Structural insights into “acid blobs and negative noodles” – The androgen receptor as a case study

Elizbieta Maria Szulc

UNIVERSITAT DE BARCELONA
FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

INSTITUTE FOR RESEARCH IN BIOMEDICINE
IRB BARCELONA

STRUCTURAL INSIGHTS INTO "ACID BLOBS AND NEGATIVE NOODLES" – THE ANDROGEN RECEPTOR AS A CASE STUDY

Elzbieta Maria Szulc
2019

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

PROGRAMA DE DOCTORAT EN BIOMEDICINA

INSTITUT DE RECERCA BIOMÈDICA

STRUCTURAL INSIGHTS INTO "ACID BLOBS AND NEGATIVE NOODLES" – THE ANDROGEN RECEPTOR AS A CASE STUDY

Memòria presentada per Elzbieta Maria Szulc per optar al títol de doctor per la Universitat de Barcelona

Director	Tutor	Doctoranda
Xavier Salvatella Giralt	Albert Tauler Girona	Elzbieta Maria Szulc

Elzbieta Maria Szulc
2019

Acknowledgements

This work could not be accomplished without help and guidance of many people I met along the way through these projects.

In the first place I would like to thank my thesis supervisor, Dr. Xavier Salvatella, for letting me conduct my PhD in his laboratory. Thank you for showing me how to dissect problems into their first principle components, a task I am still learning how to do. Thank you for letting me use the laboratory as a playground and try to find my own interests. Finally, thank you for all the hours spent on analysis, discussions and re-analysis and re-discussion.

I am grateful to La Caixa for awarding me with a fellowship that allowed me to conduct this research, attend many valuable conferences and live in this beautiful city of Barcelona.

I would like to express my gratitude to Tony Kossiakoff who prepared me for this journey. You are a great mentor to all your Polish kids.

I would like to thank Dr. Tanja Mittag for inviting me into her lab space in St. Jude Children's Research Hospital, where I met many wonderful people like Jill Bouchard, Nicole Milkovic, Joel Otero, Erik Martin and others. A month in your laboratory was full of precious lessons.

Special thanks go to Jill Bouchard, a girl with the most beautiful smile in the whole Tennessee, for teaching me about phase separation and for doing it while having great fun. I hope to meet you one day in Canada.

I am grateful to all the people that helped in the projects: Jesús García for acquiring and analysing the NMR experiments; Anna Lladó, Lídia Bardia and Nikolaos Giakoumakis for your help with microscopy experiments; Nicholas Berrow for dealing with our countless cloning requests; Isabelle Brun Heath for solving the cloning puzzles; Nuria Villegas for taking care of the precious Cary100 equipment; Marta Vilaseca and her team for all the mass spectrometry experiments; Joan Pous and Roman Bonet for your help with the DLS.

I am grateful to our collaborators Diego M. Presman, Grégory Fettweis from Gordon Hager's group and Cristina Batlle from Salvador Ventura's laboratory, Laura Tossato and Maria Pennuto.

I am especially thankful to Mireia Pesarrodona without whom the phase separation project would not be possible. Thank you for all your efforts in trying to optimize purifications of various constructs, including the beast, full-length AR. Thank you also for being a good friend and always believing in my basketball skills. Thank you and thank to Jomi Valverde and Isabel Latorre for all the protein preps. Without your work these projects would not be possible.

I would like to give special thanks to the friends I met in the Salvatella lab. Giulio Chiesa, who helped me getting accustomed to the Salvatella lab, but also supported me during the years we shared our life here. Büşra Topal for all our coffee breaks where we

posted millions of non-scientific life-oriented questions and answered very few of those. Andreea Balaceanu for your smiley face that was too short with us in the lab. Chris Phang for your support and drama comedy phone calls during the last year. Bahar Eftekharzadeh and Daniele Mungianu for your friendships. Claudio di Sanza and Víctor Buzón who shared their knowledge with me during my first year in the lab. Eva de Mol for your work that was the base of my PhD project. Many thanks also to the current members of the Salvatella lab: Albert Escobedo for forcing me to FRAP everything and for your friendship; Paula Martínez for bringing the cell biology expertise on board; Marta Frigolé for making the STED experiments reality and for all the short chats about life; Carla Garcia for being the annoyingly young PhD student; Mateusz Biesaga for bringing more of the Polishness to the lab; Marta Marín for your help in the project; Jonathan Piccirillo for the delicious cakes.

I would like to thank my TAC committee: Miquel Pons, Natàlia Carulla and Roger Gomis for each year's meeting.

I would like to thank Dr. Margarida Gairí for all the lunches full of support and many people that made the IRB corridors worth walking through: Sylwia Gawrzak and Ewelina Guca for your frendship and support, Anabel-Lise Le Roux for long chats and hikes, Alexandra Balaceanu, Antonija Kuzmanić, Gianmarco Di Mauro, Paula Climent, Daniela Ferreira, Laareb Mohammad Jabeen, Marta Rodríguez, Liudmila Filonava, Tiago Lopes and many others.

I would also like to thank Julka Skalska, Thomas Wilhelm and Rute Fernandes for making the Gràcia district homey.

I would like to give special thanks to the Polish Theater group Te-ART for bringing an incredible amount of laughter and inspiration into my life during the past year. Thank you Facetka, Facetek, Aga, Aga Anathomy, Gosia, Agata, Ania, Carles, Juan and Tom.

I thank Sam Harris and Tomasz Stawiszyński for making great podcasts that made commuting time one of my favorite moments of a day.

Finally, none of this would be possible without the support of my family, especially my father, Zygmunt, who I can always count on, Jola, who made all the visits in Poland warm and cozy with the smell of a freshly baked cake, Ala and Ola, who are the double amount of joy, Marysia, Piotr and Zeta the dog, my oldest friends, Dagmara Wądołowska and Magda Żukowska and one of my favorite humans, Berat Denizdurduran.

Thank you all.

Table of Contents

List of important abbreviations

AD	Activation Domain
AR	Androgen Receptor
CRPC	Castration Resistant Prostate Cancer
DBD	DNA-binding domain
ID	Intrinsic Disorder
IDP	Intrinsically Disordered Protein
IDR	Intrinsically Disordered Region
LBD	Ligand Binding Domain
LLPS	Liquid-liquid Phase Separation
NR	Nuclear Receptor
PC	Prostate Cancer
Rap74-CTD	RNA polymerase II-associated protein 74 C-terminal domain
SBMA	Spinal and Bulbar Muscular Atrophy
SPOP	Speckle-type POZ Protein
SR	Steroid Receptor
TF	Transcription Factor
TFIIF	Transcription Factor II F

List of tables in the Introduction and Discussion

List of figures in the Introduction and Discussion

1 INTRODUCTION

1.1 Intrinsically disordered proteins

Till not long ago intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) of proteins were just unexplored part of the "dark proteome". Analogously to the dark matter in physics, term "dark proteome" has been used by protein scientists to describe regions of proteins never observed by experimental structure determination and inaccessible to homology modeling (Perdigao et al. 2015). An IDP (and IDR, included in the IDP term in the rest of this thesis) is a protein that lacks a well-defined three-dimensional structure in its native state. The discovery of IDPs questioned the lock and key model of protein function (Dunker et al. 2001): the model proposed by Emil Fisher in 1894 and later validated experimentally by many, associates the function of a protein, for example enzymatic activity, with its well-defined shape that permits it to bind to a molecule complementary in shape, on which the enzymatic activity is to be performed (Fischer 1894; Northrop 1930; Mirsky and Pauling 1931; Dunker et al. 2001). From the beginning of protein crystallography, it was of notion that some of the protein segments did not show electron density, for which one of the explanations can be lack of structure. In 1978, nuclear magnetic resonance (NMR) studies showed that a large portion of highly charged, and what is important, functional, tail of histone H5 is disordered (Aviles et al. 1978). Since then the functional relevance of intrinsic disorder has been extensively analyzed and reviewed, with a paradigm shift of the field in the late 1990s and early 2000s (Wright and Dyson 1999; Dunker et al. 2001; Tompa 2002; Uversky 2002) thanks to developments in the NMR and computation biology fields (Christopher J Oldfield and Dunker 2014). As for 30th of May 2019, there are 803 proteins and 2167 regions of intrinsic disorder deposited in the DisProt database[1].

Intrinsic disorder (ID) is encoded in protein sequences. IDPs can be recognized by their low mean hydrophobicity and high net charge, resulting in low compaction and high repulsion (Uversky et al. 2000). In particular they are depleted in order-promoting amino acids, like Ile, Leu, Val, Trp, Tyr, Phe, Cys, Asn and enriched in disorder-promoting amino acids, like Ala, Arg, Gly, Gln, Ser, Glu, Lys, and Pro (Williams et al. 2001). Thanks to the characteristic amino acid composition of IDPs, it is currently possible to accurately predict intrinsic disorder from a linear sequence of a protein[2] (Uversky and Dunker 2010; Kozlowski and Bujnicki 2012).

IDPs can be described as dynamic ensembles, nevertheless their structure is often captured by a limited number of lower-energy conformations (Choy and Forman-Kay 2001). IDPs can exist as random coils (extended), pre-molten globules (partially collapsed with residual secondary structure), partially folded or molten globules (domains with secondary structure and compaction) and domains with poorly packed side chains (Uversky and Dunker 2010). Unlike folded proteins, IDPs are very promiscuous in binding which makes them good interactions hubs. They can form disordered complexes (often called "fuzzy") or undergo disorder-to-order transition upon binding to a partner. Many IDPs can adopt different structures, even in the same sequence motif, while binding to different

[1] http://www.disprot.org
[2] http://iimcb.genesilico.pl/metadisorder/list_of_protein_disorder_tools_programs.html

proteins. A classical, at this point, an example of a promiscuous IDP is p53, Figure 1 (Tompa and Fuxreiter 2008; Uversky 2011; Christopher J Oldfield and Dunker 2014).

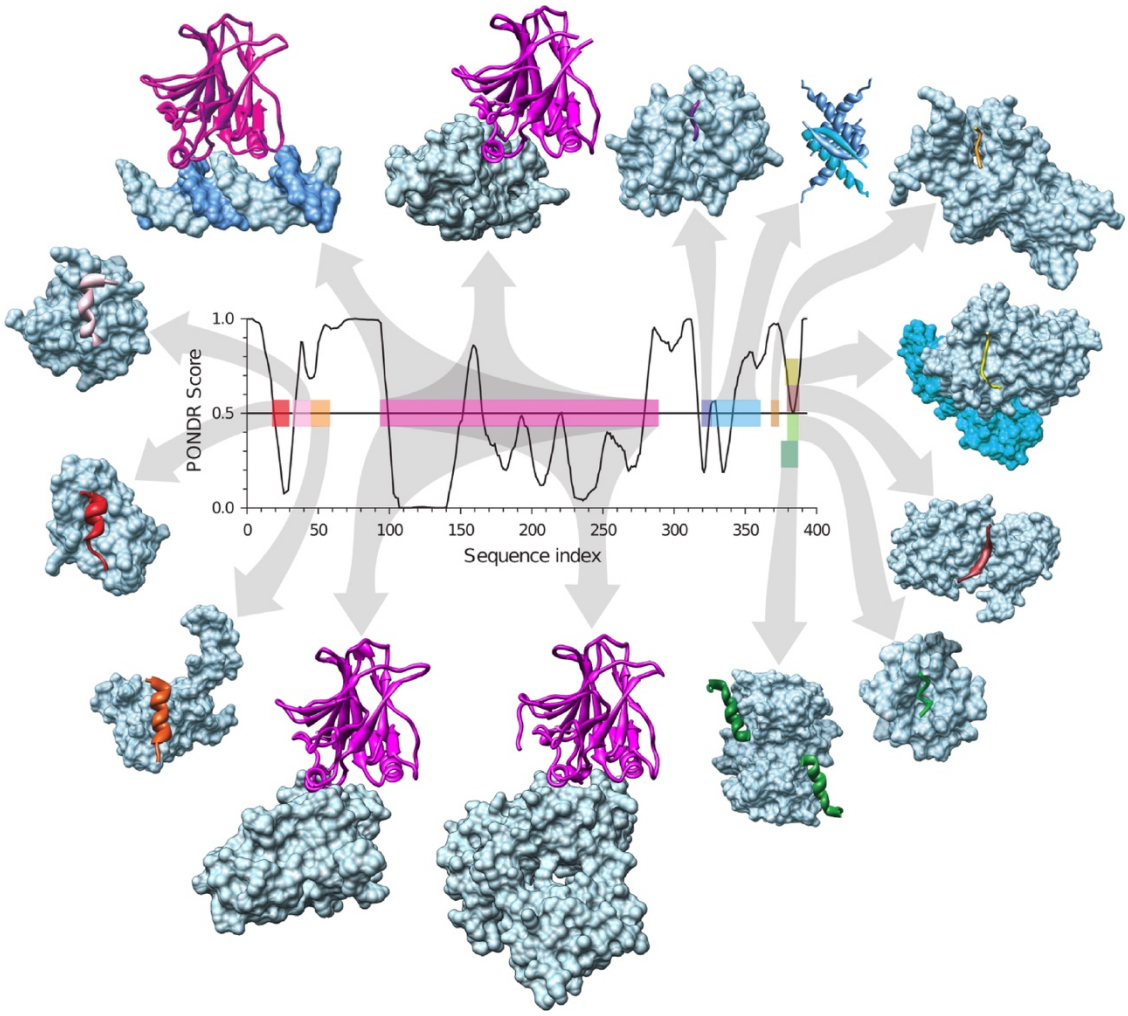


Figure 1. p53 interaction with different binding partners

A structure versus disorder prediction on the p53 amino acid sequence is shown in the center of the figure (up = disorder, down = order) along with the structures of various regions of p53 bound to fourteen different partners. The predicted central region of structure with the predicted amino and carbonyl termini as being disordered have been confirmed experimentally for p53. The various regions of p53 are color coded to show their structures in the complex and to map the binding segments to the amino acid sequence. Starting with the p53-DNA complex (top, left, magenta protein, blue DNA), and moving in a clockwise direction, the Protein Data Bank IDs and partner names are given as follows for the fourteen complexes: (1tsr – DNA), (1gzh – 53BP1), (1q2d – gcn5), (3sak – p53 (tetramerization domain)), (1xqh – set9), (1h26 – cyclinA), (1ma3 – sirtuin), (1jsp – CBP bromodomain), (1dt7 – s100ββ), (2h1l – sv40 Large T antigen), (1ycs – 53BP2), (2gs0 – PH), (1ycr – MDM2), and (2b3g – rpa70).
From (Uversky 2011).

The physicochemical and structural landscape of IDPs is affected by a plethora of post-translational modifications (PTMs). The accessibility of modification sites in IDPs polypeptide chain makes them ideal targets for modulations by PTMs, increasing the spectrum of available protein states (Bah and Forman-Kay 2016). PTMs that change

charge, phosphorylation and acetylation, affect the compactness of IDPs (Chin et al. 2016), switch between disordered and ordered states (Bah et al. 2014) or dissolve or promotes phase separation (Monahan et al. 2017) (more on this process in Chapter 1.4). Phosphorylations within binding sites have been shown to change binding affinities for target proteins (Feng et al. 2009). Beyond the most studied PTM which is phosphorylation, there are many others, such as the previously mentioned acetylation, but also methylation, ubiquitination and sumoylation of lysines, methylation and citrullination of arginines and isomerization of prolines. All of those are used for example to expand the versatility of states of histones in cells (Bah and Forman-Kay 2016; Shliaha et al. 2017).

Table 1. The 20 biological processes most strongly correlated with predicted disorder

Keyword	Z-score
1. Differentiation	18.8
2. Transcription	14.6
3. Transcription regulation	14.3
4. Spermatogenesis	13.9
5. DNA condensation	13.4
6. Cell cycle	12.2
7. mRNA processing	10.9
8. mRNA splicing	10.1
9. Mitosis	9.4
10. Apoptosis	9.3
11. Protein transport	8.8
12. Meiosis	8.7
13. Cell division	8.5
14. Ubl conjugation pathway	8.1
15. Wnt signaling pathway	6.6
16. Neurogenesis	6.6
17. Chromosome partition	6.4
18. Ribosome biogenesis	5.9
19. Chondrogenesis	5.6
20. Growth regulation	5.1

Z-score: measure of the disorder-function relationships used in (Xie et al. 2007)
Adapted from (Xie et al. 2007; A. K. Dunker et al. 2015)

IDPs have unique advantages over folded proteins for certain roles (Liu and Huang 2014). Examples of functions they perform are: constituting flexible linkers between structured domains; providing rubber-like entropic springs; containing sites for PTMs; containing sites for regulatory protease digestion; containing autoinhibitory domains; containing sites for binding to partners such as DNA, tRNA, rRNA, mRNA, protein or metal ions such as Zn^{2+}; containing signals such as the one for nuclear localization, and enabling movement through narrow pores (A.K. Dunker et al. 2015). They have been found to be evolutionary advantageous, with disordered segments found to occur in 2.0% of Archaean, 4.2% of eubacterial and 33.0% of eukaryotic proteins (Ward et al. 2004). Due to this multiplicity of roles they are found in many biological processes important in the development of multicellular organisms. Table 1 represents the 20 biological processes

most strongly correlated with predicted disorder (A.K. Dunker et al. 2015). Notice that transcription and transcriptional regulation are in positions 2 and 3 respectively.

The central role of IDPs in crucial biological pathways automatically puts them in the center of many pathologies. The concept of disorder in disorders (D2 concept) introduced by V.N. Uversky, C. J. Oldfield and A. K. Dunker in 2008 to emphasize high abundance of ID in proteins associated with various diseases (Uversky et al. 2008). Nevertheless, the inherent lack of structure of IDPs makes it impossible to target them by using the classical approaches of drug design (Uversky 2012). Two exceptions being targeting the ordered part of an IDP or targeting a binding pocket in its ordered ligand, as in the case of the ID region of p53 that folds into helix upon interaction with the folded Mdm2 (S. Wang et al. 2017). A proof-of-concept example of targeting directly an IDP comes from a high-throughput screening of Myc-Max dimer inhibitors (Yin et al. 2003), where found inhibitors were shown to bind to few amino acid stretches in the unfolded c-Myc (Hammoudeh et al. 2009). IDPs are commonly found in protein misfolding diseases, including neurodegenerative disorders, where the pathological oligomerization and aggregation of proteins trigger a cascade of events eventually leading to neurodegeneration and in many cases, death. Recently a new type of molecules called “molecular tweezers”, binding to lysine or arginine residues and preventing them from forming interactions (Fokkens et al. 2005), were shown to inhibit aggregation of a few aggregation prone proteins, including α-synuclein, an IDP involved in Parkinson’s disease (Prabhudesai et al. 2012). It has been proposed that the analysis of existing small molecule IDPs targeting drugs can provide hints for future drugs design. The captured common features of the small molecules included higher hydrophobicity and aromaticity and the presence of more aromatic rings than in conventional drugs (Ruan et al. 2019). This, rather general, description shows that the field is still in its infancy.

1.1.1 Intrinsic disorder in Transcription Factors

As mentioned in Chapter 1.1, IDPs have been found to be abundant in transcription and transcription regulation processes. The core proteins in transcriptional networks, responsible for regulation of gene expression, are called transcription factors (TFs). The characteristic domains of TFs are the DNA binding domain (DBD) and activation domain (AD) (Staby et al. 2017). The AD mediates activation or repression of gene transcription through binding to co-activator or co-repressor proteins and chromatin remodeling factors (Naar, Anders et al. 2001). In the past ADs have been described as "acid blobs and negative noodles" (Paul B. Sigler 1988), which, as the understanding of intrinsic disorder increased, has been translated to IDRs. Using a proteomic approach, it has been shown that IDPs are overrepresented in the nucleus of the cell (Skupien-Rabian et al. 2015). Bioinformatic studies showed that 94.13 to 82.63% of TFs possess extended (>30 amino acids) regions of intrinsic disorder, in comparison to 54.51 and 18.64% of the proteins in two control datasets (Liu et al. 2006).

Transcription relies on a multitude of protein-protein interactions, and, as mentioned previously, intrinsic disorder can provide many advantages over structure in interaction networks, including conformational plasticity, promiscuity, and regulation by various PTMs. For example, TF p53 contains multiple disordered regions through which it binds to diverse partners to regulate multiple mechanisms, Figure 1 (Christopher J. Oldfield and Dunker 2014). The IDP state of ADs, meaning level of compaction and presence of

secondary structure elements, is only characterized experimentally for few TFs (Kumar, Betney, Li, E Brad Thompson, et al. 2004; Kumar and Thompson 2005; Staby et al. 2017; Huang et al. 2018). Considering that ADs are important sites for TFs' protein-protein interactions, there is an unexplored expansion of the conformational landscape of TFs'ADs in complexes.

TFs have been explored as drug targets, with recent focus on ID ADs (Dunker and Uversky 2010; Tsafou et al. 2018). Examples of research in that area include: aforementioned in Section 1.1. p53-Mdm2 and Myc-Max complexes, oncogenic translocation-generated fusion protein EWS-FLI1 (Erkizan et al. 2009) and the androgen receptor (AR), the main driver in prostate cancer (PC) (Andersen et al. 2010).

1.2 Nuclear receptors

Nuclear receptors (NRs) constitute a family of TFs that regulate gene expression in response to ligands (Mangelsdorf et al. 1995). The unique property of the NRs ligands is their lipophilicity that allows them to cross the cell membrane and directly bind to their targets instead of acting on the surface receptors. That property makes the NRs signaling pathways efficient in rapid response to external stimuli. NRs ligands include steroid hormones, metabolic intermediates and products, and xenobiotics. The first notion that the steroid hormones regulate transcription came from the studies on salivary glands of insect larva in 1960s (Clever, U., Karlson 1960). Since then we have discovered 48 different NRs in humans, although a subset of those, called "orphan receptors", do not have an identified ligand yet. NRs regulate a variety of crucial biological processes including cell proliferation, development, metabolism, and reproduction (Sever and Glass 2013; Walker).

All NRs share a common domain organization: an N-terminal AD, a DBD, the hinge region and the Ligand Binding Domain (LBD), Figure 2. The LBD domain is similar in structure between NRs and is responsible for ligand-mediated activity by binding to coregulators through the activation function 2 (AF-2) surface. The DBD, the most conserved domain between NRs, has two Zn fingers allowing it to bind to specific DNA response elements. The hinge region bridges the LBD and the DBD domains. It is highly flexible and often possesses a nuclear localization signal (NLS). The N-terminal AD has an important activation function in ligand independent protein-protein interactions of receptor, called activation function 1 (AF-1). Among all domains of NRs it is the least conserved in sequence composition and length and also the least characterized due to its ID nature (Huang et al. 2010; Brélivet et al. 2012; Rastinejad et al. 2013).

Figure 2. Scheme of the NR domain architecture

AD – Activation Domain, DBD – DNA binding domain, H - hinge region, LBD – Ligand Binding Domain

1.2.1 Nuclear receptors in disease

NRs serve as master switches of complex gene regulatory networks which are crucial in many aspects of human physiology and, if misregulated, pathology. The receptors themselves are often found to be the main drivers of various diseases, in particular metabolic disorders and many cancers. Consequently, they serve as biomarkers for tumor subclassification and targets for therapy (Khan and Lingrel 2010).

In addition to playing pivotal roles in cancer, NRs natural ability to bind ligands makes them good candidates for drug design. The biggest advancement in that area has been made in treatment of ER+ breast cancer, that relies on the estrogen receptor (ER) signaling pathway, and PC, that relies on the AR signaling pathway. For both receptors there are clinically used small molecules drugs available. Few other NRs, such as the glucocorticoid receptor (GR), the progesterone receptor (PR), retinoic acid receptors (RARs) and retinoid X receptors (RXRs) have also been extensively pursued as drug targets in cancer, in some cases resulting in further marketed drugs (Lambert et al. 2018; Cheng et al. 2019; Zhao et al. 2019). In 2006 NRs constituted 13 % of all FDA approved drugs (Overington et al. 2006).

Despite the notion that for many cancers one receptor is the main driver of tumor development and progression, it has been recognized that the interplay between receptors is of undeniable importance. For example, in breast cancer it has been long known that the presence or absence of the ER and PR, along with tyrosine-protein kinase HER2, can be used for tumor classification and prognosis of patient survival (Chan et al. 2015). Nevertheless, the full spectrum of NRs cooperation or antagonism in cell type, tissue type and cancer type specific manner is being appreciated only now thanks to the development of genome-wide techniques. If we can tackle the complex networks of NRs in cancers, we can surely design better personalized therapies, Figure 3, (Dhiman et al. 2018).

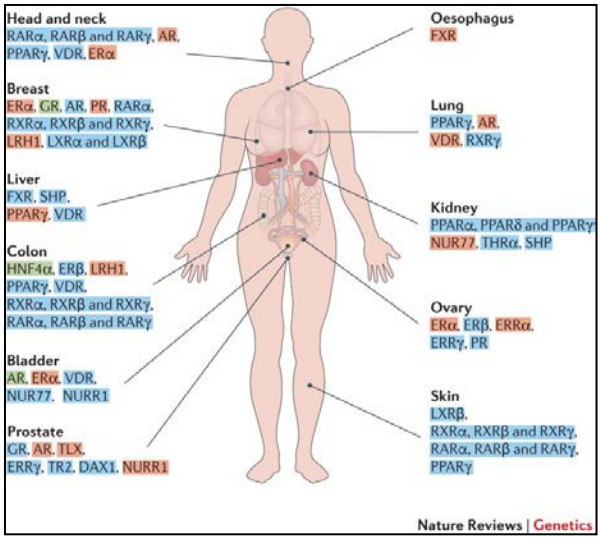


Figure 3. Crosstalk between NRs in oncogenesis

Tissue-specific expression of NRs that have been reported to play oncogenic (red), tumor-suppressive (blue) or both (green) roles in oncogenesis.
From (Dhiman et al. 2018).

1.3 Androgen receptor

AR belongs to a superfamily of steroid NRs (SRs). Other members of the family include aforementioned ER, GR, PR, thyroid receptor (,R) and mineralocorticoid receptor (MR) (McEwan 2009). The human AR is encoded in the long arm of the X chromosome (Xq11–12) (Brown et al. 1989). Expression from 8 exons results in a protein of 920 amino acids (aa)[3], although its precise length depends on the length of the two polymorphic regions in the N-terminal AD of the protein, the poly-glutamine (polyQ) and the poly-glycine (polyG). AR is expressed in most tissues. In some of the tissues its transcription is cell-type and age-specific. The primary function of AR is the development and maintenance of the male sexual phenotype (Gelmann 2002).

1.3.1 Domains

AR has the typical domain architecture of NRs: the N-terminal AD, residues 1-559, the DBD, residues 560-623, the hinge region, residues 624-670 and the LBD, residues 671-920 (the aa numbers vary depending on the protein length).

1.3.1.1 The N-terminal activation domain

As in the case of all SRs the N-terminal AD of the AR harbors the AF-1, but unlike in other SRs, the AF-1 of AR, not the AF-2, is the main transactivation function of the receptor (Jenster 1995). As mentioned in section 1.2, the AD is the least conserved domain between the SRs with only 15% homology. Even between the human and the rat AD of AR there is only 20% aa identity (Gelmann 2002). It is the largest (60% of the AR) and the least characterized domain of the protein. The lack of structural knowledge comes from its ID character and large size (almost 600 aa) that slowed down its characterization by NMR (Bain et al. 2007), Figure 4. Despite this absence of the structure-function description of the AD, deletion studies successfully identified the functional regions and linear motifs within the domain.

[3] Uniprot: P10275

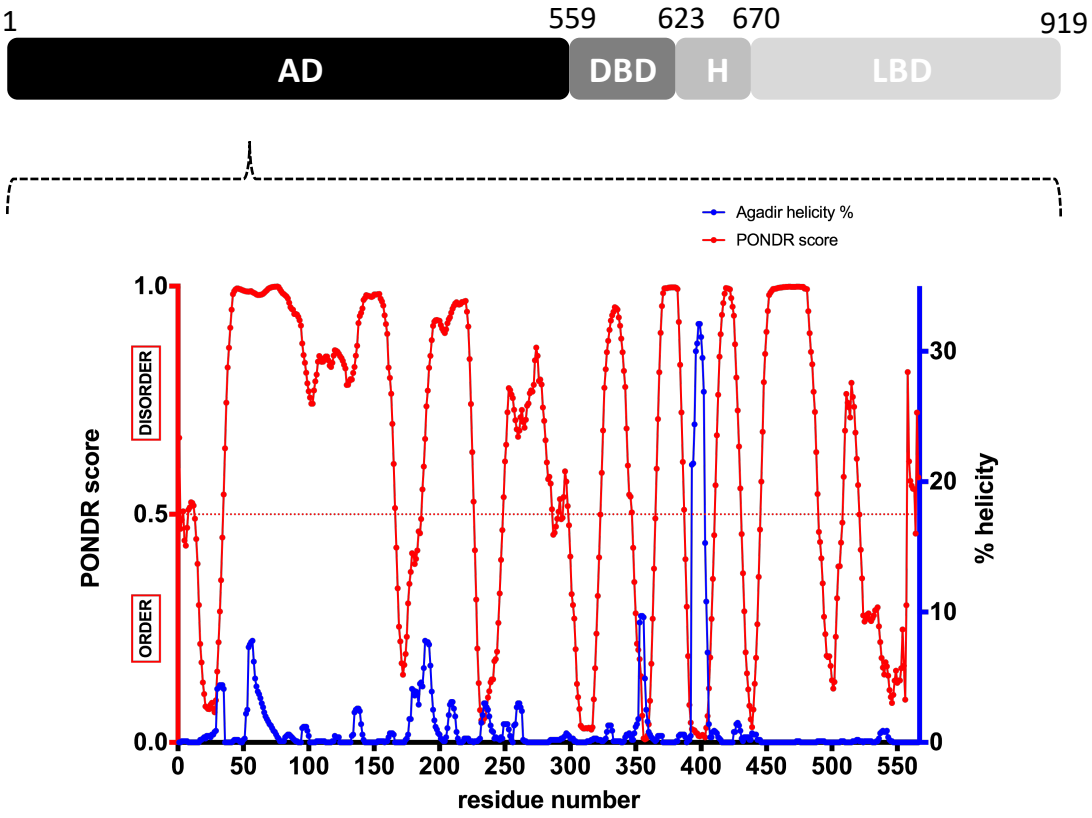


Figure 4. ID nature of the AR AD

In red – disorder prediction by PONDR[4]; In blue – helicity prediction by Agadir[5]

One of the linear motifs in the domain is in its N-terminal part. Located between residues 23 and 27 - ^{23}FQNLF^{27} motif - has been shown to be the mediator of the N/C interaction of the protein that is necessary for the androgen-dependent activation of AR (He, Gampe, et al. 2004). Upon androgen binding, the LBD undergoes a conformational rearrangement that results in the formation of a conserved protein-protein interaction surface (AF-2) and the dimerization of the protein. In most SRs AF-2 is the main activation function of the protein and binds to co-activators through their LXXLL-like motifs (Centenera et al. 2008). In the case of AR, the ^{23}FQNLF^{27} evolved to have a higher affinity for the AF-2 therefore being the main interactor of the AF-2 surface on the LBD of the same protein molecule or its dimer partner (He, Gampe, et al. 2004; Hur et al. 2004; van Royen et al. 2012a). However, the N/C interaction has been shown to occur after deletion of the motif (Schaufele et al. 2005), possibly due to binding through additional linear motifs in the AD (He et al. 2000).

AF-1 has been formally dissected into two distinct regions important for the activity of the protein in two distinct scenarios. In the presence of hormone, the activity of the protein relies on the transcription activation unit 1 (Tau-1) located between residues 102 and 371. In the absence or low levels of the hormone, the transactivation potential shifts to

[4] http://www.pondr.com
[5] http://agadir.crg.es

the transcription activation unit 5 (Tau-5) mapped to residues 361 and 537 (Jenster 1995). The core motif in the Tau-1 has been further narrowed down to a predicted helical region in the domain, residues 174-204, with a particular importance of a linear motif ^{179}LKDIL^{183}. The core Tau-1 has been shown to contribute to the N/C interaction of the protein and binding to MED1, a subunit of the Mediator complex (Callewaert et al. 2006; Jin et al. 2012). Another linear motif in Tau-1, conserved among SRs, is the ^{183}LSEASTMQLL^{192}. This Lx7LL motif serves as a binding site for TAB2, a component of the NCoR complex (Zhu et al. 2006). A third, highly conserved among species, motif in the c-terminal part of the Tau-1 ^{234}AKELCKAVSVSMGL^{247} has been shown to bind to the Hsp70-interacting protein E3 ligase CHIP (He, Bai, Andrew T Hnat, et al. 2004). Equivalently, a core motif in Tau-5 has been assigned to a ^{433}WHTLF^{437} motif. It has been shown to be the main mediator of the ligand-independent, but not the ligand-dependent, function of the protein (Scott M Dehm et al. 2007). The motif has been also shown to contribute to the N/C interaction (He et al. 2000) and to interact with histone acetyltransferase p300 in a manner dependent on the N/C interaction (Lagarde et al. 2012).

There are two amino-acid repeats in the AD. The polyQ tract, starting in the position 59, results from a polymorphic trinucleotide repeat region, (CAG)n, in the gene. The length of the polyQ in healthy individuals can vary between 6 to 39. The deviation from this length has been associated with a higher activity of the protein if shortened (increased PC risk) (Giovannucci et al. 1997) and higher aggregation propensity leading to Kennedy's disease, when above a threshold of 37 (La Spada et al. 1992; Bingham et al. 1995; Li et al. 1998). Recently, our group has shown a correlation between the length of the polyQ tract and its helical content for the lengths of the tract below the disease threshold. Longer polyQ tracts showed increased helical propensity, that could not be predicted just form amino acid sequence using predictors like Agadir[6] (Escobedo et al. 2019). The second amino-acid repeat region is a flexible polyG tract, located close to the DBD, starting at the position 449 and of average length of 22–24 glycines that accounts for about 90% of normal AR alleles (Ding et al. 2005).

1.3.1.2 The DNA binding domain

The DBD is the most conserved domain across NRs. It binds as a dimer to androgen response elements (AREs) on the DNA in the promoter and enhancer regions of androgen-related genes (Claessens et al. 2008). There are two main types of AREs, historically speaking, and variety of imperfect AREs discovered more recently thanks to chromatin immunoprecipitation assays. The classical ARE (clARE) is an inverted repeat of 5'-TGTTCT-3' sequence separated by a three-nucleotide spacer (Cato et al. 1987). The sequence of this composition is not specific to AR, it has been shown to be bound also by GR, PR and MR. The second main type of ARE is nevertheless specific to AR. The selective AREs (selARE) consist of partial direct repeats rather than inverted repeats of the same 5'-TGTTCT-3' motif and are bound by AR in a surprising head-to-head orientation (Shaffer et al. 2004), Figure 5. A variety of additional binding sites for AR have been found in studies focused on castration resistant prostate cancer (CRPC) context (Jin et al. 2014; Pomerantz et al. 2015; Wilson et al. 2016).

[6] http://agadir.crg.es

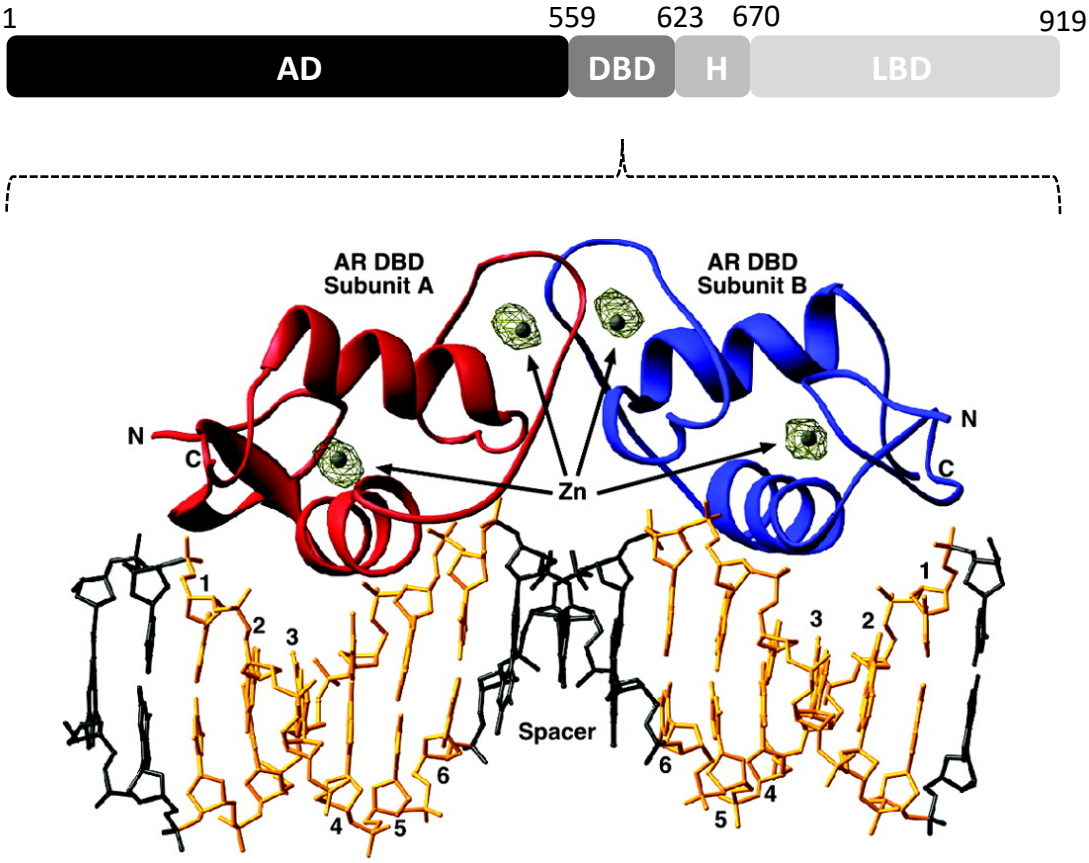


Figure 5. The AR DBD–ADR3 complex

The two DBD monomers are in red and blue, the hexameric half-site DNA is gold, and the spacer and flanking base pairs are black.
From (Shaffer et al. 2004)

The DBD is organized in two Zn fingers. An α-helix in the first Zn finger enters the major groove of the DNA and the P-box residues make interactions with the DNA to ensure selectivity for the ARE. The second Zn finger contains so called D-box residues, responsible for DNA-dependent receptor dimerization (Shaffer et al. 2004; Claessens et al. 2008), Figure 5.

1.3.1.3 The hinge region

The hinge region is poorly conserved among NRs, although it always contains the nuclear localization signal (NLS) (Evans 1988). The hinge region likely serves as a flexible linker between the DBD and the LBD, although few functionalities of the region have emerged alongside the nuclear import and export, like DNA selectivity and affinity, and transactivation potential of the AR (Haelens et al. 2007).

1.3.1.4 The ligand binding domain

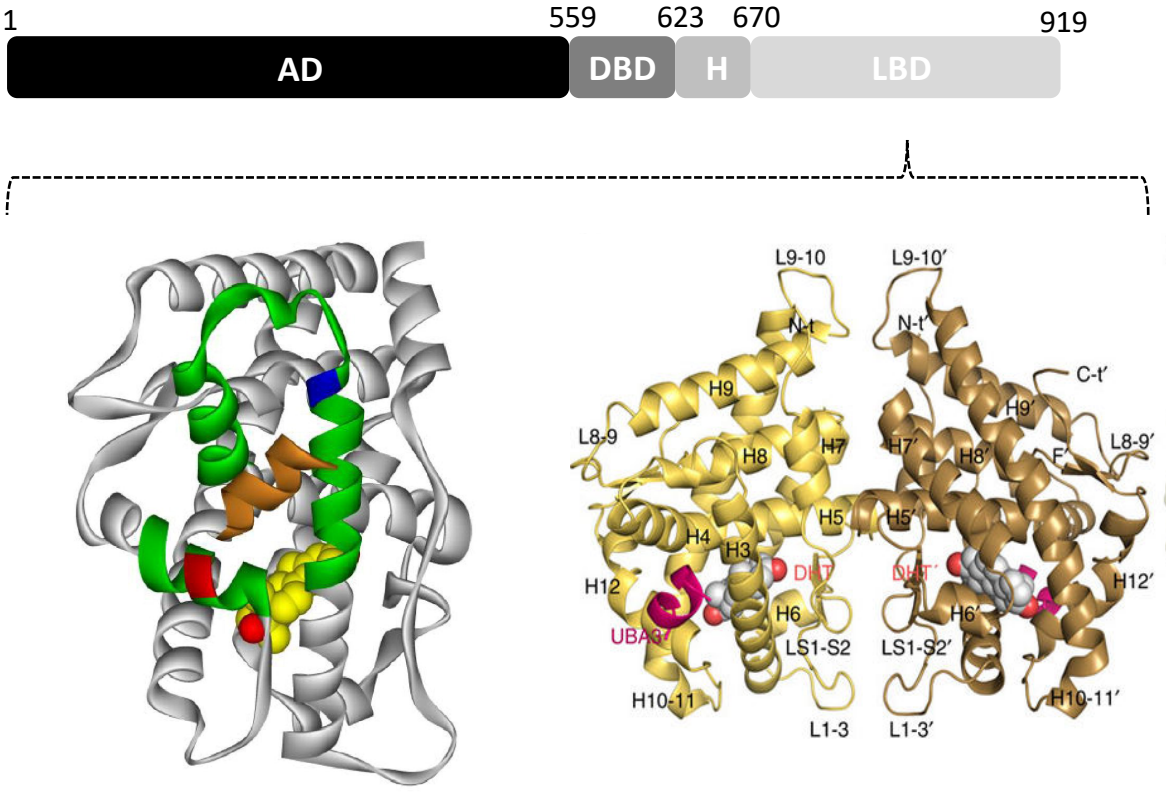


Figure 6. The AR LBD structures

Example crystallographic structures of the AR LBD. On the left: structure of the AR FxxLF peptide (brown ribbon) and R1881 (space filled atoms; yellow, carbon; red, oxygen) bound to AR LBD (white ribbon) with helices 3, 4, 5 and 12 (green ribbon) forming the coactivator binding site. Conserved charged residues (blue, positively charged lysine; red, negatively charged glutamic acid) at the opposite ends of the groove are indicated, from (van de Wijngaart et al. 2012). On the right: structure of the AR LBD core dimer. The two monomers are depicted as cartoons, with monomer B (yellow) in standard orientation and monomer C in brown; helices and loops are marked. The hormone (dihydrotestosterone, DHT) and the UBA3 peptide are shown as spheres and as a cartoon, respectively. From (Nadal et al. 2017).

The LBD is poorly conserved in sequence between various NRs, nevertheless the overall structure of the domain is preserved and consists of 12 α-helices, although the AR LBD lacks helix 2. The activation of the NRs is achieved through conformational rearrangement in the LBD upon ligand binding, where previously distant helix 12 folds back onto the ligand cavity in the LBD forming the AF2 surface (Bourguet et al. 2000). The AF2 hydrophobic cleft binds to the LXXLL motifs of co-activators, like the family of p160 (SRC-1, SRC-2 and SRC-3) (Parker et al. 1997), or, in the case of AR, to the N-terminal ^{23}FQNLF^{27} helix, forming the N/C interaction (Dubbink et al. 2004; Hur et al. 2004). There is no structural information available for the apo state of the AR LBD (with no ligand), nevertheless it is assumed to be similar between all the NRs. The hormone-bound LBD of AR has been successfully crystallized in the monomer and dimer states, Figure 6 (Pedro M Matias et al. 2000; Hur et al. 2004; Nadal et al. 2017). AF2 is not the only interaction site of the domain, binding to the androgen to the AR LBD exposes an additional surface called Binding Function 3 (BF3) that has been proposed to regulate the binding of co-regulators to the AF2 (Buzón et al. 2012). Due to the occupation of the AF2

by the N-terminal ^{23}FQNLF^{27} motif, the main activation function of the protein is taken by the AD. However, there is some evidence the N/C interaction is lost upon DNA binding making the AF2 available for co-activators binding (van Royen et al. 2007).

1.3.2 Mechanism of action

AR is present in the cytoplasm in a complex with various molecular chaperones, like Hsp90, Hsp70, Hsp40. Binding of the ligand, testosterone or dihydrotestosterone, causes conformational change in the protein, involving an intra- or intermolecular N/C interaction (binding of the ^{23}FQNLF^{27} helix of the N-terminal AD domain to the C-terminal LBD) and the dimerization of the protein. As a consequence, there is subsequent dissociation of chaperones, exposure of the nuclear localization signal (NLS) and thereupon translocation to the nucleus, Figure 7 (Tan et al. 2015; Eftekharzadeh et al. 2019).

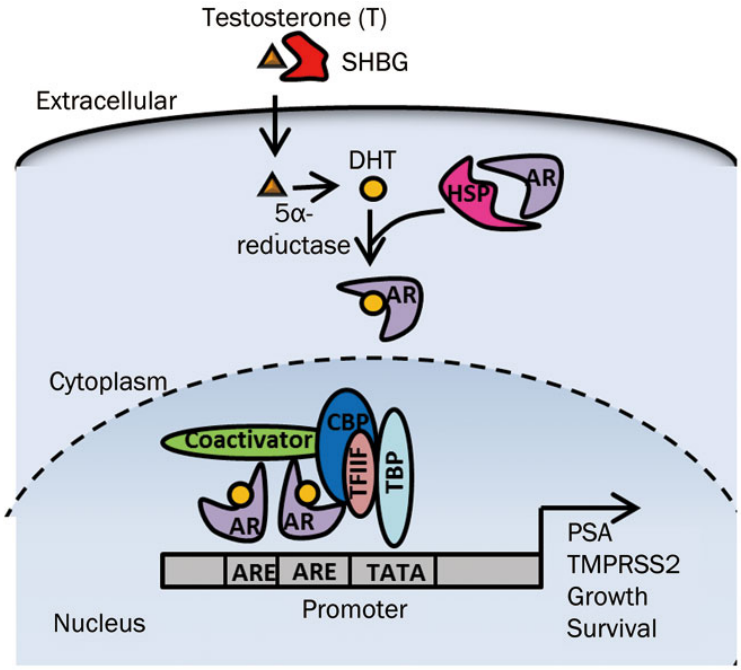


Figure 7. AR mechanism of action

Testosterone is transported to target tissues such as the prostate where it gets converted into dihydrotestosterone (DHT) by 5-α-reductase. DHT binds to the LBD pocket and promotes the dissociation of heat-shock proteins (HSPs) from the AR. The AR translocates into the nucleus, dimerizes and binds to the AREs in the promoter or enhancer region of a target gene. Once bound, the AR is able to recruit members of the basal transcription machinery and coregulators.
From (Tan et al. 2015)

1.3.2.1 Dynamics

In the cell nuclei AR binds to AREs in the promoters and enhancers of the target genes, where it recruits other components necessary to execute its transcriptional program (Shang et al. 2002). These events are followed by recruitment of general transcription factors (GTFs) and assembly of the transcription preinitiation complex (PIC), initiation of mRNA synthesis by RNA Polymerase II (Pol II), elongation of the transcript and termination of the process. This complex process is inherently dynamic and executed by

the orchestrated formation of large protein complexes and structural rearrangements of chromatin (Hager et al. 2009). AR, like other TFs, has been shown to diffuse freely through the cell nucleus, probing chromatin by transient interactions until it binds to the specific DNA site. The residence time of the TFs on the chromatin has been shown to last from minutes to hours, by biochemistry and ChIP-seq experiments, and milliseconds to seconds, by fluorescence microscopy techniques, like fluorescence correlation microscopy (FCS), fluorescence recovery after photobleaching (FRAP) and single molecule tracking (SMT). These two models are mutually exclusive, with the latter being far more convincing. Using SMT, AR has been shown to exist in three dynamic states interpreted as: un-bound, binding to non-specific sites on chromatin (fast component), and to specific response elements (slow component), with relative percentage of the hormone-activated molecules in each state equal to: 54.4 %, 39.4 % and 6.2 %. The residence times of AR within the fast and slow components were calculated to last for 1.08 ± 0.02 s and 7.02 ± 0.31 s (Paakinaho et al. 2017). It has to be pointed out that longer residence times have been reported for TFs within transcription domains characterized by high concentration of active Pol II in foci, with extended tail of longer times that fell to zero by 15 s for TF p53 and by 25 s for TF GR (Morisaki et al. 2014), making it plausible that the AR has longer residence times within transcription domains as well.

1.3.2.2 Oligomeric state

AR has been shown to be dimeric by means of crystallography (DBD and LBD) (Shaffer et al. 2004; Nadal et al. 2017) and other methods like bimolecular fluorescence complementation (BiFC) and bioluminescence resonance energy transfer (BRET) (Xu et al. 2015) and FRET (van Royen et al. 2012). However, it a study using the Number and Brightness method (N&B) it was recently possible to observe AR in higher oligomeric state, even up to octamer (Presman et al. 2016). In the nucleus fluorescently-tagged AR has been observed to form foci, possibly of higher oligomeric states of protein (Saitoh et al. 2002; Marcelli et al. 2006; van Royen et al. 2007). Formation of higher oligomeric species, AR “granules”, in the cytoplasm has been reported for nuclear translocation deficient mutants (Kumar and Tyagi 2012).

1.3.2.3 Direct interactions with transcription preinitiation complex

Once the AR is bound to DNA, it recruits, if agonist bound, coactivator or, if antagonist bound, corepressor proteins (Shang et al. 2002). Nevertheless, it has been shown that AR can interact directly with some components of the PIC. Formation of a functional PIC requires association of Pol II at promoters with the general transcription factors (GTFs), from which the core ones are: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Jishage et al. 2012). Biochemical studies demonstrated an interaction between the AD of AR and the large subunit of TFIIF, RNA polymerase II-associated protein 74 (RAP74) (Reid, Kelly, et al. 2002; Kumar, Betney, Li, E Brad Thompson, et al. 2004). In addition, TFIIH has been shown to interact with the AR AD by co-immunoprecipitation (Lee et al. 2000). TFIIH possesses kinase activity towards DNA-directed RNA polymerase II subunit (RPB1) C-terminal domain (CTD) that is required for promoter clearance (Dvir et al. 2001). Interestingly, AR has been shown to interact also with another RPB1 CTD phosphorylating component of PIC, p-TEFb (Lee et al. 2001). Activation of the P-TEFb complex is important for Pol II-Ser2p, release of paused Pol II, and gene transcription (Reines et al. 2013). Recently, it has been shown that in the case of CRPC AR is able to activate p-TEFb

by an additional mechanism, which is upregulation of enhancer RNA (eRNA) transcription. eRNAs bind to CYCLIN T1, resulting in activation of p-TEFb through its release from an inhibitory complex (Zhao et al. 2016). Finally, AR interacts directly with the RNA Pol II through association with its second largest subunit, RPB2 (Lee et al. 2003).

1.3.2.3.1 Interaction with TFIIF

TFIIF is required for stable Pol II complex formation. About 50% of Pol II is found associated with TFIIF in yeast (Rani et al. 2004). TFIIF is required for transcription initiation from TATA-containing and TATA-less promoters (Burton et al. 1988). It has been shown to contribute to the selectivity of Pol II for the DNA and start selection (Killeen and Greenblatt 1992; Ghazy et al. 2004). In the initially transcribing complex, TFIIF stabilizes the DNA-RNA hybrid and in the elongation phase reduces Pol II pausing, suppresses backtracking and RNA cleavage induced by TFIIS (Elmendorf et al. 2001; Funk et al. 2002; Zhang et al. 2003; Bengal et al. 2015).

TFIIF contains two subunits that are conserved among human, insects and yeast, named Rap74 and Rap30 in humans. The N termini of the two subunits form a dimerization domain and the C termini are winged helix (WH) domains, common DNA binding domains in Eukaryotes (Woychik and Hampsey 2002). The structures of the dimerization domain and both WH domains are available (Gaiser et al. 2000; Groft et al. 2002; Kamada et al. 2002). Biochemical and structural studies have mapped the regions of TFIIF implicated in nonspecific DNA interactions and protein-protein integrations with Pol II, TFIIB and FCP1 phosphatase (Funk et al. 2002; Chung et al. 2003; Kamada et al. 2003; Nguyen et al. 2003; Chen et al. 2010).

The cryo-EM structures of the yeast and human PIC revealed the position of the TFIIF within the complex. The dimerization domain of TFIIF was mapped onto the lob domain of the Pol II and the Rap30 WH onto the DNA, precisely the TFIIB recognition element downstream from the TATA-box. The stabilization of the downstream DNA along the cleft of Pol II via the Rap30 suggests that Rap30 rather than RAP74, is required for accurate transcription initiation. Rap74, or Tgf1 in yeast, couldn't be mapped onto structure due to its flexibility in the complex context (He et al. 2013; Murakami et al. 2013).

In vitro binding studies identified TFIIF (and specifically Rap74) as a direct interactor of the AR AD (amino acids 142–485). (McEwan and Gustafsson 2002). The interaction was also observed in a cell based study, where TFIIF was co-immunoprecipitated with the AR (Paliouras et al. 2011). The molecular details of the TFIIF and the AR AD interaction were further elaborated by the group of McEwan, that discovered the interaction in the first place. The precise binding side was mapped to the N and C termini of the Rap74 subunit of TFIIF, with the C terminal fragment (Rap74-CTD) being the main interaction site (Reid, Murray, et al. 2002a). Additional mutational analysis revealed the residues V490, L493 and L497 in helix 3 of the WH motif of the RAP74-CTD to be crucial for binding to the AR 143-494 fragment (Lavery and McEwan 2008a). On the side of the AR AD, residues M245, L247, V249 in Tau-1 and the surrounding residues were proposed to form the binding site for TFIF. In addition, residues S160 and S163 were postulated to have an indirect effect on the binding by altering the structural flexibility of the AR AD (Reid, Murray, et al. 2002b; Betney and McEwan 2003). The kinetics of the interaction was determined to be in the submicromolar range (Lavery and McEwan 2008b). Finally, a gain in helical content in the AF1 upon binding to Rap74 was shown by Fourier

transform infrared (FTIR) spectroscopy, consistent with a folding upon binding mechanism of IDPs (Kumar, Betney, Li, E. Brad Thompson, et al. 2004).

1.3.2.4 Coregulators

Almost 170 AR coregulators have been identified (Heemers and Tindall 2007a). It is important to realize that many of those were determined using yeast-two-hybrid screens with AR domains as a bait. As a consequence, for many, there is no data on their actual ability to influence receptor function (Chmelar et al. 2007). AR coregulators typically do not possess DNA binding capacity and do not significantly alter the basal transcriptional rate. They are recruited by AR to enhancer and promoter regions of the gene to enhance (coactivators) or reduce (corepressors) its transactivation. The means by which they operate include: chromatin remodeling, recruitment of GTFs and direct modulation of the AR function, for example by assuring proper AR folding or subcellular localization (Heemers and Tindall 2007a).

The function of chromatin remodeling complexes is to alter the DNA-histones interactions resulting in more accessibility of chromatin for transcription. Some of the examples shown to interact with AR involve ARIP4, a nuclear ATPase that belongs to the SNF2-like family of chromatin remodeling proteins (Rouleau et al. 2002) or BRG1 and hBRM, two core components of the SWI/SNF complex required for nucleosome repositioning (Marshall et al. 2003). In parallel with chromatin remodeling, histone modifications, usually affecting histone charge and altering its interaction with DNA, provide another level of transcriptional control. Several histone acetyltransferases (HATs), associated with transcriptional activation, and histone deacetylases (HDACs), associated with transcription repression, have been shown to interact with AR and modulate its activity (Heemers and Tindall 2007a). Examples of AR coactivators include SRC-1 and SRC-3, members of LXXLL motif-containing p160 SRC gene family. As mentioned in chapter 1.2.2.1., the LXXLL motifs of NRs' coregulators interact predominantly with AF2 of NRs, but in the AR the primary site of interaction is taken over by the AF1 in the AR AD. SRC-1, SRC-3 and SRC-2 (Yeh, S. et al. , 1998; Hong et al. 2002), beside interacting with AR, recruit other coactivators, for example p300, the p300 homolog CBP, as well as p300/CBP-associated factor (P/CAF) (Fu et al. 2000). Furthermore, p300, CBP and p/CAF can also directly interact with AR. Several HDACs have been shown to negatively affect AR transactivation, via direct interaction or through recruitment by corepressor complexes like NCoR and silencing mediator of retinoid and thyroid receptors (SMRT) (Heemers and Tindall 2007b). NCoR and SMRT bind to the AR AF2 via LxxxIxxx(I/L) motifs. They interact strongly with AR in the antagonist (hydroxyflutamide and bicalutamide) bound state (Liao et al. 2003; Hodgson et al. 2007).

The prostate specific antigen (PSA), the best-characterized androgen-responsive gene in the prostate gland, has been used as a model system to understand the AR transcription complex assembly. The proximal promoter of PSA contains two putative AREs, ARE I and ARE II (Cleutjens et al. 1996). In addition, the PSA enhancer element that is essential for full androgen response, is centered at approximately 4.2 kb and contains another ARE, ARE III, (Cleutjens et al. 2014). It has been shown that the formation of an activation complex involves recruitment of coregulators to both the promoter and enhancer regions, whereas the formation of a repression complex only involves coregulators bound at the promoter (Shang et al. 2002). The best studied components required for the AR

transcription complex assembly are aforementioned regulators of NRs including SRC-1, SRC-2, SRC-3, p300, p/CAF, BRG1, and not mentioned previously, MED1 (TRAP220) subunit of the Mediator complex and coactivator-associated arginine methyltransferase 1 (CARM-1) (Wang et al. 2002; Kang et al. 2004; Heemers and Tindall 2007a). The Mediator complex is a multi-subunit complex that plays an important role in AR, and other NRs, transcriptional activation. It has been shown that AR can interact with Mediator directly through its MED1 subunit and that this interaction is indispensable for AR mediated transcription (Wang et al. 2005; Jin et al. 2012). CARM-1, on the other hand, is a histone methyltransferase, which does not interact directly with the AR, but instead is recruited as a secondary coactivator mainly via interaction with SRC coactivators (Nef et al. 1999). A model has been proposed in which the AR and its coactivators, which are predominantly bound to the enhancer region, are able to physically contact the promoter region through a chromatin looping mechanism and as a consequence the Pol II, which is first recruited to the enhancer region, can track to the promotor region to initiate the transcription (Wang et al. 2005). The looping mechanism has been further confirmed and shown to be facilitated by bidirectional PSA (KLK3) enhancer RNA (KLK3e) (Hsieh et al. 2014). It is not clear whether the studies on the PSA hold true in the case of other AR-regulated genes.

There is no structural information on the architecture of the AR transcription complex. Presumably, it is also a simplification to call any form of AR transcription complex, the transcription complex. As mentioned previously, transcription is an inherently dynamic process, that requires subsequent exchange of cofactors. Any structure of a complex would capture a set of proteins in a particular conformation. Additionally, the structural biology tools, crystallography and cryoelectron microscopy (cryo-EM), would lack information on the ID regions of the proteins, which are so prevalent in transcription. The most comprehensive, yet very limited, view of a NR transcription complex comes from a cryo-EM study on ER alpha bound to SRC-3 and p300. Remarkably, thanks to the use of a specific antibody, the authors were able to map the AD of ER alpha on the structure and propose its contribution to the complex formation (Yi et al. 2015). Unfortunately, this structure cannot be directly translated into the AR transcription complex, due to the shift of the interaction dependence from the AF2 to the AF1 in the case of AR, which leaves the structural features of the AR transcription complex elusive.

Mediator is a coregulator worthy of attention, because its required for almost all transcription of almost all Pol II promoters (Kornberg 2007). It is a large evolutionarily conserved multisubunit protein complex (25-30 subunits in humans, >1 MDa) with functions in transcription steps ranging from chromatin remodeling to subsequent PIC formation and function (Chen and Roeder 2011). Its primary role is a functional bridge between TFs, like NRs, and the basal transcriptional machinery, including Pol II and GTFs, Figure 8, (Soutourina 2018). Mediator subunits are organized into three core modules (Head, Middle, and Tail) and a dissociable CDK8 kinase module. The organization doesn't prevent structural flexibility in the complex. The cryo-EM studies of the Mediator show rearrangements depending on the functional state of the Mediator, for example TF or PIC binding (Bernecky and Taatjes 2012; Tsai et al. 2014; Plaschka et al. 2015). The structural flexibility is also evident in its predicted high ID content (Tóth-Petróczy et al. 2008). From all the subunits, MED1, has been shown to be the main interaction partner of NRs. MED1 binds to the AF2 of NRs in a ligand dependent manner through its LXXLL motif. Although this interaction has proven to be not essential for NRs basic function, suggesting existence of alternative binding sites on MED1, other subunits of Mediator or other coactivators (Chen and Roeder 2011). In the case of AR, where AF2 is (probably) occupied by the

^{23}FQNLF^{27} motif, another sites were shown to be responsible for this interaction, Tau-1 in the AR AD and two noncanonical α-helical motifs located in the MED1 (Jin et al. 2012). Interestingly, MED1 has been shown to be important for enhanced UBE2C locus looping in the AR-negative and AR-positive CRPC cells. The enhancer-promotor looping leads to UBE2C oncogene expression and consequently, cell growth (Chen et al. 2011). Moreover, Mediator has been shown to be crucial for the AR v567es, an AR splice variant lacking the LBD domain, induced UBE2C oncogene up-regulation and subsequent PC cell growth (Liu, Sprenger, Wu, et al. 2015). Importance of the Mediator is also highlighted by the MED1 and MED17 overexpression in 50% of PC. Lowering their levels in cancer cells inhibits proliferation, slows the cell cycle and induces apoptosis (Vijayvargia et al. 2007). Recently, it has been shown that MED1 together with Bromodomain-containing protein 4 (BRD4) can form nuclear puncta at super enhancer regions in the nucleus and that these puncta exhibit properties of liquid-like condensates. These condensates were proposed to be active centers of transcription (Hnisz et al. 2017; Boija et al. 2018; Sabari et al. 2018; Nair et al. 2019). The details of these finding for the AR research will be discussed farther in the Chapter 1.4.2.

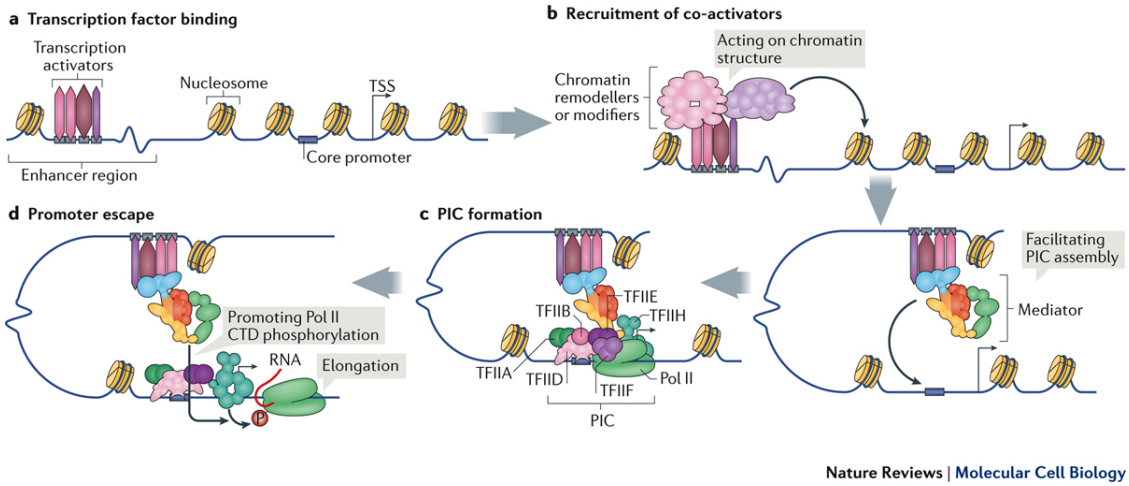


Figure 8. A simplified model for the main steps of transcription initiation by Pol II in a chromatin context

a) Transcription activation starts with the binding of TFs (here activators) on enhancer regions. The enhancer regions are located at different distances from the core promoters. The transcription start site (TSS) is indicated by an arrow. b) Activators recruit co-activator complexes that act as chromatin modifiers or remodellers to alter chromatin structure and to make it more accessible for other factors. Other co-activators are then recruited that act directly on the assembly of basal transcriptional machinery, the PIC. Mediator of Pol II transcription is one of the key co-activator complexes. c) The PIC is assembled at the core promoter. It includes Pol II, general transcription factors: transcription initiation factor IIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (10 subunits, including 7 core subunits and a 3-subunit kinase module (TFIIK) containing cyclin-dependent kinase 7 (CDK7). Multiple steps and pathways could be involved in PIC assembly *in vivo*, and Mediator acts to facilitate recruitment and/or stability of different PIC components. d) CDK7 phosphorylates (P) the carboxy-terminal domain (CTD) of the largest Pol II subunit at Ser5, which is necessary for Pol II to escape from the promoter and for the transition from the initiation step to the elongation step. This phosphorylation is also regulated by Mediator.
Form (Soutourina 2018)

1.3.2.4.1 Interaction between androgen receptor and Speckle-type POZ protein

Speckle-type POZ protein (SPOP), as the name suggests, has been identified as a BTB/POZ (poxvirus and zinc finger) domain containing component of the nuclear speckles (Nagai et al. 1997). SPOP is a substrate adaptor of a cullin-3-RING ubiquitin ligase (CRL3), the subclass of the largest group of CRLs. Once recruited by SPOP, substrates undergo ubiquitination and subsequent degradation by the proteasome system (Hernandez-Munoz et al. 2005; Zhuang et al. 2009). The ubiquitin (Ub)–proteasome pathway (UPP) constitutes a primary mechanism for regulating protein turnover and consequently enabling homeostasis of the cell (Lecker et al. 2006).

SPOP contains three domains: a MATH (meprin and traf homology) domain responsible for substrate recruitment, a BTB (bric à brac, tramtrack, broad complex) dimerization domain and a BACK (BTB and C-terminal Kelch) domain, another oligomerization domain known to form dimers, tetramers, or even pentamers (Zhuang et al. 2009; Errington et al. 2012; Van Geersdaele et al. 2013). Recently it has been shown that SPOP BTB dimers serve as building blocks for the isodesmic self-association into higher-order structures through BACK domain of the protein. This higher-order oligomerization enhanced substrate ubiquitination *in vitro* and was responsible for SPOP localization into nuclear speckles in cells (Marzahn et al. 2016).

Genome and exome sequencing of PC samples from primary or metastatic cancers allowed for identification of frequent somatic mutations (Berger et al. 2011; Grasso et al. 2012). One of the identified genes is SPOP, with a substrate binding site mutated in 6-15% of tumors tested in one of the studies but lacking many of the typical mutations, making a case for a separate class of prostate tumors (Barbieri et al. 2012; Boysen et al. 2015). PC mutants of SPOP have been shown to diminish ubiquitination of substrates and their subsequent up-regulation (Geng et al. 2013; Theurillat et al. 2014). Additionally, SPOP downregulation has been suggested to play a role in tumor progression (Kim et al. 2013).

Two SPOP-binding consensus motifs, Φ- π -S-S/T-S/T (Φ: nonpolar residues, π: polar residues), were identified in the AR hinge and the AD regions. The binding motif in the hinge region is a perfectly matched sequence ^{203}EGSSS^{207}, that was shown to be the main interaction site. The study proved the SPOP-dependent AR ubiquitination and subsequent degradation. Moreover, SPOP-mediated degradation caused inhibition of the AR transcriptional activity and consequent inhibition of PC cells proliferation. On the contrary, the PC mutants of SPOP were unable to bind AR and promote its degradation (An et al. 2014).

1.3.3 Androgen receptor related diseases

AR plays a pivotal role in few disorders including: androgen-insensitivity syndrome, spinal and bulbar muscular atrophy (SBMA), also known as Kennedy's disease, benign prostatic hyperplasia, hypogonadism and PC (Shukla et al. 2016). For the purpose of this thesis, due to the focus of the research group, I am going to focus on SBMA and PC.

SBMA is one of the nine polyglutamine disorders characterized by protein aggregation caused by a polymorphic polyglutamine (polyQ) tract expansion. The variable length of the polyQ results from the propensity of the CAG and GTC codon repeats (coding for Q amino acid) to form non classical B-DNA structures that cause DNA polymerase slippage

during replication (Mirkin 2007). In the case of SBMA patients polyQ lengths from 38 to 62 have been observed, in contrast to 9 to 36 repeats in the healthy individuals, with the threshold for the disease onset between 35 to 40 repeats (Fischbeck et al. 1991; La Spada et al. 1992; Orr 2012; Fratta et al. 2014). The expanded polyQ protein aggregates in muscles and motor neurons and forms cytoplasmic and nuclear inclusions, with the later correlating with toxicity (Adachi et al. 2005). SBMA has an estimated prevalence of 1 in 300,000 males. It only occurs in individuals of European or Asian racial background and has yet not been reported in males of African or aboriginal racial background (La Spada 2017). SBMA is characterized by degeneration of muscle tissue and motor neurons that can cause dysarthria, dysphagia, wasting and fasciculation of the tongue, weakness of the proximal muscles, and absence of tendon reflexes. It is accompanied with androgen insensitivity, causing breast enlargement, reduced fertility, and testicular atrophy. Although the quality of life of SBMA patients is lowered, the lifetime expectancy is not affected (Rhodes et al. 2012).

PC was the fourth most diagnosed cancer in Europe in 2018. In men it was the most common primary site of cancer (21.8% of the total cases), followed by lung (15.1%), colorectum (13.2%) and bladder (7.5%). The detection of PC has substantially increased in higher income countries in the early to mid 1990s due to more common use of the PSA testing. The increase of early detection due to the PSA testing could be responsible for the slight decline of the prostate cancer mortality rate in some European countries. However, the usefulness of the common use of the PSA testing has been called into question by urologists due to the high number of false positives.[7] PC remains the third most common cause of cancer death among men (constituting 10% of cases) (Ferlay et al. 2018).

AR activity is undoubtedly crucial to PC onset and progression (Huggins and Hodges 1941; Taylor et al. 2010). Consequently, preventing AR activity by androgen deprivation therapy (ADT) is the standard treatment for advanced or metastatic prostate cancer (mPC). The early stage PC, confined to the prostate capsule, can be treated by prostatectomy or radiotherapy. In the case of failure of above therapies, the disease progresses to the lethal castration resistant prostate cancer (CRPC). Even at this stage AR remains the main driver of the disease, but the patients do not respond to the ADT (Chan and Dehm 2014).

ADT is achieved by surgical castration (orchiectomy) or chemical castration by gonadotropin-releasing hormone (GnRH) agonists or GnRH antagonists. The outcome of both in the context of PC is comparable, nevertheless recently the comparison was expanded to address the adverse effects of the two approaches. The study concluded orchiectomy having significantly lower risks of experiencing any fractures, peripheral arterial disease and cardiac-related complications (Sun et al. 2016). The two methods reduce testosterone levels produced by testes, but not adrenal glands. To potentiate the therapy, ADT is often combined with a treatment with AR antagonists, molecules binding to the LBD but not activating the protein, achieving complete blockage of the testosterone activity on AR. Currently there are few nonsteroidal antiandrogens (NSAA) available, including flutamide, hydroxyflutamide, nilutamide, bicalutamide or more potent second-generation antiandrogens, enzalutamide and recently approved by the European Medical Agency apalutamide[8]. Additionally, abiraterone acetate (abiraterone), which targets a central enzyme in de novo steroidogenesis (Cytochrome P450 family 17 subfamily A

[7] https://uroweb.org/epad-2019-the-current-status-of-prostate-cancer-screening-in-eu/
[8] https://www.erleadahcp.com

polypeptide 1, CYP17A1), is used to block the testosterone production by adrenal glands. Despite a good response to the hormonal therapy for almost all patients, lasting from months to years, the disease uniformly progresses to the CRPC (Heinlein and Chang 2004; Watson et al. 2015; Fujii and Kagechika 2019).

The notion that CRPC relies on AR signaling axis came from an observation that 30% of CRPC patients that underwent ADT harbored genomic amplification of AR locus, whereas the amplification was not observed in the corresponding tumor samples collected before the treatment (Visakorpi et al. 1995). Now we also know that is not only the genomic amplification of AR locus, but also a recurrent tandem duplications involving an upstream enhancer of the AR, that drive the CRPC (Takeda et al. 2018; Viswanathan et al. 2018).

Since then, many mechanisms of AR signaling restoration and consequently, resistance to the therapy, have been described. One of these mechanisms is point mutations that occur mostly in the AR LBD. The most common ones are L702H, W742C, H875Y and T878A. Collectively, these recurrent AR mutants are present in 15–20% of CRPC cases (Beltran 2014; Robinson et al. 2015; Watson et al. 2015). The molecular mechanism by which T878A and H875Y mutants could be responsible for drug resistance came from *in vitro* studies where the mutants have been shown to be activated, rather than inhibited, by the anti-androgens nilutamide and flutamide (Valdscholte et al. 1990; Tan et al. 2014). In analogy, the W742C mutant has been shown to emerge after biclutamide treatment and to respond to the drug as to agonist (Hara et al. 2003). Nevertheless, T878A, L702H, H875Y mutants have been also found a priori to the treatment suggesting a broader effect these mutations convey (Azad et al. 2015; Chen et al. 2015). Another possible explanation is the shift of the AR to respond to other non-canonical steroid ligands such as adrenal androgens, estrogen and progesterone and in the case of L702H, by glucocorticoids (Zhao et al. 2000; Van De Wijngaart et al. 2010).

Another mechanism postulated to be implicated in CRPC progression and drug resistance is an alternative splicing of AR mRNA. Few AR splice variants (ARVs) have been detected in PC cell lines, xenografts and clinical specimens. A common feature of the ARVs is a truncation of the LBD and addition of a short sequence at the C-terminus of the protein. The most consistently expressed ARV in CRPC cell lines, tissue samples and pre-clinical PCa models of castration resistance, is AR-V7 (AR3) (Watson et al. 2015; Wadosky and Koochekpour 2017). AR-V7, despite the lack of a full NLS, is constitutively nuclear and transcriptionally active (Watson et al. 2010). The reports whether it is responsible for drug resistance clearly depend on the model system used. For example, 22Rv1 enzalutamide-resistant CRPC cell line expresses high levels of AR-V7 and siRNA-mediated knockdown of its expression restores the enzalutamide sensitivity (Li et al. 2013). On the contrary, VCaP cell line, characterized by lower expression of AR-V7, is still sensitive to androgen depletion or enzalutamide (Watson et al. 2010). Studies aimed at induction of castration or enzalutamide resistance by forced expression of AR-V7 were also dependent on the model used (Watson et al. 2015). Importantly, overexpression of AR-V7 or another ARV, AR^{v567es}, in the prostates of transgenic mice was sufficient to promote PC, but not sufficient to block castration-induced apoptosis and glandular involution (Sun et al. 2014; Liu, Sprenger, Sun, et al. 2015). The cell and mice model studies have been promptly complemented by multiple clinical investigations aiming at understanding the correlation between V7 (and AR^{v567es}) and clinical outcome in terms of resistance and prognosis. AR-V7 (often in combination with the PSA expression) has been proposed as a biomarker to predict CRPC patient response to therapy (Wadosky and Koochekpour 2017).

In 2017 there were 6 clinical trials aiming at that. An example trial[9], concluded that detection of AR-V7 in circulating tumor cells by two blood-based assays is independently associated with shorter progression free survival and overall survival with abiraterone or enzalutamide treatments and such men with metastatic CRPC should be offered alternative treatments to a standard ADT (Armstrong et al. 2019). The major caveat of that conclusion is the limited number of alternative treatments. Effective inhibition of ARVs action could be confer by development of drugs targeting either DBD (Dalal et al. 2018), or the AD of AR (an example is the terminated clinical trial of EPI-506, molecule targeting the AR AD[10]) (Antonarakis et al. 2016; Montgomery et al. 2019) or lowering the protein level of AR-V7 (currently there is one clinical trial testing the use of niclosamide in combination with abiraterone acetate and prednisone[11]) (Liu et al. 2014).

Castration reduces levels of testosterone by over 90%, but doesn't eliminate another of source of testosterone which is *de novo* steroidogenesis in the adrenal gland, keeping the testosterone level in the localized PC or metastatic CRPC on a physiologically relevant level (Titus et al. 2005; Nishiyama et al. 2015). Aforementioned, abiraterone acetate reduces the levels of necessary precursors for testosterone production, adrenal androgens dehydroepiandrosterone (DHEA) and androstenedione (AD), but the remaining pool of sulfated DHEA can be still converted to testosterone or DHT in prostate tissue (Attard et al. 2009; Tamae et al. 2015). As a consequence, there is ongoing further investigation on therapeutic blockage of enzymes downstream of CYP17A1 (Liu et al. 2014).

Another mechanism of CRPC continuous progression is a complete bypass of the AR signaling axis. Mentioned in Chapter 1.2.1. complex interplay between NRs, in specific GR, PR and MR, have been a subject of investigation in the context of CRPC (Watson et al. 2015). For example, an enzalutamide resistant LNCaP subline, LREX', was shown to be dependent on GR expression for enzalutamide resistant growth. Additionally, in VCaP cells analysis of AR and GR cistrome and transcriptome revealed high overlap between the two receptors (Arora et al. 2013). Studies using patients CRPC or localized PC samples supported the possible GR driven resistance. The potential of combined inhibition of GR and AR is currently being investigated[12] (Watson et al. 2015).

Further complications come from the molecular heterogenicity of the CRPC, not only between patients, but also within one tumor. The heterogenicity is already visible at the level of AR expression, with subpopulations of cells expressing from none, through moderate to high levels of AR (Crnalic et al. 2010). It is now clear that some men can relapse with clinically aggressive variants of CRPC with reduced or absent AR expression, although the classification of these tumor types is still ongoing and the pathways responsible for disease progression are being investigated (Beltran et al. 2011; Epstein et al. 2017).

The heterogeneity of PC is clearly visible at its most basic level through its genetic and epigenetic status, already in the primary tumors (Stelloo et al. 2019). One of the most common genomic PC alterations, TMPRSS2:ERG rearrangements, already reported in 2006, occurs in 49.2% of 118 primary prostate cancers and 41.2% of 18 hormone-naive

[9] Clinical trial: NCT02269982
[10] Clinical trial: NCT02606123
[11] Clinical trial: NCT02807805
[12] Clinical trial: NCT02012296

lymph node metastases (Perner et al. 2006). TMPRSS2 is an androgen-response gene and its promotor element in the context of the TMPRSS2:ERG fusion results in ERG oncoprotein (member of the E26 transformation-specific TFs) overexpression and subsequent contribution to PC progression (Z. Wang et al. 2017). Mentioned in Chapter 1.3.2.5., the growing availability of somatic genome or exome sequencing of cancers expands our understanding of the scope of the genomic alterations, including recurrent mutations in SPOP, MED12, FOXA1 and other factors in PC (Taylor et al. 2010; Berger et al. 2011; Barbieri et al. 2012; Robinson et al. 2015). The constantly growing data on cancers genomic is available at cBioPortal[13].

The chromatin regulatory landscape rearrangements through epigenetic changes constitute another level of PC complexity. It has been shown that AR cistrome undergoes extensive reprogramming during prostate epithelial transformation in men, with a core set of sites being consistently reprogrammed in tumors. Importantly, the reprogrammed cistrome relied on FOXA1 and HOXB13, which appeared to directly bind to the reprogrammed sites together with the AR and drive the transition (Pomerantz et al. 2015). Another study linked the AR overexpression with chromatin relaxation that was specific to CRPC. The increased chromatin accessibility in cancer cell lines was shown to be mediated by bromodomain-containing proteins (BRDs) (Urbanucci et al. 2017). Improved technological advances, methods like multicontact 4C and single-cell Hi-C, open the possibility to understand the epigenomic rearrangements of CRPC and allow for future validation for their clinical importance (Stelloo et al. 2019).

1.4 Phase separation in biology

The Protein Trinity hypothesis was introduced by Keith Dunker and Zoran Obradovic in order to put forward the idea of functional IDPs. The concept invoked the idea that the protein function can be carried out not only by its ordered state, but by any of its three native states: the ordered state, but also disordered, random coil and molten globule states (Dunker and Obradovic 2001). The authors elaborated on the idea suggesting the transitions between these states can be also functional and are conceptually analogous to phase transitions (Dunker et al. 2001). The authors probably did not predict the analogy will soon be used in a literal way to describe actual phase transitions of IDPs. In the landmark publication in 2009, Clifford Branwynne from Tony Hyman's laboratory applied the phase transition concept to explain the behavior of P granules in the germline of *Caenorhabditis Elegans* (Brangwynne et al. 2009). They were not the first ones to speculate about the importance of phase transitions in living systems or their origins (Oparin and Morgulis 1938), but the first ones to use modern tools to dissect the nature of the membranelles organelle, by physically dissecting a germline nucleus, Figure 9 (Brangwynne et al. 2009).

[13] http://www.cbioportal.org/

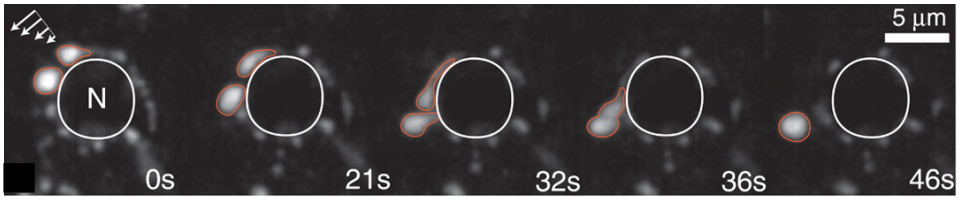


Figure 9. P granules behave like liquids

Dripping P granules (red outline) from a dissected germ line. Nucleus (N), white line.
Adapted from (Brangwynne et al. 2009)

Currently the phase separation is becoming a field on its own, where many of the known membranelles organelles, like nucleoli, Cajal bodies or stress granules are revisited by biologists with the concepts and tools borrowed from soft matter physics. The membranelles organelles formed by phase transitions vary in material properties, from liquid-like, through gel-like to solid-like assemblies, composition, formed by IDPs, folded proteins, nucleic acids and mixtures of all, and functions, for example: storage, enzymatic reactions, RNA processing or simply products of aberrant transition into solid-like aggregates. The most common and physiologically relevant phase transition appears to be the liquid-liquid phase separation (LLPS), where the resulting dense phase has properties of liquid droplets that are round, can drip, wet the surface, undergo fusion and whose components can easily rearrange.

A classic macroscopic example used to visualize the LLPS is the salad dressing, where the two liquids, oil and the vinegar, do not mix but separate into two distinct phases (Hyman et al. 2014; Mitrea and Kriwacki 2016; Shin and Brangwynne 2017). With the kitchen table in mind, the common ingredient of the liquid-like droplets in the cell are the IDPs. The property that makes IDPs the ideal candidates to undergo LLPS is their similarity to polymers, in features like multivalence and formation of rather weak interactions. Polymers can demix in response to an increase or decrease in temperature undergoing phase transitions with a lower or upper critical solution temperature (LCST or UCST), respectively (Martin and Mittag 2018).

Polymers demix only in conditions where the resulting free energy of demixing is favorable. The configurational entropy of polymer chains in solution is a collective entropy of different conformations that each chain can adopt and the relative positions that polymers and solute molecules can adopt. With an increase of concentration, the configurational entropy decreases, therefore the corresponding free energy increases. Demixing into two phases, one of high and one of low concentration of polymer, can only occur if the multivalent interactions between chains of polymer are attractive on average, so that the resulting free energy decreases with an increase of the concentration. The analytical frameworks adapted in the IDPs phase separation are the Flory−Huggins theory, Overbeek−Voorn Theory, and Random Phase Approximation Theory. Despite being useful for some systems, none of the theories accounts for sequence-specific effects. Additionally, biological phase separated condensates are rarely formed by a homopolymer and are far from thermodynamic equilibrium (Lin et al. 2018). In term of sequence-specificity of LLPS not much yet can be concluded from the limited number of sequences known to undergo phase transitions, but some patterns have been observed, like FG repeats in nucleoporins (Patel et al. 2007), the R and Y-containing repeats in FUS (Wang et al. 2018), VPGV/GVGV β-turns repeats in elastin (Reichheld et al. 2017), and charge pattern

in NPM1 (Nott et al. 2016; Mitrea et al. 2018). Primary amino acid sequence of the "low complexity regions", here referred to as IDPs, undergoing phase transition encodes for the type of the transition, whether it is LCST or UCST, and for the material properties and miscibility of the resulting dense phases. The collective observations from the existing experimental studies suggests that the UCST vs LCST behavior results from the balance of hydrophobic and polar/charged amino acids in the IDP, Figure 10 (Martin and Mittag 2018). The LLPS property of a given protein is often modified by post-translational modifications, like phosphorylation status of CPEB4 (Guillén-Boixet et al. 2016), making the direct translation of primary sequence composition to LLPS behavior less straightforward.

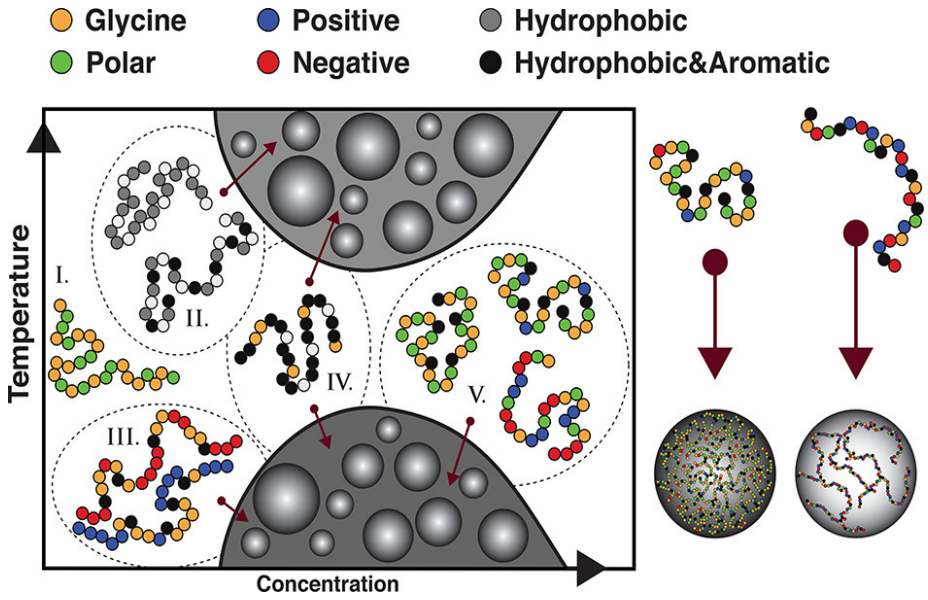


Figure 10. Sequence dependence of IDPs phase behavior

UCST or LCST behavior of an IDP depends on the balance of hydrophobic vs polar/charged amino acids. A purely polar IDP (I) is likely soluble over a wide temperature range. The addition of hydrophobic amino acids (II) results in LCST behavior. A mixture of oppositely charged polymers has UCST behavior. If the hydrophobic amino acids are predominately aromatic (IV), the LCR could have either a UCST or LCST transition. Polar IDP with aromatic amino acids and increasing charge (V) have UCST behavior. In addition, the intervening sequence space modulates the properties of the dense protein phase. For example, increasing the fraction of charged residues could result in a dense protein phase that has a higher fraction of solvent and is therefore less concentrated.
From (Martin and Mittag 2018)

1.4.1 Phase separation in nucleus

Phase separation of macromolecules into distinct compartments is becoming a widely observed mechanism of intracellular organization of the cytoplasm and the nucleus. The largest polymer in the cell nucleus, chromatin, has been shown to condense into heterochromatin domains, the high density, transcriptionally mostly silent state of chromatin, driven by the heterochromatin protein 1 (HP1) LLPS (Larson et al. 2017; Strom et al. 2017; Tatarakis et al. 2017). Studies on reconstituted nucleosome arrays, composed of histone octamers and DNA template, have shown that this chromatin model can undergo

self-organization through salt-induced LLPS with >10,000-fold concentration of nucleosomal arrays in the droplet state (Gibson et al. 2019). In the limited nuclear space next to, probably multiple, phase separated states of the chromatin there are many phase separated protein and nucleic acids organelles. Examples of those include: nucleoi (Feric et al. 2016), paraspeckles (Hennig et al. 2015), PML bodies (Shen et al. 2006), nuclear speckles (Fei et al. 2017) and Cajal bodies (Kaiser et al. 2008). These condensates are likely to form in the low density largely euchromatin regions and mechanistically exclude chromatin as shown using CasDrop, a CRISPR-Cas9-based optogenetic technology which can induce localized condensation of liquid droplets using IDRs of nuclear proteins at specific genomic loci. The condensates targeted to specific loci have been proposed to act like a mechano-active chromatin filters, by excluding the non-targeted chromatin, but bringing together distant targeted genomic loci through coalescence of the IDR driven phase separated droplets at the loci (Shin et al. 2018).

1.4.1.1 Phase separation in transcription

The mechanism of pulling distant genomic loci through phase separation of nuclear proteins involved in transcription has been proposed to be involved in the formation of super-enhancers (SE). SE are clusters of enhancers that control genes that have especially prominent roles in cell-type-specific processes. SEs are occupied by particularly high density of interacting proteins and are able to drive higher levels of transcription. The proposed model of phase separation driven transcription at super enhancers would explain several features previously related to SEs, like clustering of factors, dynamic changes, hyper-sensitivity of SEs to transcriptional inhibitors, and simultaneous activation of multiple genes by the same enhancer (Hnisz et al. 2017). That model has been validated experimentally, in a study showing enrichment of transcriptional coactivators BRD4 and MED1 at active SEs in a form of puncta observable by fluoresce microscopy. Formation of these puncta/clusters/droplets was sensitive to 1,6-hexanodiol, an alcohol perturbing hydrophobically driven phase separation (Sabari et al. 2018).

An argument for the phase separation process involvement in transcription is the observation that Pol II itself clusters through phase separation of its carboxy-terminal domain (Boehning et al. 2018; Lu et al. 2018). Interestingly, it is the disordered AD of a TF Oct4 that mediates Mediator condensation an embryonic stem cell SE, suggesting a possible role for a high ID content in TFs, Figure 11. Additionally, not only the AD of Oct4, but also few others TFs, including full length ER, were able to co-phase separate with Mediator *in vitro* (Boija et al. 2018). Contrarily, another group opposed the involvement of phase separation process in transcription in an example case of FET TFs driven transcription. Live-cell single-molecule imaging studies on FET protein family (FUS/EWS/TAF15) TFs revealed formation of local high-concentration interaction hubs at synthetic and endogenous genomic loci that were able to recruit Pol II machinery. The authors proposed a model where the ADs of TFs cluster through dynamic, multivalent and sequence specific interactions without detectable phase separation at endogenous levels of proteins. Formation of these hubs/clusters is essential to transcription by a mechanism still to be discovered (Chong et al. 2018). Finally, a study on ER in human breast cancer cells showed acute and cooperative assembly of functional enhancers in response to 17β-estradiol (E_2). The robust E_2-responsive enhancers, in contrary to weak ERα enhancers, are characterized by high levels of enhancer RNAs (eRNA) transcripts and recruitment of an ERα-dependent, megadalton-scale protein complex (MegaTrans, with components such

as GATA3, FOXA1 and AP2γ). The MegaTrans enhancers formed likely by LLPS as demonstrated by *in vitro* LLPS propensity of GATA3 and ERα, fast dynamics of ERα foci in cells measured by FRAP and 1,6-hexanediol sensitivity. An interesting feature of the MegaTrans enhancers was the ability to bring distant genomic loci into proximity in eRNA dependent manner, possibly through coalescence of droplets as proposed for mentioned chromatin filters. Additionally, optimal cooperative activation of these enhancers was additionally modulated by the ability of the enhancer loci to interact with phase separated interchromatin granules/PML bodies (Nair et al. 2019). The limited number of examples of phase separation involvement in transcription leaves its importance in the process speculative. However, it is becoming apparent that the phase separation driven enhancer assembly is a necessary or advantageous step at least in the case of a subset of enhancers.

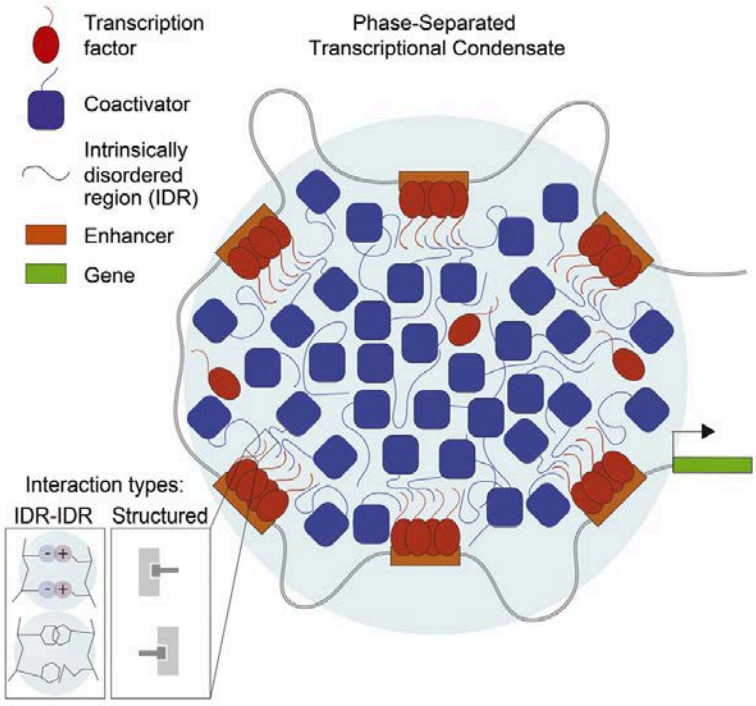


Figure 11. Model of a phase separated transcription condensate

From (Boija et al. 2018)

2 OBJECTIVES

1. Characterization of the structural features of the AR AD.
2. Determination of the specificity of the EPI-001 binding.
3. Characterization of the AD sequence and structure determinants of Rap74-CTD interaction with the AR AD.
4. Determination of SPOP-AR ability to demix in cells.
5. Determination of the ability of the AR AD to phase separate *in vitro.*
6. Study of the AR ability to phase separate in cells.
7. Understanding of the structural consequences of the AR AD phase separation.

3 CONTRIBUTION TO THE PUBLICATIONS

Student's contribution to the following publications:

1. *EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor*

- expression and purification of the AF1* construct
- help in the preparation of the 15N AF1* samples

Part of the data has been used in the thesis of Eva De Mol in a classical (not by articles) format of the thesis - Structure, dynamics and interactions of the N-terminal domain of the androgen receptor in 2014.

2. *Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators*

- expression and purification of the AF1* and Tau-5* constructs
- analysis of the structural changes caused by Rap74-CTD binding to the AR AD by NMR (ΔCα of the Tau-5* upon Rap74-CTD binding)
- help in the design of stapled peptide (Hel) and its characterization by CD
- measurements of the binding affinities of chosen peptides with Rap74-CTD by NMR
- contribution to designing the experiments
- help in writing and figure preparation

Part of the data has been used in the thesis of Eva De Mol in a classical (not by articles) format of the thesis - Structure, dynamics and interactions of the N-terminal domain of the androgen receptor in 2014.

3. *Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments*

- establishing the collaboration with Tanja Mittag's laboratory
- optimization of expression and purification of the AD construct
- *in vitro* phase separation experiments of the AD and co-phase separation with SPOP (confocal microscopy)

4. *Hormone binding causes the condensation of androgen receptor that mediates folding of activation function motifs*

- implementation of the methods to study LLPS
- design of the *in vitro* and *ex vivo* experiments
- establishing the collaboration with Gordon L. Hager's laboratory and design of the N&B experiments with Diego Presman
- preparation of the samples and partial analysis of the NMR experiments
- *in vitro* experiments of: the cloud point temperature determination (phase diagram, ionic strength dependence of phase separation, LLPS by DIC, LLPS by confocal microscopy, FRAP, CD)
- figures preparation and writing

4 PUBLICATIONS

4.1 EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor

PUBLICATIONS

ACS chemical biology

Articles

pubs.acs.org/acschemicalbiology

EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor

Eva De Mol,[†] R. Bryn Fenwick,[†] Christopher T. W. Phang,[†] Victor Buzón,[‡] Elzbieta Szulc,[†] Alex de la Fuente,[†] Albert Escobedo,[†] Jesús García,[†] Carlos W. Bertoncini,[†,#] Eva Estébanez-Perpiñá,[‡] Iain J. McEwan,[§] Antoni Riera,[†,∥] and Xavier Salvatella*[,†,⊥]

[†]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain

[‡]Departament de Bioquímica i Biología Molecular, and Institute of Biomedicine, University of Barcelona (IBUB), Baldiri Reixac 10, 08028 Barcelona, Spain

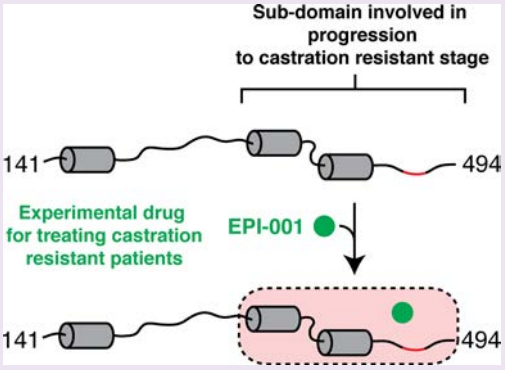
[§]Institute of Medical Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, IMS Building, Foresterhill, Aberdeen AB25 2ZD, Scotland, United Kingdom

[∥]Departament de Química Orgànica, Universitat de Barcelona, Martí i Franqués 1-11, 08028 Barcelona, Spain

[⊥]ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain

Supporting Information

ABSTRACT: Castration-resistant prostate cancer is the lethal condition suffered by prostate cancer patients that become refractory to androgen deprivation therapy. EPI-001 is a recently identified compound active against this condition that modulates the activity of the androgen receptor, a nuclear receptor that is essential for disease progression. The mechanism by which this compound exerts its inhibitory activity is however not yet fully understood. Here we show, by using high resolution solution nuclear magnetic resonance spectroscopy, that EPI-001 selectively interacts with a partially folded region of the transactivation domain of the androgen receptor, known as transactivation unit 5, that is key for the ability of prostate cells to proliferate in the absence of androgens, a distinctive feature of castration-resistant prostate cancer. Our results can contribute to the development of more potent and less toxic novel androgen receptor antagonists for treating this disease.



Prostate cancer (PC) is the second most common cancer in men and can be cured by surgery or radiotherapy in *ca.* 70% of cases. The first line of pharmacological treatment for the remaining cases targets the androgen receptor (AR) because prostate cells depend on its activation by androgens for their growth and proliferation.[1] Activation can be prevented by combining androgen deprivation therapy, which inhibits the secretion of androgens by the testes, with the administration of antagonists that competitively bind to the binding site for androgens in the ligand-binding domain (LBD) of AR.[2]

Two to three years into this treatment, PC patients inevitably develop castration-resistant prostate cancer (CRPC) as prostate cancer cells acquire the ability to activate AR at low levels of circulating androgens and in the presence of antagonists.[3] The mechanisms of aberrant activation are not well understood but appear to include the amplification of the *AR* gene and AR overexpression, the expression of constitutively active AR splice variants lacking the LBD, cell signaling cross-talk, and mutations in both AR and transcriptional co-regulators.[4]

AR is a large multidomain protein composed of globular ligand- and DNA-binding domains (LBD and DBD) and an N-terminal transactivation domain (NTD) that is intrinsically disordered (ID)[5,6] (Figure 1a). The function of the NTD (residues 1 to 558) is to recruit the basal transcription machinery by binding to general transcription factors either directly or assisted by transcriptional co-activators.[1] These protein−protein interactions are thought to cause the folding of binding motifs in a region of the NTD called activation function 1 (AF-1) that has not yet been characterized at high resolution (Figure 1a,b).[6] Inhibiting these interactions is considered a potential therapeutic approach for both PC and CRPC,[7] but the NTD has not been considered a suitable target for drug discovery due to its apparent lack of persistent secondary and tertiary structure.

The development of drugs that interact with ID regions has however recently been met with some success,[8,9] and has shown that targeting them with small molecules may be a viable therapeutic approach.[10,11] A particularly important develop-

Received: February 25, 2016
Accepted: June 29, 2016
Published: June 29, 2016

DOI: 10.1021/acschembio.6b00182
ACS Chem. Biol. 2016, 11, 2499−2505

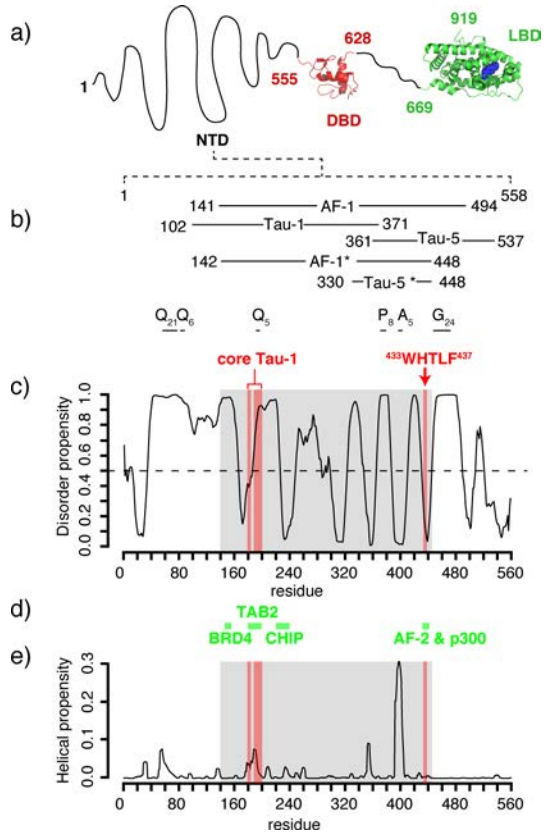


Figure 1. Predicted properties of the sequence of the transactivation domain of AR: (a) Domain organization of AR[34,39] with an indication of the position of Zn atoms in the DBD (gray) and of dihydrotestosterone (DHT) in the LBD (blue). (b) Definitions of activation function 1 (AF-1), transcription activation units 1 and 5 (Tau-1 and Tau-5), AF-1* and Tau-5* (the regions of sequence studied in this work) with an indication of the regions of low sequence complexity, such as polyGln (Q_n), polyPro (P_n), polyAla (A_n), and polyGly (G_n) tracts. (c) Propensity to disorder of the NTD predicted by PONDR VL-XT[26] with an indication of the functional motifs defining the core of Tau-1 and Tau-5, shaded in red, and of the region of sequence studied in this work, shaded in gray. (d) Positions of the motifs of the NTD of AR involved in protein−protein interactions and acronyms of the binding partners (see main text for details). (e) Propensity to adopt α-helical conformations predicted by Agadir,[36] as a function of residue number, with an indication of the core of Tau-1 and Tau-5 (shaded in red) and AF-1* (shaded in gray).

ment in this area was the recent discovery of EPI-001, an experimental drug for the treatment of CRPC identified by phenotypic screening that is efficacious both in cell lines and in animal models of this disease.[12] *In vivo* EPI-001 binds irreversibly to the AR NTD and weakens its interaction with general transcription factors and transcriptional co-activators.[13]

The discovery of EPI-001 represents an important milestone, and a derivative of this compound is currently in clinical trials for CRPC (NCT02606123). The lack of a detailed understanding of the structural features of the domain and of the mechanism of action of this class of compounds represents, however, a hurdle for the rational design of optimized inhibitors. Here we reveal, by using solution nuclear magnetic resonance (NMR) spectroscopy, the structural properties of the ID regions of the NTD predicted to have a high propensity to fold and show that EPI-001 targets a region of sequence, known as transcription activation unit 5 (Tau-5), that is key for the ability of prostate cells to proliferate in the absence of androgens.

RESULTS AND DISCUSSION

As previously reported,[5,6,15] the sequence of the NTD has features typically encountered in ID regions.[16] It has a high content of Gly, Pro, and polar and charged residues and, as shown in Figure 1b, possesses regions of low sequence complexity, such as polyGln (residues 58−78, 84−89, 193−197), polyPro (372−379), polyAla (398−402), and polyGly (449−472) tracts. Such disordered tracts are commonly flanked by motifs of relatively low disorder propensity, with some helical propensity and rich in hydrophobic side chains, that are often involved in interactions with binding partners.[17] An analysis of the sequence of the NTD with predictors of disorder and helical propensity indicates that several such motifs are indeed present in the domain, and it has been proposed that they may correspond to the regions of sequence recognized by general transcription factors and transcriptional co-regulators[15] (Figure 1c,d,e).

A number of studies have aimed at identifying the regions of sequence of AF-1 that are essential for transcriptional activity.[18,19] Two large regions of sequence, known as transcription activation units 1 and 5 (Tau-1, 102−371, and Tau-5, 361−537) have emerged as especially important (Figure 1b). Tau-1 is crucial for the transcriptional activity of AR when the receptor is activated by androgens.[18] Although Tau-5 is less well characterized, it has been shown to be more important than Tau-1 when activation occurs via androgen independent mechanisms in androgen depletion independent cell lines derived from CRPC patients and mouse xenograft models of PC.[20,21] Efforts have also been directed at identifying the critical stretches of residues within these regions and suggest that for Tau-1 they correspond to residues 174 to 204 (core Tau-1),[19] and for Tau-5 to residues 433 to 437 (WHTLF motif)[20] (Figure 1c).

AF-1* is partially folded. To investigate the structural properties of the NTD and its interaction with EPI-001, we cloned, expressed, purified, and studied by NMR at 278 K a 306-residue construct (AF-1*, residues 142−448) containing the part of AF-1 predicted to have a low disorder propensity (Figure 1b and 1c), that is flanked by polyGln and polyGly tracts (Figure 1b). In agreement with previous reports based on the use of other biophysical methods,[6] the $^{1}H,^{15}N$-HSQC spectrum of AF-1* had the features expected in an ID region such as low H^N chemical shift dispersion (Figure 2a). In addition, and contrary to what is the case for NTD constructs containing residues 1 to 141, which include the polyQ tract (Figure 1b), this region of the NTD was sufficiently stable to allow a characterization of its structural properties by solution NMR.

In spite of its ID nature, the AF-1 region of the NTD has been shown to have helical propensity by circular dichroism (CD) in buffer and in the presence of cosolvents that stabilize intramolecular hydrogen bonds, as well as in the presence of the natural osmolyte trimethylamine oxide (TMAO).[5,22] To identify the regions of sequence that adopt this secondary structure, we assigned the resonances of AF-1* by a *divide and conquer* approach (see Methods), compared the Cα and Cβ chemical shifts, which are reliable reporters of secondary structure, to those predicted for disordered AF-1* (Figure 2b),[23] and used the SSP algorithm of Forman-Kay and co-workers[24] to quantify the secondary structure propensity for the

DOI: 10.1021/acschembio.6b00182
ACS Chem. Biol. 2016, 11, 2499−2505

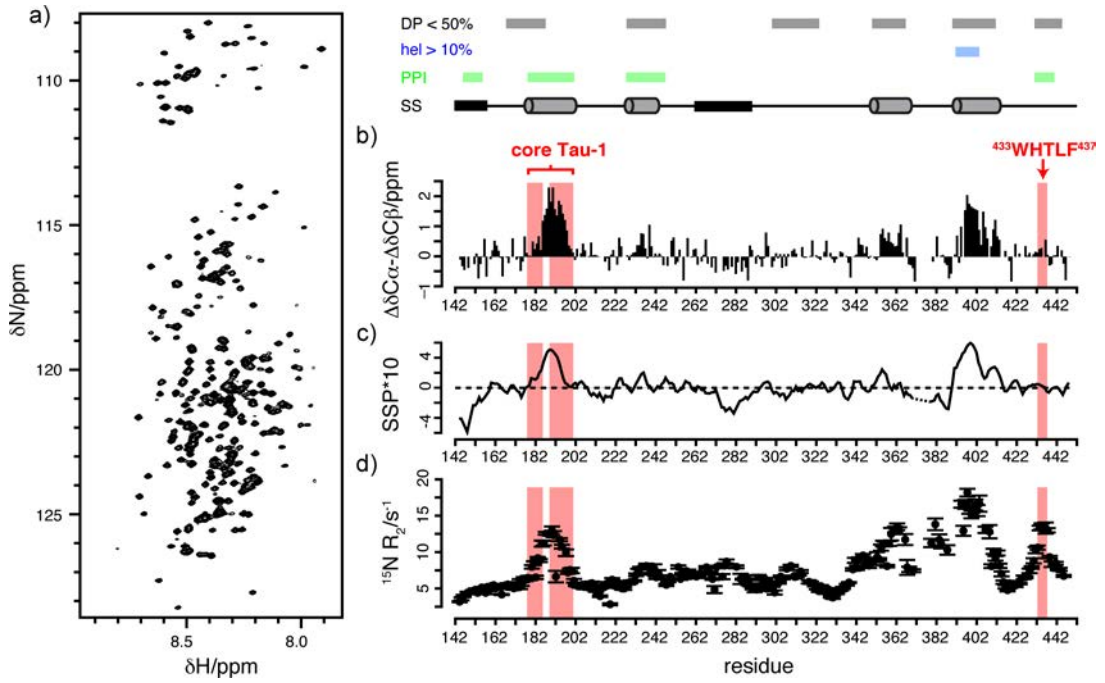


Figure 2. Structural properties of the transactivation domain revealed by solution NMR: (a) $^{1}H,^{15}N$-HSQC spectrum of AF-1*. (b) Plot of the difference between the secondary Cα and Cβ chemical shifts measured for AF-1* with an indication of the regions of sequence with predicted disorder propensity (DP) lower than 50%, in gray, for AF-1* with predicted helical propensity (hel) higher than 10%, in blue, and for AF-1* involved in protein−protein interactions (PPIs), in green and, at the bottom of the top panel, an indication of the nascent secondary structure identified experimentally (SS, cylinders indicate helical propensity; black rectangles indicate propensity to adopt an extended conformation). (c) Plot of the secondary structure propensity of the residues of AF-1*, where SSP = 1 indicates a fully formed helix and SSP = 1 an extended conformation, obtained by using the algorithm SSP to extract the information on secondary structure contained in the backbone ($^{13}C\alpha$, $^{13}C\beta$, ^{13}CO, ^{15}N, HN) chemical shifts. (d) Plot of the ^{15}N transverse relaxation rates (R_2) of the residues of AF-1* at 250 μM. The cores of Tau-1 and Tau-5 are shaded in red.

various residues of this construct from analysis of the backbone ($^{13}C\alpha$, $^{13}C\beta$, ^{13}CO, ^{15}N, H^N) chemical shifts.

The results that we obtained indicated the presence of two regions of high helical propensity (defined as $\Delta\delta C\alpha-\Delta\delta C\beta > 1$ ppm and SSP $\approx$ 0.5, corresponding to a helical propensity of 50%) which correspond to residues 185−200, in Tau-1, and 390−410, in Tau-5. Other regions of intermediate helical propensity (defined as $\Delta\delta C\alpha-\Delta\delta C\beta \approx 0.5$ ppm and SSP $\approx$ 0.2) could also be identified, such as the region 230−240, in Tau-1, and 355−365 in Tau-5 (Figure 2b, c). In addition to identifying regions of helical secondary structure, the analysis of the chemical shifts also suggests that residues 144−154 and 270−290 of AF-1* adopt an extended conformation ($\Delta\delta C\alpha-\Delta\delta C\beta \approx -0.5$ ppm and SSP ≈ -0.3) (Figure 2b, c).

To further characterize the structural properties of AF-1*, we characterized the dynamics of AF-1* by measuring the transverse relaxation rates (R_2) of the backbone ^{15}N nuclei. Such relaxation rates are good reporters of nascent secondary structure and transient tertiary contacts in chemically denatured and ID proteins.[25] The results that we obtained, presented in Figure 2d, indicate that the regions with nascent secondary structure revealed by the analysis of the ^{13}C chemical shifts (Figure 2b) also display relatively high R_2 values. These are especially high, reaching values of *ca.* 15 s^{-1}, for three regions of sequence found in Tau-5 predicted to have low disorder propensity[26] and presenting, in two cases, high helical propensity.

Some of the regions of sequence that we identified as partially folded in AF-1* correspond, in fact, to the epitopes of binding partners of AR or are known to be important for the function of this receptor. In Tau-1, for example, the region that is most structured corresponds to *core Tau-1* (Figure 2b). This region encompasses the motif ^{183}LSEASTMQLL^{192}, which is the binding epitope of TAB2, a component of the NCoR corepressor complex.[27] Partially folded residues 144 to 154, also in Tau-1, are part of the binding epitope of the amino-terminal bromo domains of BRD4.[28] The region containing residues 230 to 240, of intermediate helicity, overlaps with the binding site of CHIP, a protein known to recruit the chaperone machinery by interaction with Hsc70, Hsp70, and Hsp90, which mediates the degradation of AR by the proteasome[29] (Figure 1d).

In Tau-5 the regions that we found to be partially folded are separated by motifs of low sequence complexity, such as the Pro-rich region ^{371}GPPPPPPPPHP^{381} and the Gly-rich region ^{414}GAGAAGPGSGSPS^{426}. The former has been proposed to be the binding site of the SH3 domain of Src,[30] and the latter harbors a phosphosite, S424, which is thought to be important for transcriptional activity.[31] Although we have not assigned the backbone resonances corresponding to the Pro residues in this Pro-rich motif, an analysis of the secondary chemical shifts of residues in its flanking regions (367−370, 385−388, Figure 2b) suggests that it may adopt an extended conformation, likely of the polyproline II type, given the high Pro content; the polyproline II conformation is rather unusual but has been documented for several other IDPs.[32] The chemical shifts and transverse ^{15}N relaxation rates of the Gly-rich region (414−426, Figure 2b,c,d) indicate, by contrast, that it is disordered, as expected due to its high Gly content.

DOI: 10.1021/acschembio.6b00182
ACS Chem. Biol. 2016, 11, 2499−2505

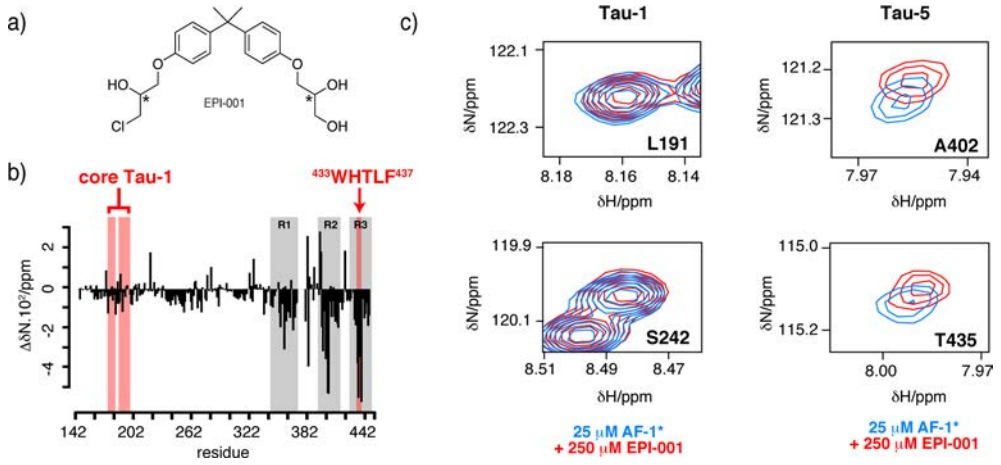


Figure 3. EPI-001 selectively interacts with transcription activation unit 5 of the transactivation domain of AR: (a) Structure of EPI-001 with an indication of the two stereogenic centers with the symbol *. (b) Plot, as a function of residue number, of the change in ^{15}N chemical shift of AF-1* caused by addition of EPI-001. (c) Selected regions of the ^{1}H,^{15}N-HSQC spectrum of 25 μM AF-1* in the absence (blue) and in the presence (red) of 10 mol equiv of EPI-001.

No binding partners have yet been identified for regions 355−365 and 390−410, that are partially helical, but the ^{433}WHTLF^{437} motif in region 433−447, corresponding to the core of Tau-5, has been proposed to bind both to activation function 2 (AF-2) in the LBD of androgen-bound AR[33,34] and to histone acetyltransferase p300;[35] in addition, it has been shown to be indispensable for transcriptional activity in androgen depletion independent cell lines.[20]

We noted that the helical propensity of the partially folded regions identified in both Tau-1 and Tau-5 (SSP ≈ 0.5, corresponding to a population of helix of ca. 50%) is substantially higher than that predicted by algorithms that solely consider local interactions such as Agadir[36] (Figures 1e and 2b, c), indicating that nonlocal interactions may contribute to stabilizing the secondary structure. We thus investigated whether such nonlocal interactions included intramolecular contacts between residues in Tau-1 and Tau-5, similarly to what has been shown to occur in globular proteins under mild denaturing conditions.[37] For this analysis we used an NTD construct encompassing the part of Tau-5 contained in AF-1*, corresponding to residues 330−448 (termed Tau-5*, Figure 1b). We compared the H^N and ^{15}N chemical shifts of Tau-5* to those of AF-1* and observed that they were highly similar (Figure S2). These results indicate that the chemical environment that residues in the Tau-5 region experience is the same in both the presence and absence of Tau-1. We conclude that the secondary structure present in AF-1* does not rely on long-range interactions between residues in Tau-1 and residues in Tau-5.

EPI-001 interacts reversibly with Tau-5. It has been proposed that the mechanism of action of EPI-001 (Figure 3a) involves two steps,[13] the first step being the formation of a reversible complex between this compound and a specific conformation of AF-1 and the second the nucleophilic attack, by a protein side chain, on the C−Cl bond of EPI-001 to form an adduct unable to activate transcription. To investigate the first step of this mechanism and characterize the putative reversible complex, we synthesized EPI-001 (details available as SI) and used NMR to analyze its effect on the resonances of AF-1* at 278 K. As shown by MS under these conditions, the irreversible reaction of EPI-001 with AF-1* is sufficiently slow to allow the study of the reversible interaction by NMR (Figure S5).

The results that we obtained, shown in Figures 3b and c, indicate that EPI-001 causes small but highly reproducible changes in ^{15}N chemical shifts in residues 354 to 448 of AF-1*, that correspond to Tau-5. Interaction with EPI-001 affects the resonances of a large number of residues, which are found in the three regions of sequence in Tau-5 identified as partially folded by the combined analysis of the backbone ^{13}C chemical shifts and ^{15}N transverse relaxation rates (Figures 2b, 2d, and 3b). We detected much smaller perturbations of the resonances of residues corresponding to Tau-1 even though this region contains partially helical regions such as residues in core Tau-1 (Figures 2b and 3b). These results indicate that EPI-001 does not simply interact with any region of sequence with helical propensity but, rather, that this compound targets Tau-5 due to the ability of this subdomain to adopt a conformation or, most likely, an ensemble of conformations that have affinity for it.

With the aim of further investigating the nature of the interaction, we monitored the effect of AF-1* on the resonances of EPI-001. We observed that substoichiometric amounts (25 μM and 50 μM) of AF-1* caused small but reproducible chemical shift changes in the ^{1}H NMR spectrum of EPI-001 at 250 μM, as well as increases in line width similar to those commonly observed in small drug-like molecules transiently associating with macromolecules (Figure S3). Together with our observation of perturbations in a large number of residues of AF-1*, this result suggests that EPI-001 interacts with an ensemble of conformations adopted by AF-1* where these regions of sequence adopt a partially folded structure.

A large number of residues in Tau-5 experience chemical shift perturbations, more than what is expected for the binding of a small molecule. This can be due to the inspecific interaction of one or various molecules of EPI-001 with three independent interaction sites or to the combination of direct and indirect chemical shift changes due to structural changes occurring in Tau-5 upon interaction with EPI-001.[38] To exclude that one or more EPI-001 molecules interact independently with the three regions of sequence of AF-1* where chemical shift perturbations are observed, we monitored the effect on its ^{1}H NMR spectrum of adding three peptides

DOI: 10.1021/acschembio.6b00182
ACS Chem. Biol. 2016, 11, 2499−2505

(R1 to R3) with sequences corresponding to regions 341−371, 391−414, and 426−446, respectively (details available as SI). We found that these peptides caused no changes in the spectrum of EPI-001, indicating that EPI-001 does not interact with them (Figure S3) and confirming that residues 354−448 must simultaneously be present because they are necessary for binding this compound or for stabilizing the bound state of Tau-5* (Figure 4).

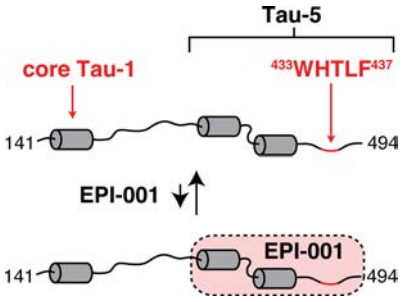


Figure 4. Scheme of the interaction of EPI-001 with partially folded Tau-5.

It is interesting to note that Tau-5, the region of sequence of the NTD targeted by EPI-001, is partially folded, but it is important to emphasize that the data that we have obtained do not provide us with a mechanism of molecular recognition for this interaction. It is possible that a conformational selection mechanism operates in which EPI-001 interacts with a subset of the conformations that Tau-5* samples in equilibrium, but it is also possible that EPI-001 induces a new conformation in Tau-5*. Further work will be necessary to differentiate these two scenarios.

EPI-001 has two stereogenic centers and can therefore be found as four stereoisomers. To investigate whether the interaction between this compound and Tau-5 is stereospecific, we synthesized the four stereoisomers (details available as SI) and studied their interaction with AF-1* by NMR. The results that we obtained (Figure S4) indicate that the four compounds can interact with the NTD of AR, and thus the interaction appears to occur with little or no stereoselectivity. These results are in agreement with results obtained *in vivo* by Myung et al., who found that, although one of the stereoisomers tested was slightly more active than the other ones, the inhibitory activity of the four stereoisomers was similar.[13]

Put together, our results indicate that AF-1 is partially folded in regions of sequence that correspond to those which are functionally relevant for interacting with the transcription machinery and co-regulators of transcription. In addition, they reveal that Tau-1 and Tau-5, the two independent transcription activation units that are found in AF-1, correspond to different subdomains that appear not to be involved in long-range interactions, i.e. are structurally and dynamically, at least under our conditions, independent. Finally, and most importantly, they show that Tau-5, which plays a particularly important role in AR activation in the absence of androgens, can be targeted by compounds such as EPI-001 (Figure 4). Although the lack of stereoselectivity that we observe suggests that the binding mode of EPI-001 may not be sufficiently well-defined for conventional drug development, our results are of relevance for drug discovery for CRPC because they suggest that the NTD, and Tau-5 in particular, may represent a suitable therapeutic target.

METHODS

Protein expression and purification. The DNA sequences coding for human WT AR residues 265 to 340 (AF-1*$_{265-340}$), 330 to 448 (Tau-5*), and 142 to 448 (AF-1*) were cloned into Gateway pDEST17 vectors (Invitrogen) with an N-terminal His_6-tag and a TEV cleavage site. Transformed *E. coli* Rosetta cells were grown at 37 °C in LB medium for the production of nonisotopically labeled protein. For single (^{15}N) or double (^{15}N,^{13}C) isotopic labeling, cells were grown in minimal MOPS medium[14] containing $^{15}NH_4Cl$ or $^{15}NH_4Cl$ and ^{13}C-glucose, respectively. The AF-1* fusion protein accumulated in inclusion bodies which were solubilized in lysis buffer containing 8 M urea and fragmented by a pass through a cell disruptor at 25 kpsi. The fusion protein was purified by Ni^{2+} affinity chromatography in urea, which was removed by two dialysis steps, after which the His_6-tag was cleaved by the TEV protease. The cleaved AF-1* was further purified by reverse Ni^{2+} affinity and size exclusion chromatography in 20 mM sodium phosphate buffer with 1 mM tris(2-carboxyethyl)-phosphine hydrochloride (TCEP) at pH 7.4.

Peptides. The synthesis of peptides R1, R2, and R3 (see SI) was performed by solid phase peptide synthesis by GenScript (peptide R1) or by ICTS NANBIOSIS, more specifically by the peptide synthesis unit of the CIBER in bioengineering, biomaterials, and nanomedicine (CIBER-BBN) at the Barcelona science park (peptides R2 and R3). The lyophilized peptides were dissolved in deionized water, and the pH of the resulting solution was adjusted by addition of a concentrated NaOH solution. The concentration of these solutions was determined by amino acid analysis. The absence of intermolecular disulfide bonds in the R2 peptide, which contains one Cys residue, was confirmed by mass spectrometry (MS).

Chemical synthesis of EPI-001 and stereoisomers. EPI-001 contains two stereogenic centers and can therefore be found as two pairs of enantiomers. We synthesized the four isomers with high diastereo- and enantioselectivity (Chiral HPLC). The synthesis followed the sequence detailed in Figure S1 of the SI. Bisphenol A was treated with enantiomerically pure glycidol. The resulting diol was protected as dimethyl acetal, and the free phenol was allowed to react with another isomer of glycidol. The corresponding diol was transformed into the epoxide and opened with $CeCl_3$. The treatment gave the final product by concomitant deprotection of the acetal. Full experimental details as well as the complete characterization of all isomers can be found in the SI.

NMR. The assignment of AF-1* was obtained by using a *divide and conquer* approach. The resonances of fragments AF-1*$_{265-340}$, Tau-5* (330 to 448), and AF-1* (142−448) were obtained by analyzing conventional three-dimensional triple resonance experiments acquired with standard Bruker pulse sequences on Bruker 600 and 800 MHz spectrometers at 278 K in 20 mM sodium phosphate buffer with 1 mM TCEP at pH 7.4. The resonances of AF-1*$_{265-340}$ and Tau-5* were equivalent to those of AF-1*, except for residues near the termini (Figure S2), and allowed transferring the assignments from the former to the latter; the residues that were unique to AF-1* were assigned directly by analysis of the relevant spectra. To measure the perturbations caused by EPI-001, EPI-002, EPI-003, EPI-004, and EPI-005 (details available as SI) on the resonances of AF-1*, appropriate volumes of a 50 mM stock solution of these compounds in 100% dioxane-d_8 were added to aliquots containing 25 μM AF-1*, 20 mM sodium phosphate, 1 mM TCEP, 10% D_2O, and 30 μM DSS-d_6 at pH 7.4.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.6b00182.

Experimental procedures, NMR data, peptide sequences, and mass spectrometry data (PDF)

DOI: 10.1021/acschembio.6b00182
ACS Chem. Biol. 2016, 11, 2499−2505

AUTHOR INFORMATION

Corresponding Author

*Phone: +34 934020459. E-mail: xavier.salvatella@irbbarcelona.org.

Present Address

#Instituto de Biología Molecular y Celular de Rosario (IBR), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ocampo y Esmeralda, Rosario, Argentina.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank J. M. Valverde (IRB) as well as the NMR facilities of the University of Barcelona (CCiT UB) and the Instituto de Química Física Rocasolano (IQFR, CSIC) for their assistance in, respectively, protein production and NMR. This work was supported by IRB, ICREA (X.S.), Obra Social "la Caixa" (Fellowship to E.D.M. and CancerTec grants to X.S.), MICINN (CTQ2009-08850 to X.S.), MINECO (BIO2012-31043 to X.S.; CTQ2014-56361-P to A.R), Marató de TV3 (102030 to X.S. and 102031 to E.E.P.,) the COFUND programme of the European Commission (C.T.W.P., A.R., and X.S.), the European Research Council (CONCERT, contract number 648201, to X.S.), the Ramón y Cajal program of MICINN (RYC-2011-07873 to C.W.B.,) the Serra Hunter Programme (E.E.P.), and AGAUR (SGR-2014-56RR14 to E.E.P). IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO (Government of Spain).

REFERENCES

(1) Gelmann, E. P. (2002) Molecular Biology of the Androgen Receptor. *J. Clin. Oncol. 20*, 3001−3015.

(2) Akaza, H., Hinotsu, S., Usami, M., Arai, Y., Kanetake, H., Naito, S., and Hirao, Y. (2009) Combined androgen blockade with bicalutamide for advanced prostate cancer. *Cancer 115*, 3437−3445.

(3) Harris, W. P., Mostaghel, E. A., Nelson, P. S., and Montgomery, B. (2009) Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. *Nat. Clin. Pract. Urol. 6*, 76−85.

(4) Karantanos, T., Corn, P. G., and Thompson, T. C. (2013) Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. *Oncogene 32*, 5501−5511.

(5) Reid, J., Kelly, S. M., Watt, K., Price, N. C., and McEwan, I. J. (2002) Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions. *J. Biol. Chem. 277*, 20079−20086.

(6) Lavery, D. N., and McEwan, I. J. (2008) Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding. *Biochemistry 47*, 3360−3369.

(7) Quayle, S. N., Mawji, N. R., Wang, J., and Sadar, M. D. (2007) Androgen receptor decoy molecules block the growth of prostate cancer. *Proc. Natl. Acad. Sci. U. S. A. 104*, 1331−1336.

(8) Lamberto, G. R., Torres-Monserrat, V., Bertoncini, C. W., Salvatella, X., Zweckstetter, M., Griesinger, C., and Fernandez, C. O. (2011) Toward the discovery of effective polycyclic inhibitors of alpha-synuclein amyloid assembly. *J. Biol. Chem. 286*, 32036−32044.

(9) Krishnan, N., Koveal, D., Miller, D. H., Xue, B., Akshinthala, S. D., Kragelj, J., Jensen, M. R., Gauss, C.-M., Page, R., Blackledge, M., Muthuswamy, S. K., Peti, W., and Tonks, N. K. (2014) Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nat. Chem. Biol. 10*, 558−566.

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

(11) Heller, G. T., Sormanni, P., and Vendruscolo, M. (2015) Targeting disordered proteins with small molecules using entropy. *Trends Biochem. Sci. 40*, 491−496.

(12) Andersen, R. J., Mawji, N. R., Wang, J., Wang, G., Haile, S., Myung, J.-K., Watt, K., Tam, T., Yang, Y. C., Bañuelos, C. A., Williams, D. E., McEwan, I. J., Wang, Y., and Sadar, M. D. (2010) Regression of castrate-recurrent prostate cancer by a small-molecule inhibitor of the amino-terminus domain of the androgen receptor. *Cancer Cell 17*, 535−546.

(13) Myung, J.-K., Bañuelos, C. A., Fernandez, J. G., Mawji, N. R., Wang, J., Tien, A. H., Yang, Y. C., Tavakoli, I., Haile, S., Watt, K., McEwan, I. J., Plymate, S., Andersen, R. J., and Sadar, M. D. (2013) An androgen receptor N-terminal domain antagonist for treating prostate cancer. *J. Clin. Invest. 123*, 2948−2960.

(14) Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) Culture medium for enterobacteria. *J. Bacteriol. 119*, 736−747.

(15) McEwan, I. J. (2012) Intrinsic disorder in the androgen receptor: identification, characterisation and drugability. *Mol. BioSyst. 8*, 82−90.

(16) Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) Intrinsically disordered protein. *J. Mol. Graphics Modell. 19*, 26−59.

(17) Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K., and Uversky, V. N. (2006) Analysis of molecular recognition features (MoRFs). *J. Mol. Biol. 362*, 1043−1059.

(18) Jenster, G., van der Korput, H. A. G. M., Trapman, J., and Brinkmann, A. O. (1995) Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. *J. Biol. Chem. 270*, 7341−7346.

(19) Callewaert, L., Van Tilborgh, N., and Claessens, F. (2006) Interplay between two hormone-independent activation domains in the androgen receptor. *Cancer Res. 66*, 543−553.

(20) Dehm, S. M., Regan, K. M., Schmidt, L. J., and Tindall, D. J. (2007) Selective role of an NH2-terminal WxxLF motif for aberrant androgen receptor activation in androgen depletion independent prostate cancer cells. *Cancer Res. 67*, 10067−10077.

(21) Metzger, E., Müller, J. M., Ferrari, S., Buettner, R., and Schüle, R. (2003) A novel inducible transactivation domain in the androgen receptor: implications for PRK in prostate cancer. *EMBO J. 22*, 270−280.

(22) Kumar, R., Betney, R., Li, J., Thompson, E. B., and McEwan, I. J. (2004) Induced alpha-helix structure in AF1 of the androgen receptor upon binding transcription factor TFIIF. *Biochemistry 43*, 3008−3013.

(23) Tamiola, K., Acar, B., and Mulder, F. A. A. (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc. 132*, 18000−18003.

(24) Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci. 15*, 2795−2804.

(25) Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein. *Science 295*, 1719−1722.

(26) Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins: Struct., Funct., Genet. 42*, 38−48.

(27) Zhu, P., Baek, S. H., Bourk, E. M., Ohgi, K. A., Garcia-Bassets, I., Sanjo, H., Akira, S., Kotol, P. F., Glass, C. K., Rosenfeld, M. G., and Rose, D. W. (2006) Macrophage/cancer cell interactions mediate hormone resistance by a nuclear receptor derepression pathway. *Cell 124*, 615−629.

(28) Asangani, I. A., Dommeti, V. L., Wang, X., Malik, R., Cieslik, M., Yang, R., Escara-Wilke, J., Wilder-Romans, K., Dhanireddy, S., Engelke,

DOI: 10.1021/acschembio.6b00182
ACS Chem. Biol. 2016, 11, 2499−2505

C., Iyer, M. K., Jing, X., Wu, Y.-M., Cao, X., Qin, Z. S., Wang, S., Feng, F. Y., and Chinnaiyan, A. M. (2014) Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature 510*, 278−282.
(29) He, B., Bai, S., Hnat, A. T., Kalman, R. I., Minges, J. T., Patterson, C., and Wilson, E. M. (2004) An androgen receptor NH2-terminal conserved motif interacts with the COOH terminus of the Hsp70-interacting protein (CHIP). *J. Biol. Chem. 279*, 30643−30653.
(30) Migliaccio, A., Castoria, G., Giovannelli, P., and Auricchio, F. (2010) Cross talk between epidermal growth factor (EGF) receptor and extra nuclear steroid receptors in cell lines. *Mol. Cell. Endocrinol. 327*, 19−24.
(31) Gioeli, D., Ficarro, S. B., Kwiek, J. J., Aaronson, D., Hancock, M., Catling, A. D., White, F. M., Christian, R. E., Settlage, R. E., Shabanowitz, J., Hunt, D. F., and Weber, M. J. (2002) Androgen receptor phosphorylation. Regulation and identification of the phosphorylation sites. *J. Biol. Chem. 277*, 29304−29314.
(32) Adzhubei, A. A., Sternberg, M. J. E., and Makarov, A. A. (2013) Polyproline-II helix in proteins: structure and function. *J. Mol. Biol. 425*, 2100−2132.
(33) He, B., Kemppainen, J. A., and Wilson, E. M. (2000) FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. *J. Biol. Chem. 275*, 22986−22994.
(34) Hur, E., Pfaff, S. J., Payne, E. S., Grøn, H., Buehrer, B. M., and Fletterick, R. J. (2004) Recognition and accommodation at the androgen receptor coactivator binding interface. *PLoS Biol. 2*, E274.
(35) Lagarde, W. H., Blackwelder, A. J., Minges, J. T., Hnat, A. T., French, F. S., and Wilson, E. M. (2012) Androgen receptor exon 1 mutation causes androgen insensitivity by creating phosphorylation site and inhibiting melanoma antigen-A11 activation of NH2- and carboxyl-terminal interaction-dependent transactivation. *J. Biol. Chem. 287*, 10905−10915.
(36) Muñoz, V., and Serrano, L. (1994) Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol. 1*, 399−409.
(37) Bruun, S. W., Iešmantavičius, V., Danielsson, J., and Poulsen, F. M. (2010) Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein. *Proc. Natl. Acad. Sci. U. S. A. 107*, 13306−13311.
(38) Fenwick, R. B., Esteban-Martín, S., and Salvatella, X. (2011) Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur. Biophys. J. 40*, 1339−1355.
(39) Shaffer, P. L., Jivan, A., Dollins, D. E., Claessens, F., and Gewirth, D. T. (2004) Structural basis of androgen receptor binding to selective androgen response elements. *Proc. Natl. Acad. Sci. U. S. A. 101*, 4758−4763.

DOI: 10.1021/acschembio.6b00182
ACS Chem. Biol. 2016, 11, 2499−2505

4.2 Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators

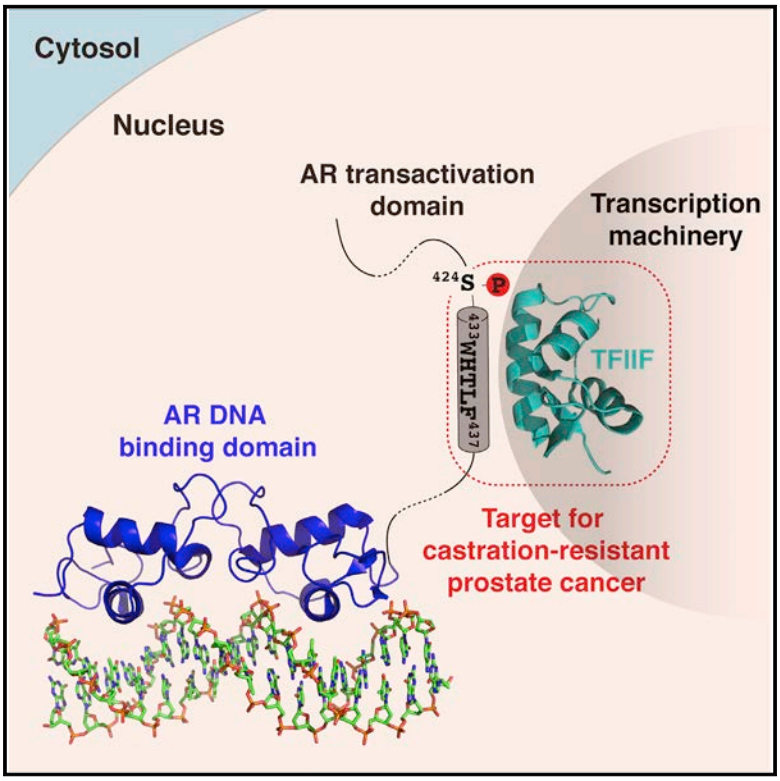
PUBLICATIONS

Short Article

Structure

Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators

Graphical Abstract



Highlights

- A short motif in transactivation unit 5 recruits the transcription machinery to the AR
- The motif is intrinsically disordered but folds into a helix upon binding
- Phosphorylation of Ser 424 of AR is essential for recruitment
- The interaction may be a target for castration-resistant prostate cancer

Authors

Eva De Mol, Elzbieta Szulc, Claudio Di Sanza, ..., Iain J. McEwan, Ángel R. Nebreda, Xavier Salvatella

Correspondence

xavier.salvatella@irbbarcelona.org

In Brief

Identifying ways to inhibit the androgen receptor (AR) is key for developing treatments for castration-resistant prostate cancer. Here, De Mol, Szulc et al. show that AR activity relies on transient interactions of a disordered motif with the transcription machinery and suggest therapeutic strategies for this disease.

De Mol et al., 2018, Structure *26*, 145–152
January 2, 2018
https://doi.org/10.1016/j.str.2017.11.007

CellPress

Structure
Short Article

Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators

Eva De Mol,[1,8] Elzbieta Szulc,[1,8] Claudio Di Sanza,[1] Paula Martínez-Cristóbal,[1] Carlos W. Bertoncini,[1,6] R. Bryn Fenwick,[1,7] Marta Frigolé-Vivas,[1] Marianela Masín,[1,6] Irene Hunter,[2] Víctor Buzón,[1] Isabelle Brun-Heath,[1] Jesús García,[1] Gianni De Fabritiis,[3,4] Eva Estébanez-Perpiñá,[5] Iain J. McEwan,[2] Ángel R. Nebreda,[1,4] and Xavier Salvatella[1,4,9,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain

[2]Institute of Medical Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, IMS Building, Foresterhill, Aberdeen AB25 2ZD, UK

[3]Computational Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), Doctor Aiguader 88, 08003 Barcelona, Spain

[4]ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain

[5]Departament de Bioquímica i Biomedicina Molecular, Universitat de Barcelona and Institute of Biomedicine of the University of Barcelona (IBUB), Baldiri Reixac 15-21, 08028 Barcelona, Spain

[6]Present address: Instituto de Biología Molecular y Celular de Rosario (IBR), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ocampo y Esmeralda, Rosario, Argentina

[7]Present address: Department of Integrative Structural and Computational Biology and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[8]These authors contributed equally

[9]Lead Contact

*Correspondence: xavier.salvatella@irbbarcelona.org

https://doi.org/10.1016/j.str.2017.11.007

SUMMARY

The androgen receptor is a transcription factor that plays a key role in the development of prostate cancer, and its interactions with general transcription regulators are therefore of potential therapeutic interest. The mechanistic basis of these interactions is poorly understood due to the intrinsically disordered nature of the transactivation domain of the androgen receptor and the generally transient nature of the protein-protein interactions that trigger transcription. Here, we identify a motif of the transactivation domain that contributes to transcriptional activity by recruiting the C-terminal domain of subunit 1 of the general transcription regulator TFIIF. These findings provide molecular insights into the regulation of androgen receptor function and suggest strategies for treating castration-resistant prostate cancer.

INTRODUCTION

The activation of transcription relies on interactions between specific transcription factors and general transcription regulators that can be mediated by transcriptional co-activators (Fuda et al., 2009). It is important to characterize them because their inhibition by small molecules or other biological tools offers opportunities for therapeutic intervention in many disease areas, including oncology (Darnell, 2002). Since they involve intrinsically disordered (ID) transactivation domains, the associated complexes are however transient, marginally stable, and challenging to study (Wright and Dyson, 2015).

One case where inhibiting these interactions is appealing is castration-resistant prostate cancer (CRPC). This condition affects prostate cancer patients who are refractory to hormone therapy, which is based on preventing the activation of the androgen receptor (AR). The mechanisms that allow cell proliferation under these conditions are not yet fully characterized, but it is becoming clear that they include expression of constitutively active AR isoforms lacking the ligand binding domain (Robinson et al., 2015).

The complexes formed by the transactivation domain of AR (Lavery and McEwan, 2008a) and general transcription regulators are targets to interfere with CRPC (Sadar, 2011) because inhibiting their formation can lead to decreases in AR transcriptional activity and in the proliferation of prostate cancer cells. Here, we report the structural basis for the interaction of the transactivation domain of AR and the C-terminal domain of subunit 1 of the general transcription regulator TFIIF (RAP74-CTD), which involves the folding upon binding of a short motif in this receptor and contributes to transcriptional activity (Choudhry et al., 2006; McEwan and Gustafsson, 1997).

RESULTS

A Motif in Transcriptional Activation Unit 5 of AR Adopts a Helical Conformation to Recruit RAP74-CTD

To identify the regions of the AR involved in recruiting RAP74-CTD, we used solution nuclear magnetic resonance (NMR). NMR is appropriate for characterizing protein-protein interactions involving ID proteins because it provides residue-specific

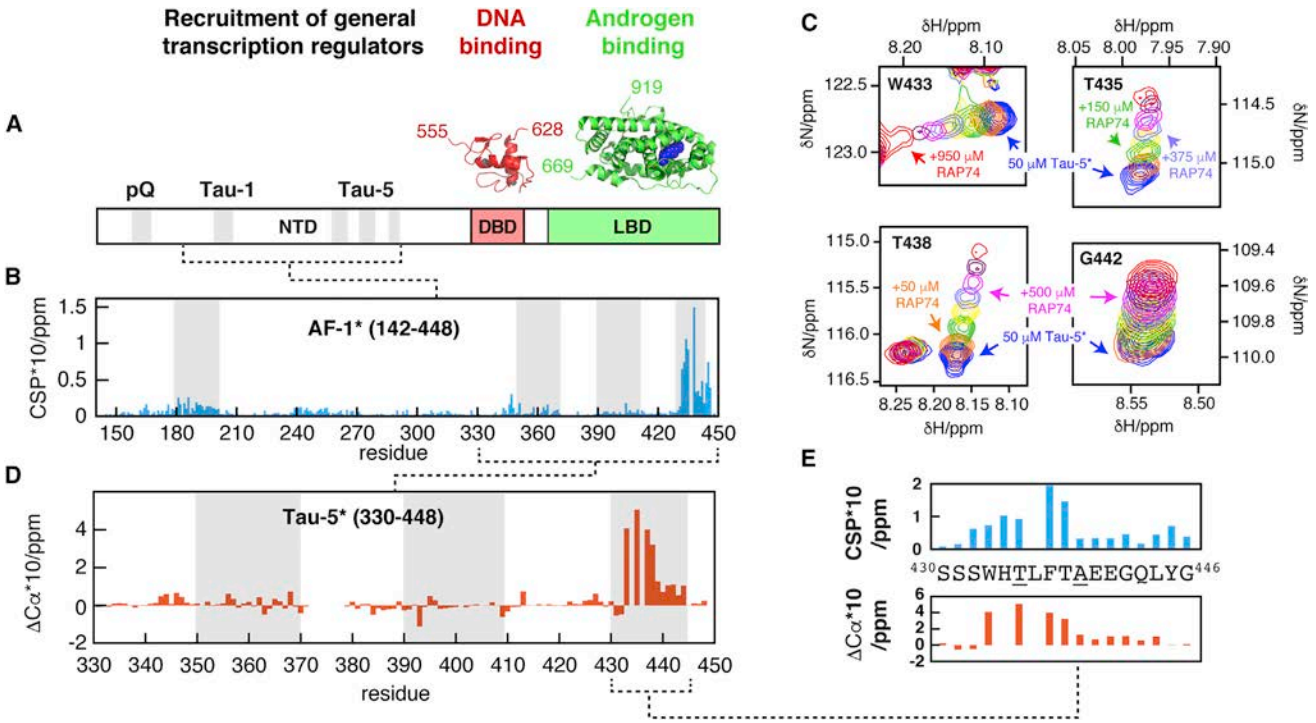


Figure 1. Identification of a Short Motif in AR that Recruits RAP74 by Adopting a Helical Structure

(A) Domain structure of AR indicating the regions that form the transactivation (NTD), DNA binding (DBD), and ligand binding (LBD) domains, as well as the polyGln tract (pQ) and the various partially folded motifs of the NTD (in gray) defined as those with a locally high transverse ^{15}N relaxation rate in NMR experiments (De Mol et al., 2016).

(B) Plot of the chemical shift perturbations (CSP) caused by 500 μM RAP74-CTD on the resonances of 50 μM AF-1* (residues 142–448) as a function of residue number ($CSP = \sqrt{\Delta\delta(H)^2 + (\Delta\delta(N)/5)^2}$) with an indication of the partially folded motifs (in gray).

(C) Changes in the resonances corresponding to four residues of Tau-5* (residues 330–448) at 50 μM titrated with 50 (orange), 150 (green), 250 (yellow), 375 (blue), 500 (pink), 710 (purple), and 950 (red) μM RAP74-CTD.

(D) Plot of the changes in ^{13}Cα chemical shift (ΔCα) caused by 500 μM RAP74-CTD* on the resonances of 100 μM Tau-5* with an indication of the partially folded motifs (in gray).

(E) Detail of the CSP and the ΔCα values obtained for specific residues in the interaction motif of AR with an indication, as underlined, of the positions used for hydrocarbon stapling (see below and Figure 3D).

See also Figure S1.

information in the absence of the long-range order required for crystallization (Dyson and Wright, 2004). In addition, it is well suited for the characterization of weak protein-protein interactions, which can occur when one of the partners is ID (Wright and Dyson, 2009).

We used a construct of the AR transactivation domain (AF1*, AR residues 142–448, Figure 1A) that contains two known functional subdomains, transcriptional activation units 1 and 5 (Tau-1 and -5) (Callewaert et al., 2006). AF1* is ID with regions of helical propensity in the structurally independent Tau-1 and Tau-5 subdomains (De Mol et al., 2016). We measured 2D ^{1}H,^{15}N heteronuclear single quantum coherence (HSQC) NMR spectra of AF-1* in the presence and in the absence of RAP74-CTD (RAP74 residues 450–517) (Lavery and McEwan, 2008a) and observed chemical shift perturbations in a region of Tau-5 with the sequence ^{431}SSWHTLFTAEEGQLYG^{446} (Figure 1B). To confirm that the interaction does not involve residues in Tau-1, we repeated the experiments with a shorter AR construct (Tau-5*, AR residues 330–448) and obtained an equivalent result (Figure S1A).

In order to estimate the stability of the complex, we performed a titration of Tau-5* with RAP74-CTD at 278 K (Figure 1C) and found that the affinity between these two proteins was in the millimolar range (Figure S1B). This is in agreement with the notion that the protein-protein interactions that activate transcription are weak due to their multivalent and transient nature (Melcher, 2000; Uesugi et al., 1997). To investigate whether binding of RAP74-CTD induces a conformational change in Tau-5, we compared the ^{13}Cα chemical shifts of Tau-5* in the presence and in the absence of its binding partner (RAP74-CTD*, see STAR Methods) by using 3D HNCA NMR experiments (Figure 1D). We observed increases in ^{13}Cα chemical shift in several residues of the motif, in agreement with the induction of a helical conformation (Neal et al., 2003), which were particularly large (ca. 0.5 ppm) for residues S432 to T438, which define two turns of an α helix (Figure 1E).

To identify the binding site of AR on the surface of RAP74-CTD, we performed ^{1}H,^{15}N HSQC experiments using RAP74-CTD (Nguyen et al., 2003a) and a peptide with sequence Ac-^{426}SA AASSSWHTLFTAEEGQLYG^{446}-NH_2. We observed chemical

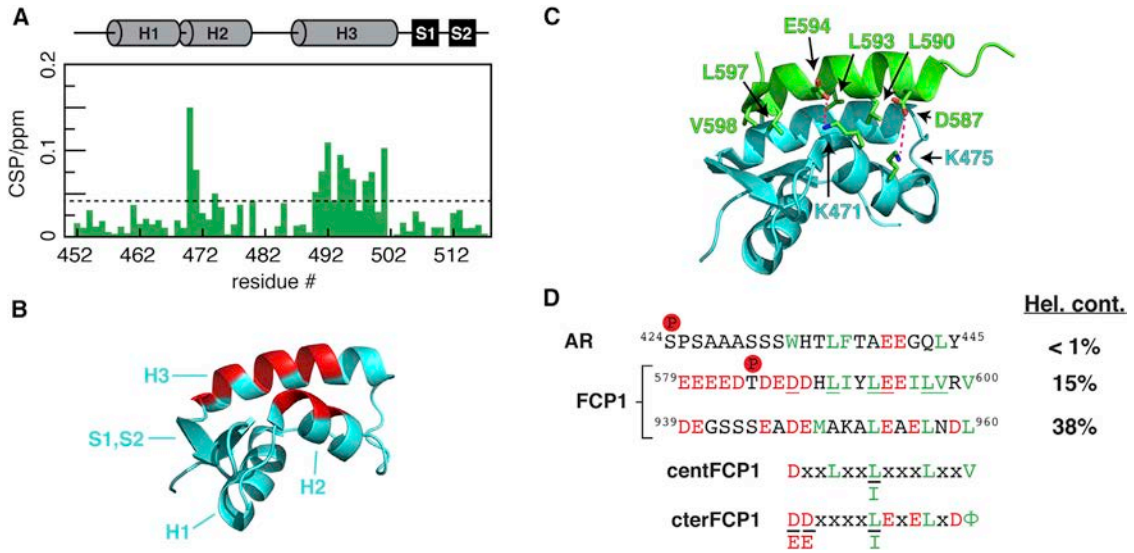


Figure 2. AR Binds to the Same Groove of RAP74-CTD as FCP1

(A) Plot of the chemical shift perturbations (CSP) caused by a peptide spanning the sequence of the AR motif (Ac-^{426}SAAASSSWHTLFTAEEGQLYG^{446}-NH_2) on the resonances of RAP74-CTD as a function of residue number with an indication, with a horizontal dashed line, of the threshold used for preparing (B).

(B) Solution structure of apo RAP74-CTD (Nguyen et al., 2003a) (PDB: 1NHA) indicating, in red, the residues whose resonances are most affected by binding of the AR motif (CSP >0.04 ppm).

(C) Structure of the complex formed by RAP74-CTD and the central motif of FCP1 (Yang et al., 2009) with an indication of the residues that are key for binding and of the main electrostatic interactions, shown in purple (PDB: 2K7L).

(D) Alignment of the sequences of the FCP1 and AR motifs interacting with RAP74-CTD indicating the acidic residues (in red), the hydrophobic residues (in green), the phosphosites identified so far (with the symbol P), the helical propensity predicted by Agadir (Muñoz and Serrano, 1994), and the consensus sequences centFCP1 and cterFCP1. The residues that are underlined in (D) correspond to those represented as sticks in (C).

See also Figure S2.

shift perturbations in helices H2 and H3 (Figures 2A and 2B), which define the binding groove of two ID motifs of FCP1 that fold into an α helix upon binding (Figure 2C) (Kamada et al., 2001, 2003; Nguyen et al., 2003a, 2003b; Yang et al., 2009). FCP1 is a nuclear phosphatase that dephosphorylates the C-terminal domain of RNA polymerase II and is recruited by RAP74-CTD at the termination of transcription (Archambault et al., 1997).

An analysis of the sequences of the disordered motifs of FCP1 (Figure 2D) and their conservation across species shows that the RAP74-CTD groove can accommodate motifs when these fulfill the requirements summarized in two consensus sequences (centFCP1 and cterFCP1) (Abbott et al., 2005; Yang et al., 2009). This emphasizes that the interaction with RAP74 relies on electrostatic interactions involving acidic residues at the N terminus and at the center of the motif and on hydrophobic interactions involving residues at relative positions *i/i+3/i+4*, which are buried in the interface (Figure 2C) (Kamada et al., 2001, 2003; Nguyen et al., 2003a, 2003b; Yang et al., 2009).

The AR motif partially fulfills the requirements summarized in centFCP1 and cterFCP1 (Figure 2D). It possesses hydrophobic residues in positions *i/i+3/i+4* (W433, L436, and F437) that can interact with the groove defined by helices H2 and H3 of RAP74-CTD and two acidic residues (E440 and E441) that can interact with the basic ones that surround the binding site (K471, K475, and K510, Figure 2C). To confirm their relevance for binding, we measured the chemical shift perturbations caused by two peptides, one with the three hydrophobic residues of the motif at positions *i/i+3/i+4* mutated to Ala (AHTAA, Ac-^{426}SAAASSSAHTAATAEEGQLYG^{446}-NH_2) and another with the two acidic residues mutated to Lys (KK, Ac-^{426}SAAASSSWHTLFTAKKGQLYG^{446}-NH_2). In both cases, we observed no chemical shift perturbations, confirming that these two features are key for binding (Figures S2A and S2B).

Helical Propensity and Phosphorylation State Determine the Stability of the Complex

The interaction between specific transcription factors and general transcription regulators can be enhanced by the binding of transcriptional co-activators (Fuxreiter et al., 2008) and by post-translational modifications (Gioeli and Paschal, 2012). The former can induce secondary structures in transactivation domains to facilitate their interaction with the basal transcription machinery (Lavery and McEwan, 2008a), and the latter can either stabilize the structural changes induced by binding (Bah et al., 2015) or directly stabilize the relevant complex (Bah and Forman-Kay, 2016). We used NMR to measure the affinity between RAP74-CTD and chemically modified peptides. This experimental setup allowed us to mimic the site-specific phosphorylations that occur during AR activation as well as by hydrocarbon stapling (Schafmeister et al., 2000) the helical secondary structure induced for example by co-activators.

The regions at the N terminus of the FCP1 motifs are rich in acidic side chains. However, the equivalent region in AR, which binds with lower affinity, is instead rich in Ser residues (^{424}SPSAAASSS^{432}, Figure 2D). We hypothesized that phosphorylations of this region contribute to stabilizing the transient

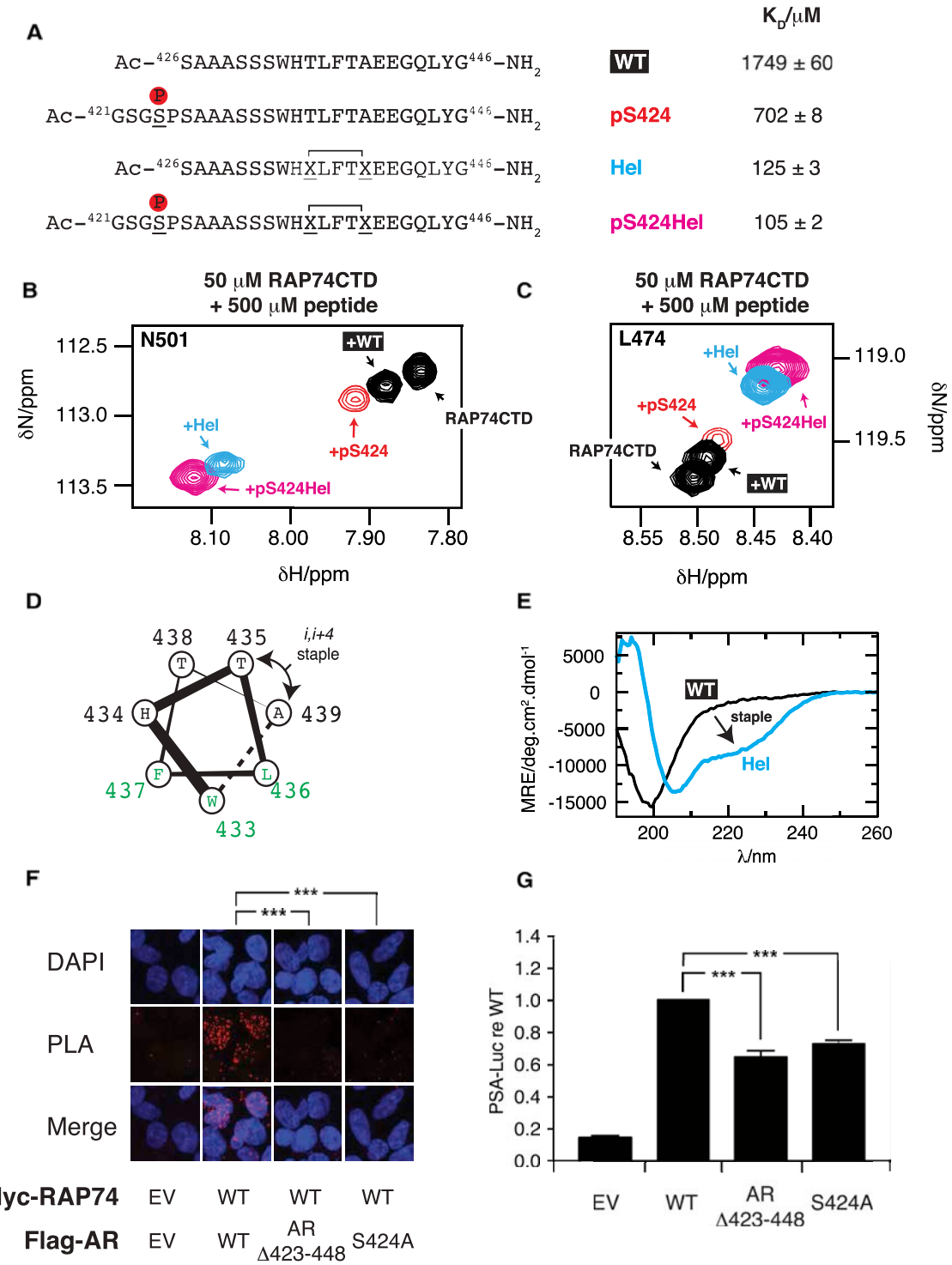


Figure 3. Determinants of the Interaction between AR and RAP74-CTD

(A) Sequences of peptides derived from the AR motif that recruits TFIIF, indicating affinities for RAP74-CTD (K_D). In Hel and pS424Hel, X represents the amino acid (*S*)-2-(4′-pentenyl)alanine and a continuous line links the residues stapled.

(B and C) Regions of the $^{1}H,^{15}N$ HSQC NMR spectrum of RAP-CTD illustrating the chemical shift perturbations caused by the peptides listed in (A) in residues N501 (B) and L474 (C) of RAP74-CTD.

(D) Helical wheel representation of the hydrophobic residues of the AR motif that recruits TFIIF with an illustration of the residues replaced by (*S*)-2-(4′-pentenyl) alanine and used for *i*,*i*+*4* stapling.

(E) CD spectrum of 50 μM solutions of the peptides WT and Hel.

(legend continued on next page)

complex that it forms with RAP74-CTD. An analysis of the known phosphosites of AR revealed that S424 is phosphorylated upon AR activation (Gioeli and Paschal, 2012). To determine whether this increases the stability of the transient complex, we measured the affinity for RAP74-CTD of a peptide phosphorylated on this position (pS424, Figure 3A–3C). The results indicated that phosphorylation of Ser 424 increased the affinity of the peptide from K_D = 1749 ± 60 μM to K_D = 702 ± 8 μM.

As shown in Figure 2D, an additional difference between the FCP1 and AR motifs is helical propensity. Whereas the central and C-terminal motifs of FCP1 have some helical propensity according to the predictor Agadir (Muñoz and Serrano, 1994) (15% and 38%), the AR motif has not (<1%), in agreement with our characterization of the structural properties of the NTD of AR (De Mol et al., 2016). Given that binding to the groove defined by helices H2 and H3 of RAP74-CTD involves the adoption of a helical conformation by the motif (Figures 1D, 1E, and 2C), we hypothesized that the low helical propensity of the AR motif contributes to its low affinity.

To determine the effect of increasing the helicity, we used hydrocarbon stapled peptides (Hel and pS424Hel, Figure 3A) where residues Thr 435 and Ala 439, which are in the face of the helix opposite the hydrophobic residues that interact with RAP74-CTD (Figure 3D), were replaced by (*S*)-2-(4′-pentenyl) alanine and stapled by olefin metathesis (Schafmeister et al., 2000). A comparison of the secondary structure of the wild-type (WT) and Hel peptides by circular dichroism (CD) (Figure 3E) confirmed that stapling indeed increased helical propensity. We analyzed the chemical shift perturbations caused in RAP74-CTD by Hel and pS424Hel, a stapled peptide including also the phosphorylation at S424, and confirmed that both interacted with higher affinity compared with their non-stapled counterparts (K_D = 125 ± 3 μM for Hel and K_D = 105 ± 2 μM, for pS424Hel, Figures 3A–3C). This confirms that phosphorylation facilitates binding of the AR motif and that processes that increase helical propensity can enhance AR transcriptional activity.

The Interaction Can Be Observed in Cells and Contributes to AR Transcriptional Activity

Although it is well established that TFIIF and the RAP74-CTD domain in particular interact with the AF1 domain of AR *in vitro* (Kumar et al., 2004; Lavery and McEwan, 2008b; McEwan and Gustafsson, 1997; Reid et al., 2002), there is little evidence that the interaction occurs in cells. To investigate this, we used biochemical techniques, such as co-immunoprecipitation, but failed to detect robust interaction presumably due to its transient nature. We next used the proximity ligation assay (PLA) (Söderberg et al., 2006), an immunofluorescence-based technique that allows the detection of proteins in close proximity inside cells.

For these studies, we tagged full-length AR and RAP74-CTD with Flag and Myc, respectively, transfected them in HEK293T cells, which do not express AR (see STAR Methods), and treated the cells with dihydrotestosterone (DHT) to cause the activation of the receptor. The results indicated that the two proteins interact (Figure 3F). To validate that the interaction takes place via the identified AR motif (AR 423–448), we carried out equivalent experiments with a mutant of AR with residues 423 to 448 removed (ARΔ423-448). In agreement with the results obtained *in vitro*, we observed a reduction of the interaction between ARΔ423-448 and RAP74-CTD in cells of ca. 40% (Figures 3F and S3D). We also investigated by PLA whether the phosphorylation of Ser 424 regulates the interaction with RAP74-CTD in cells by using an AR with Ser 424 mutated to Ala (S424A) and found that this was indeed the case (Figures 3F and S3D). Importantly, control immunofluorescence experiments showed that mutants ARΔ423–448 and S424A, like WT AR, are expressed and translocate to the nucleus upon activation by DHT (Figures S3A and S3B).

Finally, to further assess the functional relevance of the interaction between AR and RAP74-CTD via the motif identified in this work, we measured the transcriptional activity of WT AR and mutants ARΔ423–448 and S424A in HEK293T cells by means of a gene reporter assay (see STAR Methods). The results showed that deleting the motif or removing the phosphorylation site at position 424 lowered the transcriptional activity of AR by ca. 30% (Figures 3G and S3C).

DISCUSSION

The interactions between transactivation domains of specific transcription factors and transcriptional co-activators or general transcription regulators are among the best characterized examples of complexes involving ID proteins (Brzovic et al., 2011; Di Lello et al., 2006; Feng et al., 2009; Uesugi et al., 1997). Key features of these, which are present in the interaction studied here, are the induction of secondary structure upon interaction, their relatively weak nature, and the important role played by post-translational modifications in their regulation (Fuxreiter et al., 2008).

The interaction between AR and the C-terminal domain of subunit 1 of TFIIF is mediated by hydrophobic interactions between residues at positions *i/i+3/i+4* of the AR motif and a hydrophobic cleft on the surface of RAP74-CTD, with an important contribution of electrostatic interactions. This relative position of hydrophobic residues in the AR motif is common in transactivation domains, indicating that there could be a generic mechanism by which these domains recruit their binding partners and highlighting the general importance of regulatory mechanisms to provide specificity. We provide evidence that the phosphorylation of Ser 424 is important for the interaction between AR and RAP74-CTD and for transcriptional activity, illustrating how post-translational modifications can enhance the affinity and the specificity of ID proteins for their binding partners (Stein and Aloy, 2008).

Our results indicate that AR and FCP1 interact with the same groove in the structure of RAP74-CTD via similar motifs. The interaction between AR and RAP74-CTD is, as we have shown, important for transcriptional activity, whereas that between the latter and FCP1 is important for transcription termination (Archambault et al., 1997). The role of FCP1 in termination, which is carried out by its phosphatase domain, is to dephosphorylate the C-terminal

(F and G) The effect of deleting the AR motif (Δ423–448) and mutating Ser 424 to Ala (S424A) was assessed in HEK293T cells treated with 1 nM DHT by PLA to measure the interaction between AR and RAP74-CTD (F, see Figure S3D for a quantitative analysis) and by a reporter assay to measure AR transcriptional activity (G). In (F), EV stands for empty vector, and DAPI indicates the location of nuclei. In (G), the error bars represent the SE; ***$p < 0.001$.
See also Figure S3.

tail of RNA polymerase II, causing it to dissociate from the DNA and therefore allowing it to become involved in a subsequent round of transcription. We conclude that RAP74-CTD, which is a particularly dynamic part of the transcription machinery and is tethered to it via a very flexible linker (Sainsbury et al., 2015), uses a single binding mechanism to interact with different ID motifs at different stages of the process of transcription.

Our work indicates that the motif ^{431}SSWHTLFTAEEGQ LYG^{446} is important for the formation of the transient complex in cells but that its interaction with RAP74-CTD *in vitro* is weak, even after phosphorylation of Ser 424, unless a helical conformation is induced. Several mechanisms to induce a helical conformation in the motif can operate in cells, including an allosteric mechanism coupling the DNA-binding and transactivation domains upon DNA binding (Brodie and McEwan, 2005), a change in the conformation of the motif or the whole Tau-5 sub-domain after co-activator binding (Fuxreiter et al., 2008), or the effect of extrinsic factors that cannot be easily accounted for by *in vitro* studies. It is in fact possible that several of these mechanisms operate simultaneously as this would provide a means of regulating transcriptional activity (Hilser and Thompson, 2011; Wu and Fuxreiter, 2016).

The relevance of Tau-5, the sub-domain of the AR NTD where the ^{431}SSWHTLFTAEEGQLYG^{446} motif is found, for transcriptional activity in cells depends on the cell line used for the experiments and on the concentration of androgens to which the cells are exposed. Tau-5 inhibits transcriptional activity in prostate cancer cell lines expressing AR in the presence of physiological concentrations of androgens (Dehm et al., 2007). By contrast, it stimulates transcriptional activity in cell lines that do not express AR (Jenster et al., 1995) and, most importantly, in CRPC cell lines expressing AR in the absence of androgens or in their presence at castrate levels, where residues 433–437, at the core of the motif identified in this work, can act as an independent transactivation domain (Dehm et al., 2007). The reasons for this difference in behavior are currently not known and further work will be necessary to determine whether it is due, for example, to the action of specific co-regulators or to differences in post-translational modifications in AR.

Nevertheless, from a translational medicine point of view, our results and those available in the literature indicate that the motif that recruits RAP74-CTD can contribute to transcription activation by AR and, therefore, that the complex that it forms with this subunit of TFIIF is a potential therapeutic target for CRPC, although we cannot exclude the possibility that other interactions contribute to its function in transcription activation (Cato et al., 2017; He et al., 2000; Li et al., 2014). In summary, although protein-protein interactions involving ID proteins represent challenging targets for drug discovery, our work indicates that inhibitors of the recruitment of RAP74-CTD by AR, which could be either small molecules or peptides, could lead to treatments for CRPC (Yap et al., 2016).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Protein Expression and Purification
  - Handling of Peptides
  - NMR
  - Fitting of the NMR Data to Obtain the Values of K_D
  - Cell Culture and Transfection
  - Plasmids
  - Western Blot Analyses
  - Proximity Ligation Assay (PLA)
  - Localization Studies
  - Transcriptional Activity Assay
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and can be found with this article online at https://doi.org/10.1016/j.str.2017.11.007.

AUTHOR CONTRIBUTIONS

E.D.M., E.S., C.D.S., P.M.-C., and C.W.B. designed, carried out, analyzed and interpreted the experiments, and contributed to writing the manuscript. R.B.F., M.F.-V., M.M., I.H., V.B., J.G., G.D.F., and E.E.-P. provided experimental support. I.B.-H., I.J.M., and A.R.N. contributed to analyzing and interpreting the experiments as well as to writing the manuscript. X.S. designed and contributed to interpreting the experiments as well as to writing the manuscript.

ACKNOWLEDGMENTS

The authors thank Miquel Pons, Philipp Selenko, and Peter Wright for helpful discussions as well as Joan Miquel Valverde (IRB) and the ICTS NMR facility, managed by the scientific and technological centers of the University of Barcelona (CCiT UB), for their help in protein production and NMR. They also thank the IRB protein expression core facility for its assistance in cloning, the IRB advanced microscopy facility for assistance in immunofluorescence and PLA, and the IRB biostatistics and bioinformatics facility for assistance in the analysis of the gene reporter and PLA data. This work was supported by IRB, ICREA (X.S.), Obra Social "la Caixa" (E.D.M., E.S., and X.S.), MINECO (BIO2012-31043 and BIO2015-70092-R to X.S., BIO2014-53095-P to G.D.F.), Marató de TV3 (102030 to X.S. and 102031 to E.E.-P), the COFUND program of the European Commission (C.D.S.), the European Research Council (CONCERT, contract number 648201 to X.S.), the Ramón y Cajal program of MICINN (RYC-2011-07873 to C.W.B.), the Serra Hunter Program (E.E.-P.), AGAUR (SGR-2014-56RR14 to E.E.-P), and FEDER (G.D.F.). I.H. was supported by funding from the Chief Scientist's Office of the Scottish Government (ETM-258, ETM-382). IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO (Government of Spain).

Received: August 1, 2017
Revised: October 12, 2017
Accepted: November 10, 2017
Published: December 7, 2017

REFERENCES

Abbott, K.L., Renfrow, M.B., Chalmers, M.J., Nguyen, B.D., Marshall, A.G., Legault, P., and Omichinski, J.G. (2005). Enhanced binding of RNAP II CTD phosphatase FCP1 to RAP74 following CK2 phosphorylation. Biochemistry *44*, 2732–2745.

Archambault, J., Chambers, R.S., Kobor, M.S., Ho, Y., Cartier, M., Bolotin, D., Andrews, B., Kane, C.M., and Greenblatt, J. (1997). An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *94*, 14300–14305.

Bah, A., and Forman-Kay, J.D. (2016). Modulation of intrinsically disordered protein function by post-translational modifications. J. Biol. Chem. *291*, 6696–6705.

Bah, A., Vernon, R.M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L.E., and Forman-Kay, J.D. (2015). Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature *519*, 106–109.

Brodie, J., and McEwan, I.J. (2005). 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.

Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E., et al. (2011). The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. Mol. Cell *44*, 942–953.

Callewaert, L., Van Tilborgh, N., and Claessens, F. (2006). Interplay between two hormone-independent activation domains in the androgen receptor. Cancer Res. *66*, 543–553.

Cato, L., Neeb, A., Sharp, A., Buzón, V., Ficarro, S.B., Yang, L., Muhle-Goll, C., Kuznik, N.C., Riisnaes, R., Nava Rodrigues, D., et al. (2017). Development of Bag-1L as a therapeutic target in androgen receptor-dependent prostate cancer. Elife *6*, e27159.

Choudhry, M.A., Ball, A., and McEwan, I.J. (2006). The role of the general transcription factor IIF in androgen receptor-dependent transcription. Mol. Endocrinol. *20*, 2052–2061.

Darnell, J.E., Jr. (2002). Transcription factors as targets for cancer therapy. Nat. Rev. Cancer *2*, 740–749.

Dehm, S.M., Regan, K.M., Schmidt, L.J., and Tindall, D.J. (2007). Selective role of an NH2-terminal WxxLF motif for aberrant androgen receptor activation in androgen depletion independent prostate cancer cells. Cancer Res. *67*, 10067–10077.

De Mol, E., Fenwick, R.B., Phang, C.T.W., Buzón, V., Szulc, E., de la Fuente, A., Escobedo, A., García, J., Bertoncini, C.W., Estébanez-Perpiñá, E., et al. (2016). EPI-001, a compound active against castration-resistant prostate cancer, targets transactivation unit 5 of the androgen receptor. ACS Chem. Biol. *11*, 2499–2505.

Di Lello, P., Jenkins, L.M.M., Jones, T.N., Nguyen, B.D., Hara, T., Yamaguchi, H., Dikeakos, J.D., Appella, E., Legault, P., and Omichinski, J.G. (2006). Structure of the Tfb1/p53 complex: insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. Mol. Cell *22*, 731–740.

Dyson, H.J., and Wright, P.E. (2004). Unfolded proteins and protein folding studied by NMR. Chem. Rev. *104*, 3607–3622.

Feng, H., Jenkins, L.M.M., Durell, S.R., Hayashi, R., Mazur, S.J., Cherry, S., Tropea, J.E., Miller, M., Wlodawer, A., Appella, E., et al. (2009). Structural basis for p300 Taz2-p53 TAD1 binding and modulation by phosphorylation. Structure *17*, 202–210.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.

Fuxreiter, M., Tompa, P., Simon, I., Uversky, V.N., Hansen, J.C., and Asturias, F.J. (2008). Malleable machines take shape in eukaryotic transcriptional regulation. Nat. Chem. Biol. *4*, 728–737.

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

He, B., Kemppainen, J.A., and Wilson, E.M. (2000). FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. J. Biol. Chem. *275*, 22986–22994.

Hilser, V.J., and Thompson, E.B. (2011). Structural dynamics, intrinsic disorder, and allostery in nuclear receptors as transcription factors. J. Biol. Chem. *286*, 39675–39682.

Jenster, G., van der Korput, H.A., Trapman, J., and Brinkmann, A.O. (1995). Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. J. Biol. Chem. *270*, 7341–7346.

Kamada, K., De Angelis, J., Roeder, R.G., and Burley, S.K. (2001). Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. Proc. Natl. Acad. Sci. USA *98*, 3115–3120.

Kamada, K., Roeder, R.G., and Burley, S.K. (2003). Molecular mechanism of recruitment of TFIIF- associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. Proc. Natl. Acad. Sci. USA *100*, 2296–2299.

Kumar, R., Betney, R., Li, J., Thompson, E.B., and McEwan, I.J. (2004). Induced alpha-helix structure in AF1 of the androgen receptor upon binding transcription factor TFIIF. Biochemistry *43*, 3008–3013.

Lavery, D.N., and McEwan, I.J. (2008a). Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding. Biochemistry *47*, 3360–3369.

Lavery, D.N., and McEwan, I.J. (2008b). 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*, 3352–3359.

Li, N., Chen, M., Truong, S., Yan, C., and Buttyan, R. (2014). Determinants of Gli2 co-activation of wildtype and naturally truncated androgen receptors. Prostate *74*, 1400–1410.

McEwan, I.J., and Gustafsson, J. (1997). Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF. Proc. Natl. Acad. Sci. USA *94*, 8485–8490.

Melcher, K. (2000). The strength of acidic activation domains correlates with their affinity for both transcriptional and non-transcriptional proteins. J. Mol. Biol. *301*, 1097–1112.

Muñoz, V., and Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. Nat. Struct. Biol. *1*, 399–409.

Neal, S., Nip, A.M., Zhang, H., and Wishart, D.S. (2003). Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. J. Biomol. NMR *26*, 215–240.

Nguyen, B.D., Chen, H.-T., Kobor, M.S., Greenblatt, J., Legault, P., and Omichinski, J.G. (2003a). Solution structure of the carboxyl-terminal domain of RAP74 and NMR characterization of the FCP1-binding sites of RAP74 and human TFIIB. Biochemistry *42*, 1460–1469.

Nguyen, B.D., Abbott, K.L., Potempa, K., Kobor, M.S., Archambault, J., Greenblatt, J., Legault, P., and Omichinski, J.G. (2003b). NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. Proc. Natl. Acad. Sci. USA *100*, 5688–5693.

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

Robinson, D., Van Allen, E.M., Wu, Y.-M., Schultz, N., Lonigro, R.J., Mosquera, J.-M., Montgomery, B., Taplin, M.-E., Pritchard, C.C., Attard, G., et al. (2015). Integrative clinical genomics of advanced prostate cancer. Cell *161*, 1215–1228.

Sadar, M.D. (2011). Small molecule inhibitors targeting the "Achilles' heel" of androgen receptor activity. Cancer Res. *71*, 1208–1213.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 129–143.

Schafmeister, C.E., Po, J., and Verdine, G.L. (2000). An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides. J. Am. Chem. Soc. *122*, 5891–5892.

Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.-J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.-G., et al. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. Nat. Methods *3*, 995–1000.

Stein, A., and Aloy, P. (2008). Contextual specificity in peptide-mediated protein interactions. PLoS One *3*, e2524.

Stenoien, D.L., Cummings, C.J., Adams, H.P., Mancini, M.G., Patel, K., DeMartino, G.N., Marcelli, M., Weigel, N.L., and Mancini, M.A. (1999).

Polyglutamine-expanded androgen receptors form aggregates that sequester heat shock proteins, proteasome components and SRC-1, and are suppressed by the HDJ-2 chaperone. Hum. Mol. Genet. *8*, 731–741.

Uesugi, M., Nyanguile, O., Lu, H., Levine, A.J., and Verdine, G.L. (1997). Induced alpha helix in the VP16 activation domain upon binding to a human TAF. Science *277*, 1310–1313.

Vandepoele, K., Van Roy, N., Staes, K., Speleman, F., and van Roy, F. (2005). A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. Mol. Biol. Evol. *22*, 2265–2274.

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

Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. *16*, 18–29.

Wu, H., and Fuxreiter, M. (2016). The structure and dynamics of higher-order assemblies: amyloids, signalosomes, and granules. Cell *165*, 1055–1066.

Yang, A., Abbott, K.L., Desjardins, A., Di Lello, P., Omichinski, J.G., and Legault, P. (2009). NMR structure of a complex formed by the carboxyl-terminal domain of human RAP74 and a phosphorylated peptide from the central domain of the FCP1 phosphatase. Biochemistry *48*, 1964–1974.

Yap, T.A., Smith, A.D., Ferraldeschi, R., Al-Lazikani, B., Workman, P., and de Bono, J.S. (2016). Drug discovery in advanced prostate cancer: translating biology into therapy. Nat. Rev. Drug Discov. *15*, 699–718.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-β-actin-HRP	Abcam	Ab8224; RRID: AB_449644
Anti-AR (N-20)	Santa Cruz	sc-816; RRID: AB_1563391
Rabbit-anti-Flag	Sigma	F7425; RRID: AB_439687
Mouse-anti-Myc	Abcam	ab32; RRID: AB_303599
Goat anti-rabbit Alexa-Fluor 568	Invitrogen	A11036; RRID: AB_10563566
Bacterial and Virus Strains		
Rosetta™(DE3)pLysS Competent Cells	Novagen	Cat# 70956
Chemicals, Peptides, and Recombinant Proteins		
Recombinant protein: human AR AF-1* (aa 142-448, ref# P10275)	(De Mol et al., 2016)	N/A
Recombinant protein: human AR Tau-5* (aa 330-448, ref# P10275)	(De Mol et al., 2016)	N/A
Recombinant protein: human RAP74-CTD (aa 450-517, ref# NP35269)	This paper	N/A
Recombinant protein: human RAP74-CTD* (aa 363-517, ref# NP35269)	(Lavery and McEwan, 2008b)	N/A
Synthetic WT peptide: Ac-SAAASSSWHTLFTAEEGQLYG-NH2	Synthesized by ICTS NANBIOSIS	N/A
Synthetic AHTAA peptide: Ac-SAAASSSAHTAATAEE GQLYG-NH2	Synthesized by ICTS NANBIOSIS	N/A
Synthetic KK peptide: Ac-SAAASSSWHTLFTAKKGQLYG-NH2	Synthesized by ICTS NANBIOSIS	N/A
Synthetic pS424 peptide: Ac-GSGpSPSAAASSSWHTLFTA EEGQLYG-NH2	Synthesized by Genscript	N/A
Synthetic Hel peptide: Ac-SAAASSSWHXLFTXEEGQ LYG446-NH2	Synthesized by Genscript and ICTS NANBIOSIS	N/A
Synthetic pS424Hel peptide: Ac-GSGpSPSAAASSSWHXLFTXE EGQLYG-NH2	Synthesized by ICTS NANBIOSIS	N/A
Dulbecco's Modified Eagle Medium (DMEM)	Gibco	Cat# 41966-052
Fetal Bovin Serum (FBS)	Gibco	Cat#10270-106
Charcoal Stripped FBS (CSS)	Gibco	Cat# 12676011
Critical Commercial Assays		
Duolink In Situ PLA Probe Anti-Rabbit PLUS	Sigma	Cat# DUO92002
Duolink In Situ PLA Probe Anti-Mouse MINUS	Sigma	Cat# DUO92004
Duolink In Situ Detection Reagents Orange	Sigma	Cat# DUO92007
Dual-Luciferase Reporter Assay System	Promega	Cat# E1910
Deposited Data		
NMR assignments of RAP74-CTD	BMRB (www.bmrb.wisc.edu)	27288
Experimental Models: Cell Lines		
Human Embryonic Kidney 293T cells	ATCC	CRL-3216
Recombinant DNA		
pDONR221-RAP74CTD	GeneArt	N/A
pDEST-HisBMP	(Nallamsetty et al., 2005)	Addgene 11085
pDEST-Myc	(Vandepoele et al. 2005)	LMBP 4541
pDEST-Myc-RAP74CTD	This paper	N/A
pEFGFP-C1-AR	(Stenoien et al., 1999)	Addgene 28235
pCMV5-FLAG	Sigma-Aldrich	Cat# E6908
Flag-AR	This paper	N/A
ARΔ423-448	This paper	N/A

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
S424A	This paper	N/A
PSA(6.1)–Luc	Hsieh lab	N/A
pTK-Renilla	Promega	Cat# E224A
Oligonucleotides		
Primers for S424A mutant generation Fw: 5'-GGACCCGGTTCTGGGGCACCCTCAGCCGCCGC-3' Rv: 5'-GCGGCGGCTGAGGGTGCCCCAGAACCGGGTCC-3'	This paper	N/A
Primers for ARΔ423-448 mutant generation Fw: 5'-CGGGACCCGGTTCTGGGTCAGGTGGTGGGGGTG GTGGCGG-3' Rv: 5'-CCGCCACCACCCCCACCACC↓TGACCCAGAACC GGGTCCCG-3'	This paper	N/A
Software and Algorithms		
GraphPad Prism	GraphPad Software (www.graphpad.com)	7.0

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Xavier Salvatella (xavier.salvatella@irbbarcelona.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Rosetta™(DE3)pLysS Competent Cells were used for protein expression. Human Embryonic Kidney 293T cells (HEK293T cells) were maintained in DMEM containing 4.5 g/L D-glucose, pyruvate and L-glutamine (Life Technologies) supplemented with 10% Charcoal-stripped FBS (Life Technologies), 100 g·ml-1 of penicillin and 100g·ml-1 of streptomycin. Cells were cultured in a humidified atmosphere containing 5% CO2 at 37°C.

METHOD DETAILS

Protein Expression and Purification

For the preparation of samples of RAP74-CTD for NMR experiments a synthetic gene corresponding to residues 450 to 517 of RAP74 and including a cleavage site for TEV protease, at the N-terminus, was purchased from GeneArt cloned to a pDONR221vector and sub-cloned into a pDEST-HisMBP vector for protein expression, obtained from Addgene (Nallamsetty et al., 2005). Cells Rosetta cells were grown at 37°C in LB medium for the production of non-isotopically labeled samples. For single (^{15}N) or double (^{15}N,^{13}C) isotopic labeling, cells were grown in minimal MOPS medium containing $^{15}NH_4Cl$ or $^{15}NH_4Cl$ and ^{13}C-glucose, respectively as nitrogen and carbon sources. Protein expression was induced with 1 mM IPTG at OD_{600nm} 0.7. After 3 hours cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10 mM imidazole). The soluble fraction was loaded onto a Ni^{2+} affinity chromatography column (GE) and eluted in lysis buffer with an imidazole gradient. The eluted RAP74-CTD was pooled, concentrated and dialyzed for 16 hours against 50 mM Tris-HCl pH 8.0, 200 mM NaCl. After dialysis, EDTA was added to a final concentration of 0.5 mM and the protein was incubated with TEV protease for 16 hours at 4 °C. The HisMBP moiety and the uncleaved material were removed by reverse Ni^{2+} chromatography, which was followed by cationic exchange and size exclusion chromatography steps. AF-1*, Tau-5* and RAP74-CTD* were produced following procedures reported previously (De Mol et al., 2016; Lavery and McEwan, 2008b).

Handling of Peptides

The lyophilized peptides were dissolved in deionized water or directly in 20 mM sodium phosphate, 0.01% (w/v) NaN_3 and the pH of the resulting solution was adjusted by addition of concentrated NaOH. The concentration of these solutions was determined by amino acid analysis. The CD spectra of 50 μM solutions of peptides WT and Hel in 20 mM sodium phosphate pH 7.4, were measured in a JASCO spectropolarimeter at 293 K by using a 1 mm path length quartz cuvette.

NMR

NMR spectra were recorded on 600 and 800 MHz Bruker Avance spectrometers equipped with cryoprobes. Backbone assignments for RAP74-CTD were obtained using three-dimensional HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCANH and CBCA(CO)NH spectra acquired on a 0.5 mM ^{15}N,^{13}C-double labeled sample and have been deposited in the BMRB with accession code 27288. Chemical

shifts were referenced by using 3-(trimethylsilyl)-1-propanesulfonic acid sodium salt (DSS) as internal reference. $^{1}H,^{15}N$-HSQC and HNCA spectra of the AF-1* and Tau-5* in the presence of increasing amounts of unlabeled RAP74-CTD were obtained at 278 K in 20 mM sodium phosphate (pH 7.4), 1 mM TCEP, 10% D_2O, and 30 μM DSS. To study the interaction of RAP74-CTD with peptides, $^{1}H,^{15}N$-HSQC spectra were acquired at 298 K with samples containing uniformly ^{15}N-labeled RAP74-CTD (50 μM) and the indicated amount of peptide dissolved in 20 mM sodium phosphate, 0.01% (w/v) NaN_3, 30 μM DSS, 10% D_2O at pH 7.4.

Fitting of the NMR Data to Obtain the Values of K_D

The changes in ^{1}H and ^{15}N chemical shift ($\Delta\delta_i$) caused by synthetic peptides in RAP74-CTD were globally fit to the following isotherm by using GraphPad Prism to obtain the value of K_D, where C is the concentration of titrand, n is the ratio between the concentration of titrant and that of titrand and $\Delta\delta_i^{sat}$ is the difference in chemical shift, for nucleus i, between the free and bound titrand.

$$\Delta\delta_i = \frac{\Delta\delta_i^{sat}}{2}\left(1 + K_D/C + n - \sqrt{(1 + n + K_D/C)^2 - 4n}\right)$$

Cell Culture and Transfection

Transient transfection of HEK293T cells was performed with polyethylenimine (PEI, Polysciences) at a ratio of 1 μg DNA to 3 μl PEI.

Plasmids

The sequence coding for AR from the plasmid pEGFP-C1-AR (Stenoien et al., 1999), was subcloned into *Bgl* II/ *Sal* I sites of a pCMV5-FLAG vector to give Flag-AR. The ARΔ423-448 and S424A mutants were generated by site-directed mutagenesis. The sequence coding for RAP74 CTD from the plasmid pDONR221-RAP74 CTD was sub-cloned into a pDEST-Myc vector to give pDEST-Myc-Rap74 CTD.

Western Blot Analyses

Flag-AR wild type as well as AR mutants ARΔ423-448 and S424A were ectopically expressed together with Myc-RAP74-CTD in HEK293T cells. After 48 hours, cells were lysed in hypotonic protein lysis buffer, containing 0.5 % NP-40, 10 mM Tris-HCl pH8, 60 mM KCl, 1 mM EDTA and complete protease inhibitors (Roche). DMSO or 1nM DHT (Sigma) were administered to the medium and the cells treated for 24h. Total cellular lysate was fractionated in mini-Protean TGX 4-20% acrylamide gel (Biorad) and blotted onto a nitrocellulose membrane (Amersham). Protein levels were assessed by means of the following antibodies: anti-β-actin-HRP, anti-Androgen Receptor (N-20), rabbit-anti-Flag, mouse-anti-Myc.

Proximity Ligation Assay (PLA)

HEK293T cells were grown on 12 mm-diameter coverslips (Thermo Scientific) in 6-well plates and transfected with 0.5 μg of each plasmid following a ratio of 1 μg DNA to 3 μl PEI for 48 hours. When mentioned, transfected cells were then incubated with 1 nM DHT for an additional 24 hours. Cells were fixed in PBS containing 4% Paraformaldehyde (EMS) for 10 minutes, subsequently washed in PBS and permeabilized with methanol for 5 minutes. Slides were blocked in PBS containing 0.1% Tween and 2% Bovine Serum Albumin (Sigma), and incubated with Flag and Myc-specific antibodies (Rabbit-anti-Flag and Mouse-anti-Myc). Cells were subsequently incubated with Duolink II PLA probes and stained according to manufacturer's protocol. Cells were analyzed with a 63× objective lens on a Leica SP5 or SPE confocal microscopes.

Localization Studies

HEK293T cells were grown on 12 mm-diameter coverslips and fixed in PBS containing 4% paraformaldehyde, and permeabilized in methanol. Cells were stained with a Flag-specific antibody (Rabbit-anti-Flag) in PBS containing 0.1% Tween, 2% BSA. Subsequently, cells were washed in PBS containing 0.1% Tween and incubated with goat anti-rabbit Alexa-Fluor 568 secondary antibody. After washing with PBS, cells were treated for 5 minutes with 5 mg/ml DAPI and mounted in Prolong Gold antifade (Thermo Fisher Scientific). Cells were analyzed with a 63× objective on a Leica SP5 or SPE confocal microscopes.

Transcriptional Activity Assay

To assess AR-mediated transcriptional activity on the Prostate Specific Antigen (PSA) promoter, HEK293T cells were co-transfected with pCMV5-Flag-AR WT or mutants, pTK-Renilla and PSA (6.1)-Luc plasmids, and 48 hours later were treated with 1 nM DHT for 24 hours. Samples were assayed for luciferase activity using the Dual-Luciferase Reporter Assay System according to manufacturer's instructions. pTK-Renilla was used for normalization of luciferase expression.

QUANTIFICATION AND STATISTICAL ANALYSIS

For the quantification of the PLA results, reported in Figure S3D, the spot count (foci) was used to assess differences between groups and treatments using a general linear model, including image batch of each experimental observation group as random effect.

For the reporter assay, reported in the Figure 3G, a general linear model was used to compare differences in log transformed PSA-Luc vs Renilla ratio between groups of interest using experiment batch, total cell count and replicate as covariates.

4.3 Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments

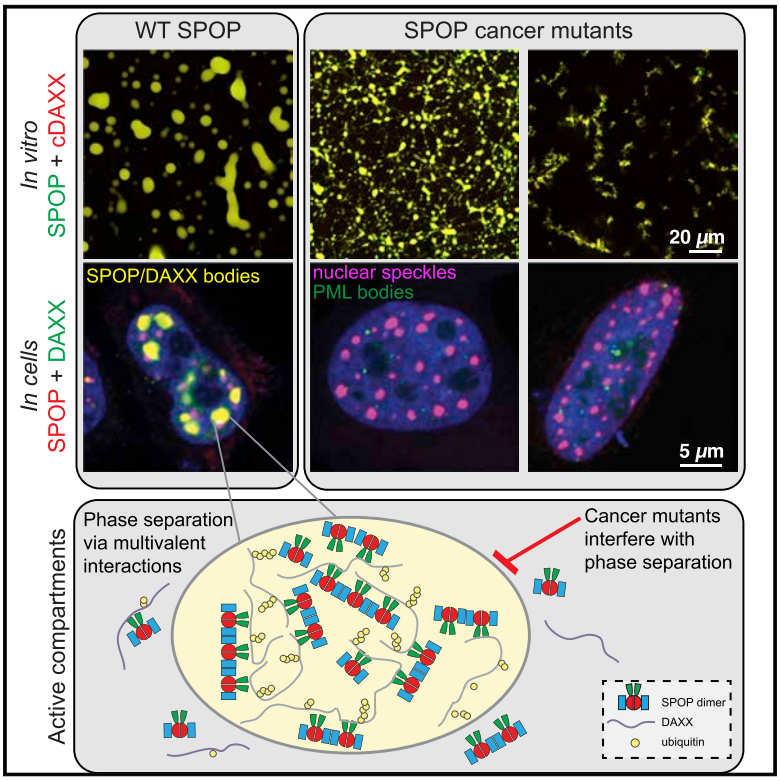
PUBLICATIONS

Article

Molecular Cell

Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments

Graphical Abstract



Authors

Jill J. Bouchard, Joel H. Otero, Daniel C. Scott, ..., Xavier Salvatella, Brenda A. Schulman, Tanja Mittag

Correspondence

tanja.mittag@stjude.org

In Brief

Mutations in the tumor suppressor SPOP are known to cause solid tumors. Bouchard and Otero et al. show that SPOP phase separates with substrates *in vitro*; the same interactions mediate co-localization in membraneless organelles in cells. SPOP cancer mutations disrupt liquid-liquid phase separation, which correlates with loss of function.

Highlights

- Substrates drive phase separation of the tumor suppressor SPOP
- Phase separation and co-localization of SPOP and substrate depend on multivalency
- Mesoscale SPOP-substrate assemblies mediate enzymatic activity
- SPOP cancer mutations disrupt phase separation, co-localization, and activity

Bouchard et al., 2018, Molecular Cell *72*, 19–36
October 4, 2018
https://doi.org/10.1016/j.molcel.2018.08.027

CellPress

Molecular Cell
Article

Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments

Jill J. Bouchard,[1,7] Joel H. Otero,[1,7] Daniel C. Scott,[1] Elzbieta Szulc,[2] Erik W. Martin,[1] Nafiseh Sabri,[1] Daniele Granata,[3] Melissa R. Marzahn,[1] Kresten Lindorff-Larsen,[3] Xavier Salvatella,[2,4] Brenda A. Schulman,[1,5,6] and Tanja Mittag[1,8,*]
[1]Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
[2]Institute for Research in Biomedicine, The Barcelona Institute of Science and Technology, Barcelona 08028, Spain
[3]Structural Biology and NMR Laboratory, The Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark
[4]ICREA, Barcelona, Spain
[5]Department of Molecular Machines and Signaling, Max-Planck Institute of Biochemistry, Martinsried/Munich, Germany
[6]Howard Hughes Medical Institute, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: tanja.mittag@stjude.org
https://doi.org/10.1016/j.molcel.2018.08.027

SUMMARY

Mutations in the tumor suppressor SPOP (speckle-type POZ protein) cause prostate, breast, and other solid tumors. SPOP is a substrate adaptor of the cullin3-RING ubiquitin ligase and localizes to nuclear speckles. Although cancer-associated mutations in SPOP interfere with substrate recruitment to the ligase, mechanisms underlying assembly of SPOP with its substrates in liquid nuclear bodies and effects of SPOP mutations on assembly are poorly understood. Here, we show that substrates trigger phase separation of SPOP *in vitro* and co-localization in membraneless organelles in cells. Enzymatic activity correlates with cellular co-localization and *in vitro* mesoscale assembly formation. Disease-associated SPOP mutations that lead to the accumulation of proto-oncogenic proteins interfere with phase separation and co-localization in membraneless organelles, suggesting that substrate-directed phase separation of this E3 ligase underlies the regulation of ubiquitin-dependent proteostasis.

INTRODUCTION

Cancer-driving mutations in enzymes typically reduce their activity or introduce an aberrant activity. However, mislocalization of mutant enzymes is another form of loss of function. Here, we explore the mechanism underlying SPOP loss-of-function cancer mutations and show that mutants fail to co-localize with substrates via disruption of phase separation.

SPOP is frequently mutated in solid tumors, particularly in prostate cancer (Cerami et al., 2012; Kim et al., 2011, 2013; Lawrence et al., 2014; Le Gallo et al., 2012). Its gene product, SPOP (speckle-type POZ protein), is a substrate adaptor of a cullin3-RING ubiquitin ligase (CRL3) that recruits substrates to the ligase for ubiquitination and subsequent proteasomal degradation (Hernández-Muñoz et al., 2005; Kent et al., 2006; Kwon et al., 2006; Li et al., 2014). Cancer-associated mutations in SPOP are typically found in the substrate-binding meprin and TRAF homology (MATH) domain (Figure 1A); accordingly, they interfere with substrate binding and ubiquitination, which increases substrate levels (Gan et al., 2015; Geng et al., 2017). SPOP substrates include the death-domain-associated protein (DAXX), androgen receptor (AR), and other important signaling cascade effectors, epigenetic modifiers, and hormone signaling effectors (Gan et al., 2015; Gao et al., 2015; Geng et al., 2013, 2017; Janouskova et al., 2017; Li et al., 2014; Theurillat et al., 2014; Zhang et al., 2009, 2018; Zhuang et al., 2009). Increased levels of proto-oncogenic substrates as a consequence of SPOP mutations can oncogenically transform sensitive cell types (An et al., 2014; Dai et al., 2017a, 2017b; Gan et al., 2015; Geng et al., 2013, 2014, 2017; Janouskova et al., 2017; Theurillat et al., 2014).

Understanding how SPOP targets its substrates is important for the development of therapeutic interventions for SPOP-driven cancers. However, unlike many prototypic cullin-RING ligases, which rely on post-translational modifications for substrate targeting (Petroski and Deshaies, 2005), mechanisms that could regulate SPOP-substrate interactions are largely elusive. SPOP typically localizes to nuclear speckles (Nagai et al., 1997) and has also been reported to localize to DNA damage loci, promyelocytic leukemia (PML) bodies, and other substrate-containing bodies (Boysen et al., 2015; Gan et al., 2015; Kwon et al., 2006; Marzahn et al., 2016; Zhang et al., 2014). Substrate level dynamics in cells may therefore direct SPOP localization. However, how SPOP assembles with substrates and gets recruited to nuclear bodies is not well understood.

The nuclear bodies with which SPOP associates are so-called membraneless organelles, i.e., mesoscale bodies that

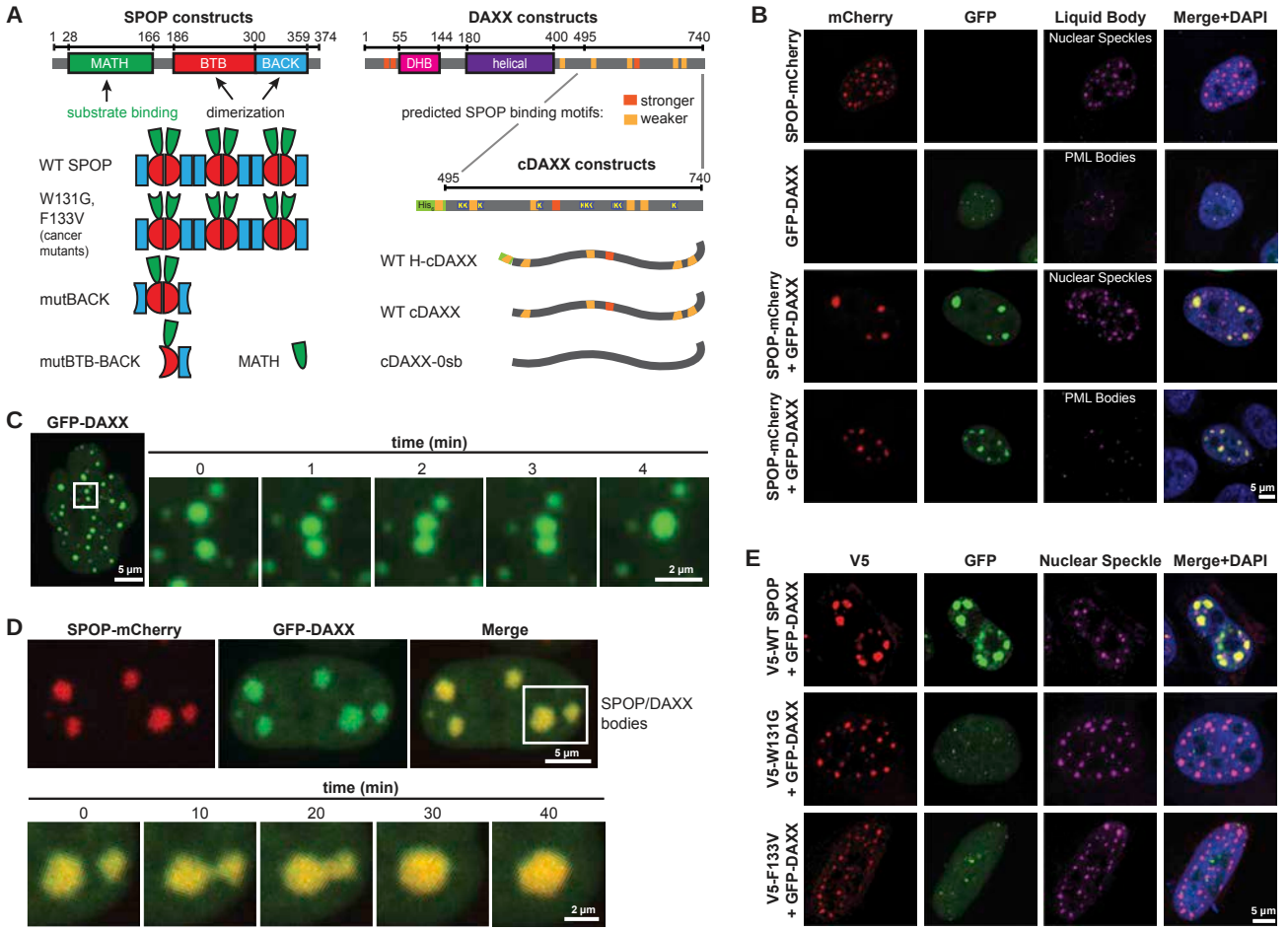


Figure 1. SPOP and DAXX Co-localize in Liquid Organelles, and SPOP Cancer Mutations Disrupt Co-localization

(A) Sequence and cartoon schematics for SPOP (left) and DAXX (right) constructs used in this study. Top: Sequence cartoons represent domain architecture. SPOP contains a substrate-binding MATH domain and two dimerization domains, BTB and BACK. DAXX contains a DAXX helical bundle (DHB) domain, a helical region, and a C-terminal disordered domain (Escobar-Cabrera et al., 2010). Predicted SPOP-binding motifs (based on the consensus sequence motif, nonpolar-polar-S-S/T-S/T; Zhuang et al., 2009) are shown in orange, with stronger binding sites shaded darker. Lys residues available for ubiquitination are shown as K. (Bottom) Cartoon schematics represent self-association and substrate binding behavior; mutated interfaces in SPOP shown curved instead of straight indicate the inability to self-associate or bind substrate. WT SPOP forms higher-order oligomers of different sizes (Marzahn et al., 2016); we show hexamers as an example. Cancer mutations W131G and F133V in the MATH domain (green) reduce substrate binding. Mutation of the dimerization interface of the BACK domain (blue, mutation Y353E; van Geersdaele et al., 2013) results in SPOP mutBACK dimers (Marzahn et al., 2016); the addition of mutations of the BTB interface (red, L186D, L190D, L193D, and I217K; Zhuang et al., 2009) results in SPOP mutBTB/BACK monomers (Marzahn et al., 2016). SPOP constructs for expression in mammalian cells encode the full-length protein; those for expression in bacteria encode residues 28–359. DAXX mammalian expression constructs encode full-length protein unless labeled cDAXX, which comprises residues 495–740, as the bacterial expression constructs. H-cDAXX harbors a His_6 tag. In cDAXX-0sb, the major SB motifs are mutated. For details, see Figure 3 and Table S1.

(B) SPOP and DAXX localize to SPOP/DAXX bodies. HeLa cells were transfected with the indicated GFP-DAXX and/or SPOP-mCherry and analyzed by confocal microscopy. mCherry (red) and GFP fluorescence (green) were observed for SPOP and DAXX, while SC-35 and PML (both magenta) were used as markers for nuclear speckles and PML bodies, respectively, and detected by immunofluorescence (IF). DAPI (blue) marks the nucleus. See also Figures S1A–S1D.

(C) PML bodies behave like liquid droplets. HeLa cells were transfected with GFP-DAXX and GFP monitored in live cells. Snapshots at the indicated time points show a PML body fusion event from the area boxed on the left. See also Video S1.

(D) SPOP and DAXX form nuclear bodies with liquid properties. HeLa cells were transfected with GFP-DAXX and SPOP-mCherry and analyzed as in (C). See also Video S2.

(E) SPOP cancer mutants fail to localize to SPOP/DAXX bodies. HeLa cells were transfected with GFP-DAXX and either WT V5-SPOP or mutants F133V or W131G and analyzed as in (B). See also Figure S1G.

concentrate specific components within them without being enclosed by a membrane. Evidence is mounting that membraneless organelles are formed through phase separation processes. Liquid-droplet-like organelles result from liquid-liquid phase separation (LLPS) (Berry et al., 2015; Brangwynne et al., 2009; Nott et al., 2015), while solid assemblies may result from

oligomerization and/or polymerization processes or maturation of liquid assemblies into solid ones (Boke et al., 2016; Cai et al., 2014; Kato et al., 2012). While mesoscale assemblies formed through different processes are on a spectrum between solid and dynamic assemblies and have different material properties (Halfmann, 2016), they all have the potential to co-localize enzymes with their substrates and therefore regulate activity in cells (Wu and Fuxreiter, 2016).

Here, we systematically examine the mechanism of SPOP-substrate co-localization. We have shown previously that SPOP self-associates into large higher-order oligomers through the synergistic function of two dimerization domains, the BTB (broad-complex, tramtrack, and bric-à-brac) and the BACK (BTB and C-terminal kelch) domains (Figure 1A). This self-association is required for localization to membraneless organelles (Marzahn et al., 2016) but is insufficient to drive LLPS, because it mediates the formation of assemblies whose size distribution continuously shifts to larger sizes with increasing concentration but never leads to the cooperative formation of mesoscale assemblies. Here, we show substrates act as the trigger for co-localization of SPOP and substrates in cells and phase separation with each other *in vitro*. CUL3 ubiquitin ligase activity is found in the resulting mesoscale assemblies. We find that cancer-associated mutations disrupt co-localization and liquid phase separation, which correlates with reduced protein ubiquitination. Our results suggest the possibility that substrate-mediated phase separation of this ubiquitin ligase is essential for concentrating the enzyme and its substrates in active liquid organelles in cells, and that such organelles may have functions essential to proteostasis.

RESULTS

SPOP and DAXX Co-localize in Liquid Nuclear Organelles

Because many substrates contain multiple SPOP-binding (SB) motifs in intrinsically disordered regions (IDRs) (Pierce et al., 2016; Zhang et al., 2009; Zhuang et al., 2009) and can mediate multivalent interactions with SPOP, substrates are candidates for mediating phase separation with SPOP and regulating its subcellular localization and function. Increased substrate levels, or increased apparent levels via other signals, seem to mediate the redistribution of SPOP from nuclear speckles into different compartments. To examine roles of substrates in SPOP localization, we used transient expression of DAXX and SPOP in cultured HeLa cells. Increased DAXX levels caused by SPOP cancer mutations likely contribute to cancer pathogenesis by increasing angiogenic factors such as VEGFR2 (Sakaue et al., 2017). Furthermore, DAXX contains at least 8 predicted SB motifs in its IDRs (Figure 1A), making it a likely candidate to regulate SPOP localization. SPOP-mCherry localized to nuclear bodies that stain for the typical nuclear speckle marker SC-35 (Figures 1B and S1A), as previously reported (Marzahn et al., 2016; Nagai et al., 1997). Transiently expressed GFP-DAXX (and its intrinsically disordered C-terminal region, GFP-cDAXX) localized to small, spherical PML bodies (Figures 1B, S1A, and S1B), in agreement with previous reports (Kwon et al., 2006; Li et al., 2000; Weidtkamp-Peters et al., 2008). However, we found that transient co-expression of SPOP-mCherry and GFP-DAXX resulted in their co-localization in a different, largely spherical type of nuclear body distinct from nuclear speckles, PML bodies, nucleoli, and Cajal bodies (Figures 1B, S1C, and S1D); we will refer to them as SPOP/DAXX bodies from here on. Different expression tags did not influence the co-localization of SPOP and DAXX (Figure S1E). Importantly, co-expression of V5-tagged SPOP and GFP-DAXX in a PC-3 prostate cancer cell line also resulted in co-localization (Figure S1F).

DAXX and SPOP therefore both re-localized from their original subcellular location when expressed together. The SPOP/DAXX bodies possess liquid properties as evidenced by their ability to undergo fusion events within minutes (Figures 1C and 1D; Videos S1 and S2). These properties place them into the category of liquid membraneless organelles.

SPOP Cancer Mutants Disrupt Co-localization

The typical prostate cancer mutations in SPOP, W131G and F133V in the substrate-binding MATH domain, disrupted SPOP and DAXX co-localization upon co-expression; V5-SPOP W131G and F133V localized to nuclear speckles, and GFP-DAXX localized to PML bodies, as they do when expressed alone (Figures 1E and S1G). The SPOP cancer mutations thus disrupt re-localization of both proteins.

It is clear from these observations not only that SPOP and DAXX bind to each other but also that binding shifts them to a different liquid organelle. We therefore hypothesized that SPOP and DAXX undergo LLPS with each other, a process in which a macromolecule (or a set of macromolecules) demixes from the solution and forms a separate, condensed liquid phase, often visible as liquid droplets.

SPOP and DAXX Undergo Phase Separation *In Vitro*

To test this hypothesis, we purified $SPOP^{28–359}$ (the extreme termini are missing in *in vitro* purified protein to improve protein behavior; Marzahn et al., 2016) and the intrinsically disordered C-terminal region of DAXX, $DAXX^{495–740}$ (cDAXX from here on) (Figure 1A), and we studied their interaction *in vitro*. We have previously reported that SPOP self-associates into linear higher-order oligomers (Marzahn et al., 2016). In the presence of molecular crowders such as Ficoll-70, these oligomers are large enough to be observed by light microscopy (Figure 2A, top, and Figures S2A–S2C). At concentrations of above ~100 μM, H-cDAXX forms condensed droplets (Figure S2A). However, the tendency of H-cDAXX to undergo phase separation is strongly enhanced in the presence of SPOP, and this was the case in the presence of both polymer and protein crowders (Figure S2B). This observation suggests that weak multivalent interactions between SPOP and DAXX result in a sol-gel transition coupled to phase separation, as defined by Harmon et al. (2017).

SPOP Oligomerization Promotes Phase Separation and Co-localization of DAXX and SPOP

We hypothesized that phase separation of DAXX and SPOP was mediated by the interaction of multiple SB motifs in DAXX with multiple MATH domains in oligomeric SPOP, as described previously for multivalent systems (Li et al., 2012). If this was in fact the

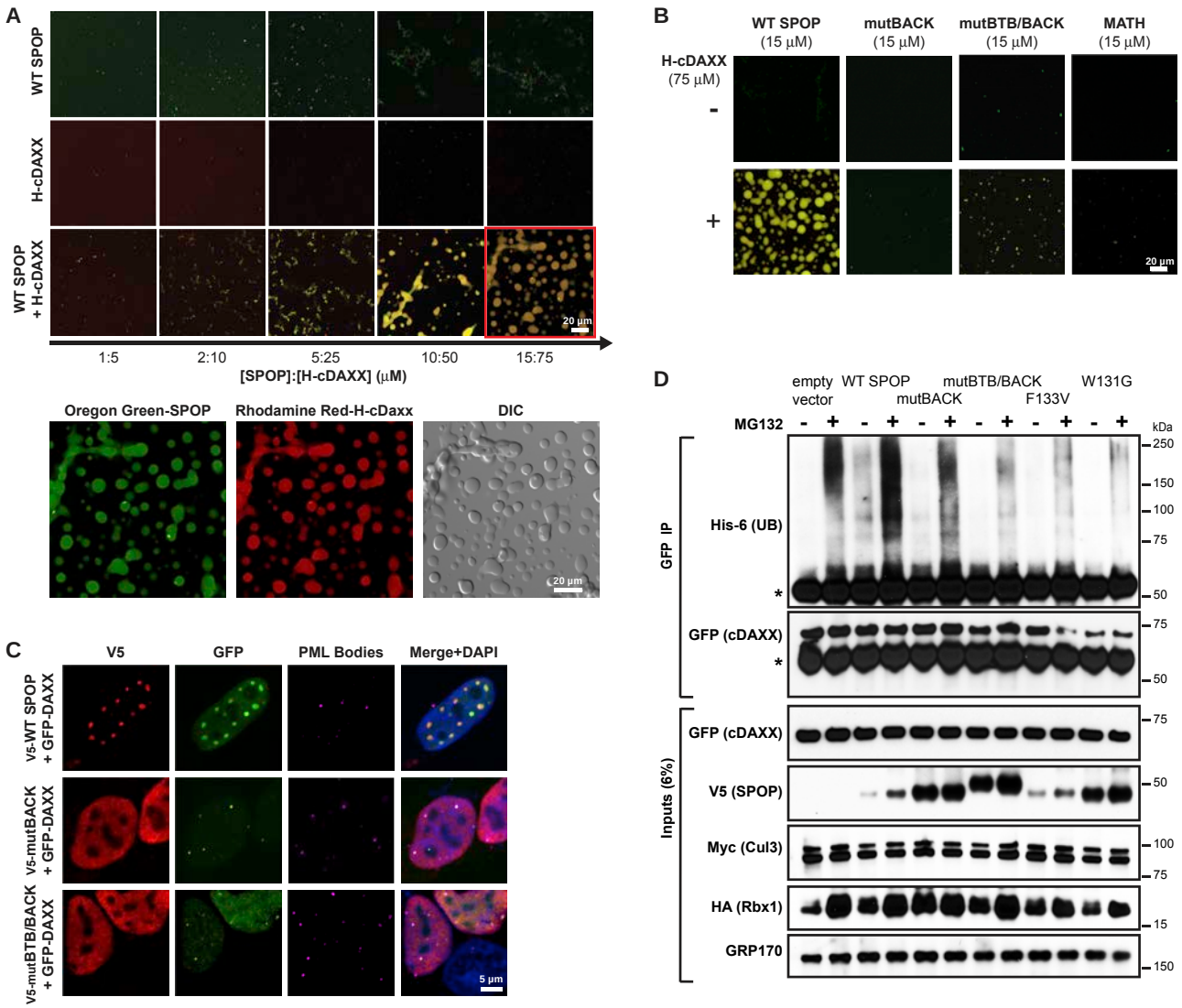


Figure 2. SPOP and DAXX Undergo Phase Separation *In Vitro* and in Cells, which Depends on SPOP Oligomerization

(A) SPOP and DAXX undergo phase separation *in vitro*. Fluorescence microscopy images of increasing concentrations of WT SPOP (green) alone, H-cDAXX (red) alone, and SPOP + H-cDAXX at a constant molar ratio of 1 SPOP to 5 DAXX. Camera settings are the same across rows. The panel boxed red is shown as separate green, red, and DIC channels below. All samples in (A) and (B) contain 10% w/v ficoll 70, 500 nM ORG-SPOP, and/or Rhodamine-H-cDAXX. Samples were in 25 mM Tris (pH 7.6), 150 mM NaCl, and 1 mM T-CEP. See also Figures S2A–S2C.

(B) SPOP multivalency is required for SPOP-DAXX phase separation *in vitro*. Fluorescence microscopy images of SPOP variants (green) with reduced self-association ability in the presence or absence of H-cDAXX (red). Camera settings are the same down columns.

(C) SPOP multivalency is required for SPOP-DAXX co-localization in cells. HeLa cells were transfected with GFP-DAXX and V5-WT SPOP or the mutants mutBACK or mutBTB/BACK and analyzed by confocal microscopy. GFP fluorescence was observed for DAXX (green), while V5-SPOP (red), and PML bodies (magenta) were detected by IF. See also Figure S2D.

(D) SPOP mutants are defective at DAXX ubiquitination in cells. Western blots showing GFP-cDAXX ubiquitination in HEK293T cells that were transfected with His_6-ubiquitin, Myc-Cul3, HA-Rbx1, and one of the SPOP variants each. The asterisk indicates the immunoglobulin G (IgG) heavy chain. See Figure S2E for the in-cell ubiquitination assay with pull-down on His_6-ubiquitin.

case, then reducing the multivalency of SPOP and DAXX should decrease their propensity to undergo phase separation. We therefore tested this hypothesis using previously established SPOP mutants of the BTB and BACK interfaces that lack the ability to form higher-order oligomers and are instead constitutively dimeric ($SPOP^{mutBACK}$, in which Y353 in the BACK interface is mutated; van Geersdaele et al., 2013), or monomeric ($SPOP^{mutBTB/BACK}$, in which the additional substitutions L186D, L190D, L193D, and I217K are made in the BTB interface; Zhuang et al., 2009) (Figure 1A; Marzahn et al., 2016). Indeed, these SPOP mutants do not or only slightly enhance H-cDAXX phase separation (Figure 2B).

Similarly to our *in vitro* observations, $SPOP^{mutBACK}$ and $SPOP^{mutBTB/BACK}$ were unable to re-localize DAXX to SPOP-DAXX bodies in HeLa cells. Instead, $SPOP^{mutBACK}$ and $SPOP^{mutBTB/BACK}$ were diffuse in the nucleus, in agreement with our earlier observation that SPOP self-association was required for recruitment to liquid organelles (Marzahn et al., 2016). Consequently, DAXX was found in PML bodies in the presence of $SPOP^{mutBACK}$ and $SPOP^{mutBTB/BACK}$ (Figures 2C and S2D). Therefore, SPOP-self-association into higher-order oligomers via its BTB and BACK domains promotes phase separation with DAXX *in vitro* and co-localization in cells.

We tested whether the lack of co-localization of DAXX and SPOP mutants resulted in decreased substrate ubiquitination. While DAXX and cDAXX were readily ubiquitinated by wild-type (WT) SPOP, ubiquitination was substantially reduced with both SPOP mutants, $SPOP^{mutBACK}$ and $SPOP^{mutBTB/BACK}$ (Figures 2D and S2E), even though their levels in cells were higher. This defect of oligomerization-deficient SPOP mutants is consistent with our previous reports on the same mutants in an *in vitro* ubiquitination assay (Marzahn et al., 2016; Pierce et al., 2016). Our data suggest that the underlying mechanism of reduced ubiquitination in the cell is mislocalization due to disrupted phase separation.

Multiple Weak SB Motifs in DAXX Mediate Phase Separation with SPOP

Since SPOP multivalency is critical for phase separation, DAXX multivalency might also be critical. We identified SB motifs in cDAXX and scrambled the strongest sites in the construct cDAXX-0sb (see STAR Methods for details) to test the ability of SPOP to enhance phase separation of cDAXX versus cDAXX-0sb.

To identify SB motifs with sequence homology to the SB consensus motif (Φ-Π-S-S/T-S/T, where Φ is a nonpolar and Π is a polar residue; Zhuang et al., 2009) in cDAXX, we performed a bioinformatic search allowing for one mismatched position per motif. We found 5 potential binding sites with different agreement with the consensus sequence and thus varying predicted strength (Figure 1A; Table S1). To characterize binding of these SB motifs by two complementary biophysical methods, we mapped regions in cDAXX that bound to $SPOP^{MATH}$ by nuclear magnetic resonance (NMR) spectroscopy and quantified the binding of each predicted site, and the full cDAXX constructs, by fluorescence anisotropy (FA).

We assigned the NMR resonances of $^{15}N,^{13}C$ cDAXX (Figures S3A and S3B), which have the limited chemical shift dispersion and sharp line widths of a typical intrinsically disordered region (IDR) (Figure 3A (Eliezer, 2009; Mittag and Forman-Kay, 2007)). These spectral properties indicate that many cDAXX conformations are in fast exchange and that the IDR exhibits little structure. Any SB motifs present in cDAXX should thus be accessible for binding SPOP. A titration of unlabeled $SPOP^{MATH}$ into $^{15}N,^{13}C$ cDAXX and monitored by CON spectra resulted in a dose-dependent loss of signal intensity along several cDAXX regions (Figures 3A and 3B). The regions of broadening coincided with most of the predicted motifs and revealed one other weak site (Figure 3B, solid and dashed orange lines, respectively, and 3C, dark and light orange sections). Additional broad regions experienced intensity loss, indicating there may be additional cryptic binding motifs.

To determine the affinity of the five predicted SB motifs to the $SPOP^{MATH}$ groove, we generated 13-residue-long peptides encompassing the motifs (Table S2) and performed FA competition binding experiments. Indeed, all candidate SB motifs interacted with SPOP weakly (Figure 3D) with dissociation constant (K_D) values from 40 μM to 1 mM (Figure 3D; Table S2), in agreement with the role of weak multivalent interactions in LLPS.

Next, we scrambled the motifs, and they indeed lost the ability to interact with $SPOP^{MATH}$; only one mutant motif still interacted with a K_D value of 2 mM (Figure 3E; Table S2). Consequently, cDAXX-0sb, the cDAXX version with scrambled motifs, bound to $SPOP^{MATH}$ with a K_D in the hundreds of micromolar range while WT cDAXX had a K_D of 40 μM (Figure 3F; Table S3). The interaction of cDAXX-0sb with multivalent $SPOP^{28–359}$ was even more dramatically decreased compared to WT cDAXX, with hundreds of micromolar K_D values versus 1.7 μM, respectively (Figure 3G; Table S3). Residual weak SB motifs must still be present in cDAXX-0sb (Figure 3C), and they result in weak binding to SPOP and $SPOP^{MATH}$. But, as expected, the reduced multivalency of cDAXX-0sb prevented SPOP-enhanced phase separation (Figure 3H). Our results therefore confirm that cDAXX contains multiple SB motifs that are critical for its ability to form condensed droplets with SPOP.

We wondered whether the multivalency of DAXX is similarly important for co-localization in cells as it is for phase separation *in vitro*. First, the localization of GFP-cDAXX in cells showed the same dependence on SPOP expression as that of full-length DAXX; GFP-cDAXX was diffuse with some fraction of the protein in PML bodies (Figure S1B), and it co-localized with SPOP when transiently expressed together (Figure 3I). Second, cDAXX-0sb was less enriched in SPOP/DAXX bodies than WT cDAXX (Figures 3I, 3J, and S3C). Residual co-localization is likely caused by the same weak multivalent interactions that we detected by NMR and FA (Figure 3B, E, C). Therefore, the localization in cells and the ability to phase separate *in vitro* depend similarly on SPOP oligomerization and DAXX multivalency. Furthermore, ubiquitination of cDAXX-0sb is significantly reduced compared to cDAXX, as is the case for SPOP oligomerization-deficient mutants compared to SPOP (Figures 3K, 2D, and S2E). Based on the dependence on multivalency for DAXX and SPOP that we have shown, it is reasonable to conclude that SPOP and DAXX also undergo phase separation in the cell.

Material Properties of SPOP/DAXX Mesoscale Assemblies

When we explored formation of mesoscale SPOP/H-cDAXX assemblies in more detail, we observed that droplets formed at H-cDAXX/SPOP molar ratios above ~3 (Figure 4A and magnification at top right). At molar ratios below ~2, the assemblies have a filamentous morphology, as especially noticeable in differential interference contrast (DIC) images (Figure 4A and magnification on bottom right; the DIC image is incorporated in the composite image to highlight the texture of the filamentous assemblies). We wondered whether these two different types of mesoscale assemblies were formed through two different

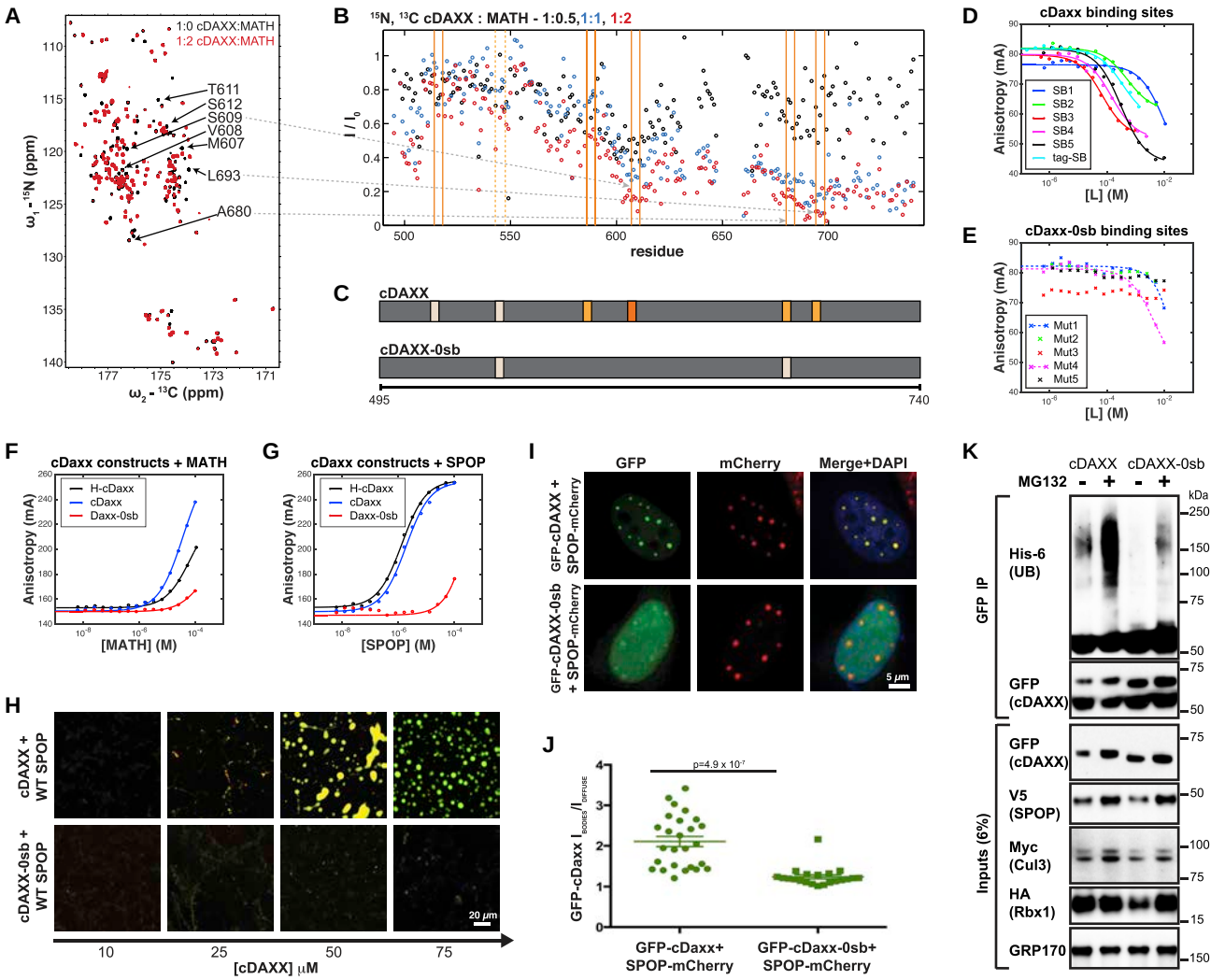


Figure 3. Multiple Weak SPOP-Binding Motifs in DAXX Mediate Phase Separation with SPOP

(A) cDAXX is intrinsically disordered and binds SPOP via several SB motifs. $^{15}N,^{13}C$ CON NMR spectrum of cDAXX at 600 MHz and 25°C, without $SPOP^{MATH}$ (black) and in the presence of 2 molar equivalents of $SPOP^{MATH}$ (red). For spectra annotated with all assignments, see Figures S3A and S3B.

(B) Titration of $SPOP^{MATH}$ into cDAXX leads to identification of SB motifs. Intensity ratios of CON correlations for cDAXX upon titration with $SPOP^{MATH}$ (I/I_0) are plotted as a function of residue number. Broadening of CON resonances of cDAXX in the presence of $SPOP^{MATH}$ reveals multiple SB motifs, the 5 predicted (solid orange lines), one unpredicted (dashed orange lines), and other broadened regions.

(C) Sequence schematic for cDAXX constructs updated based on binding data in (B), (D), and (E). Stronger SB motifs are shown in darker shades of orange. In cDAXX-0sb, the nonpolar residue in each SB motif was replaced with a polar residue, the second residue replaced with a proline, and the rest of the motif sequence scrambled. See Tables S1 and S2 for sequences.

(D–G) cDAXX binds SPOP in an SB-motif-dependent manner. Representative fluorescence anisotropy competition binding isotherms for peptides containing cDAXX-binding sites (D) or mutated binding sites (E) into $SPOP^{MATH}$ and fluorescein-$Puc^{91–106}$ and direct binding isotherms for $SPOP^{MATH}$ (F) and WT SPOP (G) into full-length Rhodamine-cDAXX constructs. Symbols are experimental data points; continuous or dashed lines are nonlinear least-squares fits (Roehrl et al., 2004). All measurements were conducted in triplicate. Average K_D values are shown in Tables S2 and S3, respectively.

(H) DAXX-0sb does not phase separate with SPOP *in vitro*. Fluorescence microscopy images of SPOP with cDAXX or cDAXX-0sb. All samples contain 10% w/v ficoll 70, 500 nM ORG-SPOP and/or Rhodamine-cDAXX.

(I) DAXX-0sb does not localize predominantly to SPOP/DAXX bodies in cells. HeLa cells were transfected with GFP-cDAXX or GFP-cDAXX-0sb and SPOP-mCherry and analyzed by confocal microscopy. cDAXX-0sb in the absence of endogenous SPOP localizes to PML bodies (Figure S3C).

(J) Quantification of partition coefficient of GFP-cDAXX and GFP-cDAXX-0sb into SPOP/cDAXX bodies in (I). Each point in the whisker plot signifies an individual cell, and the mean is shown as a line. Error bars indicate the SEM.

(K) The cDAXX-0sb mutant is defective for ubiquitination in cells. Western blots showing GFP-cDAXX and GFP-cDAXX-0sb ubiquitination in HEK293T cells that were transfected and analyzed as in Figure 2D.

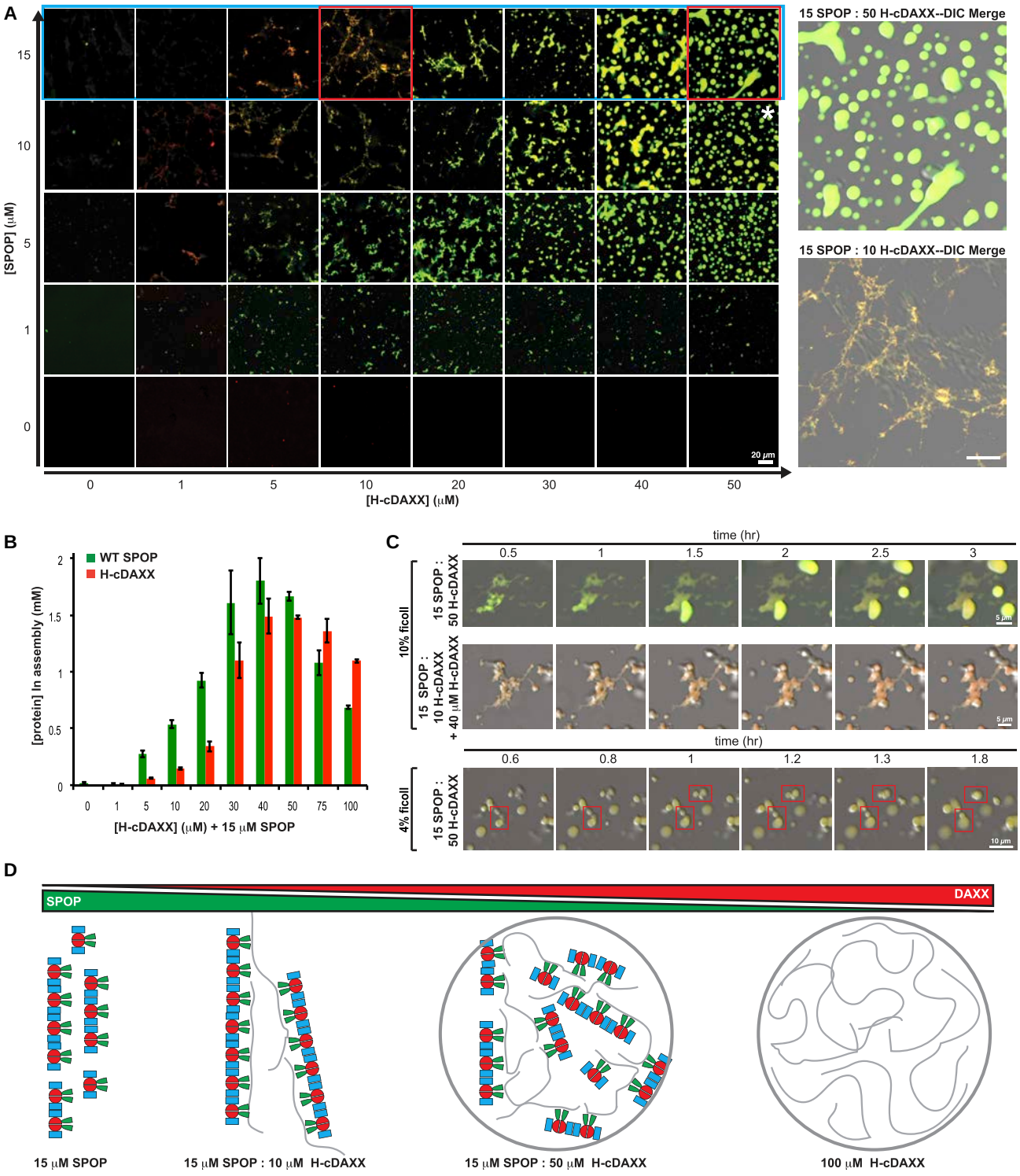


Figure 4. Material Properties of SPOP/DAXX Mesoscale Assemblies

(A) SPOP and H-cDAXX form filamentous assemblies as well as liquid droplets. Fluorescence microscopy images of SPOP/H-cDAXX as a function of protein concentration. All samples contain 10% w/v ficoll 70, 500 nM ORG-SPOP and/or Rhodamine-H-cDAXX. Images in red boxes are enlarged with DIC overlaid at the right.

(B) Quantification of protein concentration in mesoscale assemblies in the top row of (A, blue box). Error bars represent the SD from three replicate images. For standard curves and additional conditions see Figures S4A–S4C.

(legend continued on next page)

assembly processes or whether they had different material properties because of their different composition and whether the filamentous assemblies represented non-native, irreversible aggregates from maturation of liquid droplets.

Some dense liquid assemblies can mature into gel-like states or nucleate fibrillization over time (Lin et al., 2015; Mackenzie et al., 2017; Molliex et al., 2015; Monahan et al., 2017; Murakami et al., 2015; Patel et al., 2015; Zhang et al., 2015). Importantly, this maturation of dense liquid states has been proposed as the precipitating event in protein aggregation diseases (Lin et al., 2015; Mackenzie et al., 2017; Molliex et al., 2015; Monahan et al., 2017; Murakami et al., 2015; Patel et al., 2015). Interestingly, we indeed observed morphological changes over time before assemblies reached their final state; i.e., samples that result in assemblies with droplet character after incubation can have filamentous character when first mixed (Figure 4C, top). Notably, it was also possible to manipulate the morphology of samples by adding one of the components after incubation; when we added additional H-cDAXX to a sample with filamentous assemblies (as in Figure 4A, bottom inset), the sample evolved to a droplet state (Figure 4C; as in Figure 4A, starred condition; please note camera settings are different, and thus, assemblies appear to have different colors). This evolution suggests that the filamentous assemblies do not represent irreversible, non-native aggregates. We attribute these morphological changes to slow off-rates of SPOP building blocks from SPOP oligomers in the presence of crowders and multivalent substrate, as we saw previously with the multivalent substrate $Gli3^{1-90}$ (Pierce et al., 2016). All samples were therefore imaged after 4–6 hr of incubation. The assemblies with droplet appearance undergo fusion events (Figure 4C, bottom), which demonstrate liquid-like properties.

We wondered whether the filaments and droplets were both formed through LLPS and merely represented different material states or whether they formed through different assembly processes. We generated samples consisting of 15 μM SPOP and increasing concentrations of H-cDAXX, equivalent to the top row in the phase diagram in Figure 4A (blue box). After incubation, the protein concentrations within the mesoscale assemblies were determined via their fluorescence intensity in confocal micrographs (Figures 4B and S4A–S4C). The concentrations followed a biphasic behavior. They first increased with the H-cDAXX concentration in the sample, consistent with increasing levels of protein incorporated into assemblies in a typical oligomerization and/or polymerization process. The H-cDAXX concentration in assemblies then plateaued above a H-cDAXX input concentration of 30 μM (Figure 4B), while the SPOP concentration decreased continuously. This is in agreement with a mechanism in which H-cDAXX is able to mediate phase separation independently and SPOP promotes H-cDAXX phase separation further (Figures S2A, S2B, and 4D).

These results support a scenario in which SPOP forms linear, filamentous higher-order oligomers, as previously described (Marzahn et al., 2016) (Figure 4D, left). The addition of multivalent H-cDAXX (at levels below the saturation concentration for phase separation) leads to the stabilization of the oligomeric state, observed as larger species in crosslinking assays (Figures S4D and S4E), resulting in solid filaments through an oligomerization and/or polymerization process (Figure 4D, left middle, and Figure S4E, left and middle boxes). With increasing H-cDAXX concentrations, H-cDAXX molecules compete for SPOP, and SPOP oligomers are less stabilized; this is observed as smaller species in crosslinking experiments (Figure S4E, right box). Eventually, more strongly networked complexes form, resulting in the formation of dense SPOP/DAXX-containing droplets via LLPS (Figure 4D, middle right). High concentrations of H-cDAXX can undergo LLPS alone (Figure 4D, right).

In support of this interpretation, the H-cDAXX mobility as measured by fluorescence recovery after photobleaching (FRAP) is larger than the SPOP mobility, notwithstanding the type of assembly DAXX is incorporated in (Figure S4F; Table S4). Interestingly, the DAXX mobility is nearly identical in filaments and liquids, suggesting a steady flux of DAXX in and out of assemblies. The SPOP mobility is similar in the absence of DAXX and in the filamentous assemblies with DAXX. This points to SPOP as the scaffold, formed from large, adhesive oligomers, with DAXX binding to it from the outside. DAXX forms the liquid-promoting contacts. Only in SPOP/DAXX droplets does SPOP have an increased mobility, suggesting a change in the underlying structure from filaments to more networked complexes. FRAP analysis of GFP-DAXX in cells shows a higher mobility of DAXX in PML bodies than in SPOP/DAXX bodies, reflecting the different nature of the interactions driving DAXX localization into the different nuclear bodies. SPOP has a low mobile fraction in nuclear speckles as well as in SPOP/DAXX bodies, indicating viscoelastic properties of SPOP-containing bodies (Figure S4G).

Together, our results suggest that the concentration-dependent multivalency of SPOP, coupled to the phase separation propensity encoded in its substrate, results in a rich phase diagram with several different types of mesoscale assemblies (Figure S4H).

SPOP Cancer Mutants Disrupt Phase Separation and DAXX Ubiquitination

Given the role of phase separation for SPOP-substrate colocalization, we hypothesized that SPOP cancer mutations would disrupt this behavior, explaining the cellular mislocalization shown in Figure 1E. Since SPOP cancer mutants disrupt substrate binding (Gan et al., 2015, Geng et al., 2017), we first tested

(C) Filamentous assemblies are not irreversible aggregates. (Top) Time course of fluorescence microscopy and DIC images of a 15 μM SPOP to 50 μM H-cDAXX sample, which develops its typical droplet appearance over time. (Middle) Addition of extra H-cDAXX to a filamentous sample incubated for 2 hr. The assemblies change from the filamentous to the droplet-like morphology. (Bottom) Fusion events between SPOP/H-cDAXX droplets (red boxes).

(D) Schematic of the proposed nature of assemblies at different SPOP/H-cDAXX molar ratios. SPOP alone forms oligomers (left). Oligomers are stabilized in the presence of low molar ratios of H-cDAXX, leading to large filamentous assemblies (middle left). At higher molar ratios of H-cDAXX, intermolecular interactions are favored, SPOP oligomers are smaller, and H-cDAXX contributes to liquid behavior (middle right). H-cDAXX alone forms droplets (right). See also Figures S4D–S4G.

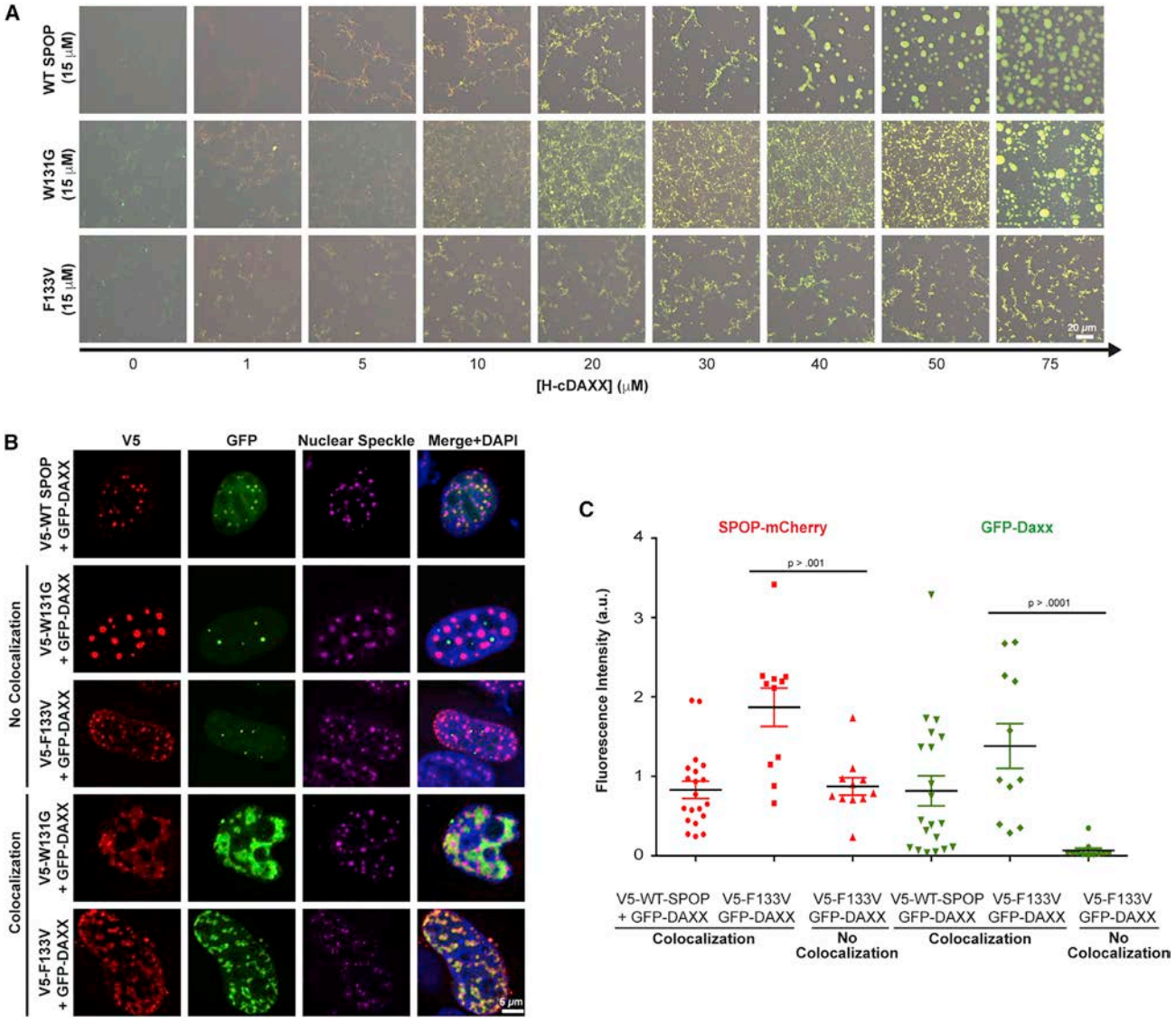


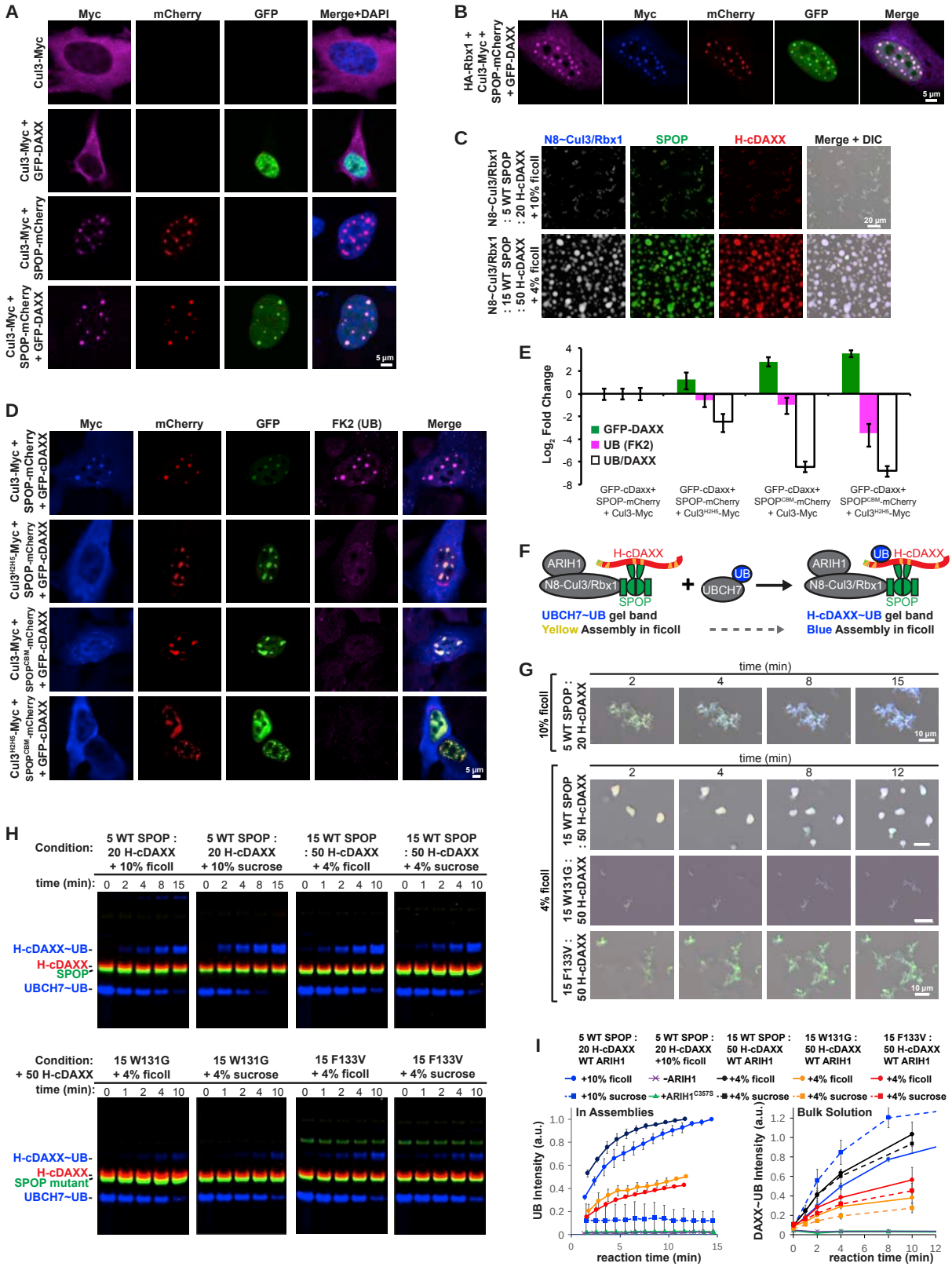
Figure 5. SPOP Cancer Mutants Disrupt Phase Separation

(A) SPOP cancer mutants are defective at phase separation *in vitro*. Fluorescence microscopy and DIC images of WT SPOP or cancer mutants as a function of H-cDAXX concentration. All samples contain 10% w/v ficoll 70, 500 nM ORG-SPOP construct and/or Rhodamine-cDAXX. Camera settings were optimized in samples containing ~1:1 molar ratios for each row.

(B) SPOP cancer mutants are defective at co-localization with DAXX in HeLa cells. SC-35 (magenta) was used as marker for nuclear speckles. Cells with SPOP-DAXX co-localization or lack thereof are indicated. See also Figure S5B.

(C) SPOP cancer mutants co-localize with DAXX when expressed at high levels. Whisker plot showing the signal intensity of V5-SPOP or V5-F133V (red points) and GFP-DAXX (green points) from (B) in which the V5-SPOP construct and GFP-DAXX co-localize or fail to co-localize. Each point represents a single cell. Horizontal lines indicate the mean; error bars indicate SEM.

the binding of SPOP W131G to a peptide with a single SB motif (fPuc). Indeed, SPOP W131G did not bind fPuc (Figure S5A; Table S5), and we therefore expected that the phase separation propensity of SPOP cancer mutants with substrates would be reduced. Indeed, *in vitro*, SPOP W131G did not form droplet assemblies up to a concentration of ~75 μM H-cDAXX, while only 30 μM was required for the formation of droplets with SPOP WT (at a constant SPOP concentration of 15 μM; Figure 5A). SPOP F133V failed to form droplet assemblies with H-cDAXX over the full concentration range tested and maintained filamentous character throughout. The propensity of SPOP to form condensed liquid droplets with DAXX was therefore markedly decreased by cancer mutations. These results demonstrate that SPOP-substrate interactions are substantially weakened by cancer mutations and that multivalency of both binding partners can still result in a physical association and the formation of



(legend on next page)

large, solid assemblies, but LLPS requires increased protein concentrations.

The observation that SPOP W131G was able to form droplet assemblies with H-cDAXX at high concentrations *in vitro* prompted us to test whether the mislocalization defect of SPOP cancer mutants with DAXX could be rescued at high protein concentrations in cells as well, as would be expected if their normal co-localization is dependent on phase separation. Indeed, while V5-SPOP W131G and V5-SPOP F133V typically remained in nuclear speckles when co-expressed with GFP-DAXX, a fraction of cells with significantly higher SPOP and DAXX levels showed co-localization in SPOP/DAXX bodies (Figures 5B, 5C, and S5B). These results also suggest that the filamentous assemblies do not mediate the formation of liquid-like bodies in the cell and that these instead require a phase transition process.

We next tested whether the lack of co-localization of DAXX and SPOP cancer mutants resulted in decreased substrate ubiquitination. While GFP-cDAXX was readily ubiquitinated by WT SPOP, ubiquitination was much reduced with the SPOP cancer mutants (Figure 2D). We also observed reduced ubiquitination of full-length FLAG-tagged DAXX in the presence of SPOP mutants (Figure S2E), in agreement with previous reports showing reduced substrate ubiquitination by SPOP cancer mutants (An et al., 2014; Dai et al., 2017a, 2017b; Gan et al., 2015; Geng et al., 2013, 2014, 2017; Janouskova et al., 2017; Theurillat et al., 2014). Together, our data suggest that the underlying mechanism of reduced ubiquitination is the disruption of phase separation, which results in a failure of the SPOP cancer mutants to co-localize with DAXX.

SPOP/DAXX Bodies Are Active Ubiquitination Compartments

Given that SPOP and substrates co-localize in liquid compartments and that this co-localization is disrupted by functionally deficient, cancer-associated mutants, we tested whether the liquid compartments are indeed active for SPOP-mediated ubiquitination.

We first tested whether other subunits of the $CRL3^{SPOP}$ ubiquitin ligase are present in the SPOP/DAXX bodies. Transient expression of Myc-Cul3, the scaffold that bridges the substrate and the E2 ubiquitin-conjugating enzyme in cullin-RING ubiquitin ligases, resulted in its diffuse localization in the cytoplasm in the absence of SPOP. In contrast, Myc-Cul3 localized in nuclear speckles in the presence of SPOP and in SPOP/DAXX bodies in the presence of both SPOP and DAXX (Figure 6A). Therefore, Cul3 is recruited to liquid organelles by SPOP. HA-Rbx1, the RING protein associating with Cul3, also localizes to SPOP/DAXX bodies (Figure 6B). *In vitro*, the active form of the scaffold, neddylated Cul3/Rbx1 (N8∼Cul3/Rbx1), likewise partitioned into the mesoscale assemblies (Figure 6C). Therefore, it is plausible that $CRL3^{SPOP}$ could carry out its function in cellular SPOP-containing liquid assemblies.

We next tested whether ubiquitination activity resides in the SPOP-containing bodies. Indeed, SPOP-cDAXX-Cul3-containing bodies in cells were positive for conjugated ubiquitin as determined by positive staining with the FK2 ubiquitin antibody, which does not stain free ubiquitin (Figure 6D). To determine if the conjugated ubiquitin signal is dependent on $CRL3^{SPOP}$, we quantified ubiquitin conjugation in cells expressing Cul3 or SPOP mutants. For Cul3, we used $Cul3^{H2H5}$, a previously described mutant that interferes with SPOP binding (Furukawa and Xiong, 2005). For SPOP, we mutated residues that we predicted would affect the SPOP-Cul3 interaction (Figure S6A) and called the resulting mutant $SPOP^{CBM}$ or Cul3-binding mutant. Indeed, both $Cul3^{H2H5}$ and $SPOP^{CBM}$ led to a reduced ubiquitin signal in the SPOP/cDAXX bodies (Figures 6D and 6E). The combination of the Cul3 and the SPOP mutant further reduced the level of conjugated ubiquitin in the bodies, demonstrating that a significant fraction of conjugated ubiquitin in the SPOP/DAXX bodies stems from $CRL3^{SPOP}$-mediated ubiquitination. Furthermore, cDAXX levels in cells rose with decreasing

Figure 6. SPOP/DAXX Bodies Are Active Ubiquitination Compartments

(A) SPOP recruits Cul3 to SPOP/DAXX bodies. HeLa cells were transfected with the indicated constructs. Cul3-Myc (magenta) was detected by IF.

(B) SPOP recruits Cul3 and Rbx1 to SPOP/DAXX bodies. Cul3-Myc (blue) and Rbx1-HA (magenta) were detected by IF.

(C) Cul3 partitions into SPOP/DAXX assemblies *in vitro*. Fluorescence microscopy images of N8∼Cul3/Rbx1 (blue channel), SPOP (green), and H-cDAXX (red). The blue-only channel images were pseudo-colored to black/white for clarity. All samples contain 500 nM of each Alexa Fluor 647-N8∼Cul3/Rbx1, ORG-SPOP, and Rhodamine-H-cDAXX. Samples were in 25 mM HEPES (pH 7.5) and 150 mM NaCl (top row) and 25 mM Tris (pH 7.6), 150 mM NaCl, and 1 mM T-CEP (bottom row).

(D) Conjugated ubiquitin in SPOP/DAXX bodies depends on SPOP-Cul3 interaction. HeLa cells were transfected with the indicated constructs. Cul3-Myc (blue) and conjugated ubiquitin (magenta, with FK2 antibody) were detected by IF.

(E) Disruption of the SPOP-Cul3 interaction results in increased GFP-cDAXX levels and decreased conjugated ubiquitin levels. Quantification of signals from GFP (green bars), conjugated ubiquitin (magenta bars), and conjugated ubiquitin normalized by GFP (open bars) for n = 20 cells per condition in (D). Error bars represent SEM.

(F) Schematic representation of *in vitro* ubiquitination assay. Transfer of ubiquitin is monitored by SDS-PAGE and the incorporation of fluorescent *UB into assemblies microscopically.

(G) Ubiquitinated H-cDAXX accumulates in SPOP/H-cDAXX assemblies *in vitro*. Fluorescence microscopy and DIC images showing the time course of *in vitro* ubiquitination assays described in (F) at the indicated SPOP (green)/H-cDAXX (red) molar ratios plus 1.25 μM N8∼Cul3/Rbx1, 20 nM ARIH1, and 1.5 μM UbcH7∼*UB (blue). All reactions contain ficoll 70 as indicated, 500 nM ORG-SPOP and Rhodamine-H-cDAXX; *UB denotes stoichiometrically labeled Alexa647-Ubiquitin. See Figure S6C for images of control reaction conditions.

(H) Ubiquitination can occur in the presence or absence of SPOP-DAXX assemblies. Representative fluorescent scan of non-denaturing gels showing time course of *in vitro* ubiquitination reactions described in (F). Blue UBCH7∼*UB band diminishes and blue H-cDAXX∼*UB band appears over the course of reactions containing WT ARIH1 + ficoll70 or sucrose, but less appears in reactions containing SPOP cancer mutants.

(I) Quantification of *in vitro* ubiquitination assay from blue fluorescence intensity of assemblies in fluorescence microscopy images in (G) and Figure S6C (left) and gel band intensity of product H-cDaxx∼*UB in (H) and Figure S6C (right). Data points represent average of triplicate experiments. Error bars indicate SD.

ubiquitination in the bodies (Figure 6E), supporting our hypothesis that DAXX is ubiquitinated within the bodies and subsequently degraded. Immunoprecipitation of cDAXX and blotting for ubiquitin confirmed that cDAXX was ubiquitinated in a Cul3- and SPOP-dependent manner in these cells (Figure S6B).

To exclude the possibility that other ubiquitin ligases in the SPOP/DAXX bodies are responsible for ubiquitination, we moved to an *in vitro* ubiquitination assay with purified, recombinant proteins. We tested an array of E2 conjugating enzymes for their activity toward H-cDAXX in the presence of neddylated $CRL3^{SPOP}$, and observed the strongest activity with ARIH1/UBCH7 (Figure S6D; Scott et al., 2016). We then designed an *in vitro* ubiquitination assay to test whether transfer of ubiquitin onto H-cDAXX correlated with the appearance of ubiquitin in mesoscale assemblies. We charged UBCH7 with an equimolar amount of fluorescently labeled ubiquitin (N-terminally labeled Alexa647-UB [from here on *UB]), quenched the charging reaction with EDTA, and added *UB∼UBCH7 to pre-formed SPOP/H-cDAXX/N8∼Cul3/Rbx1 assemblies in the presence and absence of ARIH1 or a catalytically inactive mutant ($ARIH1^{C357S}$; Scott et al., 2016) (Figure 6F). This reaction mixture is competent for a single turnover (i.e., discharge of *UB and potentially transfer onto an acceptor lysine of a substrate). With live fluorescence imaging, we observed the appearance of *UB in the assemblies over time in the presence of enzymatically active ARIH1, resulting in co-localization of H-cDAXX, SPOP, and *UB (Figures 6G and 6I, left). Visualization of the reaction products by SDS-PAGE showed ubiquitinated H-cDAXX in the presence of active ARIH1, but not with an enzymatically inactive mutant or in its absence (Figures 6H and 6I, right, and Figure S6C). Both filamentous and droplet-like assemblies with WT SPOP were able to mediate activity. However, we also observed effective transfer of *UB onto DAXX in the presence of sucrose instead of ficoll (Figure 6I, right, and Figure S6C). Under these conditions, SPOP and DAXX did not form mesoscale assemblies, but the reaction went to completion with similar kinetics. The SPOP cancer mutants W131G and F133V were hardly able to mediate DAXX ubiquitination. (Figures 6G–6I).

We come to four conclusions from these results: (1) Since most of the SPOP/DAXX is concentrated in the assemblies and little protein is diffuse under phase separation conditions, the reaction must occur largely within the assemblies. (2) Filamentous and droplet-like WT SPOP/DAXX assemblies have similar activities, in agreement with the similar protein mobilities we observed within them. (3) The assemblies do not enhance enzymatic turnover compared to diffuse reactions. This may be explained by competition between enhancing and decelerating factors such as high local concentrations and increased viscosity in the dense assemblies, respectively. FRAP indeed shows a considerable immobile fraction of DAXX and SPOP, in particular, in all assemblies (Figure S4G). (4) SPOP cancer mutants show reduced activity, in agreement with their reduced ability to assemble and their expected higher off-rates.

Our results therefore support a model in which the UB-charged E2 diffuses into the SPOP/DAXX assemblies and $CRL3^{SPOP}$-mediated ubiquitination occurs within the assemblies. We therefore propose that SPOP-mediated ubiquitination occurs largely within membraneless organelles in the cell (e.g., within SPOP/DAXX bodies).

SPOP Phase Separates with AR

To determine whether the synergistic recruitment of SPOP and DAXX to liquid organelles is a general feature of other SPOP substrates, we predicted SB motifs in a number of known SPOP substrates from the literature. Most of SPOP's substrates likely harbor weak SB motifs in their intrinsically disordered regions, in addition to the experimentally confirmed motifs in the literature (Table S6). We thus hypothesized that SPOP recruits at least a subset of its substrates via phase separation. To test this hypothesis, we generated a large N-terminal fragment of AR that contained the majority of the predicted SB motifs (nAR; Figure 7A). AR is a well-known substrate of SPOP, and increased AR levels caused by SPOP cancer mutations likely contribute to prostate cancer pathogenesis (An et al., 2014). nAR had the ability to form liquid-like droplets at high concentrations *in vitro*, and their formation was strongly enhanced by SPOP (Figure 7B). The nAR/SPOP assemblies had similar properties to SPOP/DAXX assemblies, ranging from filamentous to liquid droplet-like depending on the nAR/SPOP molar ratio (Figure 7B). Similarly, transiently expressed full-length GFP-AR and SPOP-mCherry co-localized in punctate membraneless bodies in HeLa cells (Figure 7C). nAR binding to $SPOP^{28–359}$ is enhanced compared to binding to the MATH domain (Figure S7A), supporting the existence of multiple SB motifs in nAR, which can mediate phase separation with multivalent SPOP, analogously to DAXX. We thus propose that SPOP has the ability to undergo phase separation with multivalent substrates as a general mechanism for targeting substrates.

SPOP Oligomeric Interfaces Are Evolutionarily Conserved

Since SPOP multivalency was required for phase separation with substrates, we wondered whether the ability to form higher-order oligomers was under evolutionary pressure. We employed a co-evolutionary coupling analysis (Ekeberg et al., 2013; Marks et al., 2011), which assesses statistical coupling of correlated mutations between all residue positions, from the covariation in sequence alignments of ∼2,600 SPOP orthologs (see Table S7 for their taxonomy). In particular, we assessed whether there was any covariation between residues in SPOP that could not be explained by intramolecular contacts within a SPOP monomer and that instead coincide with the known intermolecular interfaces.

The contact map of the SPOP oligomer highlights pairs of residues that are in proximity across the oligomerization interfaces, but not within SPOP monomers (Marzahn et al., 2016; van Geersdaele et al., 2013) (Figure 7D). Comparison with the 600 strongest nonlocal ($|i - j| > 3$) couplings reveal close agreement between coevolving residue pairs and contacts in the protein structure, confirming that co-evolution reports on spatial proximity. Notably, we also observe numerous couplings across the BTB and the BACK interfaces (Figure 7D).

We further analyzed the patterns of the evolutionary couplings to find natural groups of coevolving residues, also called evolutionary domains (Granata et al., 2017). A subdivision into two

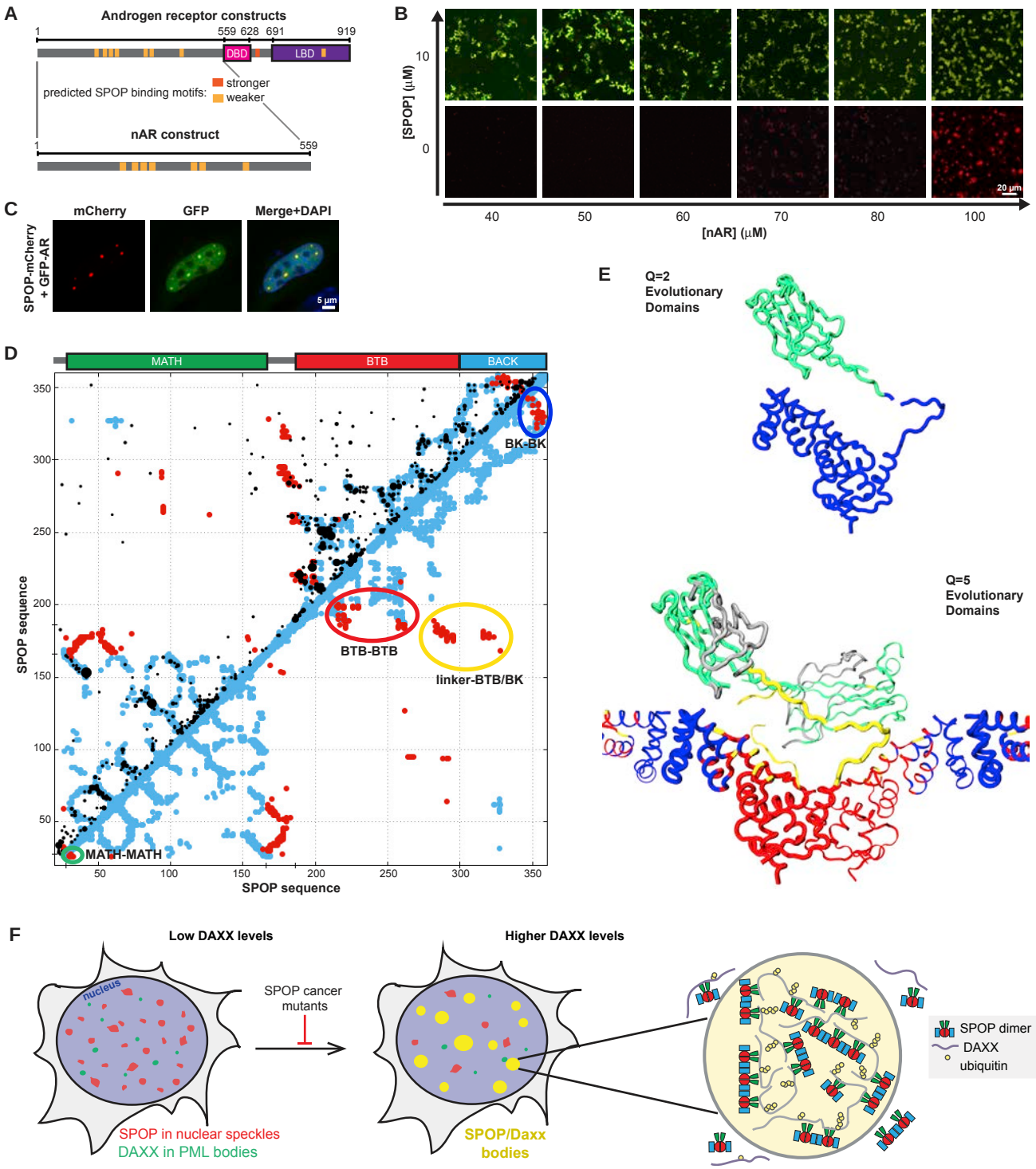


Figure 7. SPOP Phase Separates with Androgen Receptor and May Phase Separate with Other Substrates in an Evolutionarily Conserved Fashion

(A) Sequence schematic for AR and the N-terminal fragment used for *in vitro* experiments (nAR). AR contains an N-terminal disordered domain, DNA-binding domain (DBD), and ligand-binding domain (LBD) (Centenera et al., 2008).

(B) SPOP phase separates *in vitro* with nAR. Fluorescence microscopy images of SPOP (green) and nAR (red). All samples contain 8% w/v ficoll 70 and 500 nM ORG-SPOP and/or Rhodamine-nAR. Samples were in 20 mM $NaPO_4$ (pH 7.4), 60 mM NaCl, 2 mM T-CEP, and 1 mM EDTA.

(C) SPOP co-localizes with the AR in cells. HeLa cells were transfected with the indicated constructs and analyzed by confocal microscopy.

(legend continued on next page)

evolutionary domains results in a good fit to the data (Figure S7B) and points to the BTB and BACK domains as a single coevolving unit, while the MATH domain forms a separate unit (Figure 7E, top, and Figure S7C). Progressively finer subdivision next results in a good fit for 5 evolutionary domains (Figures S7B and S7C). With this subdivision, it becomes clear that the residues of the BACK domain interfacial helices coevolve across the interface, while the rest of the BACK domain coevolves with the BTB domain (Figure 7E, bottom). Even finer subdivision shows a similar co-evolution of BTB interfacial residues across the interface (Figure S7D). Based on this analysis, the BTB and BACK domains in SPOP appear to have coevolved their ability to dimerize synergistically into higher-order oligomers, and our results suggest that this property is under evolutionary pressure. This finding supports the possibility that SPOP has evolved multivalent properties to target substrates through phase separation.

DISCUSSION

SPOP localizes to several membraneless organelles in the nucleus, but the mechanism underlying its redistribution between these organelles, and how this is related to substrate targeting, have so far been unclear. Here, we show that substrates trigger SPOP co-localization and CRL activity by mediating phase separation and that SPOP cancer mutants disrupt both processes (Figure 7F). We propose the disruption of phase separation as the mechanism underlying SPOP-mutant-driven oncogenesis. The ability of SPOP to form higher-order oligomers, where the SPOP concentration determines the size of the oligomers and therefore the valency of SPOP for substrates, gives rise to a rich phase diagram; SPOP can undergo oligomerization and/or polymerization processes and LLPS with multivalent substrates, depending on concentrations and molar ratios. LLPS resulting in droplet-like assemblies seems to be the basis for co-localization and activity in cells.

Phase Separation May Allow Targeting Substrates in Membraneless Organelles

Phase separation is a potential mechanism for concentrating enzymes and substrates and enhancing turnover (Banani et al., 2017; Li et al., 2012). *In vitro*, condensed BugZ droplets enhance tubulin polymerization (Jiang et al., 2015), and Nephrin/Grb2 phase separation enhances actin polymerization (Li et al., 2012); whether the respective liquid compartments in the cell also enhance polymerization is unclear. While the nucleolus is a phase-separated, enzymatically active compartment (Berry et al., 2015; Mitrea et al., 2016), it is less clear whether phase separation is needed for activity. Smaller oligomeric assemblies may be able to sustain the activities found in typical membraneless organelles (Smith et al., 2016; Wallace et al., 2015). Many membraneless organelles may be sequestering proteins, RNA, or DNA rather than forming active compartments. Here, we show strong evidence that SPOP- and substrate-containing membraneless organelles contain enzymatic activity.

But given our observation that SPOP ubiquitinates the substrate DAXX equally efficiently in the absence of large assemblies, why has SPOP evolved the ability to phase separate? In cells, when not incorporated in SPOP/DAXX bodies, SPOP and DAXX are not largely diffuse but incorporated into two different nuclear bodies (i.e., nuclear speckles and PML bodies). Presumably only a small fraction of diffuse protein has the possibility to interact. In contrast, in our *in vitro* ubiquitination assay, the fraction of unassembled protein is not sequestered in separate bodies but is available for interaction and turnover. Importantly, there is evidence for several SPOP substrates to localize to nuclear bodies (Kwon et al., 2006; Li et al., 2000; Weidtkamp-Peters et al., 2008; Klokk et al., 2007; Tomura et al., 2001; Tyagi et al., 2000; Zhang et al., 2014). ERG and BET proteins may also participate in phase separation processes proposed for transcription factories and chromatin (Hnisz et al., 2017; Larson et al., 2017; Strom et al., 2017). Indeed, SPOP substrates are predicted to have a high propensity to phase separate, which is well above those of proteins with PDB structures (Figure S7E) (Vernon et al., 2018). SPOP may have evolved an ability to phase separate in order to target substrates localized to membraneless organelles. Phase separation may be an efficient mechanism to achieve specific co-localization with substrates in various membraneless organelles. While diffuse samples without mesoscale assemblies are active *in vitro*, the full activity in cells requires co-localization of SPOP and substrates via phase separation.

Redistribution between Membraneless Organelles as a Function of Available Substrates

Our results that SPOP cancer mutants with decreased substrate binding maintained localization to nuclear speckles support a view in which nuclear speckles act as storage sites for SPOP in the absence of high substrate levels in the cell, triggered either by the presence of pseudo-substrates in nuclear speckles or by molecules retaining oligomeric SPOP via other interactions. We propose that the rise of substrate levels, potentially accompanied by additional signals, leads to substrate phase separation with SPOP and recruitment of SPOP to the respective organelle.

(D) Covariation analysis of SPOP shows evolutionary coupling across the BTB and BACK interfaces. Co-evolutionary couplings in SPOP from covariation of ~2,600 SPOP homologs (Table S7) sharing all three structural domains. The co-evolutionary couplings (top 600) are reported in the upper triangle of the matrix as black dots with size proportional to the relative coupling strength, overlapping both intra- and intermolecular contacts. The couplings are compared to contacts between pairs of residues with a distance of up to 5 Å between side-chain heavy atoms based on a structural model of $SPOP^{28–359}$ (built using two available crystal structures, PDB: 3HQI [Zhuang et al., 2009] and PDB: 4HS2 [van Geersdaele et al., 2013], and no further assumptions); intra- and intermolecular contacts are shown in blue and red, respectively.
(E) BTB and BACK interface residues coevolve with residues across the interface, not with domain core residues. Evolutionary domains, obtained by the analysis of the patterns of couplings (Granata et al., 2017) are reported on the SPOP monomer structure model for the subdivision in 2 groups of coevolving residues (Q = 2) (top), and on the oligomer structure model for 5 groups of coevolving residues (Q = 5) (bottom). Other meaningful subdivisions are reported in Figures S7B–S7D.
(F) Schematic of proposed mechanism. SPOP phase separates with multivalent substrates and is able to target and ubiquitinate substrates localized to membrane-less organelles. SPOP cancer mutants are defective at phase separation and therefore co-localization and ubiquitination.

Since the signals driving SPOP redistribution toward substrates are not well understood, we here used transient expression of SPOP and substrates to trigger assembly. The resulting dense phase may form a separate organelle, be miscible with a preexisting organelle, or localize to the original organelle of the substrate. Several cases have recently been described in which phase separation of proteins or protein complexes supports their recruitment to preexisting membraneless organelles, including the recruitment of the miRNA-mediated silencing complex (miRISC) to P bodies (Sheu-Gruttadauria and MacRae, 2018) and of UBQLN2 to stress granules (Dao et al., 2018).

Additional regulatory processes, such as post-translational modification, could serve to change the apparent substrate concentration sensed by SPOP. Phosphorylation of SB motifs negatively impacts SPOP binding (Zhuang et al., 2009) and could increase the tolerated substrate concentration before it phase separates with SPOP. Furthermore, phosphorylation of SPOP was recently shown to modulate its function (Zhang et al., 2018). Nonequilibrium regulatory mechanisms may influence phase separation and therefore fine-tune SPOP substrate levels. Future studies will provide further insights into how SPOP moves from one organelle to another.

SPOP Evolution and Conservation Highlight the Functional Importance of SPOP Oligomerization

The multivalency of SPOP toward substrates, generated through linear SPOP oligomerization, is evolutionarily encoded in the sequence. We have previously reported that self-association deficient SPOP mutants can disrupt normal SPOP function through dominant-negative effects in a fly model (Marzahn et al., 2016). Furthermore, prostate cancer patients with SPOP mutations typically have one normal allele (i.e., no loss of heterozygosity). Intriguingly, we find that SPOP is extremely highly conserved within the human population. For example, we find no common missense variants of SPOP present at an allele frequency greater than 10^{-4} (Table S8) in a database of more than 100,000 sequenced exomes (Lek et al., 2016). In contrast, we find multiple missense mutations in the very similar SPOPL, which is not able to form higher-order oligomers (Errington et al., 2012). We therefore hypothesize that dominant-negative phenotypes resulting from SPOP mutations cause this protein to be extremely conserved. SPOP self-association is evolutionary conserved and functionally critical, and the resulting multivalency is required for SPOP-mediated phase separation.

Conclusion

We propose that phase separation is an evolutionary adaptation of the SPOP CRL substrate adaptor to target substrates in membraneless organelles. SPOP mutants disrupt not only substrate binding but also phase separation, resulting in the failure to colocalize with and turn over the substrate. There is precedence that mislocalization of SPOP can result in oncogenesis; incorrect localization of SPOP into the cytoplasm under hypoxic conditions unleashes CRL activity on the tumor suppressor PTEN, which is not usually a SPOP substrate (Li et al., 2014). Other ubiquitin ligases also form higher-order oligomers (Yin et al., 2009), and multivalent interactions are prevalent in the ubiquitin proteasome pathway (Liu and Walters, 2010). Triggering activation of ubiquitin ligases by their substrates via phase separation may therefore be a common principle for attaining proteostasis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
- METHOD DETAILS
  - Prediction of SB motifs
  - Plasmids
  - Transfections
  - Immunofluorescence
  - Live Cell Imaging
  - Immunoprecipitation and Western Blots
  - Protein Purification
  - Microscopy analysis for *in vitro* LLPS
  - *In vitro* crosslinking reactions
  - NMR data collection and analysis
  - Peptide Synthesis and Preparation
  - Fluorescence Anisotropy
  - FRAP measurements
  - *In vitro* ubiquitination assays
  - Co-evolutionary analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Partition Coefficient for cDAXX and cDAXX^{0SB}
  - GFP-DAXX and V5-SPOP Signal Intensity
  - K_D Determination from Anisotropy Measurements
  - Quantification of *in vitro* assemblies
  - FRAP Analysis and Fitting
  - Quantification of *in vitro* ubiquitination

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, eight tables, and two videos and can be found with this article online at https://doi.org/10.1016/j.molcel.2018.08.027.

ACKNOWLEDGMENTS

We thank Anne Bremer, Nicole Milkovic, Yasmine Valentin-Vega, Kevin Freeman, Melissa Mann, Linda Hendershot, Rohit Pappu, Richard Kriwacki, and J. Paul Taylor for technical help and useful discussions. Microscopy images were acquired at the Cell & Tissue Imaging Center, which is supported by SJCRH and NCI (grant P30 CA021765). We thank Victoria Frohlich, Aaron Pitre, Jennifer Peters, and Sharon King for technical help with microscopy. We thank Haojie Huang, Mayo Clinic College of Medicine, for the generous gift of a PC-3 cell line. This work was funded by a V Foundation Scholar Grant (T.M.), NIH grant R01GM112846 (T.M.), the American Lebanese Syrian Associated Charities (T.M. and B.A.S.), the Howard Hughes Medical Institute, and NIH grant R37GM069530 (B.A.S.). B.A.S. is a Howard Hughes Medical Institute Investigator. X.S. acknowledges support from IRB, ICREA, Obra Social "la Caixa," MINECO (BIO2015-70092-R), and ERC (CONCERT, contract 648201). IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO (Government of Spain). K.L.L acknowledges a Hallas-Møller and a Biotechnology grant from the Novo Nordisk Foundation.

AUTHOR CONTRIBUTIONS

Conceptualization, J.J.B., J.H.O., E.W.M., D.C.S., K.L.-L., B.A.S., and T.M.; Investigation, J.J.B., J.H.O., D.C.S., E.S., E.W.M., N.S., D.G., and M.R.M.; Formal Analysis, J.J.B., J.H.O., D.C.S, E.W.M., D.G., and T.M.; Resources, K.L.-L., X.S., B.A.S., and T.M.; Funding Acquisition and Supervision, K.L.-L., X.S., B.A.S., and T.M.; Visualization and Writing, J.J.B., J.H.O., D.C.S., E.W.M., D.G., K.L.-L., B.A.S., and T.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 15, 2017
Revised: May 29, 2018
Accepted: August 3, 2018
Published: September 20, 2018

REFERENCES

An, J., Wang, C., Deng, Y., Yu, L., and Huang, H. (2014). Destruction of full-length androgen receptor by wild-type SPOP, but not prostate-cancer-associated mutants. Cell Rep. *6*, 657–669.

Anthis, N.J., and Clore, G.M. (2013). Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. Protein Sci. *22*, 851–858.

Axelrod, D., Koppel, D.E., Schlessinger, J., Elson, E., and Webb, W.W. (1976). Mobility measurement by analysis of fluorescence photobleaching recovery kinets. Biophys. J. *16*, 1055–1059.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. *18*, 285–298.

Bermel, W., Bertini, I., Duma, L., Felli, I.C., Emsley, L., Pierattelli, R., and Vasos, P.R. (2005). Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. Angew. Chem. Int. Ed. Engl. *44*, 3089–3092.

Bermel, W., Bertini, I., Felli, I.C., Kümmerle, R., and Pierattelli, R. (2006a). Novel 13C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. J. Magn. Reson. *178*, 56–64.

Bermel, W., Bertini, I., Felli, I.C., Lee, Y.M., Luchinat, C., and Pierattelli, R. (2006b). Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J. Am. Chem. Soc. *128*, 3918–3919.

Berry, J., Weber, S.C., Vaidya, N., Haataja, M., and Brangwynne, C.P. (2015). RNA transcription modulates phase transition-driven nuclear body assembly. Proc. Natl. Acad. Sci. USA *112*, E5237–E5245.

Boke, E., Ruer, M., Wühr, M., Coughlin, M., Lemaitre, R., Gygi, S.P., Alberti, S., Drechsel, D., Hyman, A.A., and Mitchison, T.J. (2016). Amyloid-like self-assembly of a cellular compartment. Cell *166*, 637–650.

Boysen, G., Barbieri, C.E., Prandi, D., Blattner, M., Chae, S.S., Dahija, A., Nataraj, S., Huang, D., Marotz, C., Xu, L., et al. (2015). SPOP mutation leads to genomic instability in prostate cancer. eLife *4*, e09207.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729–1732.

Cai, X., Chen, J., Xu, H., Liu, S., Jiang, Q.X., Halfmann, R., and Chen, Z.J. (2014). Prion-like polymerization underlies signal transduction in antiviral immune defense and inflammasome activation. Cell *156*, 1207–1222.

Centenera, M.M., Harris, J.M., Tilley, W.D., and Butler, L.M. (2008). The contribution of different androgen receptor domains to receptor dimerization and signaling. Mol. Endocrinol. *22*, 2373–2382.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.

Dai, X., Gan, W., Li, X., Wang, S., Zhang, W., Huang, L., Liu, S., Zhong, Q., Guo, J., Zhang, J., et al. (2017a). Prostate cancer-associated SPOP mutations confer resistance to BET inhibitors through stabilization of BRD4. Nat. Med. *23*, 1063–1071.

Dai, X., Wang, Z., and Wei, W. (2017b). SPOP-mediated degradation of BRD4 dictates cellular sensitivity to BET inhibitors. Cell Cycle *16*, 2326–2329.

Dao, T.P., Kolaitis, R.M., Kim, H.J., O'Donovan, K., Martyniak, B., Colicino, E., Hehnly, H., Taylor, J.P., and Castaneda, C.A. (2018). Ubiquitin modulates liquid-liquid phase separation of UBQLN2 via disruption of multivalent interactions. Mol. Cell *69*, 965–978.e966.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Duda, D.M., Borg, L.A., Scott, D.C., Hunt, H.W., Hammel, M., and Schulman, B.A. (2008). Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. Cell *134*, 995–1006.

Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M., and Aurell, E. (2013). Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *87*, 012707.

Eliezer, D. (2009). Biophysical characterization of intrinsically disordered proteins. Curr. Opin. Struct. Biol. *19*, 23–30.

Errington, W.J., Khan, M.Q., Bueler, S.A., Rubinstein, J.L., Chakrabartty, A., and Privé, G.G. (2012). Adaptor protein self-assembly drives the control of a cullin-RING ubiquitin ligase. Structure *20*, 1141–1153.

Escobar-Cabrera, E., Lau, D.K., Giovinazzi, S., Ishov, A.M., and McIntosh, L.P. (2010). Structural characterization of the DAXX N-terminal helical bundle domain and its complex with Rassf1C. Structure *18*, 1642–1653.

Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A., and Eddy, S.R. (2015). HMMER web server: 2015 update. Nucleic Acids Res. *43* (W1), W30–W38.

Furukawa, M., and Xiong, Y. (2005). BTB protein Keap1 targets antioxidant transcription factor Nrf2 for ubiquitination by the Cullin 3-Roc1 ligase. Mol. Cell. Biol. *25*, 162–171.

Gan, W., Dai, X., Lunardi, A., Li, Z., Inuzuka, H., Liu, P., Varmeh, S., Zhang, J., Cheng, L., Sun, Y., et al. (2015). SPOP promotes ubiquitination and degradation of the ERG oncoprotein to suppress prostate cancer progression. Mol. Cell *59*, 917–930.

Gao, K., Jin, X., Tang, Y., Ma, J., Peng, J., Yu, L., Zhang, P., and Wang, C. (2015). Tumor suppressor SPOP mediates the proteasomal degradation of progesterone receptors (PRs) in breast cancer cells. Am. J. Cancer Res. *5*, 3210–3220.

Geng, C., He, B., Xu, L., Barbieri, C.E., Eedunuri, V.K., Chew, S.A., Zimmermann, M., Bond, R., Shou, J., Li, C., et al. (2013). Prostate cancer-associated mutations in speckle-type POZ protein (SPOP) regulate steroid receptor coactivator 3 protein turnover. Proc. Natl. Acad. Sci. USA *110*, 6997–7002.

Geng, C., Rajapakshe, K., Shah, S.S., Shou, J., Eedunuri, V.K., Foley, C., Fiskus, W., Rajendran, M., Chew, S.A., Zimmermann, M., et al. (2014). Androgen receptor is the key transcriptional mediator of the tumor suppressor SPOP in prostate cancer. Cancer Res. *74*, 5631–5643.

Geng, C., Kaochar, S., Li, M., Rajapakshe, K., Fiskus, W., Dong, J., Foley, C., Dong, B., Zhang, L., Kwon, O.J., et al. (2017). SPOP regulates prostate epithelial cell proliferation and promotes ubiquitination and turnover of c-MYC oncoprotein. Oncogene *36*, 4767–4777.

Granata, D., Ponzoni, L., Micheletti, C., and Carnevale, V. (2017). Patterns of coevolving amino acids unveil structural and dynamical domains. Proc. Natl. Acad. Sci. USA *114*, E10612–E10621.

Halfmann, R. (2016). A glass menagerie of low complexity sequences. Curr. Opin. Struct. Biol. *38*, 18–25.

Harmon, T.S., Holehouse, A.S., Rosen, M.K., and Pappu, R.V. (2017). Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. eLife *6*, e30294.

Hernández-Muñoz, I., Lund, A.H., van der Stoop, P., Boutsma, E., Muijrers, I., Verhoeven, E., Nusinow, D.A., Panning, B., Marahrens, Y., and van Lohuizen, M. (2005). Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. Proc. Natl. Acad. Sci. USA *102*, 7635–7640.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A phase separation model for transcriptional control. Cell *169*, 13–23.

Huang, D.T., Zhuang, M., Ayrault, O., and Schulman, B.A. (2008). Identification of conjugation specificity determinants unmasks vestigial preference for ubiquitin within the NEDD8 E2. Nat. Struct. Mol. Biol. *15*, 280–287.

Janouskova, H., El Tekle, G., Bellini, E., Udeshi, N.D., Rinaldi, A., Ulbricht, A., Bernasocchi, T., Civenni, G., Losa, M., Svinkina, T., et al. (2017). Opposing effects of cancer-type-specific SPOP mutants on BET protein degradation and sensitivity to BET inhibitors. Nat. Med. *23*, 1046–1054.

Jiang, H., Wang, S., Huang, Y., He, X., Cui, H., Zhu, X., and Zheng, Y. (2015). Phase transition of spindle-associated protein regulate spindle apparatus assembly. Cell *163*, 108–122.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

Keller, R.L.J. (2004). The Computer Aided Resonance Assignment Tutorial, First Edition (Cantina Verlag).

Kent, D., Bush, E.W., and Hooper, J.E. (2006). Roadkill attenuates Hedgehog responses through degradation of Cubitus interruptus. Development *133*, 2001–2010.

Kim, B., Nam, H.J., Pyo, K.E., Jang, M.J., Kim, I.S., Kim, D., Boo, K., Lee, S.H., Yoon, J.B., Baek, S.H., and Kim, J.H. (2011). Breast cancer metastasis suppressor 1 (BRMS1) is destabilized by the Cul3-SPOP E3 ubiquitin ligase complex. Biochem. Biophys. Res. Commun. *415*, 720–726.

Kim, M.S., Je, E.M., Oh, J.E., Yoo, N.J., and Lee, S.H. (2013). Mutational and expressional analyses of SPOP, a candidate tumor suppressor gene, in prostate, gastric and colorectal cancers. APMIS *121*, 626–633.

Klokk, T.I., Kurys, P., Elbi, C., Nagaich, A.K., Hendarwanto, A., Slagsvold, T., Chang, C.Y., Hager, G.L., and Saatcioglu, F. (2007). Ligand-specific dynamics of the androgen receptor at its response element in living cells. Mol. Cell. Biol. *27*, 1823–1843.

Kwon, J.E., La, M., Oh, K.H., Oh, Y.M., Kim, G.R., Seol, J.H., Baek, S.H., Chiba, T., Tanaka, K., Bang, O.S., et al. (2006). BTB domain-containing speckle-type POZ protein (SPOP) serves as an adaptor of Daxx for ubiquitination by Cul3-based ubiquitin ligase. J. Biol. Chem. *281*, 12664–12672.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.

Le Gallo, M., O'Hara, A.J., Rudd, M.L., Urick, M.E., Hansen, N.F., O'Neil, N.J., Price, J.C., Zhang, S., England, B.M., Godwin, A.K., et al.; NIH Intramural Sequencing Center (NISC) Comparative Sequencing Program (2012). Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. Nat. Genet. *44*, 1310–1315.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Li, H., Leo, C., Zhu, J., Wu, X., O'Neil, J., Park, E.J., and Chen, J.D. (2000). Sequestration and inhibition of Daxx-mediated transcriptional repression by PML. Mol. Cell. Biol. *20*, 1784–1796.

Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

Li, G., Ci, W., Karmakar, S., Chen, K., Dhar, R., Fan, Z., Guo, Z., Zhang, J., Ke, Y., Wang, L., et al. (2014). SPOP promotes tumorigenesis by acting as a key regulatory hub in kidney cancer. Cancer Cell *25*, 455–468.

Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. Mol. Cell *60*, 208–219.

Liu, F., and Walters, K.J. (2010). Multitasking with ubiquitin through multivalent interactions. Trends Biochem. Sci. *35*, 352–360.

Mackenzie, I.R., Nicholson, A.M., Sarkar, M., Messing, J., Purice, M.D., Pottier, C., Annu, K., Baker, M., Perkerson, R.B., Kurti, A., et al. (2017). TIA1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. Neuron *95*, 808–816.e809.

Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R., and Sander, C. (2011). Protein 3D structure computed from evolutionary sequence variation. PLoS ONE *6*, e28766.

Marzahn, M.R., Marada, S., Lee, J., Nourse, A., Kenrick, S., Zhao, H., Ben-Nissan, G., Kolaitis, R.M., Peters, J.L., Pounds, S., et al. (2016). Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. EMBO J. *35*, 1254–1275.

Mitrea, D.M., Cika, J.A., Guy, C.S., Ban, D., Banerjee, P.R., Stanley, C.B., Nourse, A., Deniz, A.A., and Kriwacki, R.W. (2016). Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. eLife *5*, e13571.

Mittag, T., and Forman-Kay, J.D. (2007). Atomic-level characterization of disordered protein ensembles. Curr. Opin. Struct. Biol. *17*, 3–14.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. Cell *163*, 123–133.

Monahan, Z., Ryan, V.H., Janke, A.M., Burke, K.A., Rhoads, S.N., Zerze, G.H., O'Meally, R., Dignon, G.L., Conicella, A.E., Zheng, W., et al. (2017). Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. EMBO J. *36*, 2951–2967.

Murakami, T., Qamar, S., Lin, J.Q., Schierle, G.S., Rees, E., Miyashita, A., Costa, A.R., Dodd, R.B., Chan, F.T., Michel, C.H., et al. (2015). ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. Neuron *88*, 678–690.

Nagai, Y., Kojima, T., Muro, Y., Hachiya, T., Nishizawa, Y., Wakabayashi, T., and Hagiwara, M. (1997). Identification of a novel nuclear speckle-type protein, SPOP. FEBS Lett. *418*, 23–26.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., and Baldwin, A.J. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Mol. Cell *57*, 936–947.

Ohta, T., Michel, J.J., Schottelius, A.J., and Xiong, Y. (1999). ROC1, a homolog of APC11, represents a family of cullin partners with an associated ubiquitin ligase activity. Mol. Cell *3*, 535–541.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. Cell *162*, 1066–1077.

Petroski, M.D., and Deshaies, R.J. (2005). Function and regulation of cullin-RING ubiquitin ligases. Nat. Rev. Mol. Cell Biol. *6*, 9–20.

Pierce, W.K., Grace, C.R., Lee, J., Nourse, A., Marzahn, M.R., Watson, E.R., High, A.A., Peng, J., Schulman, B.A., and Mittag, T. (2016). Multiple weak linear motifs enhance recruitment and processivity in SPOP-mediated substrate ubiquitination. J. Mol. Biol. *428*, 1256–1271.

Roehrl, M.H., Wang, J.Y., and Wagner, G. (2004). A general framework for development and data analysis of competitive high-throughput screens for

small-molecule inhibitors of protein-protein interactions by fluorescence polarization. Biochemistry *43*, 16056–16066.

Sahu, D., Bastidas, M., and Showalter, S.A. (2014). Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. Anal. Biochem. *449*, 17–25.

Sakaue, T., Sakakibara, I., Uesugi, T., Fujisaki, A., Nakashiro, K.I., Hamakawa, H., Kubota, E., Joh, T., Imai, Y.K., Izutani, H., and Higashiyama, S. (2017). The CUL3-SPOP-DAXX axis is a novel regulator of VEGFR2 expression in vascular endothelial cells. Sci. Rep. *7*, 42845.

Scott, D.C., Sviderskiy, V.O., Monda, J.K., Lydeard, J.R., Cho, S.E., Harper, J.W., and Schulman, B.A. (2014). Structure of a RING E3 trapped in action reveals ligation mechanism for the ubiquitin-like protein NEDD8. Cell *157*, 1671–1684.

Scott, D.C., Rhee, D.Y., Duda, D.M., Kelsall, I.R., Olszewski, J.L., Paulo, J.A., de Jong, A., Ovaa, H., Alpi, A.F., Harper, J.W., et al. (2016). Two distinct types of E3 ligases work in unison to regulate substrate ubiquitylation. Cell *166*, 1198–1214.e1124.

Sheu-Gruttadauria, J., and MacRae, I.J. (2018). Phase transitions in the assembly and function of human miRISC. Cell *173*, 946–957.e916.

Small, E., Eggler, A., and Mesecar, A.D. (2010). Development of an efficient E. coli expression and purification system for a catalytically active, human Cullin3-RINGBox1 protein complex and elucidation of its quaternary structure with Keap1. Biochem. Biophys. Res. Commun. *400*, 471–475.

Smith, J., Calidas, D., Schmidt, H., Lu, T., Rasoloson, D., and Seydoux, G. (2016). Spatial patterning of P granules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3. eLife *5*, e21337.

Stenoien, D.L., Cummings, C.J., Adams, H.P., Mancini, M.G., Patel, K., DeMartino, G.N., Marcelli, M., Weigel, N.L., and Mancini, M.A. (1999). Polyglutamine-expanded androgen receptors form aggregates that sequester heat shock proteins, proteasome components and SRC-1, and are suppressed by the HDJ-2 chaperone. Hum. Mol. Genet. *8*, 731–741.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241–245.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B., and Wu, C.H.; UniProt Consortium (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics *31*, 926–932.

Takeuchi, K., Heffron, G., Sun, Z.Y., Frueh, D.P., and Wagner, G. (2010). Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments. J. Biomol. NMR *47*, 271–282.

Theurillat, J.P., Udeshi, N.D., Errington, W.J., Svinkina, T., Baca, S.C., Pop, M., Wild, P.J., Blattner, M., Groner, A.C., Rubin, M.A., et al. (2014). Prostate cancer. Ubiquitylome analysis identifies dysregulation of effector substrates in SPOP-mutant prostate cancer. Science *346*, 85–89.

Tomura, A., Goto, K., Morinaga, H., Nomura, M., Okabe, T., Yanase, T., Takayanagi, R., and Nawata, H. (2001). The subnuclear three-dimensional image analysis of androgen receptor fused to green fluorescence protein. J. Biol. Chem. *276*, 28395–28401.

Tyagi, R.K., Lavrovsky, Y., Ahn, S.C., Song, C.S., Chatterjee, B., and Roy, A.K. (2000). Dynamics of intracellular movement and nucleocytoplasmic recycling of the ligand-activated androgen receptor in living cells. Mol. Endocrinol. *14*, 1162–1174.

van den Ent, F., and Löwe, J. (2006). RF cloning: a restriction-free method for inserting target genes into plasmids. J. Biochem. Biophys. Methods *67*, 67–74.

van Geersdaele, L.K., Stead, M.A., Harrison, C.M., Carr, S.B., Close, H.J., Rosbrook, G.O., Connell, S.D., and Wright, S.C. (2013). Structural basis of high-order oligomerization of the cullin-3 adaptor SPOP. Acta Crystallogr. D Biol. Crystallogr. *69*, 1677–1684.

Vernon, R.M., Chong, P.A., Tsang, B., Kim, T.H., Bah, A., Farber, P., Lin, H., and Forman-Kay, J.D. (2018). Pi-Pi contacts are an overlooked protein feature relevant to phase separation. eLife *7*, e31486.

Walden, H., Podgorski, M.S., and Schulman, B.A. (2003). Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. Nature *422*, 330–334.

Wallace, E.W., Kear-Scott, J.L., Pilipenko, E.V., Schwartz, M.H., Laskowski, P.R., Rojek, A.E., Katanski, C.D., Riback, J.A., Dion, M.F., Franks, A.M., et al. (2015). Reversible, specific, active aggregates of endogenous proteins assemble upon heat stress. Cell *162*, 1286–1298.

Weidtkamp-Peters, S., Lenser, T., Negorev, D., Gerstner, N., Hofmann, T.G., Schwanitz, G., Hoischen, C., Maul, G., Dittrich, P., and Hemmerich, P. (2008). Dynamics of component exchange at PML nuclear bodies. J. Cell Sci. *121*, 2731–2743.

Weisemann, R., Rüterjans, H., and Bermel, W. (1993). 3D triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins. J. Biomol. NMR *3*, 113–120.

Wu, H., and Fuxreiter, M. (2016). The structure and dynamics of higher-order assemblies: amyloids, signalosomes, and granules. Cell *165*, 1055–1066.

Yin, Q., Lin, S.C., Lamothe, B., Lu, M., Lo, Y.C., Hura, G., Zheng, L., Rich, R.L., Campos, A.D., Myszka, D.G., et al. (2009). E2 interaction and dimerization in the crystal structure of TRAF6. Nat. Struct. Mol. Biol. *16*, 658–666.

Zhang, Q., Shi, Q., Chen, Y., Yue, T., Li, S., Wang, B., and Jiang, J. (2009). Multiple Ser/Thr-rich degrons mediate the degradation of Ci/Gli by the Cul3-HIB/SPOP E3 ubiquitin ligase. Proc. Natl. Acad. Sci. USA *106*, 21191–21196.

Zhang, D., Wang, H., Sun, M., Yang, J., Zhang, W., Han, S., and Xu, B. (2014). Speckle-type POZ protein, SPOP, is involved in the DNA damage response. Carcinogenesis *35*, 1691–1697.

Zhang, H., Elbaum-Garfinkle, S., Langdon, E.M., Taylor, N., Occhipinti, P., Bridges, A.A., Brangwynne, C.P., and Gladfelter, A.S. (2015). RNA controls PolyQ protein phase transitions. Mol. Cell *60*, 220–230.

Zhang, J., Bu, X., Wang, H., Zhu, Y., Geng, Y., Nihira, N.T., Tan, Y., Ci, Y., Wu, F., Dai, X., et al. (2018). Cyclin D-CDK4 kinase destabilizes PD-L1 via cullin 3-SPOP to control cancer immune surveillance. Nature *553*, 91–95.

Zhuang, M., Calabrese, M.F., Liu, J., Waddell, M.B., Nourse, A., Hammel, M., Miller, D.J., Walden, H., Duda, D.M., Seyedin, S.N., et al. (2009). Structures of SPOP-substrate complexes: insights into molecular architectures of BTB-Cul3 ubiquitin ligases. Mol. Cell *36*, 39–50.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse Monoclonal anti-SC35	Abcam	Cat# ab11826; RRID:AB_298608
Mouse Monoclonal anti-PML	Santa Cruz	Cat#sc-966; RRID:AB_628162
Rabbit Polyclonal anti-Coilin	Santa Cruz	Cat#sc32860; RRID:AB_2081431
Mouse Monoclonal anti-B23	Santa Cruz	Cat# sc-56622, RRID:AB_784888
Mouse Monoclonal anti-V5	Invitrogen	Cat# R960-25 (previously Cat#46-0705), RRID:AB_2556564
Polyclonal Chicken anti-V5	Novus Biologicals	Cat# NB600-379, RRID:AB_10003214
Mouse Monoclonal anti-HA	Linda Hendershot's Laboratory	N/A
Rabbit Polyclonal anti-HA	Linda Hendershot's Laboratory	N/A
Rat Monoclonal anti-HA	Roche	Cat# 11867423001, RRID:AB_390918
Rabbit Monoclonal anti-Myc tag	Cell Signaling Technology	Cat# 2278S, RRID:AB_10693332
Mouse Monoclonal anti-polyUb (FK2)	Enzo Life Sciences	Cat# BML-PW8810, RRID:AB_10541840
Mouse Monoclonal anti-GFP	Cell Signaling Technology	Cat# 2955S, RRID:AB_1196614
Mouse Monoclonal anti-GFP (B-2)	Santa Cruz	Cat# sc-9996, RRID:AB_627695
Mouse Monoclonal anti-FLAG (M2)	Sigma	Cat# P2983, RRID:AB_439685
Rabbit Polyclonal anti-GAPDH	Abcam	Cat# ab9485, RRID:AB_307275
Rabbit Polyclonal anti-mCherry	Abcam	Cat# ab167453, RRID:AB_2571870
Mouse Monoclonal anti-His6	Invitrogen	Cat# MA1-21315, RRID:AB_557403
Bacterial and Virus Strains		
BL21 RIPL *E. coli*	Agilent	Cat#230280
BL21 Gold (DE3) *E. coli*	Agilent	Cat#230132
Rosetta (DE3) *E. coli*	EMD Millipore	Cat#71400
Chemicals, Peptides, and Recombinant Proteins		
Oregon Green 488 Carboxylic Acid, Succinimidyl Ester, 5 isomer	Thermo Fisher	Cat #O6147
Rhodamine Red C2 maleimide	Thermo Fisher	Cat #R6029
Alexa 647 maleimide	Thermo Fisher	Cat# A20347
Rhodamine Red-X Succinimidyl Ester, 5 isomer	Thermo Fisher	Cat# R6160
MG132 Proteasome inhibitor	EMD Millipore	Cat#474791
fPuc peptide: Ac-ENLACDEVTSTTSSSST-NH2	Genscript	Pierce et al., 2016
cDAXX peptides (see Table S2 for sequences)	This paper	N/A
Critical Commercial Assays		
Lipofectamine 3000 Transfection Reagent	Thermo Fisher	Cat# L3000015
Effectene Transfection Reagent	QIAGEN	Cat#301425
Experimental Models: Cell Lines		
HeLa	Yasmine Valentin-Vega, St Jude	
HEK293T	Kevin Freeman, St Jude	
PC-3	Haojie Huang, Mayo Clinic	
Oligonucleotides		
SPOP M233 to E mutagenesis FWD: CAGTCTTCA GTGCCATGTTTGAACATGAAGAGGAGGAGAGC		
SPOP M233 to E mutagenesis Rev: CTCAACCCGA TTCTTTTTGCTCTCCTCCTCTTCATGTT		

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SPOP D278, K279 to A278, A279 mutagenesis FWD: TGACGACTTGCTGGCAGCTGCCGCCGCATACGC		
SPOP D278, K279 to A278, A279 mutagenesis Rev: CCTTCAGACGCTCCAGGGCGTATGCGGCGGCA		
SPOP F133 to V mutagenesis FWD: GGCAAAGACTGGGGAGTGAAGAAATTCATCCG		
SPOP F133 to V mutagenesis Rev: CGGATGAATTTCTTCACTCCCCAGTCTTTGCC		
SPOP W131 to G mutagenesis FWD: GCAAGGCAAAGACGGCGGATTCAAGAAATTC		
SPOP W131 to G mutagenesis Rev: GAATTTCTTGAATCCGCCGTCTTTGCCTTGC		
Recombinant DNA		
pDONR221-DAXX	DNASU	HsCD00079589
pcDNA6.2-EmGFP-DAXX	This paper	N/A
pcDNA6.2-EmGFP-cDAXX	This paper	N/A
pcDNA6.2-EmGFP-cDAXX$^{\Delta SB}$	This paper	N/A
pDEST17-cDAXX	This paper	N/A
pDEST17-cDAXX$^{\Delta SB}$	This paper	N/A
pCDNA3-V5-SPOP	(Marzahn et al., 2016)	N/A
pCDNA3-V5-SPOP F133V	This paper	N/A
pCDNA3-V5-SPOP W131G	This paper	N/A
pCDNA3-V5-SPOP mutBACK	(Marzahn et al., 2016)	N/A
pCDNA3-V5-SPOP mutBTBBACK	(Marzahn et al., 2016)	N/A
pmCherry-N1-SPOP	(Marzahn et al., 2016)	N/A
pmCherry-N1-SPOP^{CBM}	This paper	N/A
pcDNA3-myc-CUL3	(Ohta et al., 1999)	Addgene #19893
pcDNA3-HA2-ROC1 (Rbx1)	(Ohta et al., 1999)	Addgene #19897
pcDNA3-Myc3-Cul3 H2/H5M	(Furukawa and Xiong, 2005)	Addgene #21591
pEGFP-C1-AR	(Stenoien et al., 1999)	Addgene #28235
pCDNA-His6-Ubiquitin	Wenyi Wei	N/A
pDEST17-nAR	This paper	N/A
pFastbac GST-UB E1	(Huang et al., 2008)	N/A
pGEX4T1 GST-Thrombin- UBCH7	(Huang et al., 2008)	N/A
pGEX4T1 GST-Thrombin-UBC12	(Duda et al., 2008)	N/A
pGEX4T1 APPBP1-UBA3	(Duda et al., 2008)	N/A
GST pGEX4T1 GST-ThrombinNEDD8	(Walden et al., 2003)	N/A
pGEX2TK GST- Thrombin-UB (C > S)	(Scott et al., 2014)	N/A
pGEX4T1 GST-TEV-ARIH1	(Scott et al., 2016)	N/A
pGEX4T1 GST-TEV-ARIH1 C375S	(Scott et al., 2016)	N/A
pET-DUET-1-His-Cul3/Rbx1	(Small et al., 2010)	N/A
Software and Algorithms		
PattinProt	PRABI-Lyon-Gerland Institute of Biology and Protein Chemistry	https://npsa-prabi.ibcp.fr/NPSA/npsa_pattinprot.html
RF cloning	Steve Bond	http://rf-cloning.org
NIS Elements	https://www.nikoninstruments.com/Products/Software	RRID:SCR_014329
FIJI	http://fiji.sc/	RRID:SCR_002285

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Zen software	http://www.zeiss.com/corporate/en_de/global/home.html	RRID:SCR_013672
SlideBook	https://www.intelligent-imaging.com/slidebook	RRID:SCR_014300
GraphPad Prism	https://www.graphpad.com/	RRID:SCR_002798
Protein Parameter Calculator	Anthis and Clore, 2013	http://nickanthis.com/tools/a205.html
MATLAB	https://www.mathworks.com/products/matlab/	RRID:SCR_001622
TopSpin	https://www.bruker.com/products/mr/nmr/nmr-software/nmr-software/topspin/overview.html	RRID:SCR_014227
CARA	http://cara.nmr-software.org/portal/	
plmDCA algorithm	Ekeberg et al., 2013	

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Tanja Mittag (Tanja.Mittag@stjude.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

HeLa and HEK293T cells were grown under sterile conditions on Dulbecco's Modified Eagle's Medium (DMEM). PC-3 cells were grown under sterile conditions on RPMI 1640 Medium. Culture media were supplemented with 10% fetal bovine serum (FBS), 2mM L-Glutamine and antibiotics. Cells were grown at 37°C with 5% CO_2.

METHOD DETAILS

Prediction of SB motifs

The position of SB motifs was predicted using a bioinformatics search with the online tool PattinProt (https://npsa-prabi.ibcp.fr/NPSA/npsa_pattinprot.html), using the query sequence [GAVLIMFWP]-[STCYNQ]-S-[ST]-[ST], and allowing for 1 mismatch to the consensus SB motif sequence (Φ-Π-S-S/T-S/T, where Φ is a nonpolar and Π is a polar residue; Zhuang et al., 2009). In order to create the cDAXX-0sb sequence, the SB motifs were scrambled in the following way to preserve the length and charge distribution of the cDAXX sequence: 1) The first position was mutated from a hydrophobic to polar residue. 2) The second position was mutated from a polar residue to a proline. 3) The third position was mutated to a charged residue (if one was in the original sequence). 4) The fourth or fifth position was mutated from S/T to a hydrophobic residue. The resulting mutated sequences slightly increased the Pro content from 11 to 12.6% and reduced the Ser content from 18.3 to 16.7% in cDAXX-0sb relative to cDAXX.

Plasmids

We purchased full-length DAXX cDNA in vector pDONR221 from DNASU (clone: HsCD00079589) and introduced it into pcDNA6.2-EmGFP by Gateway technology (Life Technologies) for expression in mammalian cells. We also excised the cDAXX (495-740) sequence by nested PCR and introduced it into pDEST17 (Thermo Fisher) by Gateway technology for bacterial expression of His-tagged protein and protein purification for NMR experiments. For all other *in vitro* experiments, we purchased synthetic, codon-optimized genes for cDAXX, and cDAXX^{0SB}, which included Cys to Ser mutations at positions 629, 664, 699, and 720 (Thermo Fisher), and introduced them into pCDNA6.2-EmGFP by Gateway technology for expression in mammalian cells and into pDEST17 for bacterial expression and protein purification. Site directed mutagenesis was used to add a TEV protease site to remove the His_6 tag from bacterially expressed cDAXX. Plasmids for V5-tagged SPOP-WT, mutBACK, and mutBTB-BACK were generated by switching the HA tag to V5 in pcDNA3-HA-SPOP (Marzahn et al., 2016) by site directed mutagenesis. We generated V5-SPOP F133V, and W131G by site directed mutagenesis of V5-SPOP-WT. The SPOP-mCherry plasmid was constructed by using our previously described HA-SPOP plasmid (Marzahn et al., 2016) as a template for PCR and SPOP was introduced into vector pmCherry-N1 by restriction digest followed by ligation. To generate the SPOP^{CBM}-mCherry mutant, we mutated M233 to E, D278 to A, and K279 to A by site directed mutagenesis using the restriction free cloning method (van den Ent and Löwe, 2006). Plasmids pcDNA3-myc-CUL3,

pcDNA3-HA2-ROC1 (Rbx1) (Ohta et al., 1999), pcDNA3-Myc3-Cul3 H2M/H5M (Furukawa and Xiong, 2005), and pEGFP-C1-AR (Stenoien et al., 1999) were obtained from Addgene. The plasmid for His_6-Ubiquitin was a kind gift from Wenyi Wei (Harvard). We purchased the nAR (AR 1-559 Uniprot P10275) gene for bacterial expression (Thermo Fisher) and introduced it into pDEST17 (Thermo Fisher) by Gateway technology (Life Technologies).

The following plasmids were previously published: pFastbac GST-UB E1, pGEX4T1 GST-Thrombin-UBCH7 (Huang et al., 2008), pGEX4T1 GST-Thrombin-UBC12, pGEX4T1 APPBP1-UBA3, GST pGEX4T1 GST-ThrombinNEDD8 (Duda et al., 2008; Walden et al., 2003), pGEX2TK GST-Thrombin-UB (C > S) (Scott et al., 2014), pGEX4T1 GST-TEV-ARIH1 (and C357S mutant) (Scott et al., 2016), pET-DUET-1-His-Cul3/Rbx1 (Small et al., 2010)

Transfections

Cells were transfected with Lipofectamine 3000 (Thermo Fisher) or with Effectene (QIAGEN) according to the manufacturer conditions.

Immunofluorescence

Cells were transfected in 8-well glass chambers (Millipore) and fixed with 4% paraformaldehyde 24 hours later. Cells were permeabilized with 0.1% Triton X-100 and blocked with 10% donkey serum. GFP and mCherry fluoresce was detected directly. The primary antibodies used were: SC-35 (1:300; Abcam, ab11826), PML (1:50; Santa Cruz, sc-966), coilin (1:500; Santa Cruz Biotechnology, sc-32860), B23 (1:200, Santa Cruz Biotechnology, sc-56622) Myc-tag (1:500 Cell Signaling Technologies, 71D10), HA-tag (1:250; Clone 3F10, Roche, 11867423001), FK2 (1:50; Enzo Life Sciences, BML-PW8810), V5-tag (1:300, Novus Biological, NB600-379). The secondary antibodies used were Alexa 555, 647 (1:5,000; Thermo-Fisher), and CF405S (1:1000; Biotium). Samples were mounted on ProLong Gold antifade with or without DAPI and cured before imaging on a Zeiss LSM 780 NLO microscope. Images were prepared with the Fiji software.

Live Cell Imaging

Cells were transfected in 4-well borosilicate chambers (LAB-TEK). Twenty-four hours after transfection, media was changed to phenol red-less DMEM and imaged in a Marianas spinning disk confocal microscope at 37 ^{o} C in the presence of CO_2 for the duration of the experiment. Analysis, image and video preparation was done with the Slidebook software.

Immunoprecipitation and Western Blots

HEK293T cells were transfected with plasmids expressing the indicated proteins. 24 h post-transfection, cells were incubated with MG132 at a final concentration of 20 μM for 4 h. For immunoprecipitation of the substrates, cells were lysed 24 h after transfection in Nonidet P-40 lysis buffer (50 mM Tris/HCl, pH 7.5, 150 mM NaCl, 0.5% Nonidet P-40, 50mM NaF) supplemented with 10 mM N-Ethylmaleimide (NEM) and protease inhibitors (Roche Applied Science). Immunoprecipitation was performed on clarified cell lysates with GFP antibody (Santa Cruz sc-9996) overnight at 4°C and the resulting proteins were analyzed by immunoblotting with anti-GFP and anti-His_6 antibody. Input materials were also checked by immunoblotting using anti-GFP, anti-V5, anti-Myc, anti-HA, and anti-GRP170 antibodies (loading control). Immune complexes were isolated with protein A-agarose beads, washed with NP-40 buffer in 2 × reducing Laemmli buffer. Whole cell lysates were mixed with 4 × reducing Laemmli buffer and analyzed by SDS-PAGE and followed by immunoblotting with the indicated antibodies. Cells were lysed, GFP-cDAXX was pulled down

For pull-down of ubiquitinated proteins, cells were incubated with MG132 or DMSO at 20 μM for 4 hours cells were lysed 24h after transfection in buffer A [6 M guanidine-HCl, 0.1 M Na_2HPO_4/NaH_2PO_4 (pH 8.0) and 10 mM imidazole]. The lysates were sonicated, cleared, and incubated with Ni-NTA Sepharose (QIAGEN) for 3 hr at room temperature. The beads were washed twice with buffer A, twice with A/T composed of one volume of buffer A and three volumes of buffer T [25 mM Tris (pH 8.0) and 20 mM imidazole], and twice with buffer T. The beads were finally boiled in SDS-PAGE loading buffer containing 100 mM imidazole.

The antibodies used for western blots were: GFP (1:10,000; Cell Signaling Technologies, 4B10), V5 (1:5,000; Thermo-Fisher, R960), Myc-tag (1:100 Cell Signaling Technologies, 71D10), FLAG-tag (1:1,000; Sigma, M2), GAPDH (1:5,000; Abcam ab9485). HA-tag (1:500), and GRP170 (1:1,000) were a gift from Linda Hendershot. Anti-mouse and anti-rabbit HRP conjugated antibodies (Jackson immunoresearch) were used at 1:20,000

Protein Purification

All His-SUMO-$SPOP^{28–359}$ constructs were expressed in BL21 RIPL cells in ZYM-5052 auto-induction media (Studier, 2005), and purified as previously described by Ni column, TEV cleavage during dialysis, Ni-pass-back, and SEC (Marzahn et al., 2016). GST-$SPOP^{MATH}$ was expressed in BL21 RIPL cells in LB medium, and purified as previously described by Glutathione-Sepharose, TEV cleavage during dialysis, ion-exchange by SP column, and SEC (Pierce et al., 2016). Small fractions of each construct were labeled with 4X Oregon Green 488 Carboxylic Acid, Succinimidyl Ester, 5 isomer (Cat #O6147 Thermo Fisher) at 4°C overnight. Un-reacted label was removed from the proteins by PD-10 column (Thermo Fisher 45-000-67).

All cDAXX constructs were expressed in BL21 RIPL cells in LB media at 18°C for ~20 hr. Protein expressed for NMR experiments were isotopically labeled with ^{15}N, ^{13}C by growing cells in M9 minimal media supplemented with ^{13}C glucose and ^{15}N ammonium chloride (Cambridge Isotopes). at 37°C until an OD_{600} = 0.8. Expression was then induced with 0.6 mM IPTG for 18 hours at

20°C. Cells were lysed in 50 mM Tris pH 8.0, 500 mM NaCl, 30 mM Imidazole, 2 mM β-ME, and Complete protease inhibitor cocktail (Roche) with a microfluidizer at 20,000 psi. The cleared lysate was loaded onto a gravity Ni-NTA column, washed with lysis buffer, and eluted in 50 mM Tris pH 8.0, 50 mM NaCl, 300 mM Imidazole, and 2 mM β-ME. When indicated, the tag was cleaved with 1mg TEV protease/100 mg protein at 4°C overnight. Both cleaved and uncleaved solutions were diluted 2-4 fold, passed over a Q column, and eluted by salt gradient. Peak fractions were then concentrated (Amicon 10 MWCO) and subjected to SEC on a Superdex 200 16/60 column (GE Healthcare) equilibrated with 25 mM Tris pH 7.5, 150 mM NaCl, and 5 mM DTT (or 1 mM T-CEP). Small fractions of each cDAXX construct were labeled with 10X Rhodamine Red C2 maleimide (cat #R6029 from Thermo Fisher) at 4°C overnight. Un-reacted label was removed from the proteins by PD-10 column.

UBA1 was expressed in insect cells as a GST-fusion protein, and purified by glutathione affinity, thrombin cleavage, and ion-exchange, as previously described (Huang et al., 2008; Scott et al., 2016). UBCH7 was expressed in *E. coli* BL21 Gold (DE3) cells as a GST-fusion protein, and purified by glutathione affinity, thrombin cleavage, ion-exchange, and SEC chromatography, as previously described (Huang et al., 2008; Scott et al., 2016). NEDD8 and UB were expressed in *E. coli* BL21 Gold (DE3) cells as GST-fusion proteins, and purified by glutathione affinity, thrombin cleavage, glutathione pass-back, and SEC chromatography, as previously described (Scott et al., 2014; Walden et al., 2003). UB was fluorescently labeled with 4X Alexa 647 maleimide (cat # from A20347 Thermo Fisher) for 2 hr at RT, quenched with 10 mM DTT, and passed over a PD-10 column to remove excess unreacted label. Cul3/Rbx1 were co-expressed in *E. coli* BL21 Gold (DE3) and purified by Ni, and SEC chromatography, as previously described (Small et al., 2010). Neddylation of 12 μM Cul3/Rbx1 was accomplished by incubating with 1 μM UBC12, 0.1 μM APPBP1-UBA3, and 20 μM NEDD8 with ATP and $MgCl_2$ as previously described (Duda et al., 2008). A small fraction of N8∼Cul3/Rbx1 was fluorescently labeled with 4X Alexa 647 Acid, NHS (Succinimidyl) Ester (cat #A2006 from Thermo Fisher) at RT for ∼1.5 hr, and excess unreacted label was removed by PD-10 column. ARIH1 and ARIH1^{C375S} were expressed in *E. coli* BL21 Gold (DE3) cells as GST-fusion proteins, and purified by glutathione affinity, TEV cleavage, ion-exchange, and SEC chromatography as previously described (Scott et al., 2016).

His-nAR was expressed in *E. coli* Rosetta (DE3) cells and purified from inclusion bodies. Cells were lysed by sonication. The insoluble fraction was washed with PBS buffer pH = 7.4 containing 1% Triton, and dissolved in 8 M Urea, 20 mM Tris pH 7.8, 500 mM NaCl, and 14 mM β-ME. The solution was then passed over a Ni-HP or gravity column and eluted with the same buffer with 500 mM Imidazole. The tag was cleaved with 1mg TEV protease/100 mg protein at 4°C overnight under dialysis, and removed by passing back over Ni resin in 8 M Urea, 50 mM Tris pH 7.8, 100 mM NaCl, and 1 mM β-ME. The flow-through was then concentrated (Amicon 3 MWCO) and subjected to SEC on a Superdex 200 16/60 column (GE Healthcare) equilibrated with 20 mM Phosphate buffer pH 7.4, 2 mM T-CEP, and 1 mM EDTA. Due to protein stability, protein was immediately used, and microscope samples were imaged within 1 hour of set-up. A small fraction of the protein stock labeled with 10X Rhodamine Red-X Succinimidyl Ester, 5 isomer (cat # R6160 from Thermo Fisher) for 1 hr at 4°C. Un-reacted label was removed from the proteins by PD-10 column.

Microscopy analysis for *in vitro* LLPS

Samples were prepared by mixing the determined amount of protein, buffer, and ficoll PM 70 (Sigma). Sealed sample chambers containing protein solutions comprised coverslips sandwiching two layers of 3M 300 LSE high-temperature double-sided tape (0.34 mm). For each given cDAXX and/or SPOP concentration, the sample was equilibrated at room temperature and incubated for 4-6 hours. Samples were imaged on a Nikon C2 laser scanning confocal microscope with a 20X (0.8NA) Plan Apo objective. Images and movies were processed with the Nikon NIS Elements software. All images within figures were taken with the same camera settings, unless otherwise noted.

In vitro crosslinking reactions

Samples were prepared at 15 μM SPOP and the indicated concentrations of H-cDAXX in buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT. The amine crosslinker BS^3 (bis(sulfosuccinimidyl)suberate, cat #21580 from Fisher Scientific) was added for a final concentration of 0.15 mM in samples from Figure S2C and 0.3 mM in samples from Figure S4E. Reactions were incubated at room temperature for 30 min. The reactions were quenched by the addition of 100 mM Tris pH 7.6 and were incubated at room temperature for at least 15 min prior to loading samples onto SDS-PAGE gel.

NMR data collection and analysis

NMR data were acquired on Bruker Avance 600 and 800 MHz spectrometers equipped with TCI triple-resonance cryogenic probes and pulsed-field gradient units at 5°C. All samples were prepared in an NMR buffer consisting of phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4), 10 mM DTT pH 6.9 and 10% D_2O.

Assignment of cDAXX backbone resonances were done in two steps. Initially, approximately 0.75 mM ^{15}N,^{13}C cDAXX samples were prepared and standard 3D assignment experiments based on sensitivity enhanced ^{1}H,^{15}N HSQC (8 scans, 2048 × 320 complex data points) were collected. These included a HNCACB and CBCA(CO)NH (16 scans, 1024 (^{1}H) × 32 (^{15}N) × 128 (^{13}C) complex data points, with 11 ppm, 24 ppm, and 70 ppm as ^{1}H, ^{15}N and ^{13}C sweep width, respectively), a HN(CA)CO (16 scans, 1024 (^{1}H) × 32 (^{15}N) × 75 (^{13}C) complex data points, with 11 ppm, 24 ppm, and 18 ppm as ^{1}H, ^{15}N and ^{13}C sweep widths), a HNCO (16 scans, 1024 (^{1}H) × 32 (^{15}N) × 75 (^{13}C) complex data points, with 11 ppm, 32 ppm, and 22 ppm as ^{1}H, ^{15}N and ^{13}C sweep widths), and a HNCA (16 scans, 1024 (^{1}H) × 32 (^{15}N) × 95 (^{13}C) complex data points, with 16 ppm, 25 ppm, and 30 ppm as ^{1}H, ^{15}N and ^{13}C sweep

widths). Additionally, HN(CA)NNH (16 scans, 1024 (^{1}H) × 32 (^{15}N F1) × 60 (^{15}N F2) complex data points, with 11 ppm, 24 ppm, and 24 ppm as ^{1}H, ^{15}N F1 and ^{15}N F2 sweep widths) provided connectivity between N, N+1 and N-1 facilitating a "backbone walk"(Weisemann et al., 1993). Using these methods, approximately 70% of the sequence was confidently assigned.

Gaps in assignments were filled using carbon-detect experiments. Initial HSQC based resonance assignments were transferred to a ^{13}C,^{15}N 2D carbon-detect CON spectrum using IPAP decoupling (32 scans, 1024 (^{13}C) × 256 (^{15}N) complex data points with 18 ppm and 35 ppm as ^{13}C and ^{15}N sweep widths). Carbon-detect assignments were performed using 3D CANCO (16 scans, 1024 (^{13}C) × 32 (^{15}N) × 48 (^{13}C) complex data points, with 20 ppm, 35 ppm, and 35 ppm as ^{13}C F3, ^{15}N and ^{13}C F1 sweep widths), COCON (8 scans, 1024 (^{13}C) × 40 (^{15}N) × 32 (^{13}C) complex data points, with 18 ppm, 35 ppm, and 12 ppm as ^{13}C F3, ^{15}N and ^{13}C F1 sweep widths) and CCCON (8 scans, 1024 (^{13}C) × 42 (^{15}N) × 40 (^{13}C) complex data points, with 12 ppm, 35 ppm, and 72 ppm as ^{13}C F3, ^{15}N and ^{13}C F1 sweep widths) spectra (Bermel et al., 2005, 2006a, 2006b). Overlapping peaks were resolved by using a combination of nitrogen-detect 2D NCO experiments and carbon-detect amino acid specific 2D experiments (Sahu et al., 2014; Takeuchi et al., 2010).

Titration of the SPOP^{MATH} domain into cDAXX was performed using samples containing 0.4 mM ^{15}N, ^{13}C cDAXX and 0.2 and 0.4 mM SPOP^{MATH} for 1:0.5 and 1:1 molar ratios, respectively, and 0.2 mM cDAXX and 0.4 mM SPOP^{MATH} for the 1:2 molar ratio. Carbon detect CON-IPAP spectra were recorded with identical parameters to assignment experiments however the number of scans were increased to 80 and 128 for 0.4 mM and 0.2 mM cDAXX samples respectively.

Data were processed using BRUKER Topspin version 3.4, NMRPipe (v.7.9) (Delaglio et al., 1995) and analyzed using CARA (v.1.8.4) (Keller, 2004). All spectra were referenced directly using DSS for the ^{1}H dimension; ^{13}C and ^{15}N frequencies were referenced indirectly.

Peptide Synthesis and Preparation

Peptides encompassing the SB motifs of cDAXX and the mutated motifs of cDAXX-0sb were synthesized at the Hartwell Center at St. Jude Children's Research Hospital. Each peptide contained the 5 residue SB/mut motif and 4 residues on either side, with N-terminal acetyl and C-terminal carboxy modifications. Each peptide was solubilized in water, pH-corrected, lyophilized, and re-solublized in minimal amounts of water. Stock concentrations were calculated from A_{205} readings using the extinction coefficient predicted by the online "Protein Calculator" (Anthis and Clore, 2013). The fPuc peptide, comprising residues 91–107 from the protein Puckered [amino acid sequence Ac-ENLACDEVTSTTSSSST-NH2 (Pierce et al., 2016; Zhuang et al., 2009)] and N-terminally labeled with fluorescein was purchased from GenScript, and solubilized in buffer containing 20 mM Tris pH 7.6, 150 mM NaCl, and 5 mM DTT.

Fluorescence Anisotropy

All DAXX assays were performed in 20 mM Tris pH 7.6, 150 mM NaCl, 5 mM DTT, 0.01% Triton X-100, and 10 mg/mL BSA. nAR FP assays were performed in the same buffer minus BSA. For direct binding measurements, serial dilutions of each SPOP construct, ranging from from 0.006 to 100 μM, were prepared on a 384-well plate (Greiner BioOne). Then fluorescently tagged- cDAXX construct, nAR, or fPuc peptide was added for a final concentration of 40 nM into each well. For competition binding measurements, serial dilutions of each peptide were prepared in 384-well plates ranging from ~10 mM to ~0.5 μM. MATH domain and fPuc were added to final concentrations of 6 μM and 40 nM, respectively. Anisotropy was measured using a CLARIOstar plate reader (BMG LABTECH).

FRAP measurements

FRAP experiments were performed using a Marianas spinning disk confocal (SDC) imaging system on a Zeiss Axio Observer inverted microscope platform using a Zeiss Plan-Apochromat 63X (1.4 NA) oil objective and Evolve 512 EMCCD camera (Photometrics). For *in vitro* samples, time-lapse images were collected with 200 ms exposure time for 4 to 6 min, every 300 ms for the first min, then every 5 s for the remainder of the time-lapse. Photobleach settings were: 1 ms, 1 repetition, 20%–50% of 488 channel intensity or 35%–80% of 561 channel intensity. For in cell FRAP, time-lapse images were collected every 500 ms for 3.5 to 4 min. Photobleach settings were: 1 ms, 1 repetition, 60% of 488 or 561 channels. Images were analyzed with SlideBook 6 software (3i).

In vitro ubiquitination assays

Singe-turnover assays were conducted to monitor the paths of UB transfer. E2~*UB was prepared by mixing 15 μM UBCH7, 0.6 μM UB E1, and 20 μM *UB in 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM ATP, 10 mM $MgCl_2$ at room temperature for 15 minutes. The reaction was quenched by the addition of EDTA to a final concentration of 50 mM. The single-turnover reactions consisted of mixing the E2~*UB thioester conjugate (1.5 μM final concentration) with pre-incubated complexes (30 min – 2 hr at RT) of the indicated combinations of 20 nM ARIH1 or indicated mutants, 1.25 μM N8~Cul3/RBX1, 5 or 15 μM SPOP and 20 or 50 μM H-cDAXX at room temperature in 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT, in the presence of 4 or 10% w/v ficoll-70 or sucrose. Samples were visualized as for LLPS analysis or loaded on SDS-PAGE gels and analyzed in a Typhon FLA scanner (GE Healthcare).

Co-evolutionary analysis

We obtained a multiple sequence alignment containing SPOP homologs by first building a hidden Markov model of the protein family, based on 4 iterations of jackhmmer (Finn et al., 2015), and extracting the sequences from the Uniprot Uniref100 database (Suzek

et al., 2015). We refined the alignment by requiring that all sequences contain all three structural domains (MATH, BTB and BACK) and excluding all positions that contain more than 50% of gaps. We then used the asymmetric plmDCA algorithm (Ekeberg et al., 2013), using default input parameters (including a 90% cutoff in sequence similarity resulting in 2603 sequences), to find pairs of residues with direct correlated mutations along evolution. We used the derived couplings to divide the SPOP sequence into Evolutionary Domains (Granata et al., 2017), i.e., to find groups of residues that evolved together and almost-independently from each other. In this analysis with used the webserver at spectrus.sissa.it/spectrus-evo_webserver with default parameters.

QUANTIFICATION AND STATISTICAL ANALYSIS

Partition Coefficient for cDAXX and cDAXX^{0SB}

Signal intensities were obtained with the Fiji software. The cellular average intensity for GFP-cDAXX and GFP-cDAXX^{0SB} in nuclear condensates and in the diffuse nuclear fractions were calculated by determining the signal intensities in 3 regions of interest (ROIs) per cell for condensates and for the diffuse signal. Background was subtracted by using an ROI of the same size in the area outside of the cell. Statistical significance was determined with the paired Student's t test. Signal intensities were plotted using the GraphPad Prism software.

GFP-DAXX and V5-SPOP Signal Intensity

Signal intensities for V5 (SPOP) and GFP-DAXX in cells expressing WT or F133V V5-SPOP were obtained with the Fiji software. The nuclear signal intensity for GFP-DAXX and V5-SPOP was measured by using an ROI the size of the whole nucleus to determine the signal intensity for each cell. Background was subtracted by using an ROI of the same size in the area outside of the cell. Statistical significance was determined with the paired Student's t test. Signal intensities were plotted using the GraphPad Prism software.

K_D Determination from Anisotropy Measurements

K_D values were obtained by fitting to the following equations, adapted from Roehrl et al.,

$$A_{ods} = QF_bA_b + \frac{A_f(1-F_b)}{1-F_b(1-Q)}$$

where A_{obs} is the observed anisotropy, A_b is the anisotropy of the bound state, A_f is the anisotropy of the free state, Q is the ratio of $Intensity_{free}/Intensity_{bound}$, and F_b is the fraction bound of the fluorescent species, which is given by the following equation for direct binding assays,

$$F_b = \frac{(K_D + SPOP + substrate) - \sqrt{(K_D + SPOP + substrate)^2 + 4(SPOP)(substrate)}}{2(substrate)},$$

where *substrate* is the total concentration of fluorescently labeled cDAXX construct or fPuc peptide, *SPOP* is the total concentration of SPOP (Roehrl et al., 2004).

For competitive binding assays, F_b is given by the following equations adapted from Roehrl et al.,

$$F_b = \frac{2\sqrt{(a^2-3b)}\cos(\theta/3) - a}{3K_{D,Puc} + 2\sqrt{(a^2-3b)}\cos(\theta/3) - a}$$

where

$$a = K_{D,Puc} + K_D + Puc + peptide - MATH,$$

$$b = (Puc - MATH)K_{D,Puc} + (peptide - MATH)K_D + K_{D,Puc}K_D,$$

$$c = -K_{D,Puc}K_DMATH,$$

$$\theta = \cos^{-1}\left[\frac{2a^3 + 9ab - 27c}{2\sqrt{(a^2-3b)^3}}\right],$$

and $K_{D,Puc}$ is the dissociation constant determined from a direct binding assay of fPuc and MATH, *Puc* is the total concentration of fluorescently labeled fPuc peptide, *peptide* is the variable concentration of each SB/mut peptide, and *MATH* is the total

concentration of MATH domain. For each FA assay, three independent experiments were performed and fit; the average K_D and standard deviation are reported in Tables S2, S3, and S5.

Quantification of *in vitro* assemblies

Signal intensities of proteins in assemblies were measured in NIS Elements from images without signal saturation. For quantification of protein concentrations, overlapping intensity-based threshold layers in each channel were applied to select the assemblies and measure the mean intensities within them. A standard curve of the intensity versus fluorescent protein was then used with the ratio of labeled to unlabeled protein in each sample to calculate the concentration of each protein construct in the assemblies.

FRAP Analysis and Fitting

For FRAP analysis, mean fluorescence intensities from ROIs were background-corrected, and corrected for photobleaching due to imaging. Fluorescence intensity versus time graphs were expressed in fractional form normalized by the pre-photobleach intensity (Axelrod et al., 1976) and fitted to equations for single- or double-exponential recovery. See also Table S4.

Quantification of *in vitro* ubiquitination

Quantification of *UB appearance in assemblies from microscope images was conducted in NIS Elements, overlapping intensity-based threshold layers in the green and red channels were applied to select assemblies. The average intensity in each channel within the assemblies was then measured at each time point, as well as the average intensity of the background (from a location in the image with no assemblies). The intensity within assemblies was determined by plotting the background-corrected blue intensity divided by the background-corrected average of green and red intensities, for each time point. Due to the variability in signal intensity, reactions with increases in the blue channel within assemblies over time were normalized to the intensity of the last point in the WT SPOP + ficoll conditions.

4.4 Hormone binding causes the condensation of androgen receptor that mediates folding of activation function motifs - *manuscript in preparation*

PUBLICATIONS

Hormone binding causes the condensation of androgen receptor that mediates folding of activation function motifs

Elzbieta Szulc[1,2,#], Paula Martínez-Cristóbal[1,2,#], Mireia Pessarodona[1,2], Marta Frigolé-Vivas[1,2], Diego M. Presman[3,4] , Grégory Fettweis[3], Karina Pombo-García[5], Hernan Morales-Navarrete[5], Cristina Batlle[6], Jill J. Bouchard[7], Tanja Mittag[7], Isabelle Brun-Heath[1], Jesús García[1], Alf Honingmann[5], Gordon L. Hager[3],' Salvador Ventura[6], Xavier Salvatella[1,2,8,*].

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain.
[2]Joint BSC-IRB Research Programme in Computational Biology, Baldiri Reixac 10, 08028 Barcelona, Spain.
[3] Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, NIH, Bethesda, MD 20892
[4]UBA-CONICET, Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Argentina
[5]Max Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany
[6]Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autónoma de Barcelona, Bellaterra, 08193, Spain
[7]Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105-3678, United States
[8]CREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain

#Equally contributed

*Correspondence to xavier.salvatella@irbbarcelona.org

The activation of transcription relies on physical interactions between the activation domains of transcription factors and the transcription machinery. Activation domains are intrinsically disordered and the mechanisms by which they gain affinity for transcriptional co-regulators, general transcription factors and RNA Pol II are not known. We have studied how the structural properties of the activation domain of the androgen receptor, a transcription factor that regulates the male phenotype, are altered upon its activation by androgens. Our results indicate that activation triggers oligomerization and condensation of the receptor that causes gains in secondary structure in activation function motifs that mediate the recruitment of members of the transcription complex.

Introduction

Transcription factors (TFs) are responsible for determining the specificity of transcriptional activity (Ptashne 1988). They perform this function by binding to promoters and enhancers through their DNA binding domains (DBDs) and recruiting the transcription complex through their activation domains (ADs) (Keegan, Gill, and Ptashne 1986). The specificity of DNA binding is well understood thanks to the availability of structures for the complexes that DBDs form with transcription factor-specific DNA sequences (Fulton et al. 2009). The specificity of recruitment of the transcription complex and the necessary co-activators is instead not well understood, rendering our knowledge of this key biological process incomplete. One reason for this is that ADs are generally intrinsically disordered (ID) (Minezaki et al. 2006) and form transient protein-protein interactions (Fuxreiter et al. 2008). This makes it challenging to determine the structures of the relevant complexes (Radhakrishnan et al. 1997; De Mol et al. 2018), understand the determinants of recruitment specificity and, despite their potential as drug targets, target TFs with small molecules (Koehler 2010; Metallo 2010).

Recently it has become clear that many proteins involved in transcription, including RNA Polymerase II (RNA Pol II), the MED1 subunit of the Mediator complex and some TFs, can form transcriptionally active condensates (Cisse et al. 2013; B. R. Sabari et al. 2018; Cho et al. 2016; Boija et al. 2018; Boehning et al. 2018; Wollman et al. 2017; Nair et al. 2019) formed by liquid-liquid phase separation (LLPS) (Hyman, Weber, and Jülicher 2014). In cells LLPS of proteins and nucleic acids allows for membraneless compartmentalization in a condensed phase that resembles liquid droplets in coexistence with a dilute phase composed of a solution of the same macromolecules (Alberti, Gladfelter, and Mittag 2019). The proposed functions of TF condensation via LLPS in transcription include: facilitating contacts of multiple enhancers and promoters within super-enhancers through the dense phase (Hnisz et al. 2017; B. R. Sabari et al. 2018), providing a mode of Pol II and co-activator recruitment by TFs to the transcription site (Boija et al. 2018), enabling a rate-limiting step of transcription complex assembly (Cisse et al. 2013) or cooperative assembly of a chromosomal enhancer (Nair et al. 2019) and increasing the rate (Sokolova et al. 2013) and/or efficiency of transcription (Cho et al. 2016).

The observation that the transcription complex can function in a condensed state adds another layer of complexity in elucidating the molecular details of the protein-protein interactions taking place in it. The general features of the interactions stabilizing condensates are multivalency and lack of specificity (Hyman, Weber, and Jülicher 2014). By contrast, many interactions within the transcription complex are necessarily specific and rely on disorder-to-order transition of particular linear motifs within the ID AD of the TFs (Lee et al. 2000; De Mol et al. 2018). Additionally, the specificity of the interactions of TFs with their partners could in part account for the specificity of gene regulation (Eeckhoute, Métivier, and Salbert 2009). This apparent contradiction makes us formulate the following questions, focused on the TF ADs for simplification:

1) Are the TFs scaffolds or clients of the phase separated transcription complex? In other words, are the ADs drivers of the transcription complex condensation or do they instead partition into the condensed state?

2) Are the sequence motifs in the TF ADs responsible for the non-specific multivalent interactions allowing their condensation the same sequence motifs that are responsible for the specific interactions formed by the ADs with their globular partners (commonly known as activation functions)?

3) Is the LLPS state advantageous for the AD-specific partner interaction by promoting folding of the activation function motifs and consequently allowing for conformational selection for the specific binding partner?

Here, we address these questions by studying the phase separation properties of the Androgen Receptor (AR), a TF from a steroid receptor (SR) family, that regulates the development of the male phenotype and is activated by androgens such as testosterone (Gelmann 2002). In addition to activation and DNA binding domains, SRs harbor a C-terminal ligand binding domain (LBD) that is globular. In most SRs the LBD plays an important role in activation as, after hormone binding, it displays a surface patch called activation function 2 (AF-2) that has affinity for intrinsically disordered LxxLL motifs found in transcriptional co-activators (Huang, Chandra, and Rastinejad 2010) such as TIF2 (Voegel et al. 1996) and SRC-1a (Needham et al. 2000). The functions of these co-activators, that are key for transcriptional activity, include interacting with general transcription factors, RNA Pol II and chromatin remodeling. By contrast, in the case of AR, AF-2 has a higher affinity for a ^{23}FQNLF^{27} motif, located in the AD of AR so that activation of the protein by androgen binding results in the N/C interaction, that can occur in intra or inter-molecular fashion. As a consequence, the unavailability of the AF-2 surface shifts the main activation function of AR to the activation function 1 (AF-1) located in its AD that is in a contrary to AF-2 ID (He et al. 2004).

AR has been shown to be dimeric by many fluorescence methods directly aimed at addressing its dimerization state (Xu et al. 2015; M. E. van Royen et al. 2012). However, a more unbiased approach, the number and brightness (N&B) method, was recently used to show a higher order structure of SRs, including AR, questioning the notion that homodimers are their functional state (Presman et al. 2016). Additionally, we recently showed that AR can undergo condensation via liquid-liquid demixing with the Speckle-type POZ protein in the nuclei of HEK293T cells (Bouchard et al. 2018). It is therefore possible that AR can adopt a continuum of oligomeric functional states, ranging from dimers through oligomers to condensates. These AR condensates could manifest themselves as the previously described speckled-type distribution pattern observed in cell nuclei (Saitoh et al. 2002; Martin E. van Royen et al. 2013; Arihiro Tomura et al. 2001; Klokk et al. 2007). One of the factors that would contribute to this phenomenon is the clustering of androgen receptor elements (AREs) in enhancer regions and androgen-responsive genes (Bolton et al. 2007) that would allow for a local increase in concentration of the protein leading to its oligomerization or condensation, depending on the effective concentration.

In our hands, the ID AD domain of AR was sufficient to partially recapitulate the process of AR phase separation *in vitro* (Bouchard et al. 2018)**.** Here, we exploit this system to look into the molecular interactions that drive its phase separation and the structural features arising from it. Our results in cells show that androgen binding causes AR to form highly dynamic condensates produced by LLPS. Our experiments *in vitro* show that condensation causes folding into α-helices of otherwise ID activation function motifs that are key for transcriptional activity. This *folding upon condensation* mechanism can facilitate recruitment of the transcription machinery.

Results

In cells AR forms oligomers and condensates stabilized by the N/C interaction

AR is kept soluble in the cytoplasm by forming a complex with molecular chaperones such as Hsp40 and Hsp70 (Pratt and Toft 1997; Kim et al. 2013; Eftekharzadeh et al. 2019) until it is activated by androgen (testosterone or more potent dihydrotestosterone) binding. Ligand-bound protein undergoes conformational changes that include N/C interaction and oligomerization. Activated AR translocates to the nucleus where it binds to the AREs and performs its transcriptional regulatory functions (Fig. 1A).

In order to investigate the oligomerization state and condensation propensity of the AR in cells, we expressed EGFP-tagged full length AR (EGFP-AR) (Stenoien et al. 1999) in different cell lines. We observed that overexpressed EGFP-AR can form droplet-like condensates in the cytosol immediately upon activation by androgens (Fig. 1B, Fig. S1 and Movies M1, M2, M3). AR condensation could be observed in around 12% of PC3 cells. A comparison of EGFP-AR expression levels in the cytosol prior to activation, by mean fluorescence intensity of the cells, and the incidence of droplets formation showed no correlation, suggesting a contribution of unknown factors, beyond the critical concentration of the protein, to the formation of condensates (Fig. S2). The condensate droplets have liquid-like properties, are spherical, fuse (Fig. 1C) and are highly dynamic as shown by fluorescence recovery after photobleaching (FRAP) in Fig. 1D, with T-half 3,69 ± 2,23 s and mobile fraction of 0,85 ± 0,14, suggesting they are formed by LLPS. After 30 min to 1 h after activation, AR translocates to the nucleus (Fig. 1B and Fig. S1 and Movies M1, M2, M3), in agreement with previous findings (R. K. Tyagi et al. 2000).

Nuclear, overexpressed AR is known to form nuclear compartments (R. K. Tyagi et al. 2000; A. Tomura et al. 2001; Saitoh et al. 2002) although the nature of these protein clusters is not clear. In order to look more closely at the nuclear oligomeric state of AR we performed a number and brightness (N&B) analysis to investigate the relevance of the various domains of AR and the influence of DNA binding on AR oligomerization (Fig. 1E) and stimulated emission depletion (STED) microscopy to study the ability of the protein to form nuclear clusters (Fig. 1G). The oligomerization state of AR was measured in 3617 mouse cells with an integrated MMTV array (Presman et al. 2016) and the three constructs tested were: the full length EGFP-AR, shown to form cytoplasmic phase separated droplets in the 3617 mouse cells (Fig. S1A), the EGFP-V7 (AR variant 7), a nuclear splice variant of the AR containing the AD and DBD associated with an incurable late stage of prostate known as castration resistant prostate cancer (Nagabhushan et al. 1996) and the EGFP-AD (Fig 1E). The EGFP-DBD-LBD construct was also tested, but appeared to be highly insoluble. As reported previously (Presman et al. 2016), full length EGFP-AR formed high oligomeric species in the cell nucleoplasm and upon DNA binding, with a median of heptamers and hexamers respectively (Fig. 1E) and showed speckled sub-nuclear distribution (Fig. 1F). EGFP-V7 existed in much lower oligomeric states than the full-length protein: as a mixture of monomers and dimers in the nucleoplasm and as a mixture of dimers and some higher oligomeric species upon binding to DNA (Fig. 1E); its sub-nuclear distribution was more diffused than that of the full length EGFP-AR (Fig. 1F). EGFP-AD, similarly to EGFP-V7, was in a mixture of monomers and dimers (Fig. 1E) and its distribution was either diffused in the nucleus Fig. 1F) or the protein formed non-spherical highly immobile aggregates in the cytoplasm (Fig. S3). We concluded that only the protein possessing both the AD and the LBD has a high propensity to form large soluble oligomers, suggesting that the LBD dimerization

and/or the N/C interaction play a role in AR oligomerization. To investigate whether the former plays a significant contribution to the AR oligomerization, we studied the LBD dimerization-deficient mutant P767A (Nadal et al. 2017) and observed that its oligomerization state is comparable to the WT protein, indicating that LBD dimerization does not appear to be a major contribution to the high-order oligomerization in the nucleus (Fig. S4). Finally, to investigate whether the interaction between the AD and the LBD by the most potent N/C interaction AD motif - ^{23}FQNLF^{27} - is responsible for the formation of high-order nuclear oligomers, we mutated the motif to ^{23}AQNAA^{27} sequence (Schaufele et al. 2005; He et al. 2004). Again, we observed that it forms oligomers larger than dimers although the median size of oligomers was reduced compared to the wild type protein suggesting a contribution of the ^{23}FQNLF^{27}-mediated N/C interaction to high-order oligomerization of AR. It is possible, in agreement with previous findings, that the N/C interaction can be additionally mediated by motifs other than ^{23}FQNLF^{27}, such as ^{433}WHTLF^{437} and ^{179}LKDIL^{183} (He, Kemppainen, and Wilson 2000; Alen et al. 1999) (Fig. S4).

To complement these results and further investigate the ability of the protein to form chromatin-associated clusters we used STED microscopy on stably integrated AR-EGFP. We observed that the activated protein formed clusters of 105-125 nm diameter in the cells nuclei (Fig. 1G). However, no condensates were observed in the cytoplasm of these cells, likely due to the lower expression of the construct (Fig. S5): the need for a critical concentration for the cytoplasmic condensates formation agrees with the notion that the cytoplasmic droplets are formed by a process of LLPS. Collectively our observations confirm that that the cytosolic AR *aggregates* (Kumar and Tyagi 2012) and likely the nuclear AR *compartments* (Saitoh et al. 2002; Martin E. van Royen et al. 2013; Arihiro Tomura et al. 2001; Klokk et al. 2007; Rakesh K. Tyagi et al. 2000) described in the literature are therefore AR condensates formed by LLPS.

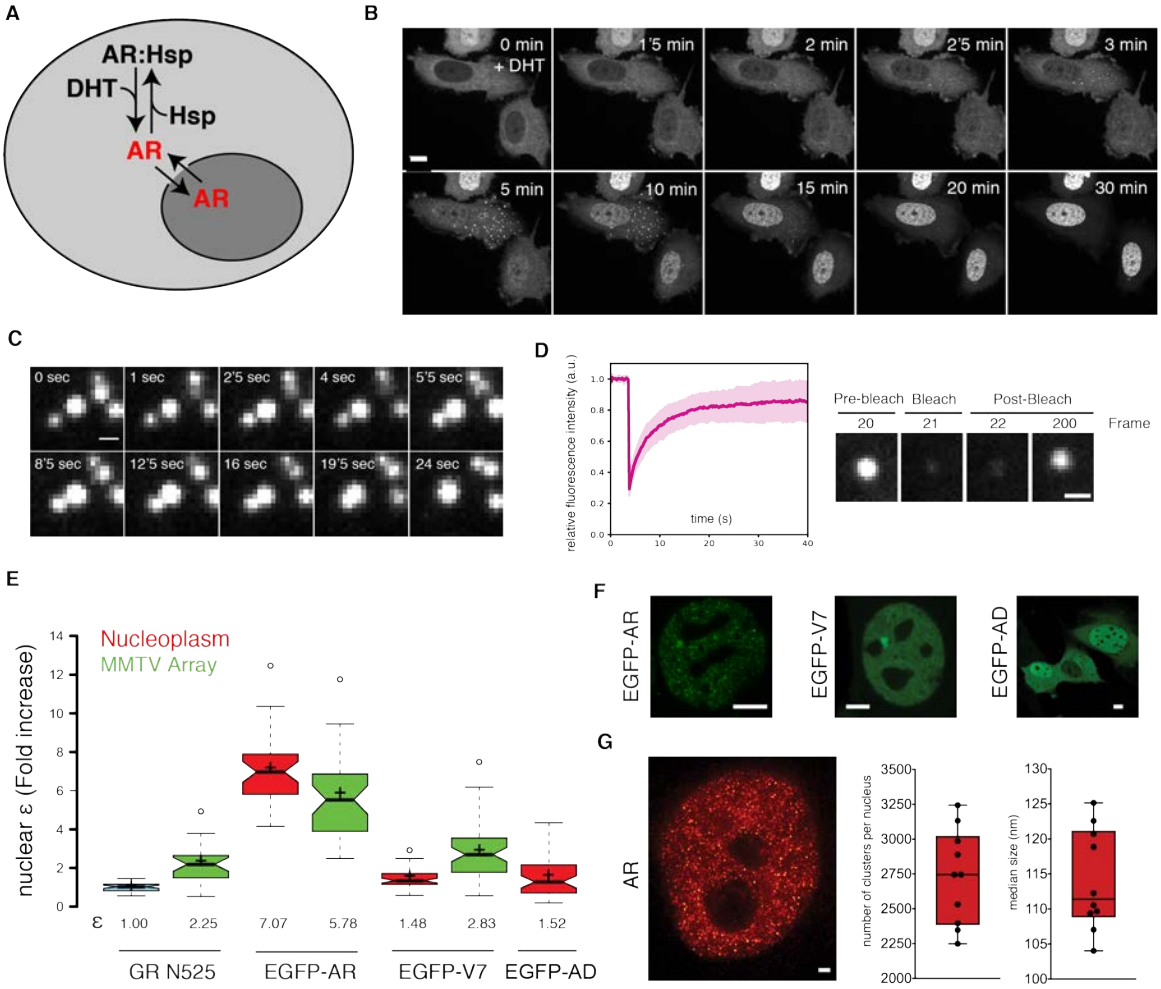


Figure 1 - AR can form oligomers and condensates upon activation: A) Scheme with AR activation pathway, where a change in color from black to red indicates the conformational change that occurs upon activation. DHT - dihydrotestosterone, Hsp - Heat shock protein, AR - androgen receptor. **B)** Time resolved fluorescence microscopy of EGFP-AR condensation upon activation with 1 nM DHT in PC3 cells. Scale bar 10 µm. See also Movie M1. **C)** Snapshots at the indicated time points show a fusion event of EGFP-AR droplets in the cytoplasm of PC3 cells. Scale bar 1 µm. **D)** FRAP of the EGFP-AR droplets in the cytoplasm of 1 nM DHT-treated PC3 cells. Left panel: average relative fluorescence intensity curve of the EGFP-AR cytoplasmic droplets as a function of time following photobleaching. Error bars represent s.d. of n=18 cells per time point. Right panel: representative images of EGFP-AR droplets before and after photobleaching. Scale bar 1 µm. **E)** Oligomeric state of the EGFP-AR variants: WT, AR V7 (AR variant 7) and AR AD (AR activation domain) in the nucleoplasm (red) and upon DNA binding (green) measured by the N&B assay in 3617 cells. Glucocorticoid receptor mutant (GR N525) was used as a reference. The figure shows the fold increase of the molecular brightness (ε) in the nucleus, relative to the control. Centered lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5-fold the interquartile range from the 25th and 75th percentiles, with outliers represented by dots; and crosses represent sample means (n=20). **F)** Fluorescence microscopy images of representative nucleus from 3617 cells transiently transfected with the EGFP-AR variants, after DHT treatment: WT, EGFP-V7 and EGFP-AD. Scale bar 5 µm **G)** On the left: representative STED image of a nucleus from AR-EGFP stably transfected HeLa cell activated with 1nM DHT for 4h. Scale bar 1 µm. On the right: quantification of the number and size of nuclear clusters (mean of 10 nucleus). Each point in the box plot

corresponds to an individual cell; centered lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend to the minimum and maximum value.

The AR AD oligomerizes and undergoes LCST type phase separation in vitro

In order to investigate how AR oligomerization and condensation may affect the main activation function of the protein we used the full-length AD construct, AR_{1-558}, in our *in vitro* experiments. Both the DBD (Shaffer et al. 2004) and the LBD (Nadal et al. 2017) of AR contain well-defined dimerization interfaces that have been characterized by X-ray crystallography and can therefore contribute to AR multimerization (Fig. 2A). The AD is by contrast ID (Lavery and McEwan 2008) of a weak polyampholyte nature (a net negative charge per residue = -0.039) with a relatively well mixed charge patterning (κ = 0.183) (Holehouse et al. 2017) and enrichment of some hydrophobic and aromatic amino acids: Ala, Leu and Tyr, although depletion of Val and Ile in comparison with a set of ID proteins deposited in DisProt 3.4 database (Vacic et al. 2007) (Fig. S6 and Fig. S7). The structural properties of the domain are only beginning to be understood, mainly through the use of solution nuclear magnetic resonance (NMR), a technique that can provide reside-specific structural information of ID proteins (Dyson and Wright 2004) (Fig. 2B). Our previous NMR studies showed that the AD contains several regions of sequence that, although ID, have moderate to high α-helix propensity (De Mol et al. 2016, 2018; Eftekharzadeh et al. 2016; Escobedo et al. 2019) and correspond to AR motifs and sub-domains that are important for transcriptional activity such as the cores of transactivation units 1 and 5 (Tau-1 and Tau-5) (Jenster et al. 1995; Callewaert, Van Tilborgh, and Claessens 2006) (Fig. 2C, top).

To study the oligomerization properties of the AD we used NMR. We assigned the main chain NMR chemical shifts of the full-length AD by transferring the assignments of two previously studied constructs (AD_{1-155} and $AD_{142-448}$) and those of a construct spanning the rest of the domain ($AD_{441-558}$) to the spectrum of the full-length domain (Fig. 2B and 2C). In order to determine whether the structural properties of the residues of the various constructs changed in the context of the full-length domain we compared their main chain chemical shifts in terms of helical propensity by using the δ2D algorithm (Camilloni et al. 2012). The resulting helicity profiles are highly comparable as shown in Fig. 2C indicating that no major change in local structure takes place upon inclusion of the sub-domains in the full-length AD. However, we immediately realized that the NMR signals corresponding to some regions, including some of the ones with helical propensity, were of lower intensity in the $^{1}H,^{15}N$ HSQC spectrum of the AD than in that of AD_{1-155}, $AD_{142-448}$ and $AD_{441-558}$. In some cases, highlighted in Fig. S8, resonances that were observable and assigned in the context of the various subdomains of the AD ceased to be observable under equivalent solution conditions in the full-length AD context. Low signal intensity in NMR spectra of ID proteins due to increased transverse relaxation is often observed in residues involved in transient inter-residue long-range interactions (Klein-Seetharaman et al. 2002). To determine whether the interactions responsible for the low peak intensity observed in the AD are intra or inter-molecular we measured $^{1}H,^{15}N$ HSQC spectra of the AD at a range of concentrations (25, 50, 75 and 100 μM). We observed that the relative intensities of some resonances decreased with concentration, indicating that they mainly stem from inter-molecular interactions (Fig. 2D). We conclude that the AD has propensity to oligomerize in solution.

An analysis of the regions of sequence involved in the inter-molecular interactions revealed contribution of the ^{23}FQNLF^{27} motif, the helical ^{54}LLLL^{57} sequence before the polyQ tract (polyQ tract could not be assigned), the two partially helical motifs of the Tau-5 region, the sequence in a proximity to ^{433}WHTLF^{437} motif (the motif itself ceased to be observable) that can undergo folding-upon-binding (De Mol et al. 2018) and

the highly hydrophobic and aromatic region after the polyglycine (polyG) tract (Fig. 2C and Fig. S6). The NMR experiments were performed at 5ºC and low ionic strength to assure solubility of the protein, since as we showed previously the AR AD can undergo phase separation at room temperature upon addition of a crowding agent *in vitro* (Bouchard et al. 2018). The amino acid bias of the domain, with a large number of Leu, Ala and Tyr residues, and the high content of hydrophobic and aromatic amino acids in the motifs involved in the inter-molecular interactions observed by NMR, made us hypothesize that the hydrophobic and aromatic interactions play a dominant role in the inter-molecular interactions of the AD molecules in the diluted and condensed phase. The hydrophobic or hydrophobic and aromatic interactions driven phase separation should manifest a low critical solution temperature (LCST) behavior that allows for phase separation upon increased temperature conditions (Martin and Mittag 2018). We also noticed that the domain acquires more collapsed states with increase of ionic strength as shown in Fig. S9 by size exclusion chromatography at 4°C, suggesting the screening by salt of repulsive charge-charge interactions and/or strengthening the hydrophobic effect leading to increased intra-molecular hydrophobic interactions.

In order to probe the nature of interactions driving the AD phase separation behavior without the use of crowding agent we first incubated the protein at high ionic strength, 500 mM NaCl, at room temperature and indeed observed formation of the phase separated droplets in concentrations including and above 37.5 μM protein (Fig. 2E). We partially mapped the phase diagram for the AD construct at 500 mM NaCl showing the predicted LCST behavior (Fig. 2F). To further confirm the ionic strength dependence of the phase separation of the domain, we incubated 50 μM of AD at three ionic strengths, 250 mM, 500 mM and 750 mM NaCl, at room temperature and observed formation of phase separated droplets at 500 mM and 750 mM NaCl, with apparent larger number and larger size of droplets at higher salt concentration, Fig. 2G, left. Complementarily, the temperature at which the process occurs, the cloud point temperature, lowered with increased ionic strength, Fig. 2G, right. Together, these observations confirmed the importance of hydrophobic interactions in the AR AD phase separation process.

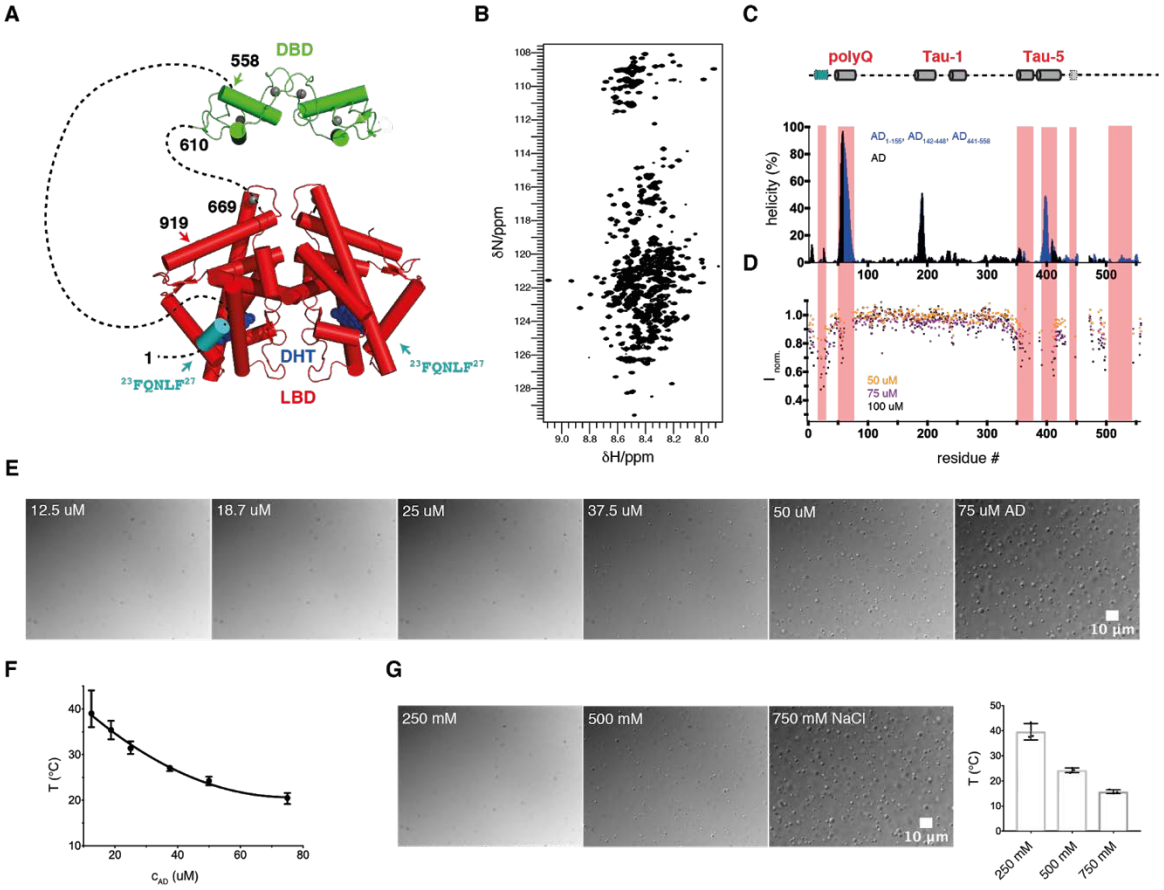


Figure 2 - **AR AD oligomerizes and phase separates *in vitro*: A)** Scheme with the domain structure of AR with the indication of the globular domains in their dimeric states determined by X-ray crystallography (in green DNA Binding Domain with grey spheres representing Zn^{2+} ions, in red Ligand Binding Domain, dashed black lines represent the hinge region, residues 610-669, and the Activation Domain, residues 1-558, with the ^{23}FQNLF^{27} motif in light blue). **B)** ^{1}H,^{15}N HSQC spectrum of the AD (AR_{1-558}). **C)** On the top, a scheme of the AD helical motifs with the indication of the polyQ tract, transactivation units Tau-1 and Tau-5 and two motifs that undergo folding upon binding (^{23}FQNLF^{27} and ^{433}WHTLF^{437} - dashed lines). Below, a plot of the residue-specific helicity of the AD based on the analysis with δ2D algorithm (Camilloni et al. 2012) using the NMR backbone chemical shifts of the three AD_{1-155}, $AD_{142-448}$ and $AD_{441-558}$ constructs of the domain in dark blue and of the full length AD construct in black. **D)** Residue specific loss of the peaks intensity in the 1H-15N HSQC spectra upon concentration of the AD construct (50 µM, 75 µM, 100 µM) normalized by their intensity at 25 µM. Red rectangles indicate regions undergoing loss of peaks intensity with concentration. **E)** AD droplets at 500 mM NaCl at room temperature observed by differential interference contrast (DIC) microscopy at different concentrations of the protein (from 12.5 µM till 75 µM). **F)** Phase diagram of the AD at 500 mM NaCl obtained by turbidimetry. **G)** On the left: AD droplets at 50 µM of the protein at three ionic strengths (250 mM, 500 mM and 750 mM) observed by DIC microscopy at room temperature. On the right: temperatures of the AD cloud points obtained by turbidimetry.

The AR AD droplets undergo fast maturation in vitro

In vitro the AD droplets have, initially, liquid-like character, grow and wet the surface, but readily i.e. within minutes evolve into a gel-like state, in which fusion events are arrested (Fig. 3A and Movie M4). Such liquid-to-solid transitions have been observed in a number of proteins that undergo LLPS and can be associated to changes in the strength and kinetic stability of the inter-molecular interactions that stabilize the condensed phase (Patel et al. 2015; Molliex et al. 2015). In previous studies we found that prior to activation the ^{23}FQNLF^{27} motif in the AR AD is aggregation-prone and that its interaction with the molecular chaperone Hsp70 enhances its solubility *in vitro*, in cells and *in vivo* (Eftekharzadeh et al. 2019). During activation, the motif folds into a helix upon binding to AF-2 in the LBD (Hur et al. 2004; He, Kemppainen, and Wilson 2000), as shown in Fig. 2A, leading to the N/C interaction (Doesburg et al. 1997; M. E. van Royen et al. 2012) and, as we have shown, the oligomerization and condensation of AR (Fig. 1E and 1B). Based on these findings, which illustrate that the ^{23}FQNLF^{27} is either bound to Hsp70 or AF-2 in the LBD during its functional cycle i.e. most probably is never free, we hypothesized that this motif may be responsible for the gelation process of the AR AD *in vitro*.

In order to strongly decrease the aggregation propensity of the ^{23}FQNLF^{27} sequence we decided to replace the Leu residue with a Pro (^{23}FQNPF^{27}). Pro residues cannot form inter-strand hydrogen bonds that stabilize amyloid fibrils and, as a consequence, mutations to Pro strongly decrease the aggregation propensity of amyloidogenic sequences as predicted for the ^{23}FQNLF^{27} to ^{23}FQNPF^{27} mutation by AmylPred2 (consensus-5) algorithm (Tsolis et al. 2013). To investigate whether the aggregation propensity of this motif underlined the gelation process that we observe *in vitro* we studied the effect of this mutation on the phase separation properties of full-length AD (L26P mutant). We observed that the mutation decreased the phase separation propensity of the AD as it increased the temperature of the cloud point of a 50 uM sample at 500 mM NaCl by 8.2°C (Fig. 3B on the left). The mutation, in agreement with our hypothesis, had also a dramatic effect on the material properties of the condensates, increasing the speed of droplets fusions, and also preventing the gelation within the experimental times used. Indeed, the droplets of L26P showed continued fusion events at the incubation times above 5 min, maintaining a liquid-like character (Fig. 3B, movie M5).

We labeled the L26P mutant with Alexa-647 dye and monitored the protein behavior by confocal microscopy after 1 h. We adjusted the conditions of the phase separation to a more physiological-like buffer (150 mM NaCl, 10 % ficoll). In these conditions the droplets formed at 15 µM of the protein at room temperature. The dense phase manifested liquid-like character (fusion, surface wetting). A monolayer of the dense phase sedimented at the bottom of the imaged well was observed at 50 µM of the protein (Fig. 3C, top). To confirm the liquid-like character of the L26P AD droplets we performed FRAP experiments and observed fast recovery, within seconds, of the bleached region (Fig. 3C, bottom). Collectively, we demonstrated the ^{23}FQNLF^{27} motif aggregation causes gelation of the AD droplets and designed a mutant of the AD, L26P, that allowed maintenance of the liquid-like character of the droplets.

To investigate the molecular basis of the ^{23}FQNLF^{27} aggregation we carried out experiments with a short peptide, with sequence Ac-FQNLFQ-NH_2, and a corresponding peptide with the Leu to Pro mutation. The Ac-FQNPFQ-NH_2 solubility increase could be predicted by the CamSol software (Sormanni, Aprile, and Vendruscolo 2015) (Fig. 3D). We confirmed that the ^{23}FQNLF^{27} motif has a high propensity to aggregate into amyloid fibrils and that the Leu to Pro mutation prevents this process (Fig. 3E and F). The

evolution of the Ac-FQNLFQ-NH_2 peptide sample could already be observed after an overnight incubation at 125 µM, as shown by synchronous light scattering in Fig. 3E, and Thioflavin T (ThT) binding in Fig. S10. The fibrillar species of the peptide imaged by Transmission Electron Microscopy (TEM) in Fig. 3F were formed by the preferential parallel disposition of β–strands as shown in FTIR spectrum of the amide I region with a strong band at 1620–1630 cm−1 (Fig. 3G, dark blue shaded area) with no detectable anti-parallel β–sheet band (~1690 cm−1) (Fig 3G). We used the AD L26P as a default system in the further experiments.

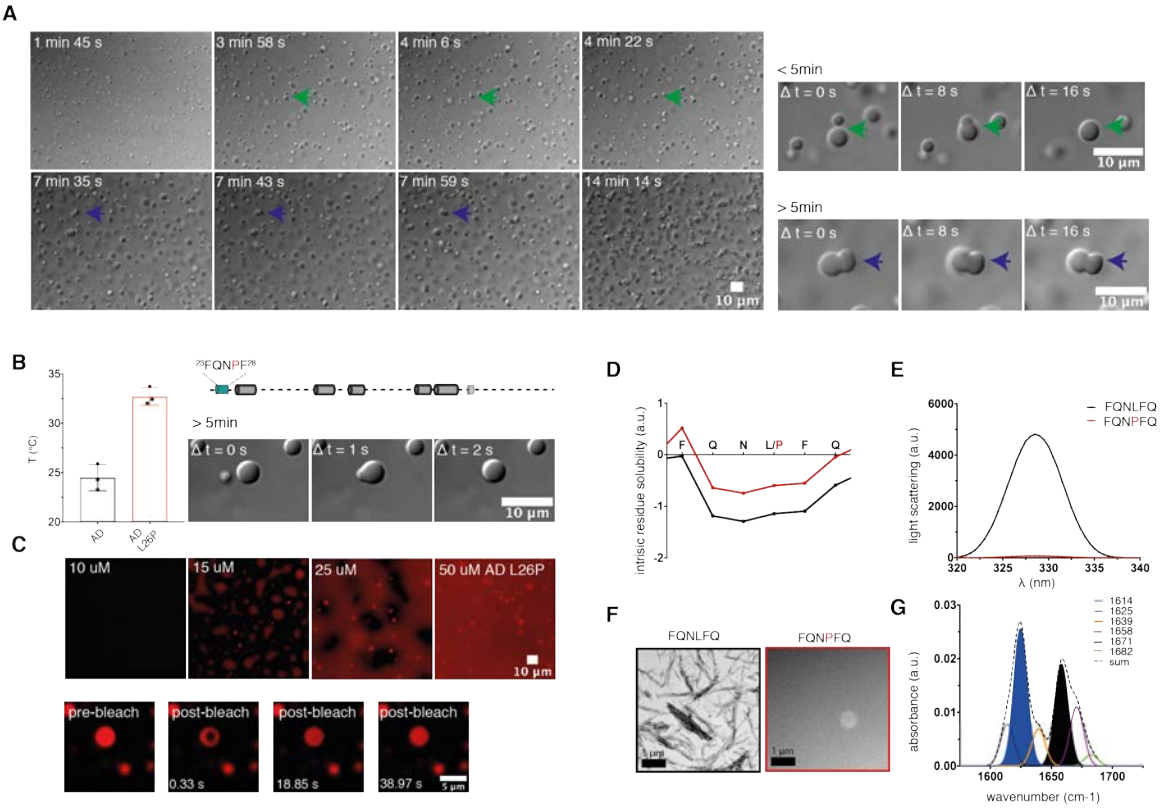


Figure 3 - Aggregation of the ^{23}FQNLF^{27} motif is responsible for droplets maturation *in vitro*: A) DIC images from a time-lapse of the phase separation of 50 µM AD sample at 500 mM NaCl at room temperature. Green arrow points at a fusion event, blue arrow shows the arrest of the fusion in the matured (> 5 min) sample. Zoom in on the right. See also Movie M4. **B)** On the top: a scheme of the AD with the position of the L26P mutation**.** On the left: AD (black bar) and AD L26P (red bar) temperatures of the cloud point of 50 µM samples at 500 mM NaCl obtained by turbidimetry. On the right: DIC images from a time-lapse of the AD L26P phase separated sample incubated for > 5 min. See also Movie M5. **C)** On the top: confocal microscopy images of the AD L26P droplets labeled with Alexa-647 at 150 mM NaCl and 10 % ficoll. On the bottom: representative images of the AD L26P droplet before and after photobleaching in FRAP experiment. **D)** FQNLFQ (black line) and FQNPFQ (red line) peptides solubility predicted by CamSol. **E)** Synchronous light scattering of the FQNLFQ (black line) and FQNPFQ (red line) peptides after an overnight incubation. **F)** Representative TEM micrographs of the FQNLFQ and FQNPFQ peptides after the overnight incubation. **G)** FT-IR absorbance spectrum in the amide I region (dashed line) of the FQNLFQ peptide aggregates. Blue shaded area indicates the intermolecular β-sheet signal contribution to the total spectrum area.

AR AD gains helicity upon condensation

We took advantage of the L26P mutation to address the question of the importance of condensation for the AD secondary structure propensity. In other words, is the phase separation of the AD an inducer of the secondary structure in the activation function motifs? It is known that ID regions often acquire higher secondary structure content with increased temperature (Uversky, Li, and Fink 2001; Uversky et al. 2002; Permyakov et al. 2003), it is then possible that increased hydrophobic contacts and therefore increased folding of the domain persists in the phase separated state of the AD.

In order to test that hypothesis, we first used a known helix inducer, 2,2,2-Trifluoroethanol (TFE). Upon incubation of the domain, AD L26P, with an increasing TFE concentration in the absence of salt at low temperature (5°C), we observed induction of the helical content from 4.7 % at 0% TFE to 22.5 % at 15% TFE by circular dichroism (CD) (Fig 4A). To gain resolution into the regions of helicity gain in the AD L26P, we performed a series of NMR experiments in identical solution conditions at 5°C and measured the Cα and C' chemical shift differences at 2.5 and 5 % TFE versus 0 % TFE (Fig. 4B and S11). Above 5 % of TFE many resonances in the affected regions were not observable. As expected, the mapped regions of increased helicity upon TFE addition corresponded to the regions of existing helical propensity (Fig. 2C): the ^{54}LLLL^{57} sequence before the polyQ tract, the two motifs in the Tau-1 and the less helical sequence in the Tau-5 (residues 341-358). We suspect the remaining prone to structuration regions of the Tau-5, partially helical residues 390-410 and the ^{433}WHTLF^{437} motif, also gain helicity upon incubation with TFE, nevertheless their resonances are not observable in these experiments. In order to probe phase separation propensity of the domain in the presence of the helix inducer, we increased the salt concentration to 500 mM NaCl and measured the temperature of the cloud point of 25 uM sample at different TFE concentrations. We observed a shift of the cloud point temperature of phase separation to lower values with the increase of TFE concentration (Fig. 4C). The observation that the increased helicity is promoting the phase separation of the AD L36P is equivalent with the notion of increased helical content of the domain in the phase separated state.

In order to complement the observations on the increased helical content in the phase separated state of the domain, we generated four mutants in the background of the L26P mutation: L26P polyQ P, L26P Tau-1 P, L26P Tau-5 P and the L26P fullP, where we decreased the helical propensity of the motifs by introducing P mutations in the following positions: L26P polyQ P - L56P; L26P Tau-1 P - A186P, L192P, C238P; L26P Tau-5 P - A356P, A398P, T435P and fullP - all of the former. We hypothesized that the structure-breaking Pro mutations in centers of all the helix forming motifs would disrupt a cooperative helical structure as predicted using the Agadir software (Muñoz and Serrano 1997) (Fig S.12). Indeed, the decrease in the helical content of the mutants resulted in a decreased propensity of the domain to undergo phase separation (Fig. 4D) by 14.7°C for the fullP mutant. Interestingly, the disruption of the LLL-polyQ helix and the disruption of both helical motifs in the Tau-1 region resulted in a similar cloud point temperature shift by 5.7 and 5.8 °C respectively. On the contrary, the helicity in the three motifs of the Tau-5 region had a larger effect resulting in the temperature shift of the cloud point by 9.6°C, suggesting a larger contribution of helicity in the Tau-5 region to the phase separation propensity of the domain.

Experiments with the TFE and the mutants show that the helical character of the activation motifs is important for the phase separation propensity of the domain (Fig. 4C and 4D). These additional experiments also show that the Tau-1 helical regions are also

important for the inter-molecular interactions of the domain, suggesting that NMR is biased to detecting oligomerization mediated by aromatic residues (Conicella et al. 2016) (Fig. 1C). Collectively, the experiments show the helicity is promoted in the phase separated state of the domain.

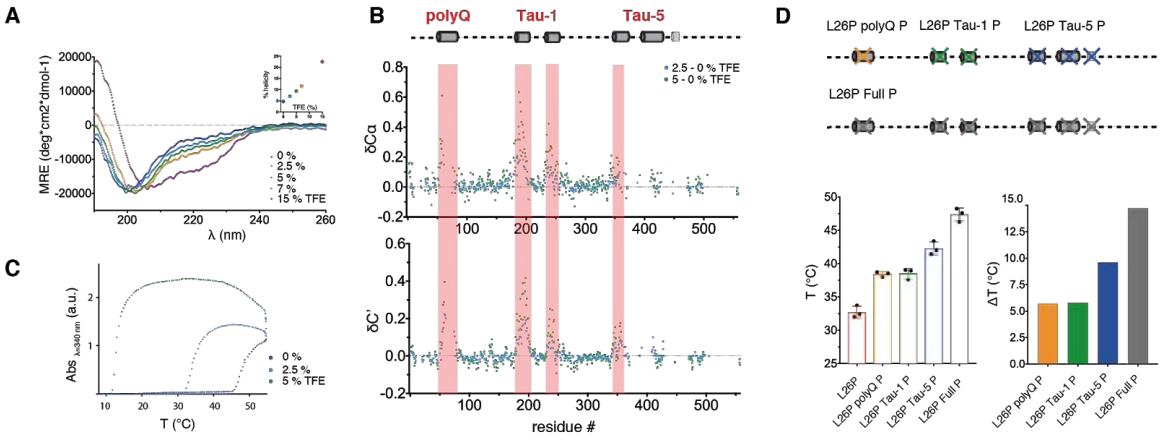


Figure 4 - AR AD acquires helical structure in functionally relevant activation motifs upon condensation: A) CD spectra of the AD L26P upon TFE titration measured at 5°C and plot of the AD L26P helicity determined by CD as a function of TFE concentration. **B)** On the top: a scheme of the AD L26P helical motifs with the indication of the polyQ tract, partially helical activation motifs in the Tau-1 and Tau-5. ^{433}WHTLF^{437} motif that undergoes folding upon binding is represented with dashed lines. Below: plots of the difference of the Cα and C' NMR chemical shifts between the AD L26P incubated with 2.5 or 5 % TFE and 0 % TFE at 5°C. Red boxes indicate regions of gain in helical propensity. **C)** Shift in the temperature of the AD L26P cloud point as a function of TFE concentration obtained by turbidimetry at 25 uM and 500 mM NaCl. **D)** On the top: a scheme of the positions of Pro mutations in the AD L26P activation motifs: L26P polyQ P, L26P Tau-1 P, L26P Tau-5 P and L26P fullP. Below: the effect of the mutations on the cloud point temperature of the AD L26P phase separation monitored by turbidimetry at 50 uM protein and 500 mM NaCl.

Discussion

Recent developments have shown that some transcriptionally active compartments have properties equivalent to those of condensates produced by LLPS (Benjamin R. Sabari et al. 2018). Such condensates are generally stabilized by transient interactions between multi-domain proteins with ID regions (IDRs) and RNA or DNA. A key property of these condensates is their highly dynamic nature that allows them to readily assemble and disassemble as a function of specific signals such as phosphorylation, interactions with small molecules or changes in solution conditions. Since TF ADs are rich in ID and contain specific activation motifs for interacting with members of the transcription complex it has been hypothesized that they have the propensity to phase separate into liquid droplets at the initiation of transcription. Recently, estrogen receptor (ER), a member of the SR family, has been shown to co-phase separate with the Mediator complex subunit, MED1 IDR, and the ER incorporation into the droplets was greatly enhanced by the MED1 LxxLL motif interaction with the AF-2 surface on the LBD upon estrogen activation of the ER (Boija et al. 2018). The same SR was also shown to promote a cooperative enhancer assembly via phase separation

in human breast cancer cells (Nair et al. 2019). It is possible then that all SRs can co-phase separate with other members of transcription complex in order to execute an efficient cooperative initiation of transcription upon hormone stimulation when needed and that this transcription complex assembly relies on non-specific multivalent interactions within the phase separated state of proteins. Nevertheless, this picture lacks explanation for the specific selection for binding partners by TF ADs. Here, we show that AR, another member of the SR family, can undergo LLPS and forms cytoplasmic and likely nuclear condensates, previously reported as AR *compartments, chromatin-associated bodies, nuclear foci* or *speckled distribution* of the protein (Georget et al. 1997; R. K. Tyagi et al. 2000; A. Tomura et al. 2001). We prove that the formation of the higher than dimer oligomers and condensates of the protein depends on the presence of the ID AD and the globular LBD, suggesting a contribution of the intra- or intermolecular N/C interaction or interactions with binding partners to their formation. We also provide an explanation for the specificity of the AD-partner interactions that occur in the transcription complex condensates. *In vitro* the AD of AR has a strong propensity to phase separate and its phase separation is driven by hydrophobic or hydrophobic and aromatic inter-molecular interactions formed by ID region after the polyG tract and partially folded activation motifs, including ΩxxLΩ and LxxLL motifs. Mutations that decrease the helical propensity of these motifs strongly decrease the propensity of AR to phase separate, indicating that they gain helical structure upon phase separation. Gain of the helicity in the functionally relevant motifs would increase their specificity for the binding partners allowing for efficient conformational selection. Additionally, the gain of helicity in each linear motif of the AD would depend on droplet formation providing a cooperative mechanism of the domain folding upon condensation, which as a consequence would allow for efficient cooperative assembly of the transcription complex.

Acknowledgements

The authors wish to thank the ICTS NMR facility, managed by the scientific and technological centers of the University of Barcelona (CCiT UB), for their help in NMR. We also thank the Advanced Digital Microscopy (ADM) and Protein Expression Core Facility (IRB Barcelona). We wish to thank Maria Penutto (University of Padova, Italy) for providing us with the AR-EGFP stably transfected HeLa cell line. X.S. acknowledges funding from AGAUR (2017 SGR 324), Marató TV3 (102030), MINECO (BIO2012-31043 and BIO2015-70092-R) and the European Research Council (CONCERT, contract number 648201). IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO (Government of Spain).

Methods

Protein purification. $AD_{441-558}$ was expressed as a his-tag fusion in E. coli Rosetta (DE3) cells and purified from inclusion bodies. Cells were lysed by sonication. The insoluble fraction was washed with PBS buffer pH= 7.4 containing 1% Triton, and dissolved in 8M Urea, 20 mM Tris pH 7.8, 500 mM NaCl, and 14 mM b-ME. The solution was then passed over a Ni-HP and eluted with the same buffer with 500 mM Imidazole. The tag was cleaved with 1mg TEV protease/100 mg protein at 4 °C overnight under dialysis, and removed by passing back over Ni resin in 8 M Urea, 50 mM Tris pH 7.8, 100 mM NaCl, and 1 mM b-ME. The flow-through was then concentrated (Amicon 3 MWCO) and subjected to SEC on a Superdex 75 16/60 (GE Healthcare) equilibrated with appropriate buffer and concentrated (Amicon 3 MWCO). $AD_{142-448}$, AD_{1-155} and the full-length AD were obtained as previously described (De Mol et al. 2016; Eftekharzadeh et al. 2016; Bouchard et al. 2018). The AD construct was subjected to SEC on a Superdex 200 16/60

or Superdex 200 10/300 Increase (GE Healthcare) equilibrated with appropriate buffer, concentrated (Amicon 10 MWCO) and used immediately.

Nuclear magnetic resonance (NMR). The assignment of the AD was obtained by a divide and conquer approach using AD_{1-155} and $AD_{142-448}$ and $AD_{441-558}$ fragments and confirmed with conventional three-dimensional triple resonance experiments. All experiments in the study were acquired with standard Bruker pulse sequences on Bruker 600 and 800 MHz spectrometers at 278 K in 20 mM sodium phosphate buffer with 1 mM TCEP and 0.05 % NaN_3 at pH 7.4. The TFE titration in experiment in Fig. 4B was performed with deuterated TFE (Sigma). The residue-specific peak intensities in the experiment in Fig. 2D and S8 were obtained by dividing the peaks intensities at indicated concentrations by the intensity of residue 154 Asp at the same concentration (internal normalization) and divided by their values at 25 uM.

Turbidimetry of protein samples. Samples were prepared on ice in 20 mM HEPES buffer pH 7.4 with 2 mM TCEP, 1 mM EDTA and 0.05 % NaN_3 with indicated concentration of NaCl or in 20 mM phosphate buffer pH 7.4 with 2 mM TCEP, 1 mM EDTA, 0.05 % NaN_3 at indicated NaCl concentration and indicated percentage of TFE in experiment in Fig. 4C. Samples were centrifuged for 20 min at 21130 rcf at 4 °C. Supernatant was subjected to measurement. The phase separation cloud point of protein samples was monitored by their optical density as a function of temperature on a Cary 100 ultraviolet–visible spectrophotometer equipped with a multicell thermoelectric temperature controller at a rate of $1\,°C\,min^{-1}$ heating. The plotted cloud point temperatures were obtained as 1^{st} order derivatives of the obtained curves from 3 samples.

Circular dichroism (CD). Spectra were acquired on 50 µM samples in 20 mM sodium phosphate buffer pH 7.4 with 1 mM TCEP at indicated TFE (Alfa Aesar) concentration in a Jasco 815 UV spectrophotopolarimeter at 285 K with a 0.01 mm optical path cuvette. Spectra deconvolution to determine secondary structure propensity was performed with the analysis software CONTIN (reference set 7) hosted at DichroWeb61 (dichroweb.cryst.bbk.ac.uk).

Microscopy of protein samples. Samples were prepared in PCR low binding 200 ul tubes (Eppendorf) on ice, in the case of fast evolving WT AD or RT in the case of L26P AD mutants. The proteins were in 20 mM HEPES buffer pH 7.4 with 2 mM TCEP, 1 mM EDTA and 0.05 % NaN_3 with indicated concentration of NaCl. In experiment in Fig. 3C the samples were prepared in 20 mM HEPES buffer pH 7.4 with 2 mM TCEP, 1 mM EDTA, 0.05 % NaN_3, 150 mM NaCl and 10 % Ficoll 70 (Sigma). Samples were transferred into sealed sample chambers containing protein solutions comprised coverslips sandwiching two layers of 3M 300 LSE high-temperature double-sided tape (0.34 mm) or 384 WELL NON-BINDING MICROPLATE, µCLEAR® from Greiner Bio-One (781906). The DIC images were taken using an Automated Inverted Microscope TIRF - ScanR Olympus with a Hamamatsu Orca-ER camera and a 60x/1.20 water UPlan SAPo objective. The confocal images were recorded using Zeiss LSM780, confocal microscope system with Plan ApoChromat 63x 1.4 oil objective. The samples were imaged immediately in order to monitor samples evolution or after 30 min to 1 h 30 min for other purposes. The sample in Movie M4 was recorded at 100 uM AD and 500 mM NaCl, the sample in Movie M5 was recorder at 50 uM AD L26P and 1 M NaCl.

Fluorescence recovery after photobleaching (FRAP) of protein samples. The FRAP experiments were recorded using Zeiss LSM780, confocal microscope system with Plan ApoChromat 63x 1.4 oil objective. Coverslips and slides for FRAP experiments were PEGylated according to the published protocol (Alberti et al. 2018) and the

measurements were collected in 3 independent experiments within 30 min to 1 h from sample preparation. The bleached region was around 30 % of the droplet diameter in the middle section.

Peptides. FQNLFQ and FQNPFQ synthetic peptides were obtained as lyophilized powders with >95% purity from GenScript (Piscataway, NJ, USA) with Amidated C and Acetylated N termini. The lyophilized peptides were solubilized at a final concentration of 5 mM in dimethyl sulfoxide (DMSO). Right before each experiment, the stock solutions were diluted to 125 µM in 20 mM HEPES buffer pH 7.5 with 150 mM NaCl. For aggregation assays the samples were incubated overnight at 37ºC at 600 rpm agitation.

Synchronous light scattering. Synchronous light scattering was monitored using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 360 nm, emission range from 350 to 370 nm, slit widths of 5 nm, 0.5 nm interval and 1000 nm/min scan rate. The peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before measurement.

Fourier transform infrared (FT-IR) spectroscopy. FTIR experiments were performed using a Bruker Tensor 27 FT-IR spectrometer (Bruker Optics Inc) with a Golden Gate MKII ATR accessory. Each spectrum consists of 16 independent scans, measured at a spectral resolution of 4 cm−1 within the 1800–1500 cm−1 range. All spectral data were acquired and normalized using the OPUS MIR Tensor 27 software. Data was afterwards deconvoluted using the Peak Fit 4.12 program. The buffer without peptide was used as a control and subtracted from the absorbance signal before deconvolution.

Transmission electron microscopy (TEM). The morphology of the aggregated FQNLFQ peptide was evaluated by negative staining and using a JEOL JEM-1400Plus Transmission Electron Microscope. 5 µl of peptide solution was placed on carbon-coated copper grids and incubated for 5 min. The grids were then washed and stained with 5 µl of 2% w/v uranyl acetate for 5 min. Then, grids were washed again before analysis. Images and movies were processed with ImageJ.

Thioflavin-T (Th-T) binding. The fluorescence spectra of Th-T were recorded using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 440 nm, emission range from 460 to 600 nm, slit widths of 5 nm, 0,5 nm interval and 1000 nm/min scan rate. Peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before dye addition. 5 µl of the sonicated aggregated peptide was added to 100 µl of 25 µM ThT in buffer. The same buffer with 25 µM ThT and without peptide was employed as a control.

Cell culture. PC3 cells were obtained from ATCC (CRL-1435) and cultured in RPMI containing 4.5 g L^{-1} glucose (Glutamax, Gibco) supplemented with 10% (v/v) charcoal stripped FBS (Gibco) and antibiotics. HEK293T cells (ATCC; CRL-3216) and AR-EGFP HeLa stable cells (gift from Pennuto lab) were maintained in DMEM containing 4.5 g L^{-1} glucose supplemented with 10% (v/v) charcoal stripped FBS and antibiotics. 3617 mouse cell line (McNally et al. 2000) was routinely cultured in DMEM high glucose supplemented with 10% (v/v) charcoal stripped FBS and 2 mM l-glutamine (Life Technologies) and in the presence of 5 µg/mL tetracycline (Sigma–Aldrich) to prevent expression of a stably integrated GFP-GR (Presman et al. 2014). Cells were cultured in a humidified atmosphere containing 5% CO_2 at 37 °C.

Antibodies. For Western blot analyses, the following antibodies were used: anti-β-actin-HRP (ab8224, 1:10000) and anti-androgen receptor (441) (sc-7305, 1:1000). Anti-mouse HRP-conjugated secondary antibody was from Bio-Rad (170-6516, 1:10000).

Western blot (WB). Cells were washed and harvested in PBS 1x, lysed in Passive Lysis buffer 1x (Promega) containing phosphatase and protease inhibitors (Roche). Lysates were centrifuged at 15000 g to separate soluble and pellet fractions. Total protein was quantified using BCA assay (Pierce Biotechnology). Proteins were resolved by 7.5 or 15% SDS-PAGE, transferred to PVDF membranes and incubated with specific antibodies.

Live-cell imaging. PC3 cells were seeded in collagen I-coated µ-slide 4-well Glass Bottom plates (Ibidi 80426) at 60% confluency 24 hours before transfection. Then, 170 ng of expression vectors encoding androgen receptor (AR) tagged with eGFP (eGFP-AR) or mutants were transfected per well using polyethylenimine (PEI) (Polysciences) at a ratio of 1 µg DNA to 3 µl PEI. Four hours after transfection, media was changed to RPMI supplemented with 10% charcoal stripped FBS and cells were cultured for 16 hours before imaging. Transiently transfected PC3 cells expressing eGFP-AR were imaged in 3D for one minute every 15 seconds to acquire a baseline readout of AR expression. Cells were then immediately treated with 1 nM of DHT and imaged consecutively for 1 h every 15-sec time interval. Time lapse imaging was performed with a Spinning Disk confocal system (Olympus IX81 + Yokogawa CSU-XI scan head) from Andor (Belfast, Ireland) and a 60x/1.42 Oil Plan Apo N objective. A stable temperature (37°C) was maintained during imaging in a CO_2 and temperature regulated incubation chamber (EMBL, Heidelberg, Germany). eGFP was excited with a 488 nm laser and Z-stack images were acquired every 0.45 µm step size. Time lapse images were compiled, processed and edited with Fiji (ImageJ). Intensity thresholds were set manually and uniformly to minimize background noise.

FRAP assay in live cells. PC3 cells were transfected and prepared for microscopy in identical conditions to those of live cell imaging experiments. Before performing Fluorescence Recovery after Photobleaching (FRAP), cells were treated with 1 nM DHT. FRAP data for each condition were acquired over the course of approximately 1 hour, combining results for each condition as no trend was observed between FRAP data acquired at the beginning versus the end of the hour. FRAP measurements were performed on a Spinning Disk Confocal Microscope equipped with an iXon EMCCD Andor DU-897 camera at 100x/1.40 Oil UPlanSApo. Pre-bleaching and fluorescence recovery images of the EGFP-AR were acquired with the same 488 nm laser power with an exposure time of 180 msec in a 10x10 pixel square format. Bleaching was done for 5 times repetition at maximum intensity 488 nm laser power at 100 msec. Twenty pre-bleached images and 200 post-bleached images were taken in total. Post-bleached images were acquired immediately after the 5 bleaching point steps. Mean intensity measurements were quantified in three different Regions of Interest (ROIs) in each FRAP experiment: A bleached region, a background region outside the cells and a region spanning the whole cell. Fiji (ImageJ) was used to measure fluorescence intensities. Exported csv tables were normalized and fitted in EasyFrap software (Rapsomaniki et al. 2012) in order to extract kinetic parameters such as T-half and mobile fraction. Double normalization was used to correct for fluorescence bleaching during imaging and for intensity level differences. The curves were fitted to equations for double-exponential recovery.

Sub-cellular localization and Number & Brightness (N&B) analysis. Images were taken using an LSM 780 laser scanning microscope (Carl Zeiss, Inc.) equipped with an environmental chamber. Cells were imaged from 20 min after hormone addition up to a maximum of 2 h. We used a 63× oil immersion objective (N.A. = 1.4). The excitation source was a multiline Ar laser tuned at 488 nm. Fluorescence was detected with a gallium arsenide phosphide (GaAsP) detector in photon-counting mode. N&B

measurements were done as previously described (Presman et al. 2014). Briefly, for each studied cell, a single-plane stack of 150 images (256 × 256 pixels) was taken in the conditions mentioned above, setting the pixel size to 80 nm and the pixel dwell time to 6.3 µs. In every case, we discarded the first 10 images of the sequence to reduce overall bleaching. The frame time under these conditions is 0.97 s, which guarantees independent sampling of molecules according to previously reported fluorescence correlation spectroscopy (FCS) measurements (Mikuni, Tamura, and Kinjo 2007). Each stack was further analyzed using the N&B routine of the "GLOBALS for Images" program developed at the Laboratory for Fluorescence Dynamics (University of California, Irvine, CA). In this routine, the average fluorescence intensity (<I>) and its variance (σ2) at each pixel of an image are determined from the intensity values obtained at the given pixel along the image stack. The apparent brightness (B) is then calculated as the ratio of σ2 to <I>, whereas the apparent number (N) of moving particles corresponds to the ratio of <I> to B. Classification of pixels according to their intensity values easily allows splitting of the cytoplasm, nucleus, and array for further analysis. Selection of cells for analysis followed these criteria: (i) in the case of stimulated cells, an accumulation of signal at the array must be visible; (ii) the average N of molecules in the nuclear compartment must have a range of three to 18 in all cases; (iii) no saturation of the detector at any pixel (n < 60); and (iv) bleaching cannot be more than 5–10%. In a previous work, it has been demonstrated that B is equal to the real brightness, ε, of the particles plus 1 (Digman et al. 2008). Therefore, ε at every pixel of images can be easily extracted from B measurements. Importantly, this analysis only provides information regarding moving or fluctuating fluorescent molecules because fixed molecules (relative to our frame time) will give B values equal to 1. The experiments were independently repeated two to three times for each treatment/condition.

Stimulated emission depletion (STED) super-resolution microscopy on fixed cells. AR-EGFP HeLa stable cells (a kind gift from Pennuto lab) were cultured on a glass bottomed cellview cell culture dish (Greiner bio-one) in DMEM media provided with 10% Charcoal-Stripped Serum (A33821-01, Life) for 24h prior addition of 1nM DHT for 4h. Then, cells were fixed with PBS containing 4% p-Formaldehyde for 15 min. After washing with PBS, the cells quenched with 300mM Glycine and 0.3% Triton in PBS and permeabilized with 0.5% Triton in PBS. Cells were subsequently blocked with 2% BSA in PBS. Finally, AR-EGFP was stained with GFP-Booster Abberior AS 635P 1:50.

Confocal and STED imaging, on fixed and live samples, was performed using an Abberior 3D-2 Color-STED system (Abberior Instruments, Goettingen) with a 100x/1.4 NA oil or a 60x/1.2 NA water objectives (Olympus). Star580 was imaged with a pulsed laser at 560 nm, and excitation of Abberior Star Red probe was performed at 640 nm (Göttfert et al. 2013). The depletion laser for both colors was a 775 nm pulsed laser (Katana HP, 3W, 1ns pulse duration, NKT Photonics).

All images were processed, segmented and analysed using the custom-designed image analysis software MotionTracking (Morales-Navarrete et al. 2015), as previously described (Rink et al. 2005; Murray et al. 2016; Senigagliesi et al. 2019). First the background intensity was subtracted from the images and high frequency noise was removed using a Gaussian low-pass filter (sigma = 0.02 µm). Then, the objects of interest were reconstructed using the model-based segmentation approach (Rink et al. 2005), i.e. objects were recognized by fitting the image intensity with a sum of powered Lorenzian functions (Rink et al. 2005). The different features (e.g. number of objects, size, mean intensity) were calculated only on the segmented objects inside a Region of Interest (ROI). ROIs were automatically defined using a threshold-based masking. Object size (L) was defined as $L=(A/n)^{1/2}$, where A is the area of the object.

References

Alberti, Simon, Amy Gladfelter, and Tanja Mittag. 2019. “Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates.” *Cell* 176 (3): 419–34.

Alberti, Simon, Shambaditya Saha, Jeffrey B. Woodruff, Titus M. Franzmann, Jie Wang, and Anthony A. Hyman. 2018. “A User’s Guide for Phase Separation Assays with Purified Proteins.” *Journal of Molecular Biology*, #pagerange#.

Alen, P., F. Claessens, G. Verhoeven, W. Rombauts, and B. Peeters. 1999. “The Androgen Receptor Amino-Terminal Domain Plays a Key Role in p160 Coactivator-Stimulated Gene Transcription.” *Molecular and Cellular Biology* 19 (9): 6085–97.

Boehning, Marc, Claire Dugast-Darzacq, Marija Rankovic, Anders S. Hansen, Taekyung Yu, Herve Marie-Nelly, David T. McSwiggen, et al. 2018. “RNA Polymerase II Clustering through Carboxy-Terminal Domain Phase Separation.” *Nature Structural & Molecular Biology* 25 (9): 833–40.

Boija, Ann, Isaac A. Klein, Benjamin R. Sabari, Alessandra Dall’Agnese, Eliot L. Coffey, Alicia V. Zamudio, Charles H. Li, et al. 2018. “Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains.” *Cell*, 1–14.

Bolton, Eric C., Alex Y. So, Christina Chaivorapol, Christopher M. Haqq, Hao Li, and Keith R. Yamamoto. 2007. “Cell- and Gene-Specific Regulation of Primary Target Genes by the Androgen Receptor.” *Genes and Development* 21 (16): 2005–17.

Bouchard, Jill J., Joel H. Otero, Daniel C. Scott, Elzbieta Szulc, Erik W. Martin, Nafiseh Sabri, Daniele Granata, et al. 2018. “Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments.” *Molecular Cell* 72 (1): 19–36.e8.

Callewaert, Leen, Nora Van Tilborgh, and Frank Claessens. 2006. “Interplay between Two Hormone-Independent Activation Domains in the Androgen Receptor.” *Cancer Research* 66 (1): 543–53.

Camilloni, Carlo, Alfonso De Simone, Wim F. Vranken, and Michele Vendruscolo. 2012. “Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts.” *Biochemistry* 51 (11): 2224–31.

Cho, Won Ki, Namrata Jayanth, Brian P. English, Takuma Inoue, J. Owen Andrews, William Conway, Jonathan B. Grimm, et al. 2016. “RNA Polymerase II Cluster Dynamics Predict mRNA Output in Living Cells.” *eLife* 5 (MAY2016). https://doi.org/10.7554/eLife.13617.

Cisse, Ibrahim I., Ignacio Izeddin, Sebastien Z. Causse, Lydia Boudarene, Adrien Senecal, Leila Muresan, Claire Dugast-darzacq, and Bassam Hajj. 2013. “Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells.” *Science* 245 (August): 664–67.

Conicella, Alexander E., Gül H. Zerze, Jeetain Mittal, and Nicolas L. Fawzi. 2016. “ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain.” *Structure* 24 (9): 1537–49.

De Mol, Eva, R. Bryn Fenwick, Christopher T. W. Phang, Victor Buzón, Elzbieta Szulc, Alex de la Fuente, Albert Escobedo, et al. 2016. “EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor.” *ACS Chemical Biology* 11 (9): 2499–2505.

De Mol, Eva, Elzbieta Szulc, Claudio Di Sanza, Paula Martínez-Cristóbal, Carlos W. Bertoncini, R. Bryn Fenwick, Marta Frigolé-Vivas, et al. 2018. “Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators.” *Structure* 26 (1): 145–52.e3.

Digman, Michelle A., Rooshin Dalal, Alan F. Horwitz, and Enrico Gratton. 2008. “Mapping the Number of Molecules and Brightness in the Laser Scanning Microscope.” *Biophysical Journal* 94 (6): 2320–32.

Doesburg, P., C. W. Kuil, C. A. Berrevoets, K. Steketee, P. W. Faber, E. Mulder, A. O. Brinkmann, and J. Trapman. 1997. “Functional in Vivo Interaction between the Amino-

Terminal, Transactivation Domain and the Ligand Binding Domain of the Androgen Receptor." *Biochemistry* 36 (5): 1052–64.
Dyson, H. Jane, and Peter E. Wright. 2004. "Unfolded Proteins and Protein Folding Studied by NMR." *Chemical Reviews* 104 (8): 3607–22.
Eeckhoute, Jéröme, Raphaël Métivier, and Gilles Salbert. 2009. "Defining Specificity of Transcription Factor Regulatory Activities." *Journal of Cell Science* 122 (Pt 22): 4027–34.
Eftekharzadeh, Bahareh, Varuna C. Banduseela, Giulio Chiesa, Paula Martínez-Cristóbal, Jennifer N. Rauch, Samir R. Nath, Daniel M. C. Schwarz, et al. 2019. "Hsp70 and Hsp40 Inhibit an Inter-Domain Interaction Necessary for Transcriptional Activity in the Androgen Receptor." *Nature Communications* 10 (1): 3562.
Eftekharzadeh, Bahareh, Alessandro Piai, Giulio Chiesa, Daniele Mungianu, Jesús García, Roberta Pierattelli, Isabella C. Felli, and Xavier Salvatella. 2016. "Sequence Context Influences the Structure and Aggregation Behavior of a PolyQ Tract." *Biophysical Journal* 110 (11): 2361–66.
Escobedo, Albert, Busra Topal, Micha B. A. Kunze, Juan Aranda, Giulio Chiesa, Daniele Mungianu, Ganeko Bernardo-Seisdedos, et al. 2019. "Side Chain to Main Chain Hydrogen Bonds Stabilize a Polyglutamine Helix in a Transcription Factor." *Nature Communications* 10 (1). https://doi.org/10.1038/s41467-019-09923-2.
Fulton, Debra L., Saravanan Sundararajan, Gwenael Badis, Timothy R. Hughes, Wyeth W. Wasserman, Jared C. Roach, and Rob Sladek. 2009. "TFCat: The Curated Catalog of Mouse and Human Transcription Factors." *Genome Biology* 10 (3): R29.
Fuxreiter, Monika, Peter Tompa, István Simon, Vladimir N. Uversky, Jeffrey C. Hansen, and Francisco J. Asturias. 2008. "Malleable Machines Take Shape in Eukaryotic Transcriptional Regulation." *Nature Chemical Biology* 4 (12): 728–37.
Gelmann, E. P. 2002. "Molecular Biology of the Androgen Receptor." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 20 (13): 3001–15.
Georget, V., J. M. Lobaccaro, B. Terouanne, P. Mangeat, J. C. Nicolas, and C. Sultan. 1997. "Trafficking of the Androgen Receptor in Living Cells with Fused Green Fluorescent Protein-Androgen Receptor." *Molecular and Cellular Endocrinology* 129 (1): 17–26.
Göttfert, Fabian, Christian A. Wurm, Veronika Mueller, Sebastian Berning, Volker C. Cordes, Alf Honigmann, and Stefan W. Hell. 2013. "Coaligned Dual-Channel STED Nanoscopy and Molecular Diffusion Analysis at 20 Nm Resolution." *Biophysical Journal* 105 (1): L01–03.
He, Bin, Robert T. Gampe, Adam J. Kole, Andrew T. Hnat, Thomas B. Stanley, Gang An, Eugene L. Stewart, Rebecca I. Kalman, John T. Minges, and Elizabeth M. Wilson. 2004. "Structural Basis for Androgen Receptor Interdomain and Coactivator Interactions Suggests a Transition in Nuclear Receptor Activation Function Dominance." *Molecular Cell* 16 (3): 425–38.
He, Bin, Jon a. Kemppainen, and Elizabeth M. Wilson. 2000. "FXXLF and WXXLF Sequences Mediate the NH2-Terminal Interaction with the Ligand Binding Domain of the Androgen Receptor." *The Journal of Biological Chemistry* 275 (30): 22986–94.
Hnisz, Denes, Krishna Shrinivas, Richard A. Young, Arup K. Chakraborty, and Phillip A. Sharp. 2017. "A Phase Separation Model for Transcriptional Control." *Cell* 169 (1): 13–23.
Holehouse, Alex S., Rahul K. Das, James N. Ahad, Mary O. G. Richardson, and Rohit V. Pappu. 2017. "CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins: Biophysical Journal." *Biophysical Journal* 112 (1): 16–21.
Huang, Pengxiang, Vikas Chandra, and Fraydoon Rastinejad. 2010. "Structural Overview of the Nuclear Receptor Superfamily: Insights into Physiology and Therapeutics." *Annual Review of Physiology* 72: 247–72.
Hur, Eugene, Samuel J. Pfaff, E. Sturgis Payne, Hanne Grøn, Benjamin M. Buehrer, and Robert J. Fletterick. 2004. "Recognition and Accommodation at the Androgen Receptor Coactivator Binding Interface." *PLoS Biology* 2 (9): E274.

Hyman, Anthony A., Christoph A. Weber, and Frank Jülicher. 2014. “Liquid-Liquid Phase Separation in Biology.” *Annual Review of Cell and Developmental Biology* 30 (1): 39–58.
Jenster, G., H. A. van der Korput, J. Trapman, and A. O. Brinkmann. 1995. “Identification of Two Transcription Activation Units in the N-Terminal Domain of the Human Androgen Receptor.” *The Journal of Biological Chemistry* 270 (13): 7341–46.
Keegan, L., G. Gill, and M. Ptashne. 1986. “Separation of DNA Binding from the Transcription-Activating Function of a Eukaryotic Regulatory Protein.” *Science* 231 (4739): 699–704.
Kim, Yujin E., Mark S. Hipp, Andreas Bracher, Manajit Hayer-Hartl, and F. Ulrich Hartl. 2013. “Molecular Chaperone Functions in Protein Folding and Proteostasis.” *Annual Review of Biochemistry* 82: 323–55.
Klein-Seetharaman, Judith, Maki Oikawa, Shaun B. Grimshaw, Julia Wirmer, Elke Duchardt, Tadashi Ueda, Taiji Imoto, Lorna J. Smith, Christopher M. Dobson, and Harald Schwalbe. 2002. “Long-Range Interactions within a Nonnative Protein.” *Science* 295 (5560): 1719–22.
Klokk, T. I., P. Kurys, C. Elbi, A. K. Nagaich, A. Hendarwanto, T. Slagsvold, C-Y Chang, G. L. Hager, and F. Saatcioglu. 2007. “Ligand-Specific Dynamics of the Androgen Receptor at Its Response Element in Living Cells.” *Molecular and Cellular Biology* 27 (5): 1823–43.
Koehler, Angela N. 2010. “A Complex Task? Direct Modulation of Transcription Factors with Small Molecules.” *Current Opinion in Chemical Biology* 14 (3): 331–40.
Kumar, Sanjay, and Rakesh K. Tyagi. 2012. “Androgen Receptor Association with Mitotic Chromatin – Analysis with Introduced Deletions and Disease-Inflicting Mutations.” *The FEBS Journal* 279 (24): 4598–4614.
Lavery, Derek N., and Iain J. McEwan. 2008. “Structural Characterization of the Native NH2-Terminal Transactivation Domain of the Human Androgen Receptor: A Collapsed Disordered Conformation Underlies Structural Plasticity and Protein-Induced Folding.” *Biochemistry* 47 (11): 3360–69.
Lee, H., K. H. Mok, R. Muhandiram, K. H. Park, J. E. Suk, D. H. Kim, J. Chang, Y. C. Sung, K. Y. Choi, and K. H. Han. 2000. “Local Structural Elements in the Mostly Unstructured Transcriptional Activation Domain of Human p53.” *The Journal of Biological Chemistry* 275 (38): 29426–32.
Martin, Erik W., and Tanja Mittag. 2018. “Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions.” *Biochemistry* 57 (17): 2478–87.
McNally, James G., Waltraud G. Müller, Dawn Walker, Ronald Wolford, and Gordon L. Hager. 2000. “The Glucocorticoid Receptor: Rapid Exchange with Regulatory Sites in Living Cells.” *Science Signaling* 287 (5456): 1262.
Metallo, Steven J. 2010. “Intrinsically Disordered Proteins Are Potential Drug Targets.” *Current Opinion in Chemical Biology* 14 (4): 481–88.
Mikuni, Shintaro, Mamoru Tamura, and Masataka Kinjo. 2007. “Analysis of Intranuclear Binding Process of Glucocorticoid Receptor Using Fluorescence Correlation Spectroscopy.” *FEBS Letters* 581 (3): 389–93.
Minezaki, Yoshiaki, Keiichi Homma, Akira R. Kinjo, and Ken Nishikawa. 2006. “Human Transcription Factors Contain a High Fraction of Intrinsically Disordered Regions Essential for Transcriptional Regulation.” *Journal of Molecular Biology* 359 (4): 1137–49.
Molliex, Amandine, Jamshid Temirov, Jihun Lee, Maura Coughlin, Anderson P. Kanagaraj, Hong Joo Kim, Tanja Mittag, and J. Paul Taylor. 2015. “Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization.” *Cell* 163 (1): 123–33.
Morales-Navarrete, Hernán, Fabián Segovia-Miranda, Piotr Klukowski, Kirstin Meyer, Hidenori Nonaka, Giovanni Marsico, Mikhail Chernykh, Alexander Kalaidzidis, Marino Zerial, and Yannis Kalaidzidis. 2015. “A Versatile Pipeline for the Multi-Scale Digital Reconstruction and Quantitative Analysis of 3D Tissue Architecture.” *eLife* 4 (December). https://doi.org/10.7554/eLife.11214.

Muñoz, V., and L. Serrano. 1997. “Development of the Multiple Sequence Approximation within the AGADIR Model of Alpha-Helix Formation: Comparison with Zimm-Bragg and Lifson-Roig Formalisms.” *Biopolymers* 41 (5): 495–509.

Murray, David H., Marcus Jahnel, Janelle Lauer, Mario J. Avellaneda, Nicolas Brouilly, Alice Cezanne, Hernán Morales-Navarrete, et al. 2016. “An Endosomal Tether Undergoes an Entropic Collapse to Bring Vesicles Together.” *Nature* 537 (7618): 107–11.

Nadal, Marta, Stefan Prekovic, Nerea Gallastegui, Christine Helsen, Montserrat Abella, Karolina Zielinska, Marina Gay, et al. 2017. “Structure of the Homodimeric Androgen Receptor Ligand-Binding Domain.” *Nature Communications* 8 (February): 14388.

Nagabhushan, M., C. M. Miller, T. P. Pretlow, J. M. Giaconia, N. L. Edgehouse, S. Schwartz, H. J. Kung, et al. 1996. “CWR22: The First Human Prostate Cancer Xenograft with Strongly Androgen-Dependent and Relapsed Strains Both in Vivo and in Soft Agar.” *Cancer Research* 56 (13): 3042–46.

Nair, Sreejith J., Lu Yang, Dario Meluzzi, Soohwan Oh, Feng Yang, Meyer Friedman, Susan Wang, et al. 2019. “Phase Separation of Ligand-Activa Ted Enhancers Licenses Cooperative Chromosomal Enhancer Assembly.” *Nature Structural & Molecular Biology* 26 (March). https://doi.org/10.1038/s41594-019-0190-5.

Needham, M., S. Raines, J. McPheat, C. Stacey, J. Ellston, S. Hoare, and M. Parker. 2000. “Differential Interaction of Steroid Hormone Receptors with LXXLL Motifs in SRC-1a Depends on Residues Flanking the Motif.” *The Journal of Steroid Biochemistry and Molecular Biology* 72 (1-2): 35–46.

Patel, Avinash, Hyun O. Lee, Louise Jawerth, Shovamayee Maharana, Marcus Jahnel, Marco Y. Hein, Stoyno Stoynov, et al. 2015. “A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation.” *Cell* 162 (5): 1066–77.

Permyakov, Sergei E., Ian S. Millett, Sebastian Doniach, Eugene A. Permyakov, and Vladimir N. Uversky. 2003. “Natively Unfolded C-Terminal Domain of Caldesmon Remains Substantially Unstructured After the EffectiveBinding to Calmodulin.” *PROTEINS: Structure, Function, and Genetics* 53: 855–62.

Pratt, W. B., and D. O. Toft. 1997. “Steroid Receptor Interactions with Heat Shock Protein and Immunophilin Chaperones.” *Endocrine Reviews* 18 (3): 306–60.

Presman, Diego M., Sourav Ganguly, R. Louis Schiltz, Thomas A. Johnson, Tatiana S. Karpova, and Gordon L. Hager. 2016. “DNA Binding Triggers Tetramerization of the Glucocorticoid Receptor in Live Cells.” *Proceedings of the National Academy of Sciences of the United States of America*, 2–7.

Presman, Diego M., M. Florencia Ogara, Martín Stortz, Lautaro D. Alvarez, John R. Pooley, R. Louis Schiltz, Lars Grøntved, et al. 2014. “Live Cell Imaging Unveils Multiple Domain Requirements for in Vivo Dimerization of the Glucocorticoid Receptor.” *PLoS Biology* 12 (3): e1001813.

Ptashne, M. 1988. “How Eukaryotic Transcriptional Activators Work.” *Nature* 335 (6192): 683–89.

Radhakrishnan, I., G. C. Pérez-Alvarado, D. Parker, H. J. Dyson, M. R. Montminy, and P. E. Wright. 1997. “Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model for Activator:coactivator Interactions.” *Cell* 91 (6): 741–52.

Rapsomaniki, Maria Anna, Panagiotis Kotsantis, Ioanna-Eleni Symeonidou, Nickolaos-Nikiforos Giakoumakis, Stavros Taraviras, and Zoi Lygerou. 2012. “easyFRAP: An Interactive, Easy-to-Use Tool for Qualitative and Quantitative Analysis of FRAP Data.” *Bioinformatics* 28 (13): 1800–1801.

Rink, Jochen, Eric Ghigo, Yannis Kalaidzidis, and Marino Zerial. 2005. “Rab Conversion as a Mechanism of Progression from Early to Late Endosomes.” *Cell* 122 (5): 735–49.

Royen, Martin E. van, Dennis J. van de Wijngaart, Sónia M. Cunha, Jan Trapman, and Adriaan B. Houtsmuller. 2013. “A Multi-Parameter Imaging Assay Identifies Different Stages of Ligand-Induced Androgen Receptor Activation.” *Cytometry. Part A: The Journal of the International Society for Analytical Cytology* 83 (9): 806–17.

Royen, M. E. van, W. a. van Cappellen, C. de Vos, a. B. Houtsmuller, and J. Trapman. 2012. “Stepwise Androgen Receptor Dimerization.” *Journal of Cell Science* 125 (8): 1970–79.

Sabari, Benjamin R., Alessandra Dall'Agnese, Ann Boija, Isaac A. Klein, Eliot L. Coffey, Krishna Shrinivas, Brian J. Abraham, et al. 2018. "Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control." *Science* 361 (6400). https://doi.org/10.1126/science.aar3958.

Sabari, B. R., Benjamin R. Sabari, Alessandra Dall Agnese, Ann Boija, Isaac A. Klein, Eliot L. Coffey, Brian J. Abraham, et al. 2018. "Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control" 3958 (June): 1–17.

Saitoh, Masayuki, Ryoichi Takayanagi, Kiminobu Goto, Akiyoshi Fukamizu, Arihiro Tomura, Toshihiko Yanase, and Hajime Nawata. 2002. "The Presence of Both the Amino- and Carboxyl-Terminal Domains in the AR Is Essential for the Completion of a Transcriptionally Active Form with Coactivators and Intranuclear Compartmentalization Common to the Steroid Hormone Receptors: A Three-Dimensional." *Molecular Endocrinology* 16 (4): 694–706.

Schaufele, F., and X. Carbonell. 2005. "The Structural Basis of Androgen Receptor Activation: Intramolecular and Intermolecular Amino–carboxy Interactions." *Proceedings of the*. https://www.pnas.org/content/102/28/9802.short.

Senigagliesi, Beatrice, Carlotta Penzo, Luisa Ulloa Severino, Riccardo Maraspini, Sara Petrosino, Hernan Morales-Navarrete, Enrico Pobega, et al. 2019. "The High Mobility Group A1 (HMGA1) Chromatin Architectural Factor Modulates Nuclear Stiffness in Breast Cancer Cells." *International Journal of Molecular Sciences* 20 (11). https://doi.org/10.3390/ijms20112733.

Shaffer, Paul L., Arif Jivan, D. Eric Dollins, Frank Claessens, and Daniel T. Gewirth. 2004. "Structural Basis of Androgen Receptor Binding to Selective Androgen Response Elements." *Proceedings of the National Academy of Sciences of the United States of America* 101 (14): 4758–63.

Sokolova, Ekaterina, Evan Spruijt, Maike M. K. Hansen, Emilien Dubuc, Joost Groen, Venkatachalam Chokkalingam, Aigars Piruska, Hans A. Heus, and Wilhelm T. S. Huck. 2013. "Enhanced Transcription Rates in Membrane-Free Protocells Formed by Coacervation of Cell Lysate." *Proceedings of the National Academy of Sciences of the United States of America* 110 (29): 11692–97.

Sormanni, Pietro, Francesco A. Aprile, and Michele Vendruscolo. 2015. "The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility." *Journal of Molecular Biology* 427 (2): 478–90.

Stenoien, D. L., C. J. Cummings, H. P. Adams, M. G. Mancini, K. Patel, G. N. DeMartino, M. Marcelli, N. L. Weigel, and M. A. Mancini. 1999. "Polyglutamine-Expanded Androgen Receptors Form Aggregates That Sequester Heat Shock Proteins, Proteasome Components and SRC-1, and Are Suppressed by the HDJ-2 Chaperone." *Human Molecular Genetics* 8 (5): 731–41.

Tomura, A., K. Goto, H. Morinaga, M. Nomura, T. Okabe, T. Yanase, R. Takayanagi, and H. Nawata. 2001. "The Subnuclear Three-Dimensional Image Analysis of Androgen Receptor Fused to Green Fluorescence Protein." *The Journal of Biological Chemistry* 276 (30): 28395–401.

Tomura, Arihiro, Kiminobu Goto, Hidetaka Morinaga, Masatoshi Nomura, Taijiro Okabe, Toshihiko Yanase, Ryoichi Takayanagi, and Hajime Nawata. 2001. "The Subnuclear Three-Dimensional Image Analysis of Androgen Receptor Fused to Green Fluorescence Protein." *The Journal of Biological Chemistry* 276 (30): 28395–401.

Tsolis, Antonios C., Nikos C. Papandreou, Vassiliki A. Iconomidou, and Stavros J. Hamodrakas. 2013. "A Consensus Method for the Prediction of 'Aggregation-Prone' Peptides in Globular Proteins." *PloS One* 8 (1): e54175.

Tyagi, Rakesh K., Yan Lavrovsky, Soon C. Ahn, Chung S. Song, Bandana Chatterjee, and Arun K. Roy. 2000. "Dynamics of Intracellular Movement and Nucleocytoplasmic Recycling of the Ligand-Activated Androgen Receptor in Living Cells." *Molecular Endocrinology* 14 (8): 1162–74.

Tyagi, R. K., Y. Lavrovsky, S. C. Ahn, C. S. Song, B. Chatterjee, and A. K. Roy. 2000. “Dynamics of Intracellular Movement and Nucleocytoplasmic Recycling of the Ligand-Activated Androgen Receptor in Living Cells.” *Molecular Endocrinology* 14 (8): 1162–74.

Uversky, Vladimir N., Jie Li, and Anthony L. Fink. 2001. “Evidence for a Partially Folded Intermediate in α-Synuclein Fibril Formation.” *The Journal of Biological Chemistry* 276 (14): 10737–44.

Uversky, Vladimir N., Sergey E. Permyakov, Vasily E. Zagranichny, Igor L. Rodionov, Anthony L. Fink, Alexandra M. Cherskaya, Lyubov A. Wasserman, and Eugene A. Permyakov. 2002. “Effect of Zinc and Temperature on the Conformation of the Gamma Subunit of Retinal Phosphodiesterase: A Natively Unfolded Protein.” *Journal of Proteome Research* 1 (2): 149–59.

Vacic, Vladimir, Vladimir N. Uversky, A. Keith Dunker, and Stefano Lonardi. 2007. “Composition Profiler: A Tool for Discovery and Visualization of Amino Acid Composition Differences.” *BMC Bioinformatics* 8 (June): 211.

Voegel, J. J., M. J. Heine, C. Zechel, P. Chambon, and H. Gronemeyer. 1996. “TIF2, a 160 kDa Transcriptional Mediator for the Ligand-Dependent Activation Function AF-2 of Nuclear Receptors.” *The EMBO Journal* 15 (14): 3667–75.

Wollman, Adam J. M., Sviatlana Shashkova, Erik G. Hedlund, Rosmarie Friemann, Stefan Hohmann, and Mark C. Leake. 2017. “Transcription Factor Clusters Regulate Genes in Eukaryotic Cells.” *eLife* 6: 1–36.

Xu, D., Y. Zhan, Y. Qi, B. Cao, S. Bai, W. Xu, S. S. Gambhir, et al. 2015. “Androgen Receptor Splice Variants Dimerize to Transactivate Target Genes.” *Cancer Research* 75 (17): 3663–71.

5 GENERAL SUMMARY OF RESULTS

We have successfully cloned, expressed and purified the subdomains and the full-length AD constructs of the AR. We characterized the domain structural properties by various biophysical methods including size exclusion chromatography, circular dichroism, dynamic light scattering and nuclear magnetic resonance. This characterization allowed us to understand the general properties of the disordered domain and map the regions of structural (helical) propensity, see Publications 3.1 and 3.4, some previously known to be important for AR function. We mapped the interaction of the Rap74-CTD onto the AR AD to a ^{433}WHTLF^{437} motif, see Publication 3.2. We showed that the motif, completely disordered in solution (<1% helicity), gains helicity upon binding to the Rap74-CTD and accordingly, increase in the motif's helicity increases the strength of the interaction. We also showed the interaction can take place in cells and depends on the phosphorylation status of the S424 located in the N-terminus of the ^{433}WHTLF^{437} motif. Collectively, we provided insights into the structural properties of the ID AD ensemble of AR by applying knowledge of general features of IDPs: sequence determinants and environmental effects on the IDPs (Uversky 2009), presence of molecular recognition features (MoRFs) (see more in the Discussion chapter) (Mohan et al. 2006), folding-upon-binding and conformational selection (Mészáros et al. 2007) and posttranslational modifications (Liu and Huang 2014). We showed the AR can undergo oligomerization and condensation through LLPS in cells, see Publication 3.3. and Manuscript 3.4., and studied the consequences of this phenomenon on the AR AD *in vitro*. LLPS has been known to be often linked to ID (Hyman et al. 2014). We proved the LLPS of the ID AD is driven by hydrophobic interactions resulting in the LCST transition. We identified an aggregation prone motif (^{23}FQNLF^{27}) in the AD and introduced L26P mutation that prevented the fast evolution of the sample. That allowed us to investigate the relationship between the structural properties of the domain and its LLPS propensity. We showed the helicity of the previously characterized activation function motifs of the domain is an important factor in the AD propensity to undergo LLPS and that the helical content of these motifs is increased in the condensed state.

6 DISCUSSION

The AR is a SR of high biomedical importance. This pharmaceutical interest comes mostly from its involvement in prostate cancer onset and progression, as mentioned in Chapter 1.3.3.3. AR has been extensively studied since its first identification attempts in 1969-1973 (when several groups focused on isolation of receptor-proteins for DHT) (Liao, S, Fang 1969; Baulieu, E.E., Jung, I., Blondeau, J.P., Robel 1970; Mainwaring, W.I.P., Mangan 1970; Tveter, K.J., Unhjem, O., Attramadal, A., Aakvaag, A.,Hansson 1970; Wilson, JD, Gloyna 1970; Liao et al. 1973) and since its cDNA cloning (Chang et al. 1988; Lubahn et al. 1988; Trapman et al. 1988; Tilley et al. 1989). “Androgen receptor” appears in the title or abstract of 18387[14] publications in pubmed.gov, with around 1000 publications every year in the past six years, Fig. 12.

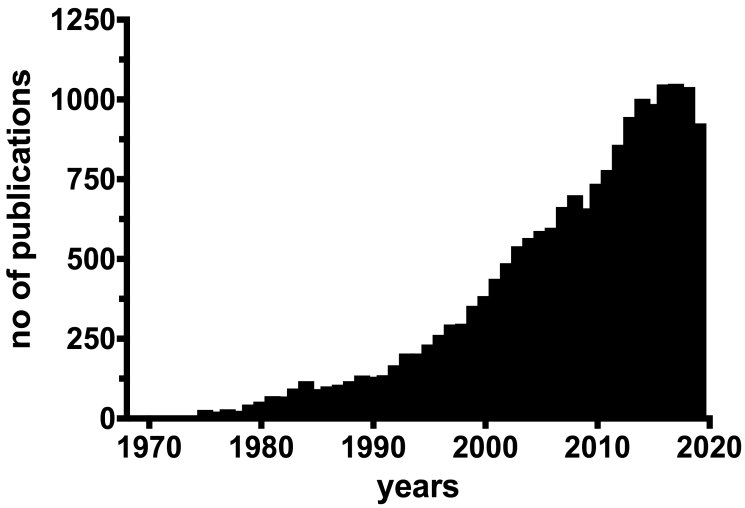


Figure 12. Number of publications with "Androgen receptor" phrase in a title or abstract in pubmed.gov by year.

The pharmaceutical interest in the receptor has boosted also the structural side of the AR research. However, the development in this area has been slowed down by the protein complexity. AR is a 919 AA protein with two globular domains (DBD and LBD), flexible Hinge region and a large ID AD, Chapter 1.3.1. AR structural complexity is also a derivative of its ligand dependence, posttranslational modifications (van der Steen et al. 2013), allosteric inter-domain communication (He et al. 1999; Helsen et al. 2012) and a plethora of binding partners (Chmelar et al. 2007) that can directly change the protein structural state by folding-upon-binding of the ID regions (De Mol et al. 2017), acting like a bridge between the domains (Bai et al. 2005) or possibly allosterically change the protein structure. To try to capture all these aspects of AR structure, studies on the full length (FL) construct of the protein would have to be conducted. Nevertheless, despite the number of years (effort, money and students!) spent on the AR research, the output in publications on the FL construct is not impressive. There are only several groups who reported production of the FL protein in insect or bacterial cells (Xie et al. 1992; Liao and Wilson 2001; Juzumiene et al. 2005; Zhou et al. 2010a) and the structural studies by X-ray

[14] 05.09.2019

crystallography have proven to be unsuccessful (Zhou et al. 2010b). On the other hand, the structures of the separate domains have been solved, the DBD dimer by X-ray crystallography in 2004 (Shaffer et al. 2004), the LBD in 2000 (Pedro M. Matias et al. 2000) and the LBD dimer in 2017 (Nadal et al. 2017) by X-ray crystallography. In 2002 Iain J. McEwan and coworkers took the AR AD under the loop and proved experimentally that the domain (its fragment from 142 to 485 residue) has ID character and adopts more helical structure upon incubation with TFE or the natural osmolyte trimethylamine N-oxide (TMAO) and upon binding to the Rap74-CTD (Reid, Kelly, et al. 2002; Kumar, Betney, Li, E Brad Thompson, et al. 2004). These observations were confirmed in our laboratory and set the base for further research with a focus on the understanding the structural properties of the ID AR AD. The work in the scope of this thesis is the first, to our knowledge, report on the structural properties of the FL AD construct with (almost) single amino acid resolution and an attempt to understand the phase separation properties of the domain.

The purification of the AR FL AD construct, as for many IDPs (Lebendiker and Danieli 2014), had to be performed from inclusion bodies. The stability and reproducible behavior of the AD protein sample was the biggest challenge in the experiments. We have established a protocol for the protein purification in which the final step of exchanging the protein from 8M-urea buffer into the native buffer by size exclusion chromatography has assured protein stability for up to 7 days if kept within a threshold concentration of 100 uM (evolution followed by NMR) for the solution experiments. Optimization of the purification protocol for the AD sample included also a his-tag induced aggregation of the construct, resulting in a crucial step of assuring a maximum cleavage of the tag and a carful check of the tagged protein removal. Although this sounds like a standard procedure, the low Ni affinity of the construct made the removal of uncleaved protein neither efficient nor predictable, resulting in an increased importance of the quality check of the final sample. Unfortunately, expression trials with untagged protein were unsuccessful. Another interesting observation or a very unsettling experimental variable was the nucleation of the phase separation. We observed three types of nucleation: 1) time dependent nucleation, 2) air surface nucleation and 3) unknown source nucleation. Point 1 was addressed with a very precise time preparation of the sample. Point 2 is most probably a consequence of increased protein concentration in the air surface (Lu et al. 1999) that was practically tackled by imaging the center of the droplet/well of the sample and multiple repetitions of the experiments. Point 3 was addressed by a careful cleaning of the glass slides before samples imaging and 20 min centrifugation at 21130 rcf at 4 °C for the cloud point temperature measurements, to assure removal of any heteronuclei. Such preparation of protein for the experiments assured reproducible results.

It has been proposed that IDPs, including the "acid blobs and negative noodles" of TFs (Paul B. Sigler 1988), can perform their interaction hub function thanks to the MoRFs, short ID sequences that have propensity to fold upon binding to the specific partner. MoRFs could exist as completely disordered sequences and adapt a structure only upon contact with the binding partner or they can be in ensamble of conformations with pre-existing secondary structure which population increases upon binding to the partner (Mohan, Oldfield, Radivojac, Vacic, Cortese, A.K. Dunker, et al. 2006). For the ADs of SRs the second scenario seems to be more frequent. ER (Métivier et al. 2000; Peng et al. 2019), PR (Kumar et al. 2013), MR (Kumar et al. 2013), GR (Kumar, Volk, et al. 2004; Kumar et al. 2007) and here, AR, have been shown to contain MoRFs of some helical propensity that can become more helical upon biding to their partners. We have shown a substantial helical

propensity of the AR AD regions that have been previously recognized as functional: a known helix forming sequence, ^{23}FQNLF^{27}, that is the main N/C interaction motif, the ^{179}LKDIL^{183} motif that plays a role in the N/C interaction (Alen et al. 1999), the core Tau-1 with the ^{183}LSEASTMQLL^{192} sequence, a binding epitope for TAB2 (Zhu et al. 2006), the residues from 230 to 240 that overlap with a biding epitope for CHIP (He, Bai, Andrew T. Hnat, et al. 2004), a polyQ containing region and two helical potential MoRFs in Tau-5 to which binding partners are yet to be discovered. We also showed that the ^{433}WHTLF^{437} motif, important for protein activity in androgen-independent scenario (Scott M. Dehm et al. 2007), has no helical structure (<1%) but can adopt it upon binding to the Rap74-CTD (domain of TFIIF). This observation confirmed previous work of J. McEwan and coworkers, where they showed increase in the AD helicity upon binding to the Rap74-CTD (Kumar, Betney, Li, E Brad Thompson, et al. 2004) and provides molecular details of the interaction. Interactions of IDPs are often modulated by post-translational modifications (Bah and Forman-Kay 2016; Dahal et al. 2017). That is also a case for the Rap74-CTD biding to the AR AD, where a phosphorylation on the S424 of the AR AD increased its binding affinity *in vitro* and the same mutation abolished the interaction in cells, as shown by the PLA assay, and significantly lowered the AR activity, as shown by the luciferase assay. Unfortunately, we were not able to show the interaction between the proteins on the endogenous levels. Further work is also needed to investigate the importance of the Rap74-CTD binding to the ^{433}WHTLF^{437} motif in the androgen-independent context.

Some of the motifs identified as partially folded appeared to be also important for the AR AD oligomerization and phase separation *in vitro*. The main feature driving the phase separation of the domain was its hydrophobicity and presumably also its aromatic character. The resulting LCST behavior has been reported only for few proteins, including elastin-like polypeptide (ELP) (Muiznieks et al. 2018) and yeast Pab1 (Riback et al. 2017) (not taking into account polymers phase separation, where LCST behavior is common). Both examples are proteins that only in its initial state form liquid-like assemblies, with Pab1 proceeding into hydrogel structure under persistent stress conditions and elastin forming its final state of fibrillar, supramolecular structures of extracellular matrix. In the case of the AR AD we also observed gel-like structure formation, that was primarily a consequence of the ^{23}FQNLF^{27} motif aggregation. However, experiments with multiple rounds of phase separation induced by heating and subsequent dissolution by cooling down with the use of the AD L26P mutant showed that a small fraction of the protein sample was consistently forming irreversible droplets. Additionally, long incubation of the droplets resulted in their lower dynamics (measured by FRAP) and prevented droplets characterization by NMR, probably due to formation of more structured assemblies that could be stabilized by disulfide bridges (the construct has 11 cysteine residues). Whether it is an artifact of experiments *in vitro* or LLPS is an important intermediate step in AR aggregation in cells (although, here explored only with the limited construct of AD alone) important in a context of SBMA remains to be answered.

Another aspect of the LLPS of the AR AD was its structural gain. We showed that helical propensity of the domain motifs is linked to its phase separation propensity suggesting preferential helix formation in the droplet state. Similarly, Fawzi and coworkers demonstrated that the helix-helix inter-molecular contacts between the ID of the RNA-binding protein TDP-43 are modulators of its phase separation (Domain et al. 2016; Conicella et al. 2019). We hypothesize that the low populated helical motifs are stabilized by the helix-helix interactions in the droplet state of TF ADs and can be replaced by specific helix-co-activator interactions in the transcription complex via conformational selection.

Moreover, the helical gain upon condensation would allow for cooperative folding of otherwise independent MoRFs allowing for an efficient transcriptional complex formation. Unfortunately, we do not provide a direct evidence of increased helix population in the droplet state for which NMR would be a method of choice. The few available reports on protein conformation in the LLPS droplets by NMR have shown so far examples of a maintained disorder in the droplets, as in the case of FUS or Elastin (Ii et al. 2015; Reichheld et al. 2017) or increase of β-structure in the case of Tau protein (Ambadipudi et al. 2017).

Experiments on AR oligomerization in cells provided evidence for the AD-LBD dependence for higher oligomer formation of the protein. The AR oligomers can be an intermediate form between a monomer and its phase separated state as shown for the HP1α, a protein important for heterochromatin formation (Larson et al. 2017; Strom et al. 2017).

We showed AR condensation by LLPS in cell cytoplasm and nucleus in the case of SPOP-AR demixing. Nevertheless, all the experiments were performed with an overexpressed protein leaving the physiological relevance of the phenomenon not addressed in the scope of this thesis. The challenges in providing a compelling evidence for LLPS mediated transcription is a small size of the transcription clusters already reported (below 1 um diameter) and their association with chromatin (FRAP would be influenced by the DNA binding of the studied components and indirect DNA binding by tethering to other DNA-bound components). The currently accepted way of proving these transcriptional clusters are formed by LLPS is their sensitivity to 1,6-hexanediol (Sabari et al. 2018; Nair et al. 2019), a compound originally shown to perturb FG repeat interactions between nucleoporins in nuclear pores (Ribbeck and Gorlich 2002) and currently thought to perturb weak hydrophobic interactions between proteins and proteins and RNA in their LLPS state (Kroschwald et al. 2017). Nevertheless, the exact mechanism of hexanediol dissolution of the LLPS droplets is not known leaving the evidence to be looked at with precaution.

The research gathered in this thesis is just a small contribution to the AR knowledge. However, I hope it opened up a new phase separation prism which we should at least consider looking through.

7 CONCLUSIONS

1. The AR AD is intrinsically disordered with sequence motifs of low to high helical propensity.

2. AR interacts with Rap74-CTD of TFIIF through the ^{433}WHTLF^{437} motif in the AR AD and this interaction depends on the phosphorylation status of serine 424 in AR.

3. The AR AD ^{433}WHTLF^{437} motif undergoes folding-upon-binding to Rap74-CTD. The interaction is weak but can be improved by the phosphorylation of S424 in the N-terminus of the ^{433}WHTLF^{437} motif and by increase in its helicity.

4. AR and SPOP can undergo liquid-liquid demixing in cells and co-phase separate *in vitro*.

5. AR can undergo oligomerization and liquid-liquid phase separation in cells after activation by androgens and both processes depend on the presence of the AD and LBD.

6. The AR AD oligomerizes and undergoes LCST type phase separation *in vitro*. Both processes are driven by hydrophobic inter-molecular interactions.

7. The main hydrophobic inter-molecular interactions of the AR AD are between partially helical motifs of the domain (activation function motifs) and the sequence after the polyG region.

8. ^{23}FQNLF^{27} sequence is highly aggregation prone and its aggregation causes fast gelation of the AR AD droplets.

9. The helicity of the AR AD activation function motifs increases in the condensed phase.

8 BIBLIOGRAPHY

Adachi H, Katsuno M, Minamiyama M, Waza M, Sang C, Nakagomi Y, Kobayashi Y, Tanaka F, Doyu M, Inukai A, et al. 2005. Widespread nuclear and cytoplasmic accumulation of mutant androgen receptor in SBMA patients. Brain. 128(3):659–670. doi:10.1093/brain/awh381.

Alen P, Claessens F, Verhoeven G, Rombauts W, Peeters B. 1999. The androgen receptor amino-terminal domain plays a key role in p160 coactivator-stimulated gene transcription. Mol Cell Biol. 19(9):6085–97. doi:10.1128/mcb.19.9.6085. [accessed 2019 Sep 7]. http://www.ncbi.nlm.nih.gov/pubmed/10454556.

Ambadipudi S, Biernat J, Riedel D, Mandelkow E, Zweckstetter M. 2017. Liquid–liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. Nat Commun. 8(1):275. doi:10.1038/s41467-017-00480-0.

An J, Wang C, Deng Y, Yu L, Huang H. 2014. Article Destruction of Full-Length Androgen Receptor by Wild-Type SPOP , but Not Prostate-Cancer-Associated Mutants. CellReports. 6(4):657–669. doi:10.1016/j.celrep.2014.01.013.

Andersen RJ, Mawji NR, Wang J, Wang G, Haile S, Myung JK, Watt K, Tam T, Yang YC, Bañuelos CA, et al. 2010. Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor. Cancer Cell. 17(6):535–546. doi:10.1016/j.ccr.2010.04.027.

Antonarakis ES, Chandhasin C, Osbourne E, Luo J, Sadar MD, Perabo F. 2016. Targeting the N-Terminal Domain of the Androgen Receptor: A New Approach for the Treatment of Advanced Prostate Cancer. Oncologist. 21:1427–1435. doi:10.1634/theoncologist.2016-0161.

Armstrong AJ, Halabi S, Luo J, Nanus DM, Giannakakou P, Szmulewitz RZ, Danila DC, Healy P, Anand M, Rothwell CJ, et al. 2019. Prospective Multicenter Validation of Androgen Receptor Splice Variant 7 and Hormone Therapy Resistance in High-Risk Castration-Resistant Prostate Cancer: The PROPHECY Study. J Clin Oncol. 37(13):1120–1129. doi:10.1200/JCO.18.01731.

Arora VK, Schenkein E, Murali R, Subudhi SK, Wongvipat J, Balbas MD, Shah N, Cai L, Efstathiou E, Logothetis C, et al. 2013. Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. Cell. 155(6):1309–1322. doi:10.1016/j.cell.2013.11.012.

Attard G, Reid AHM, A'Hern R, Parker C, Oommen NB, Folkerd E, Messiou C, Molife LR, Maier G, Thompson E, et al. 2009. Selective inhibition of CYP17 with abiraterone acetate is highly active in the treatment of castration-resistant prostate cancer. J Clin Oncol. 27(23):3742–3748. doi:10.1200/JCO.2008.20.0642.

Aviles F, Chapman G, Kneale G, Crane-Robinson C, Bradbur y E. 1978. The Conformation of Histone H5. Eur J Biochem. 371(88):363–371.

Azad AA, Volik S V., Wyatt AW, Haegert A, Le Bihan S, Bell RH, Anderson SA, McConeghy B, Shukin R, Bazov J, et al. 2015. Androgen receptor gene aberrations in circulating cell-free DNA: Biomarkers of therapeutic resistance in castration-resistant prostate cancer. Clin Cancer Res. 21(10):2315–2324. doi:10.1158/1078-0432.CCR-14-2666.

Bah A, Forman-Kay JD. 2016. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. J Biol Chem. 291(13):6696–6705. doi:10.1074/jbc.R115.695056.

Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, Zhao C, Sonenberg N, Kay LE, Forman-Kay JD. 2014. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature. advance on(2):110–7. doi:10.1038/nature13999.

Bai S, He B, Wilson EM. 2005. Melanoma antigen gene protein MAGE-11 regulates androgen receptor function by modulating the interdomain interaction. Mol Cell Biol. 25(4):1238–57. doi:10.1128/MCB.25.4.1238-1257.2005.

Bain DL, Heneghan AF, Connaghan-Jones KD, Miura MT. 2007. Nuclear Receptor Structure: Implications for Function. Annu Rev Physiol. 69(1):201–220. doi:10.1146/annurev.physiol.69.031905.160308.

Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, White TA, Stojanov P, Van Allen E, Stransky N, et al. 2012. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet. 44(6):685–689. doi:10.1038/ng.2279.

BIBLIOGRAPHY

Baulieu, E.E., Jung, I., Blondeau, J.P., Robel P. 1970. Advances in the Biosciences 7: Schering Workshop on Steroid Hormone "Receptors."

Beltran H et al. 2014. Targeted NGS of Advanced PCa. NIH Public Access. 63(5):920–926. doi:10.1016/j.eururo.2012.08.053.Targeted.

Beltran H, Rickman DS, Park K, Chae SS, Sboner A, Macdonald TY, Wang Yuwei, Sheikh KL, Terry S, Tagawa ST, et al. 2011. Molecular Characterization of Neuroendocrine Prostate Cancer and Identification of New Drug Targets. Cancer Discov. 1(6):487–495. doi:10.1158/2159-8290.CD-11-0130.

Bengal E, Flores O, Krauskopf A, Reinberg D, Aloni Y. 2015. Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. Mol Cell Biol. 11(3):1195–1206. doi:10.1128/mcb.11.3.1195.

Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, et al. 2011. The genomic complexity of primary human prostate cancer. Nature. 470(7333):214–220. doi:10.1038/nature09744.

Bernecky C, Taatjes DJ. 2012. Activator-mediator binding stabilizes RNA polymerase II orientation within the human mediator-RNA polymerase II-TFIIF assembly. J Mol Biol. 417(5):387–394. doi:10.1016/j.jmb.2012.02.014.

Betney R, McEwan IJ. 2003. Role of conserved hydrophobic amino acids in androgen receptor AF-1 function. J Mol Endocrinol. 31(3):427–439. doi:10.1677/jme.0.0310427.

Bingham PM, Scott M, Wang S, