

## Supplementary Information

### Polyglutamine expansion induced dynamic misfolding of Androgen Receptor

Laurens W.H.J. Heling<sup>1,2</sup>, Vahid Sheikhhassani<sup>1,2</sup>, Julian Ng<sup>1,2</sup>, Morris van Vliet<sup>1</sup>, Alba Jiménez-Panizo<sup>3</sup>, Andrea Alegre-Martí<sup>3</sup>, Jaie Woodard<sup>4</sup>, Willeke van Roon-Mom<sup>5</sup>, Iain J McEwan<sup>6</sup>, Eva Estébanez-Perpiñá<sup>3</sup>, Alireza Mashaghi<sup>1,2\*</sup>

<sup>1</sup> Medical Systems Biophysics and Bioengineering, Division of Systems Pharmacology and Pharmacy, Leiden Academic Centre for Drug Research, Leiden University, the Netherlands

<sup>2</sup> Laboratory for Interdisciplinary Medical Innovations, Centre for Interdisciplinary Genome Research, Leiden University, the Netherlands

<sup>3</sup> Department of Biochemistry and Molecular Biomedicine, Institute of Biomedicine (IBUB) of the University of Barcelona (UB), Barcelona, Spain

<sup>4</sup> Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA

<sup>5</sup> Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

<sup>6</sup> Institute of Medical Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Scotland, UK

\*Correspondence to Alireza Mashaghi. Email: [a.mashaghi.tabari@lacdr.leidenuniv.nl](mailto:a.mashaghi.tabari@lacdr.leidenuniv.nl)

### Supplementary materials: Wild-type pQ/LBD interaction analysis

#### A. pQ20 interacts with BF3 pocket in LBD which may promote FQNLF/AF-2 interactions

In order to understand the impact of pQ expansion on the interaction between NTD and LBD of AR (termed the N/C interaction), we first explored the behaviour of wt-PQ in this process. Thus, we investigated the interaction between isolated pQ20 and AR-LBD (PDB: 1T7T) using docking analysis. Our analysis revealed three distinct interacting regions on the surface of the protein, located mainly between amino acids 670-730, 770-795, and 830-900 (Fig S10A). Of these, the first region, previously identified as the BF-3 region, interacts with the main cluster of pQ20 poses (Fig S10B), composed of 12 members, and was thus selected for additional interaction analysis. Our findings showed that seven distinct amino acids on the LBD surface are involved in the peptide-protein binding interface, involving both polar and nonpolar contacts with an almost equal interface area (Fig S10B). Furthermore, two hydrogen bonds were observed between the side chain of Asn833 and the main chain of Gly724 with two glutamine residues (Gln 1 and 4) in the peptide structure, which may contribute to stabilizing the pQ20-LBD interactions.

To investigate possible conformational changes induced by pQ20 binding to BF-3, we performed an all-atom MD simulation for 300 ns using the best pQ20 pose in complex with LBD as the initial structure. Interestingly, LBD showed almost the same RMSD changes over the simulation time, independent of pQ20 interaction, indicating that global conformational dynamics of LBD did not alter upon interaction with pQ20. However, higher average RMSF values were recorded in three regions, amino acids 690-695, 770-774, and 855-857, upon pQ20-LBD interaction. These residues are in helices 3, 5 and 9, respectively. Finally, representative structures of AR-LBD and pQ20-LBD complex were superimposed to examine the structural differences (Fig S11). Several structural changes were identified in LBD due to the pQ20 binding, compared with LBD in isolation. In particular, helices number 4, 5, and 9 rotated 3° and helix 12 6° with respect to their original positions. Moreover, the area of the AF-2 groove was expanded by moving helices 3, 4, 9, and 12 away from the groove, accompanied by moving the helix 5

