Biophysics of disordered nuclear receptors and their DNA binding regulation

Heling, L.W.H.J.

Chapter 1

Beyond the Structure: the role of Intrinsic Disorder in Nuclear Receptors

Nuclear receptors comprise a superfamily of transcription factors (TFs) that are widely expressed throughout the body. They control a vast array of physiological processes, including metabolism, reproduction, inflammation, and cell proliferation[1,2]. Since the cloning of the first nuclear receptor 40 years ago[3], 48 members have been identified in humans[4,5]. These proteins function as master regulators of gene expression, translating diverse stimuli into specific cellular responses. Consequently, dysregulation of nuclear receptor signalling pathways is implicated in a range of pathologies, from cancer and diabetes to chronic inflammatory conditions[6–10]. Because of their central role in health and disease, nuclear receptors represent a highly valuable class of therapeutic targets.

Nearly all nuclear receptors share a common modular architecture, typically composed of five distinct domains (A–E)[11,12] **(Figure 1A)**. The central DNA-binding domain (DBD) contains 8 conserved cysteine residues that coordinate 2 zinc ions into finger motifs, necessary for the receptor's specific binding to DNA motifs called response elements (REs) configured as single, inverted or direct repeats[7,11,12]. The first zinc finger contains the "P-box," a motif that makes direct contact with the DNA major groove, thereby governing sequence recognition. The second finger contains the "D-box" motif, which is involved in receptor dimerisation on DNA. Sequence specificity has been linked to DBD dimerisation[13,14]. ChIP-seq analyses have allowed for genome-wide identification of nuclear receptor binding sites, the cooperation between nuclear receptors and different TFs, and the genome wide effects of nuclear receptor activation on acetylation and methylation of histones markers[15].

The ligand binding domain (LBD) of nuclear receptors is a complex allosteric signalling domain, typically composed of 12 α-helices and 4 ß-sheets that fold into a three-layered sandwich, creating a hydrophobic ligand binding pocket (LBP) at the base of the domain which enables interactions with cholesterol, steroid hormones and with cellular metabolic products[7,8]. Steroid ligand binding to the LBD is suggested to induce a conformational change, repositioning the C-terminal helix 12 to create a surface known as activation function 2 (AF-2)[16]. This surface is essential for protein–protein interactions (PPIs) with transcriptional co-regulators, which enhances nuclear receptor activity[17,18]. Steroid ligand-induced folding of the LBD creates a distinct surface pocket known as binding function 3 (BF-3)[19]. This site is of particular interest as it is conserved among steroid hormone receptors, and mutations in residues lining this pocket are implicated in disease. BF-3 may serve as an alternative surface for PPIs, and several small molecule modulators have been reported to bind to this region[19].

A short, flexible hinge region connects the DBD with the LBD[7]. Long considered merely a linker region, the hinge contains the nuclear localisation signal motifs shown to be essential regulators of nuclear translocation[20]. Although poorly conserved across nuclear receptors, this region is rich in post-translational modifications (PTM) sites that influence receptor activity[21–23]. Several mutations for diseases like steroid insensitive syndrome and

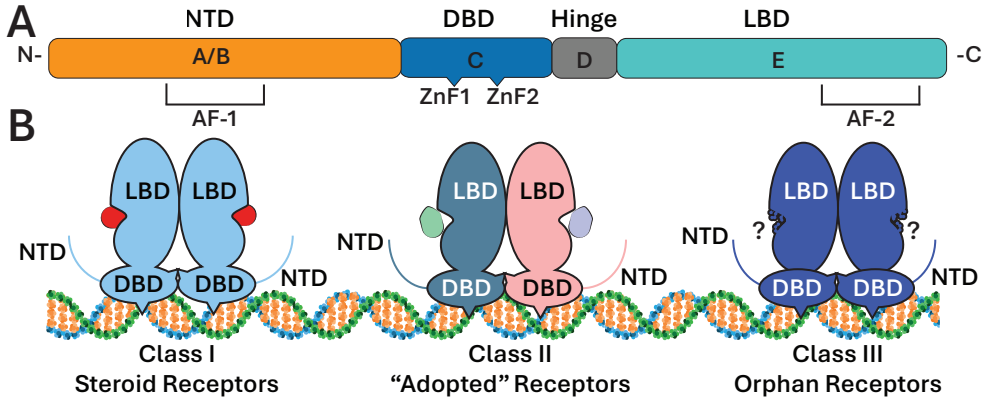

Figure 1. Canonical nuclear receptor domains and classification. (A) *Cartoon of the linear protein structure of nuclear receptors, divided in 5 sections: A/B encodes the N-terminal domain (NTD) including the AF-1 region, C the DNA-binding domain (DBD) with its two zinc finger (Znf) domains, D the hinge region, and E the ligand-binding domain (LBD) with the AF-2 region.* ***(B)*** *Nuclear receptors can be divided in three classes based on their ligand and DNA binding characteristics. For class I receptors, the ligand binds with high affinity, while for class II this affinity is attenuated. Question marks next to the class III orphan receptor ligand binding pocket denote that physiological ligands are not known. The ligand binding pocket is lined with bulky amino acids, making ligand binding probably more difficult.*

cancer are mapped to this region, further illustrating its functional role[23].

NRs are often classified into three or four major groups based on their ligand-binding, and DNA bound dimerisation characteristics[8] **(Figure 1B)**. Class I includes classical steroid hormone receptors like the androgen receptor (AR), estrogen receptor (ER) and glucocorticoid receptor (GR). Class II consists of "adopted" receptors originally thought to be orphans (i.e. no endogenous ligands had been identified) for which naturally occurring ligands have been found since, including the retinoid X receptor (RXR) and peroxisome proliferator-activated receptors (PPARs). The ligands for these receptors often have a lower reported affinity for their LBDs than class I nuclear receptor ligands[7]. Class III is made up of orphan nuclear receptors, for which no definitive endogenous ligands have yet been identified[24]. For many of these orphans, such as the NR4A subfamily members, the LBD is packed with bulky amino acid side chains, which is thought to obstruct conventional ligand binding[24]. As a result, it is suggested that the regulation of Class III (and IV) nuclear receptors depend primarily on the control of their expression, PTMs, and PPIs, rather than direct ligand activation[24].

The structures of the DBD and LBD domains of many nuclear receptors have been experimentally determined, albeit typically in complex with ligands or DNA. Some nuclear receptors are shown to be in a complex with molecular chaperones. These proteins (reviewed here[25]) are key to cellular protein folding, control the conformational landscape and protect against aggregation. For steroid nuclear receptors, chaperones are essential

throughout the cellular lifetime[26], and consequently, the structure of the hormone-lacking (apo) state of the protein remains enigmatic. Furthermore, the presence of a variable, intrinsically disordered N-terminal transactivation domain (NTD) has hindered complete structural elucidation of nuclear receptors.

