plasticity via allostery. *Nat Rev Mol Cell Biol* **18**, 159–174 (2017).
- Meijsing, S. H. *et al.* DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity. *Science* **324**, 407–410 (2009).
- Hudson, W. H., Youn, C. & Ortlund, E. A. The structural basis of direct glucocorticoid-mediated transrepression. *Nat Struct Mol Biol* **20**, 53–58 (2013).
- Surjit, M. *et al.* Widespread Negative Response Elements Mediate Direct Repression by Agonist- Liganded Glucocorticoid Receptor. *Cell* **145**, 224–241 (2011).
- Nordeen, S. K., Suh, B. J., Kühnel, B. & Hutchison, C. A. Structural determinants of a glucocorticoid receptor recognition element. *Mol. Endocrinol.* **4**, 1866–1873 (1990).
- Luisi, B. F. *et al.* Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505 (1991).
- Haerd, T., Dahlman, K., Carlstedt-Duke, J., Gustafsson, J. A. & Rigler, R. Cooperativity and specificity in the interactions between DNA and the glucocorticoid receptor DNA binding domain. *Biochemistry* **29**, 5358–5364 (2002).



10. Bittencourt, D. *et al.* G9a functions as a molecular scaffold for assembly of transcriptional coactivators on a subset of Glucocorticoid Receptor target genes. *Proc Natl Acad Sci USA* **109**, 19673–19678 (2012).
11. Rogatsky, I., Luecke, H. F., Leitman, D. C. & Yamamoto, K. R. Alternate surfaces of transcriptional coregulator GRIP1 function in different glucocorticoid receptor activation and repression contexts. *Proc Natl Acad Sci USA* **99**, 16701–16706 (2002).
12. Chinenov, Y., Sacta, M. A., Cruz, A. R. & Rogatsky, I. GRIP1-associated SET-domain methyltransferase in glucocorticoid receptor target gene expression. *Proc Natl Acad Sci USA* **105**, 20185–20190 (2008).
13. Yang, C. K., Kim, J. H. & Stallcup, M. R. Role of the N-terminal activation domain of the coiled-coil coactivator in mediating transcriptional activation by beta-catenin. *Mol. Endocrinol.* **20**, 3251–3262 (2006).
14. Watson, L. C. *et al.* The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. *Nat Struct Mol Biol* **20**, 876–883 (2013).
15. Morgan, D. J. *et al.* Glucocorticoid receptor isoforms direct distinct mitochondrial programs to regulate ATP production. *Sci Rep* **1–10** <https://doi.org/10.1038/srep26419> (2017).
16. Birth, P., Schöne, S., Stelzl, U. & Meijnsing, S. H. Identification and characterization of BATF3 as a context-specific coactivator of the glucocorticoid receptor. *PLoS ONE* **12**, e0181219–15 (2017).
17. Thomas-Chollier, M. & Watson, L. C. A naturally occurring insertion of a single amino acid rewires transcriptional regulation by glucocorticoid receptor isoforms. In <https://doi.org/10.1073/pnas.1316235110/-/DCSupplemental> (2013).
18. Baumann, H. *et al.* Refined solution structure of the glucocorticoid receptor DNA-binding domain. *Biochemistry* **32**, 13463–13471 (1993).
19. van Tilborg, M. A. *et al.* Mutations in the glucocorticoid receptor DNA-binding domain mimic an allosteric effect of DNA. *J Mol Biol* **301**, 947–958 (2000).
20. van Tilborg, M. A. *et al.* Structure refinement of the glucocorticoid receptor-DNA binding domain from NMR data by relaxation matrix calculations. *J Mol Biol* **247**, 689–700 (1995).
21. Hård, T. *et al.* Solution structure of the glucocorticoid receptor DNA-binding domain. *Science* **249**, 157–160 (1990).
22. Schöne, S. *et al.* Sequences flanking the core-binding site modulate glucocorticoid receptor structure and activity. *Nat Commun* **7**, 12621 (2016).
23. Álvarez, L. D., Presman, D. M. & Pecci, A. Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain suggest a role of the lever-arm mobility in transcriptional output. *PLoS ONE* **12**, e0189588–18 (2017).
24. Mossessova, E. & Lima, C. D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* **5**, 865–876 (2000).
25. Hudson, W. H. *et al.* Conserved sequence-specific lincRNA-steroid receptor interactions drive transcriptional repression and direct cell fate. *Nat Commun* **5**, 5395 (2014).
26. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* **66**, 213–221 <https://doi.org/10.1107/S0907444909052925> 1–9 (2010).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst D* **60**, 2126–2132 <https://doi.org/10.1107/S0907444904019158> 1–7 (2004).
28. Case, D. A. *et al.* AMBER 2016. *University of California, San Francisco* (2016).
29. Peters, M. B. *et al.* Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **6**, 2935–2947 (2010).
30. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327–341 (1977).
31. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
32. Feig, M., Karanicolas, J. & Brooks, C. L. III MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modelling* **22**, 377–395 (2004).

## Acknowledgements

This work was supported by a W.M. Keck Foundation Medical Research Grant. C.D.O. is funded by an NIH-NIGMS Institutional Research and Career Development Award [Grant K12 GM000680]. Crystallographic data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract W-31-109-Eng-38.

## Author Contributions

F.F. performed protein expression, purification, crystallization, X-ray data collection, and solved the structure. C.D.O. performed molecular dynamics simulations and analysis. F.F., C.D.O. and E.A.O. designed experiments and discussed the data. F.F., C.D.O. and E.A.O. wrote the manuscript. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-31812-9>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018