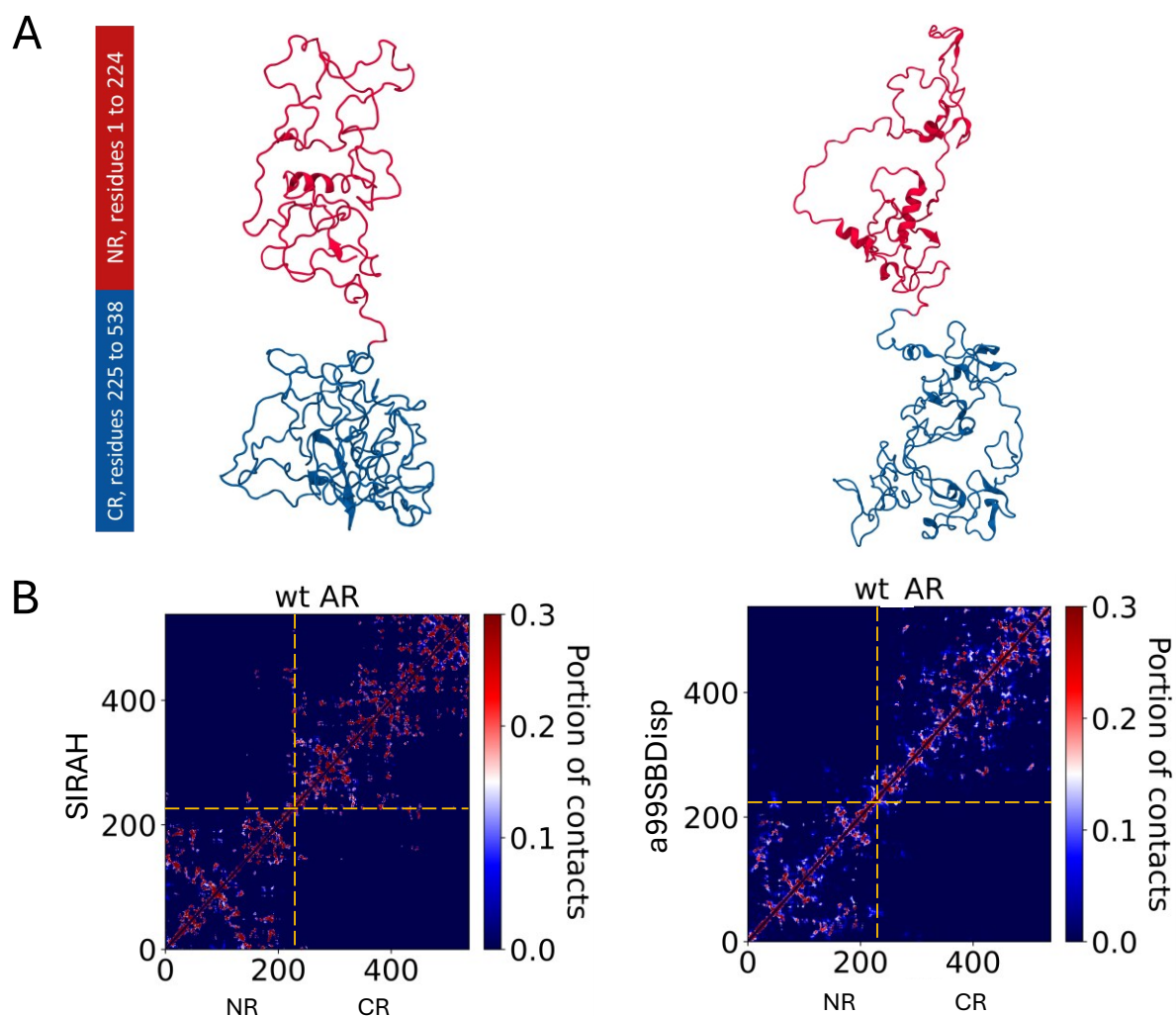
towards the groove, which kept the general configuration of the AF-2 intact. Overall, our findings provide insights into the structural basis of pQ20-LBD interactions and potential conformational changes induced by this interaction.

## B. Methodology: AR-LBD interactions analysis with polyQ tract

AR-LBD structure in complex with its natural ligand was available through the protein data bank. Prior to performing molecular docking, a total of 100 ns were simulated with a time step of 2 fs for relaxing the AR-LBD structure. All MD simulations were carried out using GROMACS package. Neighbour searching was performed every 20 steps. The PME algorithm was used for electrostatic interactions with a cut-off of 1.2 nm. A single cut-off of 1.221 nm was used for Van der Waals interactions. Temperature and pressure coupling were done with the V-rescale and Parrinello-Rahman algorithms respectively.

Due to the technical limitation of peptide-protein interaction analysis algorithms, which accept a maximum of 20 residues for peptide structures, we performed wt pQ-LBD interactions using a truncated length of 20 Q residues. In this study, we used pepATTRACT<sup>1</sup>, a flexible peptide-protein docking approach in ATTRACT, to find the interaction sites of the wt-polyQ tract (Q20) with AR-LBD. We used the sequence of the peptide as the initial input for the peptide structure. Despite these constraints, to broaden our analysis, we overcame the limitations by employing the ClusPro rigid body protein-protein docking algorithms to dock Q23 against LBD. Remarkably, the highly ranked poses from ClusPro align well with the results obtained from pepATTRACT. However, acknowledging the higher precision of flexible docking algorithms in predicting residue coordinates at the binding interface, we specifically selected the top 50 highest-ranked peptide poses derived from pepATTRACT for further detailed analysis. The energy-minimized structure of the LBD in both analysis was used as the receptor protein.

