Received: 25 January 2022 | Revised: 2 May 2022 | Accepted: 3 May 2022

DOI: 10.1002/pro.4334

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

Topological dynamics of an intrinsically disordered N-terminal domain of the human androgen receptor

Vahid Sheikhhassani[1,2] | **Barbara Scalvini**[1,2] | **Julian Ng**[1,2] | **Laurens W. H. J. Heling**[1,2] | **Yosri Ayache**[1,2] | **Tom M. J. Evers**[1,2] | **Eva Estébanez-Perpiñá**[3] | **Iain J. McEwan**[4] | **Alireza Mashaghi**[1,2]

[1]Medical Systems Biophysics and Bioengineering, Leiden Academic Centre for Drug Research, Faculty of Science, Leiden University, Leiden, The Netherlands

[2]Centre for Interdisciplinary Genome Research, Faculty of Science, Leiden University, Leiden, The Netherlands

[3]Department of Biochemistry and Molecular Biomedicine, Institute of Biomedicine (IBUB) of the University of Barcelona (UB), Barcelona, Spain

[4]Institute of Medical Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Scotland, UK

Correspondence
Alireza Mashaghi, Medical Systems Biophysics and Bioengineering, Leiden Academic Centre for Drug Research, Faculty of Science, Leiden University, Leiden, The Netherlands.
Email: a.mashaghi.tabari@lacdr.leidenuniv.nl

Review Editor: John Kuriyan

Abstract

Human androgen receptor contains a large N-terminal domain (AR-NTD) that is highly dynamic and this poses a major challenge for experimental and computational analysis to decipher its conformation. Misfolding of the AR-NTD is implicated in prostate cancer and Kennedy's disease, yet our knowledge of its structure is limited to primary sequence information of the chain and a few functionally important secondary structure motifs. Here, we employed an innovative combination of molecular dynamics simulations and circuit topology (CT) analysis to identify the tertiary structure of AR-NTD. We found that the AR-NTD adopts highly dynamic loopy conformations with two identifiable regions with distinct topological make-up and dynamics. This consists of a N-terminal region (NR, residues 1–224) and a C-terminal region (CR, residues 225–538), which carries a dense core. Topological mapping of the dynamics reveals a traceable time-scale dependent topological evolution. NR adopts different positioning with respect to the CR and forms a cleft that can partly enclose the hormone-bound ligand-binding domain (LBD) of the androgen receptor. Furthermore, our data suggest a model in which dynamic NR and CR compete for binding to the DNA-binding domain of the receptor, thereby regulating the accessibility of its DNA-binding site. Our approach allowed for the identification of a previously unknown regulatory binding site within the CR core, revealing the structural mechanisms of action of AR inhibitor EPI-001, and paving the way for other drug discovery applications.

KEYWORDS

androgen receptor, circuit topology, conformational dynamics, intrachain contacts, nuclear hormone receptors

Vahid Sheikhhassani and Barbara Scalvini contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

1 | INTRODUCTION

A key structural element of nuclear hormone receptors (NHRs) is the N-terminal transactivation domain (NTD), which is critical for receptor function, yet its structure is unknown. This is due to its intrinsic disorder and high dynamics that prevent its structural determination. As such, molecular mechanism behind NTD function remains elusive.[1–3] In the case of the androgen receptor (AR), recent cryo-electron microscopy (cryo-EM) analysis of the full-length receptor in complex with an interacting coregulatory partner, suggest that the NTD is a disordered conformation that surrounds the ligand-binding domain (LBD).[4] However, additional studies show the AR-NTD can also fold in functional state, as truncated forms of AR, devoid of the LBD, act as constitutively fully active form,[5] suggesting a central role of the NTD as a transcriptional driver.[6,7] This makes the NTD an interesting target for drug development,[8–10] however the absence of high-resolution information has challenged this development.

Studying the conformation of highly dynamic intrinsically disordered protein regions (IDRs) is a major challenge in structural biology.[11,12] Some important steps have been taken including insights from single-molecule fluorescence resonance energy transfer (smFRET), electron paramagnetic resonance (EPR), and nuclear magnetic resonance (NMR).[13] However, none have been applied successfully to resolve the structure of the full-length AR-NTD due to technical difficulties.[14–16]

The intrinsic disorder poses a challenge not only for experimental analysis of the conformation but also for computational modeling of the chain due to the size of the conformation space and lack of stable folds. For example, the state-of-the art artificial intelligence-based prediction approaches fail to identify the conformation of AR-NTD.[17] As such, new technological innovations are needed to study intrinsically disordered protein chains. Despite being disordered, it is reasonable to assume that at least some residues in the disordered protein regions will adopt certain arrangements to function.[18] Topological analysis might be able to capture those invariant arrangements hidden within a dynamic geometry. However, the field of protein topology has been primarily focused on knotted conformations and ignored intra chain contacts,[19,20] and the field of structural biology mainly concentrated on geometric analysis of protein chains. Circuit topology (CT) has recently emerged as a complementary topology approach to knot theory[21,22] and is able to fully characterize folded molecular chains.[23,24] Circuit topology is a theoretical framework which only focuses on intra-chain contacts and their arrangement within the chain. If we pick any pair of contacts in the chain, their topological relation can either be one of three possible arrangements: series (S), parallel (P), and cross (X) (Figure S1). Contacts which are in series do not intersect and belong to different sub-sections of the chain. Parallel relation implies that the first contact is formed by a sub-section of the chain included in the two contact sites of the second contact. Two contacts in cross relation intersect but belong to a different class with respect to parallel relations, since the contact sites of one loop do not fully encompass the other. One can establish the topological relation of any two contacts in the chain: the total number of S, P, and X can be used as the simplest measure of the topological state of the chain. The CT approach has led to unprecedented insights into the role of topology in molecular folding processes.[25–34] However, circuit topology has not yet been used to study intrinsically disordered proteins, and thus the topology of IDPs/IDRs is an unexplored area.

Here, we combine an innovative CT-based analysis with molecular dynamics (MD) simulations to study an intrinsically disordered protein chain for the first time. We focus on the AR-NTD, given its key role in receptor physiology and mediating diseases.[35–37] We search for topological patterns and investigate their implications for AR receptor assembly, target DNA binding, and interaction with AR inhibitor drugs.

2 | METHODS

2.1 | Modeling of three-dimensional structure of AR-NTD

Due to its disordered conformation, there is no resolved structure of the NTD domain deposited on the Protein Data Bank (PDB). To start from a computationally efficient initial structure, I-TASSER server[38] was used for modeling the three-dimensional structure of the AR-NTD (residues 1–538). I-TASSER is a hierarchical approach to protein structure prediction and structure-based function annotation. It is ranked the best protein structure prediction method by the Critical Assessment of Protein Structure Prediction (CASP) community. The *Z*-score and normalized *Z*-score >1 indicates a template with a good alignment quality. The Confidence-score (*C*-score, values −5 to 2) reflects the threading alignment and convergence of structure refinement simulations with higher score corresponding to a high confidence model. A Template Modeling (TM) score >0.5 indicates correct conformation of the model, while scores <0.17 corresponds to a model of random topology. Therefore, models with the highest *Z*-score, *C*-score and TM-score were selected for our MD studies.

In order to further optimize and validate the initial structure, an energy minimization step using steepest descent method was performed followed by a conjugate gradient with a ff99SB all atom force field to perform a total of 100,000 steps using the GROMACS software package.[39] For validation of energy refined structures, WHATIF[40] was used for anomalous bond angle and bond length detection. RAMPAGE[41] was used for Ramachandran plot analysis and detection of incorrectly modeled protein regions was done using ERRAT.[30] The compatibility of the three-dimensional protein model to amino acid sequence was verified using VERIFY3D[42] and the global quality of the models was calculated using QMEAN.[43] Structural disorder was analyzed using the PONDR web server and plots directly obtained from the server.[44]

2.2 | Molecular dynamics setup

In this study, 5 μs of MD simulations was carried out using the SIRAH coarse-grained force field for proteins in combination with the WT4 explicit coarse-grained water model.[45] The advantage of using this model is to reproduce the roughly tetrahedral ordering of bulk water due to the existence of hydrogen bonds between atomistic water molecules. Three MD simulation with different random seeds started from energy minimized predicted structure obtained from structure prediction pipeline. The entire AR-NTD (residues 1–538) was mapped to a coarse-grained representation according to the standard SIRAH mapping. Rhombic dodecahedron box was used to dissolve NTD structure by adding WT4 water molecules. Electroneutrality and physiological concentration of salt were achieved by replacing corresponding amount of water molecules with NaW and ClW (coarse-grained representations of Na+ and Cl− ions, respectively). All systems were minimized using the steepest descent algorithm and then went through a 5 ns NVT equilibration, a 5 ns NPT equilibration, and a NPT production run. The leapfrog integrator with a 20 fs time step was used throughout. Protein beads were constrained with the LINCS algorithm[46] during the equilibration, and no constraints were used during the minimization and production. The temperature was kept at 310 K with a velocity rescale thermostat,[47] and the pressure at 1 bar with the Parrinello−Rahman barostat.[48] τ_T for the thermostat was set to 1.0 ps during the equilibration phases and to 2.0 ps during the production. τ_P for the barostat was set to 10.0 ps during both the NPT equilibration and the production. Both nonbonded cut-offs (van der Waals and shortrange electrostatics) were set to 1.2 nm. Long-range electrostatics were treated with the PME method with a 0.2 nm grid spacing during the equilibration and 0.25 nm during the production. Nonbonded interactions were calculated using a 1.2 nm cut-off neighbor list, updated every 25 steps (in the production and the NPT equilibration) or 10 steps (in the NVT equilibration). Both energy and pressure dispersion corrections were applied. Periodic boundary conditions and the minimum image convention were used. Snapshots were collected every 1,000 steps (20 ps). All simulations and analyses were carried out with the GROMACS 2020 software.[39]

2.3 | Molecular docking

The binding interactions of NTD with LBD and DBD was studied using the ClusPro web server[49] with the aim of identifying interaction sites. This docking suite has been consistently rated among the best global docking procedures in the CAPRI challenge (Critical Assessment of Predicted Interaction).[49] ClusPro first performs rigid body docking by sampling a wide range of conformations and subsequently clusters the thousand lowest-energy structures based on root-mean-square deviation (RMSD) to identify the largest clusters that represent the most likely models. It furthermore refines the represented structures by energy minimizing the CHARMM potential.[49] We supplied ClusPro with previously reported crystal structures for DBD (PDB code: 1R4I), LBD (PDB code: 1T7T), and RAP74-CTD (PDB code: 1I27) to dock with our energy minimized structure of the NTD. The top 10 best ranked models in ClusPro were selected for further analysis.

