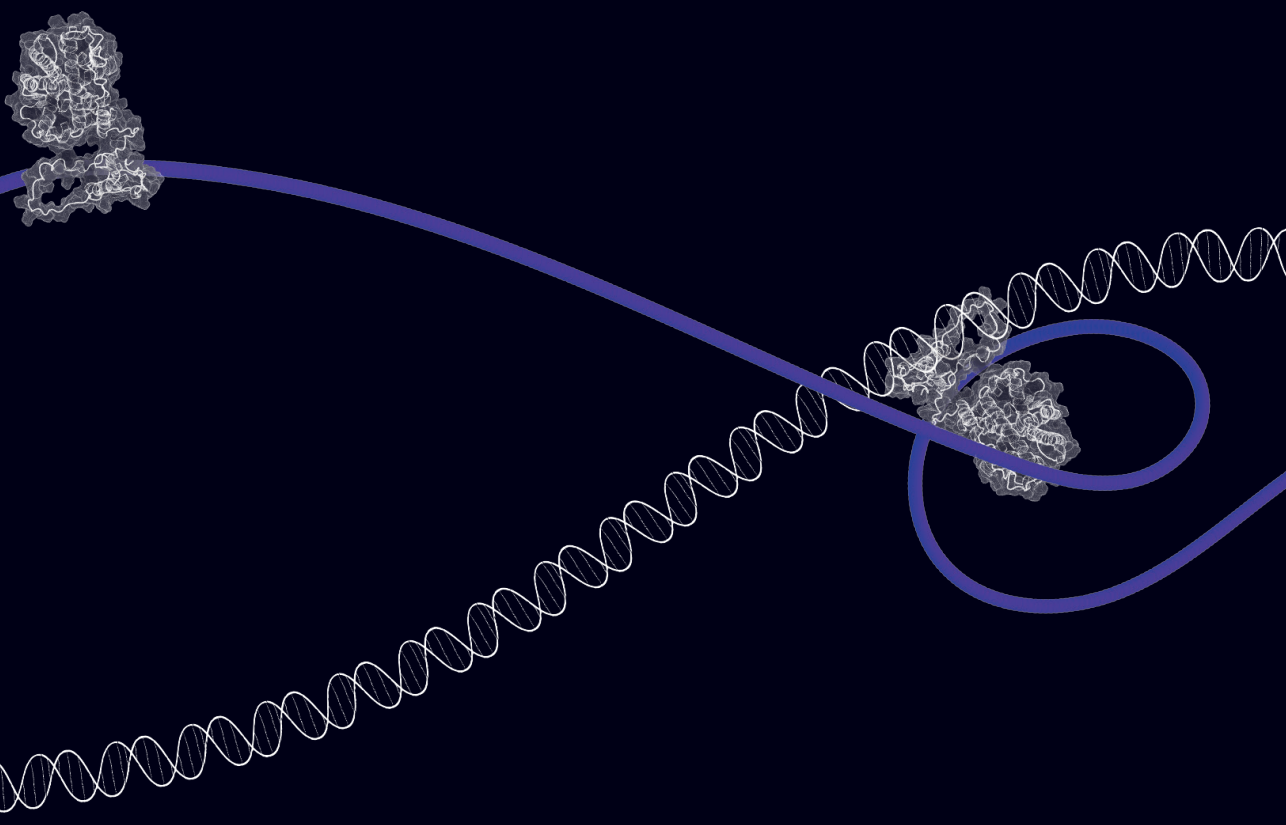
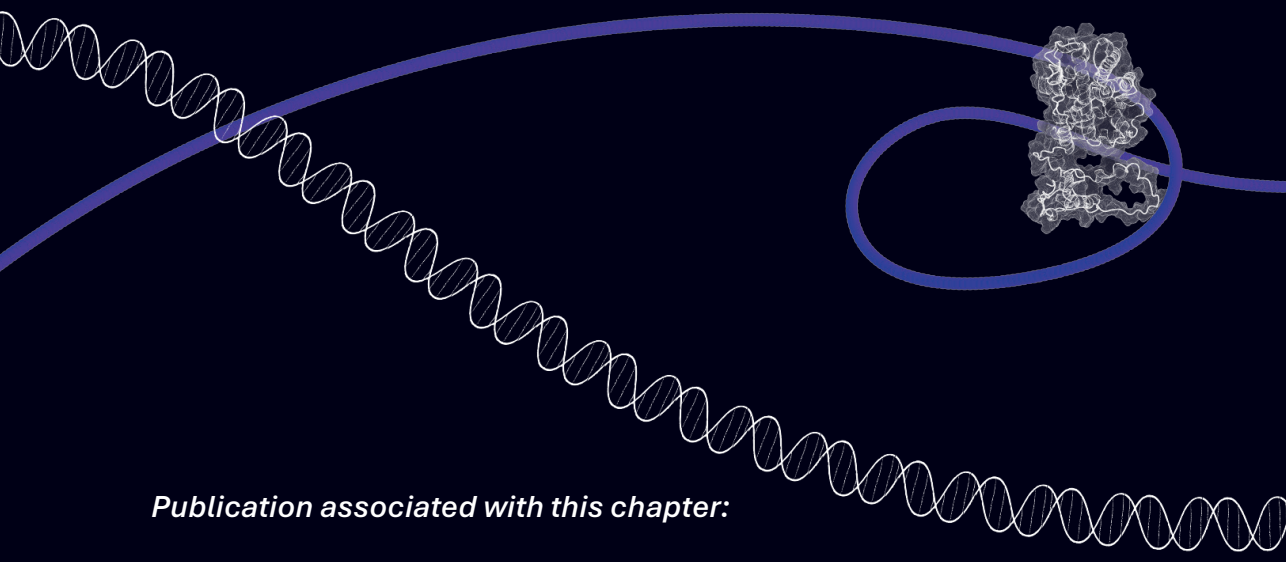
Biophysics of disordered nuclear receptors and their DNA binding regulation

Heling, L.W.H.J.



Chapter 3

Deciphering the Allosteric Control of Androgen Receptor DNA Binding by its Disordered N-Terminal Domain



Publication associated with this chapter:

Laurens W.H.J. Heling, Jolieke van der Veen, Adam Rofe, Eric West, Alba Jiménez-Panizo, Andrea Alegre-Martí, Vahid Sheikhhassani, Julian Ng, Thomas Schmidt, Eva Estébanez-Perpiñá, Iain J. McEwan, Alireza Mashaghi. *Deciphering the Allosteric Control of Androgen Receptor DNA Binding by its Disordered N-Terminal Domain.* Mol Cell Endocrinol 608, 112634 (2025).

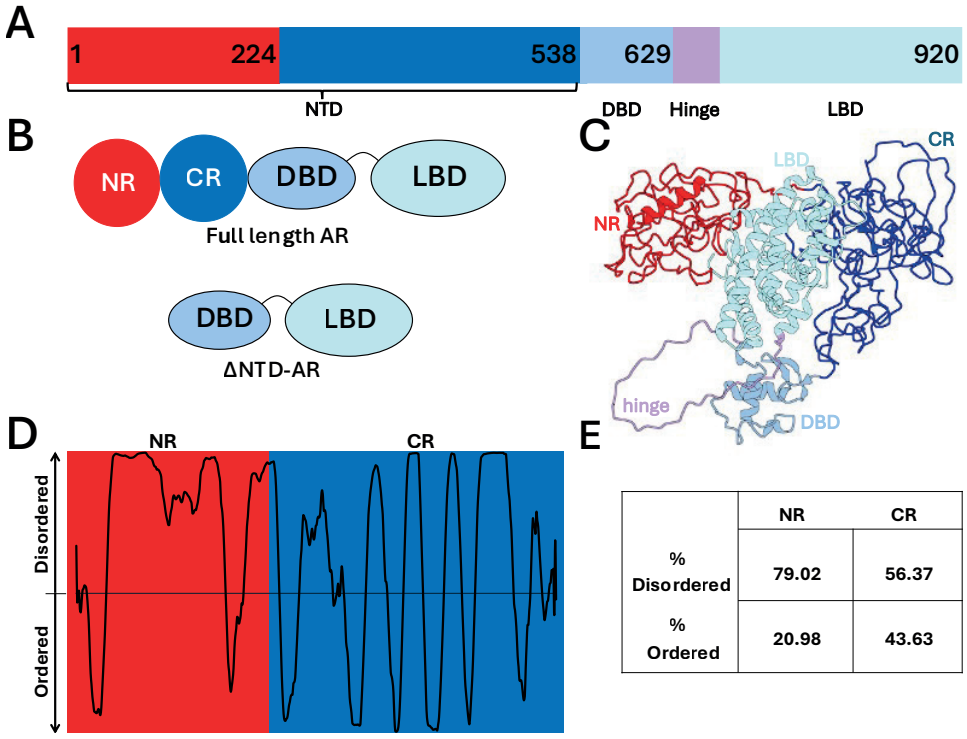
Abstract

The androgen receptor (AR) is pivotal in male physiological development and drives the pathogenesis of diseases such as prostate cancer. While its intrinsically disordered N-terminal domain (NTD) is essential for transcriptional activation, the mechanisms by which it regulates AR DNA-binding activity remain poorly understood. Here, we elucidate the allosteric control that specific NTD subregions—the N-terminal region (NR) and C-terminal region (CR)—exert over the DNA-binding properties of a truncated AR construct (**Δ**NTD-AR). Using microscale thermophoresis (MST) and single-molecule fluorescence imaging, we demonstrate that these subregions differentially modulate the kinetics and affinity of AR–DNA interactions. MST revealed that the CR drives a concentration-dependent reduction of DNA-binding affinity in **Δ**NTD-AR. However, single-molecule imaging showed that while the NR accelerates dissociation, the CR markedly diminishes binding frequency alongside accelerating dissociation. When combined, the NR and CR exert synergistic effects at high concentrations, reducing affinity and altering binding kinetics distinctly from either subregion alone. Collectively, our findings define the unique functional roles of the NR and CR in the allosteric modulation of AR–DNA interactions, providing critical mechanistic insights into intrinsic receptor regulation and exposing potential allosteric sites for therapeutic intervention.