Intrinsically disordered protein regions - central to the functional characterisation of nuclear receptors

In 1894, Emil Fischer's lock-and-key model established the paradigm that protein function, specifically enzymatic activity, is governed by 3D structure. This structural conformation dictates the protein's ability to selectively bind a complementary substrate, thereby facilitating the catalytic process[27]. Over the course of the 20th century, this lock-and-key model was further refined through experimental observations[28–37], incorporating configurational adaptability or induced fit theory. This model posits that substrate binding induces structural alignment within the active site to facilitate function[38–40]. These two frameworks strongly supported the structure–function paradigm of protein, which became the leading focus in the field of structural biology[30,37]. From the beginning, structural biology techniques (X-ray crystallography and later nuclear magnetic resonance (NMR) and (cryo-) electron microscopy) have captured regions of proteins with low electron density. These were considered to represent the absence of structure and, likely due to a lack of understanding, it remained largely ignored within the field of structural biology[30].

Advances in computational biology in the 1990s and early 2000s led to the development of protein structure prediction methods[41–43]. This revealed that a large number of proteins in the eukaryotic proteome contained long natively unstructured/disordered[4,43–47] stretches, over 30 amino acids long[48]. Because these disordered sections often lie within highly functional proteins, focus on these regions has accelerated and led to new discoveries. Intrinsic disorder is encoded in the protein sequence, with an enrichment of amino acids that lead to low mean hydrophobicity and high net charge[49]. This makes it possible to accurately predict intrinsic disorder from the amino acid sequence of a protein[44,50].

The conceptual shift from stable protein structures to dynamic conformational ensembles is fundamental for understanding the roles of these intrinsically disordered proteins (IDPs) and regions (IDRs). Despite lacking a fixed three-dimensional structure, IDRs facilitate key cellular processes, including signalling and regulatory pathways[51]. Their structural flexibility provides distinct functional advantages[44]. They are required for mediating receptor-dependent transcription, facilitating intradomain interactions, and enabling allosteric regulation[51–55]. The inherent disorder allows IDRs to engage in transient interactions with a wide array of binding partners, which can vary depending on cell type and cell-specific signalling cues[56–58]. This adaptability is often achieved through a mechanism known as “coupled folding and binding,” where the disordered region undergoes a local

conformational transition, forming an α-helix for instance, upon interacting with its target[53,55]. This process permits the burial of large intermolecular surface areas, even when the interacting domains are small, thus achieving high-affinity binding. The resulting protein complexes are sometimes described as "fuzzy," as the IDR may retain significant disorder even in the bound state[59]. This allows IDPs to adopt different structures, within the same motif, while binding to varying partners[59–61]. Another characteristic feature in IDRs is the high density of PTM sites, which can alter the physicochemical and structural landscape, further increasing conformational and functional adaptability[62]. IDRs are frequently subjected to phosphorylation, acetylation, methylation, ubiquitination and SUMOylation, which affect the compactness[63], cause a switch from a disordered to an ordered state[64] or alter the phase separation qualities *in vitro*[65].

The combination of conformational plasticity, binding promiscuity and temporal regulation through PTMs explains the abundance of IDRs in TFs. Once described as "acid blobs and negative noodles"[66], the IDR of TFs are core elements regulating gene expression. Transcription relies on the spatiotemporal composition of a multitude of PPIs[67–69], which is uniquely suited for IDRs.

The NTD of nuclear receptors are classified as IDRs. In contrast to the generally stable, globular structures of the DBD and LBD, the NTD exhibits a high degree of structural plasticity. It is best described as a collection of conformations, comprising varying levels of unfolded, secondary and tertiary structures[4,7]. The NTD is the least evolutionarily conserved domain of nuclear receptors across species and superfamily members, and plays an essential role in transcriptional activation and regulation, through its structural plasticity and propensity for PTMs. Given the central role of nuclear receptors in key biological pathways, the NTD is implicated in many pathologies[10,21,70–72]. However, its elusive structural nature has complicated detailed conformational and functional characterisation[7,8,10,71–74]. The androgen receptor and Nur77 are nuclear receptors implicated in severe pathologies, including castration-resistant prostate cancer and chronic inflammatory disease, that are driven by their large, disordered NTDs. Understanding the dynamic mechanisms of these domains is therefore a critical frontier for novel drug development.

Androgen receptor - allosteric hub defined by disorder

The Androgen Receptor (AR/NR3C4) is a 110 kDa, classic steroid hormone receptor, belonging to Class I of the nuclear receptor superfamily[7,75] **(Figure 2)**. Testosterone and dihydrotestosterone (DHT) hormones primarily regulate AR action which is important for male sexual development and reproductive health[76]. Its dysregulation is linked to various disorders, including neuromuscular disease and cancer[75].

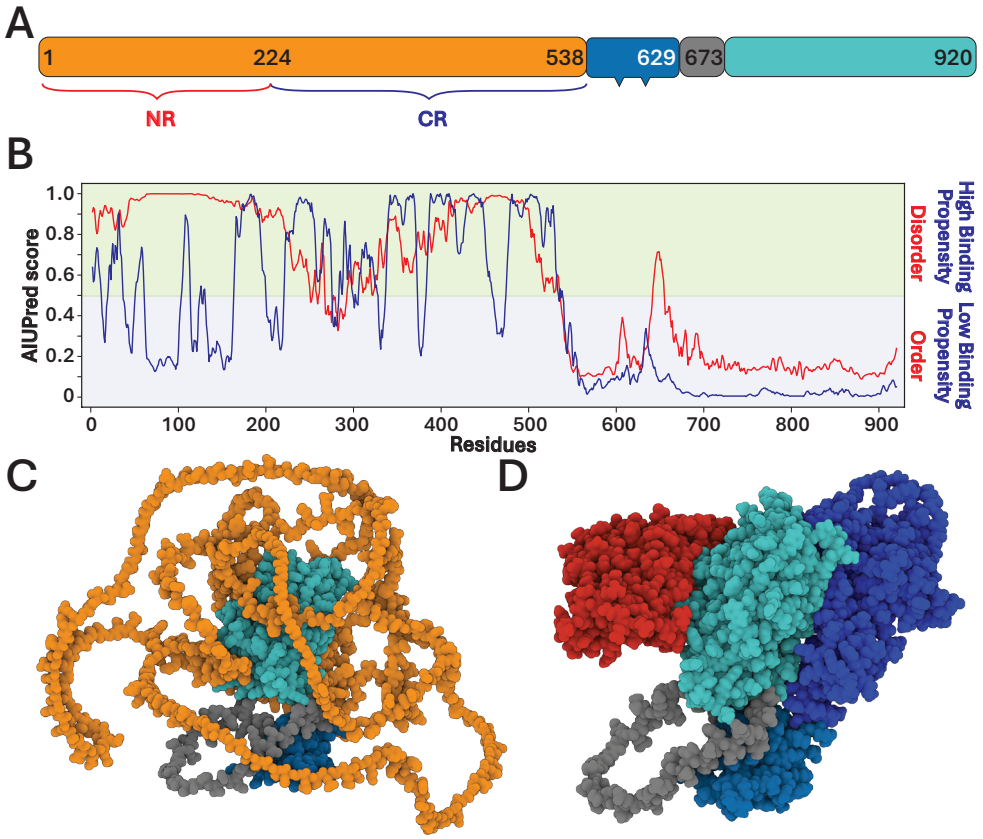

Figure 2. Structural and functional classification of disorder in the Androgen Receptor. (A) *Cartoon of the linear protein structure of AR with the residue numbers depicting the borders of each region. The colours align with the linear structure in* ***Figure 1A. (B)*** *Plot from AIUPred[99,100] (https://aiupred.elte.hu/), which predicts the likelihood of residues being more structurally ordered or disordered (in red), and which residues and regions have a propensity to be functional binding motifs (in blue, scores >0.5).* ***(C)*** *Alphafold predicted model of the full-length AR protein (AF-P10275-F1-v4), colour coded according to the linear structure in panel A. The structural dynamics of the NTD cannot be captured by the algorithm and present as a characteristic ribbon around the more stable structures of the protein.* ***(D)*** *Full length AR model from Sheikkhassani et al. (2022)[98], which identified, using molecular dynamics simulations and topological analyses, that the NTD can form two highly dynamic regions, the NR (residues 1–224) and CR (residues 225–538), coloured in red and darkblue respectively. The NTD and full-length model showed remarkable agreement with experimental evidence.*

