PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

# The marionette mechanism of domain–domain communication in the antagonist, agonist, and coactivator responses of the estrogen receptor

Xun Chen[a,b,1], Shikai Jin[a,c,1], Mingchen Chen[a,d,1], Carlos Bueno[a,e], and Peter G. Wolynes[a,b,c,e,2]

Contributed by Peter G. Wolynes; received October 3, 2022; accepted December 28, 2022; reviewed by John A. Katzenellenbogen, Yaakov Levy, and Zaida Luthey-Schulten

**The human estrogen receptor α (hERα) is involved in the regulation of growth, development, and tissue homeostasis. Agonists that bind to the receptor's ligand-binding domain (LBD) lead to recruitment of coactivators and the enhancement of gene expression. In contrast, antagonists bind to the LBD and block the binding of coactivators thus decreasing gene expressions. In this work, we carry out simulations using the AWSEM (Associative memory, Water mediated, Structure and Energy Model)-Suite force field along with the 3SPN.2C force field for DNA to predict the structure of hERα and study its dynamics when binding to DNA and coactivators. Using simulations of antagonist-bound hERα and agonist-bound hERα by themselves and also along with bound DNA and coactivators, principal component analyses and free energy landscape analyses capture the pathway of domain–domain communication for agonist-bound hERα. This communication is mediated through the hinge domains that are ordinarily intrinsically disordered. These disordered segments manipulate the hinge domains much like the strings of a marionette as they twist in different ways when antagonists or agonists are bound to the ligand-binding domain.**

breast cancer | estrogen receptor | DNA binding | domain–domain communication | marionette mechanism

## Significance

The status of estrogen receptors is used as an indicator of the proper approach to the clinical management of breast cancer. Knowledge of the structure of full-length estrogen receptors and their mechanisms of conformational change when binding with different ligands and peptides can help to design therapies. In this work, we use coarse-grained simulations that employ the AWSEM-3SPN.2C forcefield to predict the structure and dynamics of the human estrogen receptor α (hERα) when bound to ligands and to DNA. The flexible regions of the receptors undergo a collective motion, much like the dancing of a marionette that amplifies domain–ligand conformational motions so as to communicate ligand binding to the DNA-binding domain in different ways when antagonists or agonists bind.

Estrogen receptors (ERs) belong to the nuclear receptor superfamily of transcription factors that, in response to binding to small molecule hormonal signals, initiate diverse molecular events culminating in the activation or repression of genes (1–3). Estrogen receptors dimerize when they encounter hormones and then are transported to the nucleus (4). The crucial role of estrogen receptors in breast cancer makes them therapeutic targets for tumor growth inhibition. ER-positive tumors account for at least 75% of all breast cancers (5), and thus, many antitumor drugs aim at the regulation of estrogen receptor functions (6).

Estrogen signaling requires a subtle balance between two different receptors, ERα and ERβ (4, 7, 8). ERα and ERβ are products of different genes, and they have different transcription activities (9). Full-length ERα has 595 amino acids, comprising five different domains (10, 11): an N-terminal domain (NTD), a DNA-binding domain (DBD), a hinge domain, a ligand-binding domain (LBD), and a C-terminal domain (CTD). The N-terminal domain includes the activation function 1 region (AF-1) which does not rely on the presence of a ligand (12). The DNA-binding domain contains two zinc finger motifs that also bind to the DNA-specific sequences that are called hormone response elements (HRE) (13). The hinge domain is intrinsically disordered and resembles a polymeric string. It connects the DBD and the LBD and also binds to a chaperone (14). The LBD not only enhances the dimerization of estrogen receptors but also binds to coactivators or corepressor proteins. The LBD by itself contains the activation function 2 region (AF-2), whose action also depends on the ligand (15). The function of the C-terminal domain is not entirely clear and its length varies (16).

X-ray crystallography has already given us multiple structures of ERα. Most of these structures contain only the DBD or the LBD (17, 18). Recent cyro-EM experiments help us gain a better picture of the full-length ERα complex but lack many structural details (19). In this work, we use our coarse-grained force field and its structure prediction algorithms, AWSEM-suite, to build a structural model of human estrogen receptor α (hERα) and compare this model with a model built based on SAXS data in a previous paper (20). We will occasionally refer to this as the "full-length" receptor although it still misses two terminal domains. The AWSEM-3SPN2 forcefield with template and coevolutionary constraints can be used to produce high-resolution structures from low-resolution cryo-EM structures such as the recently studied progesterone

Author contributions: X.C., M.C., and P.G.W. designed research; X.C. and S.J. performed research; X.C., M.C., and C.B. contributed new reagents/analytic tools; X.C., S.J., and M.C. analyzed data; and X.C., S.J., C.B., and P.G.W. wrote the paper.

Reviewers: J.A.K., University of Illinois at Urbana-Champaign; Y.L., Weizmann Institute of Science; and Z.L.-S., University of Illinois at Urbana-Champaign.

The authors declare no competing interest.



[1]X.C., S.J., and M.C. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: pwolynes@rice.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2216906120/-/DCSupplemental.

Published February 2, 2023.

receptor complex (21), a publication which appeared just as we finished this manuscript.

Classically, agonists bind to the LBD of the estrogen receptor and thereby enable the receptor to bind to its cognate estrogen-response element, which is located near the promoter of the corresponding genes (22, 23). After activation by an "agonist" ligand, the DNA-bound ER recruits coactivators which then enhance gene expression. ER$\alpha$ and ER$\beta$ also indirectly stimulate gene expression by binding to other transcription factors such as AP-1 and SP-1 (24, 25). In contrast, when antagonists bind to the LBDs, the binding of the coactivator is blocked, which decreases gene expression. Based on this action of antagonists, the antagonist tamoxifen has been in clinical use for many years as endocrine therapy for breast cancer treatment (26, 27). Most designed ER ligands do not behave as "pure" antagonists as tamoxifen mostly does (28). To better understand the different responses to the antagonists and agonists, we explore different conformational ensembles of these large complexes using coarse-grained force fields and energy landscape theory.

Due to the large size of the estrogen receptor, there have only been a handful of attempts to simulate estrogen receptor binding processes. MD simulations of ER$\alpha$ primarily focus on the DNA-binding domain or on the ligand-binding domain by themselves (29, 30). In this work, we use an efficient coarse-grained, transferable, and realistic protein–DNA force field to survey the energy landscape of ER$\alpha$ both by itself and interacting with DNA and coactivators. We have already used AWSEM-3SPN.2 force field to study multiple protein–DNA systems including nucleosomes (31), NF$\kappa$B complexes (32), complexes of the Fis protein (33), the PU.1 complex (34), and a single-stranded DNA helicase system (35). This force field has succeeded in predicting both the structures and mechanisms of these systems. Frustration analysis using energy landscape theory to highlight regions with possible alternative configurations has been shown to not only shed considerable light on the folding and misfolding of proteins but also give insight into motions including allostery, electron transfer processes, protein–protein interactions, and enzyme catalysis (36–40).