Introduction

The androgen receptor (AR) is a ligand-activated transcription factor (TF) that plays a pivotal role in male development, tissue homeostasis, and the pathogenesis of numerous diseases[1] **(Figure 1)**. As a member of the nuclear receptor superfamily, AR orchestrates complex gene expression programs in response to androgens like testosterone and dihydrotestosterone (DHT)[1]. In the absence of ligand binding, AR is in a transcriptionally inactive conformation, regulated by chaperones[2], residing in the cytoplasm. Upon ligand binding, AR may undergo a series of conformational changes, which lead to dissociation from chaperone and cochaperone complexes, and translocates to the nucleus, where it binds to specific DNA sequences known as androgen response elements (AREs) within the regulatory regions of target genes[3,4]. This process, further modulated by a cascade of post-translational modifications—including phosphorylation, acetylation, and SUMOylation—fine-tune AR's transcriptional activity[5] and its interaction with a diverse array of co-regulatory proteins[6]. Aberrant AR signalling is implicated in a wide range of pathologies. In particular, it is the primary driver of prostate cancer, with sustained AR activity promoting tumour growth and progression even in the context of castrate levels of androgens, a condition known as castration-resistant prostate cancer (CRPC)[7]. Mechanisms underlying CRPC are multifaceted, but often involve AR gene amplification, overexpression, mutations that broaden ligand specificity, or the emergence of constitutively active AR splice variants that lack the ligand-binding domain[8].

Despite decades of research, a comprehensive understanding of the structural and molecular mechanisms governing the regulatory network of AR remains elusive. Until recently, structural information regarding AR was limited to isolated domains **(Figure 1B)**. This hindered the understanding of interdomain and intradomain cooperativity in AR, and particularly its role in transactivation and DNA binding. While the structures of its ligand-binding (LBD) and DNA-binding (DBD) domains have been resolved[9,10], the large N-terminal domain (NTD), comprising more than half of the protein, has defied structural characterisation due, in part, to its intrinsically disordered nature **(Figure 1C–E)**. This knowledge gap is particularly significant as the NTD acts as a hub for protein–protein interactions, is subject to extensive post-translational modifications, and frequently harbours mutations and drives constitutive activity of AR splice variants found in CRPC[11]. The NTD is vital for transactivation of AR, as removal leads to a loss of all AR mediated expression[12,13]. Furthermore, the interplay between NTD regions and LBD exposed surfaces, such as the AF-2 coactivator binding pocket, may serve as an intrinsic regulatory mechanism[3,4,14,15]. Early biochemical studies indicated that the AR-NTD had a negative impact on DNA binding to both selective and non-selective androgen receptor elements (ARE)[16] and our previous computational modelling efforts suggested that the NTD transiently adopts two distinct subregion conformations, a N-terminal region (NR) and a C-terminal region (CR) **(Figure 1B–E)**, which may differentially modulate AR[17].



	NR	CR
% Disordered	79.02	56.37
% Ordered	20.98	43.63

Figure 1. AR has a standard nuclear receptor domain architecture with a large intrinsically disordered NTD. (A) *Linear domain structure of AR, colour coded by region. Red: NR, blue: CR, turquoise: DBD, purple: hinge region, cyan LBD. Numbers depict amino acid numbers following the UniProt sequence.* ***(B)*** *Cartoon representation of full length and ΔNTD-AR.* ***(C)*** *Full length model of AR-NTD colour coded by region. Model taken from Sheikhhassani et al (2022)* ***(D)*** *PoNDR VLS2 predicted score of the NTD region, indicating the predicted probability of stretches being disordered and ordered.* ***(E)*** *Table with the percentage of residues being ordered or disordered taken from the PoNDR VSL2 predictions.*

Here, we build upon this previous work to study the modulatory effects of NTD subregions by visualising AR DBD–LBD (**Δ**NTD-AR **(Figure 1B)**) interactions with DNA using *in vitro* binding assays and single-molecule fluorescence microscopy. We demonstrate that the NTD subregions directly regulate **Δ**NTD-AR's interaction with DNA. The NR and CR subregions differentially modulate the ability of **Δ**NTD-AR to bind and release from DNA. These findings support the hypothesis of allosteric regulation of AR through its NTD and illuminate a novel mechanism of the NTD in modulating AR's dynamic interaction with DNA. A combination of rigid body docking with flexible protein complex structure prediction revealed that the interactions of NTD subregions can be mapped to the allosterically linked AF2 and BF-3[18] surfaces on the LBD, with additional binding sites on the DBD dimerisation interface. Our data thus provides crucial insights into the molecular mechanisms of intrinsic AR regulation.

Materials and methods

Protein expression and purification

NTD subfragments

The NR (amino acid 1 to 224) and CR (amino acids 225–538) fragments were amplified by PCR using the primers (Eurofins) **(Table S1)** and cloned into *pET-23a* and *pET-19b* His-tag expression plasmids by Gibson cloning. The constructs were confirmed by restriction enzyme digest and sequencing. The protein expression and purification of recombinant AR-NTD polypeptides, NR and CR was as previously described. Briefly, *E.coli* (BLR cells) were treated with 1 mM IPTG, for 2 hours at 37 °C. cells were collected and lysed using 0.5–1 mg/ml (50 mg/ml stock). The expressed proteins were then purified by Ni-affinity chromatography and subsequently eluted with Tris buffer (20 mM Tris, pH7.9, 500 mM NaCl, 200 mM imidazole and 5% glycerol). Purity of expressed proteins were checked by SDS polyacrylamide gel electrophoresis and protein concentrations determined by Bradford assay (BioRad).

ΔNTD-AR

Recombinant human AR DBD–LBD (**Δ**NTD-AR, residues 550–920) was cloned into a pet28a vector and expressed as a fusion protein with a N-terminal hexahistidine (6xHis) tag. The protein was purified to homogeneity using standard chromatographic procedures after cleavage with TEV protease[10].

Microscale Thermophoresis (MST)

A fluorescently labelled duplex DNA (fDNA) containing the IR3 ARE[9] was used as a substrate. The top strand was 31 base pair long, while the bottom strand was 32 base pairs long, creating an overhang on the 5' end of the duplex. This overhang was filled with an aminoallyl-dUTP-ATTO-647N (NU-803-647N-S, Jena Bioscience) using DNA Polymerase I, Large (Klenow) Fragment (M0210, New England Biolabs). Free nucleotides were subsequently removed using the Monarch® Spin PCR & DNA Cleanup Kit (T1130, England Biolabs) and the fDNA was checked for homogeneity on a 2% agarose gel. **Δ**NTD-AR was dialysed in imaging buffer (100 mM Tris-HCl (pH 7.5), 100 mM NaCl, 20 mM $MgCl_2$) overnight before serial diluted from 500 nM to 31.25 nM in imaging buffer supplemented with 0.4% Tween-20 and 50 nM DHT (D-073 Sigma Aldrich). These samples were mixed 1:1 with the 80 nM of the fDNA substrate in milliQ water to ensure all samples contained 40nM fDNA and AR concentrations ranged from 250 nM to 15.625 nM in 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$, 0.2% Tween-20 and 25 nM DHT. For experiments with the addition of NTD subfragments, **Δ**NTD-AR and the subfragment were mixed prior to adding with DNA. Each sample was incubated at room temperature for 10 minutes and transferred into MST capillaries (Standard Monolith Capillaries, MO-K022, NanoTemper, Germany). MST measurements were done on a Monolith NT.115 instrument (NanoTemper, Germany) at 40% LED power and medium MST power. Total measurement time was 30 s, with 5 s laser off, 20 s laser on and 5 s laser off. F_{norm} values were evaluated after 10s of laser on and normalised against baseline (DNA alone).

Single-molecule DNA tightrope assay

The DNA tightrope assay was described previously in detail[19]. In short, a custom microfluidic flow chamber was made by combining a standard microscope slide with two holes drilled in it, a double-sided tape gasket and a silanised coverslip. Two tubes connected to a syringe on a peristaltic pump (World Precision Instruments AL1000-220) and a custom Eppendorf tube permitted the controlled movement of reagents through the flow chamber. To reduce non-specific surface interactions of proteins, DNA or fluorophores, the flow chamber was blocked by incubating overnight in mPEG buffer (25 mg/ml mPEG5000 in 250 mM $NaHCO_3$, pH 8.2) followed by incubation overnight in ABT buffer (10 mg/ml BSA, 0.1% Tween-20 & 0.1% NaN_3). 5 µm silica microspheres (Bang's Laboratories, USA) coated with poly-L-lysine (P5899, Sigma-Aldrich) are deposited on the surface and λ-phage DNA (500 ng) is flowed back and forth through the chamber at a constant velocity (300 µL/min) for 40 minutes, elongate the DNA molecules and forming single-molecule DNA tightropes between the beads.

Proteins were labelled with fluorescent Quantum dots. The his-tag located on the **Δ**NTD-AR was tagged through a Qdot antibody sandwich. 1 µM AR **Δ**NTD-AR was dialysed against the imaging buffer (described in the MST section) overnight. An equimolar concentration of hexahis primary antibody (10001-0-AP, Proteintech) was added to the protein and incubated on ice for 30 minutes. A F(ab')2-Goat anti-Mouse IgG (H+L) Secondary Antibody with a 525 Qdot (Q-11041MP, Invitrogen) was then added to the protein:primary solution at a threefold higher concentration and incubated for 30 minutes on ice. Labelled proteins were diluted in imaging buffer, supplemented with 25 nM DHT and 100 mM DTT, and ready for introduction into the tightrope assay.

Super Resolution Fluorescence Microscopy

Single-molecule imaging was performed on a home-built wide-field setup, based on an Axiovert S100 (Zeiss, Germany) inverted microscope equipped with a 100x 1.4NA oil-immersion objective (Zeiss, Germany) at room temperature (20°C). The sample was excited by a 488nm laser (Sapphire CDRH, Coherent Inc., USA). The intensity and timing were set through an acousto-optic modulator (AOTFnC-VIS, AA-OptoElectronic, France) such that the intensity was set at 1 kW/cm^2 and the illumination time at 100 ms per frame. The light was detected through a dichroic/emission combination (Di01-R405_488_561_635/zet405_488_561_640m, Semrock, USA). The signal of individual dye molecules was captured on a sCMOS camera (Orca Flash 4.0V2, Hamamatsu, Japan). The comparison of the single-molecule fluorescence signal to background yielded a signal-to-noise ratio of 14 and were spatially distributed according to the microscope's point-spread function (440 nm FWHM), allowing for localisation of individual fluorophores with sub-30 nm precision.

Data analysis

Single-molecule microscopy videos were converted into kymographs using ImageJ[20]. To identify tightrope positions between beads, videos were projected as Z-stacks using maximum intensity projection. A 25-pixel rolling ball radius background subtraction was applied to the kymographs, to average the intensity of bright and dim binding events. Kymograph projections allow the identification of different DNA search mechanisms: protein sliding along the DNA to search for its target sequence will appear as movement along the y-axis over time, while a 3-dimensional search mechanism is represented by horizontal streaks. For the lifetime analysis, molecules that exhibited binding which exceeded the video length were excluded. To mitigate bin size bias, attached lifetimes were plotted as cumulative frequency histograms and fitted to exponential functions. The necessity of single versus double exponential fits was assessed via an F-test. Data fitting was performed in Microsoft Excel, using sum-of-squares difference fitting with the built in Solver function and the Statistics add-in package. Fitting errors were determined in Origin (Version 2024, OriginLab, Northampton, MA, USA) software packages, using the parameters derived from the Excel fits.

Structural modelling

The atomic coordinates of NR and CR were taken from our all-atom aa99SB-disp MD models[21] and used as input for protein–protein interactions using biased rigid-body docking with ClusPro 2.0 webserver[22]. For **Δ**NTD-AR, residues 550–920 were extracted from the AlphaFold model (AF-P10275-F1-v4). ClusPro predicted 120 poses, which were clustered using Gromacs with a 10Å cutoff. We used a local version of ColabFold implementation of AlphaFold-Multimer[23], which was shown to have a performance similar to the original implementation of AlphaFold-Multimer while using MMSeqs2 to search genomic databases orders of magnitude faster. ColabFold v1.5.2 is initialised with model_type set to alphafold_multimer_v3, num_recycles is equal to 10 and num_seeds is 5. The rest of the values were set to default. We applied a sliding window bin approach to the NR and CR structures to improve the prediction success. To reduce bias for the nuclear receptor boxes, the search sensitive parameter -s was set to 3. These boxes, like the FQNLF on NR and the LXXLF on CR can, due to their presence in the protein data bank—and therefore part of the training data for AlphaFold—dominate the structural predictions. The constructed MSAs were used as an input for LocalColabFold to predict multimer structures for **Δ**NTD-AR with the NR and CR bins. The structures were relaxed by AlphaFold's AMBER forcefield.

All poses predicted by ClusPro and AlphaFold were inspected and analysed using a custom-made Python script; utilizing different libraries: madplotlib[24], mdanalysis[25,26], mdtraj[27], numpy[28], pandas[29]. The best docking pose and AlphaFold predictions were overlaid and examined in ChimeraX[30].

Results

Microscale thermophoresis suggests allosterically regulation of DNA binding by AR-NTD

To investigate the influence of the NR and CR subregions on the DNA binding affinity of **Δ**NTD-AR **(Figure 2A)**, we employed microscale thermophoresis (MST) **(Figure 2B)** using a fluorescently labelled, 32 basepair long, DNA substrate containing the palindromic AR response element. This approach enabled the specific monitoring of protein–DNA interactions, thereby facilitating the discrimination of potential physiological irrelevant higher-order protein complexes and aggregate formation that may otherwise dominate observations. To eliminate any potential impact of NR and CR on DNA diffusion, we conducted initial experiments where NR and CR were added to DNA. Our findings reveal no discernible difference in diffusion compared to DNA alone **(Figure S1)**, confirming that alterations in DNA diffusion are not attributable to NTD binding directly to DNA. Next, we titrated **Δ**NTD-AR in the absence of NTD fragments against DNA, for which the steady state k_D was determined at 64.4 ± 2.2 nM **(Figure 2C and Table 1)**.

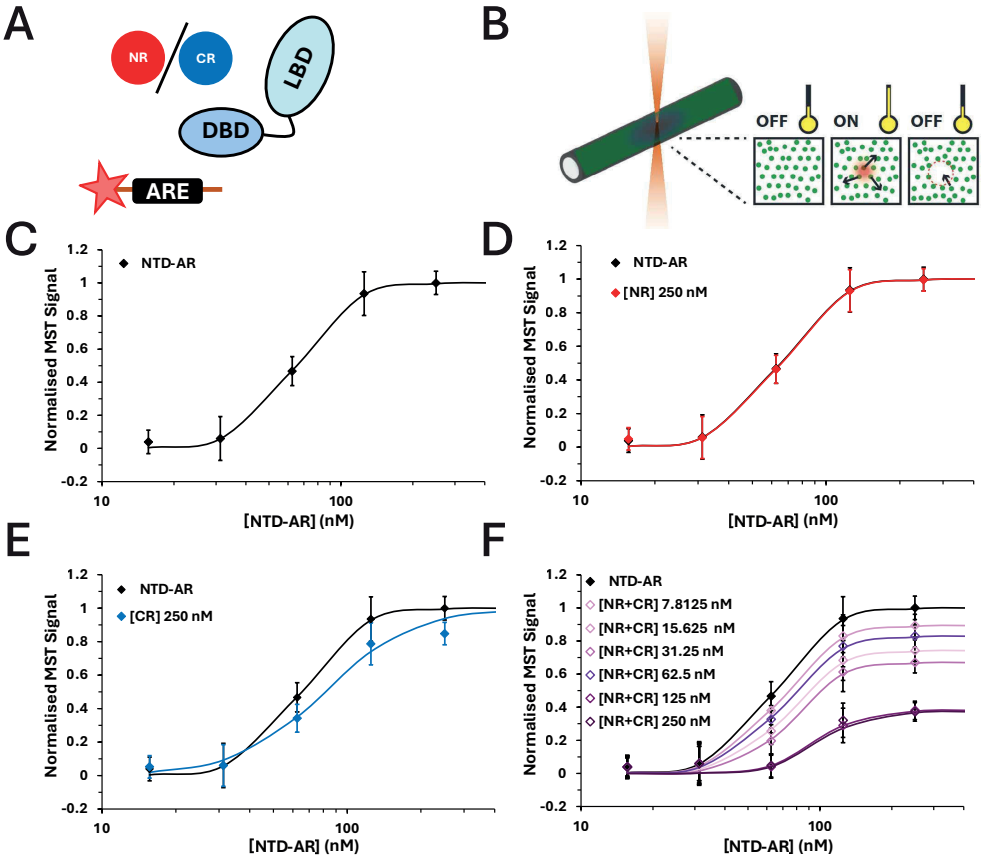


Figure 2. Microscale thermophoresis of ΔNTD-AR titrated against a short DNA fragment in the presence and absence of NTD. (A) *The components in the MST assay. A short DNA fragment, containing an ARE sequence, is fluorescently labelled. This is mixed with different concentrations of ΔNTD-AR and NR, CR, NR+CR or no NTD, and loaded into capillaries.* ***(B)*** *Graphical representation of an MST assay. The fluorescence emission of the capillary is continuously followed. Upon application of heat, molecules, depending on their hydrodynamic size, diffuse away from the heat point. Smaller fragments diffuse faster than larger fragments.* ***(C)*** *Normalised MST binding curve for ΔNTD-AR alone.* ***(D)*** *Binding curve of ΔNTD-AR in the presence of 250 nM NR in red.* ***(E)*** *Binding curve of ΔNTD-AR in the presence of 250 nM CR in blue.* ***(F)*** *Binding curves of ΔNTD-AR in the presence of different concentrations of NR+CR. Means ± SEM, n = 8.*

The assay was repeated with different concentrations of NR, CR and NR+CR **(Table 1)**. Titrating NR to **Δ**NTD-AR revealed no significant impact on the k_D **(Figure 2D, S2 and Table 1)**, suggesting that NR alone does not directly influence the DNA binding affinity of **Δ**NTD-AR. In contrast, CR demonstrated a concentration-dependent inhibitory effect on **Δ**NTD-AR–DNA interactions **(Figure 2E, S3 and Table 1)**. While lower CR concentrations had a minimal impact on k_D, higher concentrations led to a marked increase in k_D and decreased the maximum response, indicative of reduced binding affinity and overall binding capacity. To probe potential cooperative effects, we next examined the combined impact of equimolar NR+CR on **Δ**NTD-AR–DNA binding. At lower concentrations of the combined NTD fragments (62.5 nM and below), the k_D and maximum response showed a moderate effect, but did not show a linear response based on the concentration used **(Figure 2F and S4)**. At higher concentrations of NR and CR, a more pronounced inhibitory effect emerged, characterised by a considerable increase in k_D (99.9 ± 6.12 nM) and decrease of the maximum response.

