



OPEN

Decoding the dual recognition mechanism of the glucocorticoid receptor for DNA and RNA: sequence versus shape

Johanna Hörberg & Anna Reymer

Transcription factors (TFs) regulate eukaryotic transcription through selective DNA-binding, can also specifically interact with RNA, which may present another layer of transcriptional control. The mechanisms of the TFs-DNA recognition are often well-characterised, while the details of TFs-RNA complexation are less understood. Here we investigate the dual recognition mechanism of the glucocorticoid receptor (GR), which interacts with similar affinities with consensus DNA and diverse RNA hairpin motifs but discriminates against uniform dsRNA. Using atomic molecular dynamics simulations, we demonstrate that the GR binding to nucleic acids requires a wide and shallow groove pocket. The protein effectively moulds its binding site within DNA major groove, which enables base-specific interactions. Contrary, the GR binding has little effect on the grooves geometry of RNA systems, most notably in uniform dsRNA. Instead, a hairpin motif in RNA yields a wide and shallow major groove pocket, allowing the protein to anchor itself through nonspecific electrostatic contacts with RNA backbone. Addition of a bulge increases RNA hairpin flexibility, which leads to a greater number of GR-RNA contacts and, thus, higher affinity. Thus, the combination of structural motifs defines the GR-RNA selective binding: a recognition mechanism, which may be shared by other zinc finger TFs.

Transcription factor proteins, via selective binding to DNA, regulate genetic transcription¹. Multiple experiments suggest that transcription factors (TFs) also specifically interact with RNA, which may present an additional layer of transcription regulation^{2–4}. RNA molecules are constantly present at the actively transcribed sites^{5–9}. As our understanding of the role of TFs-RNA interactions still evolves, the available data suggest that TF-RNA interactions can facilitate the association of TFs with the chromatin fibre and the search for their corresponding DNA binding sites. Examples include, nuclear TF Y (NF-YA) interactions with PANDA long noncoding RNA¹⁰, which titrates the TF from the target genes to reduce their expression; Yin Yang 1 TF (YY1) interactions with nascent transcripts to retain its presence at actively transcribed promoters¹¹; heat shock factor 1 (HSF1) that employs RNA as a scaffold for the TFs assembly on DNA¹². Furthermore, a recent study by Oksuz et al. reports that nearly half (48%) of TFs identified in human K562 cells appear to interact with RNA¹³. As TF proteins do not possess the characteristic structural domains of well-studied RNA binding proteins^{14,15}, little is known on the structural basis of RNA-TF associations.

Here we aim to provide atomic level insights into what drives the TF-RNA association, using as a model system the glucocorticoid receptor that can interact with RNA^{4,16–18}. The glucocorticoid receptor (GR) is a ubiquitous TF, which regulates the transcription of thousands of genes associated with development, immune responses, metabolism, inflammation, apoptosis, etc^{19–21}. To regulate a transcriptional response, GR needs a glucocorticoid binding, which induces the factor translocation from the cytoplasm to the nucleus to enable interactions with DNA. GR is widely dysregulated in disease, and thus is targeted by synthetic glucocorticoids to combat a range of autoimmune disorders and many cancers, as a part of combinatorial therapy^{22–24}. GR has a modular structure²⁵, including two well-characterised domains: a ligand-binding domain and a DNA-binding domain. The GR DNA-binding domain (DBD) consists of two Zn-fingers of Cys4-type, and associates on DNA in several configurations that lead to distinct transcriptional responses^{21,26–29}. The GR dimer-DNA association, in a head-to-head orientation, leads to the transcription induction. The GR monomer-DNA association, collaboratively or tethered to other TFs, e.g., from AP-1 or COUP families, leads to the transcription repression.

Department of Chemistry and Molecular Biology, University of Gothenburg, 405 30 Göteborg, Sweden. email: anna.reymer@gu.se

The GR-DBD-DNA binding is highly sequence-specific. The protein recognises the AGAACA sequence, known as the glucocorticoid response element (GRE).

GR interactions with RNA are much less explored. Several reports describe GR associations with tRNA, mRNA, and Gas5 long noncoding RNA (lncRNA)^{16–18}. It has been proposed that GR binds the GRE-like elements within RNA molecules that act as molecular decoys for the GR-DBD, but details on the mechanism and selectivity remain elusive. An NMR study by Parsonnet et al.⁴, focusing on the mechanism of Gas5 lncRNA-GR interactions in vitro, reported intriguing results. Firstly, GR-DBD interacts with RNA as a monomer and adopts a distinct RNA-bound state, different from the DNA-bound and free states. Secondly, the protein recognises a range of RNA hairpin motifs, both synthetic and biologically derived, but discriminates against uniform dsRNA. And finally, the experiments suggested that GR recognises RNA in a structure-specific and not in sequence-specific manner, as is in the DNA case.

Motivated by these experimental insights, we want to further explore the mechanistic details of the GR-RNA recognition, using all-atom molecular dynamics simulations. As target systems for the GR binding, we employ several RNA structures: a fully paired hairpin, a hairpin with a bulge, and uniform dsRNA; and compare those to the GR interactions with its native DNA target. All studied nucleic acids contain the GRE-site. We observe that the presence of a hairpin and a bulge in an RNA molecule creates the most suitable pocket for the GR binding, which results in a similar interaction energy as is in the DNA-GR monomer case. However, contrary to GRE-DNA-GR recognition, which requires the formation of several specific contacts, the GR binding to RNA is secured by nonspecific electrostatic interactions of long-chain positively charged residues to RNA backbone. Furthermore, our data suggest that it is rather the 3D structural motif of an RNA molecule than its sequence that determines the stable binding by GR.

Methods

Systems preparation

We simulate four different nucleic acids systems where the glucocorticoid receptor (GR) interacts with DNA (CCAGAACAGAGTGTCTGA), dsRNA (GGAGAACAAAAUGUUCUUU), a fully paired RNA hairpin “GRE_RNA” (GGCAAAAUGUUCUUUCGAGAACAUUUGCC) and a bulge-containing hairpin “Gas5” (GGCCCA GUGGUCUUUGUAGACUGCCUGAUGGCC) (Fig. 1), and their unbound counterparts⁴. We derive the 3D structures for the GRE_RNA and Gas5_RNA hairpins through the secondary structure prediction with RNAfold webserver^{30,31} (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) followed by the 3D structure prediction with 3dRNA webserver^{32,33} (<https://bio.tools/3dRNA>). We create the A-form dsRNA structure in USCF Chimera³⁴. To derive GR-RNA complexes, we perform the macromolecular docking in HDOCK³⁵, where GR (PDB ID: 6CFN³⁶) is assigned as the receptor and the RNA molecule—as the ligand. The modelling of Gas5_RNA provides a larger structural variation compared to GRE_RNA (Figure S1), thus we use two models of Gas5_RNA termed, model1 and model2, for the docking. In addition, we exclude helix 4 of GR (Fig. 1) during the docking, to obtain a similar GR-RNA interacting pose as is in the GR-DNA complex, as helix 4 might provide a steric hindrance. We employ this docking setup, as the NMR experiments⁴ imply that GR binds RNA in a similar

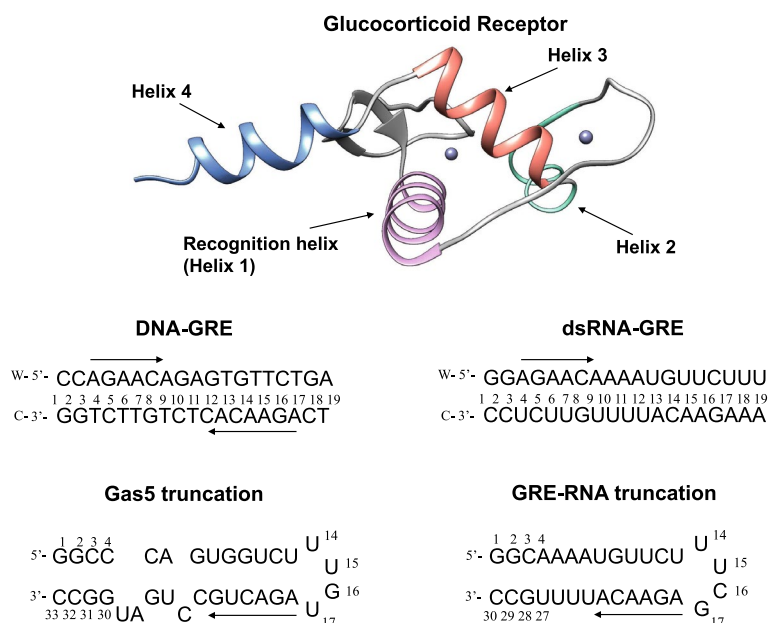


Figure 1. The 3D structure of the glucocorticoid receptor (PDB ID: 6CFN³⁶), with indicated recognition helix, central for the selective DNA binding; and helix 4, implicated in binding of RNA. Below are the 2D structures of the studied nucleic acids systems with their nucleotide sequences. The direction of the GR-response element (GRE) is highlighted with an arrow.

fashion as DNA. The docking generates the GR-RNA complexes matching this criterion among the top-ten scored decoys (Figure S2). Experiments also report that GR binds DNA as a dimer, whereas RNA as a monomer. The dimeric GR-DNA complex is built in USCF Chimera from the GR-DNA complex crystal structure (PDB ID: 5CBX)³⁷, where each GR monomer was exchanged for helix 4 containing GR crystal structure (PDB ID: 6CFN).

