Biophysics of disordered nuclear receptors and their DNA binding regulation

Heling, L.W.H.J.

Chapter 2

Polyglutamine Expansion induced Dynamic Misfolding of Androgen Receptor

Publication associated with this chapter:

Laurens W. H. J. Heling, Vahid Sheikhhassani, Julian Ng, Morris van Vliet, Alba Jiménez-Panizo, Andrea Alegre-Martí, Jaie Woodard, Willeke van Roon-Mom, Iain J. McEwan, Eva Estébanez-Perpiñá, Alireza Mashaghi. *Polyglutamine expansion induced dynamic misfolding of androgen receptor.* Protein Science. **34**(6):e70154 (2025).

Abstract

Spinal and Bulbar Muscular Atrophy (SBMA) is caused by a polyglutamine expansion (pQe) in the N-terminal transactivation domain of human androgen receptor (AR-NTD), resulting in a combination of toxic gain- and loss-of-function mechanisms. The structural basis of these processes has not been resolved due to the disordered nature of the NTD, which hinders experimental analyses of its detailed conformations. Here, using extensive computational modelling, we show that AR-NTD forms dynamic compact regions, which upon pQe re-organise dynamically, mediated partly by direct pQ interaction with the Androgen N-Terminal Signature (ANTS) motif. The altered dynamics of the NTD result in a perturbation of interdomain interactions, with potential implications for binding of the receptor protein to its response element. Oligomeric aggregation of the dynamic misfolded NTD exposes pQe, but blocks tau-5 and the FQNLF motif, which could lead to aberrant receptor transcriptional activity. These observations suggest a structural mechanism for AR dysfunction in SBMA.

Introduction

The Androgen Receptor (AR/NR3C4) is a crucial ligand-activated transcription factor, which is mutated in several human pathologies[1]. Spinal and Bulbar Muscular Atrophy (SBMA)[2,3], also known as Kennedy's Disease, is a neuromuscular condition that affects approximately 1 in 40,000 men[4] and is characterised by the progressive loss of motor neurons in the brainstem and spinal cord, leading to muscle atrophy and weakness in bulbar and extremity muscles[5]. The toxicity of AR in SBMA is dependent on the presence of androgens, testosterone, and its more potent derivative, dihydrotestosterone, while alterations in cellular processes ultimately lead to cell dysfunction and cell death[6–11]. The onset and progression of SBMA is associated with an expansion of the CAG trinucleotide repeat in the part of the gene that encodes the N-terminal transactivation domain

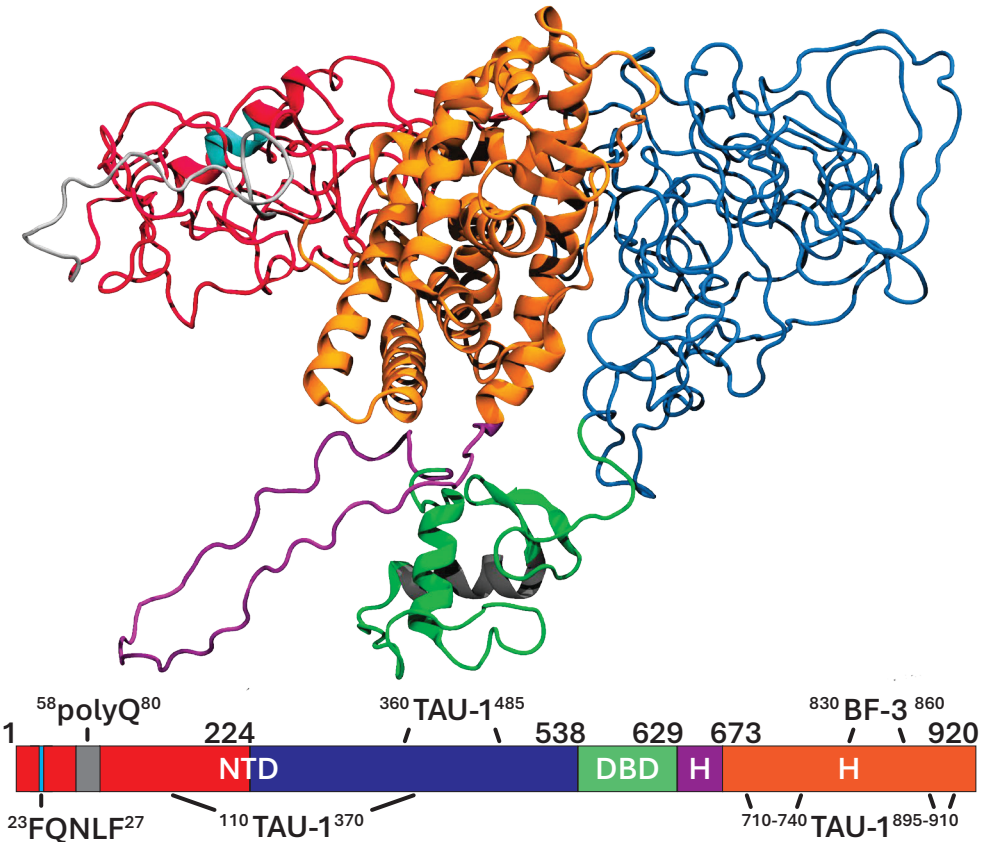


Figure 1. Full length model of AR-NTD colour coded by region. *Red: NR; blue: CR; green: DBD; purple: hinge region and orange: LBD. The grey α-helix in the DBD contains the DNA-interacting P-Box. Model combines the Alphafold structure of DBD-LBD and the computationally derived wt-NTD model published previously[18]. Model taken from Sheikhhassani et al (2022)[18]. The graphic underneath highlights the linear structure of AR, with the subregions and corresponding residue numbers. It also contains some motifs important for AR function.*

(NTD), the intrinsically disordered protein (IDP) region of AR. This longer CAG repeat is translated in an expanded polyglutamine (pQ) stretch (henceforth referred to as pQe) in the protein. However, the molecular mechanisms and cellular pathways that lead to neuronal dysfunction and cell death remain poorly understood. For SBMA and other AR-related pathologies the underlying mechanisms remain unclear, in part, due to the limited structural models of AR.

AR is a member of the nuclear receptor superfamily[12] and displays a distinct modular structure consisting of three structural domains **(Figure 1)**. While the ligand binding domain (LBD)[13–15] and DNA binding domain (DBD)[16] have been structurally characterised to atomic level as isolated modules, a high-resolution experimental structure of the NTD and full-length AR remains elusive, though low-resolution models are available[17]. In a recent computational study, we presented a first insight into a high-resolution atomistic model of wild-type (wt) AR-NTD[18]. Our study revealed that the AR-NTD forms two spatially segregated 3D regions, the N-terminal Region (NR) (residues 1–224) and C-terminal Region (CR) (225-538) **(Figure S1A)**. The overall shape and orientation of the computational model agreed well with the cryo-EM image of the AR[17]. This work provided a first 3D conformational model of the NTD, paving the way for the analysis of native and mutated forms of AR.

pQ tracts occur frequently in the human proteome, predominantly in the intrinsically disordered regions of transcription factors[19], yet their biological function is poorly understood[20]. One hypothesis is that the pQ tracts regulate activity by altering the stability of the complexes they form[21] and that contractions and expansions have functional implications subject to natural selection[22]. Variations in length of pQ tracts have been linked to nine hereditary neurodegenerative diseases, including Huntington's disease and SBMA, colloquially known as the pQ diseases[23]. The pQ tract in AR-NTD is also evidently involved in several other key AR pathologies. Apart from SBMA, variable lengths of this tract in AR have been linked to the onset and severity of prostate cancer[24,25] and infertility[26]. The structural and functional consequences of pQ expansion towards the progression of SBMA is still a matter of debate with several hypotheses proposed. Some suggest that pQe AR-NTD are a neurotoxic species which lead to protein aggregate formation in the cytoplasm and nucleus[27], while others have suggested that pQe AR-NTD itself is neurotoxic, affecting signalling functions[24] and autophagy[28], possibly leading to apoptosis. AR-NTD forms key protein-protein interactions (PPIs) with the other AR domains (DBD and LBD) as well with over 250 different proteins to exert its function[29]. Through intradomain interactions with DBD, the NTD was previously shown to act as an allosteric regulator of DNA binding[30], while LBD interactions, also termed the N/C interaction[31], are seemingly important for transcriptional activity of AR. The interdomain interactions are with an extensive range of co-regulator proteins[32–36]. For example, RNA polymerase II-associating protein 74 (RAP74, a subunit of the TFIIF transcription factor) promotes transcriptional activities of AR while the C-terminal of Heat Shock Protein (HSP)70 Interacting Protein (CHIP) mediates AR ubiquitination. While unknown, any mechanistic process through which pQe affects the

interaction dynamics with these partners is plausible to contain important information regarding molecular mechanisms behind SBMA.

In this study, we employed extensive molecular dynamics (MD) simulations, topological analysis and advanced docking approaches to model the impact of pQe in AR-NTD (pQe: 45 glutamines, a length that has been associated with SBMA) compared to physiologically wild type (wt: 23 glutamines) present in healthy subjects. We investigated whether disease-related pQe tracts can alter conformational dynamics of NTD, monitored its consequences for interactions with DBD and co-regulators, and revealed how the pQe adds to aggregation propensity of AR in SBMA. We built a structural model that describes how the dynamic misfolding (a term we use to make a distinction from misfolding as it happens in structured proteins) of NTD leads to the gain/loss-of function of AR in SBMA.

Results and Discussion

Polyglutamine expansion alters local and global conformations and dynamics

We conducted all-atom MD simulations using a99SB-disp on isolated pQ tracts to understand the effect of expansion on their structural dynamics. Initially, we simulated pQ23 and pQ45 for 500 ns and found that pQ23 adopts an α-helical rod shape **(Figure 2A)**, consistent with previous reports[37]. Our simulations indicate that pQ45 adopts α-helical structures with varying lengths and orientations. This aligns with observations that pQ tracts are structurally plastic, adopting conformations ranging from random coils to α-helices, influenced by factors such as repeat length and sequence context[2,21,38,39] Our observation highlights that flanking sequences modulate pQ structure, either stabilizing helical conformations or promoting aggregation-prone assemblies[2,37,40,41]. Studies have shown that coiled-coil domains flanking pQ tracts play a role in stabilizing aggregates, enhancing their formation[21]. The conformational heterogeneity and helical propensity of pQ tracts further supports the importance of flanking regions[42], which influences aggregation. These studies suggest that factors such as charged or polar residues and the propensity to form coiled-coil structures adjacent to pQ tracts influence solubility and structural stability, thereby modulating aggregation pathways[43]. The pQ repeat length has also been shown to directly influence the folding and conformational dynamics of the AR-NTD, affecting its structural and functional state[2]. These context-dependent behaviours thus raise the question of how an expanded pQ tract impacts the global architecture and dynamicity of AR-NTD.