The mechanism of AR activation is a multi-step, allosteric process[77]. Upon ligand binding, the AR-LBD is thought to undergo conformational changes, which lead to the dissociation from chaperone and co-chaperone complexes[78], repositioning helix 12 to form the AF-2 pocket[79,80]. This surface is an interaction site for co-activator proteins. A unique feature of AR is the preference of its AF-2 pocket to the ^{23}FQNLF^{27} motif located within AR-

NTD[81]. This N/C (N-terminal to C-terminal) interdomain interaction, which is induced by androgen binding, is required for full AR activity[82] and involves the FQNLF motif folding into an α-helix upon binding the LBD. Once activated, AR translocates to the nucleus and binds to specific DNA sequences known as androgen response elements (AREs) as a homodimer. Recent structural work has shown that this binding is asymmetric and cooperative, with the DNA itself acting as an allosteric effector to stabilise the dimer on the response element[83].

The AR-NTD, which comprises more than half of the protein sequence[20,77] **(Figure 2A)**, contains the AF-1 and is essential for transactivation[74,82]. The AF-1 has further been dissected into two transactivation units (TAU-1 and TAU-5)[84,85]. In addition, several linear motifs have been identified along the NTD sequence, the aforementioned FQNLF, the ^{179}LKDIL^{183}, and the ^{433}WHTLF^{437} [75]. The NTD also contains binding sites for the TFIIF transcription factor, the p160 coactivators and co-chaperone proteins[86–90], such as the fully conserved AR NTD signature sequence (ANTS)[91] between residues 233 and 246. Several amino acid repeats, including the polymorphic polyglutamine (pQ; residues 58–80) and polyglycine (pG; residues 372–379) stretches, are also encoded within the NTD. The physiological roles of these polymorphic regions are unclear, but they have been proposed to provide inhibitory regulation on AR[20]. These stretches are involved in disease. For instance, expansion of the pQ repeat beyond 38 residues is the mutation leading to spinal and bulbar muscular atrophy (SBMA; also known as Kennedy's disease), a neurodegenerative disease[92], while the length of the pQ and pG stretch is inversely correlated with prostate cancer risk[93,94]. The NTD is furthermore indispensable for the activity of LBD-truncated AR splice variants (AR-Vs), which are key drivers of castration-resistant prostate cancer (CRPC)[95].While the structural information of the DBD and LBD of AR have been experimentally characterised, structural information on the NTD has remained elusive, as a result of the high propensity for disorder **(Figure 2B)**. State-of-the-art approaches based on artificial intelligence have also not been able to overcome this **(Figure 2C)**. Given its key role in AR function and its structural plasticity, the AR-NTD has become a major focus for drug screening programs for CRPC[77]. Small molecule inhibitors, such as EPI-001, target the AR-NTD, blocking its transactivation function and indicating a potential mechanism for overcoming therapy resistance[99].

Nur77 - Regulation through Interaction

The NR4A subfamily, which comprises Nur77 (NR4A1), Nurr1 (NR4A2) and NOR-1 (NR4A3) belongs to Class III of orphan nuclear receptors[24]. While the DBDs have high sequence homology, the LBD and NTD are more divergent between the subfamily members, with 58–68% and <30% similarity respectively. Nur77 **(Figure 3)**, also known as NGFI-B and TR3, is predominantly localised in the nucleus where it functions as a transcription factor, modulating gene expression through binding to specific DNA motifs.

Nur77 has an important role in immune regulation, influencing pro-inflammatory signalling pathways and cellular metabolism[100]. It generally acts as an anti-inflammatory factor, with a protective role in chronic inflammatory diseases like atherosclerosis and inflammatory bowel disease. Its anti-inflammatory effects include negative crosstalk with the NF-κB signalling pathway, where it can directly bind and trans-repress the p65 subunit[101]. Recent work has expanded Nur77's regulatory influence to immunometabolism and has identified it as a novel regulator of macrophage immunomechanics[102]. This study showed that Nur77-deficient macrophages exhibit altered actin networks, leading to a stiffer cellular phenotype, enhanced migration, and increased phagocytic activity. Transcriptomic analysis revealed that Nur77 directly regulates the expression of genes related to the cytoskeleton and cellular mechanics. This direct regulation of mechanobiology genes by Nur77 shows an additional mechanism through which it restrains inflammation by modulating the mechanical characteristics of macrophages[102].

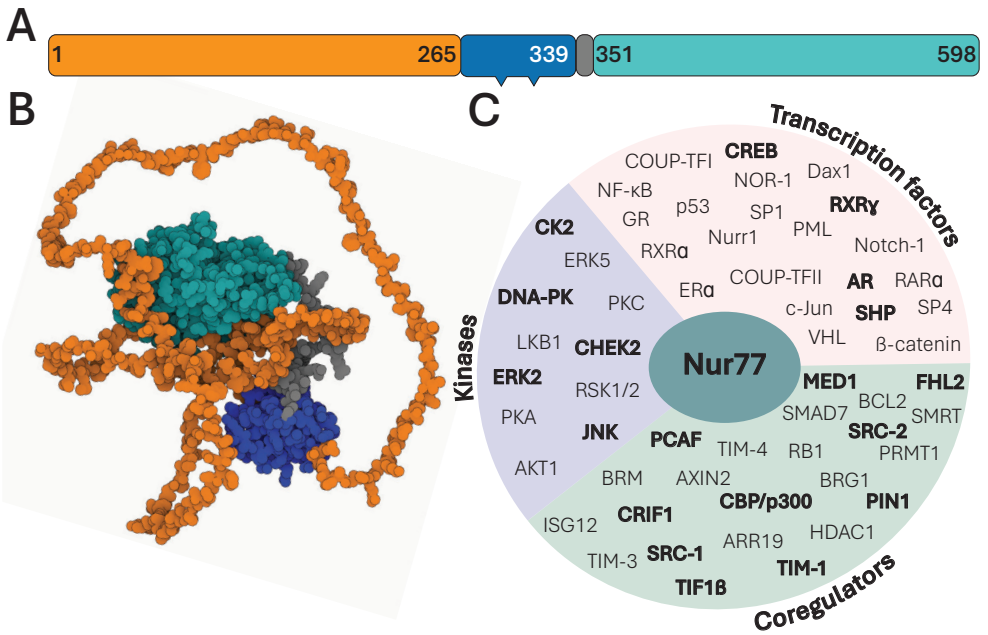

Figure 3. Nur77 activity is likely regulated through protein-protein interactions. ***(A)*** *Cartoon of the linear protein structure of Nur77, with the residue numbers depicting the borders of each region. The hinge region is encoded by residue 340 to 350. The colours align with the linear structure in* ***figure 1A****.* ***(B)*** *Alphafold predicted model of the full-length Nur77 protein (AF-P22736-F1-v4), colour coded according to the linear structure in panel* ***A****.* ***(C)*** *Transcription factors, coregulators and kinases identified to interact with Nur77. Proteins with bold names have been identified to interact through the NTD of Nur77. List of interactome taken from Kurakula et al. (2014)[24].*

A defining characteristic of Nur77 is the absence of an identified endogenous activating ligand[24,100]. This is likely because its LBD is packed with bulky hydrophobic amino acid sidechains, which obstruct conventional ligand binding. However, recent studies did

suggest potential ligands that may bind and regulate Nur77[103–108], although the molecular mechanisms behind the regulation are unclear, and the potential has not been consistently verified. The LBD of Nur77 is dispensable, while like AR, the NTD plays a major role in transactivation[109,110]. Another unique characteristic of Nur77 is the ability to bind specific DNA motifs as a monomer or as a homodimer[111], or form activating heterodimers with retinoid X receptors (RXRs).

As the endogenous ligand for Nur77 remains unknown, the current consensus is that activity is likely regulated through expression levels, PTMs, and PPIs[24] **(Figure 3C)**. Expression of Nur77 is transiently and rapidly induced by diverse cell-specific signals, like growth factors, cytokines, kinases and mechanical stress[112]. Various other transcription factors and co-activators have been found in complex with Nur77, including NF-κB, SP1, p300, Bcl-2, p53 and GR[24,113–116]. These interactions can enhance or repress the transcriptional activity of Nur77, while vice versa, Nur77 has been found to induce or inhibit the activity of the interacting receptors[24,117–119].

Cell-type specific, it has been shown that Nur77's roles also include non-transcriptional functions. Nur77 interaction with Bcl-2 at the mitochondria induces apoptosis[114,115]. These processes can be modulated by PTMs and PPIs. As most of the PPI and PTM sites lie within the NTD, this further establishes the critical importance of this domain.

Drugging the dynamics - tackling challenges for the therapeutic frontier

The characterisation of AR and Nur77 shows a need to move beyond static structural descriptions and to study their transient folding and dynamic molecular interactions. The evidence concerning the intrinsic disorder within their NTDs shows that conformational flexibility and dynamic transitions are active determinants of function and regulation. The complex interplay between the AR-NTD's intrinsic disorder and its implications in diseases exemplifies the effect of dynamic structural properties on biological function. For Nur77, the absence of a conventional ligand-binding pocket means that its activity is modulated by its transient PPIs and PTMs through the NTD. This inherent disorder in IDRs presents both challenges and opportunities for developing new therapeutic interventions[70,120–122].

Current therapeutic strategies targeting nuclear receptors often have limitations, including off-target effects and the emergence of resistance, like AR-Vs in prostate cancer[7,84]. The NTDs of nuclear receptors, particularly, are emerging as promising drug targets despite their disordered nature. The challenge lies in the rational design of small molecules that can precisely modulate the transient conformational states and dynamic interactions of these IDRs. Traditional structure-based drug design is often not applicable to IDRs Unlike the lock-and-key or induced-fit models applicable to globular domains, the

elusive structural nature of IDRs makes them a very challenging target for homology and functional classification. State-of-the-art AI approaches, like AlphaFold[123], have emerged as a prominent tool for predicting the 3D structures of proteins and protein complexes. However these fail to accurately capture the conformational dynamics of intrinsically disordered sequences. This has hindered drug development, and the general "druggability" of IDRs remains unclear[124]. Despite the limited knowledge on the conformational ensemble of IDRs, several small molecules have been found to bind and interact with disordered sequences[125–131], and some were shown to affect biological activity[128,129,131]. One of the most promising pharmacological compounds is the aforementioned EPI family, such as EPI-001[99], which targets the AR-NTD and has shown efficacy in blocking AR transactivation and mitigating resistance in prostate cancer[132]. This highlights the importance for a better understanding of these dynamical regions of nuclear receptors and important features within.

Overcoming this challenge requires new developments. In the last two decades, several tools have emerged. Computational approaches have been developed to readily identify regions and residues within IDPs as functional binding sites[97,98]. Recently, a machine learning tool allowed for the identification of protein binders targeting disordered regions, and functionally modulating IDPs *in vitro*[133]. Molecular dynamics (MD) simulations have emerged as an indispensable technology that can characterise the very dynamics that render the NTD elusive to classical approaches. By simulating the atomic motions of the protein over time, MD, coupled with complementary experimental assays such as NMR spectroscopy, can map the dynamic conformational landscape of IDPs, revealing the distribution of states it populates[134–136]. NMR spectroscopy is continuously developed to characterise disorder and the functional atomistic details within[137] and has been used to localise sequences within IDPs which are stabilised by compounds through chemical shift perturbations[99].

Combining MD simulations with topological analysis allows for the identification of motifs within the conformational noise. Topology is a mathematical framework that is capable to detect invariant shapes within geometric dynamics. The field of protein topology was traditionally focussed on knotted conformations and ignored intrachain amino acid contacts[138,139]. The emergence of circuit topology **(**CT, **Figure 4)** provides a complementary topological approach to knot theory which can fully characterise folded molecular chains[140–144]. CT in its simplest form characterises a pair of intramolecular contacts as one of three configuration: series (S), parallel (P) and cross (X). A pair of contacts in the S configuration are independent from each other and can be distant, while P and X contacts indicate more entanglement of protein chain. The CT framework, which had already provided insights into molecular folding processes and the 3D genome architecture[145–155], was enhanced with a lifetime distribution parameter to create dynamic CT (dCT)—ideally suited to capture the transient nature of IDPs. MD simulations and dCT were first combined in our study on the full length wild-type (wt) AR-NTD[96]. This revealed with

atomistic detail that AR-NTD form two highly dynamic disjoint regions, a N-terminal region (NR; residues 1–224) and C-terminal region (CR; 225-538) **(Figure 2D)**. Beyond the structural observations from the MD simulation results, the topological analysis revealed stark differences between the two subregions. CR contacts showed more entanglement (as evidenced by higher P and X relations) and had longer lifetimes, indicating higher stability and compaction compared to the NR, which explored the topological space in what appeared a more random fashion[96]. This study presented a novel and a protocol validated by experimental evidence for studying IDP and IDRs, as exemplified by the comparative study on the AR, ER and GR[156]. Recently, a CT-based fold analysis approach was able to detect the emergence of disordered compact states in a study on the effects of phosphorylation on the AF1 core of GR[157].

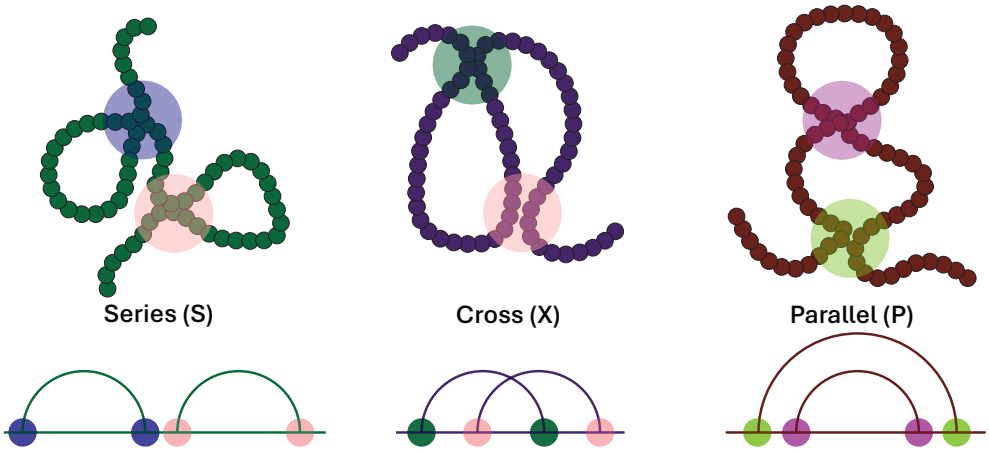

Figure 4. Circuit topology identifies relations between contacts. *A pair of contacts can be in one of three relations, with varying levels of complexity. Two pairs in series relations are entirely independent of each other. When one half of the binding site is within another pair, this is described as a cross relation, while if one contact lies entirely within another contact the relations is described as parallel. Respective circuit diagrams below offer a simpler overview of these relations in linear space.*

These approaches have provided new tools for the studying of IDPs and IDRs, offering frameworks for the experimental study of these highly dynamic and functionally important protein regions.

Scope of this thesis

Nuclear receptors are invaluable mediators of biological processes, and their dysregulation is central to numerous pathologies, including cancer and inflammatory diseases. Consequently, they have been a major focus of drug development. However, the efficacy of developments has been hindered by the limited knowledge on the conformational and allosteric dynamics of nuclear receptors. This thesis aims to gain a more comprehensive understanding of how the structural and functional dynamics mediate the Androgen Receptor and Nur77 nuclear receptors.

In **Chapter 2**, we look at effects of the disease-causing mutation, pQ expansion, on the folding dynamics of AR-NTD using extensive coarse-grained and all-atom molecular dynamics simulations, topological analyses and advanced docking approaches. We investigate how expansion of the pQ tract leads to misfolding of AR-NTD not seen in wt AR-NTD. we examine the consequences for the delicate balance of PPI and oligomerisation patterns, and how this disrupts critical interactions and regulatory mechanisms.

Our wt AR-NTD computational model suggested a distinct differential regulatory role for NR and CR on AR's ability to interact with DNA. In **Chapter 3** we utilise a combination of real-time single-molecule fluorescence microscopy and a thermophoretic approach to study the effects of these subregions on AR DBD-LBD (ΔNTD-AR) DNA binding capabilities.

Nur77 is an orphan receptor, meaning it does not have an (identified) endogenous ligand. The main consensus among researchers is that the activity of Nur77 is regulated through a plethora of external signals, including PTMs and PPIs. A critical interaction occurs with the glucocorticoid receptor (GR). These receptors are thought to engage in reciprocal repression, yet the molecular mechanism driving this cross-talk is unknown. **Chapter 4** demonstrates the DNA binding behaviours of Nur77, and how activated GR affects this process. Additionally, we look at the effects of dexamethasone, a common synthetic glucocorticoid, on Nur77's DNA interactions.

No endogenous ligands are known for Nur77, but some potential compounds have been identified to regulate Nur77's expression and/or activity. In **Chapter 5**, we examine cytosporone-B to determine its impact on DNA binding specificity and to propose a mechanistic model for its modulation of Nur77.

Finally, in **Chapter 6** the results of this thesis will be summarised and discussed within the wider scope of nuclear receptor and transcription factors and directions for future research into nuclear receptor structure and function will be proposed.

References

1. Kininis, M. & Kraus, W. L. A global view of transcriptional regulation by nuclear receptors: gene expression, factor localization, and DNA sequence analysis. *Nucl Recept Signal* **6**, nrs-06005 (2008).

2. Gustafsson, J. A. Historical overview of nuclear receptors. *Journal of Steroid Biochemistry and Molecular Biology* **157**, (2016).

3. Hollenberg, S. M. *et al*. Primary structure and expression of a functional human glucocorticoid receptor cDNA. *Nature* **318**, 635–641 (1985).

4. Kumar, R. & McEwan, I. J. Allosteric modulators of steroid hormone receptors: structural dynamics and gene regulation. *Endocr Rev* **33**, 271–299 (2012).

5. Pawlak, M., Lefebvre, P. & Staels, B. General Molecular Biology and Architecture of Nuclear Receptors. *Curr Top Med Chem* **12**, (2012).

6. Sever, R. & Glass, C. K. Signaling by nuclear receptors. *Cold Spring Harb Perspect Biol* **5**, (2013).

7. Weikum, E. R., Liu, X. & Ortlund, E. A. The nuclear receptor superfamily: A structural perspective. *Protein Science* **27**, 1876–1892 (2018).

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

9. Olefsky, J. M. Nuclear receptor minireview series. *Journal of Biological Chemistry* **276**, 36863–36864 (2001).

10. Xu, P. Nuclear receptors in health and diseases. *Int. J. Mol. Sci.* **24** (2023).

11. Germain, P., Staels, B., Dacquet, C., Spedding, M. & Laudet, V. Overview of nomenclature of nuclear receptors. *Pharmacol Rev* **58**, 685–704 (2006).

12. Rastinejad, F., Huang, P., Chandra, V. & Khorasanizadeh, S. Understanding nuclear receptor form and function using structural biology. *J Mol Endocrinol* **51**, T1–T21 (2013).

13. Schauwaers, K. *et al*. Loss of androgen receptor binding to selective androgen response elements causes a reproductive phenotype in a knockin mouse model. *Proc Natl Acad Sci U S A* **104**, (2007).

14. Claessens, F., Joniau, S. & Helsen, C. Comparing the rules of engagement of androgen and glucocorticoid receptors. *Cellular and Molecular Life Sciences* **74**, 2217–2228 (2017).

15. Meyer, C. A., Tang, Q. & Liu, X. S. Minireview: applications of next-generation sequencing on studies of nuclear receptor regulation and function. *Molecular Endocrinology* **26**, 1651–1659 (2012).

16. Heery, D. M., Kalkhoven, E., Hoare, S. & Parker, M. G. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**, (1997).

17. Moras, D. & Gronemeyer, H. The nuclear receptor ligand-binding domain: Structure and function. *Curr Opin Cell Biol* **10**, (1998).

18. Weatherman, R. V, Fletterick, R. J. & Scanlan, T. S. Nuclear-receptor ligands and ligand-binding domains. *Annu Rev Biochem* **68**, 559–581 (1999).

19. 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).

20. Shen, H. C. & Coetzee, G. A. The androgen receptor: unlocking the secrets of its unique transactivation domain. *Vitam Horm* **71**, 301–319 (2005).

21. Anbalagan, M., Huderson, B., Murphy, L. & Rowan, B. G. Post-translational modifications of nuclear receptors and human disease. *Nucl Recept Signal* **10**, e001 (2012).

22. Haelens, A., Tanner, T., Denayer, S., Callewaert, L. & Claessens, F. The hinge region regulates DNA binding, nuclear translocation, and transactivation of the androgen receptor. *Cancer Res* **67**, (2007).

23. Clinckemalie, L., Vanderschueren, D., Boonen, S. & Claessens, F. The hinge region in androgen receptor control. *Mol Cell Endocrinol* **358**, 1–8 (2012).

24. Kurakula, K., Koenis, D. S., van Tiel, C. M. & de Vries, C. J. M. NR4A nuclear receptors are orphans but not lonesome. *Biochim Biophys Acta* **1843**, 2543–2555 (2014).

25. Wruck, F. *et al*. Probing Single Chaperone Substrates. in *Biophysics of Molecular Chaperones: Function, Mechanisms and Client Protein Interactions* (eds. Hiller, S., Liu, M. & He, L.) 278-318 (Royal Society of Chemistry, 2023).

26. Echeverria, P. C. & Picard, D. Molecular chaperones, essential partners of steroid hormone receptors for activity and mobility. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1803**, 641–649 (2010).

27. Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* **27**, (1894).

28. Mirsky, A. E. & Pauling, L. On the Structure of Native, Denatured, and Coagulated Proteins. *Proceedings of the National Academy of Sciences* **22**, (1936).

29. Northrop, J. H. Crystalline pepsin: I. Isolation and tests of purity. *Journal of General Physiology* **13**, (1930).

30. Dunker, A. K. *et al*. Intrinsically disordered protein. *J Mol Graph Model* **19**, 26–59 (2001).

31. Anson, M. L. & Mirsky, A. E. The effect of denaturation on the viscosity of protein systems. *Journal of General Physiology* **15**, (1932).

32. Wu, H. Studies on denaturation of proteins xiii. a theory of denaturation. *Adv Protein Chem* **46**, (1995).

33. Anson, M. L. Protein Denaturation and the Properties of Protein Groups. *Adv Protein Chem* **2**, (1945).

34. Sela, M., White, F. H. & Anfinsen, C. B. Reductive cleavage of disulfide bridges in ribonuclease. *Science (1979)* **125**, (1957).

35. Kendrew, J. C. *et al*. Structure of myoglobin. *Nature* **185**, 427–1960 (1960).

36. Perutz, M. F. *et al*. Structure of hæmoglobin: a three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. *Nature* **185**, 416–422 (1960).

37. Brunori, M. 1960 Annus mirabilis: The birth of structural biology. *Rendiconti Lincei* **21**, (2010).

38. Karush, F. Heterogeneity of the Binding Sites of Bovine Serum Albumin. *J Am Chem Soc* **72**, (1950).

39. Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences* **44**, (1958).

40. Lund, M., Bjerrum, O. J. & Bjerrum, M. J. Structural heterogeneity of the binding sites of HSA for phenyl-

groups and medium-chain fatty acids: Demonstration of equilibrium between different binding conformations. *Eur J Biochem* **260**, 470–476 (1999).

41. Wootton, J. C. Non-globular domains in protein sequences: Automated segmentation using complexity measures. *Comput Chem* **18**, (1994).

42. Romero, P. *et al*. Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput* (1998).

43. Dunker, A. K. *et al*. What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disord Proteins* **1**, (2013).

44. Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1804**, 1231–1264 (2010).

45. Babu, M. M., van der Lee, R., de Groot, N. S. & Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol* **21**, 432–440 (2011).

46. Tompa, P. Unstructural biology coming of age. *Curr Opin Struct Biol* **21**, 419–425 (2011).

47. Choi, W. J., Haratipour, Z. & Blind, R. D. Full-length nuclear receptor allosteric regulation. *J Lipid Res* **64**, 100406 (2023).

48. Midic, U., Oldfield, C., Dunker, A., Obradovic, Z. & Uversky, V. Unfoldomics of Human Genetic Diseases: Illustrative Examples of Ordered and Intrinsically Disordered Members of the Human Diseasome. *Protein Pept Lett* **16**, (2009).

49. Williams, R. M. *et al*. The protein non-folding problem: amino acid determinants of intrinsic order and disorder. in *Biocomputing 2001* 89–100 (World Scientific, 2000).

50. Kozlowski, L. P. & Bujnicki, J. M. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **13**, (2012).

51. Garza, A. S., Ahmad, N. & Kumar, R. Role of intrinsically disordered protein regions/domains in transcriptional regulation. *Life Sci* **84**, 189–193 (2009).

52. Kumar, R. A. J. & Thompson, E. B. Transactivation functions of the N-terminal domains of nuclear hormone receptors: protein folding and coactivator interactions. *Molecular endocrinology* **17**, 1–10 (2003).

53. Vacic, V. *et al*. Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* **6**, (2007).

54. Dunker, A. K. & Uversky, V. N. Signal transduction via unstructured protein conduits. *Nat Chem Biol* **4**, 229–230 (2008).

55. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr Opin Struct Biol* **19**, 31–38 (2009).

56. Ball, L. J. *et al*. Cell type- and estrogen receptor-subtype specific regulation of selective estrogen receptor modulator regulatory elements. *Mol Cell Endocrinol* **299**, (2009).

57. Spillman, M. A. *et al*. Tissue-specific pathways for estrogen regulation of ovarian cancer growth and metastasis. *Cancer Res* **70**, (2010).

58. Smith, C. L. & O'malley, B. W. Coregulator function: a key to understanding tissue specificity of selective receptor modulators. *Endocr Rev* **25**, 45–71 (2004).

59. Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **33**, (2008).

60. Oldfield, C. J. & Dunker, A. K. Intrinsically

disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**, (2014).

61. Uversky, V. N. Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* **43**, 1090–1103 (2011).

62. Bah, A. & Forman-Kay, J. D. Modulation of intrinsically disordered protein function by post-translational modifications. *Journal of Biological Chemistry* **291**, (2016).

63. Chin, A. F., Toptygin, D., Elam, W. A., Schrank, T. P. & Hilser, V. J. Phosphorylation Increases Persistence Length and End-to-End Distance of a Segment of Tau Protein. *Biophys J* **110**, (2016).

64. Bah, A. *et al*. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, (2015).

65. Monahan, Z. *et al*. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity . *EMBO J* **36**, (2017).

66. Sigler, P. B. Acid blobs and negative noodles. *Nature* **333**, 210–212 (1988).

67. Lambert, S. A. *et al*. The human transcription factors. *Cell* **172**, 650–665 (2018).

68. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613–626 (2012).