Table 1. Summary of equilibrium dissociation constants. *k_D of ΔNTD-AR to DNA in the presence of different NTD regions determined by the MST fits.*

NTD subregion	k_D (nM)	NTD subregion	k_D (nM)	NTD subregion	k_D (nM)
None	64.4 ± 2.8				
7.8125 nM **NR**	65.4 ± 2.4	7.8125 nM **CR**	66.3 ± 2.3	7.8125 nM **NR + CR**	72.2 ± 3.6
15.625 nM **NR**	67.2 ± 2.5	15.625 nM **CR**	67.2 ± 2.5	15.625 nM **NR + CR**	67.0 ± 2.5
31.25 nM **NR**	66.8 ± 2.4	31.25 nM **CR**	66.8 ± 2.4	31.25 nM **NR + CR**	75.5 ± 4.2
62.5 nM **NR**	68.8 ± 3.3	62.5 nM **CR**	66.8 ± 2.4	62.5 nM **NR + CR**	68.9 ±2.9
125 nM **NR**	64.4 ± 2.4	125 nM **CR**	82.4 ± 8.6	125 nM **NR + CR**	97.0 ± 6.1
250 nM **NR**	64.8 ± 2.6	250 nM **CR**	80.9 ± 8.1	250 nM **NR + CR**	99.9 ± 9.1

Single-molecule imaging reveals NTD increases dissociation of ΔNTD-AR

To better understand how the NR and CR fragments affect DNA interaction kinetics of AR, we imaged individual fluorescently labelled **Δ**NTD-AR proteins in real-time using single-molecule microscopy. For this we employed the *in vitro* DNA tightrope assay[31], where individual DNA molecules are suspended between glass beads in a microfluidic chamber **(Figure 3A)**. Videos of **Δ**NTD-AR proteins binding to DNA were transformed into kymographs **(Figure 3B)** from which diffusion behaviour can be characterised.

Our data made clear that **Δ**NTD-AR uses a 3-dimensional diffusive search for its target on DNA as the vast majority (>90%) of events showed no positional movement along the tightrope **(Figure 3B)**. The lifetime of the molecules could be extracted from the length of the streaks, which were plotted as a cumulative frequency and fitted with exponentials **(Figure 3C)**. In the absence of NTD fragments, the data fit best to a double exponential (F-test = >1 and residual plots **(Figure S5)**), suggesting two populations of attachments **(Figure 3C** inset**)**. The majority of events (87%) indicated a rate constant of 0.461 s^{-1}. The population of longer-lived interactions (13%; 0.072 s^{-1}) most likely are **Δ**NTD-AR molecules binding to partial response elements along the lambda (λ)-phage DNA as there are 20 instances of half the palindromic repeat in the sequence.

Next, we investigated AR–DNA interactions in the presence of equimolar concentrations of NTD fragments (NR alone, CR alone or NR+CR together). Fitting of the cumulative frequencies of the lifetimes for these conditions also fitted best to double exponentials **(Figure S5 and Table S2)**. Modulation of **Δ**NTD-AR by NTD fragments increased the dissociation rates of both populations bound to DNA **(Figure 3D)**. The addition of CR led to the most dramatic increase, halving the dwell-time of **Δ**NTD-AR on DNA **(Figure 3E**, effective k_{off} = 0.840 s^{-1}**)**. Simultaneously, qualitative observations indicated a significant reduction in **Δ**NTD-AR binding when CR was present. While the combined effect of NR and CR partially restored the rate constant, aligning it more closely with the results observed for NR modulation (effective k_{off} 0.536 s^{-1} and 0.573 s^{-1} respectively), the frequency of observed events remained lower than those with NR or without NTD fragments. Notably, the introduction of NR led to an increase in the number of longer-lived interactions (32% for NR and 17% for NR+CR; 13% for **Δ**NTD-AR and 9% for CR).

Structural basis for allostery modelled to the AF-2/BF-3 regions

To elucidate the physical mechanisms driving these altered dissociation rates, we sought a structural explanation. Since structural models derived from experimental evidence on full-length AR are limited, we employed our previously computationally derived atomistic model of AR-NTD combined with ClusPro and the AlphaFold-Multimer Machine Learning prediction approaches to model the interactions between the NTD regions and **Δ**NTD-AR. While we previously elucidated the strengths of computational modelling and ClusPro, here we added AlphaFold Multimer to elucidate the conditional folding of the NTD subregions with **Δ**NTD-AR. The AlphaFold algorithm has difficulties with accurately

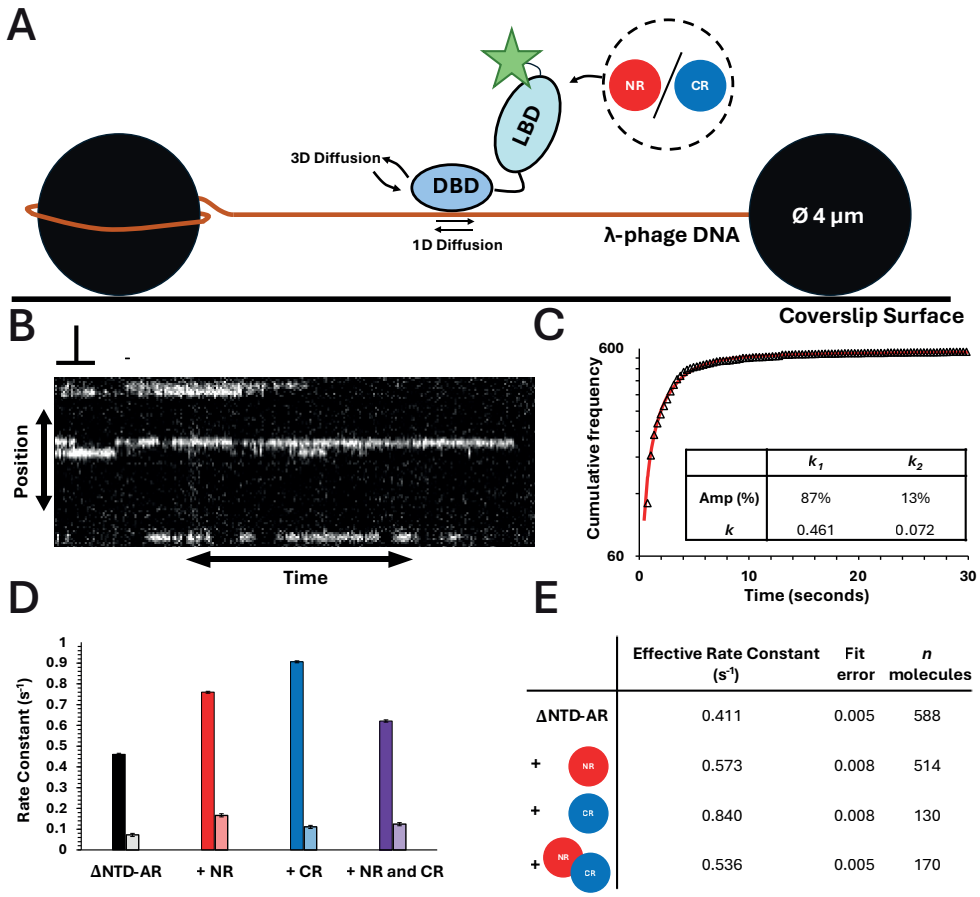


Figure 3. The single-molecule binding kinetics of ΔNTD-AR. ***(A)*** *Graphical representation of the DNA tightrope assay.* ***(B)*** *Example of a kymograph transformation from a video showing Q-Dot labelled AR DBD-LBD binding along the DNA tightrope. Y-axis denotes the position along the DNA, the X-axis the time. Scale bare represents 1 second and 0.727 µm.* ***(C)*** *Length of streaks were compiled into a logarithmic cumulative frequency plot against time. Non-linearity of the data indicates that there are two processes, fit with a double exponential with rate constants given in the inset.* ***(D)*** *Comparison of the rate constants of AR DBD-LBD and the modulatory effect of NTD. Each condition contained two populations of events, with different average dissociation rate constants. The fast population is depicted by the darker bars, the slow population by the lighter bars. Error bars indicate SEM of the fit.* ***(E)*** *Summary of the effective rate constants and fit errors. "n molecules" refers to molecules analysed from at least 3 independent flowcells.*

predicting the structures of IDPs, often representing them with a distinctive ribbon-like appearance. However, recent studies revealed that the success rate of IDP structural prediction in protein–protein interactions can be significantly improved—from 42.9% to >85%—by implementing a sliding window bin strategy[32,33]. We adopted a similar approach, dividing the NR and CR sequences into 10 bins (amino acid stretches) with a 30% sequence overlap. The resulting predictions were then cross-referenced with our

ClusPro results, and the structure with the highest-scoring combination of pLDDT, ipTM and cluster score is presented **(Figure 4A)**. This revealed that CR's highest ranked position brings it close to the hydrophobic AF-2 pocket on the LBD, while NR shows more affinity to the close-by and allosterically linked BF-3 binding pocket. The residue-based interaction counts were weighed by the pLDDT scores of the prediction and plotted **(Figure 4B)**. This reveals that apart from the BF-3, NR also has a narrow peak at the D-box on the DBD, a five-residue stretch that is responsible for DBD dimerisation.