To address this, we performed coarse-grained SIRAH MD simulations on full-length NTD with a pQe stretch, using our previously published AR modelling approach[18]. To ensure our initial conditions were independent of the results, we performed these simulations multiple times (n=7) with different starting velocities. In contrast to wt-NTD **(Figure S1)**,

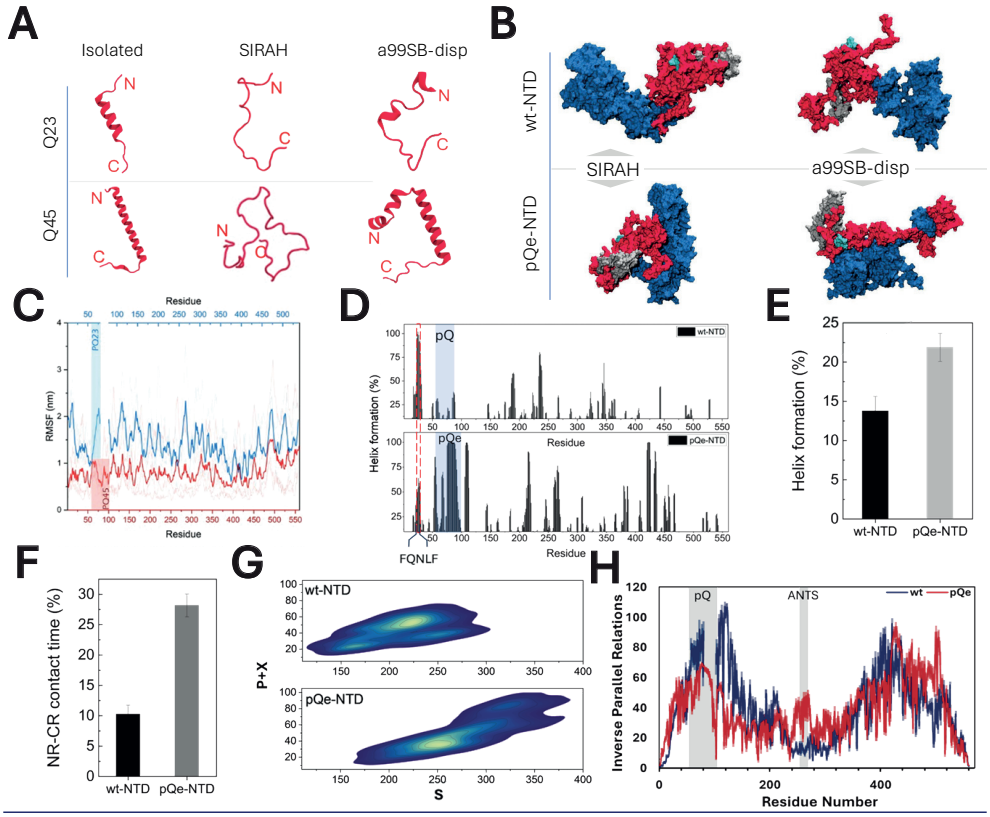


Figure 2. pQe results in local misfold and global alterations of protein dynamics. (A) *pQ tracts at the end of the simulations taken from "isolated simulations", the SIRAH and a99SB-disp models. The results indicate that Q23 and Q45 tracts adopt different conformations.* ***(B)*** *pQ expansion results in an altered global conformation, both from SIRAH and a99SB-disp forcefields. NR is coloured red, CR blue, the pQ region in silver and the FQNLF motif in cyan in all models.* ***(C)*** *Root mean squared fluctuations (RMSF) calculated per residue indicates that pQ expansion reduces chain dynamics.* ***(D)*** *Residue-based α-helix analysis of wt- and pQe-NTD during the last 850 ns of all-atom MD simulations. Residues that are involved in a stable helical formation (probability >30%) are observed in the FQNLF region and sections of the pQ and pQe tracts.* ***(E)*** *Quantitative secondary structure analysis reveals that pQ expansion increases the α-helical content from 13.7 to 21.8% (error: STD over three simulations).* ***(F)*** *Contact time between the CR and NR regions of the wt-NTD and pQe-NTD. It is defined as the fraction of time during the last 850 ns of simulations when residues from the NR and CR were in contact. (error: STD over three simulations).* ***(G)*** *Two-dimensional density plots comparing the topological configuration space of the wt-NTD and pQe-NTD. Each contour plot shows the distribution of combined P+X topological relations as a function of S topological relations, aggregated over three independent simulation replicates. The contour levels represent the probability density of sampled configurations, with lighter shades indicating regions of higher probability. The axes represent the number of residues (× 1000).* ***(H)*** *Residue-based local circuit topology for wt and pQe AR NTD. The mean number of inverse parallel relations (± SEM, shown as shades) is shown as a function of residue number. The pQ tract and ANTS motif are highlighted by grey-shaded areas.*

we observed that within 2 μs, the expanded pQ tract forms a globular structure different from what was observed for isolated pQ45 **(Figure 2A)**, suggesting local misfolding upon expansion influenced by flanking residues. Importantly, the pQe-NTD forms a dynamic globular conformation whereby the segregation of the dynamic subregions is less prominently observed as in wt-NTD **(Figure 2B)**. Furthermore, the pQ tract as well as the other physiologically important motifs, like the ^{23}FQNLF^{27} and the KELCKAVSVSM (also known as the Androgen N-terminal Signature (ANTS) motif[44,45]: UniProt residues 233–246, pQe-NTD residues 255–268) remain surface exposed, despite the drastic architectural changes. We continued the simulations a further 3 μs (for all 7 runs) to follow its dynamicity over a longer time. The molecule showed high levels of structural dynamics, yet the dynamicity of pQe-NTD was reduced when compared to wt-NTD **(Figure 2C)**. However, it still maintained a larger root-mean-square fluctuation (RMSF) when compared to a folded protein, like AR-LBD[46].

Given the importance of our observations, we decided to validate the conformational patterns of the NTD with highly accurate but computationally expensive all-atom simulations using a99SB-disp, a forcefield which is tailored to IDP modelling[47]. To mitigate potential bias introduced by initial structure assumptions, we initiated these all-atom simulations from fully extended conformations for both stretched wt- and pQe-NTD. The all-atom simulations corroborated the dynamic, compact conformations observed in the coarse-grained simulations, particularly the distinct segregation of the N-terminal and C-terminal regions in wt-NTD **(Figure S1)**. Notably, the consistent absence of interregional contacts between these domains across both simulation methodologies **(Figure S2)** reinforces the robustness of our observed structural dynamics. While both simulation approaches effectively captured the overall dynamic nature of the NTD, we observed differences in the representation of secondary structure, specifically regarding α-helix formation within the NR. Quantitatively, the all-atom simulations revealed a significantly higher percentage of α-helical content (13.3% ± 2.1%) compared to the coarse-grained simulations (6.6% ± 0.6%). This discrepancy is consistent with the established accuracy of the a99SB-disp force field in modelling helical structures, a critical feature for IDP dynamics. Remarkably, the 13.3% helical content observed in our all-atom simulations is consistent with circular dichroism results, which reported an α-helical content of 14%[2], providing compelling experimental validation for our computational findings.

Upon pQ expansion, the NR dynamically misfolds and interacts with the CR **(Figure 2B and S1)**. This is reflected by an increased contact time between NR and CR **(Figure 2F)**. In addition, as a result of the global restructuring caused by pQ expansion, the α-helical content in NTD increased by >8% as determined through structural analysis quantification **(Figure 2D and E)** and comparable to experimental observations[2].

To understand the role of the glutamine residues in these altered fold dynamics, we mapped its interaction network with other NTD residues. It shows that the pQe tract formed longer-

range contacts, including with residues in the CR region, a behaviour not observed in wt-NTD **(Figure S2)**. These long-range contacts also included directional hydrogen bonds, which likely explains the lack of regional segregation and reduced dynamicity in pQe. The CR residues that the pQ tract interacts with, partially overlap with the ANTS motif **(Figure S3)**[45].

To quantify the structural differences between wt- and pQe-NTD, we sought to capture key topological features of the protein chain. Unlike traditional geometric methods, topological analysis is particularly well-suited to quantify and categorise the residual structural properties of proteins that are continuously deforming, such as IDPs. By mapping the pQe NTD dynamics onto a topological space, we identified recurring motifs and patterns within the dynamics while showing a recognisable difference with wt-NTD **(Figure 2G)**. To probe these differences further, we tracked the topological dynamics of the NTD over time, employing the circuit topology (CT) framework[48–51] to calculate the residue-based parallel (P), series (S), and cross (X) topologies within the chain's conformations at 25 ns intervals **(Figure S4)**. Residue contacts, defined by spatial proximity, serve as a fundamental topological representation of the chain, offering deep insights into the dynamics, and how they are altered in pQe. This revealed that, in agreement with our contact analysis, the pQ region in wt is enriched in inverse parallel relations **(Figure 2G)**, indicating that it is more sequestered in the structure than in pQe NTD. In contrast, the ANTS motif is more sequestered in pQe NTD **(Figure 2G)**, revealing a long-range contact between the pQ tract and this motif in pQe NTD.

Overall, both SIRAH and a99SB-disp simulations support a model in which wt-NTD adopt segregated dynamically compact regions. Importantly, local misfolding due to pQ expansion propagates to global reorganisation in NTD, in which these segregated regions dynamically merge and unmerge. Next, we investigated the implications of the observed dynamic misfolding for interdomain interactions.

Gain of function, loss of regulation: pQ expansion induced dynamic misfolding alters protein–protein interactions

The roles of AR span genomic and non-genomic pathways[1], which are both tightly regulated in cells. Our previous wt-NTD model suggested a structural mechanism for the regulation of AR-DBD binding to DNA through competing NR and CR interactions. We showed that CR binding would cover the P-Box, the zinc finger of the DBD that binds to the major groove of DNA and thus affect the ability of the DBD to interact with DNA[18]. These observations are consistent with previously published experimental studies, which identified the NTD and stretches of the CR as having DNA binding regulatory functions[30,31,52–54].

To evaluate the impact of pQ expansion on NTD–DBD interactions, we conducted molecular docking using models derived from both coarse-grained and all-atom simulations. Initially, we clustered our simulations, selecting cluster representatives that accounted for at

least 50% of the total simulation frames to ensure adequate conformational sampling. These representatives were then docked against DBD using ClusPro, generating up to 120 potential interaction poses per docking experiment (420 poses for wt-NTD and 406 poses for pQe NTD). Given their ability to capture broader conformational sampling, essential for representing the inherent flexibility of NTD interactions, we prioritised the coarse-grained results in our primary analysis **(Figure 3)**. The all-atom derived models, while providing higher atomic resolution, offer limited conformational sampling. These results are included in the supplementary information **(Figure S5–7)** to provide an additional layer of validation, specifically focusing on the detailed structural aspects of the interactions.