69. Khan, S. H. & Okafor, C. D. Interactions governing transcriptional activity of nuclear receptors. *Biochem Soc Trans* **50**, 1941–1952 (2022).

70. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **37**, 215–246 (2008).

71. Xu, H. E. Family reunion of nuclear hormone receptors: structures, diseases, and drug discovery. *Acta Pharmacol Sin* **36**, 1–2 (2015).

72. Tenbaum, S. & Baniahmad, A. Nuclear receptors: structure, function and involvement in disease. *Int J Biochem Cell Biol* **29**, 1325–1341 (1997).

73. McEwan, I. J., Lavery, D., Fischer, K. & Watt, K. Natural Disordered Sequences in the Amino Terminal Domain of Nuclear Receptors: Lessons from the Androgen and Glucocorticoid Receptors. *Nucl Recept Signal* **5**, nrs.05001 (2007).

74. 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).

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

76. McEwan, I. J. & Brinkmann, A. O. Androgen physiology: receptor and metabolic disorders. *Endotext [Internet]* (2021).

77. Hunter, I., Jamieson, C. & McEwan, I. J. The androgen receptor amino-terminal domain: structure, function and therapeutic potential. *Endocrine Oncology* **5**, e240061 (2025).

78. 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).

79. Askew, E. B., Gampe, R. T., Stanley, T. B., Faggart, J. L. & Wilson, E. M. Modulation of Androgen Receptor Activation Function 2 by Testosterone and Dihydrotestosterone. *Journal of Biological Chemistry* **282**, 25801–25816

(2007).

80. 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).

81. 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).

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

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

84. 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**, 7341–7346 (1995).

85. Christiaens, V. *et al.* Characterization of the two coactivator-interacting surfaces of the androgen receptor and their relative role in transcriptional control. *Journal of Biological Chemistry* **277**, 49230–49237 (2002).

86. McEwan, I. J. & Gustafsson, J.-Å. Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF. *Proceedings of the National Academy of Sciences* **94**, 8485–8490 (1997).

87. 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 (2018).

88. Bevan, C. L., Hoare, S., Claessens, F., Heery, D. M. & Parker, M. G. The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1. *Mol Cell Biol* **19**, 8383–8392 (1999).

89. Reid, J., Murray, I., Watt, K., Betney, R. & McEwan, I. J. The androgen receptor interacts with multiple regions of the large subunit of general transcription factor TFIIF. *Journal of Biological Chemistry* **277**, 41247–41253 (2002).

90. Cato, L. *et al.* Development of Bag-1L as a therapeutic target in androgen receptor-dependent prostate cancer. *Elife* **6**, e27159 (2017).

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

92. Spada, A. R. La, Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**, 77–79 (1991).

93. Giovannucci, E. *et al.* The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc Natl Acad Sci U S A* **94**, (1997).

94. Ingles, S. A. *et al.* Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. *J Natl Cancer Inst* **89**, (1997).

95. Cato, L. & Shomali, M. AR structural variants and prostate cancer. in *Nuclear Receptors in Human Health and Disease* 195–211 (Springer, 2022).

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

97. Erdős, G., Deutsch, N. & Dosztányi, Z. AIUPred – Binding: Energy Embedding to Identify Disordered Binding Regions. *J Mol Biol* **437**, 169071 (2025).

98. Erdős, G. & Dosztányi, Z. AIUPred: combining energy estimation with deep learning for the enhanced prediction of protein disorder. *Nucleic Acids Res* **52**, W176–W181 (2024).

99. De Mol, E. *et al.* EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. *ACS Chem Biol* **11**, 2499–2505 (2016).

100. Lith, S. C. & de Vries, C. J. M. Nuclear receptor Nur77: its role in chronic inflammatory diseases. *Essays Biochem* **65**, 927–939 (2021).

101. Li, L. *et al.* Impeding the interaction between Nur77 and p38 reduces LPS-induced inflammation. *Nat Chem Biol* **11**, 339–346 (2015).

102. Lith, S. C. *et al.* Nuclear receptor Nur77 regulates immunomechanics of macrophages. *Eur J Cell Biol* **103**, 151419 (2024).

103. Zhan, Y. *et al.* Cytosporone B is an agonist for nuclear orphan receptor Nur77. *Nat Chem Biol* **4**, (2008).

104. Vinayavekhin, N. & Saghatelian, A. Discovery of a protein-metabolite interaction between unsaturated fatty acids and the nuclear receptor Nur77 Using a metabolomics approach. *J Am Chem Soc* **133**, (2011).

105. Chintharlapalli, S. *et al.* Activation of Nur77 by selected 1,1-bis(3'-indolyl)-1-(p-substituted phenyl)methanes induces apoptosis through nuclear pathways. *Journal of Biological Chemistry* **280**, (2005).

106. Lakshmi, S. P., Reddy, A. T., Banno, A. & Reddy, R. C. Molecular, chemical, and structural characterization of prostaglandin A2 as a novel agonist for Nur77. *Biochemical Journal* **476**, (2019).

107. Hu, M. *et al.* Celastrol-Induced Nur77 Interaction with TRAF2 Alleviates Inflammation by Promoting Mitochondrial Ubiquitination and Autophagy. *Mol Cell* **66**, (2017).

108. Safe, S., Shrestha, R. & Mohankumar, K. Orphan nuclear receptor 4A1 (NR4A1) and novel ligands. *Essays Biochem* **65**, 877–886 (2021).

109. Wansa, K. D. S. A., Harris, J. M. & Muscat, G. E. O. The activation function-1 domain of Nur77/NR4A1 mediates trans-activation, cell specificity, and coactivator recruitment. *Journal of biological chemistry* **277**, 33001–33011 (2002).

110. Wansa, K. D. S. A., Harris, J. M., Yan, G., Ordentlich, P. & Muscat, G. E. O. The AF-1 domain of the orphan nuclear receptor NOR-1 mediates trans-activation, coactivator recruitment, and activation by the purine anti-metabolite 6-mercaptopurine. *Journal of biological chemistry* **278**, 24776–24790 (2003).

111. Lévesque, D. & Rouillard, C. Nur77 and retinoid X receptors: crucial factors in dopamine-related neuroadaptation. *Trends Neurosci* **30**, 22–30 (2007).

112. Pearen, M. A. & Muscat, G. E. O. Minireview: Nuclear hormone receptor 4A signaling: implications for metabolic disease. *Mol Endocrinol* **24**, 1891–1903 (2010).

113. Lee, S.-O. *et al.* Inactivation of the orphan nuclear receptor TR3/Nur77 inhibits pancreatic cancer cell and tumor growth. *Cancer Res* **70**, 6824–6836 (2010).

114. Moll, U. M., Marchenko, N. & Zhang, X. K. p53 and Nur77/TR3–transcription factors that directly target mitochondria for cell death induction. *Oncogene* **25**, 4725–4743 (2006).

115. Kolluri, S. K. *et al.* A short Nur77-derived peptide converts Bcl-2 from a protector to a killer. *Cancer Cell* **14**, 285–298 (2008).

116. Lin, B. *et al.* Conversion of Bcl-2 from protector to killer by interaction with nuclear orphan receptor Nur77/TR3. *Cell* **116**, 527–540 (2004).