In order to find the possible binding site of the EPI-001 compound on the CR region of the NTD, drug–protein interaction analysis was performed. The mol file of the EPI-001 was created in online molecular drawing module and then transferred to PRODRG server[50] to generate the required topologies and structure coordinates. The representative structures of three clusters obtained from clustering the CR region of NTD were prepared for docking by adding Kollman charges and hydrogen atoms to the polar groups of the protein using AutoDock tools.[51] AutoDock 4.2 package[51] was used for docking the EPI-001 to the target molecules. Blind docking was performed for the drug using the default parameters on Autodock.

2.4 | Preparing the structures for circuit topology analysis

After the trajectory of the system was obtained, atomic positions of amino acids were generated from the location of CG beads. Backmapping was done using

sirah_vmdtk.tcl plugin followed by 100 steps of steepest-descent and 50 steps of Conjugated Gradient minimization in vacuum using the sander module of AmberTools.[42] This process made the procedure robust and independent from the fine details of the back-mapping library. The obtained atomistic coordinates were used for circuit topology analysis.

2.5 | Circuit topology analysis

Circuit topology parameters were retrieved using custom-made Python code, modified for energy, length filtering, and circuit decomposition. For doing this, 1,000 frames were extracted from the trajectory containing only the protein molecule. Contacts between residues were calculated based on a distance cut-off of 4.5 Å. Residue pairs where more than five atoms were found to be at a respective distance below the distance cut-off were computed as contacts. The three closest neighbors of each residue were excluded from analysis.

In order to perform dynamic CT analysis, frames from the AR-NTD simulations were taken 5 ns apart and transformed into binary contact maps. The stack of contact maps was processed to create the contact life-time kymographs: all combinations of residues which never participate in contacts were removed. In order to create the contact lifetime distribution, the maximum lifetime was chosen for each particular combination of residues.

Time filtering for topological study was performed by creating life-time filters matrices. These matrices were then multiplied to regular contact matrices as a customization of the CT Python tool. All dynamic CT analysis was performed using a custom-made Jupyter lab data analysis file.

Accession numbers:

PDB: 1R4I, 1T7T, 1I27.

3 | RESULTS AND DISCUSSIONS

3.1 | AR-NTD folds into two disjoint highly dynamic regions

First, we performed molecular dynamics simulations of the AR-NTD in an aqueous solution with physiological salt concentration. The initial structure, shown in Figure 1a, was built using I-TASSER server.[38] The model was chosen as the best ranked model obtained from the server. The model was superior to conformations predicted by the AlphaFold and RosettaFold severs, based on confidence measures and comparative analysis with experimental data (Figure S2). Comparison between α-helix content of the models predicted by different algorithms with Circular Dichroism analysis of the NTD[18] revealed that among the predicted models, the best ranked model from I-TASSER and the AlphaFold model showed closer values to CD data. However, given the very low confidence level of the AlphaFold model, the best ranked model obtained from I-TASSER was chosen for performing MD simulations.

After minimization and relaxation, we performed MD simulations of the NTD structure. Visual examination of the trajectory and RMSD plots show that the initial conformation has undergone an extensive structural change (Figures S3 a and 1b). Figure 1c provides a kymograph of all the secondary structures within the NTD, which is mostly composed of random coils, turns, and loops despite the apparent changes in the global 3D shape of the system (Figure S3 b and c). We repeated the MD simulation three times using different initial velocities to ensure that our observations are not dependent on the initial conditions. Importantly, all three independent runs consistently led to the emergence of two disjoint regions in 3D within 2 μs of simulations: an extended N-terminal sub-region (NR, residues 1–224), and a C-terminal sub-region (CR, residues 225–538). Both sub-regions showed high level disorder and high solvent accessibility, which reached a plateau after 2 μs of simulations (Figure 1d,e). We monitored the dynamics for an additional 3 μs and computed root-mean-square fluctuations (RMSF) to quantify the fluctuations of the chain. We observed an order of magnitude larger RMSF for AR-NTD in comparison to the folded AR-LBD (Figure 1f, dashed line at 0.15 nm). Considering the average fluctuations of the chain we observed lower values of RMSF in the vicinity of residue 400. Quantification of solvent accessibility revealed that both CR and NR are highly solvent accessible. The CR however has a central segment (residues 355–469) which is relatively less solvent accessible (Figures 1e and S4), corresponding to the reduction seen in the RMSF profile. Disorder prediction data produced by PONDR analysis agrees with the solvent accessibility and RMSF profiles, with the central region within CR having less disorder than the rest of the CR (Figure S5). We denote this segment as *CR core*, to differentiate it from the rest of the CR, *CR shell*. Interestingly the core includes the well-known transcriptional activation unit 5 (Tau-5) motif of the AR-NTD. Comparing the radii of gyration of CR and NR, one can clearly see that the CR is significantly more compact than NR (Figure 1g). This is consistent with our model that CR, despite its disorder, carries residual structures in its core.

To further demonstrate that the formation of disjoint regions does not depend on the choice of initial configurations, we melted the compact NTD at 350 K and then

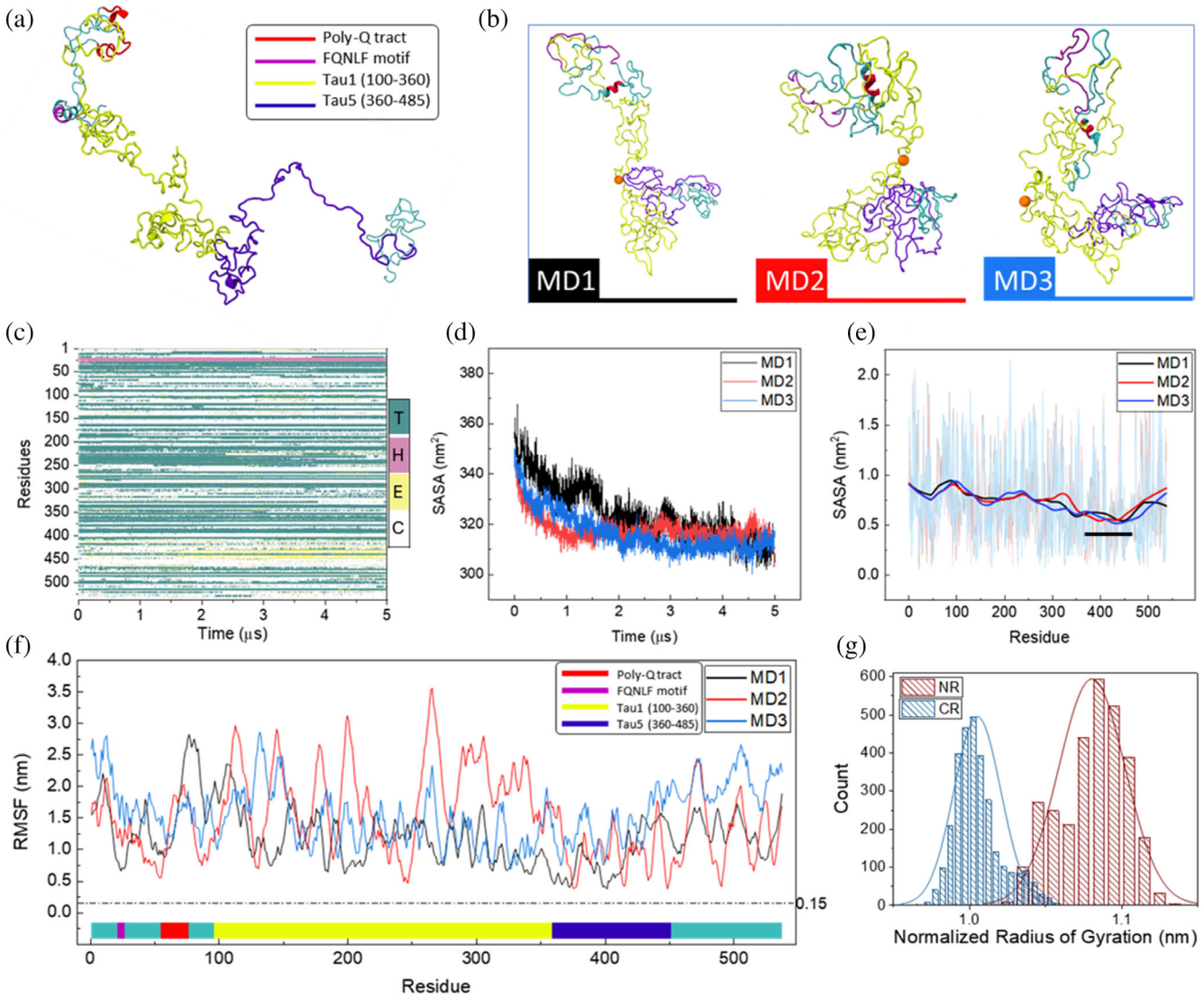


FIGURE 1 (a) Initial structure of the AR-NTD, cartoon representation is colored based on different functional motifs including PolyQ (residues 58–80) which is colored in red, FQNLF (residues 23–27) in purple, Trans activation unit 1 (Tau-1, residues 100–360) in yellow, and Trans activation unit 5 (Tau-5, residues 360–485) in blue. (b) three representative conformations from the last 10 ns of each replicate of the MD simulations. (c) Secondary structure evolution of NTD during 5 μs of simulation. With T, H, E, and C representing Turns, Helix, Strand, and Random Coil, respectively. (d) Time evolution of the solvent accessible surface area (SASA) of NTD. For all three replicates of the simulation, SASA values were decreased due to the overall compaction of the structure. (e) Solvent accessible surface area calculated for each residue during the last 3 μs of the simulation. Small black line indicates CR core (residues 355–469). (f) Average root mean square fluctuations of AR-NTD calculated per residue over the last 3 μs of the simulation. The dashed line indicates the average RMSF of folded AR-LBD (0.15 nm). (g) The size-normalized radius of the CR and NR regions during the last 3 μs of the simulation.

ran the simulation for 1 μs at physiological temperature. Melting opened up the compact structures and led to disappearance of separation between regions. The subsequent simulations, however, recovered the NR/CR in all three experiments and the low solvent accessibility of the CR core. Moreover, we find that our observation of NTD folding to two disjoint regions does not depend on the *F*-parameter of the force field, which encodes the interactions between the polypeptide and the surrounding solvent (Figure S6). We note that compaction of the chain to form the CR/NR segregation and the emergence of the CR core are AR-NTD specific, as synthetic models of AR-NTD using large poly-glutamine (polyQ) or poly-alanine (polyA) stretches, generated by mutating all residues of the structure shown in Figure 1a to Q or A respectively, did not produce these features (Figure S7, S8).