Our findings indicate that the altered structural dynamics of pQe-NTD significantly affect its interactions with the DBD. In the wt-model, the majority of identified interacting residues are localised within the CR region of the NTD. However, this interface was disrupted in the pQe variant **(Figure 3A)**, where interactions with DBD are predominantly mediated by NR residues **(Figure S6)**. Specifically, the ability of the NTD to block the P-box—a critical regulatory site for DNA binding—was markedly reduced, with only 11% of docking poses

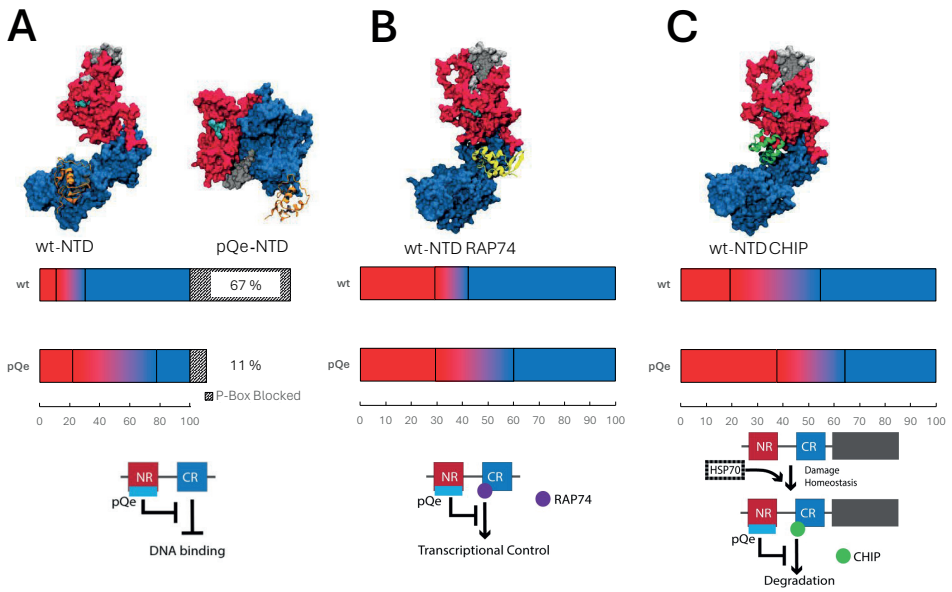


Figure 3. Protein-protein interactions are altered as a result of pQ expansion. (A) *Representative structure of wt- and pQ-NTD interacting with DBD (orange). CR-mediated DNA binding regulation was reduced in pQe-NTD due to exposure of P-Box despite (black) NTD-DBD interactions. The pQ region is coloured in silver and the FQNLF motif in cyan in all models.* ***(B)*** *Representative structure of wt-NTD interacting with RAP74. pQ expansion induces a change in binding sites, reducing transcriptional control.* ***(C)*** *Representative structure of wt-NTD interacting with CHIP. Bar charts indicate the fraction of poses interacting with NR (red) NR/CR (dual coloured) and CR (blue) residues. Graphics underneath highlight the mechanistic changes upon pQ expansion.*

achieving P-box blockade compared to 67% in the wt-model. This reduction aligns with prior reports of increased DNA binding in AR with expanded pQ tracts but contrasts with their observation of reduced AR transcriptional activity[55].

To explore this apparent discrepancy, we used our docking approach to model interactions between pQe-NTD and RAP74, a subunit of the TFIIF complex[56,57]. RAP74 plays an essential role in the formation of the pre-initiation complex, facilitating the specificity and efficiency of RNA Polymerase II recruitment to DNA[58]. Consistent with the observed DBD interactions, our model (based on 414 docking poses for pQe and 417 wt) revealed that the predominantly CR-mediated interactions with RAP74 in pQe-NTD were suppressed, instead having interactions covering both the NR and CR regions **(Figure 3B and S6)** while the types of interactions suggest similar strengths as with wt-NTD **(Figure S7)**. These findings present a structural model illustrating how pQ expansion disrupts regulatory interactions critical to AR transcriptional activity.

NTD has a propensity to aggregate as a result of altered dynamics by pQ expansion

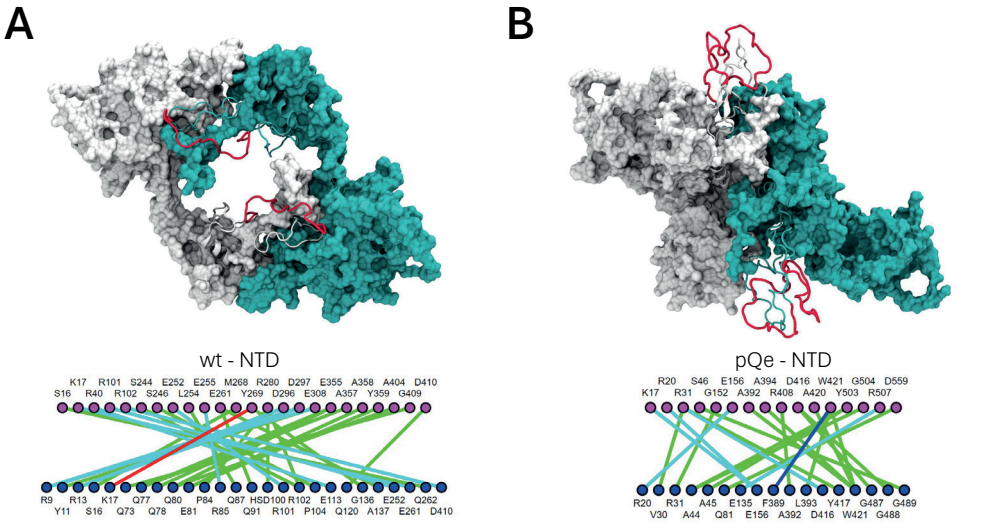
Expansion of the glutamine tracts leads to ligand-dependent protein aggregation in AR[59–64], which in turn leads to nuclear inclusions where AR proteins get proteolysed[65]. This is widely regarded as one of the pathological hallmarks of SBMA. To explore how the altered structural dynamics we have seen in our pQe-NTD model impact oligomerisation, we first performed self-docking studies based on wt-NTD. This dimer model (highest representative cluster of 396 docking poses) demonstrates that the two monomers adopt both asymmetric head-to-tail (N–C) conformations **(Figure 4A)** and symmetric head-to-head (N–N) conformations **(Figure S8A)**. These dimeric poses further reveal a spatial gap between the two molecules, in a manner that is consistent with the low-resolution cryo-EM AR structure[17]. This configuration supports the docking and interaction of the previously characterised active LBD dimer[66] in agreement with the crystal structure of the AR-LBD dimer.

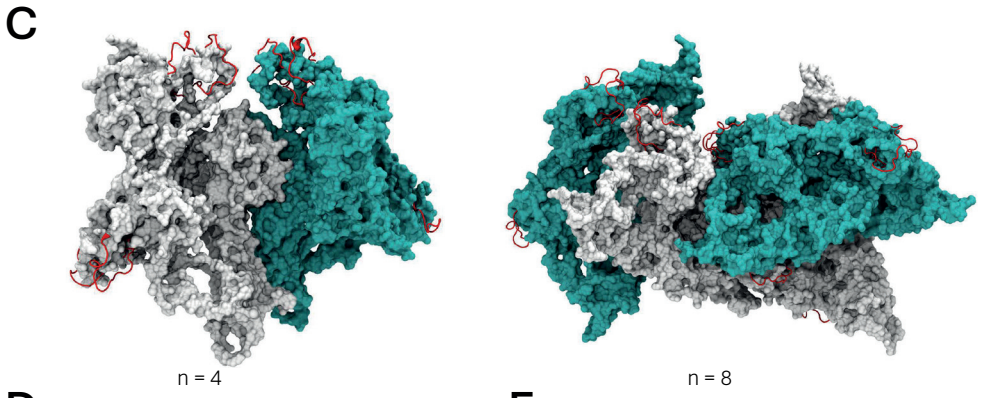
Subsequently, we generated a pQe-NTD dimer (from 426 poses), which illustrates a loss of the spatial gap **(Figure 4B)** and the conformational characteristics seen in the wt-dimer **(Figure S8B)**. Notably, over 50% of the docking poses of pQe-NTD block the FQNLF motif **(Figure 4B** - interaction map **and 4D)**, while the pQ tract and ANTS motif remain exposed on the surface of the dimer. The FQNLF is crucial for AR activation, a process partly regulated by the competition between chaperone proteins such as HSP40 and HSP70, and the AF-2 coactivator binding pocket. NMR data has shown that HSP70 directly interacts with FQNLF in wt-NTD, preventing aggregation *in vitro*[67]. Furthermore, modulating HSP70 was found to increase solubility, ubiquitination, and clearance of pQe-NTD[67]. Hsp70 cycles between ATP- and ADP-bound states, with the ADP-bound state exhibiting significantly enhanced substrate binding affinity[68]. This state allows HSP70 to discriminate between properly folded and damaged proteins. The most effective HSP70

compounds stabilised it in the ADP-bound states[69], which suggests that, in the case of pQe-AR, HSP70 could wrongly release AR-NTD. To explore chaperone interactions further, we examined the binding of HSP70 and CHIP with both wt-NTD and pQe-NTD. We observed that while HSP70 interacts through the region with the FQNLF in wt-NTD, as reported in the NMR data, this interaction is absent in pQe-NTD **(Figure S9)**. While CHIP interacted with wt-NTD primarily through CR residues and in the groove covering NR and CR, this interaction was significantly reduced in pQe-NTD **(Figure 3C)**. Instead, interactions with NR residues increased, including π-Cation bonds **(Figure S7)**. Together these results suggest that in the absence of HSP70 modulation, pQe-AR is less likely to be soluble and ubiquitinated for degradation **(Figure 3C**, model**)**.

Previous reports have shown that pQe-NTD also undergoes an altered oligomerisation process compared to wt-NTD, forming fibrils instead of annular oligomers *in vitro*[70]. To explore if our pQe-NTD model could provide atomistic details to this process, we built a higher-level oligomerisation model, factoring in tetramer and octamer pQe NTD models **(Figure 4C)**. Our results mirror those observed in dimerisation, where the pQe tracts and ANTS motifs of the individual NTDs are exposed on the surface, while motifs with functional implications like the Transactivation Unit (Tau)-5 and FQNLF, mediate interactions between individual molecules. Oligomerisation to an octamer demonstrates a clear tendency to form a fibre. While these specific oligomeric species (n=4 and n=8 AR molecules) have not been directly observed *in vitro* or *in vivo*, the protofibril model we have produced is in agreement with AFM observations[70]. This data allowed us to build a model of multimerisation of pQe-NTD **(Figure 4E)**, from a monomer, via oligomers into protofibrils. This multistep oligomerisation was also observed *in vitro* and *in vivo*, where pQ expansion in the NTD causes oligomerisation beyond dimers via SDS soluble oligomers[67,70,71]. Our model highlights that the expanded pQ tract is not the main interaction interface in this oligomerised phase. Rather it remains exposed on the surface of the protein, together with the ANTS motif. This motif was previously shown to be involved in nucleation of AR-NTD and was suggested to be a regulator for the pQ tract[45] in a manner comparable to the N17 domain in Htt[72]. Our model, which further highlights the importance of these two motifs through its exposure, could signify a further step in oligomerisation, where the protofibrils bind together through these exposed areas together to form stable bundles.

This model of AR-NTD oligomerisation shows similarities with studies on pQe Ataxin-3, a protein associated with another polyglutamine expansion disorder, Machado–Joseph disease. In Ataxin-3, the expansion of the pQ tract leads to misfolding and aggregation in a two-step process[73]. It transitions from SDS-soluble oligomers, driven by the ordinarily structured Josephin domain[74] with the pQ tract exposed, to insoluble aggregates[75], driven by exposed pQ–pQ interactions[76–78]. While the model we present here does not take ordinarily structured domains (AR LBD and DBD) into account, it could be that the altered structural dynamics caused by pQ expansion increases dynamics and aggregation propensity in these domains.