In this work, we combine the AWSEM-Suite force field and 3SPN.2C force field first to predict the structure of our human estrogen receptor construct $\alpha$(hER$\alpha$) in different states. Owing to the key part played by the "intrinsically disordered" hinge region, there is no single structure but, instead, a broad conformational ensemble of structures. The predictions from our predicted ensemble for hER$\alpha$ agree well with measured small angle X-ray data (20). We also use these simulations to build the free energy landscape for agonist-bound hER$\alpha$ with DNA and successfully calculate its binding affinity, in agreement with experiment. Using principal component analyses of the molecular dynamics sampled trajectories of antagonist-bound hER$\alpha$, agonist-bound hER$\alpha$, and agonist-bound hER$\alpha$ with coactivators, we can analyze how communications between the DBDs and LBDS are mediated through the motions of the disordered hinge regions. These simulations yield a series of free energy landscapes that provide key intermediate structures. Differences between the frustration patterns of these species explain the way in which key parts of the sequence control these motions. Highlighting the intrinsically disordered nature of the hinge domains, another series of free energy landscapes of hER$\alpha$ having different hinge properties were constructed to explore how the hinge domains facilitate domain–domain communication. They act like the strings of a marionette allowing the LBD to control the binding affinity of the DBD to DNA remotely.

## The Predicted Structure of the Estrogen Receptor Complex Verified by Experiments

We first predicted the dimeric structure of the human estrogen receptor $\alpha$(hER$\alpha$), including both the LBDs and the DNA-binding domains (DBDs). For this prediction, we combined the AWSEM-Suite model for protein and the 3SPN.2C model for DNA. Compared with the original unbiased AWSEM force field, we have also incorporated template and coevolutionary information for hER$\alpha$. The structural templates with the lowest E-value (LBDs: 3ERD,3ERT; DBDs: 1HRC) were found using the multiple sequence alignment information from the HHpred

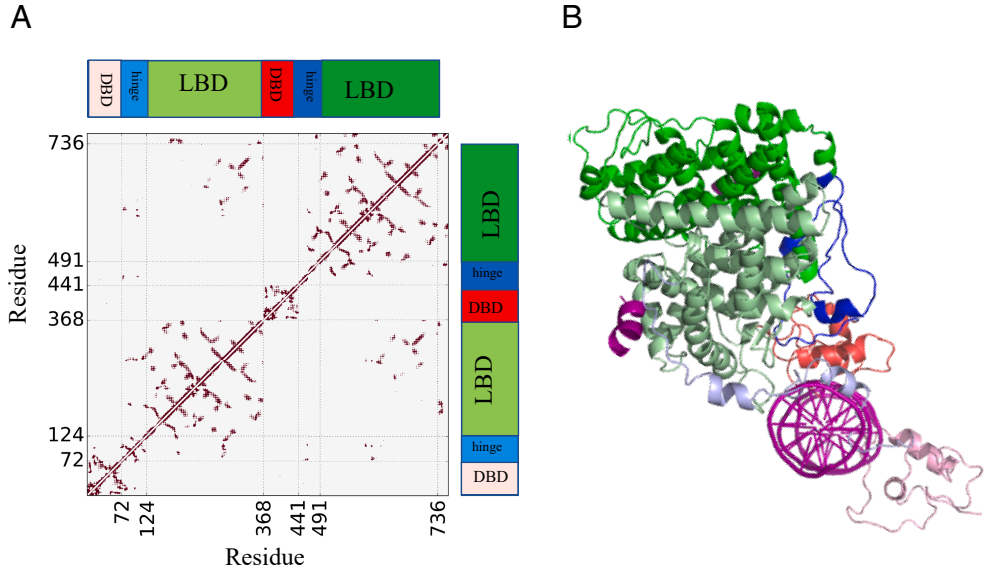


**Fig. 1.** Details of the predicted human estrogen receptor $\alpha$ (hER$\alpha$) complex structure. (*A*) A sequence bar indicates the domain assignments that we used in the text. The contact map of the estrogen receptor complex. The contacts are indicated in red. (*B*) The structure of the estrogen receptor is shown in cartoon: ligand-binding domains (LBDs), green or pale green for different chains; hinge domains, blue or ice blue for different chains; DNA-binding domains (DBDs), red and pink for different chains; DNA, Purple. Different views can be found in *SI Appendix*, Fig. S22 also.

server (41). It is worth noting that these structural templates provide only the monomeric structural restraints in our dimeric hERα prediction. The coevolutionary information from the RaptorX-contact server provides additional predicted contacts. We considered only predicted contacts whose reported contact probability was larger than 0.5. We ran 20 annealing simulations with different random velocity seeds and selected three frames having the lowest energies for each simulation. The final predicted structure is obtained as the central structure of the largest Qw cluster of 90 frames (as shown in Fig. 1*B*). Overall, the dimeric hERα has asymmetric domain–domain interactions. As shown in the contact map (as shown in Fig. 1*A*), one of the LBDs forms multiple contacts with the DBD and the hinge domain while the other one does not. The complexity of the interface between the LBDs and other domains suggests the functional importance of the LBDs. To evaluate the structural accuracy of our prediction, we scored our predicted structure against experimental SAXS data via the $\chi^2$ function, Foxs server (42). A lower $\chi^2$ indicates a better fit with the experimental SAXS profile. As shown in Fig. 1*C*, the $\chi^2$ of our predicted SAXS profile against the experimental SAXS profile is 1.215, suggesting very good agreement with experiment. In comparison, the $\chi^2$ of the models provided by Huang et al. was 1.339 (as shown in *SI Appendix*, Fig. S1) (43). Their model also shows asymmetric interfaces for the LBDs and the other domains.

To further increase our confidence in the hERα model, we evaluated the frustration patterns of the DNA-binding domains of hERα. As shown in *SI Appendix*, Fig. S2, highly frustrated interactions (red lines) indicate contacts that are energetically unfavorable so that they can easily reconfigure to other alternative structures. In contrast, the minimally frustrated interactions (green lines) suggest that these contacts are the most stable among their possible structures, so they prefer not to change during functional motions (36). We see that the interactions with DNA (*SI Appendix*, Fig. S2*B*) eliminate all the highly frustrated interactions of the DNA-binding domains found in the unbound form (*SI Appendix*, Fig. S2*A*), which agrees with their DNA-binding preference. We also calculated the binding affinity of hERα. *SI Appendix*, Fig. S3 shows the free energy profile projected onto the distance between the DBD and the DNA. The binding affinity of agonist-bound hERα turns out to be around −11.14 kcal/mol, which agrees well with the experimental values of (−10.03 to −11.93) kcal/mol (44–46).

## The Domain–Domain Collective Motions that Are Triggered by Antagonists, Agonists, and Coactivators

The predicted hERα dimeric structure agrees well with experiments at the structural and energetic levels. We now turn to the effects of binding ligands. The hERα dimer includes N-terminal domains, DBDs, hinge domains, LBDs, and CTDs. Without the LBDs, the hERα cannot activate gene transcription. When antagonists (such as 4-hydroxytamoxifen) bind to the LBD, the coactivator-LBD binding is blocked, and the gene expression is repressed. In contrast, the coactivators enhance gene repression when agonists (diethylstilbestrol) bind the LBD. In this work, we study the dynamics of hERα in several different situations: 1) with antagonist-bound LBDs; 2) with agonist-bound LBDs; and 3) with agonist-bound LBDs along with coactivators. We generated 20 trajectories at 300 K for these different situations. All trajectories start with the same initial structure (the predicted structure) with different velocities. To generate statistics, we used the last 4 million steps of the whole 6 million steps simulation, corresponding to 20 μs in lab time. The antagonist-bound hERα simulations employed template restraints on LBDs using the crystal structure of antagonist-bound LBDs (PDBID:3ERT), and the agonist-bound hERα simulations employed template restraints on LBDs based on the crystal structure of agonist-bound LBD (PBDID:3ERD). The simulations of the agonists-bound LBDs with coactivators employed template restraints on LBDs with the addition of coactivators.