It is noteworthy that we have used these two models as control analysis to better understand the complex behavior of the AR-NTD (particularly during the first 1 μs). The synthetic polyQ model used to demonstrate the

sequence specificity is not to be confused with a natural short polyQ motif that is embedded within the AR-NTD. Q and A residues are known as disorder promoting and order promoting amino acids and were thus chosen to build these synthetic models with extreme behaviours. Although it has recently been shown that short lengths of the polyQ tract (up to 25 residues) can form a stable alpha helical structure[14] there is evidence that extended lengths of the polyQ chain show a largely disorderd structure,[52] while polyA shows a high propensity to form ordered structures.[53]

3.2 | Mapping CR folding dynamics to the topology landscape showed a directional conformational evolution

A natural way to express how the NTD peptide chain folds dynamically is to describe its topological folding patterns over time. In contrast to the conventional geometric approaches, topological approaches are specifically designed to quantify and categorize residual structural (shape) properties of constantly deforming polypeptides. Thus, we decided to map the chain dynamics onto the chain topology space, to identify patterns or motifs of the folding process.[54] Contacts in the chain are defined based on the spatial proximity of residues. The arrangement of residue–residue contacts provides a first-order topological representation of a folded chain, which can be highly informative on folding processes.[25] Here, we [46] exploited the CT framework to provide a dynamic topological description of a disordered protein for the first time. In order to map the topological dynamics of the NTD over time, we calculated the percentages of P, S, and X in chain conformations at 5 ns time points with a total of 1,000 frames over 5 μs. In this way, we followed the topological dynamics of the chain in time (Figure S9).

The topological analysis revealed critical differences in the folding process of NR and CR. While a clear topological trajectory is present for CR, the NR topology map is not showing any readily visible pathway. As seen in Figure S9 c, the starting conformation of CR was rich in series (S) but low in parallel (P) and cross (X) configurations. However, over time, while the percentage of P was almost constant, major changes were observed for S and X content, over all three runs. This topological profile is consistent with the observed emergence of the CR core. Ideally, an unfolded or semi-unfolded chain will present local contacts at first,[55] which have little interaction with far away residues along the chain. Non-interacting contacts belong by definition to the S topological configuration. However, as the folding process develops, the chain will fold onto itself, creating long-range interactions between residues, with a more complex topological profile. This process is true for any folding process, as exemplified by the topological evolution of our polyQ model of AR-NTD, see Methods section for details about the polyQ model.

In contrast to the CR topological trajectory, the NR region explores the topological space in what appears to be a random fashion. As expected from such a highly dynamic region, the opportunities of forming and breaking contacts with high frequency leads to a very wide configurational space. This results in a globular pattern in the topological space. The lack of a preferred direction for topological evolution is nicely visualized by the width and overall shape of the pattern. This is in contrast with the narrow topological evolution of CR, which indicates a number of transient states with significantly different configurations: once a new configuration is reached, the set of possible contacts changes. This phenomenon can be visualized on the topological space as a "leap" from one cluster to another, resulting in a narrow trajectory with high directionality. The lower percentage of series relations relative to the CR indicates that the NR evolution is truly dynamic in nature, with long-range contacts continuously forming and breaking and the absence of sub-structures. Indeed, contacts internal to a sub-structure are most likely in series with contacts embedded into another sub-structure. This motivated us to look more closely into the topological dynamics of NR and search for patterns using our circuit topology approach.

3.3 | Dynamic CT analysis reveals temporal regimes for topological development

Dynamic topological analysis of the AR-NTD shows the presence of three different time regimes for topological evolution, characterized by the time scales of residue–residue contact lifetimes. We envisioned a dynamic CT-based topological analysis to identify the time scale of topological evolution of the two previously identified NTD sub-regions, NR and CR. This, it is possible to follow the temporal evolution of every residue–residue contact formed at any moment of the MD simulation by plotting a kymograph of all possible residue-residue contact combinations (Figure 2a). By looking at the kymograph (and zooming in on regions of interest), we see how some contacts are fairly short-lived, often breaking and reforming with high frequency. On the other hand, other contacts live for a few microseconds. Therefore, we might have one (or more) lifetimes for each residue–residue contact: by picking the maximum lifetime of each

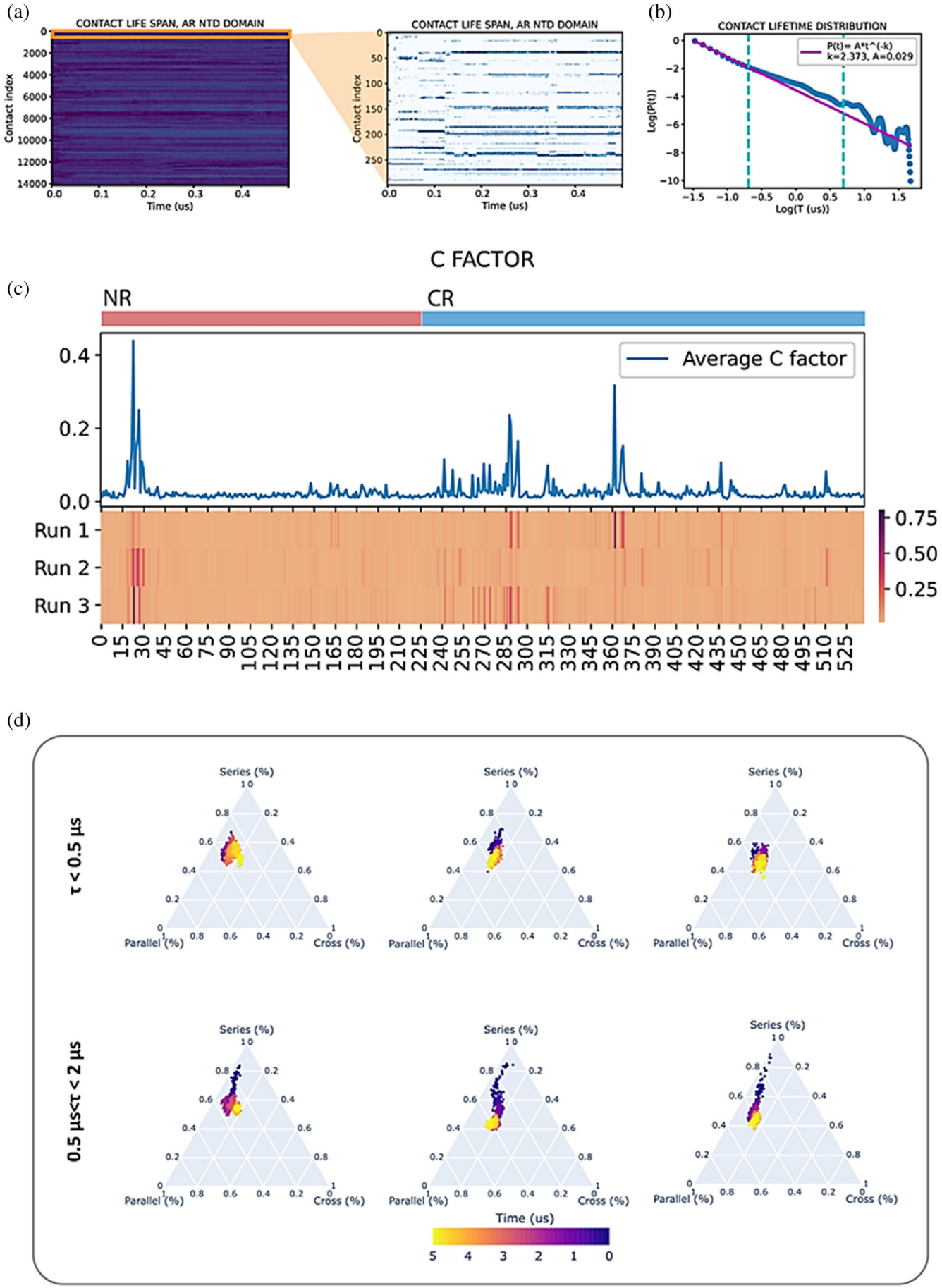


FIGURE 2 Legend on next page.

particular residue combination, we can build a lifetime distribution of all contacts (Figure 2b).

As expected by such a dynamic and heterogeneous system, contact lifetimes span across the whole range of the simulation, with the vast majority of them being short lived (lifetime distribution maxima are 35, 18, and 15 ns, respectively for the three runs). The distribution in Figure 2b respects to some extent the well-known scale free law ($P(\tau) = A\tau^{-\gamma}$), indicating there might not be a preferential duration for contact lifetimes.[56] We see, however, that the degree with which the lifetime data match the power law is not the same for all time scales: we identify a "short life" range ($\tau < 0.5\,\mu s$), where matching is maximum. After 0.5 μs the distribution starts deviating from the scale-free behavior ($0.5\,\mu s < \tau < 2\,\mu s$, "middle life"). After 2 μs, instability prevails, and deviations from the law appear with wide fluctuations, probably due to poor fitting of the distribution caused by the lack of statistics. The scale free distribution might represent a useful ideal model to quantify the behavior of IDPs. The adherence to this law is to a somewhat higher degree present for our polyQ case study (Figure S10). This regime behavior is highly conserved over all three runs, for both NR and CR (Figure S11). We can calculate the C-factor for each residue, that is, the average contact lifetime among all contacts in which the residue participates (Figures 2c, S12,13). Here, this heterogeneity becomes apparent. If we exclude residue 23–27, where the only alpha helix present within the NTD resides, higher lifetimes are way more prevalent in the CR region, indicating higher stability and compaction with respect to the NR region. This stark difference in behavior between the two regions cannot be replicated in the synthetic polyQ (Figure S10), indicating this asymmetry is specific to AR-NTD.