## Supplementary Figures



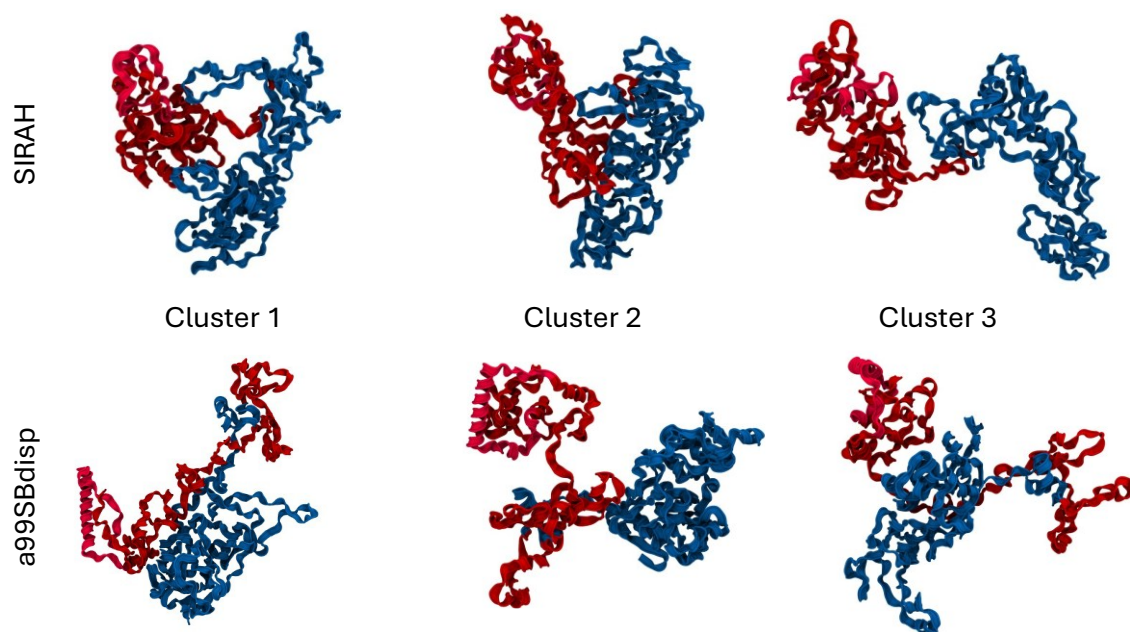


Figure S2 Three largest clusters from SIRAH and all-atom a99SBdisp simulations of pQe NTD. NR and CR residues are depicted in red and blue respectively.

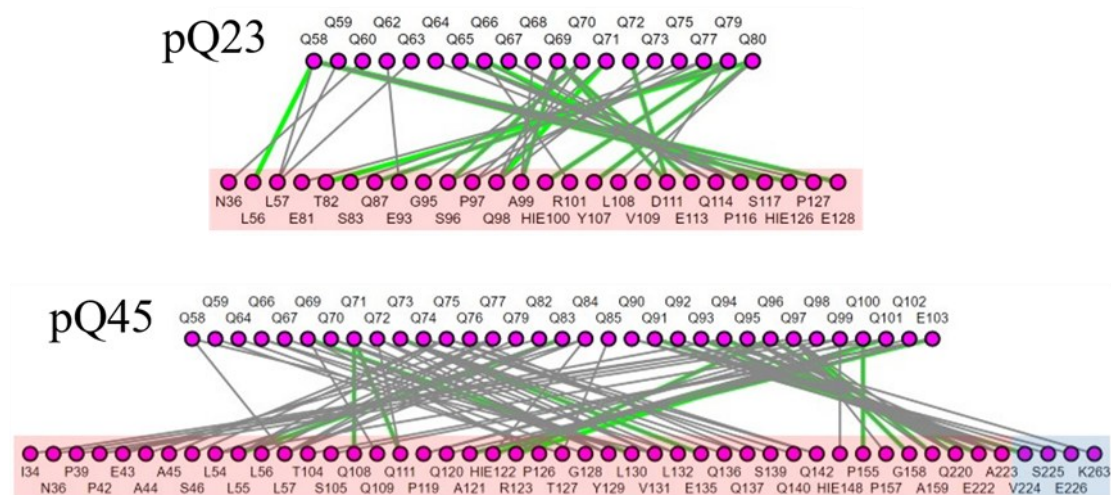


Figure S3 Interaction network of the polyQ tracts in wt-NTD and pQe-NTD with other NTD residues. NR residues and CR residues are depicted with red and blue boxes respectively. In this interaction analysis hydrogen bonds are shown in green and close proximity contacts are colored in grey.