Then, we utilized principal component analysis (PCA) on the Cartesian coordinates of these trajectories to explore the hERα structural fluctuations. As shown in *SI Appendix*, Fig. S4, the largest PC of antagonist-bound hERα, agonist-bound hERα, and agonist-bound hERα with coactivators contributed 33, 46, and 31% to the fluctuations, respectively. The largest PCs thus capture the main features of the dynamics. We display the relative motion of each residue for each PC in Fig. 2. We observed large motions of one of the DNA-binding domains but relatively small motions of the other DBD in the largest PC of antagonist-bound hERα. Meanwhile, the motions of the hinge domains and the ligand-binding domains are negligible. These observations suggest that the two DBDs move somewhat independently of the other domains (Movie S1). In the largest PC of agonist-bound hERα, the two DBDs have large motions along with moderate motions of the other domains, suggesting that the

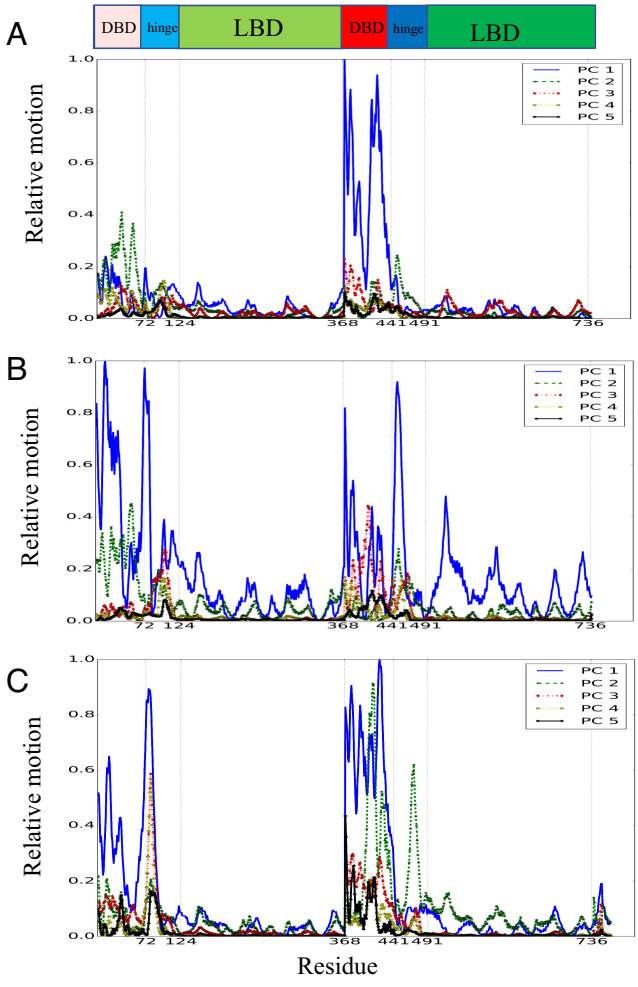


**Fig. 2.** Cartesian principal component analyses of the antagonist-bound hERα, agonist-bound hERα, and agonist-bound hERα bound with coactivators. (*A*) Cartesian PCA of the antagonist-bound hERα. (*B*) Cartesian PCA of the agonist-bound hERα. (*C*) Cartesian PCA of the agonist-bound hERα with the coactivator.

structural changes of the set of domains are in sync (Movie S2). The coordinated motion of all domains together suggests that for the agonist-bound hER$\alpha$, the two domains communicate during DNA binding. In contrast, we observed large motions of both DBDs along with only a tiny motion of the other domains in the largest PC of agonist-bound hER$\alpha$ with the coactivator bound(Movie S3). The differences between the largest PCs of the different forms of hER$\alpha$ suggest that domain–domain communication is most important in the agonist-bound form. We also investigated the structural fluctuations of apo hER$\alpha$. The apo form's largest PC behaves more like the antagonist-bound form in terms of its largest PC, while for this system, the 5th PC resembles the agonist-bound form's largest PC. These observations suggest that although the apo form has features of both the antagonist-bound and agonist-bound forms, the features of the antagonist-bound form dominate (30).

## The Domain–Domain Communication Underlying the hER$\alpha$'s Antagonist, Agonist, and Coactivator's Actions

The Cartesian PCAs of hER$\alpha$ reveal simultaneous domain motions for the agonist-bound hER$\alpha$. We constructed the free energy landscapes for all of the different hER$\alpha$ forms. We performed an umbrella sampling simulation using the distances between each of the centers of mass of the DNA-binding domains (DBDs) and the center of mass of the DNA as the biasing coordinates to build free energy profiles. Fig. 3 shows the