The lifetime analysis allows the identification of connectivity properties within the NTD under three different time regimes. This information can be exploited to disentangle the role of fast forming, short-lived contacts from that of longer-lived contacts during topological evolution. Next, we applied a lifetime filter to NTD contact maps and repeated the topological analysis. Results for short and middle life NR contacts can be seen in Figure 2d. Short life contacts seem overall to occupy only one topological state, over all the simulation, which is represented as a globule in the triangular S/P/X topological space. While these contacts oscillate around what looks like a wide topological equilibrium state, with no preferred direction for topological evolution, middle life contacts hop from one state to the other, creating narrow trajectories in the topological space. It is interesting to point out how this narrowing of the topological evolution pattern for middle lifetime contacts is not as evident in the synthetic polyQ model. This indicates there is some AR-specific coding within its secondary structures that leads to a well-defined configurational change within the middle lifetime. This is contrary to the view of the NTD as having a completely random conformation. The three runs show a clear pattern towards higher fractions of cross relation, indicating an evolution towards higher degrees of entanglement. The topology of long-life contacts is shown in Figure S14. As the scale $\tau > 2$ μs is close to the duration of the simulation, this scale is subject to heterogeneous behavior caused by kinetic traps, which manifest in the topological space as globules similar to those we have encountered for short lifetime. The same analysis carried out on CR (Figure S15) reveals a similar topological evolution, although the topological make-up of the two regions is quite different. While NR and CR have similar cross fractions, NR consistently scores higher in parallel, and the CR in series fractions.

In the context of lifetime analysis, one can consider following the evolution of a specific set of residues, and selectively visualize their contact formation dynamics. For example, cysteine residues play an important role in the folding and conformational dynamics of the proteins by forming disulfide bonds and covalently connecting two distinct parts of the chain.[57] There are 11 cysteine residues in AR-NTD, which make up for a total of 55 possible cys-cys contacts. Interestingly, only five of such possible contacts actually occur across our simulation runs. These contacts are by nature very dynamic and often form and break multiple times during a run; however, it is possible to calculate the cumulative time a certain contact is detected across frames in the simulation, in order to see which configurations are most likely to occur. Among the five outstanding contacts, we found contacts ^{239}Cys-Cys^{266}, ^{239}Cys-Cys^{326} in which Cys239 stand out as a main contributor. This agrees with previous finding that Cys239 plays an important role in oxidative hemostasis of the AR and aggregation behaviour of the chain.[58]

FIGURE 2 (a) Contact lifetime analysis reveals temporal regimes for topological evolution. A NTD contact lifetime kymograph, with zoomed detail of the first 300 contacts. (b) Log–log plot distribution of contact lifetime (maximum lifetime was selected for each contact) with power law fit. The dashed lines indicate the thresholds delimiting the three regimes: short, middle, and long lifetime. (c) Heatmap of C factor (average contact lifetime per residue) for all three runs, with plot of the C factor averaged over the three runs. (d) Triangular plots of the S/P/X topological space for NR, over all three runs, filtered for short and middle life contacts.

One might also consider including spatial information encoded in sequence separation, to shed a light on these topological differences. For this purpose, we calculate the average range of contacts for each residue. The distribution $P(\mathrm{d})$ of such spatial range (Figure S16 a) peaks at about 50 residue–residue distances, for both regions and all MD runs, indicating once again the high dynamicity of this domain. The most probable range of interaction in NTD is already larger than what is commonly identified as interaction range threshold defined for long-range contacts in proteins (typically 24 residues). Thus, there is a need for defining a relevant threshold for contact range. Based on these distributions, we identified a threshold for long-range contacts ($\lambda \geq <d> + \sigma_d \sim 82$ residues). The NR region consistently shows a higher density of such long-range contacts (Figure S16 b). This might explain the different topological profiles observed for the two regions: short range contacts tend to be in series with each other, as the contact sites are too close together to form entangled loops. Moreover, NR long-range contacts do not display series fraction at all (Figure S16 c), which might indicate the propensity of the region to consistently interact with itself along its full length (i.e., without creating self-interacting sub-domains). Furthermore, middle life long-range contacts do not consistently show the trajectories in the topological space, which we have identified in Figure 2d. This excludes long-range contacts as main drivers for conformational change, which therefore happens for smaller spatial scales.

Finally, we note that topological arrangements of contacts have implications for overall shape of the NTD domain. We sampled P-rich, S-rich, and X-rich topologies from topology space of middle-life contacts from last 500 ns of each simulation. Aligning the structures, we found highly dynamic region composed of residues 50–140 with an obvious structural shift in X-rich conformations (Figure S17). This also was approved by our RMSD analysis which showed a larger deviation for X-rich conformations in comparison to P- and S-rich ones. The most obvious finding to emerge from the analysis is that high content of X configuration had a key role in this large structural deviation. In the following section, we take a complementary approach and first identify the NTD shape prototypes and will subsequently study their differences in topology and function.

3.4 | AR-NTD adopts several prototypic shapes capable of targeting AR-LBD

Our spatial and temporal analysis study shows that AR-NTD is dynamic, yet there are regional differences in adopting certain topologies. As a next step, we asked whether these topological transitions contribute globally to the overall shape of the NTD, and whether this would have any impacts on the so-called N/C interaction which is mediated by ^{23}FQNLF^{27} motif and a hydrophobic pocket on the surface of the AR-LBD. To address this question, we modeled the dynamics of the protein after regional compaction was achieved and sampled the protein conformations. From the sampled conformations, we identified six different classes by performing structural clustering. Figure 3a represents the clusters together with their corresponding occupancy. Looking closer at the circuit topology of these structures, we could categorize them in three groups for whole NTD and two groups for NR region (Figure S18). Structures from the same circuit topology groups were similar in global shape and distinct from other groups. One can readily map the well-known NTD motifs onto the surface of these structures. The motifs are surrounded by residues that may change during shape transitions (Figure S19). For example, a motif, which is surrounded by positive charges in one NTD conformation, may be surrounded by negative charges in a different shape state. These transitions could be the regulatory mechanism for the binding interactions of these motifs with their partners and be important for their function.

Strikingly, a cleft forms between the CR and NR regions to adopt orientations that resemble opening and closing movements within the AR-NTD structure (Figure 3a). We performed docking of the identified NTD structures (Figure S20) with LBD (PDB code: 1T7T) and found that the shape of the complex shows striking resemblance to the recently published cryo-EM image of the full-length AR.[4] Particularly, the highest score docking pose from cluster 2 was nearly identical to the reported experimental structure (Figure S21). Interestingly, in the highest score docking pose from cluster 4 (Figure 3b), the ^{23}FQNLF^{27} motif is located at a close distance to the AF-2 pocket, which agrees with the well-established models of N/C interactions.[1] Studying best docking poses from all six clusters (a total of 600 poses) reveals that LBD interacts with NTD primarily via the NR region. This interaction dominantly formed either with only NR region or at the cleft. In the latter cases, we do find poses in which the ^{433}WHTLF^{437} motif is in close proximity with the AF-2 site of the LBD. Nonetheless, we do find a small number of clusters where the LBD also interacts with NTD-CR part only.

3.5 | AR NTD-DBD interaction suggests a regulatory role via two distinct binding modes

The AR DNA-binding domain (DBD) bind to AR response elements, and through this mediates the critical

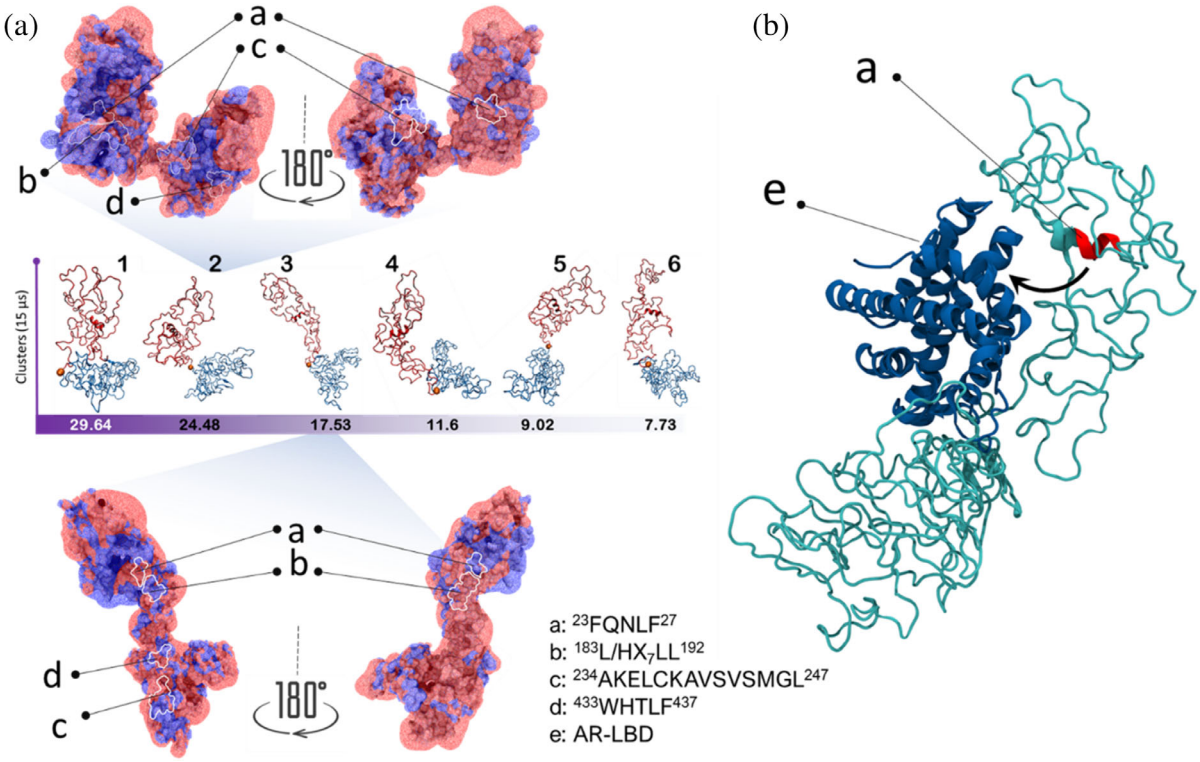


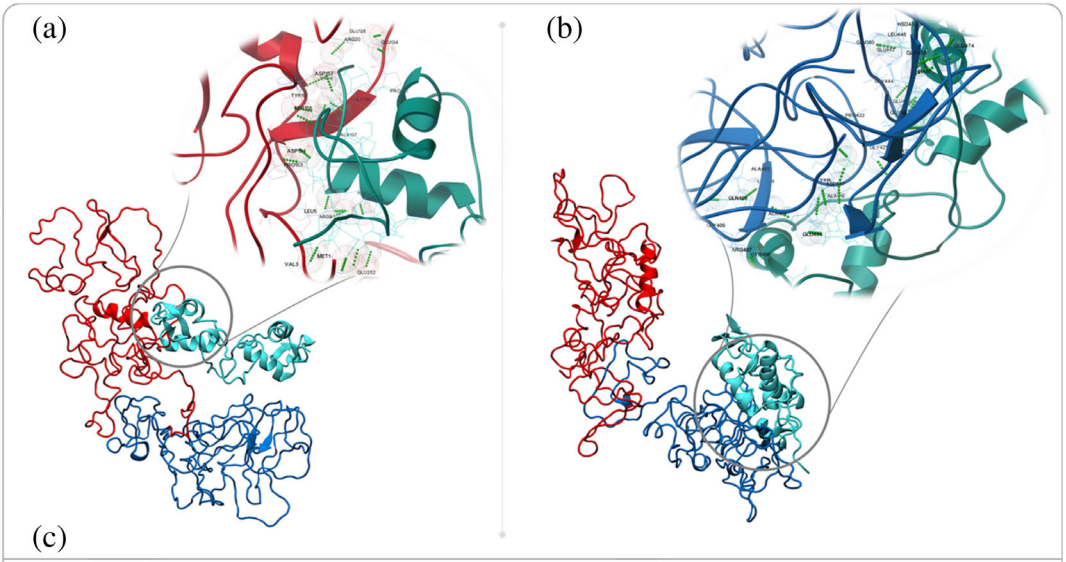
FIGURE 3 (a) Structures 1–6 are representative AR-NTD conformations identified by MD simulation and subsequent conformational sampling and clustering of the three trajectories which led to six distinct clusters. Note that the NR-CR cleft opens and closes as can be best seen in structures 3 and 2, respectively. Surface electrostatic potentials are calculated and mapped as wireframe surface in two colors, red for negative and blue for positive charges. (b) AR-LBD in complex with representative structure of the cluster 2. LBD docking was performed on all six structures shown in panel a and the highest ranked poses are shown in Figure S20. The identified NTD-LBD assemblies resemble the recently reported Cryo-EM images of AR-NTD[4] (Figure S21).

role in the AR transcriptional activity.[59] We investigated whether and how the identified AR-NTD conformations could interact with DBD. We used the previously reported crystal structure of the DBD (PDB code: 1R4I) and docked it against the AR-NTD conformations we have identified, as shown in Figure 4a,b. For each AR-NTD conformation, highest ranked poses (ranked by cluster size) were selected, and their interaction interfaces were analyzed (Figures 4 and S22). Our data revealed three interaction modes namely (1) interaction via NR, (2) interaction via CR, and (3) interactions that involves the cleft and neighboring NR/CR regions.

Strikingly, CR-mediated binding typically covers DNA-binding surface of DBD (residues His553, Tyr554, Ser561, Val564, Arg568, Tyr576, Arg591, Lys592, and Pro595), while NR-mediated interactions leave this surface exposed. When DBD interacts with NR region both D and P boxes surfaces are accessible. DBD interactions with CR region limited the access to the P box, which is responsible for DNA binding, however, the D box remained exposed and accessible for dimerization. This suggests a model in which CR-mediated NTD-DBD interactions could occlude DNA binding, while NR-mediated interaction would allow for DNA binding. Figure 4c summarizes NTD–DBD interaction modes and their overlap with DNA-binding site. Our analysis also showed that no cysteine residues were involved in these interactions.

3.6 | Compact CR binds to an EPI drug

Due to its medical importance, developing drugs that target AR NTD is an area of intense research; however, the conformational disorder significantly hinders drug development. Only recently, EPI-001 has been developed as the first inhibitor of the AR-NTD,[61] yet we do not know how the drug molecule interacts with the receptor. Interestingly, biochemical studies revealed that the binding happens in the CR part of the AR-NTD.[52] To study the drug binding, we clustered the CR region after the regional compaction (last 3 μs) and identified three representative structures. Note that the clusters found in Figure 3a are the result of clustering of the whole NTD and the NR contributes to the heterogeneity due to its dynamics. It is thus expected that CR clustering would lead to a smaller number of representative structures.



(c)

	Pose 1	Pose 2	Pose 3	Pose 4	Pose 5	Pose 6
Cluster 1	NR	NR/CR	NR	CR	CR	NR
Cluster 2	NR/CR	CR	NR/CR	CR	CR	NR/CR
Cluster 3	CR	CR	CR	CR	CR	CR
Cluster 4	CR	CR	CR	CR	CR	CR
Cluster 5	CR	CR	NR	NR/CR	CR	NR/CR
Cluster 6	CR	CR	CR	CR	CR	NR/CR

FIGURE 4 (a), (b) Docking of AR-DBD to AR-NTD conformation 1 shown in Figure 4a revealed two binding modes, mediated by NR or CR regions, respectively. In (a) DNA-binding surface of DBD is exposed and accessible for target DNA while in (b) DNA-binding surface is fully occupied upon interaction with CR. In detailed analysis of each interacting mode of the DBD with NTD showed that when DBD interacts with NR region both D and P boxes surfaces are accessible. On the other hand, DBD interactions with CR region limited the access to the P box which is responsible for DNA binding.[60] (c) AR-DBD docking and interface analysis have been performed for six representative NTD conformations shown in Figure 4a and six highest ranked poses for each conformation are tabulated in (c). When DBD-NTD interactions involved the DNA-binding surface of DBD, the interacting NTD regions are colored in red; otherwise, the NTD regions are colored in green.

Modeling the three CR conformations show how the core is enclosed by the shell and how the dynamic shell can mask or expose the core. We performed docking study on the identified CR structures and the EPI-001 drug and identified the most probable binding modes (Figure 5). The most probable drug-CR complex is established by three hydrogen bonds. No π-π or T-shaped π interactions were observed between the EPI-001 and the CR region of the NTD. In this binding mode, more than 90% of the interacting residues are located in the Tau-5 region, which suggests a high affinity of the drug to this key functional motif, in agreement with previous reports.[62,63] NMR analysis suggested that residues 354–448 are essential for drug binding,[62] which coincides with core of CR. This is consistent with our most probable model (interacting with residues 372–415). Furthermore, we see that residues from CR shell are also involved in drug binding (residues 527 and 529). This has not been seen in NMR study as the NTD construct used in those studies (residues 142–448) was shorter than the wild-type and so the shell interaction could not be probed. Moreover, we studied other likely modes and, interestingly, the third CR clusters also showed binding configurations that engage both the CR core (interacting with residues 374–381) and the CR shell (residues 348 and 450) which has a complete overlap with Tau-5 that is consistent with the NMR data. However second CR cluster showed interactions only with CR shell (residues 492–535) which involves NTD parts not evaluated in the NMR study (and thus cannot be compared).

As a case study to demonstrate the drug effect, we asked whether our identified drug-NTD interaction can interfere with AR-NTD for RAP74 interactions. Previous research has found Tau-5 to be an important interaction site on the AR-NTD for RAP74 interactions.[7] Tau-5 is found to bind helices 2, 2.5, and 3 on the C-terminal domain of RAP74 (RAP74-CTD), which form a groove where AR can bind.[2,7] Importantly, our docking analysis shows that RAP74-CTD binds the highest ranked CR structure and the binding site overlaps both with the Tau-5 and the aforementioned EPI-001 drug-binding sites (Figure S23). We see a consistent picture with other CR

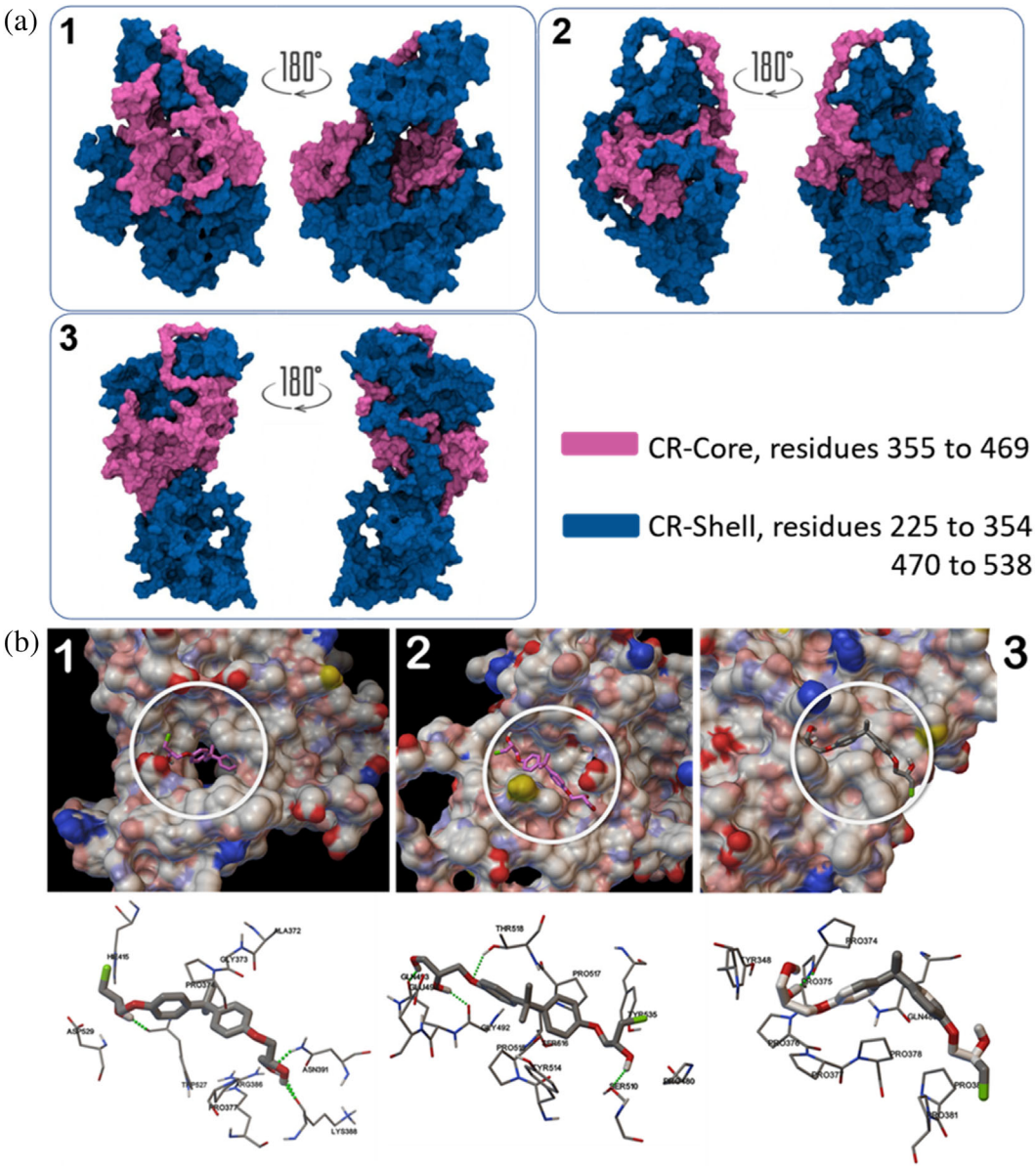


FIGURE 5 (a) Representative members of three clusters obtained for C-terminal region (CR). Clustering was done on sampled conformations from the last 3 μs of the simulations, based exclusively on the dynamics of CR region. CR regions are shown in surface representation and CR-core (residues 355–469) and CR-Shell (residues 255–354 and 470–538) are colored in pink and blue, respectively. (b) Surface palatability map representations of CR regions from three clusters that show the tri-dimensional positions of EPI-001 interacting with distinct regions on the surface of the protein. As it can be seen from the line representations, at least one hydrogen bond is formed between drug molecule and corresponding residues on the surface of the CR. Despite the slightly different interaction sites, drug molecule in all three clusters is interacting with residues that are located within, or adjacent to Tau-5 functional motif of the N-terminal transactivation domain.

models as well. The data clearly suggest that the drug binding can interfere with RAP74-CTD binding, in agreement with previous studies.

4 | CONCLUSION

Our study shows that the AR-NTD, despite being a disordered chain, can adopt certain topologies and region-specific behavior. Its structure can be subdivided into two C-terminal and N-terminal sub-regions (CR and NR). The identified topologies, when mapped against known AR interaction sites are significant in how they affect our understanding of the LBD and DBD, as well as drug-binding interactions that regulate important AR functions. Our structural data are consistent with data from hydrophobic fluorescence probe binding and size-exclusion chromatography which first suggested the

possibility of native AR-NTD existing in a collapsed disordered conformation, distinct from extended disordered (random coil) and a stable globular fold.[64] Here, for the first time, we present structural and topological evidence for those speculative models, while also finding significant differences in dynamic behavior between NR and CR. Our observed overall NTD shape dynamics enables the domain to bind to the LBD domain in a way that agrees with recent Cryo-EM data of the full-length AR. Strikingly, our structural data also suggest that AR-NTD can regulate DNA-binding enabled through competing CR- and NR-mediated DBD-binding interactions. This is in agreement with earlier biochemical studies, demonstrating that the AR-NTD can alter DNA-binding affinity[65] and the region immediately N-terminal to the DBD inhibited DNA binding.[66] Furthermore, in vitro protein–protein interaction (GST-pull down, unpublished observation) assays show that the removal of the first 187 amino acids of NTD (which corresponds to NR), increases the interaction of the truncated NTD polypeptide with the DNA-binding domain, suggesting that NR prevents an inhibitory function of CR on DNA binding. Our data suggest a model for CR in which a dynamic shell may enclose or expose the core and thus regulate its binding to cellular partners. Finally, the modeling analysis allowed for the identification of a previously unknown binding site in the CR core and shell of the AR-NTD, revealing also the functional motifs to which the EPI-001 drug binds, and how the interaction interferes with co-regulatory RAP74 protein binding to NTD.

Our NTD analysis allowed us to build a model for wild-type human AR. Figure 6 shows a model constructed using multiple structure alignment approach. Figure 6 also summarizes the key interactions and regulatory mechanisms, suggested by our data. This includes: (1) the well-established N/C interaction, (2) NTD-mediated regulation of DNA binding, and (3) drug-induced inhibition of RAP74 binding to NTD. All these findings are in close agreement with experimental data, as discussed above. The model can open up possibilities for future investigations of AR biology and disease analysis. For example, we speculate that post-translational

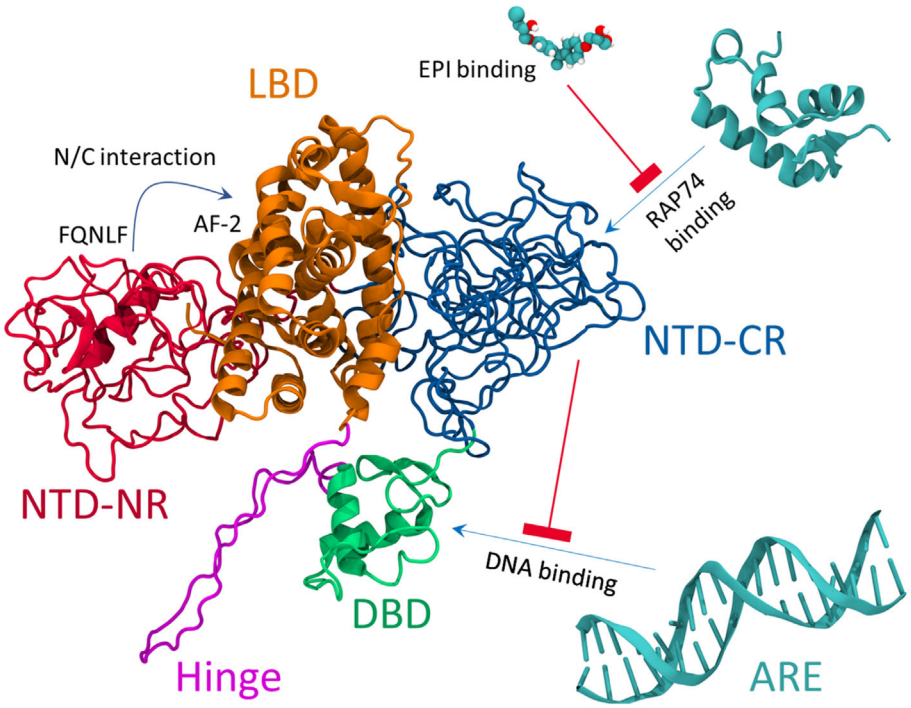


FIGURE 6 Full length AR structure constructed based on our N-terminal transactivation domain (NTD) model. In order to construct the full length AR model, full length AR structure was obtained from AlphaFold, the NTD domain (residues 1–538) was cleaved, and the rest of the AlphaFold model was aligned to our NTD–LBD complex. Connections later were made using Chimera software. Our full-length model shows the expected close proximity between the FQNLF motif on NTD and the AF-2 pocket of the AR-LBD which are the main mediators of the formation of a well-established interaction in AR known as N/C interaction. The spatial localization of the DBD with respect to the CR region is in a way that it allows DBD to move and interact with the CR region. Our model also clearly shows that the hinge region and dimerization interface of the LBD are both accessible for interaction with partner proteins. Finally, we do not see any spatial limitation for the binding of EPI-001 drug or RAP74-CTD to the CR region, in line with the proposed mechanism of action of RAP74 and the inhibitory EPI-001 drug.

modification sites as well as pathogenic mutations[67–69] are typically residing in CR and in close proximity to long life-time attractive contacts[70]; a hypothesis that can be readily addressed in follow-up studies. Furthermore, we note that the hinge region is highly exposed in the proposed model. A growing number of evidence has shown that the hinge region plays a crucial role in regulating AR function.[71] Besides mediating nuclear translocation and DNA binding, the AR hinge region acts as a target for post-translational modifications[67] and can recruit some co-regulatory proteins to affect AR activity.[72] Structural, biochemical, and cell-based analysis are required to verify the proposed interactions and mechanistic models, and their relevance to in vivo biology.

Our article sets the stage for future investigations of AR biology, and provides a workflow for analysis of other NHRs. In our current study, we focused on NTD dynamics in isolation. An important future goal would be to study the impact of interdomain interactions, and the role homo/heterodimerization and co-regulatory protein binding have on the topological dynamics of AR-NTD.[60] Highly regulated and complex interdomain interactions in AR and other NHRs could be an example of how cross-talking between structured domains and disordered regions can regulate complicated cellular functions. Despite differences in primary sequence and thus many structural properties, NTDs of NHRs share certain features and thus comparing our data with experimental data from other NHRs might also be informative.[73–75] In agreement to our data, previous experiments suggest that NTD of NHRs are largely disordered. However, importantly intrinsically disordered structure can range from fully unfolded to partially folded ensembles and in the case of the AR and the mineralocorticoid receptor[76] NTD regions with more stable structure (2, 7 and present study). One can of course apply the proposed computational analysis workflow on these NHRs and investigate topological similarities and differences between various NHR proteins.

We proposed a novel approach, termed dynamic circuit topology (dCT), to study protein dynamics from a topological perspective. Conformational disorder is one of the main challenges in structural biology,[77] and yet, disordered proteins represent very attractive drug targets.[78] Previous attempts at topological description mainly focused on the protein native conformation,[79–84] a choice inappropriate in this context due to the highly dynamic nature of AR-NTD. Dynamic systems require dynamic approaches and here, we have applied a multi time-scale analysis to highlight the topological evolution of the system, revealing that the most apparent (elongated) trajectories in the conformational space happens between 0.5 and 2 μs, while the vast majority of AR-NTD contacts are formed and broken with a lifetime shorter than 0.5 μs. A major technical advancement is the representation of the folding process on the topological landscape. Expressing conformational evolutions in terms of S, P, and X allows for clear visualization and quantification of topological evolution. This reveals that topological trajectories are developed by hopping from one configuration to the other. This happens generally in the direction of decreasing series and increasing parallel and (especially) cross fractions. Although both NR and CR display similar qualitative behavior in this respect, quantitatively they are very different, both from a timescale and a topological point of view. While a narrow topological trajectory is clearly visible for CR without the need of dynamic analysis, it is necessary to disentangle the role of short-lived contacts to be able to visualize a trajectory for NR. Moreover, the differences in topological make up between the two regions allow us to speculate about how formation of sub-structures with high contact density arise in AR-NTD. The higher series fraction in CR suggests that transiently connected sub-structures may be formed, which are mostly independent of each other. The formation of such sub-structures might be necessary for folding and target binding. On the other hand, the longer range of residue–residue interactions in NR (and the lack of series contacts for long range contacts) is suggestive of very high flexibility of this region, with possible biological implications; it could be functional, for example, for rapid opening and closing movements in the structure. Overall, we envision that the contact analysis methodology used in this study paves the way for more in depth dynamic structural/topological studies on disordered proteins including other NHRs, creating new opportunities for drug development.

AUTHOR CONTRIBUTIONS

Vahid Sheikhhassani: Formal analysis (lead); investigation (lead); writing – original draft (equal); writing – review and editing (equal). **Barbara Scalvini:** Formal analysis (equal); Investigation (equal); writing – original draft (equal); writing – review and editing (equal). **Julian Ng:** Resources (supporting); writing – review and editing (equal). **Laurens Heling:** Investigation (supporting); writing – review and editing (equal). **Yosri Ayache:** Investigation (supporting). **Tom Evers:** Investigation (supporting); writing – review and editing (supporting). **Eva Estébanez-Perpiñá:** Resources (supporting); writing – review and editing (supporting). **Iain McEwan:** Resources (supporting); writing – review and editing (supporting). **Alireza Mashaghi:** Conceptualization (lead); funding acquisition (lead); investigation (equal); supervision (lead); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGEMENTS

We are thankful to Eugene Shakhnovich (Harvard University) for critical reading of the manuscript, and to Peter Bolhuis and Ioana Ilie (University of Amsterdam) for technical discussions. The research in Mashaghi lab is supported by funding from Muscular Dystrophy Association (USA), Grant Number MDA628071, and Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) through NWA-IDG (NWA.1228.192.309) and Open Competition XS (OCENW.XS.076).

CONFLICT OF INTEREST

The authors declare no competing interests.

ORCID

Alireza Mashaghi https://orcid.org/0000-0002-2157-1211

REFERENCES

1. Schaufele F, Carbonell X, Guerbadot M, et al. The structural basis of androgen receptor activation: Intramolecular and intermolecular amino-carboxy interactions. Proc Natl Acad Sci USA. 2005;102:9802–9807. https://doi.org/10.1073/pnas.0408819102.
2. Lavery DN, McEwan IJ. Functional characterization of the native NH 2 -terminal transactivation domain of the human androgen receptor: Binding kinetics for interactions with TFIIF and SRC-1a. Biochemistry. 2008;47:3352–3359. https://doi.org/10.1021/bi702220p.
3. McEwan IJ. Intrinsic disorder in the androgen receptor: Identification, characterisation and drugability. Mol Biosyst. 2012;8:82–90. https://doi.org/10.1039/c1mb05249g.
4. Yu X, Yi P, Hamilton RA, et al. Structural insights of transcriptionally active, full-length androgen receptor coactivator complexes. Mol Cell. 2020;79:812–823.e4. https://doi.org/10.1016/j.molcel.2020.06.031.
5. Antonarakis ES, Lu C, Wang H, et al. AR-V7 and resistance to enzalutamide and Abiraterone in prostate cancer. N Engl J Med. 2014;371:1028–1038. https://doi.org/10.1056/NEJMoa1315815.
6. Wasmuth EV, Hoover EA, Antar A, Klinge S, Chen Y, Sawyers CL. Modulation of androgen receptor DNA binding activity through direct interaction with the ETS transcription factor ERG. Proc Natl Acad Sci USA. 2020;117:8584–8592. https://doi.org/10.1073/pnas.1922159117.
7. De Mol E, Szulc E, Di Sanza C, et al. Regulation of androgen receptor activity by transient interactions of its transactivation domain with general transcription regulators. Structure. 2018;26:145–152.e3. https://doi.org/10.1016/j.str.2017.11.007.
8. Peng S, Wang J, Chen H, et al. Regression of castration-resistant prostate cancer by a novel compound QW07 targeting androgen receptor N-terminal domain. Cell Biol Toxicol. 2020;36:399–416. https://doi.org/10.1007/s10565-020-09511-x.
9. Kuznik NC, Solozobova V, Jung N, et al. Development of a Benzothiazole scaffold-based androgen receptor N-terminal inhibitor for treating androgen-responsive prostate cancer. ACS Chem Biol. 2021;16:2103–2108. https://doi.org/10.1021/acschembio.1c00390.
10. Ban F, Leblanc E, Cavga AD, et al. Development of an androgen receptor inhibitor targeting the N-terminal domain of androgen receptor for treatment of castration resistant prostate cancer. Cancers (Basel). 2021;13:3488. https://doi.org/10.3390/cancers13143488.
11. Pauwels K, Lebrun P, Tompa P. To be disordered or not to be disordered: Is that still a question for proteins in the cell? Cell Mol Life Sci. 2017;74:3185–3204. https://doi.org/10.1007/s00018-017-2561-6.
12. Tompa P, Schad E, Tantos A, Kalmar L. Intrinsically disordered proteins: Emerging interaction specialists. Curr Opin Struct Biol. 2015;35:49–59. https://doi.org/10.1016/j.sbi.2015.08.009.
13. Na J-H, Lee W-K, Yu Y. How do we study the dynamic structure of unstructured proteins: A case study on Nopp140 as an example of a large, intrinsically disordered protein. Int J Mol Sci. 2018;19:381. https://doi.org/10.3390/ijms19020381.
14. Escobedo A, Topal B, Kunze MBA, et al. Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. Nat Commun. 2019;10:2034. https://doi.org/10.1038/s41467-019-09923-2.
15. Eftekharzadeh B, Banduseela VC, Chiesa G, et al. Hsp70 and Hsp40 inhibit an inter-domain interaction necessary for transcriptional activity in the androgen receptor. Nat Commun. 2019;10:3562. https://doi.org/10.1038/s41467-019-11594-y.
16. Bielskutė S, Garcia-Cabau C, Frigolé-Vivas M, et al. Low amounts of heavy water increase the phase separation propensity of a fragment of the androgen receptor activation domain. Protein Sci. 2021;30:1427–1437. https://doi.org/10.1002/pro.4110.
17. Tunyasuvunakool K, Adler J, Wu Z, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021;596:590–596. https://doi.org/10.1038/s41586-021-03828-1.
18. Davies P, Watt K, Kelly SM, Clark C, Price NC, McEwan IJ. Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. J Mol Endocrinol. 2008;41:301–314. https://doi.org/10.1677/JME-08-0042.
19. Goundaroulis D, Dorier J, Benedetti F, Stasiak A. Studies of global and local entanglements of individual protein chains using the concept of knotoids. Sci Rep. 2017;7:6309. https://doi.org/10.1038/s41598-017-06649-3.
20. Mishra R, Bhushan S. Knot theory in understanding proteins. J Math Biol. 2012;65:1187. https://doi.org/10.1007/s00285-011-0488-3.
21. Golovnev A, Mashaghi A. Circuit Topology for Bottom-Up Engineering of Molecular Knots. Symmetry. 2021;13:2353. https://doi.org/10.3390/sym13122353
22. Golovnev A, Mashaghi A. Generalized Circuit Topology of Folded Linear Chains. iScience. 2020;23:101492. https://doi.org/10.1016/j.isci.2020.101492
23. Golovnev A, Mashaghi A. Topological analysis of folded linear molecular chains. Singapore: Topol. Polym. Chem., Springer Singapore, 2022;p. 105–114. https://doi.org/10.1007/978-981-16-6807-4_7.
24. Mashaghi A. Circuit topology of folded chains. Not Am Math Soc. 2021;68:420–423. https://doi.org/10.1090/noti2241.