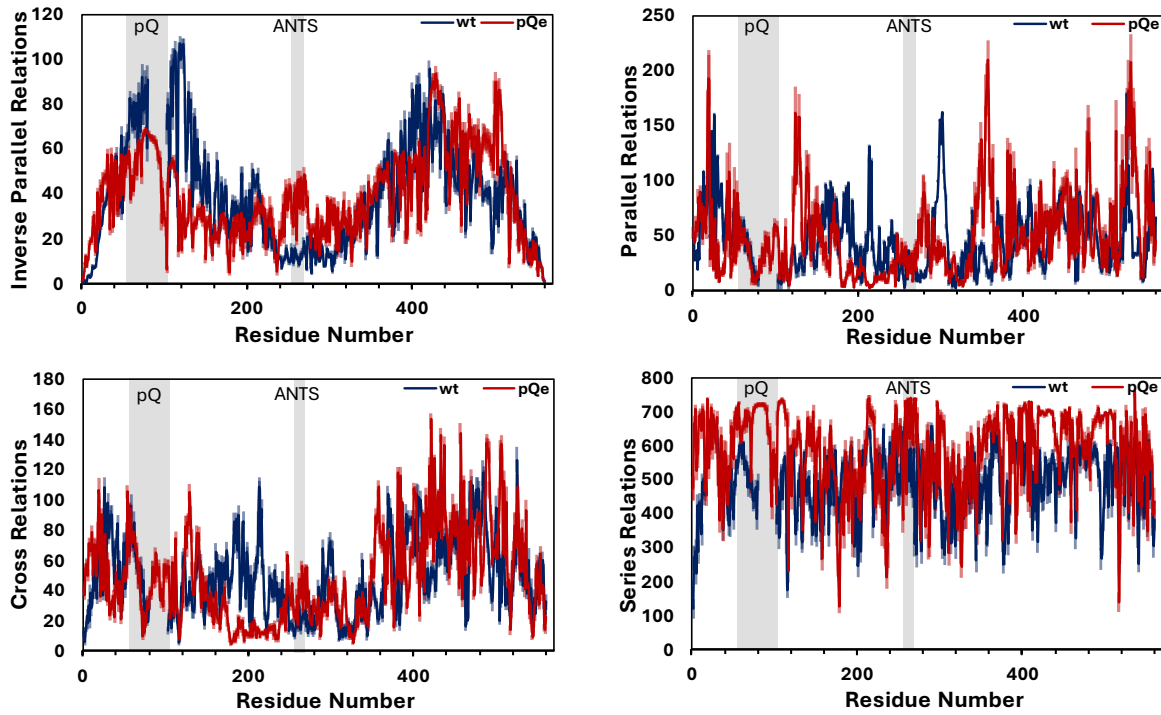


Figure S4 Residue-based local circuit topology for wt and pQe AR-NTD. Shown are mean number ( $\pm$  SEM in shading) of inverse parallel (IP), parallel (P), cross (X), and series (S) relations as a function of residue number. The pQ region and ANTS motif are highlighted by the grey boxes.

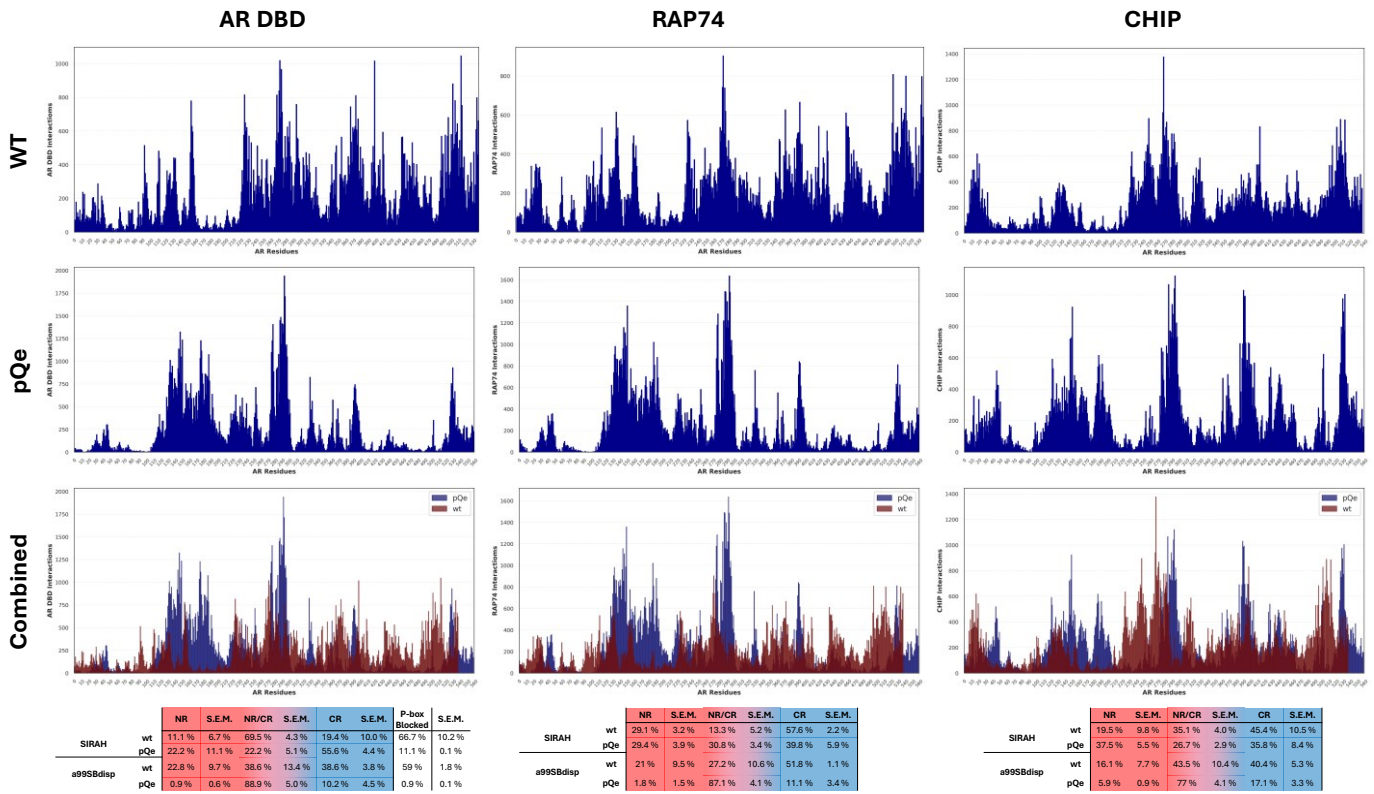
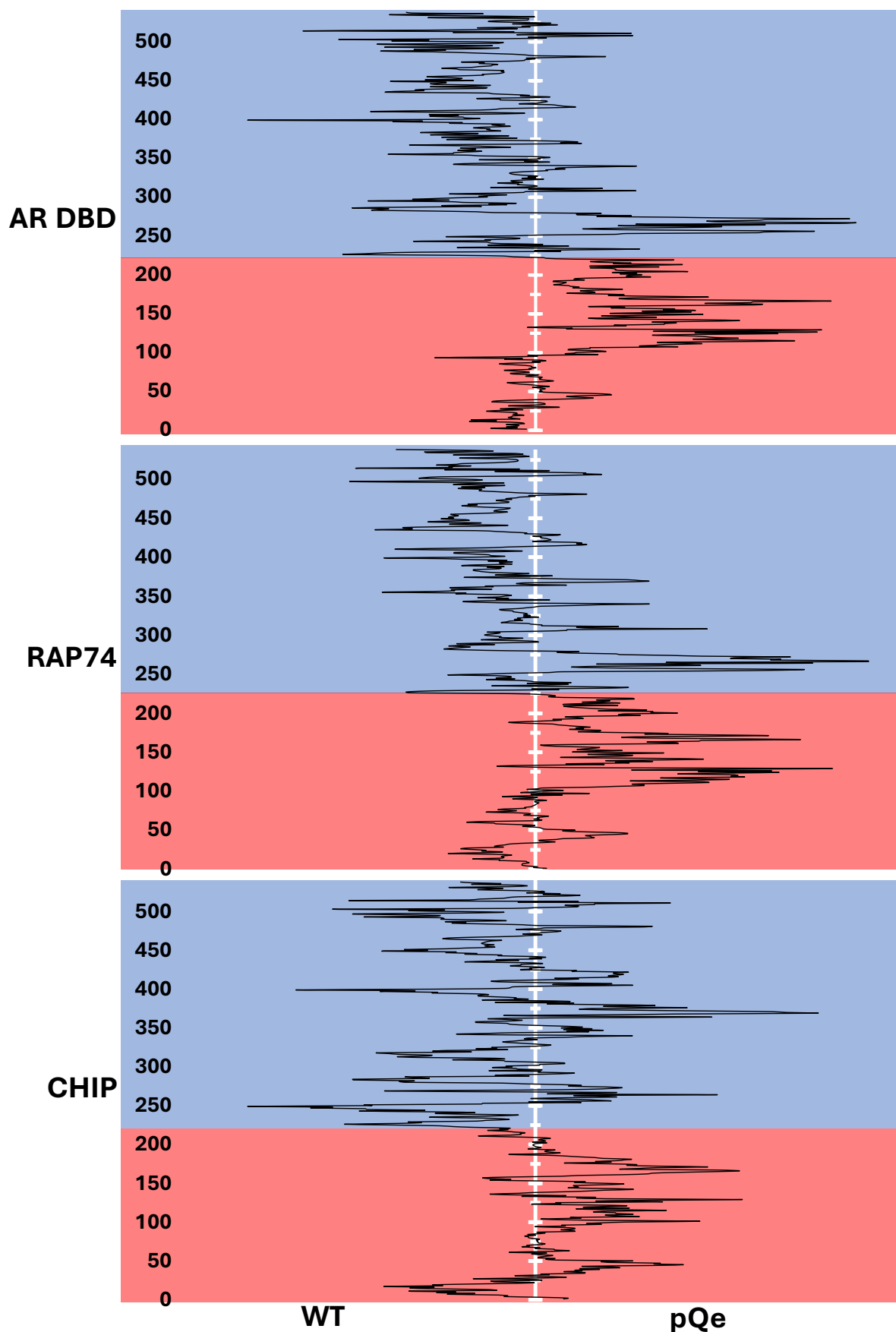


Figure S5. wt and pQe interaction histograms with DBD, RAP74 and CHIP based on ClusPro poses using our aa99SBdisp derived models. Height of the bars indicate the total interaction count of that residue. The tables below show the percentages of exclusive interactions with NR and CR or both. Errors are shown as S.E.M.





## Interaction Shift

Figure S6. *pQe* shifts interactions with partner proteins. Normalized Interaction Shift plot of *aa99SBdisp* derived models for wt and *pQe* AR NTD interacting with DBD, RAP74 and CHIP based on ClusPro models. Deviation of the black line from the central axis indicates the shift of interactions of that residue in wt (left side of the vertical axis) and *pQe* (right side of the vertical axis). Colors indicate the NR (red) and CR (blue) regions of the AR-NTD.