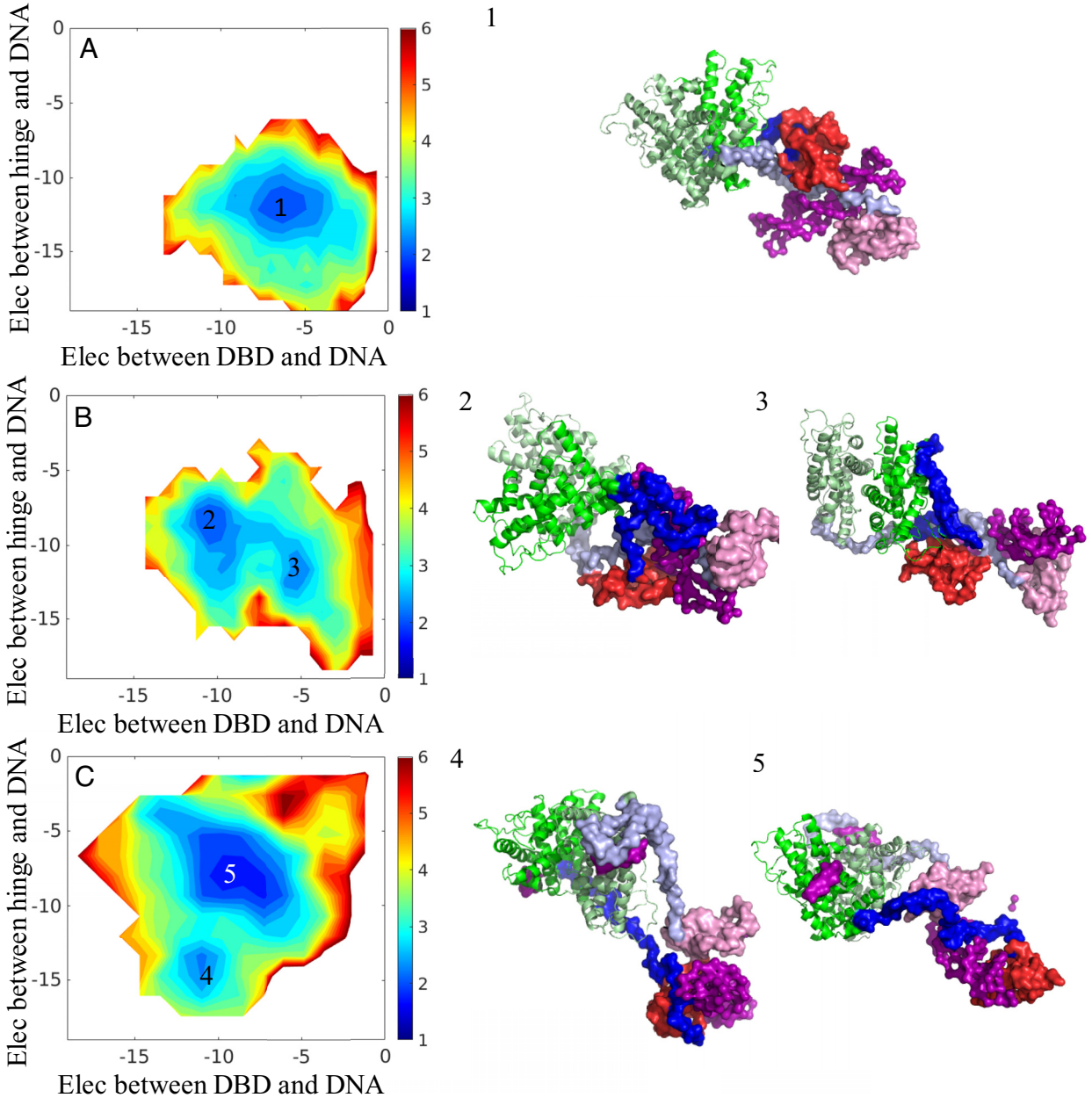


**Fig. 3.** Free energy landscapes of the antagonist-bound hER$\alpha$, the agonist-bound hER$\alpha$, and agonist-bound hER$\alpha$ with coactivators using as coordinates the electrostatic interaction between the DBDs and the DNA and the electrostatic interaction between the hinge domain and the DNA. (*A*) Free energy landscape of the antagonist-bound hER$\alpha$. (*B*) Free energy landscape of the agonist-bound hER$\alpha$. (*C*) Free energy landscape of the agonist-bound hER$\alpha$ with coactivators. Representative structures of each basin are shown in the *Right* panel. The ligand-binding domains (LBDs) are shown in the green or pale green cartoon for different chains; the hinge domains are shown in blue or ice blue surface for different chains; the DNA-binding domains (DBDs) are shown in red or pink surface for different chains; the DNA is shown in purple surface.

free energy surfaces projected using the electrostatic interaction between the DBDs and the DNA and the electrostatic interaction between the hinge domains and DNA. Representative structures from each basin are also shown in Fig. 3. Only one basin is apparent on the free energy landscapes of the antagonist-bound form. In this basin, the DNA interacts both with the DBDs and with the hinge domains. There are however several different DNA binding basins on the free energy profile of the agonist-bound form. One of the states (labeled as 3) prefers to bind the DNA using the hinge domains, while the other state (labeled as 2) prefers to bind DNA using the DBDs. As shown in the morph movie (Movie S6), the binding configurations of the DBDs change concomitantly with the significant changes of hinge binding configuration, supporting this as the mechanism of domain–domain communication. Although we observed two basins on the free energy profile of the agonist-bound hERα with coactivators, these basins have nearly the same binding preference for the two DBDs. From the morph movie (Movie S7), the DBDs–DNA remains relatively immobile, while the hinge–DNA relative configuration varies significantly, suggesting that the motions of each domain are nearly independent. We also constructed the free energy profile of apo hERα (*SI Appendix*, Fig. S7). On this landscape, the more favorable basin resembles the antagonist-bound structure, while the other resembles the agonist-bound structure. We see again that the apo form incorporates the features of the antagonist-bound and agonist-bound forms, although the features of the antagonist-bound form dominate. We see that domain–domain communication exists only through the binding process of agonist-bound hERα

To investigate the origin of domain–domain communication, the frustration patterns of the sampled hERαstructures were evaluated using the AWSEM force field. In this frustration analysis, the electrostatic interactions between hERα and DNA were found approximating the DNA as a GLU peptide with negative charges. The minimally frustrated interactions imply that the configurations are more stable than alternative structures. In contrast, highly frustrated interactions indicate that there is energetical instability allowing changes to other tertiary structures. The frustration pattern of the antagonist-bound hERα (*SI Appendix*, Fig. S8) reveals only minimally frustrated interactions on the LBD–hinge interface and the DBD–hinge interface, suggesting the stability of these surfaces. The frustration pattern of the agonist-bound hERαs incorporates only minimally frustrated interactions at the DBD–hinge interface (*SI Appendix*, Fig. S9 *A* and *B*). In contrast, for the basin where the DNA prefers to interact with the hinge, the minimally frustrated interactions are found only on the LBD–hinge interface. The differences between the frustration patterns of agonist-bound hERα species implicate the allosteric coupling of the different domains, as the means of domain–domain communication. We do not find a significant difference between the frustration patterns of agonist-bound hERα with coactivators on these interfaces (*SI Appendix*, Fig. S10). These frustration patterns of different hERα support the dominance of domain–domain communication in the dynamics of agonist-bound hERα. We also analyze the frustration patterns of apo hERα (*SI Appendix*, Fig. S7). The differences between the frustration patterns of the interfaces support that the domain–domain communication exists in the dynamics of apo hERα much as for the agonist-bound hERα. Both structures incorporate however minimally frustrated interactions on both the LBD–hinge interface and DBD–hinge interfaces, agreeing with the previous conclusion that the features of antagonist-bound hERα dominate.

To examine the connections in another way, the fluctuations in the distances between the Cα atoms of these trajectories were subjected to principal component analyses. *SI Appendix*, Fig. S11 displays the largest PCs of the antagonist-bound hERα, the agonist-bound hERα, and the agonist-bound hERα along with coactivators. These contribute 35, 37, and 26%, respectively to the overall fluctuations, suggesting that the largest PC can capture the main features of these motions. We plot the largest PC as a contact map, where the more robust contacts are colored in red, and the disappearing contacts are colored in blue. In the antagonist-bound hERα's PC, the formation of contacts on LBD–hinge interfaces indicates that the LBD hinge interfaces do not affect the behavior of the DBDs. The formation of contacts on both LBD–hinge interfaces and DBD–hinge interactions is observed in the largest PC of agonists-bound hERα, supporting domain–domain communication. There is no simultaneous formation of contacts on the LBD–hinge interface and corresponding DBD–hinge interface in the largest PC for the agonist-bound hERα with coactivators, suggesting the independence of these interfaces. These results support the critical role of the domain–domain communication during the motions of the agonist-bound hERα.

## The Hinge Domains' Roles in the Domain–Domain Communication

The hinge domain that connects the ligand-binding domain (LBD) and the DNA-binding domain (DBD) is intrinsically disordered. So, it is interesting to construct free energy landscapes of agonist-bound hERα where the hinge domain properties have been modified. We performed three sets of umbrella sampling simulations using the distances between the centers of mass of the DBDs and the center of mass of DNA as the bias coordinates: One set has the native hinge domains, another set has hinge domains without any charges, and in another set, only one hinge domain is present. Fig. 4 shows the free energy profile projected on the number of contacts between the DBDs and the DNA and the number of contacts between hinge domains and the DNA as collective coordinates. Representative structures from each basin are also displayed in Fig. 4. Two basins are identified on the free energy surface for the hERα with both native hinge domains (Fig. 4*A*). One basin incorporates more contacts between the hinge domains and DNA, while in the other more favorable basin, more contacts form between the DBDs and the DNA. The change of the pattern of contacts between the hinge domains and the DNA suggests that the motion of hinge domains helps DNA move from the hinge domains to the DBDs (Movie S8). In fact, we also see that the DNA helically translates upon binding. This would imply that receptor binding couples to large-scale DNA motions and architecture perhaps via supercoiling the DNA with motor proteins or topoisomerases. On the free energy surface of hERα having hinge domains with no charges (Fig. 4*A*), there are also two basins: The basin with fewer contacts between DBDs and DNA is similar to that found in hERα with native hinge domains. The other more stable basin having more contacts between the DBDs and the DNA now does not incorporate any contacts between hinge domains and the DNA (Movie S9). The change of the DBDs–DNA binding structures indicates that the electrostatic interaction with the hinge domains helps the DBD stabilize the DNA-binding configuration. The existence of two different binding configurations however suggests that the motion of the hinge domains helps the binding transition of DNA even without any electrostatic forces. We observed only the DBD-binding basin on the free energy surface of hERα that

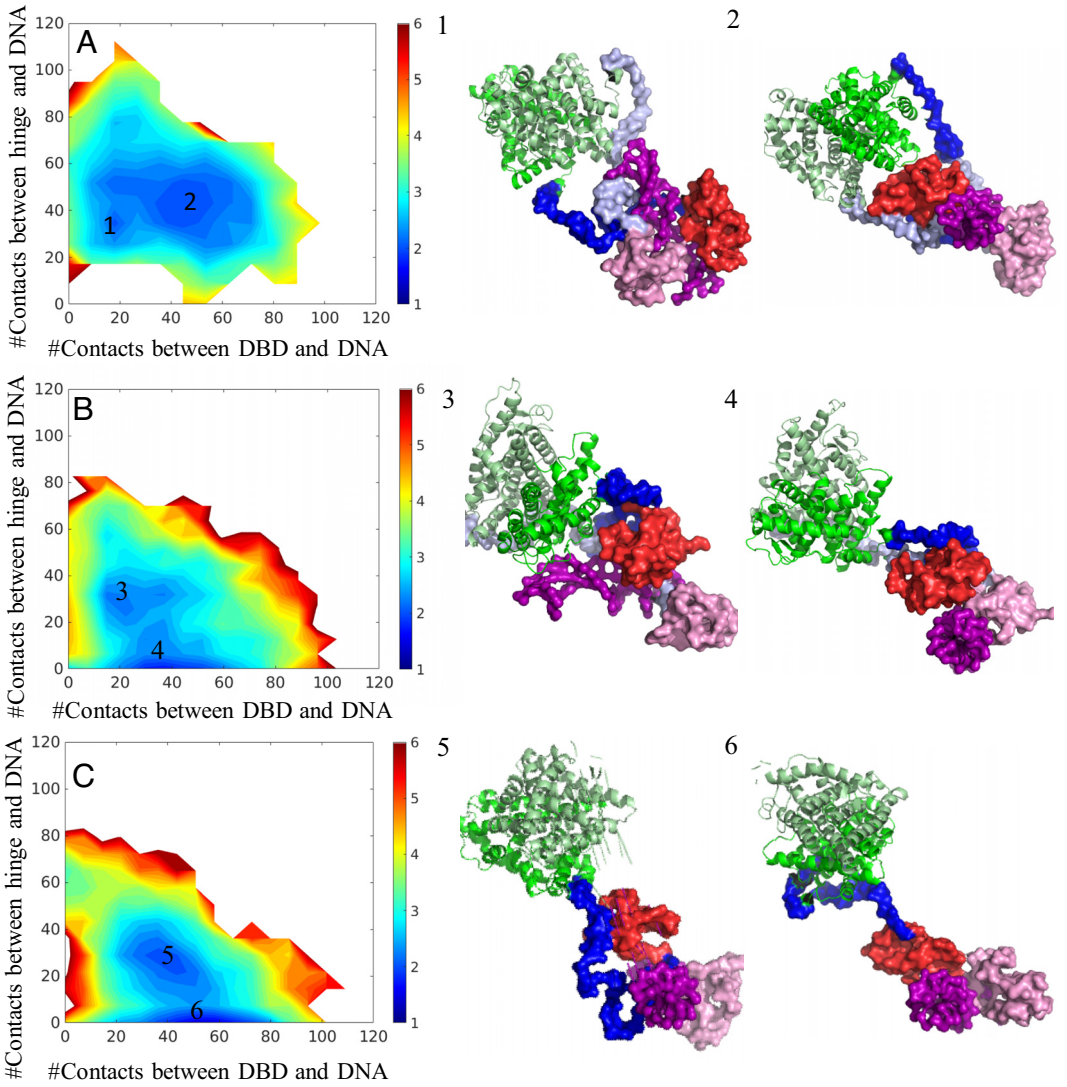


**Fig. 4.** Free energy landscapes of the agonist-bound hER$\alpha$ under different conditions projected using the number of contacts formed by the DBDs with DNA and the number of contacts formed by the hinge domains with DNA. (*A*) Free energy landscape of the regular hER$\alpha$ (*B*) Free energy landscape of an agonist-bound hER$\alpha$ with hinge domains having no charge. (*C*) Free energy landscape of an agonist-bound hER$\alpha$ that has only one hinge domain. Representative structures of each basin are shown in the *Right* panel. The ligand-binding domains (LBDs) are shown in green or pale green cartoon for different chains; the hinge domains are shown in blue or ice blue surface for different chains; the DNA-binding domains (DBDs) are shown in red or pink surface for different chains; the DNA is shown in purple surface.

lacks one of the hinge domains (Fig. 4*A*). The basin having more contacts between the hinge domains and DNA is similar to the DBD-binding basin of hER$\alpha$ with the native hinge domains, while the other basin with no contacts between the hinge domains and DNA is identical to the DBDs-binding species of hER$\alpha$ with the hinge domains with no charge (Movie S10). We see that the electrostatic forces from the hinge domains are important for the stability of DNA–DBD binding.

The frustration patterns of structures in these basins were evaluated. As shown in *SI Appendix*, Fig. S13, the DBD–DNA binding basin incorporates minimally frustrated interactions on the DBD–DNA interfaces, while the hinge–DNA binding basin incorporates minimally frustrated interactions also on the hinge–DNA interface. Minimally frustrated interactions are also found at the hinge–hinge interfaces. Examining the frustration pattern of agonist-bound hER$\alpha$ when the hinge charges are removed (*SI Appendix*, Fig. S14), we see that the number of minimally frustrated interactions on the hinge–hinge surface has decreased for both basins. Clearly, the hinge domains stabilize these basins via electrostatics. When there is only one hinge domain (*SI Appendix*, Fig. S15), the DBD–DNA binding basin has fewer minimally frustrated interactions at the DBD–DNA interfaces, and the hinge–DNA binding basin has fewer minimally frustrated interactions at the hinge–DNA interfaces compared to the basins for hER$\alpha$ with native hinge domains. These features of the frustration patterns support the critical role of the hinge domains in stabilization.

The mutations of the receptors that are found in breast cancer and psychiatric diseases (*SI Appendix*, Table S1) also are found to occur primarily in these hinges, supporting the hinge domains' important role in hER$\alpha$'s function (47). We have constructed the free energy landscapes of these pathological mutated hER$\alpha$s using perturbation theory (*SI Appendix*, Fig. S17). Compared with the native profile, in the mutants, the DNA–DBD–bound species (basin 1) and DNA–hinges–bound species (basin 3) are unfavored, suggesting that the DNA allosteric transition upon binding is hindered for these mutated hER$\alpha$s. The frustration patterns of representative structures

(*SI Appendix*, Table S2, S3, S4 and Figs. S18 and S19) also support these perturbation theory findings. In the frustration patterns of mutated hER$\alpha$s (*SI Appendix*, Figs. S18 and S19), one sees that all the mutated residues have highly frustrated interactions with the residues on the LBDs and the DNA, suggesting that domain–domain communication is hindered by these mutations during the DNA binding process (37, 48).

**The Marionette Mechanism for Domain–Domain Communication.** The critical role of the hinge domains in the domain–domain communications underlying the agonist-bound hER$\alpha$'s dynamics suggests looking at the free energy landscape using the averaged length of the hinge domains and the number of contacts between DBDs and DNA as coupled collective coordinates (as shown in Fig. 5*A*). We note that often while disordered linkers change sequence, they often take on the same length during evolution. Representative structures from each basin are shown in Fig. 5*B*. Again, two DBD–DNA binding configurations were found in the landscapes having a large number of contacts between the DBDs and the DNA. These two basins differ in the average length of the two hinge domains. Another basin with fewer contacts between DBDs and DNA has a short averaged length of hinge domains and more contacts between hinge domains and DNA. The differences in the averaged length of the hinge domain between all three basins suggest that the hinge controls DNA-binding activity (as shown in Movies S11, S12, and S13). How these two hinge domains move resembles the way puppets are manipulated by a puppeteer through strings. An interesting question is whether the strings ever tangle. We examined these structures for knots using a knot analysis server (49). These structures do not show any knots to start with and none form during binding. An interesting question is whether "topological isomers" with knotted hinge domains are possible or exist in vivo. Such topological isomers seem to exist for cohesin (50). We leave this question for future work.

To further analyze the marionette mechanism for the agonist-bound hER$\alpha$, the representative structures of the three basins were studied using frustration analysis. The DNA–hinge binding basin with the moderate averaged length of hinge domains has the largest number of minimally frustrated interactions on the LBD–hinge interface (*SI Appendix*, Fig. S20*B*). The two different DBD–DNA binding basins contain more minimally frustrated interactions at the DBD–DNA interface than does the DNA–hinge binding basin. The basin with the longest averaged length of the hinge domains has many minimally frustrated interactions and few highly frustrated interactions at the hinge–hinge surface (*SI Appendix*, Fig. S20*C*). In contrast, the basin with the shortest average length of the hinge domains has neither minimally nor highly frustrated interactions on the hinge–hinge interfaces (*SI Appendix*, Fig. S20*A*). These frustration patterns support the idea that the LBD–hinge interface changes synchronously with the change of the DBD–hinge interface. The motions of the hinge domains resemble the manipulation of string puppets, with the ligand-binding domains being the puppeteer.

## Discussion

**The Marionette Mechanism in Multidomain Proteins with Disordered Linkers.** A key question in the estrogen receptor field has been how allosteric conformational changes in the ligand-binding domain lead to changes of function at the DNA-binding domain since they are connected by disordered linkers. In the present study, we have revealed the structural changes involved in the domain–domain communication in the agonist-bound hER$\alpha$'s dynamics. Furthermore, we have identified the specific role played by the "intrinsically disordered" hinge domains in domain–domain communication and have shown how the proposed communication occurs through a marionette mechanism. As shown in Fig. 6 (*SI Appendix*, Fig. S21), our results suggest that after antagonists leave and agonists bind to the LBDs of hER$\alpha$, initially the hinge domains capture the DNA. Then, the motions of LBDs transfer their changes to the hinge domains which shorten and thereby transfer information to the DBDs finally. Ultimately, the agonists act as marionettists that first control the motions of the LBDs; leading to the variation of the LBD–hinge interfaces which then control the lengths of the hinge domains as strings, and finally change the binding between DNA and the other receptor domains (hinge domains or DBDs) much as string puppets are manipulated by a puppeteer.

The estrogen receptor dimer is asymmetric, although the sequences of the two chains are the same. The origin of asymmetry comes from the disordered regions of the long hinges, which connect the ligand-binding domain and the DNA-binding domain. The disordered hinges domains have lots of local minima on the energy landscape. The length, the critical position, and the diverse local minima of the disordered region make the dimer

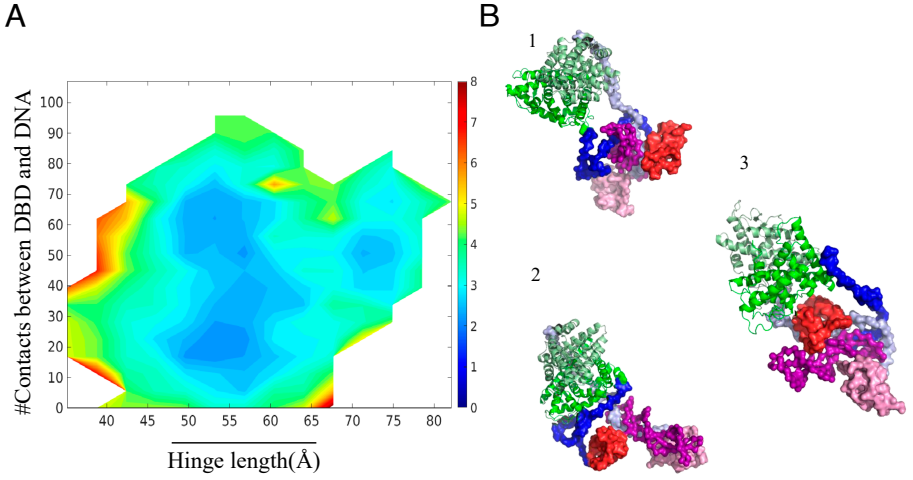


**Fig. 5.** (*A*) The free energy landscape of the agonists-bound hER$\alpha$ using the number of contacts formed by DBDs with DNA and the average length of two hinge domains. (*B*) The representative structures of each basin: The ligand-binding domains (LBDs) are shown in green or pale green cartoon for different chains; the hinge domains are shown in blue or ice blue surface for different chains; the DNA-binding domains (DBDs) are shown in red or pink surface for different chains; the DNA is shown in purple surface.