D

	Wild type	PolyQ expanded
FQNLF	Exposed	Blocked
pQ tract	Exposed	Exposed
ANTS	Partially involved	Exposed
TAU-5	Exposed	Blocked

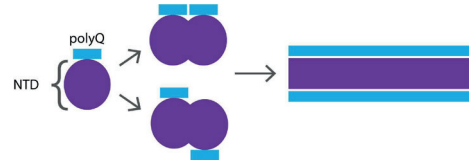


Figure 4. Multimerisation of pQe NTD leads to fibrillar aggregates. *Representative structure and residue interaction map of **(A)** wt-NTD and **(B)** pQe-NTD dimer. The white and cyan structures represent one NTD monomer, the pQ tract is depicted in red. The precise interactions we saw between the two monomers for wt- and pQe-NTD are depicted under the representative structures. Different types of interactions are coloured differently: Hydrogen bonds (green), Salt Bridge/Ionic (cyan), π-Cation (red), and π-Stacking (blue). **(C)** Models of tetramer and octamer of pQe AR-NTD. The exposed pQ tracts are depicted in red. The cyan and white in these models each depict half the total number of proteins in the oligomer. **(D)** Table of functionally important motifs and their involvement in a NTD dimer. "Exposed" means that the motif is exposed on the surface of the dimer,*

while "blocked" means that it is buried in the interaction interface. "Partially involved" means that a subpart of the motif is involved in the dimer interface. ***(E)*** *Aggregation model of pQe NTD, leading from a monomer to a fibril with the pQ tracts exposed.*

The behaviour of pQe-AR observed here therefore aligns with previous reports on polyglutamine expanded proteins and provides a foundational basis for a unifying mechanism for polyglutamine and other misfolding diseases. In these disorders, (toxic) monomeric proteins aggregate into insoluble fibres, a process initially thought to underlie pathogenicity[79]. However, other studies have questioned whether aggregation is the primary cause of cytotoxicity[80]. Could the formation of insoluble fibres be cytoprotective? Previous reports on Huntingtin (Htt) suggest that inclusion body formation reduces the diffusion of Htt protein, thereby decreasing neuronal cell death[81]. Our model, where the FQNLF and other functional motifs are blocked through multimerisation, may suggest a similar protective mechanism, limiting (aberrant) function. Oligomerisation of pQe AR leads to aggregation into insoluble fibres, which, at least initially, may act as a cytoprotective mechanism.

Conclusions

This study presents the first molecular model for the NTD of AR containing an expanded polyglutamine tract, offering unprecedented insights into the structural dynamics and functional implications of pQ expansion. Using extensive *in silico* simulations, topological mapping, and molecular docking analyses, we have provided detailed understanding of the complex intra- and interdomain interaction within AR-NTD, highlighting the effects of pQe on its structural reorganisation, interactions with other AR domain and cofactor binding, particularly RAP74, HSP70 and CHIP.

Our findings reveal that expansion of the pQ tract induces a global dynamic restructuring of the AR-NTD, driven by long-range interactions of the glutamine residues with residues in the CR region. This reorganisation disrupts the native dynamic regional segregation seen in wt AR-NTD. Importantly, these long-range interactions seem to include the ANTS motif, which was previously reported to crosstalk with the pQ tract of AR in the control of aggregation[45]. Interestingly, that study demonstrated that fibrillar aggregation of NTD fragments only occurred with pQe tracts, suggesting that the distance between the glutamine tract and ANTS was of importance. Our pQe NTD model does provide topological information on molecular interactions between the pQe and the ANTS motif. In contrast, this interaction is not seen with the wt-NTD. This solidifies the potential importance of this motif and the positive regulatory crosstalk these have.

Additionally, our model underscores the delicate balance of PPIs. While pQ expansion results in the disruption of critical interactions—such as those between AR and TFIIF through RAP74—leading to impaired gene regulation, it may simultaneously enhance

DNA binding through the loss of self-inhibitory interactions. Furthermore, our results provide atomistic details on oligomerisation of pQe-AR, a process which could represent a cytoprotective mechanism. However, this oligomerisation disrupts key PPIs, including interaction with HSP70 and CHIP, which are essential for protein homeostasis. These alterations in PPIs as a result of pQe mediated dynamic misfolding can ultimately disrupt numerous cellular processes essential for cell survival, from homeostasis to gene expression. Dysregulation of these multiple pathways could culminate in toxicity in a cumulative fashion, suggesting that therapeutic targeting of a single pathway may not provide complete amelioration. In contrast, reports of reducing cellular levels of pQe AR have emerged as an appealing therapeutic strategy to target the proximal mediator of disease pathogenesis[67]. In this regard, our model opens new therapeutic avenues by revealing the distinct interaction patterns between the pQe tract and AR motifs. Targeting these interactions may stabilise or reverse the pathological effect of pQe-AR. Further experiments will be necessary to gain a deeper understanding of the altered transcription and regulation through PPIs by AR-NTD, for which our models can serve as a framework. Preliminary findings **(Supplementary material)** suggest that the interaction between the pQ tract and the LBD may further influence AR function by modulating the N/C interaction[82], which can be affected by neighbouring motifs[83]. Although current computational methodologies limit the study of interactions between disordered regions like NTD and folded domains like LBD, our initial peptide docking results provide intriguing evidence that the pQ tract interacting with a pocket on the LBD **(Figure S10)** could modulate the FQNLF/AF-2 interaction through a conformational change **(Figure S11)**. This binding pocket was previously identified as a highly conserved structural pocket, the BF-3[84]. It is possible that expansion of the pQ tract affects this process in SBMA.

In conclusion, despite the incredible complexity of modelling nuclear receptors and large intrinsically disorder proteins, our approach has generated atomistic-level insights that are consistent with experimental observations. This work not only provides a detailed understanding of the structural dynamics of AR in the context of polyglutamine expansion but also sets the stage for future studies of other disordered proteins and their pathological mutations. By unlocking new mechanistic insights, this model paves the way for the development of novel therapeutic strategies aimed at mitigating the effects of misfolding and aggregation in neurodegenerative diseases.

Methods

Molecular dynamics simulations of AR-NTD

This study presented MD simulation data on wt- and pQe-NTD using SIRAH and a99SDisp. All details of our simulations are depicted in **Table S1**. Initial representations of pQe AR-NTD were generated using I-TASSER[85] and mapped to the coarse-grained (CG) SIRAH force field. The CG models were solved in octahedron boxes with a minimum distance of 1.2 nm from the solute using pre-equilibrated WT4 water molecules. The ionic strength was set to 0.15 M by randomly replacing WT4 molecules with Na^+ and Cl^- CG ions. The system was prepared following standard SIRAH protocols. First, energy minimisation was performed with positional restraints of 1000 kJ mol^{-1} nm^{-2}, followed by a 100 ps NVT simulation at 300 K and a 100 ps NPT equilibration. Production simulations were performed in the NPT ensemble at 310 K and 1 bar using GROMACS 2020[86] with a time step of 20 fs. PME electrostatics were used with the Verlet scheme and a cutoff of 1.2 nm. Solvent and solute were coupled separately to velocity rescale thermostats with coupling times of 2 ps. Pressure was controlled by the Parrinello–Rahman barostat with a coupling time of 10 ps. Production trajectories were generated for 5 µs.

Initial structure preparation for all-atom simulations

Initial structures for all-atom simulations were generated from the I-TASSER structure prediction. Initially, structures were subjected to pulling via steered MD simulations to a fully extended chain. Subsequently, 50 ns all-atom simulations in implicit solvent were conducted to produce new initial structures using the QuickMD module of Visual Molecular Dynamics (VMD)[87] in conjunction with the NAMD molecular dynamics package[88].

All-atom molecular dynamics simulations

All-atom simulations on wt-NTD, pQe-NTD and isolated pQ tracts (Q23 and Q45) were performed using a99SDisp forcefield with compatible amber water molecules. A total of 1000 ns were simulated with a 2 fs time step. Neighbour searching was performed every 10 steps. The PME algorithm was used for electrostatic interactions with a 1 nm cutoff. A single cutoff of 1.006 nm was used for Van der Waals interactions. Temperature coupling was done with the V-rescale algorithm and pressure coupling was done with the Parrinello–Rahman algorithm. Trajectory analysis and visualisation for all simulations were performed using VMD and GROMACS tools.

Clustering

To consider the high dynamicity of the disordered chain, we combined the dynamic phases—the phase of the simulation when the chain reached the plateau in the RMSD trajectory—which corresponds to the last 800 ns of the all-atom simulations and last 3.5 µs of our coarse-grained simulations. We clustered the combined MD trajectories with the GROMOS clustering algorithm in the GROMACS 2020 gmx cluster module. The RMSD cutoff-value for the clusters was set to 0.9 nm.

Circuit Topology analysis

Circuit topology parameters were analysed from the final 850 ns of each aa99SB-disp simulation using custom Python scripts[49,51]. These scripts were modified to include energy and length filtering, as well as circuit decomposition. Residue contacts were defined by a 4.5 Å distance cutoff, requiring at least five atoms from each residue to be within this distance. To avoid local overcounting, the three nearest neighbors of each residue were excluded from the contact analysis.

Residue-based local circuit topology was extracted using MATLAB based on established methods[49,50]. A cutoff scheme of 5 or more atom–atom contacts within 4.5 Angstroms was employed, excluding contacts within 3 residues along the protein sequence. Circuit topology was calculated over the final 700 ns of the all-atom simulations, at a 25 ns timestep, for all three simulations. Error was calculated as SEM over all extracted frames.

AR-NTD protein–protein interactions

The atomic coordinates of the pQe-NTD and wt-NTD were taken from our structural SIRAH and aa99SB-disp models and used as input for PPIs using biased rigid body docking with ClusPro 2.0 webserver[89]. The atomic coordinates of partner proteins (AR-DBD: PDB 1R4I; AR-LBD: PDB 1T7T; CHIP: residue 311–369 extracted from AlphaFold ID P50502; RAP74 C-terminal domain: PDB 1I27; HSP70: 4PO2) were selected as ligands. For multimeric docking, the atomic coordinates of our NTD models were taken as both receptor and ligand.

For the inputs of ClusPro, we took the number of clusters that cover at least 50% of the poses. ClusPro predicted up to 120 different poses per protein–protein combination. All poses predicted by ClusPro were inspected using VMD and analysed using a custom-made Python script; utilizing different libraries: madplotlib[90], mdanalysis[91–93], mdtraj[94], numpy[95], pandas[96], scipy[97] and seaborn[98].

Author contributions

L.W.H.J. Heling: Investigation, Data Curation, Visualisation, Writing – Original Draft, Writing – Review & Editing. **V. Sheikhhassani:** Investigation, Methodology, Data Curation, Visualisation, Writing – Original Draft, Writing – Review & Editing. **J. Ng:** Investigation, Funding Acquisition, Writing – Review & Editing. **M. van Vliet:** Investigation. **A. Jiménez-Panizo:** Investigation, Writing – Review & Editing. **A. Alegre-Martí:** Investigation, Writing – Review & Editing. **J. Woodard:** Investigation, Writing – Review & Editing. **W. van Roon-Mom:** Writing – Review & Editing. **I. J McEwan:** Investigation, Resources, Writing – Review & Editing. **E. Estébanez-Perpiñá:** Investigation, Resources, Writing – Review & Editing. **A. Mashaghi:** Conceptualisation, Investigation, Project Administration, Supervision, Writing – Original Draft, Writing – Review & Editing.

Acknowledgements

The authors thank Dr. Annemieke Aartsma-Rus for discussions.

References

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

2. Davies, P. *et al*. Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. *J Mol Endocrinol* **41**, 301–314 (2008).

3. Parodi, S. & Pennuto, M. Neurotoxic effects of androgens in spinal and bulbar muscular atrophy. *Front Neuroendocrinol* **32**, 416–425 (2011).

4. Kennedy, W. R., Alter, M. & Sung, J. H. Progressive proximal spinal and bulbar muscular atrophy of late onset: A sex-linked recessive trait. *Neurology* **18**, 671–671 (1968).

5. Rhodes, L. E. *et al*. Clinical features of spinal and bulbar muscular atrophy. *Brain* **132**, (2009).

6. Mhatre, A. N. *et al*. Reduced transcriptional regulatory competence of the androgen receptor in X linked spinal and bulbar muscular atrophy. *Nat Genet* **5**, (1993).

7. Morfini, G. *et al*. JNK mediates pathogenic effects of polyglutamine-expanded androgen receptor on fast axonal transport. *Nat Neurosci* **9**, (2006).

8. Szebenyi, G. *et al*. Neuropathogenic forms of huntingtin and androgen receptor inhibit fast axonal transport. *Neuron* **40**, (2003).

9. Ranganathan, S. *et al*. Mitochondrial abnormalities in spinal and bulbar muscular atrophy. *Hum Mol Genet* **18**, (2009).

10. McCampbell, A. *et al*. CREB-binding protein sequestration by expanded polyglutamine. *Hum Mol Genet* **9**, (2000).

11. Lieberman, A. P., Harmison, G., Strand, A. D., Olson, J. M. & Fischbeck, K. H. Altered transcriptional regulation in cells expressing the expanded polyglutamine androgen receptor. *Hum Mol Genet* **11**, 1967–1976 (2002).

12. Mangelsdorf, D. J. *et al*. The nuclear receptor superfamily: The second decade. *Cell* **83**, (1995).

13. Sack, J. S. *et al*. Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proceedings of the National Academy of Sciences* **98**, 4904–4909 (2001).

14. Bohl, C. E., Gao, W., Miller, D. D., Bell, C. E. & Dalton, J. T. Structural basis for antagonism and resistance of bicalutamide in prostate cancer. *Proceedings of the National Academy of Sciences* **102**, 6201–6206 (2005).

15. Estébanez-Perpiñá, E. *et al*. The molecular mechanisms of coactivator utilization in ligand-dependent transactivation by the androgen receptor. *Journal of Biological Chemistry* **280**, 8060–8068 (2005).

16. 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).

17. Yu, X. *et al*. Structural Insights of Transcriptionally Active, Full-Length Androgen Receptor Coactivator Complexes. *Mol Cell* **79**, 812-823.e4 (2020).

18. Sheikhhassani, V. *et al*. Topological

dynamics of an intrinsically disordered N-terminal domain of the human androgen receptor. *Protein Science* **31**, (2022).

19. Gemayel, R. *et al*. Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity. *Mol Cell* **59**, (2015).

20. Silva, A., de Almeida, A. V. & Macedo-Ribeiro, S. Polyglutamine expansion diseases: More than simple repeats. *J Struct Biol* **201**, (2018).

21. Fiumara, F., Fioriti, L., Kandel, E. R. & Hendrickson, W. A. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* **143**, (2010).

22. Gemayel, R., Vinces, M. D., Legendre, M. & Verstrepen, K. J. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annual Review of Genetics* vol. 44 (2010).

23. Orr, H. T. & Zoghbi, H. Y. Trinucleotide repeat disorders. *Annual Review of Neuroscience* **30**, (2007).

24. Kumar, R. *et al*. Role of the androgen receptor CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy. *Life Sci* **88**, 565–571 (2011).

25. He, Y. *et al*. Androgen receptor with short polyglutamine tract preferably enhances Wnt/β-catenin-mediated prostatic tumorigenesis. *Oncogene* **39**, (2020).

26. Pan, B. *et al*. Genetic Association Between Androgen Receptor Gene CAG Repeat Length Polymorphism and Male Infertility: A Meta-Analysis. *Medicine* **95**, e2878 (2016).

27. Adachi, H. Widespread nuclear and cytoplasmic accumulation of mutant androgen receptor in SBMA patients. *Brain* **128**, 659–670 (2005).

28. Cortes, C. J. *et al*. Polyglutamine-expanded androgen receptor interferes with TFEB to elicit autophagy defects in SBMA. *Nat Neurosci* **17**, (2014).

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

30. 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).

31. 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).

32. He, B. *et al*. An androgen receptor NH2-terminal conserved motif interacts with the COOH terminus of the Hsp70-interacting protein (CHIP). *Journal of Biological Chemistry* **279**, (2004).

33. De Mol, E. *et al*. Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators. *Structure* **26**, 145-152.e3 (2018).

34. Wafa, L. A. *et al*. Isolation and identification of L-dopa decarboxylase as a protein that binds to and enhances transcriptional activity of the androgen receptor using the repressed transactivator yeast two-hybrid system. *Biochemical Journal* **375**, (2003).

35. Tavassoli, P. *et al*. TAF1 differentially enhances androgen receptor transcriptional activity via its N-terminal kinase and ubiquitin-activating and -conjugating domains. *Molecular Endocrinology* **24**, (2010).

36. Ray, M. R. *et al*. Cyclin G-associated kinase: A novel androgen receptor-

interacting transcriptional coactivator that is overexpressed in hormone refractory prostate cancer. *Int J Cancer* **118**, (2006).

37. Escobedo, A. *et al*. Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. *Nat Commun* **10**, 2034 (2019).

38. Bhattacharyya, A. *et al*. Oligoproline effects on polyglutamine conformation and aggregation. *J Mol Biol* **355**, (2006).

39. Vitalis, A., Wang, X. & Pappu, R. V. Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories. *Biophys J* **93**, (2007).

40. Saunders, H. M. & Bottomley, S. P. Multi-domain misfolding: Understanding the aggregation pathway of polyglutamine proteins. *Protein Engineering, Design and Selection* **22**, (2009).

41. Wetzel, R. Physical Chemistry of Polyglutamine: Intriguing Tales of a Monotonous Sequence. *J Mol Biol* **421**, 466–490 (2012).

42. Barbosa Pereira, P. J., Manso, J. A. & Macedo-Ribeiro, S. The structural plasticity of polyglutamine repeats. *Current Opinion in Structural Biology* vol. 80 (2023).

43. Chiti, F. & Dobson, C. M. Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annual Review of Biochemistry* vol. 86 (2017).

44. Shen, H. C. & Coetzee, G. A. The Androgen Receptor: Unlocking the Secrets of Its Unique Transactivation Domain. *Vitamins and Hormones* vol. 71 (2005).

45. Oppong, E. *et al*. An amyloidogenic sequence at the N-terminus of the androgen receptor impacts polyglutamine aggregation. *Biomolecules* **7**, (2017).

46. Kumar, R. Role of androgen receptor polyQ chain elongation in Kennedy's disease and use of natural osmolytes as potential therapeutic targets. *IUBMB Life* **64**, 879–884 (2012).

47. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci U S A* **115**, (2018).

48. Mashaghi, A., Van Wijk, R. J. & Tans, S. J. Circuit topology of proteins and nucleic acids. *Structure* **22**, 1227–1237 (2014).

49. Moes, D. *et al*. ProteinCT: An implementation of the protein circuit topology framework. *MethodsX* **9**, (2022).

50. Woodard, J., Iqbal, S. & Mashaghi, A. Circuit topology predicts pathogenicity of missense mutations. *Proteins: Structure, Function and Bioinformatics* **90**, (2022).

51. Scalvini, B. *et al*. Circuit topology approach for the comparative analysis of intrinsically disordered proteins. *J Chem Inf Model* **63**, 2586–2602 (2023).

52. 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**, 8584–8592 (2020).

53. 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. *Journal of Biological Chemistry* **278**, 14956–14960 (2003).

54. He, B., Bowen, N. T., Minges, J. T. & Wilson, E. M. Androgen-induced NH2- and COOH-terminal Interaction Inhibits p160 Coactivator Recruitment by Activation Function 2. *Journal of Biological Chemistry* **276**, (2001).

55. Belikov, S., Bott, L. C., Fischbeck, K. H. & Wrange, Ö. The polyglutamine-expanded androgen receptor has increased DNA binding and reduced transcriptional activity. *Biochem Biophys Rep* **3**, 134–139 (2015).

56. McEwan, I. J. & Gustafsson, J. Å. Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF. *Proc Natl Acad Sci U S A* **94**, (1997).

57. Lavery, D. N. & McEwan, I. J. Functional characterization of the native NH2-terminal transactivation domain of the human androgen receptor: Binding kinetics for interactions with TFIIF and SRC-1a. *Biochemistry* **47**, (2008).

58. Lei, L., Ren, D. & Burton, Z. F. The RAP74 Subunit of Human Transcription Factor IIF Has Similar Roles in Initiation and Elongation. *Mol Cell Biol* **19**, (1999).

59. Adegbuyiro, A., Sedighi, F., Pilkington, A. W., Groover, S. & Legleiter, J. Proteins Containing Expanded Polyglutamine Tracts and Neurodegenerative Disease. *Biochemistry* **56**, 1199–1217 (2017).

60. Katsuno, M. *et al*. Testosterone reduction prevents phenotypic expression in a transgenic mouse model of spinal and bulbar muscular atrophy. *Neuron* **35**, 843–854 (2002).

61. Takeyama, K. I. *et al*. Androgen-dependent neurodegeneration by polyglutamine-expanded human androgen receptor in Drosophila. *Neuron* **35**, 855–864 (2002).

62. Chevalier-Larsen, E. S. *et al*. Castration Restores Function and Neurofilament Alterations of Aged Symptomatic Males in a Transgenic Mouse Model of Spinal and Bulbar Muscular Atrophy. *Journal of Neuroscience* **24**, 4778–4786 (2004).

63. Yu, Z. *et al*. Androgen-dependent pathology demonstrates myopathic contribution to the Kennedy disease phenotype in a mouse knock-in model. *J Clin Invest* **116**, 2663–2672 (2006).

64. Lisberg, A., Liu, Y. & Merry, D. E. Blocking the dimerization of polyglutamine-expanded androgen receptor protects cells from DHT-induced toxicity by increasing AR turnover. *Journal of Biological Chemistry* **300**, (2024).

65. Heine, E. M. *et al*. Proteasome-mediated Proteolysis of the polyglutamine-expanded androgen receptor is a late event in spinal and bulbar muscular atrophy (SBMA) pathogenesis. *Journal of Biological Chemistry* **290**, (2015).

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

67. 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).

68. Mayer, M. P. & Bukau, B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences 2005 62:6* **62**, 670–684 (2005).

69. Young, Z. T. *et al*. Stabilizing the Hsp70-Tau Complex Promotes Turnover in Models of Tauopathy. *Cell Chem Biol* **23**, (2016).

70. Jochum, T. *et al*. Toxic and non-toxic aggregates from the SBMA and normal forms of androgen receptor have distinct oligomeric structures. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1822**, 1070–1078 (2012).

71. Li, M., Chevalier-Larsen, E. S., Merry, D. E. & Diamond, M. I. Soluble androgen receptor oligomers underlie pathology in a mouse model of spinobulbar muscular atrophy. *Journal of Biological*

Chemistry **282**, (2007).

72. Cho, H. The N17 domain of huntingtin as a multifaceted player in Huntington's disease. *Front Mol Biosci* **11**, (2025).

73. Ellisdon, A. M., Thomas, B. & Bottomley, S. P. The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *Journal of Biological Chemistry* **281**, (2006).

74. Lupton, C. J. *et al.* Enhanced molecular mobility of ordinarily structured regions drives polyglutamine disease. *Journal of Biological Chemistry* **290**, (2015).

75. Natalello, A. *et al.* A major role for side-chain polyglutamine hydrogen bonding in irreversible ataxin-3 aggregation. *PLoS One* **6**, (2011).

76. Chow, M. K. M., Paulson, H. L. & Bottomley, S. P. Destabilization of a Non-pathological Variant of Ataxin-3 Results in Fibrillogenesis via a Partially Folded Intermediate: A Model for Misfolding in Polyglutamine Disease. *J Mol Biol* **335**, 333–341 (2004).

77. Ellisdon, A. M., Thomas, B. & Bottomley, S. P. The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J Biol Chem* **281**, 16888–16896 (2006).

78. Ellisdon, A. M., Pearce, M. C. & Bottomley, S. P. Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein. *J Mol Biol* **368**, 595–605 (2007).

79. Merry, D. E., Kobayashi, Y., Bailey, C. K., Taye, A. A. & Fischbeck, K. H. Cleavage, aggregation and toxicity of the expanded androgen receptor in spinal and bulbar muscular atrophy. *Hum Mol Genet* **7**, (1998).

80. Sisodia, S. S. Nuclear inclusions in glutamine repeat disorders: Are they pernicious, coincidental, or beneficial? *Cell* **95**, 1–4 (1998).

81. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431**, 805–810 (2004).

82. He, B., Kemppainen, J. A. & Wilson, E. M. FXXLF and WXXLF Sequences Mediate the NH 2 -terminal Interaction with the Ligand Binding Domain of the Androgen Receptor. *Journal of Biological Chemistry* **275**, 22986–22994 (2000).

83. Dubbink, H. J. *et al.* Distinct Recognition Modes of FXXLF and LXXLL Motifs by the Androgen Receptor. *Molecular Endocrinology* **18**, 2132–2150 (2004).

84. Buzón, V., Carbó, L. R., Estruch, S. B., Fletterick, R. J. & Estébanez-Perpiñá, E. A conserved surface on the ligand binding domain of nuclear receptors for allosteric control. *Mol Cell Endocrinol* **348**, 394–402 (2012).

85. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7–8 (2015).

86. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

87. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph* **14**, 33–38 (1996).

88. Phillips, J. C. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics* **153**, 44130 (2020).

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

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

91. Alibay, I. *et al*. Building a community-driven ecosystem for fast, reproducible, and reusable molecular simulation analysis using mdanalysis. *Biophys J* **122**, (2023).

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

93. 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).

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

95. Harris, C. R. *et al*. Array programming with NumPy. *Nature* vol. 585 (2020).

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

97. Virtanen, P. *et al*. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, (2020).

98. Waskom, M. seaborn: statistical data visualization. *J Open Source Softw* **6**, (2021).

Supplementary Information

Wild-type pQ/LBD interaction analysis

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 **(Figure S10A)**. Of these, the first region, previously identified as the BF-3 region, interacts with the main cluster of pQ20 poses **(Figure 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 **(Figure 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 **(Figure 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.

Supplementary 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 cutoff of 1.2 nm. A single cutoff 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 glutamine residues. In this study, we used pepATTRACT, 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-minimised structure of the LBD in both analysis was used as the receptor protein.

Supplementary Figures

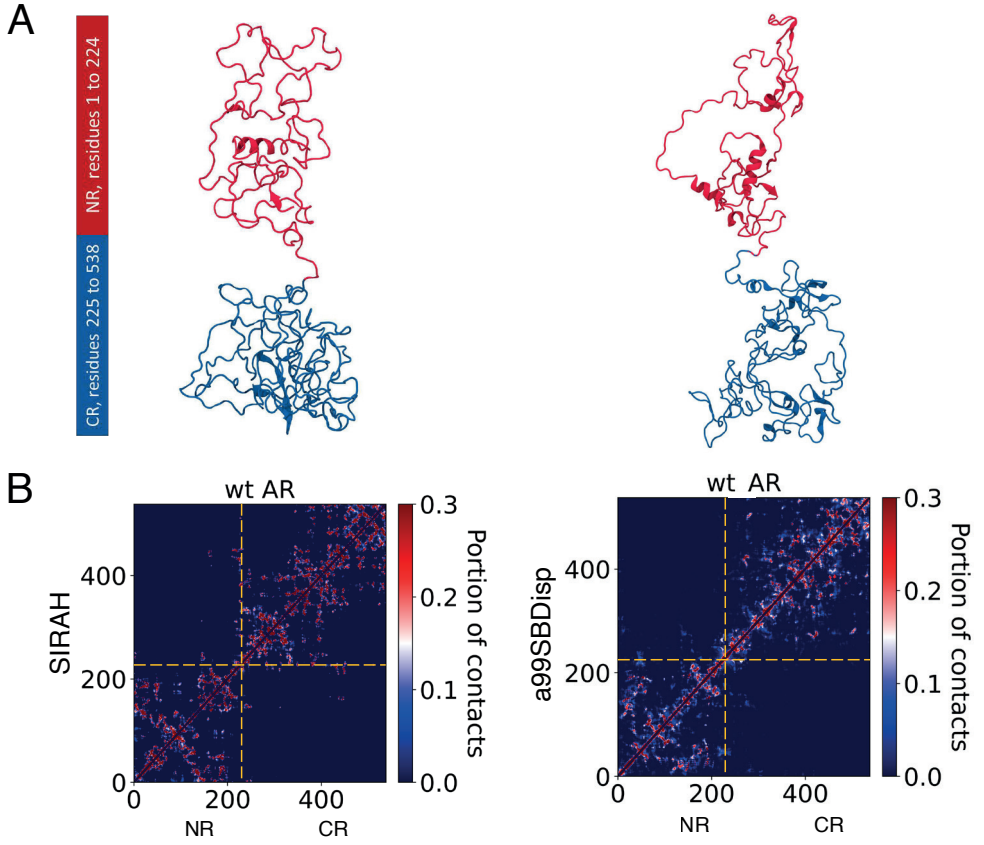


Figure S1. Regional segregation of wt-NTD. (A) *NR and CR subregion segregation in NTD is obtained with both coarse-grained SIRAH (left) and all-atom a99SB-disp (right) forcefields.* ***(B)*** *Contact maps of the three all atom simulationscombined highlights dynamic regional segregation of NTD into NR and CR as seen both in SIRAH and a99SB-disp generated models.*

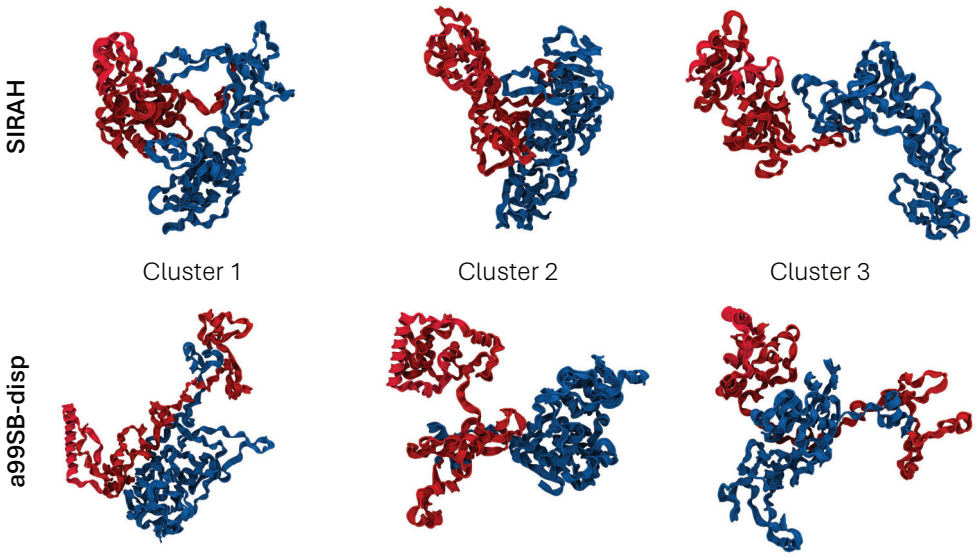


Figure S2. Three largest clusters from SIRAH and all-atom a99SB-disp simulations of pQe NTD. *NR and CR residues are depicted in red and blue respectively.*

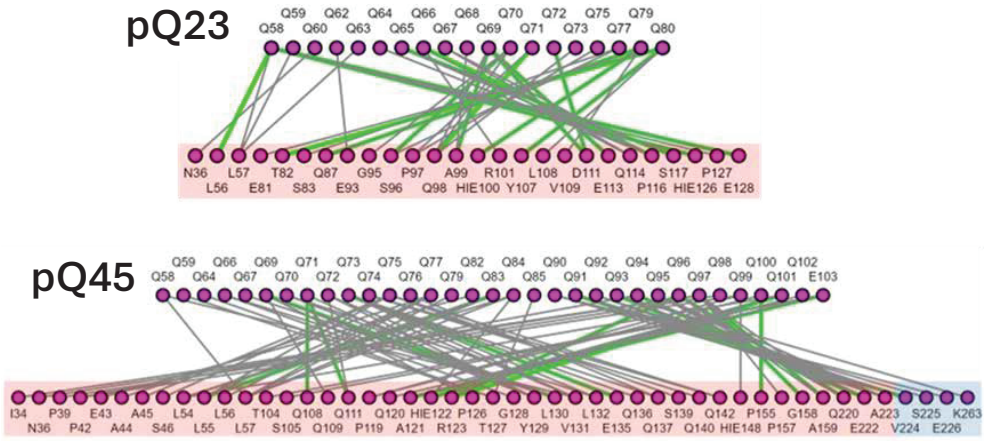


Figure S3. Interaction network of the pQ tracts in wt-NTD and pQe-NTD with other NTD residues. *NR residues and CR residues are depicted with red and blue boxes respectively. In this interaction analysis hydrogen bonds are shown in green and close proximity contacts are coloured in grey.*

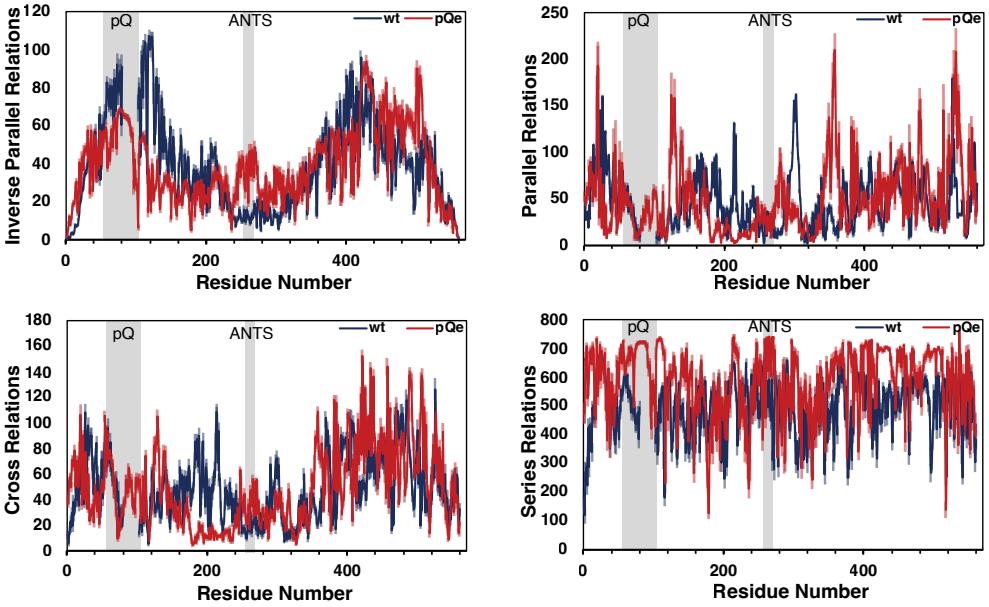


Figure S4. Residue-based local circuit topology for wt- and pQe-NTD. *Shown are mean number (± S.E.M. in shading) of inverse parallel (IP), parallel (P), cross (X), and series (S) relations as a functionof residue number. The pQ region and ANTS motif are highlighted by the grey boxes.*

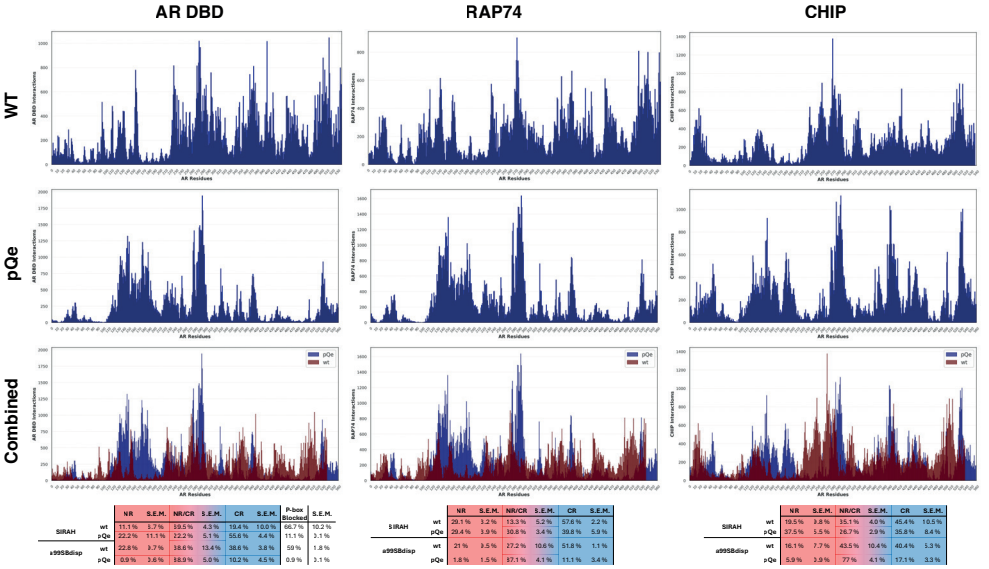


Figure S5. wt and pQe interaction histograms with DBD, RAP74 and CHIP based on ClusPro poses using our aa99SB-disp derived models. *Height of the bars indicate the total interaction count of thatresidue. The tables below show the percentages of exclusive interactions with NR and CR or both. Errors are shown as S.E.M.*

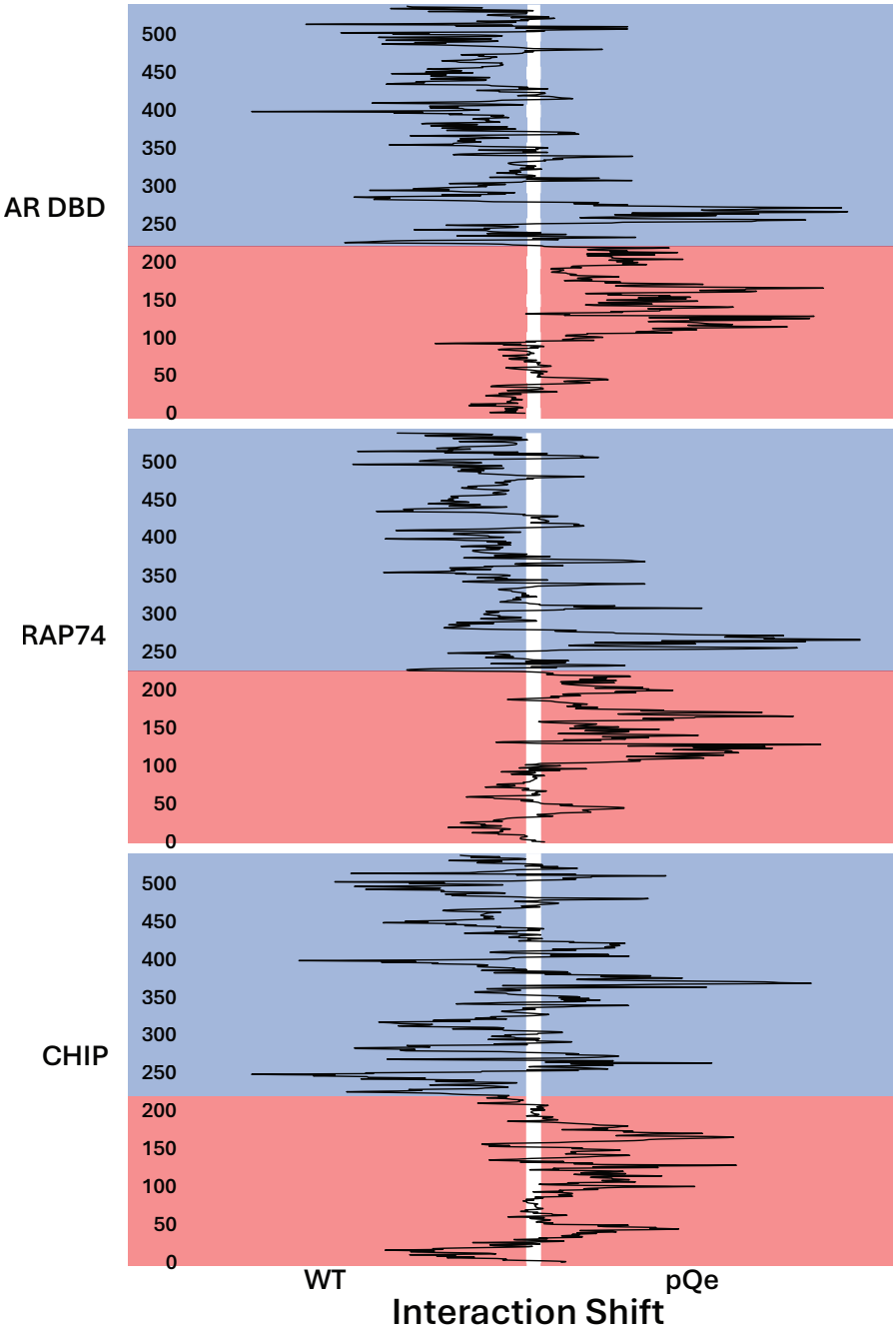


Figure S6. pQe shifts interactions with partner proteins. *Normalized Interaction Shift plot of aa99SB-disp derived models for wt and pQe AR-NTD interacting with DBD, RAP74 and CHIP based on ClusPro models. Deviation of the black line from the central axis indicates the shift of interactions of a residue towards wt or pQe. Colours indicate the NR (red) or CR (blue) of AR-NTD.*

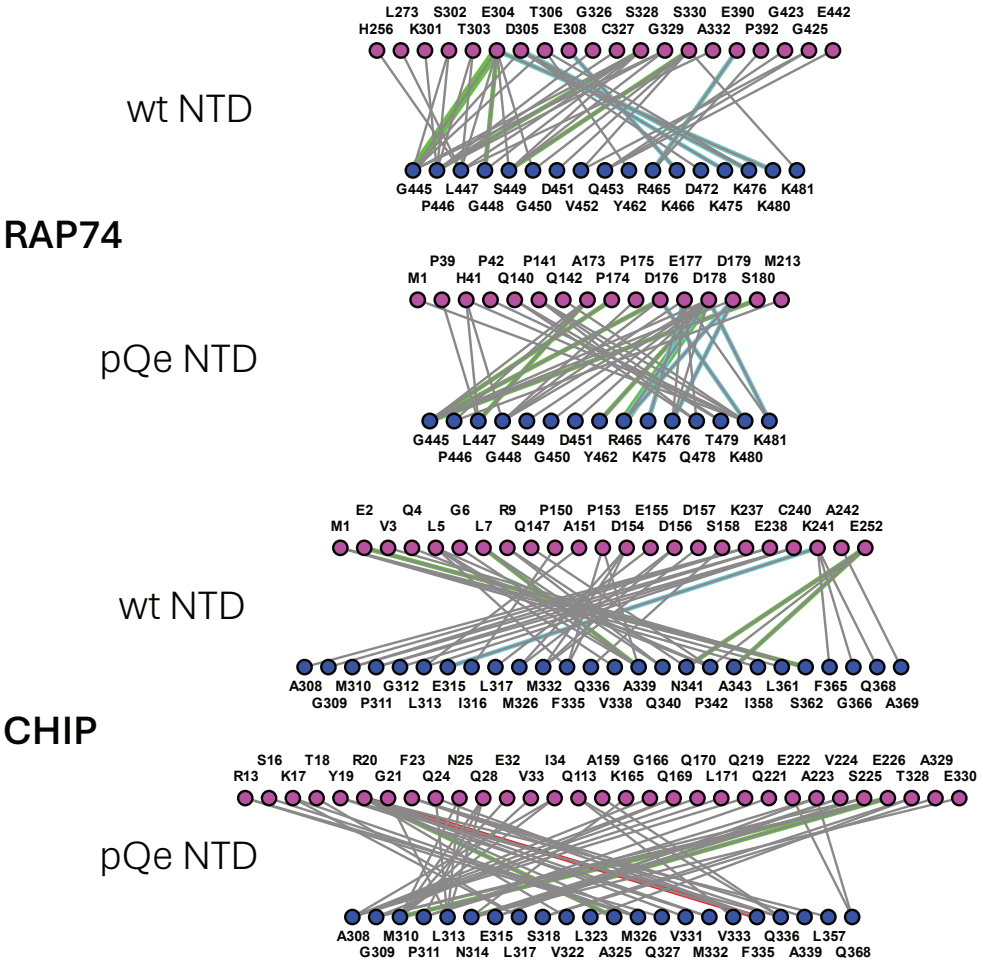


Figure S7. Interaction network of AR-NTD residues (top row, magenta spheres) with RAP74 and CHIP (bottom row, blue spheres). *Different types of interactions are coloured differently: hydrogen bonds (green), salt-bridge/Ionic (cyan), π-cation (red), and π-stacking (blue).*