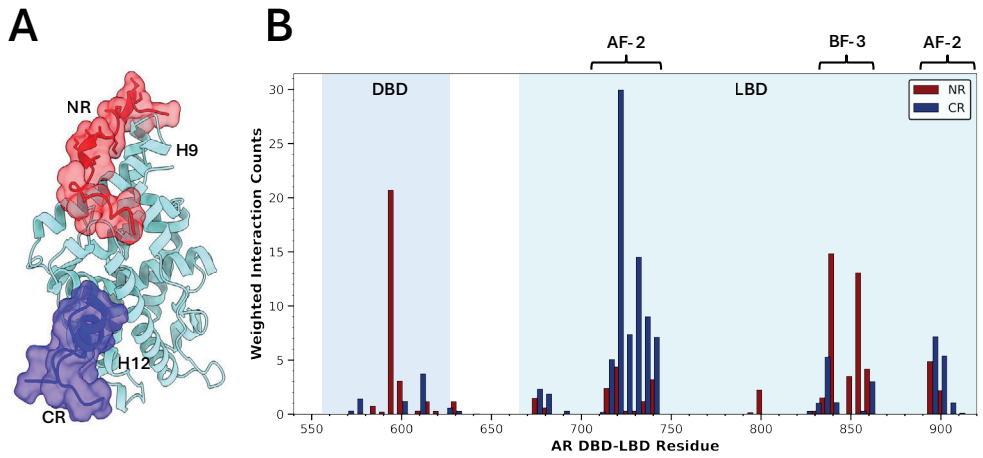


Figure 4. Structural modelling of NTD fragments binding to ΔNTD-AR. (A) *Highest ranked prediction structural model combining Alphafold and ClusPro show NR (red) and CR (blue) stretches interacting with different LBD (cyan) surfaces. Helices 9 and 12 are marked.* ***(B)*** *Weighted interactions of NR (red bars) and CR (blue bars) with AR DBD-LBD residues. Regions covering the DBD and LBD are coloured in lightblue and cyan respectively, and the AF-2 and BF-3 regions are indicated by brackets. In agreement with previous modelling data[17,21], NR appears as a major interacting partner of DBD. While NR engages D-box (residues 596–600), CR moderately restrict access to P-box (residues 577–581) and leaves the D-box exposed and accessible for dimerisation.*

Discussion

Androgen receptor (AR) signalling is a multifaceted process that is tightly controlled through an ensemble of intra- and intermolecular signals. Improving our understanding of these processes is of great importance, as several pathologies are associated with aberrant AR function—including prostate cancer and androgen insensitivity syndrome (AIS). Structural determination of separate domains has presented valuable information[34]. Yet the structural–functional cooperativity of AR domains remains controversial. In this study we took a key step towards elucidating the complex interplay between AR, its NTD, and DNA binding modulation. By integrating Microscale Thermophoresis (MST), single-molecule fluorescence imaging and structural modelling, we revealed a comprehensive

view of how NTD subregions NR and CR, individually and synergistically, modulate the binding kinetics and affinity of AR to specific and non-specific DNA in a simple *in vitro* model system.

In the cellular space, TFs can diffuse around the nucleus in search for their target sequence, and upon binding recruit other proteins to regulate target gene expression[35–37]. DNA residence times of TFs has been correlated with specific target motif interactions in cells[38,39], and search mechanisms for these target motifs are protein- and nuclear structure-dependent[35,37,40]. Our single-molecule imaging revealed the existence of two distinct populations of **Δ**NTD-AR when interacting with DNA, consistent with a 3D search mechanism for its target[41]. The λ-phage DNA construct in our assay does not contain any canonical AR target motifs, but several partial sites are present. A recent study showed that, similarly to the glucocorticoid receptor, monomeric AR binds DNA at the 5'-AGAACA-3' hexanucleotide sequence, which induces a conformational change that allows a second AR protein to cooperatively bind DNA with less sequence stringency[42]. We postulate that the longer-lived population may be AR binding to these partial AR target motifs. This also explains the increased dissociation rate constants that we observe when **Δ**NTD-AR is modulated by NR. Our results indicate **Δ**NTD–AR readily binds to DNA irrespective of NR presence, as the number of observations were similar. However, the dissociation rate constants of both the short- and long lifetime populations were 65% and 130% faster, respectively. Our structural model suggests two dominant poses, where NR is shown to interact with the D-box of DBD and the BF-3 on LBD **(Figure 3B)**. Both these poses may prohibit dimerisation and thus decrease the affinity or stability of AR interactions. Analysis of data from previous reports based on rigid docking analysis of NTD–DBD interactions suggests that the NTD may interact with the DBD through the NR, partially mediated by the D-box in its most probable conformation[17]. Previous EMSA experiments showed that D-box mutations, which inhibit DBD dimerisation, resulted in a faster release of AR from DNA[42], suggesting dimerisation is important for stabilisation of AR–DNA interactions. They also reported no change in receptor activity, despite this faster release. Another way NR could inhibit AR dimerisation is through interactions with LBD. In the reported structure of the AR-LBD homodimer, apart from a large interaction interface, interactions were also mapped to the BF-3 surface[10], although the biological relevance of this interface remains unclear. NR modulating dimerisation could therefore increase the specificity of AR to its correct response elements. Our results corroborate this, with unchanged k_D and maximum binding of **Δ**NTD-AR with the ARE but changed kinetics with unspecific and half reporter sequences. This suggests that the stability, not the energy landscape of the complex, is the mechanism of NR mediated modulation.

CR, in contrast, exerts a more pronounced effect, particularly at higher concentrations. MST data reveals a concentration dependent increase in k_D, suggesting that CR either competes with DNA for binding or induces an allosteric effect that weakens the interaction. Single-molecule analysis supports this, indicating a dramatic reduction in the number

of binding observations, while the effective dissociation rate constant has doubled. This suggests that CR not only destabilises the AR–DNA complex but may also inhibit its formation. However, our MST data does not align with this, since the maximum response of DNA diffusion is similar compared to unmodulated AR. These seemingly contradicting observations raise further questions. The interaction between CR and the AF-2 may induce a conformation that destabilises the DHT interactions, thus decreasing the number of activated AR molecules. Previous studies have indicated that interactions with the AF-2 can affect the stability of the ligand binding pocket[43]. This may explain the lower observations in our single-molecule assay, since the absence of DHT in the buffer resulted in no binding events being observed. The lack of observation could also be explained by an increased specificity induced by CR. In MST, there is a full canonical androgen response element on the DNA fragment, while in the single-molecule assay there are none. If CR induces a conformation that stabilises AR complexes bound to the response element, this would lead to fewer observations when no response elements are present. An alternative mechanism that may explain our observation was proposed in other computational studies[17,21], where CR interactions with the DBD were shown to moderately restrict access to the P-Box – the surface responsible for specific DNA interactions. This effect is also observed in our analysis and appears to suppress binding to random DNA sequences compared to high-affinity canonical response elements.

The combined action of NR and CR reveals a synergistic modulation, most evident at high concentrations. When both NTD subregions are present, the binding affinity and capacity of **Δ**NTD-AR for DNA is evidently reduced. Intriguingly enough, the effects at lower concentrations are not linearly dependent on concentration. Our single-molecule data, show a reduced number of observations similar to CR alone but dissociation kinetics resembling those with NR alone. This suggests that NR and CR and the DNA cooperatively modulate AR, likely through a combined effect on the conformational landscape. The allostery between the AF-2 and BF-3 surfaces on the conformational landscape of AR have been discussed in detail before[44], and the sequence of DNA was shown to have an allosteric effect on AR conformations[42]. Further refining this allosteric landscape, our findings demonstrate competitive binding dynamics between DNA, NR and CR for **Δ**NTD-AR. Our data indicate that CR binding hinders the subsequent interaction of **Δ**NTD-AR with DNA, consistent with an allosteric model where CR binding stabilises a conformation with reduced DNA affinity. This contrasts with NR, which primarily impacts the stability of the formed complexes. We postulate that NR may be more dominant in the competition, especially at lower concentrations. At higher concentrations, where NR, CR and **Δ**NTD-AR outnumber the DNA molecules in our MST assay, we see an enhanced effect. Preliminary MST experiments with a full-length NTD fragment **(Figure S6)**, reveals that this competition at lower concentrations is similar as with the equimolar NR+CR mix, however at higher concentrations we see an attenuated effect. This suggest that steric hindrance occurs when NR and CR are attached to each other, which affects its overall regulation on ΔNTD-AR.

In a physiological context, the NR and CR fragments do not exist as separate entities, and the NTD is present in equimolar amounts to the other AR domains. Our model, therefore, proposes a potential AR perturbation strategy, where saturating CR interactions inhibits specific interactions. Our study did not identify the specific amino acids, or peptide stretches within the NTD fragments are predominantly responsible for the effects we have described here, warranting further empirically study. We surmise that the effective delivery of peptides remains a challenge, and additional investigations on potential small compound inhibitors that achieve similar effects may be necessary. Furthermore, it is important to note that our model here describes an *in vitro* system, with only AR protein domains present. By contrast, *in vivo* and *in cellulae* involve numerous other proteins that colocalise with AR in biomolecular condensates to regulate signalling and transcription[45]. Consequently, further work is needed to determine how our findings affect these dynamics. Since the stability of transcriptional condensates influences downstream effects[46], our finding that regions in the NTD affect AR–DNA stability could have similar consequences for condensate integrity and function.

In conclusion, this study reveals a model where the NTD can allosterically regulate the DNA binding kinetics and affinity of AR through its subregions NR and CR, adding information to the structure–function relations and intradomain cooperativity in AR. Further studies based on this work should fully elucidate the molecular mechanisms underlying intradomain allosteric regulation of AR. Studying the short peptide stretches of NTD using NMR and biochemical characterisation *in vitro* and *in vivo* may shed light on the vital amino acids underpinning these processes, and provide new avenues for therapeutic modulation of aberrant AR.

Author Contributions

Laurens W. H. J. Heling: Conceptualisation; investigation; methodology; visualisation; formal analysis; data curation; validation; writing-original draft; writing-review and editing. **Jolieke van der Veen:** Investigation, methodology. **Adam Rofe:** Investigation. **Eric West:** Investigation. **Alba Jiménez-Panizo:** Investigation; writing-review and editing. **Andrea Alegre-Martí:** Investigation. **Vahid Sheikhhassani:** Resources. **Julian Ng:** Funding acquisition. **Thomas Schmidt:** Resources; writing-review and editing. **Eva Estébanez-Perpiñá:** Writing-review and editing; resources; investigation. **Iain J. McEwan:** Writing-review and editing; resources; investigation. **Alireza Mashaghi:** Conceptualisation; investigation; resources; supervision; project administration; writing-review and editing; writing-original draft; funding acquisition.

Acknowledgements

Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualisation, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

References

1. Davey, R. A. & Grossmann, M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin Biochem Rev* **37**, 3 (2016).

2. Eftekharzadeh, B. *et al*. Hsp70 and Hsp40 inhibit an inter-domain interaction necessary for transcriptional activity in the androgen receptor. *Nature Communications 2019 10:1* **10**, 1–14 (2019).

3. Schaufele, F. *et al*. The structural basis of androgen receptor activation: Intramolecular and intermolecular amino-carboxy interactions. *Proc Natl Acad Sci U S A* **102**, 9802–9807 (2005).

4. van Royen, M. E., van Cappellen, W. A., de Vos, C., Houtsmuller, A. B. & Trapman, J. Stepwise androgen receptor dimerization. *J Cell Sci* **125**, 1970–1979 (2012).

5. Gioeli, D. & Paschal, B. M. Post-translational modification of the androgen receptor. *Mol Cell Endocrinol* **352**, 70–78 (2012).

6. Brooke, G. N., Parker, M. G. & Bevan, C. L. Mechanisms of androgen receptor activation in advanced prostate cancer: Differential co-activator recruitment and gene expression. *Oncogene* **27**, (2008).

7. Tan, M. E., Li, J., Xu, H. E., Melcher, K. & Yong, E. L. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica 2015 36:1* **36**, 3–23 (2014).

8. Guo, Z. *et al*. A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. *Cancer Res* **69**, (2009).

9. Shaffer, P. L., Jivan, A., Dollins, D. E., Claessens, F. & Gewirth, D. T. Structural basis of androgen receptor binding to selective androgen response elements. *Proceedings of the National Academy of Sciences* **101**, 4758–4763 (2004).

10. Nadal, M. *et al*. Structure of the homodimeric androgen receptor ligand-binding domain. *Nature Communications 2017 8:1* **8**, 1–14 (2017).

11. Gottlieb, B., Beitel, L. K., Nadarajah, A., Paliouras, M. & Trifiro, M. The androgen receptor gene mutations database: 2012 update. *Hum Mutat* **33**, 887–894 (2012).

12. Simental, J. A., Sar, M., Lane, M. V., French, F. S. & Wilson, E. M. Transcriptional activation and nuclear targeting signals of the human androgen receptor. *Journal of Biological Chemistry* **266**, 510–518 (1991).

13. Jenster, G., Van Der Korput, H. A. G. M., Trapman, J. & Brinkmann, A. O. Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. *Journal of Biological Chemistry* **270**, (1995).

14. He, B., Kemppainen, J. A., Voegel, J. J., Gronemeyer, H. & Wilson, E. M. Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH2-terminal domain. *Journal of Biological Chemistry* **274**, 37219–37225 (1999).

15. Wasmuth, E. V. *et al*. Modulation of androgen receptor DNA binding activity through direct interaction with the ETS transcription factor ERG. *Proc Natl Acad Sci U S A* **117**, (2020).

16. Brodie, J. & McEwan, I. J. 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* **34**, 603–615 (2005).

17. Sheikhhassani, V. *et al*. Topological dynamics of an intrinsically disordered N-terminal domain of the human androgen receptor. *Protein Science* **31**, (2022).

18. Estébanez-Perpiñá, E. *et al*. A surface on the androgen receptor that allosterically regulates coactivator binding. *Proc Natl Acad Sci U S A* **104**, (2007).

19. Springall, L., Inchingolo, A. V. & Kad, N. M. DNA-protein interactions studied directly using single molecule fluorescence imaging of quantum dot tagged proteins moving on DNA tightropes. in *Methods in Molecular Biology* vol. 1431 141–150 (2016).

20. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).

21. Heling, L. W. H. J. *et al*. Polyglutamine expansion induced dynamic misfolding of androgen receptor. *Protein Science* **34**, e70154 (2025).

22. Kozakov, D. *et al*. The ClusPro web server for protein–protein docking. *Nat Protoc* **12**, 255–278 (2017).

23. Mirdita, M. *et al*. ColabFold: making protein folding accessible to all. *Nat Methods* **19**, (2022).

24. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput Sci Eng* **9**, (2007).

25. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, (2011).

26. Naughton, F. B. *et al*. MDAnalysis 2.0 and beyond: fast and interoperable, community driven simulation analysis. *Biophys J* **121**, (2022).

27. McGibbon, R. T. *et al*. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J* **109**, (2015).

28. Harris, C. R. *et al*. Array programming with NumPy. *Nature* **585**, 357–362 (2020).

29. McKinney, W. Data Structures for Statistical Computing in Python. in *Proceedings of the 9th Python in Science Conference* (2010).

30. Meng, E. C. *et al*. UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* **32**, (2023).

31. Kad, N. M., Wang, H., Kennedy, G. G., Warshaw, D. M. & Van Houten, B. Collaborative Dynamic DNA Scanning by Nucleotide Excision Repair Proteins Investigated by Single- Molecule Imaging of Quantum-Dot-Labeled Proteins. *Mol Cell* **37**, (2010).

32. Basu, S., Kihara, D. & Kurgan, L. Computational prediction of disordered binding regions. *Comput Struct Biotechnol J* **21**, 1487–1497 (2023).

33. Bret, H., Gao, J., Zea, D. J., Andreani, J. & Guerois, R. From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2. *Nat Commun* **15**, (2024).

34. Johnson, M. J. & Wasmuth, E. V. Structural perspectives on the androgen receptor, the elusive shape-shifter. *Steroids* **211**, 109501 (2024).

35. Paakinaho, V. *et al*. Single-molecule analysis of steroid receptor and cofactor action in living cells. *Nat Commun* **8**, (2017).

36. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors.

Nat Rev Genet **15**, 69–81 (2014).

37. Mazzocca, M., Fillot, T., Loffreda, A., Gnani, D. & Mazza, D. The needle and the haystack: single molecule tracking to probe the transcription factor search in eukaryotes. *Biochem Soc Trans* **49**, 1121–1132 (2021).

38. Hettich, J. & Gebhardt, J. C. M. Transcription factor target site search and gene regulation in a background of unspecific binding sites. *J Theor Biol* **454**, (2018).

39. Clauß, K. *et al*. DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Res* **45**, 11121–11130 (2017).

40. Izeddin, I. *et al*. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *Elife* **2014**, (2014).

41. Van Royen, M. E. *et al*. Androgen receptor complexes probe DNA for recognition sequences by short random interactions. *J Cell Sci* **127**, (2014).

42. Lee, X. Y. *et al*. Structural mechanism underlying variations in DNA binding by the androgen receptor. *J Steroid Biochem Mol Biol* **241**, 106499 (2024).

43. Jin, Y. *et al*. Communication between the Ligand-Binding Pocket and the Activation Function-2 Domain of Androgen Receptor Revealed by Molecular Dynamics Simulations. *J Chem Inf Model* **59**, (2019).

44. Grosdidier, S. *et al*. Allosteric conversation in the androgen receptor ligand-binding domain surfaces. *Molecular Endocrinology* **26**, (2012).

45. Zhang, F. *et al*. Dynamic phase separation of the androgen receptor and its coactivators key to regulate gene expression. *Nucleic Acids Res* **51**, (2023).

46. Gui, T. *et al*. Targeted perturbation of signaling-driven condensates. *Mol Cell* **83**, (2023).

Supplementary Material

Supplementary Methodology

Dynamic light scattering (DLS)

62.5 nM ΔNTD-AR with different concentrations of the NTD subfragments in buffer (total volume 20 µL: 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM MgCl2 and 100 nM DHT) were added directly to a DynaPro Nanostar (Wyatt Technology, CA, USA) using a disposable microcuvette (Wyatt Technology, CA, USA) at room temperature (20°C). The mean hydrodynamic diameter was calculated from 3 separate measurements. The theoretical hydrodynamic radius of ΔNTD-AR was calculated from the PDB of the Alphafold model (AF-P10275-F1-v4, residues 550–920) using PyMOL.

Supplementary Figures

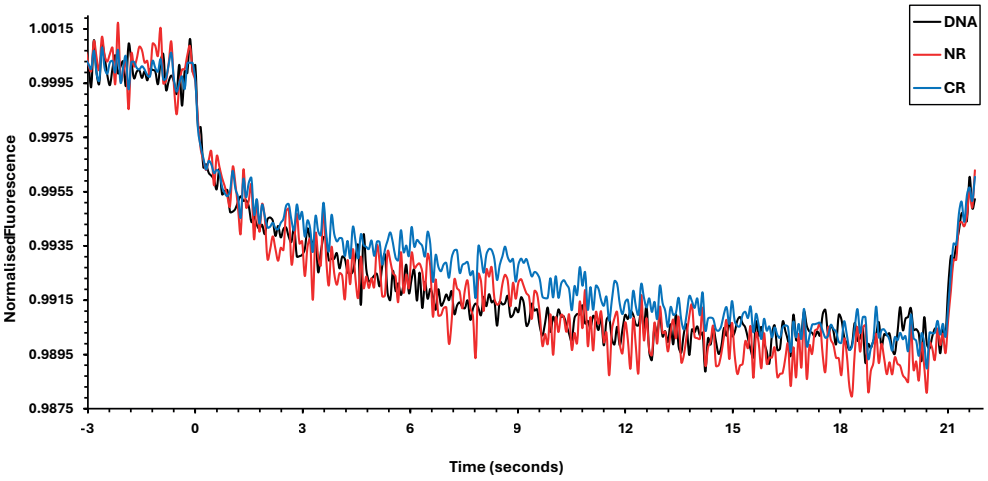


Figure S1. MST traces of fluorescent DNA. *Alone (black), with NR (red), and CR (blue). Traces are an average of 3 seperate MST experiments.*

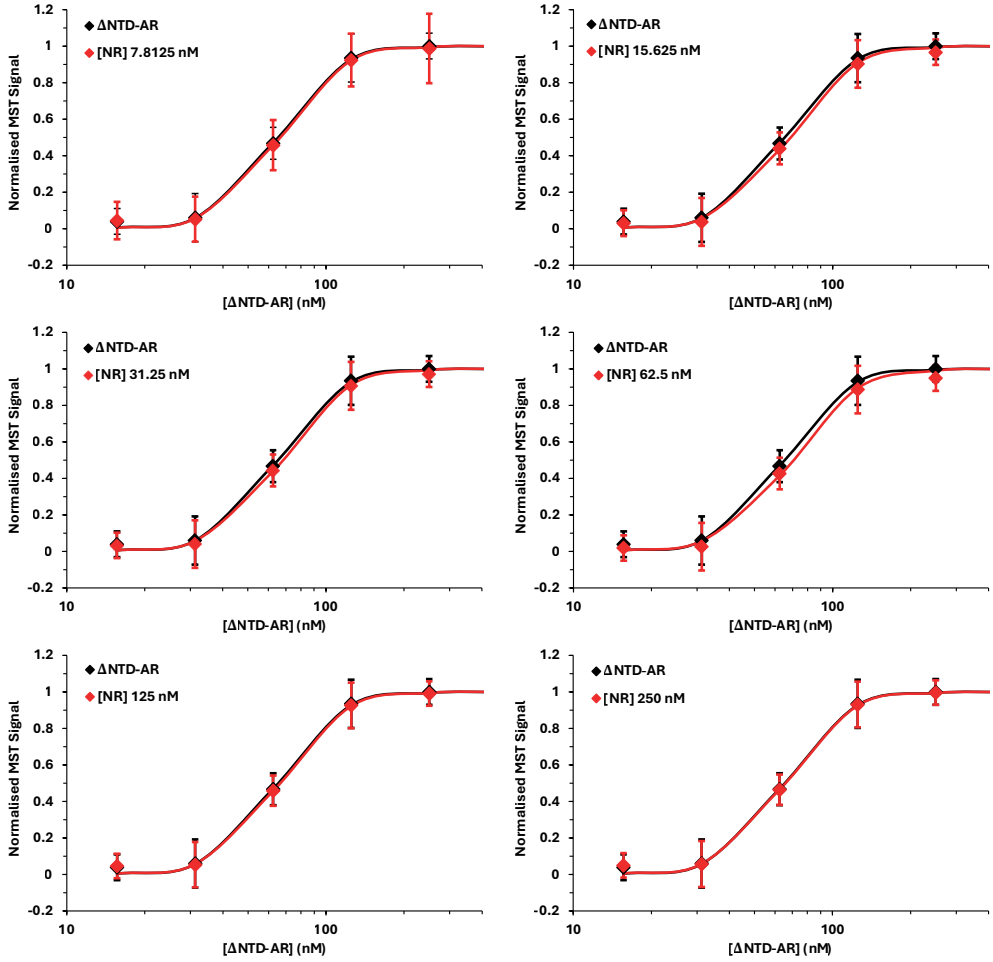


Figure S2. Normalised MST binding curves of ΔNTD-AR with NR. *Without (black) and with NR (red) in increasing concentrations.*

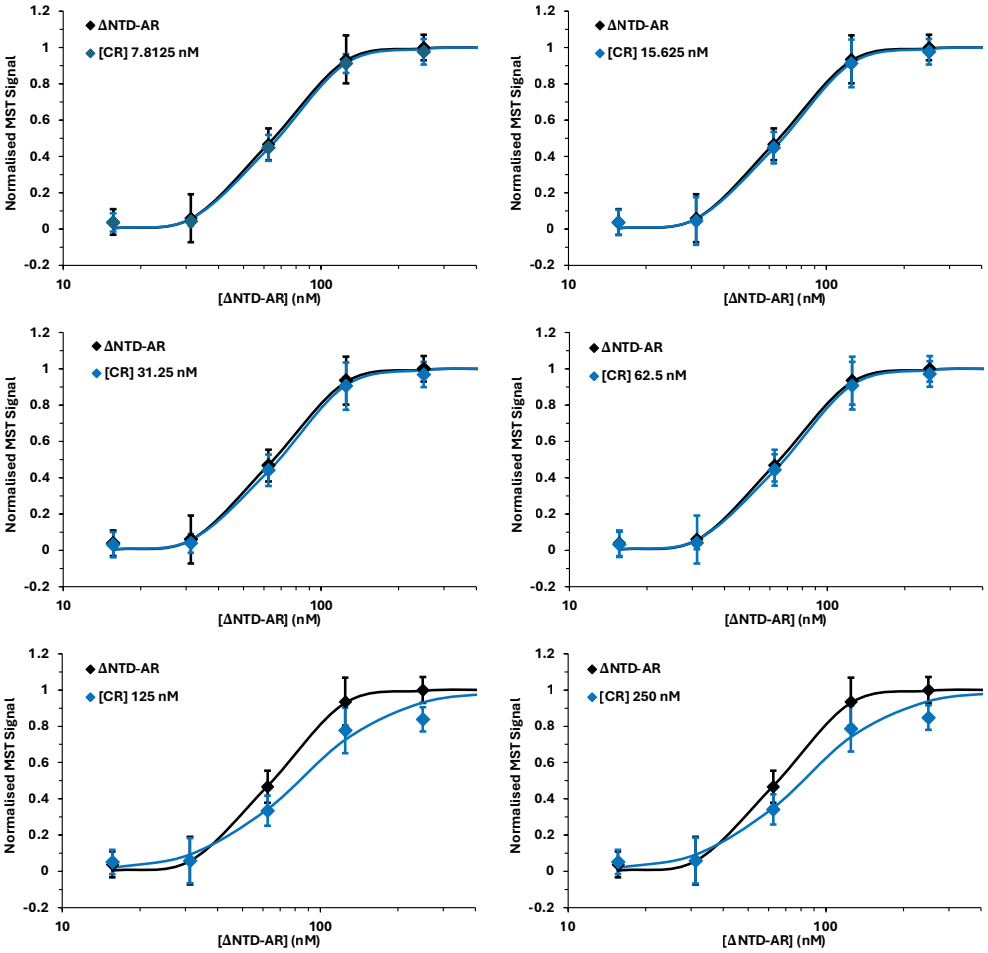


Figure S3. Normalised MST binding curves of ΔNTD-AR with CR. *Without (black) and with CR (blue) in increasing concentrations.*

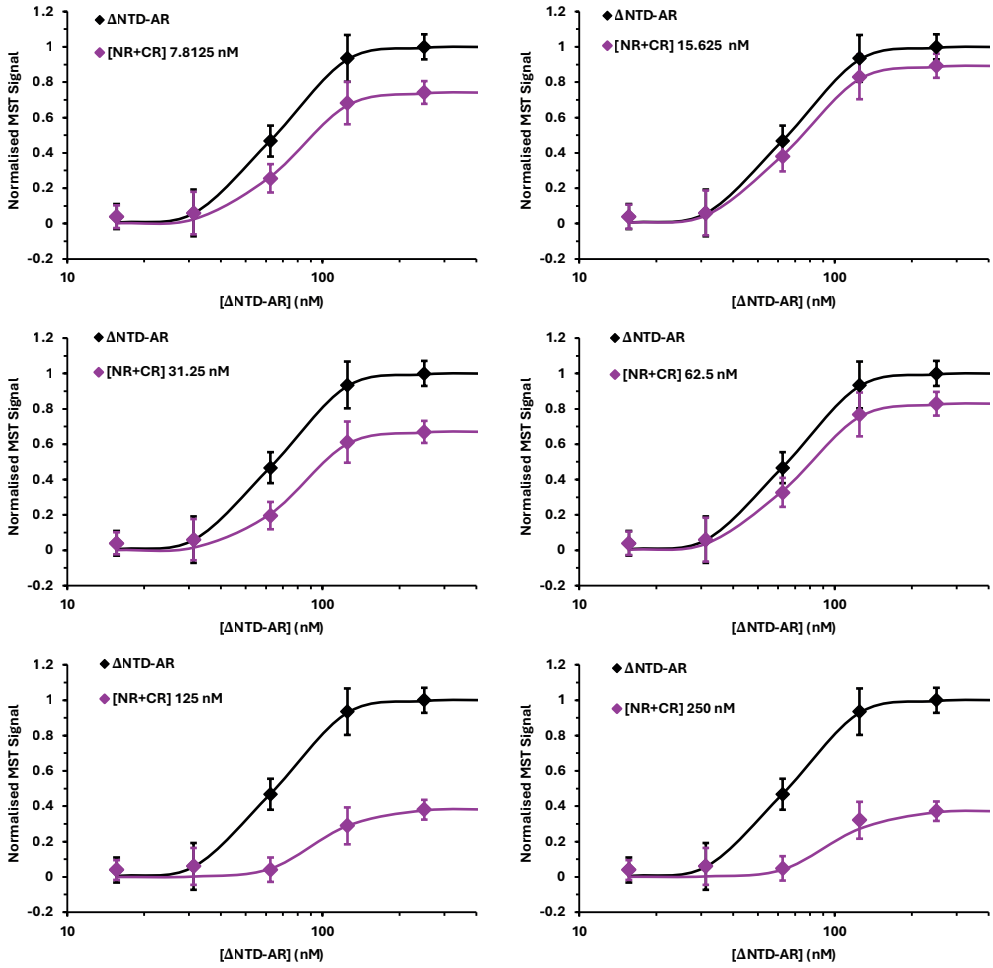


Figure S4. Normalised MST binding curves of ΔNTD-AR with NR+CR. *Without (black) and with NR+CR (purple) in increasing concentrations.*