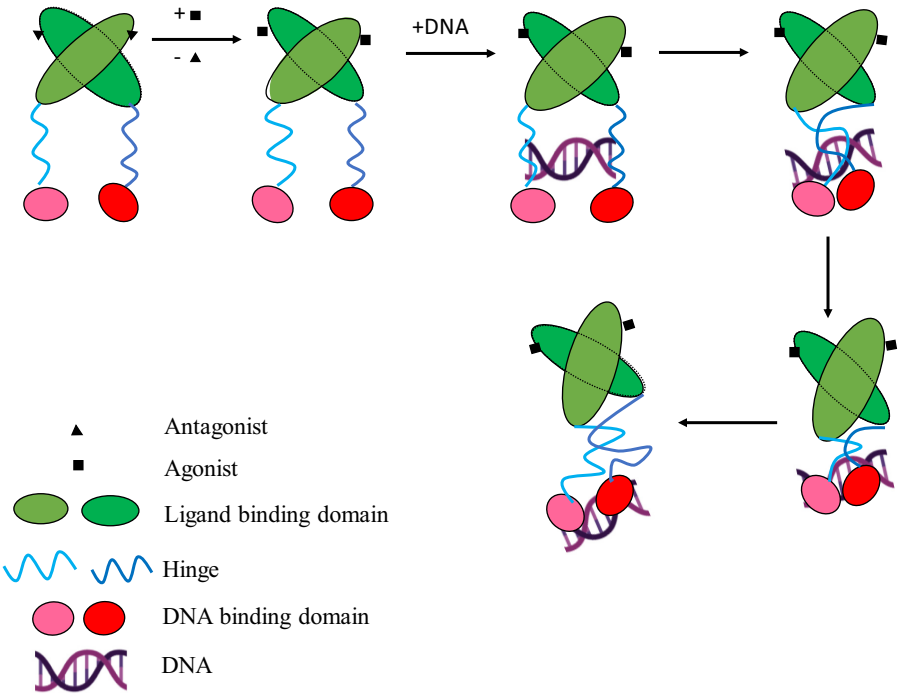


**Fig. 6.** A schematic diagram for the marionette mechanism of estrogen receptor with agonists. With the departure of the antagonist, the estrogen receptor binds the agonist and the hinge domain attracts the DNA. Accompanying the motions of the LBDs, the motions are transferred to the hinge domains and the DBDs, which change the DNA binding configurations including the hinge length as well as the contacts between DBDS and DNA (This sequence of events is shown in Movies S11, S12, and S13).

asymmetric. These features, leading to asymmetry, contribute to the marionette mechanism in the actions of the estrogen receptor.

In summary, our protocol has revealed a mechanism of action for hER$\alpha$. hER$\alpha$ belongs to the nuclear receptor family. Many nuclear receptors such as Farnesoid X receptor (51) are also multimers having multiple domains like hER$\alpha$, including LBDs and DBDs. The structural similarity and functional similarity of these receptors to hER$\alpha$ suggest that the marionette mechanism may be a universal mechanism for nuclear receptor action (52), as Fig. 6. The motion between the small parts of the nuclear receptor and other molecules is transferred to the binding domain by lengthening and shortening the string-like hinges to carry out its functions.

## Methods

A detailed description of *Materials and Methods* is given in *SI Appendix*. Briefly, the simulations were carried out using the AWSEM-3SPN.2C forcefield for protein–DNA complexes in the LAMMPS open-source software package.

**Data, Materials, and Software Availability.** The input files, simulation data, and analysis codes used in this work are deposited in GitHub: https://github.com/chemlover/ER_project. Instructions for the development and use of AWSEM-3SPN2 including templates and coevolutionary term assignments have been placed on GitHub: http://github.com/adavtyan/awsemmd/wiki/ and http://github.com/chemlover/ER_project/template.txt.

**ACKNOWLEDGMENTS.** This work is funded by the G. Harold & Leila Y. Mathers Charitable Foundation primary award No. 30086506 (to Wolynes). It is also funded by the Center for Theoretical Biological Physics, sponsored by NSF grant PHY-2019745. Additionally, we wish to recognize the D.R. Bullard Welch Chair at Rice University, Grant C-0016 (to Wolynes). This work was also supported by National Science Foundation (NSF) grants PHY-1522550 (to Bueno). We thank the Data Analysis and Visualization Cyberinfrastructure funded by National Science Foundation Grant OCI-0959097.

Author affiliations: [a]Center for Theoretical Biological Physics, Rice University, Houston, TX 77005; [b]Department of Chemistry, Rice University, Houston, TX 77005; [c]Department of Biosciences, Rice University, Houston, TX 77005; [d]Department of Research and Development, neoX Biotech, Beijing 102206, China; and [e]Systems, Synthetic, and Physical Biology, Rice University, Houston, TX 77005

1. R. Siersbæk, S. Kumar, J. S. Carroll, Signaling pathways and steroid receptors modulating estrogen receptor $\alpha$ function in breast cancer. *Genes Dev.* **32**, 1141–1154 (2018).
2. N. Fuentes, P. Silveyra, Estrogen receptor signaling mechanisms. *Adv. Protein Chem. Struct. Biol.* **116**, 135–170 (2019).
3. S. K. Jin, W. X. Yang, Factors and pathways involved in capacitation: How are they regulated? *Oncotarget* **8**, 3600–3627 (2017).
4. K. Hu, G. Zhong, F. He, Expression of estrogen receptors ER$\alpha$ and ER$\beta$ in endometrial hyperplasia and adenocarcinoma. *Int. J. Gynecol. Cancer* **15**, 537–541 (2005).
5. X. Fu, C. De Angelis, R. Schiff, Interferon signaling in estrogen receptor–positive breast cancer: A revitalized topic. *Endocrinology* **163**, bqab235 (2022).
6. L. Jin, Y. Li, Structural and functional insights into nuclear receptor signaling. *Adv. Drug Deliv. Rev.* **62**, 1218–1226 (2010).
7. D. G. Monroe *et al.*, Estrogen receptor isoform-specific regulation of endogenous gene expression in human osteoblastic cell lines expressing either ER$\alpha$ or ER$\beta$. *J. Cell. Biochem.* **90**, 315–326 (2003).
8. C. Guido *et al.*, Human sperm physiology: Estrogen receptor alpha (ER$\alpha$) and estrogen receptor beta (ER$\beta$) influence sperm metabolism and may be involved in the pathophysiology of varicocele-associated male infertility. *J. Cell. Physiol.* **226**, 3403–3412 (2011).
9. B. S. Katzenellenbogen, J. A. Katzenellenbogen, Estrogen receptor transcription and transactivation estrogen receptor alpha and estrogen receptor beta: Regulation by selective estrogen receptor modulators and importance in breast cancer. *Breast Cancer Res.* **2**, 1–10 (2000).
10. V. Kumar *et al.*, Functional domains of the human estrogen receptor. *Cell* **51**, 941–951 (1987).
11. S. Green, V. Kumar, A. Krust, P. Walter, P. Chambon, *Structural and Functional Domains of the Estrogen Receptor in Cold Spring Harbor Symposia on Quantitative Biology* (Cold Spring Harbor Laboratory Press, 1986), vol. 51, pp. 751–758.
12. P. J. Kushner *et al.*, Estrogen receptor pathways to AP-1. *J. Steroid Biochem. Mol. Biol.* **74**, 311–317 (2000).
13. J. W. Schwabe, L. Chapman, J. T. Finch, D. Rhodes, The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements. *Cell* **75**, 567–578 (1993).