Molecular dynamics simulations

We perform all molecular dynamics simulations with the GROMACS MD engine version 2020.5³⁸, using amber14SB³⁹, parmbsc1⁴⁰ and parmbsc0 + chiOL3^{41,42} forcefields for the protein, DNA and RNA, respectively. GR contains two Zn-Cys4 clusters, which we treat with the ZAFF parameters⁴³. We solvate DNA and GR-DNA systems with 15 Å SPC/E⁴⁴ water in cubic boxes and neutralise with K⁺. We add the K⁺ and Cl⁻ ions to reach a physiological concentration of 150 mM KCl. We employ a different setup for all RNA simulated systems, which we solvate with 15 Å TIP3P water⁴⁴ in cubic boxes and neutralise with Na⁺. We add Na⁺ and Cl⁻ ions to reach a concentration of 100 mM NaCl. We treat the monovalent ions with the Joung-Cheatham parameters⁴⁵. We select the KCl instead of NaCl environment for DNA since the conformational transitions such as the sequence-specific bimodality of shift, twist, and other helical parameters, are more accurately captured by the presence of K⁺ ions⁴⁶. Conversely, when it comes to RNA, the NaCl environment is preferred as it provides the greater structural stability⁴⁷. The origin of the greater structural stability lies in the differences in the charge density-related binding modes of the two cations. The smaller hydrated sodium ions preferentially bind to the highly charged phosphates that have high surface area. In contrast, the larger potassium ions can also reside in the grooves. As a result, more cations condense around the phosphate groups in the case of sodium ions, which reduces electrostatic repulsion between the backbones⁴⁸. We proceed with energy minimization with 5000 steps of steepest decent, followed by 500 ps equilibration-runs with weak position restraints on heavy atoms of the solute (1000 kJ/mol) in the NVT and NPT ensembles, adjusting temperature and pressure to 300 K and 1 atm^{49,50}, in periodic boundary conditions. For all RNA systems, both unbound and GR-bound, we perform an additional equilibration run with position restraints on heavy atoms of the solute (1000 kJ/mol) in the NPT ensemble for 20ns⁴⁷. Releasing the restraints, for each of the systems we carry out 600 ns MD simulations at constant pressure and temperature (1 atm and 300 K).

Trajectory analysis

For each of the generated MD trajectories, we discard the first 100 ns as equilibration. To analyse the protein-nucleic acids contacts, we use the CPPTRAJ⁵¹ program from AMBERTOOLS16 software package. We distinguish between intermolecular specific (hydrogen bonding and hydrophobic contacts formed between atoms of protein side chains and nucleic acids bases) and nonspecific contacts (that involve interactions with either DNA/RNA or protein backbones). For every pair of protein-DNA residues we sum up all hydrogen bonds, salt bridges, and hydrophobic (apolar) interactions. We set the contribution of each type of contact to 1. This is done for simplicity, since the energy cost of hydrogen bonds, salt bridges, and hydrophobic interactions varies greatly, depending on the nature of the atoms involved, the bond geometry and the surrounding environment. We consider only the contacts that are present for longer than 10% of the trajectory, see previous publications for further details^{52,53}. Subsequently, we apply Curves+ and Canal programs⁵⁴ to derive the helical parameters, backbone torsional angles and groove geometry parameters of the nucleic acid systems for each trajectory snapshot extracted every ps. For the RNA hairpin systems, we perform the analysis for the fully paired GRE-site region. We use the GROMACS energy tool to calculate the protein-DNA/RNA interaction energies, which operates with the GROMACS force field functional terms i.e., the short range (within a cut-off distance ~ 10 Å) electrostatic (Coulomb) and Lennard-Jones (vdW) interactions for every atom pair within the selected groups are calculated and summed up to obtain the total interaction energy. To separate interaction energies into specific and nonspecific, we calculate interactions for several atom groups. The specific interactions involve the DNA/RNA bases and the protein side chains. While the non-specific interactions involve the DNA/RNA backbone and the protein residues, as well as the protein backbone and the DNA/RNA bases. We also perform free energy analysis and estimate the solvation effects using the MMPBSA/MMGBSA method included in the AMBER CPPTRAJ module. We employ the GROMACS “covar” and “anaeig” tools to calculate the configurational entropies, using two subsets of atoms: the backbone “P” atoms of the entire DNA/RNA sequences and of only the GR-response element. From the obtained eigenvectors, we calculate the configurational entropies at 300 K using the Schlitter’s formula⁵⁵. To derive standard deviations, we calculate the configurational entropies for the entire 500 ns trajectory and for each consecutive 100 ns windows of the trajectory. We perform the principal component analysis (PCA) for every MD trajectory using the GROMACS “covar” tool. We perform PCA for the unbound and the GR-bound systems. For the GR-bound systems, we use two different PCA protocols: (1) for the heavy atoms of the protein-DNA/RNA complexes with the superposition on the nucleic acids; (2) for the heavy atoms of DNA/RNA with the superposition on the nucleic acids.

Additional information

We use MatLab software for the post-processing and plotting of all data and USCF Chimera³⁴ for all molecular graphics.

Results and discussion

In vitro binding studies demonstrated that the GR-DBD (DNA-binding domain) associates with high affinity with Gas5 lncRNA^{16,18}, recognising RNA hairpin motifs within and discriminating against dsRNA⁴. To rationalize the experimental observations and gain further insights into the mechanism of GR association with RNA, we conduct a computational study. As target systems for the GR binding, we employ B-DNA containing the protein native target sequence and several RNA structures: a fully paired hairpin, a hairpin with a bulge, and uniform

dsRNA (Fig. 1). The DNA molecule, 19 b.p. in length, contains two GRE-sites (AGAACA) separated by a three b.p. spacer. Same sequence composition is also selected for the dsRNA molecule, as it is designed to mimic the DNA system. The fully paired RNA hairpin (later 'GRE_RNA'), 30 nucleotides in length, contains the GRE-motif followed by the terminal loop UUCG (5' → 3'). The RNA hairpin with a bulge (later 'Gas5_RNA'), 33 nucleotides in length, is derived from noncoding Gas5 RNA, contains the GRE-like motif (AGACUG) followed by the same terminal loop as is in GRE_RNA.

We derive the models for Gas5_RNA and GRE_RNA through the secondary structure prediction in RNAFold^{30,31} followed by the 3D structure prediction in 3dRNA^{32,33}. We build the A-form dsRNA molecule using USCF Chimera³⁴. To derive the GR-RNA complexes we perform the protein-nucleic acids docking in HDOCK³⁵, specifying GR (PDB ID: 6CFN³⁶) as a "receptor" and RNA as a "ligand". The modelling of the 3D structure of Gas5_RNA provides larger structural variations in the hairpin region compared to GRE_RNA (Figure S1). The differences include the orientation of the terminal hairpin nucleotides and the openness (width) of the Gas5_RNA major groove. Thus, we derive the GR-Gas5_RNA complexes for the two highest scored Gas5_RNA models (later termed model1 and model2). In the two models the orientation of GR within the binding site differs slightly (Figure S3): in model1 the recognition helix of GR is tilted towards the back-face of major groove of Gas5_RNA, while in model2 the recognition helix is more tilted towards the front-face of major groove of Gas5_RNA. The NMR experiments⁴ suggest that GR-DBD binds to RNA in a similar fashion as to DNA. Namely, the chemical shift analysis of 1H–15N HSQC NMR spectra⁴ indicated that GR-DBD interacts with RNA also through the recognition helix (helix 1). Furthermore, the comparison of the chemical shift data for the RNA systems, including Gas5_RNA, GRE_RNA, and dsRNA, indicated similar binding interfaces of GR-DBD, despite different affinities. However, the stoichiometry of the binding differs: GR binds DNA as a dimer, whereas RNA as a monomer. The docking in HDOCK results in GR-Gas5_RNA, GR-GRE_RNA, and GR-dsRNA complexes matching these criteria among the top-ten scored decoys. We build the GR dimer-DNA complex in USCF Chimera from the crystal structure of the complex (PDB ID: 5CBX³⁷) where we exchange each GR monomer with the GR crystal structure (PDB ID: 6CFN) that includes helix 4, which was shown to be important for the GR-RNA association. We subsequently subject the derived five GR-nucleic acids complexes as well as all unbound RNA and DNA systems to 600 ns all-atomistic molecular dynamics (MD) simulations.

We first compare the trajectory-averaged structures for the bound and unbound DNA and RNA molecules, to see if the binding of GR contributes to any significant conformational changes in nucleic acids (Fig. 2A). DNA, GRE_RNA and dsRNA exhibit no major conformational changes, illustrated by the RMSD values of ~ 1 Å. Gas5_RNA shows greater conformational changes upon the binding of GR with the RMSD values of ~ 3 Å, however, within the GRE-site, the changes are small, with the RMSD values of ~ 1 Å. This suggests that the binding site within the nucleic acids systems is predefined by the 3D structure of a particular sequence. Although, the starting orientation of GR-DBD within Gas5_RNA is different for model1 and model2, the two trajectories result in the relatively similar average structures (Figure S3). To identify any major conformational changes in the protein between the different GR-bound DNA/RNA systems, we compare the average complex structures by aligning GR-DBDs (Figure S6). We observe no major conformational differences in GR-DBD between the different systems, except for the unfolding of helix 4 in the GR-Gas5_RNA_model2 case.

We next analyse the evolution of the RMSD values of heavy atoms of DNA/RNA with respect to the average structure along the MD trajectories. As seen in Fig. 2B,C, the binding of GR to DNA and Gas5_RNA significantly reduces the RMSD values fluctuations compared to the unbound state. For GRE_RNA, the fluctuations of the RMSD values are smaller, whereas for dsRNA, the GR binding has no impact on the RMSD behaviour. The decreased fluctuations of the RMSD values, upon the protein binding, suggests a greater surface complementarity of the interacting molecules and the stability of the intermolecular contacts networks in the GR-DNA and GR-Gas5_RNA complexes. We also analyse the evolution of RMSD for the heavy atoms of (1) GR (Figure S4) and (2) the GR-DNA/RNA complexes (Figure S5). For the protein, we see that the RMSD fluctuates below 5 Å in the different systems, where the major fluctuations are due to the movements and changes of helix 4. For the protein-DNA/RNA complexes, we see the most RMSD fluctuations for the GR-dsRNA complex, which are due to both changes in dsRNA and gain and loss of contacts to GR (see PCA analysis).

Protein–nucleic acids contacts

We continue exploring the differences in the GR-DNA and GR-RNA binding mechanisms through the analysis of the protein-DNA/RNA contact networks. For this, we employ our dynamic contacts map approach^{52,53}, where we follow the evolution of the number and strength of specific and nonspecific contacts (Figures S7–S12). By specific contacts we mean the contacts formed between atoms of nucleic acids bases and protein side chains, and by nonspecific contacts—the contacts that involve atoms of either molecule backbones.

In case of the GR–DNA binding, each GR monomer specifically interacts with the GRE-site (AGAACA) employing mainly the residues of the recognition helix: His432, Lys442, Val443, and Arg447 (Fig. 3A). The intermolecular contacts network is entangled, including specific hydrogen bonds and hydrophobic contacts as well as nonspecific electrostatic interactions (Figs. 3A, 4). In detail, for GR monomer 1, His432 interacts with the outer CA step, including the A nucleotide of the response element (C-AGAACA). Lys442 interacts with the AG step (AGAACA); and Val443 and Arg447 form hydrophobic contacts with the GT step and the last T nucleotide, respectively, on the opposite DNA strand (AGAACA/TGTTCT). Arg447 also forms nonspecific electrostatic interactions with DNA backbone. For GR monomer 2, the residues exhibit nearly identical DNA-contacts as the monomer 1, with a few differences: His432 forms no specific contacts with the response element flanking sites while Lys442 exhibits stronger specific contacts. Alanine scanning experiments⁴ provide the contribution order of the positively charged residues of the GR-DBD domain for the DNA binding: R447 > K471 > K480 ≈ R479 > K446. Based on our simulations, we conclude that R447 plays the major role for the GR-DNA binding and recognition:

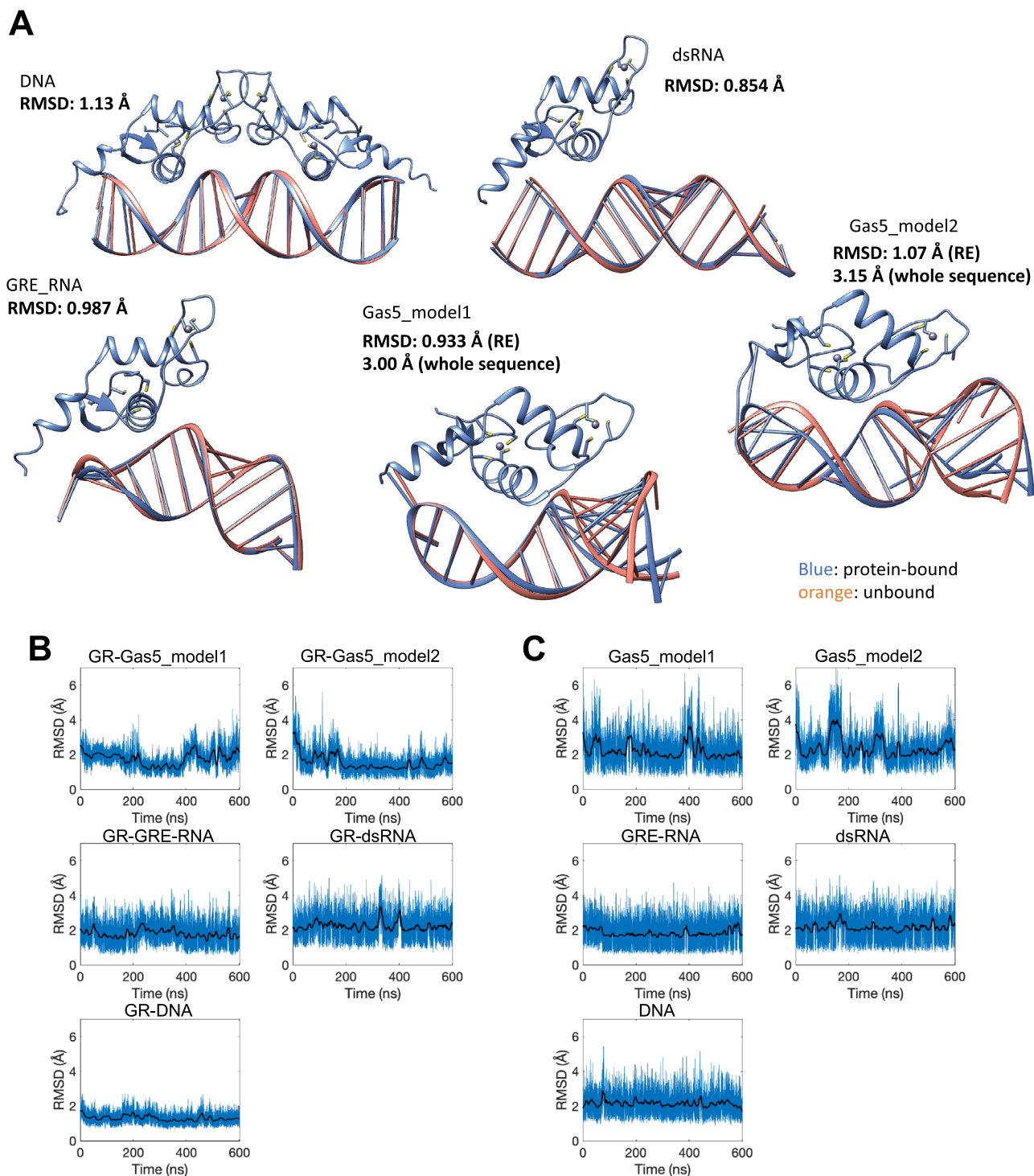


Figure 2. (A) Comparison of the trajectory-averaged structures for the unbound and GR-bound DNA, dsRNA, GRE_RNA and Gas5_RNA systems. Evolution of the RMSD values along the trajectories for (B) GR-bound-DNA/RNA and (C) unbound DNA/RNA. The RMSD values are calculated for heavy atoms of DNA/RNA, with respect to the corresponding average structures.

the residue interacts with the GRE-site through specific and nonspecific contacts. The K471 and K446 residues interact with DNA backbone, while K480 and R479 may play a structural stability role as these are not observed to directly interact with DNA. Comparing to the crystal structure of the GR-DNA complex (PDB ID: 5CBX³⁷), we notice the only defence that Arg447 interacts through hydrogen bonds with the GT step (AGAACA/TGTTCT) instead of making hydrophobic and nonspecific interactions. However, in another crystal structure (PDB ID: 4HN5), Arg477 interacts similarly to what we observe in the MD simulations. This is because Arg477 is a so-called flickering residue which can fluctuate between different conformational substates.