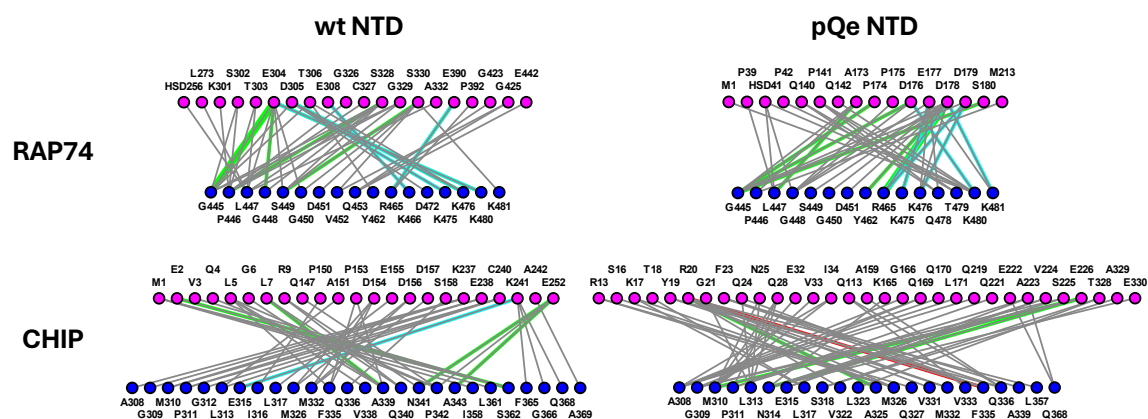


Figure S7. Interactions network of wt- and pQe-NTD residues (purple spheres) with RAP74 and CHIP (blue spheres). Different types of interactions are colored differently: Hydrogen bonds (green), Salt Bridge/Ionic (cyan),  $\pi$ -Cation (red), and  $\pi$ -Stacking (blue).

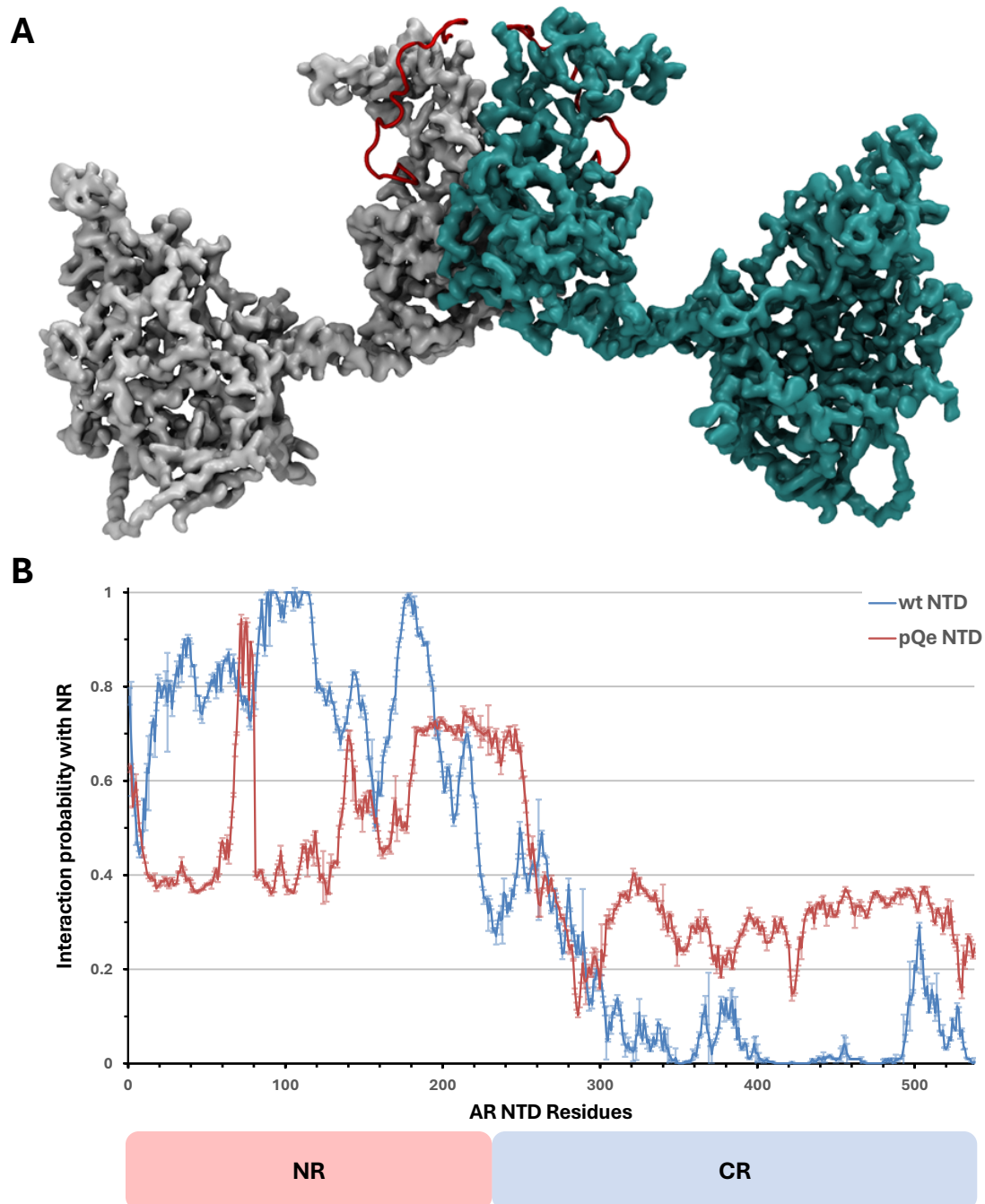


Figure S8. AR dimerization is altered in pQe (A) wt-wt dimer (based on CG models) demonstrating a head-head conformation. (B) Residue based probability of interacting with NR residues in the other AR molecule in the dimer (based on all-atom derived models). In wt, the NR shows a high probability of interacting with the NR of the other monomer, and a lower probability to interact with the CR. This trend is lost in pQe, showing a relatively similar probability across the chain. Errors are shown as S.E.M. over 3 repeats.