14. W. Zwart *et al*., The hinge region of the human estrogen receptor determines functional synergy between AF-1 and AF-2 in the quantitative response to estradiol and tamoxifen. *J. Cell Sci.* **123**, 1253–1261 (2010).
15. S. Möcklinghoff *et al*., Synthesis and crystal structure of a phosphorylated estrogen receptor ligand binding domain. *Chembiochem* **11**, 2251–2254 (2010).
16. Y. Arao, K. S. Korach, The physiological role of estrogen receptor functional domains. *Essays Biochem.* **65**, 867–875 (2021).
17. A. K. Shiau *et al*., The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927–937 (1998).
18. R. A. Goldstein, J. A. Katzenellenbogen, Z. A. Luthey-Schulten, D. A. Seielstad, P. G. Wolynes, Three-dimensional model for the hormone binding domains of steroid receptors. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9949–9953 (1993).
19. P. Yi *et al*., Structure of a biologically active estrogen receptor-coactivator complex on DNA. *Mol. Cell* **57**, 1047–1058 (2015).
20. W. Huang *et al*., Multidomain architecture of estrogen receptor reveals interfacial cross-talk between its DNA-binding and ligand-binding domains. *Nat. Commun.* **9**, 1–10 (2018).
21. X. Yu *et al*., Spatial definition of the human progesterone receptor-B transcriptional complex. *Iscience* **25**, 105321 (2022).
22. D. M. Lonard, C. L. Smith, Molecular perspectives on selective estrogen receptor modulators (SERMS): Progress in understanding their tissue-specific agonist and antagonist actions. *Steroids* **67**, 15–24 (2002).
23. S. Nilsson *et al*., Mechanisms of estrogen action. *Physiol. Rev.* **81**, 1535–1565 (2001).
24. E. J. Routledge, R. White, M. G. Parker, J. P. Sumpter, Differential effects of xenoestrogens on coactivator recruitment by estrogen receptor ER$\alpha$ and ER$\beta$. *J. Biol. Chem.* **275**, 35986–35993 (2000).
25. C. J. Gruber, D. M. Gruber, I. M. Gruber, F. Wieser, J. C. Huber, Anatomy of the estrogen response element. *Trends Endocrinol. Metab.* **15**, 73–78 (2004).
26. T. Barkhem *et al*., Differential response of estrogen receptor $\alpha$ and estrogen receptor $\beta$ to partial estrogen agonists/antagonists. *Mol. Pharmacol.* **54**, 105–112 (1998).
27. C. K. Osborne, Tamoxifen in the treatment of breast cancer. *N. Eng. J. Med.* **339**, 1609–1618 (1998).
28. D. S. Goodsell, The molecular perspective: Tamoxifen and the estrogen receptor. *Oncologist* **7**, 163–164 (2002).
29. G. Lambrinidis *et al*., The estrogen receptor and polyphenols: Molecular simulation studies of their interactions, a review. *Environ. Chem. Lett.* **4**, 159–174 (2006).
30. T. D. McGee, J. Edwards, A. E. Roitberg, Preliminary molecular dynamic simulations of the estrogen receptor alpha ligand binding domain from antagonist to apo. *Int. J. Environ. Res. Public Health* **5**, 111–114 (2008).
31. B. Zhang, W. Zheng, G. A. Papoian, P. G. Wolynes, Exploring the free energy landscape of nucleosomes. *J. Am. Chem. Soc.* **138**, 8126–8133 (2016).
32. D. A. Potoyan, W. Zheng, D. U. Ferreiro, P. G. Wolynes, E. A. Komives, Pest control of molecular stripping of NF$\kappa$B from DNA transcription sites. *J. Phys. Chem. B* **120**, 8532–8538 (2016).
33. M. Y. Tsai, W. Zheng, M. Chen, P. G. Wolynes, Multiple binding configurations of Fis protein pairs on DNA: Facilitated dissociation versus cooperative dissociation. *J. Am. Chem. Soc.* **141**, 18113–18126 (2019).
34. X. Chen, M. Y. Tsai, P. G. Wolynes, The role of charge density coupled DNA bending in transcription factor sequence binding specificity: A generic mechanism for indirect readout. *J. Am. Chem. Soc.* **144**, 1835–1845 (2022).
35. S. Jin *et al*., Computationally exploring the mechanism of bacteriophage T7 gp4 helicase translocating along ssDNA. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2202239119 (2022).
36. D. U. Ferreiro, J. A. Hegler, E. A. Komives, P. G. Wolynes, Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19819–19824 (2007).
37. X. Chen, M. Chen, P. G. Wolynes, P. Wittung-Stafshede, H. B. Gray, Frustration dynamics and electron-transfer reorganization energies in wild-type and mutant azurins. *J. Am. Chem. Soc.* **144**, 4178–4185 (2022).
38. M. Chen *et al*., Protein structure refinement guided by atomic packing frustration analysis. *J. Phys. Chem. B* **124**, 10889–10898 (2020).
39. M. Jenik *et al*., Protein frustratometer: A tool to localize energetic frustration in protein molecules. *Nucleic Acids Res.* **40**, W348–W351 (2012).
40. M. Chen *et al*., Surveying biomolecular frustration at atomic resolution. *Nat. Commun.* **11**, 1–9 (2020).
41. J. Söding, A. Biegert, A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
42. D. Schneidman-Duhovny, M. Hammel, A. Sali, FoXS: A web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).
43. H. Hua, H. Zhang, Q. Kong, Y. Jiang, Mechanisms for estrogen receptor expression in human cancer. *Exp. Hematol. Oncol.* **7**, 1–11 (2018).
44. R. L. Rich *et al*., Kinetic analysis of estrogen receptor/ligand interactions. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8562–8567 (2002).
45. P. Augereau *et al*., Characterization of the proximal estrogen-responsive element of human cathepsin D gene. *Mol. Endocrinol.* **8**, 693–703 (1994).
46. F. E. Murdoch, D. A. Meier, J. D. Furlow, K. A. Grunwald, J. Gorski, Estrogen receptor binding to a DNA response element in vitro is not dependent upon estradiol. *Biochemistry* **29**, 8377–8385 (1990).
47. M. H. Herynk, S. A. Fuqua, Estrogen receptor mutations in human disease. *Endocr. Rev.* **25**, 869–898 (2004).
48. Y. Zhang *et al*., Frustration and the kinetic repartitioning mechanism of substrate inhibition in enzyme catalysis. *J. Phys. Chem. B* **126**, 6792–6801 (2022).
49. G. Kolesov, P. Virnau, M. Kardar, L. A. Mirny, Protein knot server: Detection of knots in protein structures. *Nucleic Acids Res.* **35**, W425–W428 (2007).
50. D. Krepel, A. Davtyan, N. P. Schafer, P. G. Wolynes, J. N. Onuchic, Braiding topology and the energy landscape of chromosome organization proteins. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1468–1477 (2020).
51. W. Zheng *et al*., A novel class of natural FXR modulators with a unique mode of selective co-regulator assembly. *ChemBioChem* **18**, 721–725 (2017).
52. E. R. Weikum, X. Liu, E. A. Ortlund, The nuclear receptor superfamily: A structural perspective. *Protein Sci.* **27**, 1876–1892 (2018).