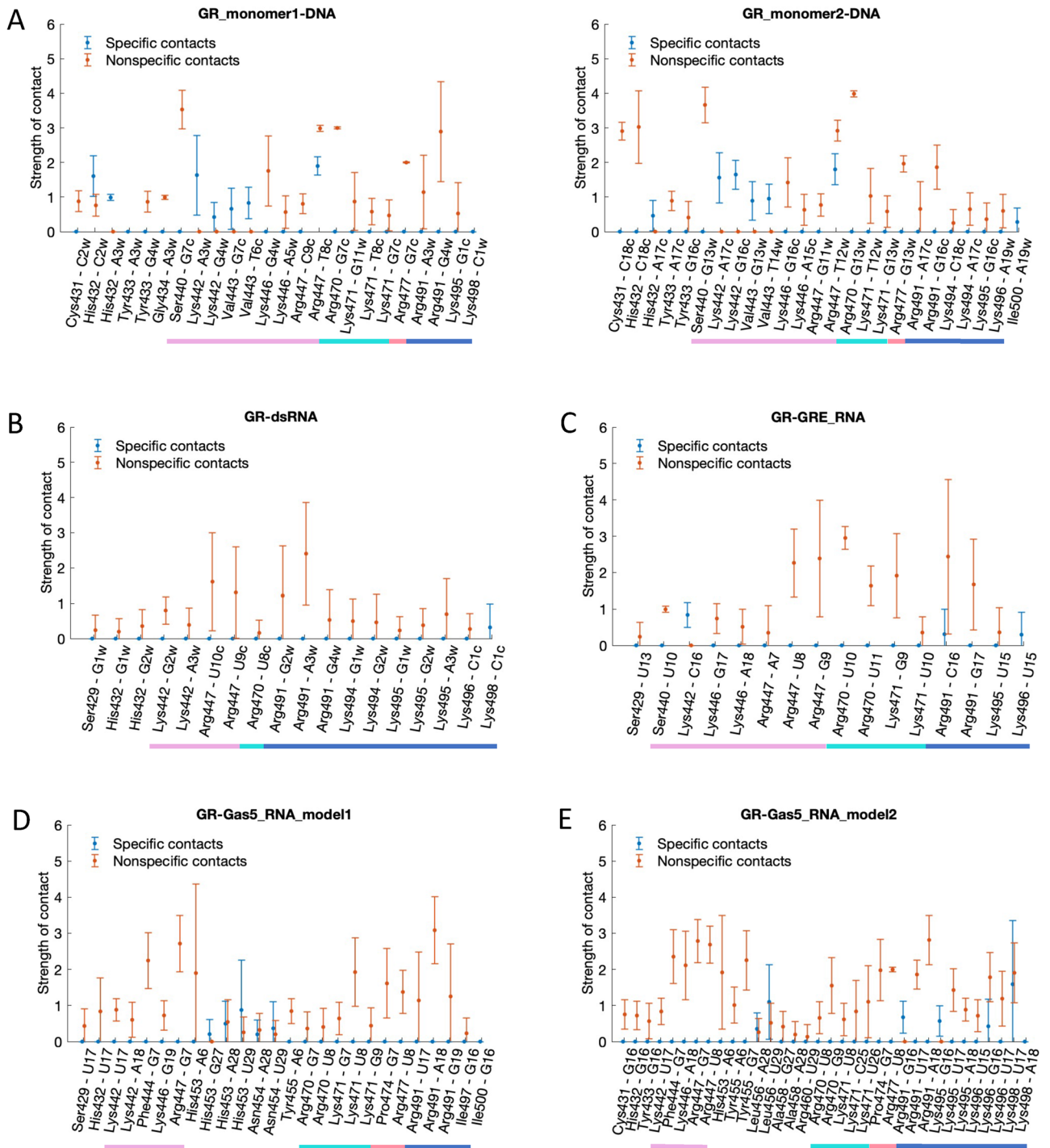


Figure 3. Average contacts strength with standard deviations for specific (blue) and nonspecific (orange) contacts for (A) GR-DNA (B) GR-dsRNA (C) GR-GRE_RNA (D) GR-Gas5_model1 (E) GR-Gas5_model2. The different helices of GR-DBD are denoted with coloured lines: H1: pink, H2: aquamarine, H3: coral, H4: blue (see also Fig. 1). Dots at 0 indicate that this contact is not present as specific (blue) or nonspecific (orange). See details in method section for the definition of a contact strength.

In case of the GR-RNA binding, our dynamic contacts maps show, in agreement with the experiment⁴, the dominating role of the electrostatic nonspecific interactions for the GR-RNA complex formation (Figs. 3B–E, 4). Though, depending on the RNA structure, GR may exhibit some specific contacts as well (Figs. 3B–E, 4). We would like to clarify that by the "specific" interactions we here mean the contacts formed between the RNA bases and the GR sidechains, exchanging a "specifically" interacting RNA base to another would have no or little impact on the binding affinity. Specific GR-RNA contacts involve the protein residues of helix 4 and the flexible nucleotides within the terminal loops of the GRE_RNA and Gas5_RNA hairpins. Additional specific contacts

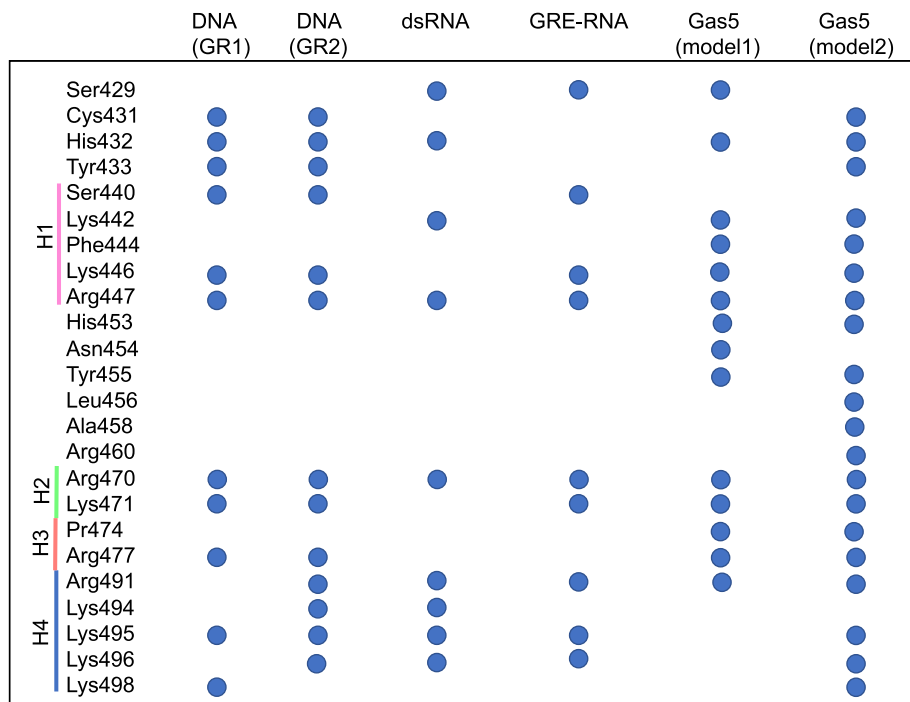


Figure 4. Residues of the glucocorticoid receptor involved in nonspecific interactions with different nucleic acids model systems.

include, for GR-GRE_RNA, the contact by Lys442 of the recognition helix to C16w of the terminal loop, for GR-Gas5_RNA—by His453, Asn454 (Gas5_model1) and Leu456 (Gas5_model2) of the region between the recognition helix and helix 2 to the U28-A29 bulge bases. The latter specific GR-Gas5_RNA contacts become possible due to the presence of the bulge, which allows the RNA molecule to bend toward the protein (see Groove and Helical Parameters for further details).

Overall, we observe that all the residues of the recognition helix, except Val443 (and Lys442 for GRE_RNA) that interact specifically with DNA, form nonspecific contacts with RNA. Also, most of the residues involved in nonspecific contacts with RNA, exhibit nonspecific interactions with DNA (Figs. 3, 4). The only unique nonspecific interactions, seen for the GR-Gas5_RNA system, include the GR residues 453–458 interacting with the Gas5_RNA bulge. Therefore, based on the analysis of the GR-nucleic acids contacts, we do not observe any specific set of GR residues, uniquely involved in the RNA binding. We would like to clarify that our observation is limited to the GR-DBD residues present in the employed crystal structure (PDB ID, 6CFN, a. a. r. 418–500).

Alanine scanning experiments⁴ have revealed a significant role of helix 4 for the GR-RNA associations with the following contribution order: Lys492 > Arg491 > Lys495 > Lys494 > Lys496 > Lys498. Our simulations show a preference for the Arg491-RNA interactions over the Lys492-RNA interactions. Lys492 is observed to electrostatically interact with GRE_RNA in less than 4% of the corresponding MD trajectory snapshots. Our simulations also indicate that it is important for helix 4 to unfold into a random-coil structure to be able to form more electrostatic interactions with RNA. For the GR-Gas5_RNA_model1 system, helix 4 does not unfold due to the nucleotide fraying within the terminal loop; this results in a fewer helix 4-Gas5_RNA interactions. Here we would like to point out that sampling of all conformational substates exploited by flexible random-coil regions require significantly longer simulations, beyond the μ s range, which could explain why we do not see long-lived Lys492-RNA interactions in our simulations. The role of Lys492 could also be to impact the flexibility of helix 4, instead of directly interacting with RNA. The alanine scanning experiments also show some contribution of the residues in helices 1–3 of GR, in the order: R479 \approx K471 \approx K480 > R447 \approx K446 \approx R477. However, this contribution is not of the same magnitude as observed for the GR-DNA recognition. Similar to the DNA-GR system, we observe no contacts with K480 and R479, as they are located at the top of helix 3, pointing towards the solvent. Thus, they may be important for the structure and flexibility of helix 3. We observe nonspecific contacts of K471 and the adjacent residue R470, similar to the GR-DNA system.

Our data cannot directly explain why helix 4 plays a greater role in the RNA recognition, as suggested by the alanine scanning experiments. However, our results allow us to hypothesize that for the RNA recognition, where the recognition helix of GR-DBD cannot reach the nucleobases for specific interactions and forms only nonspecific electrostatic contacts, the electrostatic interactions from helix 4 become an important complementary factor, which allows GR to anchor itself to RNA. Mutating the residues of helix 4 can potentially impact its flexibility, conformation, and interactions with RNA. Similarly, mutating the positively charged residues of the recognition helix will also lower the GR-RNA affinity. However, as these are single mutations, other residues from helices 2 and 3 could compensate keeping GR anchored to RNA.

In terms of the DNA recognition, any mutation that perturb the specific contacts of the recognition helix will significantly affect the GR-DNA affinity. It should be noted that the alanine scanning experiments were performed for GR-Gas5_RNA and GR-DNA systems. The order of residue contributions could be different for the GR complexes with GRE_RNA and dsRNA. Additionally, with longer DNA molecules as found in the cell nucleus, helix 4 may play a more significant role in the recognition through interactions with DNA minor groove. To gain further mechanistic insights into the role of different residues, MD simulations of various GR mutants bound to DNA/RNA would be required.

The results from our simulations point toward an importance of both the recognition helix and helix 4 for the GR-RNA interactions, where the residues within these two regions allow GR to anchor itself within the RNA molecule. Experiments show that GR exhibit poor binding affinity towards dsRNA ($K_d > 3500$ nM). Our simulations show that nonspecific interactions exploited by the recognition helix (Lys442 and Arg470) but also residues outside this region (Ser429, His432 and Arg470) are lost toward the end of the GR-dsRNA MD trajectory (Figure S6). The recognition helix cannot follow the changes of dsRNA major groove geometry (see Groove and Helical Parameters for further details). The only few GR-dsRNA contacts left towards the end of the trajectory are those exploited by unfolded helix 4 (Figure S6).

Protein–nucleic acids interaction energies and configurational entropies

The binding experiments⁴ show the following GR binding preference order: DNA > Gas5_RNA > GRE_RNA > > > dsRNA. To explore if our simulations capture the same trend, we next analyse the intermolecular interaction energies (Figs. 5, S13–S14). Here we would like to point out that the accurate calculation of protein-DNA/RNA interaction energies is challenging due to the massive negatively charged backbone of nucleic acids. Nevertheless, to see if we can capture a qualitative trend, we calculate the interaction energies including electrostatic and van der Waals components, for the specific and nonspecific protein-DNA/RNA contacts along all protein-bound trajectories with GROMACS energy analysis tool. By specific interactions we refer to the contacts formed between the protein side chains and DNA/RNA nucleobases, and by nonspecific interactions we refer to the contacts that involve atoms belonging to either RNA/DNA/protein backbones. The GR-DNA complex shows more favourable interaction energies for the specific contacts, whereas for the nonspecific contacts the energies are similar to those for the GR-RNA complexes. The cumulative interaction energies, going from more negative to less negative energies, show the following trend: GR-Gas5_RNA_model2 > > GR(mon1)-DNA > GR(mon2)-DNA > GR-Gas5_RNA_model1 ≈ GR-GRE_RNA > > GR-dsRNA. The trend is consistent with the one observed experimentally. Our simulations further detail that the unfolding of helix 4 allows for the most favourable GR-Gas5_RNA interaction energies, if compared to the interaction energies of GR-DNA monomeric complexes. However, the GR binding to DNA is dimeric. Thus, combining the protein-DNA interactions energies for both GR monomers, shows a stronger GR preference for DNA over RNA. We would like to emphasize that these are more qualitative estimations of the interaction energies that allows us to look at the trends.

We proceed with the free energy analysis of the GR-DNA/RNA trajectories using the MMPBSA/MMGBSA approach included in the AMBER CPPTRAJ module (Tables S1–S3). The MMPBSA/MMGBSA energies suggest the same binding affinity trend as the above-described analysis of the interaction energies, with one exception. The GR-dsRNA complex shows a more favorable free energy than the GR-GRE_RNA complex, but the estimated free energies of the two systems overlap within the range of standard deviations. The MMPBSA/MMGBSA approach also allows us to estimate the solvation effects. The highly positive values of the solvation free energies highlight the favorable GR-DNA/RNA complexation (Table S2). The GR-DNA dimer has a significantly higher solvation free energy. However, if we consider the GR-DNA monomeric complexes separately, the solvation free energies show the following trend, going from more positive to less positive energies: Gas5_RNA_model2 > > GR(mon1)-DNA ≈ GR(mon2)-DNA > > GR-Gas5_RNA_model1 ≈ GR-dsRNA ≈ GR-GRE_RNA. It must be noted that the differences in solvation free energies, especially for the latter three systems, are not statistically significant. Dividing the solvation free energies into polar and non-polar terms (Table S3) emphasizes the dominating role of electrostatic interactions between the solute and the solvent. However, caution must be

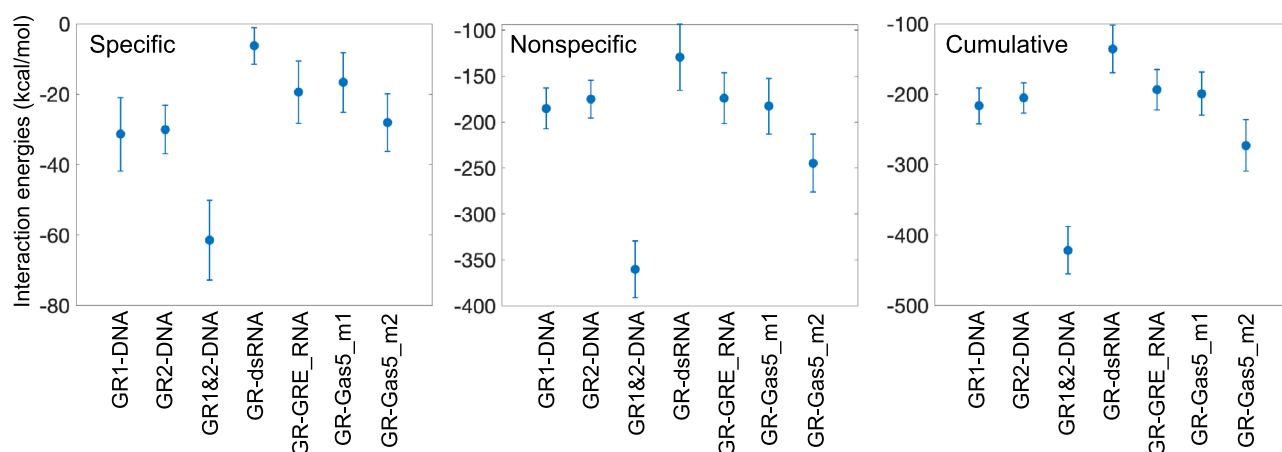


Figure 5. Protein-DNA/RNA interaction energies (kcal/mol), calculated for all GR-DNA/RNA complexes.

exercised when assessing the non-polar solvation energies, as they are approximated by a linear relation to the solvent accessible surface area (SASA), which may underestimate the true impact of the hydrophobic effects on the complexation⁵⁶.

We also calculate the configurational entropies for all GR-nucleic acids systems to explore further the thermodynamics of the binding process. The configurational entropy is a part of the total entropy change, which arises from the degrees of freedom of the solute. NMR experiments have shown that for proteins the configurational entropy and the solvent entropy contributions can be similar in magnitude^{57,58}, and thus can impact the thermodynamics of protein interactions. We calculate the configurational entropies using the Schlitter's formula⁵⁹ in the presence and absence of GR for the entire DNA/RNA sequence and only for the response element (RE). To estimate the standard deviations, we calculate the configurational entropies for the entire 500 ns trajectory and for all consecutive 100 ns windows within. Upon the GR binding to DNA/RNA, we expect the configurational entropy to decrease if the two molecules are structurally complementary and form a stable complex. As if the protein makes the nucleic acid molecule more rigid for global motions. Looking at the configurational entropies for the entire sequences (Table 1), we see that this statement is true for GR-bound DNA and both Gas5_RNA systems, where the configurational entropies decrease by ~ 15, 5 and 7.5 kcal/mol, respectively. In the case of GRE_RNA, the decrease is insignificant, 0.37 kcal/mol. While, for GR-bound dsRNA the configurational entropy remains approximately the same (increased by 0.014 kcal/mol). Similar situation occurs through the analysis of the configurational entropies of the RE-sites, where the decrease in the configurational entropy upon the GR binding to dsRNA is ~ 10–100 times smaller than in other systems. These results further support that the GR-dsRNA complex is less stable, and the protein cannot adjust to the binding site on dsRNA.

Nucleic acids helical and groove parameters

Next, we analyse changes in helical and groove parameters upon the GR binding for all studied nucleic acids systems (Figs. 6, 7 and Figs. S15–S23). We have previously shown that local changes in DNA groove geometries and helical parameters, mainly in shift and slide, facilitate the direct-read-out mechanism^{53,60}; that is the ability of proteins to form specific contacts with their genetic sites. The binding of GR to DNA, as expected, brings in alterations in shift, slide and twist within the GRE-sequence and the flanking sites (Figs. 6, S15). Additionally, the GR binding also leads to small changes in roll and tilt angles (Figs. 6, S15), due to the hydrophobic interactions of the bulky Val443 residue and the interactions of helix 4 with DNA minor groove. Furthermore, the GR monomers make DNA major groove wider within the GRE-sites, and narrower within the linking region between the two monomers (Figs. 6, S15). Several b.ps. within the GRE-site shift towards the major groove, as illustrated by changes in x-displacement (Figs. 6, S19), making the major groove shallower and more accessible for specific interactions with the protein.

In contrast to DNA, the binding of GR to the RNA systems induces no notable changes in helical parameters (Figs. S16–S18, S20–S22). This is not surprising as the helical parameters distributions appear monomodal within the GRE-site for all unbound RNA, which means that the RNA b.ps. are quite rigid and the GR residues cannot reach them to form specific contacts. Also, the values of helical parameters for the GR-bound RNA molecules (except rise) are distinct from those of GR-bound DNA (Fig. S23). Nevertheless, we observe that the presence of a hairpin and a bulge result in different values of helical parameters for GR-bound RNA, within the GRE-site (Fig. 7). Most notably, for GR-bound Gas5_RNA the last 3 b.ps. of the GRE-site (AGACUG), adjacent to the bulge, have the least difference in twist, shift, tilt, and x-displacement values (Figs. 7, S23) with respect to the same parameters of DNA. For GR-bound GRE_RNA, the only significant difference from dsRNA is in the higher values of b.ps. roll angle.

The GR binding also has little impact on the RNA systems grooves geometries. The RNA grooves geometries differ depending on the structural motifs of the RNA systems (Figs. S20–S22). When comparing the major groove width within the GRE-site across all studied GR-bound nucleic acids systems, we notice that the first 3

	Whole sequence (kcal/mol)	Response element (kcal/mol)
DNA	67.41 ± 1.37	13.48 ± 0.364 (RE1) 12.65 ± 0.370 (RE2)
GR-DNA	52.66 ± 1.83 (- 14.75)	10.03 ± 0.521 (RE1) 10.07 ± 0.771 (RE2) (- 3.45) (RE1) (- 2.58) (RE2)
dsRNA	51.98 ± 0.502	9.604 ± 0.0551
GR-dsRNA	52.00 ± 0.560 (0.0138)	9.573 ± 0.109 (- 0.0306)
GRE_RNA	37.45 ± 0.234	9.999 ± 0.0527
GR-GRE_RNA	37.08 ± 0.775 (- 0.367)	9.362 ± 0.394 (- 0.637)
Gas5_model1	46.38 ± 0.774	9.813 ± 0.116
GR-Gas5_model1	41.41 ± 1.36 (- 4.96)	9.440 ± 0.228 (- 0.372)
Gas5_model2	47.16 ± 2.36	10.16 ± 0.314
GR-Gas5_model2	39.63 ± 1.73 (- 7.54)	8.656 ± 0.299 (- 1.50)

Table 1. Configurational entropies [TS (kcal/mol), for T = 300K] for the DNA/RNA backbone P atoms. Entropies have been derived for the windows: last 500 ns, 100–200 ns, 200–300 ns, 300–400 ns and 400–500 ns to obtain the standard deviations. Values in parenthesis are the differences in mean values with respect to the naked nucleic acid molecule. RE1 and RE2 are the two response elements presence on the DNA molecule.

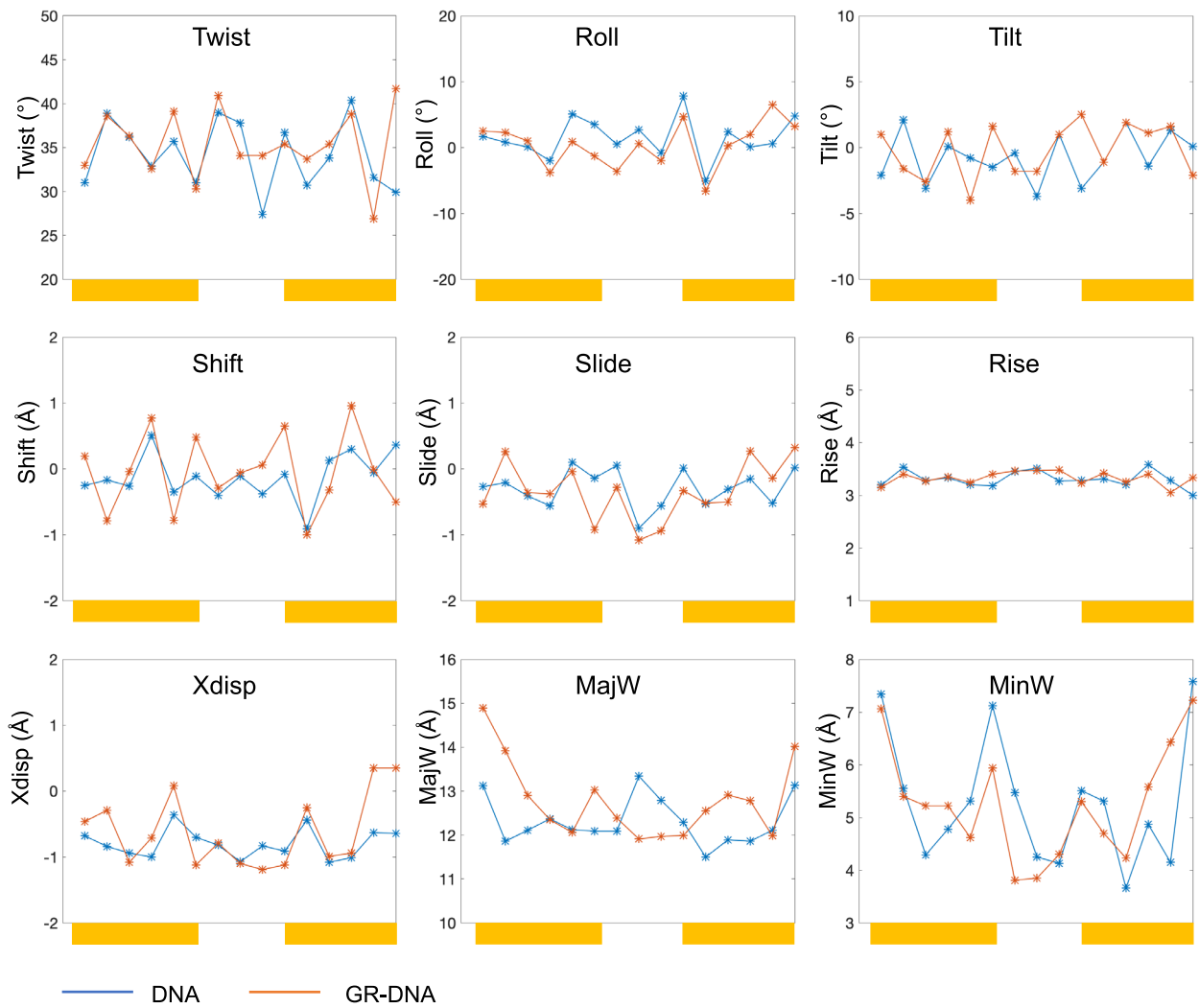


Figure 6. Average helical and groove parameters for unbound (blue) and GR-bound (orange) DNA. The binding sites for the two GR monomers are highlighted with yellow rectangular boxes.

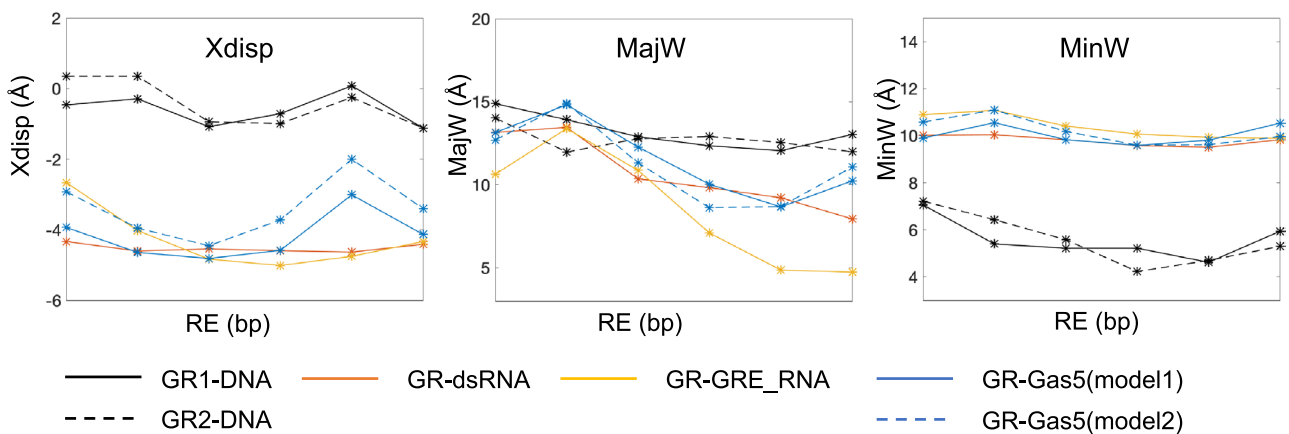


Figure 7. Comparison of average values for x-displacement and groove widths in different nucleic acids model systems for the b.p.s. within the GRE-site for the GR-bound systems.

b.ps. exhibit comparable values (Fig. 7), which suggest a suitable binding pocket for the protein initial binding. For GRE_RNA and Gas5_RNA, by the first 3 b.ps. we refer to the ones on the hairpin side. However, it should be noted that the major groove width distributions for dsRNA are broader (Figs. S19–S22), which indicates larger structural fluctuations, which contribute to the instability of the GR-dsRNA contacts network and thus low binding affinity. The major groove widths of the remaining 3 b.ps of the GRE-site differ in the RNA systems, with Gas5_RNA being the closest to DNA, where we refer to the Gas5_RNA GRE-half site adjacent to the bulge. b.ps. The conformational similarities between Gas5_RNA and DNA, we believe, contribute to the more favourable protein-Gas5_RNA contacts network, interaction energies and configurational entropy changes among the studied RNA systems.

Principal component analysis

We perform principal component analysis (PCA)^{61,62} of the trajectories to see if there are any predominant molecular motions that could explain the GR preference for DNA and hairpin RNA over dsRNA. First, we do the analysis for the unbound nucleic acids systems. The most dominant motions in the MD trajectories are represented within the first three principal components, which accounts for ~60–65% of the total variance (Table S4, Fig. S24A).

For DNA, dsRNA, and GRE_RNA the first three components describe similar motions, where the first two components correspond to sideward and upward bending, and the third component describes the twisting rotation (Fig. S25A–C, Movies S1–S9). The three motions appear to impact the open state of the major groove. However, the structural dynamics of the GRE-site binding pocket differs in the three systems. In DNA the shape fluctuations of the GRE-site binding pocket are small, whereas in dsRNA the pocket becomes extremely deep and narrow. In GRE_RNA the binding pocket appears smaller and tilted to one side of the major groove, explaining the observed asymmetry of the GR binding.

For Gas5_RNA systems the motion described by the first component dominates. It involves a bending motion that changes the size of the major groove pocket, going from an open to a closed and bent conformation (Fig. S25D–E, Movies S10 and S13), where the middle state resembles the averaged structure of the GR-bound state. The motions described by the second and third components differ between the Gas5_RNA_model1 and model2. In model1, they correspond to a rotation and a downwards bending motions. Contrary, in model2 the second and third components describe a rotational bending motion where several terminal loop residues rearrange to make the binding pocket more open (Fig. S25D–E, Movies S11–S12, S14–S15).