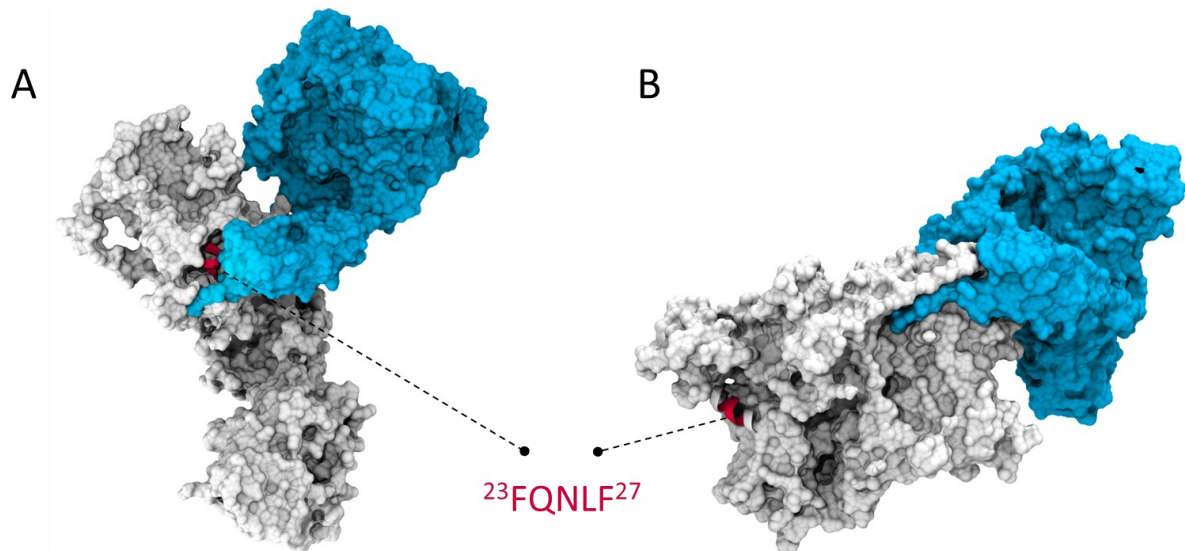


Figure S9. HSP70 interactions with NTD are altered upon pQ expansion. Representative structures of (A) wt-NTD and (B) pQe-NTD in white docked using ClusPro with HSP70 (Blue; structure from AlphaFold: AF-P11142). The interaction of HSP70 with the FQNLF motif of AR-NTD (depicted in red) seen in wt-NTD is lost upon pQ expansion. It is possible that the loss of this interaction is the basis for reduced solubility of AR in SBMA.



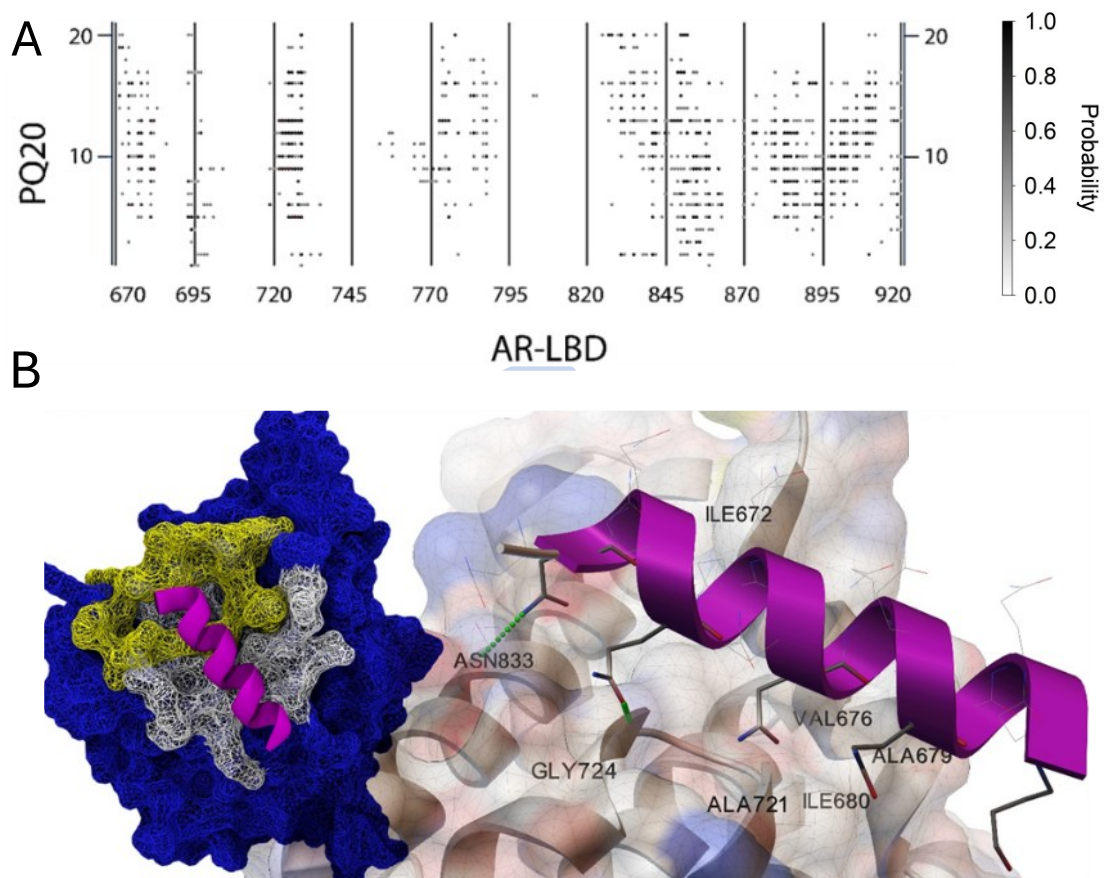


Figure S10. The pQ tract binds to the LBD through the BF-3 pocket. (A) Consensus map of the top 50 ranked poses of the PQ20AR-LBD complex. The grayscale density indicates the probability of contact formation between interacting residues. (B) interacting region of PQ20-AR-LBD (white) shows a considerable surface overlap with BF-3 pocket (yellow). Interacting residues on the surface of AR-LBD. Small spheres in green shows hydrogen bonds formed between GLN residues of the peptide and ASN 833 and GLY 727 or the LBD.

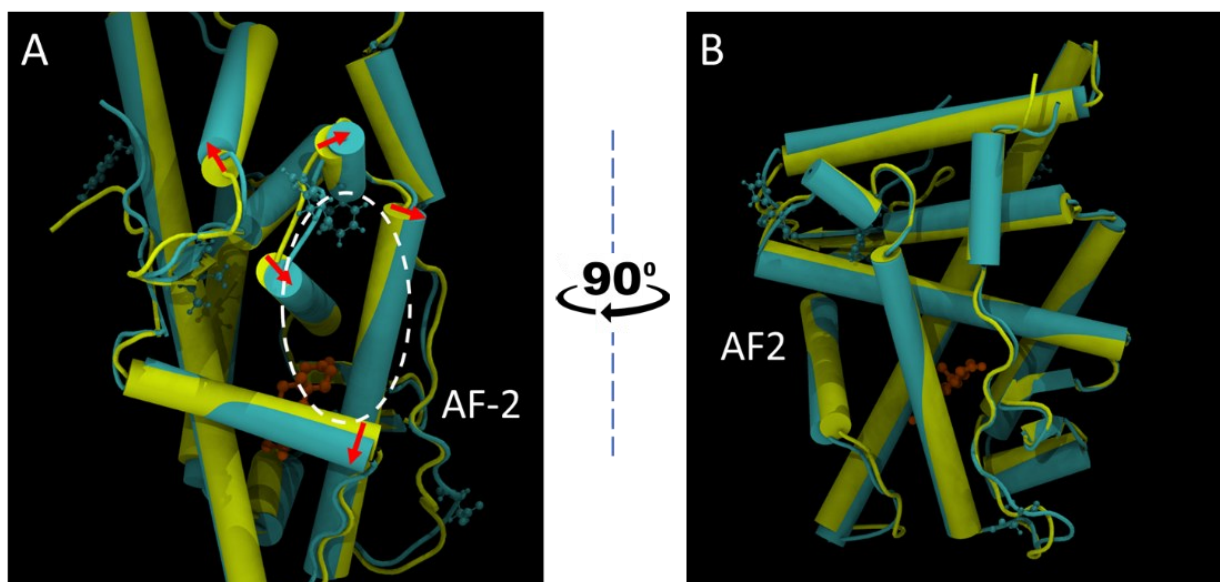


Figure S11. pQ interactions with BF-3 induce conformational change in LBD, opening the AF-2 pocket. (A) Superimposition of representatives of PQ20-LBD and AR-LBD in isolation from AF-2 view and (B) from side view by 90° clockwise rotation. AR-LBD is drawn in yellow, and LBD structure isolated from PQ20-LBD complex in cyan. The red arrows on (A) mark helices number 4, 5, and 9 which rotated 3° and helix 12 6° with respect to their original positions.

	<b>Force Field</b>	<b>Time</b>	<b>Starting Condition</b>	<b>Runs</b>
wt-NTD	SIRAH	5 $\mu$ s	I-TASSER	3
wt-NTD	a99SBDisp	1 $\mu$ s	Stretched chain	3
pQe-NTD	SIRAH	5 $\mu$ s	I-TASSER	7
pQe-NTD	a99SBDisp	1 $\mu$ s	Stretched chain	3

*Figure S12. Table of the structures and subsequent simulations with forcefields, times and starting conditions.*