25. Heidari M, Schiessel H, Mashaghi A. Circuit topology analysis of polymer folding reactions. ACS Cent Sci. 2020;6:839–847. https://doi.org/10.1021/acscentsci.0c00308.
26. Scalvini B, Sheikhhassani V, Mashaghi A. Topological principles of protein folding. Phys Chem Chem Phys. 2021;23: 21316–21328. https://doi.org/10.1039/d1cp03390e.
27. Nikoofard N, Mashaghi A. Topology sorting and characterization of folded polymers using nano-pores. Nanoscale. 2016;8: 4643–4649. https://doi.org/10.1039/c5nr08828c
28. Nikoofard N, Mashaghi A. Implications of Molecular Topology for Nanoscale Mechanical Unfolding. The Journal of Physical Chemistry B. 2018;122:9703–9712. https://doi.org/10.1021/acs.jpcb.8b09454
29. Mashaghi A, Ramezanpour A. Distance measures and evolution of polymer chains in their topological space. Soft Matter. 2015;11:6576–6585. https://doi.org/10.1039/c5sm01482d
30. Mashaghi A, Ramezanpour A. Circuit topology of linear polymers: a statistical mechanical treatment. RSC Advances. 2015; 5:51682–51689. https://doi.org/10.1039/c5ra08106h
31. Mugler A, Tans SJ, Mashaghi A. Circuit topology of self-interacting chains: implications for folding and unfolding dynamics. Phys. Chem. Chem. Phys. 2014;16:22537–22544. https://doi.org/10.1039/c4cp03402c
32. Heidari M, Satarifard V, Mashaghi A. Mapping a single-molecule folding process onto a topological space. Physical Chemistry Chemical Physics. 2019;21:20338–20345. https://doi.org/10.1039/c9cp03175h
33. Heidari M, Satarifard V, Tans SJ, Ejtehadi MR, Mashaghi S, Mashaghi A. Topology of internally constrained polymer chains. Physical Chemistry Chemical Physics. 2017;19:18389–18393. https://doi.org/10.1039/c7cp02145c
34. Satarifard V, Heidari M, Mashaghi S, Tans SJ, Ejtehadi MR, Mashaghi A. Topology of polymer chains under nanoscale confinement. Nanoscale. 2017;9:12170–12177. https://doi.org/10.1039/c7nr04220e
35. Buchanan G. Structural and functional consequences of glutamine tract variation in the androgen receptor. Hum Mol Genet. 2004;13:1677–1692. https://doi.org/10.1093/hmg/ddh181.
36. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature. 1991;352:77–79. https://doi.org/10.1038/352077a0.
37. Duff J, Davies P, Watt K, McEwan IJ. Structural dynamics of the human androgen receptor: Implications for prostate cancer and neurodegenerative disease. Biochem Soc Trans. 2006;34: 1098–1102. https://doi.org/10.1042/BST0341098.
38. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function prediction. Nat Methods. 2015;12:7–8. https://doi.org/10.1038/nmeth.3213.
39. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001.
40. Vriend G. WHAT IF: A molecular modeling and drug design program. J Mol Graph. 1990;8:52–56. https://doi.org/10.1016/0263-7855(90)80070-V.
41. Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014;42:W252–W258. https://doi.org/10.1093/nar/gku340.
42. Eisenberg D, Lüthy R, Bowie JU. [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol. 1997;277:396–404. https://doi.org/10.1016/S0076-6879(97)77022-8.
43. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics. 2011;27:343–350. https://doi.org/10.1093/bioinformatics/btq662.
44. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol. 2005;3:35–60. https://doi.org/10.1142/s0219720005000886.
45. Machado MR, Barrera EE, Klein F, Sóñora M, Silva S, Pantano S. The SIRAH 2.0 force field: Altius, Fortius, Citius. J Chem Theory Comput. 2019;15:2719–2733. https://doi.org/10.1021/acs.jctc.9b00006.
46. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comput Chem. 1997;18:1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
47. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126:014101. https://doi.org/10.1063/1.2408420.
48. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 1981;52:7182–7190. https://doi.org/10.1063/1.328693.
49. Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for protein–protein docking. Nat Protoc. 2017;12:255–278. https://doi.org/10.1038/nprot.2016.169.
50. Schüttelkopf AW, van Aalten DMF. PRODRG: A tool for high-throughput crystallography of protein–ligand complexes. Acta Crystallogr Sect D Biol Crystallogr. 2004;60:1355–1363. https://doi.org/10.1107/S0907444904011679.
51. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009;30:2785–2791. https://doi.org/10.1002/jcc.21256.
52. Wetzel R. Physical chemistry of Polyglutamine: Intriguing Tales of a monotonous sequence. J Mol Biol. 2012;421:466–490. https://doi.org/10.1016/j.jmb.2012.01.030.
53. Polling S, Ormsby AR, Wood RJ, et al. Polyalanine expansions drive a shift into α-helical clusters without amyloid-fibril formation. Nat Struct Mol Biol. 2015;22:1008–1015. https://doi.org/10.1038/nsmb.3127.
54. Mashaghi A, Van Wijk RJ, Tans SJ. Circuit topology of proteins and nucleic acids. Structure. 2014;22:1227–1237. https://doi.org/10.1016/j.str.2014.06.015.
55. Montelione GT, Scheraga HA. Formation of local structures in protein folding. Acc Chem Res. 1989;22:70–76. https://doi.org/10.1021/ar00158a004.
56. Albert R. Scale-free networks in cell biology. J Cell Sci. 2005; 118:4947–4957. https://doi.org/10.1242/jcs.02714.
57. Bhopatkar AA, Uversky VN, Rangachari V. Disorder and cysteines in proteins: A design for orchestration of conformational see-saw and modulatory functions. Prog Mol Biol Transl Sci. 2020;174:331–373. https://doi.org/10.1016/bs.pmbts.2020.06.001.

58. Oppong E, Stier G, Gaal M, et al. An Amyloidogenic sequence at the N-terminus of the androgen receptor impacts Polyglutamine aggregation. Biomolecules. 2017;7:44. https://doi.org/10.3390/biom7020044.
59. Verrijdt G, Tanner T, Moehren U, Callewaert L, Haelens A, Claessens F. The androgen receptor DNA-binding domain determines androgen selectivity of transcriptional response. Biochem Soc Trans. 2006;34:1089–1094. https://doi.org/10.1042/BST0341089.
60. Centenera MM, Harris JM, Tilley WD, Butler LM. Minireview: The contribution of different androgen receptor domains to receptor dimerization and signaling. Mol Endocrinol. 2008;22:2373–2382. https://doi.org/10.1210/me.2008-0017.
61. Yang YC, Banuelos CA, Mawji NR, et al. Targeting androgen receptor activation Function-1 with EPI to overcome resistance mechanisms in castration-resistant prostate cancer. Clin Cancer Res. 2016;22:4466–4477. https://doi.org/10.1158/1078-0432.CCR-15-2901.
62. De Mol E, Fenwick RB, Phang CTW, et al. EPI-001, a compound active against castration-resistant prostate cancer, targets transactivation unit 5 of the androgen receptor. ACS Chem Biol. 2016;11:2499–2505. https://doi.org/10.1021/acschembio.6b00182.
63. Myung J-K, Banuelos CA, Fernandez JG, et al. An androgen receptor N-terminal domain antagonist for treating prostate cancer. J Clin Invest. 2013;123:2948–2960. https://doi.org/10.1172/JCI66398.
64. Lavery DN, McEwan IJ. Structural characterization of the native NH 2 -terminal transactivation domain of the human androgen receptor: A collapsed disordered conformation underlies structural plasticity and protein-induced folding. Biochemistry. 2008;47:3360–3369. https://doi.org/10.1021/bi702221e.
65. Brodie J, McEwan IJ. Intra-domain communication between the N-terminal and DNA-binding domains of the androgen receptor: Modulation of androgen response element DNA binding. J Mol Endocrinol. 2005;34:603–615. https://doi.org/10.1677/jme.1.01723.
66. Liu G-Z, Wang H, Wang Z. Identification of a highly conserved domain in the androgen receptor that suppresses the DNA-binding domain-DNA interactions. J Biol Chem. 2003;278:14956–14960. https://doi.org/10.1074/jbc.M212229200.
67. Coffey K, Robson CN. Regulation of the androgen receptor by post-translational modifications. J Endocrinol. 2012;215:221–237. https://doi.org/10.1530/JOE-12-0238.
68. Kalfa N, Philibert P, Baskin LS, Sultan C. Hypospadias: Interactions between environment and genetics. Mol Cell Endocrinol. 2011;335:89–95. https://doi.org/10.1016/j.mce.2011.01.006.
69. Linja MJ, Visakorpi T. Alterations of androgen receptor in prostate cancer. J Steroid Biochem Mol Biol. 2004;92:255–264. https://doi.org/10.1016/j.jsbmb.2004.10.012.
70. Thomas PD, Dill KA. Statistical potentials extracted from protein structures: How accurate are they? J Mol Biol. 1996;257:457–469. https://doi.org/10.1006/jmbi.1996.0175.
71. Clinckemalie L, Vanderschueren D, Boonen S, Claessens F. The hinge region in androgen receptor control. Mol Cell Endocrinol. 2012;358:1–8. https://doi.org/10.1016/j.mce.2012.02.019.
72. Link KA, Balasubramaniam S, Sharma A, et al. Targeting the BAF57 SWI/SNF subunit in prostate cancer: A novel platform to control androgen receptor activity. Cancer Res. 2008;68:4551–4558. https://doi.org/10.1158/0008-5472.CAN-07-6392.
73. Zheng J, Corzo C, Chang MR, et al. Chemical crosslinking mass spectrometry reveals the conformational landscape of the activation helix of PPARγ; a model for ligand-dependent antagonism. Structure. 2018;26:1431–1439.e6. https://doi.org/10.1016/j.str.2018.07.007.
74. Holding AN. XL-MS: Protein cross-linking coupled with mass spectrometry. Methods. 2015;89:54–63. https://doi.org/10.1016/j.ymeth.2015.06.010.
75. Iacobucci C, Piotrowski C, Aebersold R, et al. First community-wide, comparative cross-linking mass spectrometry study. Anal Chem. 2019;91:6953–6961. https://doi.org/10.1021/acs.analchem.9b00658.
76. Fischer K, Kelly SM, Watt K, Price NC, McEwan IJ. Conformation of the mineralocorticoid receptor N-terminal domain: Evidence for induced and stable structure. Mol Endocrinol. 2010;24:1935–1948. https://doi.org/10.1210/me.2010-0005.
77. Chen CYC, Tou WI. How to design a drug for the disordered proteins? Drug Discov Today. 2013;18:910–915. https://doi.org/10.1016/j.drudis.2013.04.008.
78. Metallo SJ. Intrinsically disordered proteins are potential drug targets. Curr Opin Chem Biol. 2010;14:481–488. https://doi.org/10.1016/j.cbpa.2010.06.169.
79. Baiesi M, Orlandini E, Seno F, Trovato A. Exploring the correlation between the folding rates of proteins and the entanglement of their native states. J Phys A Math Theor. 2017;50:504001. https://doi.org/10.1088/1751-8121/aa97e7.
80. Baker D. A surprising simplicity to protein folding. Nature. 2000;405:39–42. https://doi.org/10.1038/35011000.
81. Panagiotou E, Plaxco KW. A topological study of protein folding kinetics. E Flapan & H Wong (eds), Topology and Geometry of Biopolymers. 2020;746:233–250. https://doi.org/10.1090/conm/746/15010.
82. Baiesi M, Orlandini E, Trovato A, Seno F. Linking in domain-swapped protein dimers. Sci Rep. 2016;6:1–11. https://doi.org/10.1038/srep33872.
83. Baiesi M, Orlandini E, Seno F, Trovato A. Sequence and structural patterns detected in entangled proteins reveal the importance of co-translational folding. Sci Rep. 2019;9:1–12. https://doi.org/10.1038/s41598-019-44928-3.
84. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol. 1998;277:985–994. https://doi.org/10.1006/jmbi.1998.1645.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Sheikhhassani V, Scalvini B, Ng J, Heling LWHJ, Ayache Y, Evers TMJ, et al. Topological dynamics of an intrinsically disordered N-terminal domain of the human androgen receptor. Protein Science. 2022;31(6):e4334. https://doi.org/10.1002/pro.4334