We continue with the PCA analysis of the trajectories for the protein-bound systems (Figs. S24B–D, S26–S27). We do the analysis in two ways; (1) for the heavy atoms of the protein-DNA/RNA systems and (2) for heavy atoms of DNA/RNA with the superposition on the nucleic acids in both cases. The two analyses are complementary: (1) highlights the GR movements upon changes in the RNA/DNA structures; and (2)—conformational changes in RNA/DNA upon the protein binding.

For all protein-bound systems, the first three principal components from the two analyses account for ~50–70% of the total variance (Table S4, Fig. S24B). For the GR-DNA system, analysis (1) reveals that both GR monomers are stably bound to DNA. The GR dimer makes DNA more rigid, and the three principal components correspond to small sideward bending motions (Fig. S26A, Movies S16–S18), which the GR dimer follows. The largest amplitude movements are seen for helix 4 of GR, which correspond to swinging out of DNA minor groove. Analysis (2) shows motions described by the first and third components similar to those of unbound DNA, whereas the second component shows an expansion of DNA major groove as a result of the GR dimer binding (Fig. S27A, Movies S19–S21).

For GR-dsRNA system (Fig. S26B, Movies S22–S24), analysis (1) reveals a translational movement of the protein: the first component describes a sidwards bending of dsRNA, which closes the major groove and makes GR tilt and shift from one side of the GRE-site to the other. The second component involves the opening of the major groove, which allows GR to sink into to the GRE-pocket, but not as deeply as is in the GR-DNA complex. The third component describes a transition to a closed major groove state through bending and the loss of contacts of the GR recognition helix with dsRNA. Analysis (2) (Fig. S27B, Movies S25–S27) results in the first two components describing a downward and sideward bending that impacts the open state of the major groove, which forces GR to shift to remain bound to dsRNA. The third component corresponds to the movement, similar to that of the first two components of analysis (1). Comparing the motions from PCA analyses (1) and (2), and the ones of the unbound state, we conclude that the protein stimulates the opening of the dsRNA major groove. However, the dsRNA molecule cannot maintain a suitable width of the binding pocket for the GR monomer, resulting in the protein dissociation.

For the GR-GRE_RNA and both GR-Gas5_RNA systems (Figs. S26–S27C–E, Movies S28–S30, S34–S36, S40–S42), similarly to GR-DNA, analysis (1) shows the protein stably bound to RNA. However, we observe different predominant motions in the three RNA hairpin systems, and as a result a different structural response of GR. For GRE_RNA, we observe breathing upward and sideward bending motions that impact the compactness of the GRE-site, leading to the rocking and tilting motions of the recognition helix and helix 4. For Gas5_RNA_model1, the motions described by the first three principal components resemble those for the GRE_RNA case, but the amplitude is bigger suggesting a more open and flexible structure. The GR protein adjusts by tilting and turning of its recognition helix to follow the changes in the shape of the binding pocket. The motions of the Gas5_RNA_model2 also include sidwards bending, but in this case, additionally we observe the expansion of the minor groove adjacent to the GRE-site. This motion allows residues from GR helix 2 to approach the groove and interact with the bulge region, leading to a forward tilting movement of the GR recognition helix. Uniquely, for Gas5_RNA_model2, the sideward bending, captured by the component 3 of analysis (1) (Figs. S26–S27, Movie S42), also involves movements of the terminal loop and bulge residues. For the GR-GRE_RNA and both

GR-Gas5_RNA systems, analysis (2) provides the same principal components as analysis (1); but with a slightly larger amplitude of motions (Figs. S26-S27, Movies S31-S33, S37-S39, S43-S45).

Overall, the PCA analysis confirms that to anchor within the GRE-binding site, the protein requires a wide and shallow major groove pocket. The protein effectively moulds the GRE-site within DNA major groove but has little effect on the major groove geometry of the RNA systems. Furthermore, to remain bound the protein needs to adjust to the structural dynamics of RNA through changes in the orientation of the recognition helix, which in turn requires a more open conformation of the GRE-site supported by the terminal hairpin. Addition of the bulge, in Gas5_RNA, increases the kinking motions of RNA and provides an additional interaction surface for GR helix 2.

Conclusions

Using all-atomistic MD simulations, we explore the mechanistic details of transcription factor associations with RNA. As the model system we employ the glucocorticoid receptor (GR) bound to several nucleic acids model systems, which include the native DNA target and three structurally different RNA molecules containing the glucocorticoid response element (GRE) and GRE-like sequences. The studied RNA systems include dsRNA, a fully paired RNA hairpin, and an RNA hairpin with a bulge. The latter system is derived from long-noncoding Gas5 RNA, which was shown to interact with GR *in vivo*^{16,18}. Through the analyses of the protein-DNA/RNA contacts networks, interaction energies, configurational entropies, nucleic acids helical and groove parameters, and principal component analysis, we conclude that GR DNA-binding domain (DBD) can bind both DNA and RNA, but the association mechanisms are different.

The DNA recognition by a GR dimer involves formation of several specific contacts between the residues of the GR-DBD recognition helices and the bases of the GRE-sites. These specific contacts anchor each recognition helix deep within DNA major groove, leading to the significant widening of the major groove of both GRE-sites (Movie S20). It is exactly one DNA turn that separates the centers of two GRE-sites, which allows for the strongest allosteric communication between the monomers of the GR dimer^{63,64}, explaining the observed positive cooperativity for the GR-DNA association^{4,28}.

Contrary, the GR-DBD-RNA complexation depends predominantly on nonspecific electrostatic contacts. Though, the number and strength of these nonspecific contacts depend on the structural motif of the RNA molecule (Figs. 3, S3-S6). A GR-complex with uniform dsRNA, with its characteristic A-form double helix, shows fewer contacts, mostly from the residues of helix 4. The GR-dsRNA MD trajectory shows a gradual loss of protein-RNA contacts and the beginning of the protein dissociation. Contrary, a hairpin motif allows RNA to adapt a more open conformation with a rather wide and shallow major groove pocket, allowing for more GR-RNA contacts to be formed. GR interacts with the Gas5_RNA and GRE_RNA hairpins through contacts formed by the positively charged residues of the recognition helix and helix 4. Addition of the internal loop (bulge) in Gas5_RNA increases flexibility of the off-GRE stem, which contributes with additional contacts formed by the residues adjacent to the GR dimerization motif.

The binding energies and the configurational entropies analyses of the GR-DNA/RNA complexes show the same binding affinities relations, as the binding experiments⁴: DNA > Gas5_RNA > GRE_RNA > > dsRNA. In addition, computationally we can estimate the binding energies for each GR monomer in the GR-DNA complex, which further reveals that a GR monomer can bind Gas5_RNA even stronger than DNA. This is possible due to the combination of two factors. The first being the above-mentioned GR-Gas5_RNA nonspecific interactions, because of the increased flexibility of the off-GRE stem. The second being the unfolding of helix 4, the residues of which then interact specifically with the bases of the terminal loop region. Crystal structures support that helix 4 can exist as both folded^{36,65} and unfolded^{28,66}. In our computational experiments, we observe the unfolding of helix 4 for the GR-DNA, GR-GRE_RNA, and GR-Gas5_RNA_model2 systems. Though, we believe that the unfolding of helix 4 in the GR-DNA MD trajectory is the consequence of GRE-flanking regions being too short. The unfolding of helix 4 may present a mechanism for the GR-nucleic acids structural recognition—a hypothesis that we plan to investigate further in the future.

Taken together, we demonstrate with atomic level detail that the glucocorticoid receptor employs the sequence recognition mechanism when binding DNA and the shape recognition mechanism when binding RNA. The protein interacts with either DNA or RNA using nearly same residues, yet the nature of the formed protein-nucleic acids contacts and the binding interfaces differ. We believe that the described dual recognition mechanism may be shared by other zinc finger transcription factors.

The ability of the glucocorticoid receptor to selectively interact with both DNA and RNA, demonstrating comparable binding affinities within the nanomolar range (10^{-8} to 10^{-7} M)^{4,18,28}, carries significant biological implications. This phenomenon might constitute a finely tuned evolutionary mechanism that enables organisms to promptly respond to conditions such as low blood-glucose levels and various stress conditions. Studies indicate that upon activation, GR is rapidly recruited to transcribed sites⁶⁷. This could be attributed to tethered RNA transcripts functioning as molecular beacons, actively attracting the GR protein. Moreover, RNA transcripts could serve as reservoirs, enabling the protein to maintain heightened local concentrations to sustain transcriptional activity. The competitive binding of GR-DBD to DNA and RNA may play a pivotal role in modulating DNA-GR equilibrium binding constants, potentially leading to the dynamic regulation of gene transcription rates. The DNA/RNA binding duality offers a mechanism for the protein to dissociate from high-affinity GRE-sites. Indeed, experimental studies underscore that GR exhibits a more rapid turnover on DNA compared to other transcription factors^{68,69}. However, further characterization is imperative to establish the *in vivo* role of the glucocorticoid receptor-nucleic acids binding duality.

Data availability

Generated datasets from analyses of MD trajectories, including protein–nucleic acids contacts information, interaction energies, helical parameters of unbound and protein–bound nucleic acids; also starting, last snapshot, and average structures for unbound and protein–bound nucleic acids are freely accessible at zenodo.org as <https://doi.org/10.5281/zenodo.7738133>.