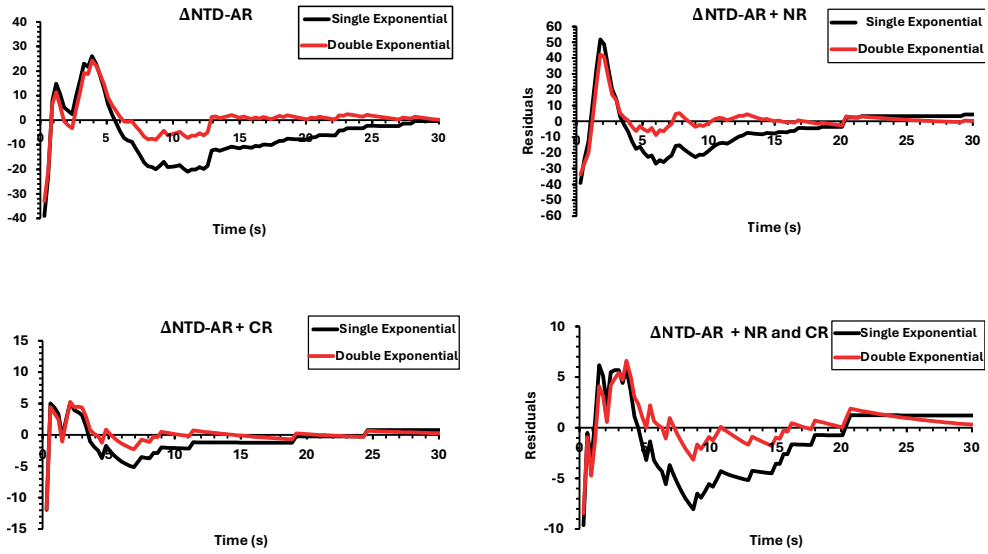


Figure S5. Residual analysis of exponential model fits. *Residuals are shown for single (black) and double (red) exponential fits to the cumulative frequency plots for each condition.*

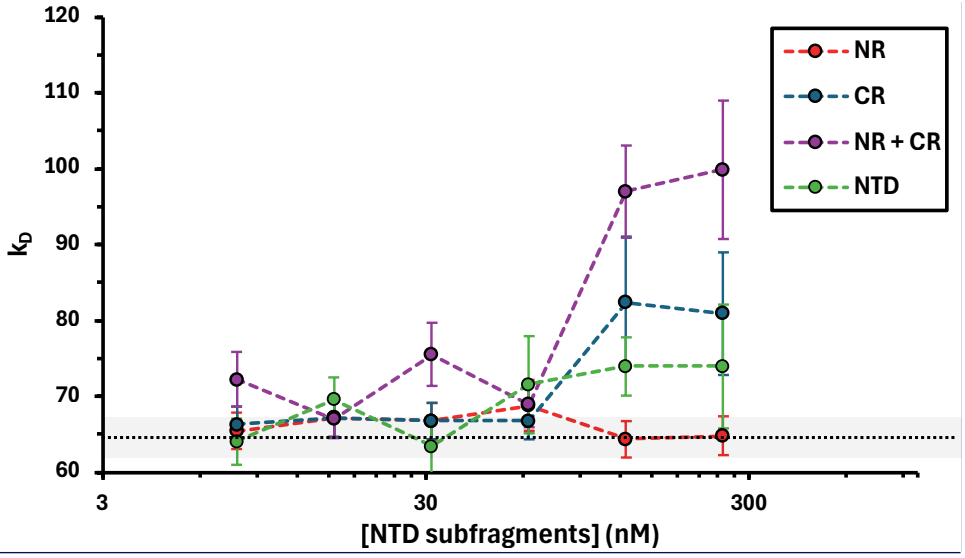


Figure S6. k_D vs [NTD subfragment] plot. *Dotted line with grey shaded area is the k_D ± error for ΔNTD-AR without NTD subfragments added.*

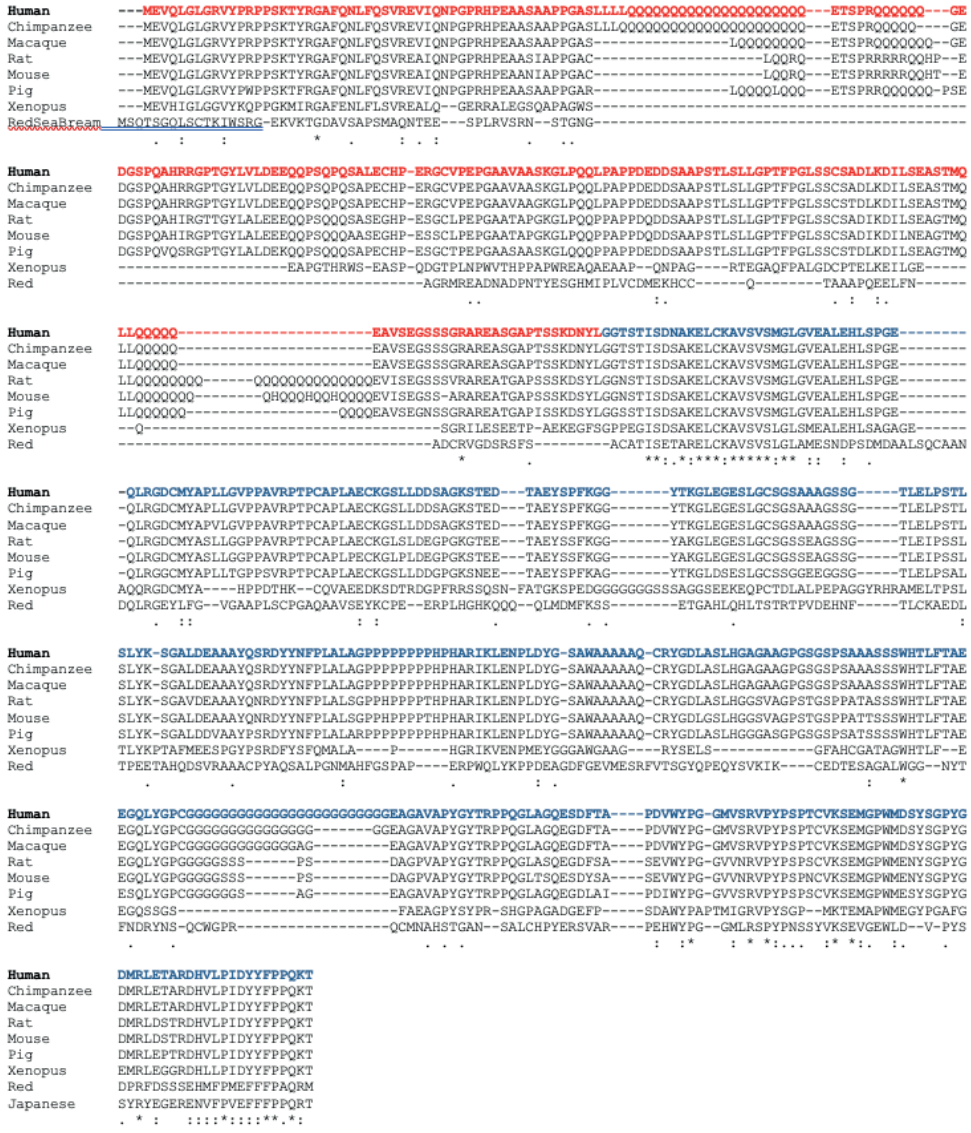


Figure S7. Alignment of the AR-NTD amino acid sequence from selective mammalian, amphibian and fish species. *Alignment generated using CLUSTAL W. The NR (amino acids 1–224) and CR (amino acids 225–538) are coloured red and blue. * denote conserved residues; : denote conservative changes.*

Table S1. Primers for cloning.

Name	Sequence 5' to 3'
NR Forward	AGCAAATGGGTCGCGGAATGGAAGTGCAGTTAGGGCTG
NR Reverse	CGGAGCTCGAATTCGAATTGTCCTTGGAGGAAGTGGGAG
CR Forward	ATATGGCGAGCTCCGGATACTTAGGGGGCACTTCGACC
CR Reverse	TCAGATCTAGAGCTAGCTATTAGCACATGTCCCCGTAAGGTCCG

Table S2. Summary of the detachment rate constants and populations sizes in the different conditions. *n molecules refers to individual molecules counted in at least three separate flowcells.*

	ΔNTD-AR	+ NR	+ CR	+ NR and CR
Fast Rate Constant (s^{-1})	0.461	0.760	0.906	0.621
Population size (%)	87	68	92	83
Slow Rate Constant (s^{-1})	0.072	0.167	0.111	0.124
Population size (%)	13	32	8	17
Fit error	0.005	0.008	0.008	0.005
***n* molecules**	588	514	130	170

Table S3. Hydrodynamic radii of ΔNTD-AR with and without NTD fragments.

	ΔNTD-AR Size (nm)	Concentration NTD fragment (nM)	NR Size (nm)	CR Size (nm)	NR/CR Size (nm)
Theoretical	2.01	7.8125	2.72 ± 0.26	4.37 ± 0.77	3.44 ± 1.57
Measurement	4.37 ± 0.52	15.625	4.37 ± 0.43	4.37 ± 0.67	4.37 ± 1.20
		31.25	3.44 ± 1.78	3.44 ± 0.39	4.37 ± 1.61
		62.5	3.44 ± 0.79	3.44 ± 0.43	7.02 ± 0.91
		125	3.44 ± 0.91	3.44 ± 0.55	4.37 ± 0.74
		250	5.54 ± 0.73	3.44 ± 0.35	4.37 ± 0.63