117. Bilodeau, S. *et al.* Role of Brg1 and HDAC2 in GR trans-repression of the pituitary POMC gene and misexpression in Cushing disease. *Genes Dev* **20**, 2871–2886 (2006).

118. Martens, C., Bilodeau, S., Maira, M., Gauthier, Y. & Drouin, J. Protein-protein interactions and transcriptional antagonism between the subfamily of NGFI-B/Nur77 orphan nuclear receptors and glucocorticoid receptor. *Molecular Endocrinology* **19**, (2005).

119. Philips, A. *et al.* Antagonism between Nur77 and glucocorticoid receptor for control of transcription. *Mol Cell Biol* **17**, 5952–5959 (1997).

120. Santofimia-Castaño, P. *et al.* Targeting intrinsically disordered proteins involved in cancer. *Cellular and Molecular Life Sciences* **77**, 1695–1707 (2020).

121. Metallo, S. J. Intrinsically disordered proteins are potential drug targets. *Curr Opin Chem Biol* **14**, 481–488 (2010).

122. Mészáros, B., Hajdu-Soltész, B., Zeke, A. & Dosztányi, Z. Mutations of intrinsically disordered protein regions can drive cancer but lack therapeutic strategies. *Biomolecules* **11**, 381 (2021).

123. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).

124. Fuertes, G., Nevola, L. & Esteban-Martín, S. Perspectives on drug discovery strategies based on IDPs. in *Intrinsically disordered proteins* 275–327 (Elsevier, 2019).

125. Heller, G. T. *et al.* Sequence specificity in the entropy-driven binding of a small molecule and a disordered peptide. *J Mol Biol* **429**, 2772–2779 (2017).

126. Ban, D., Iconaru, L. I., Ramanathan, A., Zuo, J. & Kriwacki, R. W. A small molecule causes a population shift in the conformational landscape of an intrinsically disordered protein. *J Am Chem Soc* **139**, 13692–13700 (2017).

127. Iconaru, L. I. *et al.* Discovery of small molecules that inhibit the disordered protein, p27 kip1. *Sci Rep* **5**, 15686 (2015).

128. Tatenhorst, L. *et al.* Fasudil attenuates aggregation of α-synuclein in models of Parkinson's disease. *Acta Neuropathol Commun* **4**, 39 (2016).

129. Heller, G. T. *et al.* Small-molecule sequestration of amyloid-β as a drug discovery strategy for Alzheimer's disease. *Sci Adv* **6**, eabb5924 (2020).

130. Zhao, J. *et al.* EGCG binds intrinsically disordered N-terminal domain of p53 and disrupts p53-MDM2 interaction. *Nat Commun* **12**, 986 (2021).

131. Iconaru, L. I. *et al.* Small molecule sequestration of the intrinsically disordered protein, p27Kip1, within soluble oligomers. *J Mol Biol* **433**, 167120 (2021).

132. Yang, Y. C. *et al.* Targeting Androgen Receptor Activation Function-1 with EPI to Overcome Resistance Mechanisms in Castration-Resistant Prostate Cancer. *Clinical Cancer Research* **22**, 4466–4477 (2016).

133. Liu, C. *et al.* Diffusing protein binders to intrinsically disordered proteins. *Nature* 1–9 (2025).

134. Konrat, R. NMR contributions to

structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance* **241**, 74–85 (2014).

135. Abyzov, A. *et al*. Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation. *J Am Chem Soc* **138**, 6240–6251 (2016).

136. Salvi, N., Abyzov, A. & Blackledge, M. Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation. *J Phys Chem Lett* **7**, 2483–2489 (2016).

137. Schiavina, M. *et al*. Intrinsically disordered proteins studied by NMR spectroscopy. *J Magn Reson Open* **18**, 100143 (2024).

138. Goundaroulis, D., Dorier, J., Benedetti, F. & Stasiak, A. Studies of global and local entanglements of individual protein chains using the concept of knotoids. *Sci Rep* **7**, 6309 (2017).

139. Mishra, R. & Bhushan, S. Knot theory in understanding proteins. *J Math Biol* **65**, 1187–1213 (2012).

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

141. Golovnev, A. & Mashaghi, A. Generalized Circuit Topology of Folded Linear Chains. *iScience* **23**, 1–17 (2020).

142. Scalvini, B. *et al*. Topology of Folded Molecular Chains: From Single Biomolecules to Engineered Origami. *Trends Chem* **2**, 609–622 (2020).

143. Golovnev, A. & Mashaghi, A. Circuit topology for bottom-up engineering of molecular knots. *Symmetry (Basel)* **13**, 2353 (2021).

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

145. Mashaghi, A. & Ramezanpour, A. Distance measures and evolution of polymer chains in their topological space. *Soft Matter* **11**, 6576–6585 (2015).

146. Nikoofard, N. & Mashaghi, A. Implications of Molecular Topology for Nanoscale Mechanical Unfolding. *Journal of Physical Chemistry B* **122**, 9703–9712 (2018).

147. Mashaghi, A. & Ramezanpour, A. Circuit topology of linear polymers: a statistical mechanical treatment. *RSC Adv* **5**, 51682–51689 (2015).

148. Heidari, M. *et al*. Topology of internally constrained polymer chains. *Physical Chemistry Chemical Physics* **19**, 18389–18393 (2017).

149. Satarifard, V. *et al*. Topology of polymer chains under nanoscale confinement. *Nanoscale* **9**, 12170–12177 (2017).

150. Heidari, M., Satarifard, V. & Mashaghi, A. Mapping a single-molecule folding process onto a topological space. *Physical Chemistry Chemical Physics* **21**, 20338–20345 (2019).

151. Heidari, M., Schiessel, H. & Mashaghi, A. Circuit Topology Analysis of Polymer Folding Reactions. *ACS Cent Sci* **6**, 839–847 (2020).

152. Scalvini, B., Sheikhhassani, V. & Mashaghi, A. Topological principles of protein folding. *Phys Chem Chem Phys* **23**, 21316–21328 (2021).

153. Mugler, A., Tans, S. J. & Mashaghi, A. Circuit topology of self-interacting chains: Implications for folding and unfolding dynamics. *Physical Chemistry Chemical Physics* **16**, 22537–22544 (2014).

154. Scalvini, B., Schiessel, H., Golovnev, A. & Mashaghi, A. Circuit topology analysis of cellular genome reveals signature

motifs, conformational heterogeneity, and scaling. *iScience* **25**, (2022).

155. Nikoofard, N. & Mashaghi, A. Topology sorting and characterization of folded polymers using nanopores. *Nanoscale* **8**, 4643–4649 (2016).

156. Scalvini, B. *et al*. Circuit Topology Approach for the Comparative Analysis of Intrinsically Disordered Proteins. *J Chem Inf Model* **63**, (2023).

157. Akulov, V., Jiménez Panizo, A., Estébanez-Perpiñá, E., van Noort, J. & Mashaghi, A. Phosphorylation-Regulated Conformational Diversity and Topological Dynamics of an Intrinsically Disordered Nuclear Receptor. *J Phys Chem B* (2024).