Received: 3 March 2023; Accepted: 21 September 2023

Published online: 26 September 2023

References

- Lambert, S. A. *et al.* The human transcription factors. *Cell* **172**, 650–665 (2018).
- Cassiday, L. A. Having it both ways: Transcription factors that bind DNA and RNA. *Nucleic Acids Res.* **30**, 4118–4126 (2002).
- Hudson, W. H. & Ortlund, E. A. The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **15**, 749–760 (2014).
- Parsonnet, N. V., Lammer, N. C., Holmes, Z. E., Batey, R. T. & Wuttke, D. S. The glucocorticoid receptor DNA-binding domain recognizes RNA hairpin structures with high affinity. *Nucleic Acids Res.* **47**, 8180–8192 (2019).
- Asimi, V. *et al.* Hijacking of transcriptional condensates by endogenous retroviruses. *Nat. Genet.* **54**, 1238–1247 (2022).
- Kaikkonen, M. U. & Adelman, K. Emerging roles of non-coding RNA transcription. *Trends Biochem. Sci.* **43**, 654–667 (2018).
- Seila, A. C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
- Sharp, P. A., Chakraborty, A. K., Henninger, J. E. & Young, R. A. RNA in formation and regulation of transcriptional condensates. *RNA* **28**, 52–57 (2022).
- Henninger, J. E. *et al.* RNA-mediated feedback control of transcriptional condensates. *Cell* **184**, 207–225.e24 (2021).
- Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–629 (2011).
- Sigova, A. A. *et al.* Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978–981 (2015).
- Shamovsky, I., Ivannikov, M., Kandel, E. S., Gershon, D. & Nudler, E. RNA-mediated response to heat shock in mammalian cells. *Nature* **440**, 556–560 (2006).
- Oksuz, O. *et al.* Transcription factors interact with RNA to regulate genes. *Mol. Cell* **83**, 2449–2463.e13 (2023).
- Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **8**, 479–490 (2007).
- Corley, M., Burns, M. C. & Yeo, G. W. How RNA-binding proteins interact with RNA: Molecules and mechanisms. *Mol. Cell* **78**, 9–29 (2020).
- Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA Gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* **3** (2010).
- Park, O. H., Do, E. & Kim, Y. K. A new function of glucocorticoid receptor: regulation of mRNA stability. *BMB Rep.* **48**, 367–368 (2015).
- Hudson, W. H. *et al.* Conserved sequence-specific lincRNA–steroid receptor interactions drive transcriptional repression and direct cell fate. *Nat. Commun.* **5**, 5395 (2014).
- 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).
- Vockley, C. M. *et al.* Direct GR binding sites potentiate clusters of TF binding across the human genome. *Cell* **166**, 1269–1281.e19 (2016).
- So, A. Y. L., Cooper, S. B., Feldman, B. J., Manuchehri, M. & Yamamoto, K. R. Conservation analysis predicts in vivo occupancy of glucocorticoid receptor-binding sequences at glucocorticoid-induced genes. *Proc. Natl. Acad. Sci. USA* **105**, 5745–5749 (2008).
- De Bosscher, K., Beck, I. M., Ratman, D., Berghe, W. V. & Libert, C. Activation of the glucocorticoid receptor in acute inflammation: The SEDIGRAM concept. *Trends Pharmacol. Sci.* **37**, 4–16 (2016).
- Pufall, M. A. Glucocorticoids and cancer. *Adv. Exp. Med. Biol.* **872**, 315–333 (2015).
- Kadmiel, M. & Cidlowski, J. A. Glucocorticoid receptor signaling in health and disease. *Trends Pharmacol. Sci.* **34**, 518–530 (2013).
- Simons, S. S., Edwards, D. P. & Kumar, R. Minireview: Dynamic structures of nuclear hormone receptors: New promises and challenges. *Mol. Endocrinol.* **28**, 173–182 (2014).
- Lim, H.-W. *et al.* Genomic redistribution of GR monomers and dimers mediates transcriptional response to exogenous glucocorticoid in vivo. *Genome Res.* **25**, 836–844 (2015).
- Starick, S. R. *et al.* ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors. *Genome Res.* **25**, 825–835 (2015).
- Hudson, W. H., Youn, C. & Ortlund, E. A. The structural basis of direct glucocorticoid-mediated transrepression. *Nat. Struct. Mol. Biol.* **20**, 53–58 (2013).
- Schiller, B. J., Chodankar, R., Watson, L. C., Stallcup, M. R. & Yamamoto, K. R. Glucocorticoid receptor binds half sites as a monomer and regulates specific target genes. *Genome Biol.* **15**, 418 (2014).
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).
- Lorenz, R. *et al.* ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
- Zhang, Y., Wang, J. & Xiao, Y. 3dRNA: 3D structure prediction from linear to circular RNAs. *J. Mol. Biol.* **434**, 167452 (2022).
- Zhang, Y., Wang, J. & Xiao, Y. 3dRNA: Building RNA 3D structure with improved template library. *Comput. Struct. Biotechnol. J.* **18**, 2416–2423 (2020).
- Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S.-Y. HDock: A web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **45**, W365–W373 (2017).
- Frank, F., Okafor, C. D. & Ortlund, E. A. The first crystal structure of a DNA-free nuclear receptor DNA binding domain sheds light on DNA-driven allostery in the glucocorticoid receptor. *Sci. Rep.* **8**, 13497 (2018).
- Hudson, W. H. *et al.* Distal substitutions drive divergent DNA specificity among paralogous transcription factors through subdivision of conformational space. *Proc. Natl. Acad. Sci.* **113**, 326–331 (2016).
- Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
- Maier, J. A. *et al.* ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
- Ivani, I. *et al.* Parmbsc1: A refined force field for DNA simulations. *Nat. Methods* **13**, 55–58 (2016).

41. Pérez, A. *et al.* Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **92**, 3817–3829 (2007).
42. Zgarbová, M. *et al.* Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 (2011).
43. Peters, M. B. *et al.* Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *J. Chem. Theory Comput.* **6**, 2935–2947 (2010).
44. Mark, P. & Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960 (2001).
45. Joung, I. S. & Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
46. Dans, P. D. *et al.* The static and dynamic structural heterogeneities of B-DNA: extending Calladine–Dickerson rules. *Nucleic Acids Res.* **47**, 11090–11102 (2019).
47. Reißer, S., Zucchelli, S., Gustincich, S. & Bussi, G. Conformational ensembles of an RNA hairpin using molecular dynamics and sparse NMR data. *Nucleic Acids Res.* **48**, 1164–1174 (2020).
48. Henning-Knechtel, A., Thirumalai, D. & Kirmizialtin, S. Differences in ion–RNA binding modes due to charge density variations explain the stability of RNA in monovalent salts. *Sci. Adv.* **8** (2022).
49. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
50. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
51. 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).
52. Hörberg, J. & Reymer, A. Specifically bound BZIP transcription factors modulate DNA supercoiling transitions. *Sci. Rep.* **10**, 18795 (2020).
53. Hörberg, J., Moreau, K., Tamás, M. J. & Reymer, A. Sequence-specific dynamics of DNA response elements and their flanking sites regulate the recognition by AP-1 transcription factors. *Nucleic Acids Res.* **49**, 9280–9293 (2021).
54. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37**, 5917–5927 (2009).
55. Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **215**, 617–621 (1993).
56. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449–461 (2015).
57. Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G. & Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nat. Chem. Biol.* **6**, 352–358 (2010).
58. Tzeng, S.-R. & Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **488**, 236–240 (2012).
59. Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M. & Laughton, C. A. Cooperativity in drug–DNA recognition: A molecular dynamics study. *J. Am. Chem. Soc.* **123**, 12658–12663 (2001).
60. Hörberg, J., Hallbäck, B., Moreau, K. & Reymer, A. Abnormal methylation in the NDUFA13 gene promoter of breast cancer cells breaks the cooperative DNA recognition by transcription factors. *QRB Discov.* **3**, e23 (2022).
61. Maisuradze, G. G., Liwo, A. & Scheraga, H. A. principal component analysis for protein folding dynamics. *J. Mol. Biol.* **385**, 312–329 (2009).
62. David, C. C. & Jacobs, D. J. Principal component analysis: A method for determining the essential dynamics of proteins. in 193–226 (2014). https://doi.org/10.1007/978-1-62703-658-0_11.
63. Kim, S. *et al.* Probing allostery through DNA. *Science* **339**, 816–819 (2013).
64. Balaceanu, A., Pérez, A., Dans, P. D. & Orozco, M. Allostery and signal transfer in DNA. *Nucleic Acids Res.* **46**, 7554–7565 (2018).
65. Meijnsing, S. H. *et al.* DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324**, 407–410 (2009).
66. Luisi, B. F. *et al.* Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505 (1991).
67. McDowell, I. C. *et al.* Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding. *Genome Res.* **28**, 1272–1284 (2018).
68. McNally, J. G., Müller, W. G., Walker, D., Wolford, R. & Hager, G. L. The glucocorticoid receptor: Rapid exchange with regulatory sites in living cells. *Science* **287**, 1262–1265 (2000).
69. Stavreva, D. A., Müller, W. G., Hager, G. L., Smith, C. L. & McNally, J. G. Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Mol. Cell. Biol.* **24**, 2682–2697 (2004).

Acknowledgements

The authors would like to dedicate this work to late Dr. Matti Lepistö, who laid the foundation for this project. The authors thank Prof. Deborah S. Wuttke and Prof. Robert T. Batey for insightful discussions, and Swedish National Infrastructure for Computing (SNIC) for the generous provision of computing resources.

Author contributions

J.H. and A.R. conceived and designed the study. J.H. performed the MD simulations. J.H. and A.R. analysed the data. J.H. and A.R. wrote the article.

Funding

Open access funding provided by University of Gothenburg. Swedish Foundation for Strategic Research SSF Grant [ITM170431] and Magn. Bergvalls Foundation Grant to A.R.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-43244-1>.

Correspondence and requests for materials should be addressed to A.R.

Reprints and permissions information is available at www.nature.com/reprints.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023