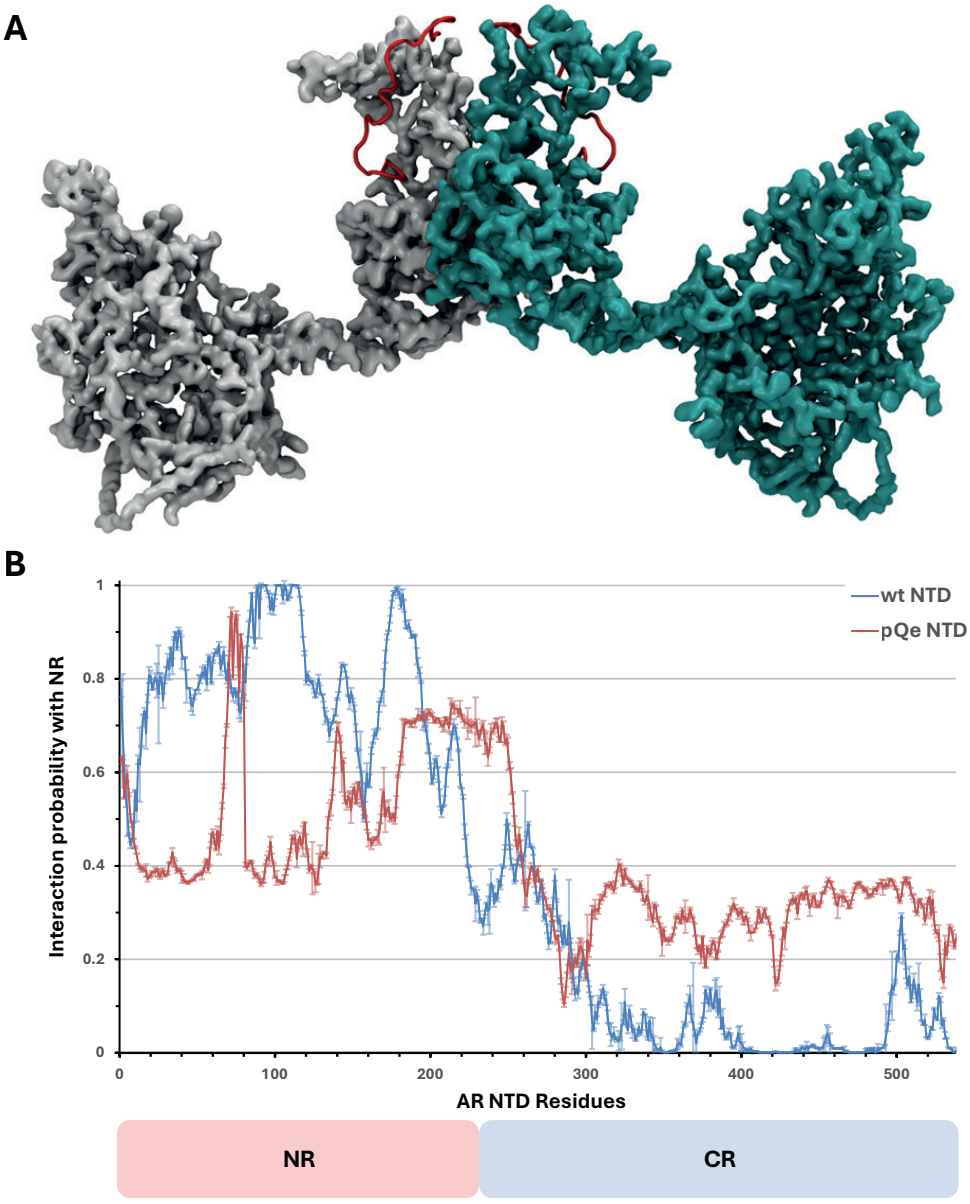


Figure S8. AR-NTD dimerisation is altered by pQ expansion. (A) *wt–wt dimer (based on CG models) demonstrating a head–head conformation.* ***(B)*** *Residue based probability of interacting with NR residues in the other AR molecule in the dimer (based on all-atom derived models). In wt, the NR shows 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.*

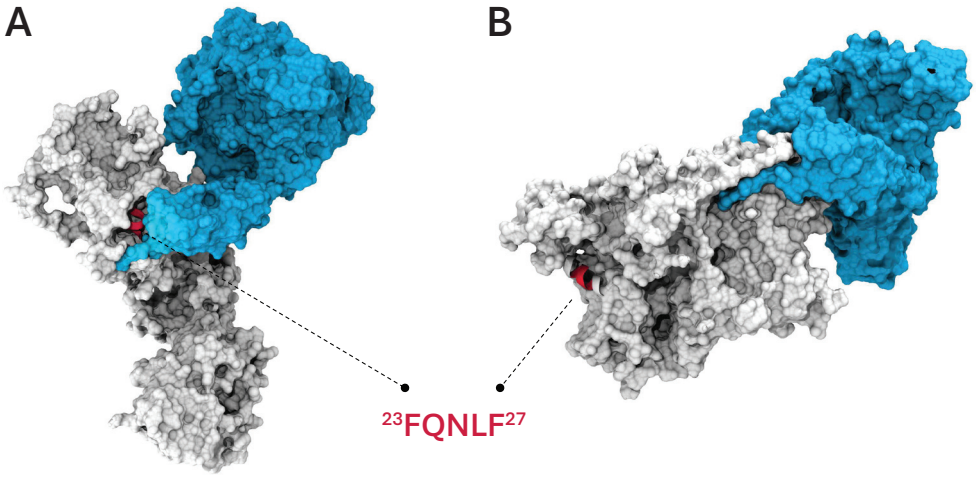


Figure S9. HSP70 interactions with NTD are altered upon pQ expansion. *Representative structures of* ***(A)*** *wt-NTD and* ***(B)*** *pQe-NTD in white docked using ClusPro with HSP70 (Blue; structure from Alphafold: AF-P11142). The interaction of HSP70 with the FQNLF motif of AR-NTD (depicted in red) 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.*

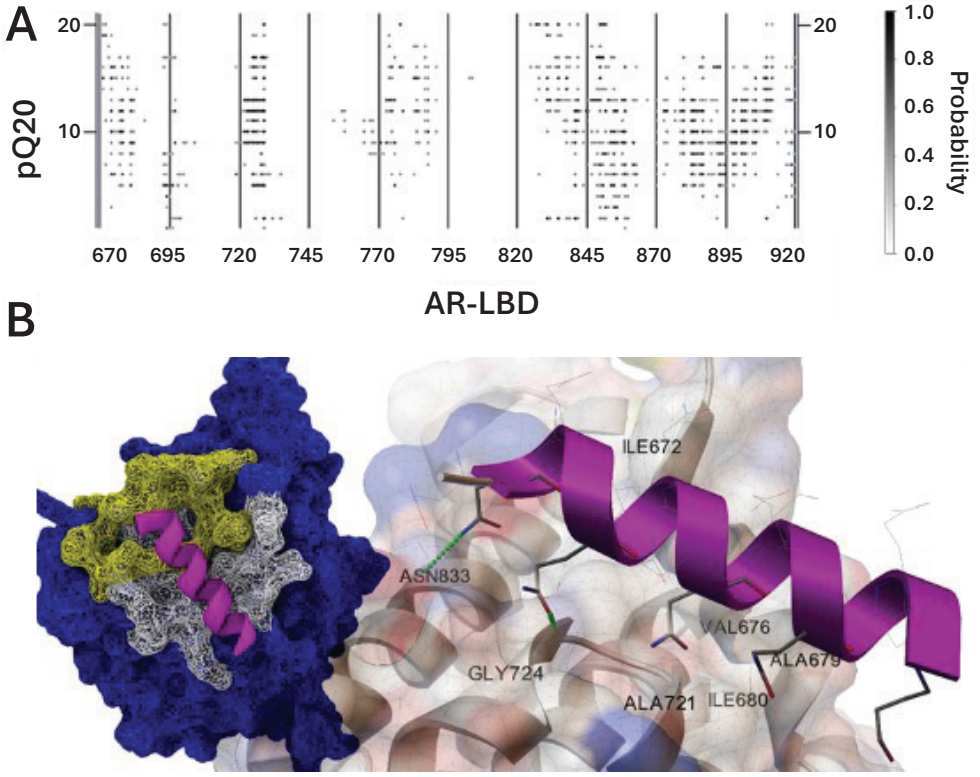


Figure S10. The pQ tract binds to the LBD through the BF-3 pocket. (A) *Consensus map of the top 50 ranked poses of the pQ20 AR-LBD complex. The grayscale density indicates the probability of contact formation between interacting residues.* ***(B)*** *Interacting region of pQ20 AR-LBD (white) shows a 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.*

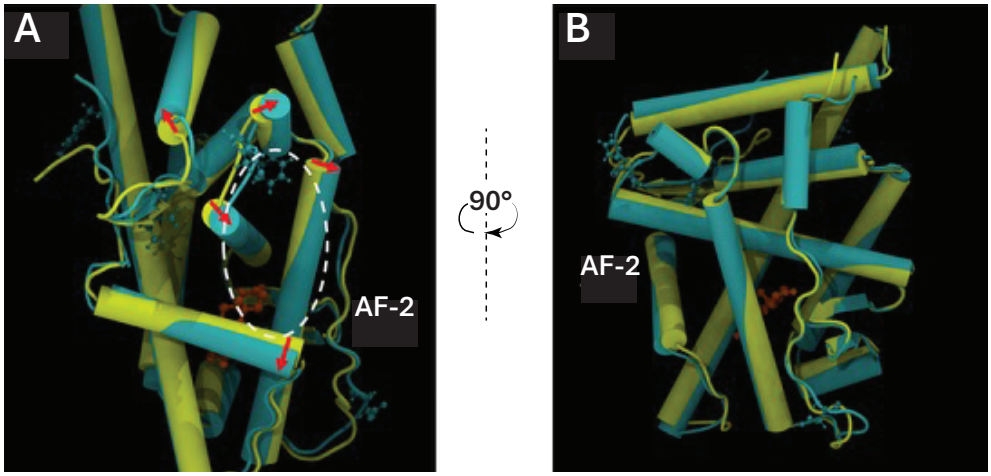


Figure S11. pQ interactions with BF-3 induce conformational change in LBD, opening the AF-2 pocket. ***(A)*** *Superimposition of representatives of pQ20–LBD and AR-LBD in isolation from AF-2 view and* ***(B)*** *from side view by 90 clockwise rotation. AR-LBD is drawn in yellow, and LBD structure isolated from pQ20–LBD complex in cyan. The red arrows in panel* ***A*** *mark helices number 4, 5, and 9 which rotated 3°, and helix 12 which rotated 6°, with respect to their original positions.*

Table S1. Structure conditions and subsequent simulation details.

	Force Field	Time	Starting Condition	Runs
wt NTD	SIRAH	5 μs	I-TASSER	3
wt NTD	a99SB-disp	1 μs	Stretched Chain	3
pQe NTD	SIRAH	5 μs	I-TASSER	7
pQe NTD	a99SB-disp	1 μs	Stretched Chain	3

2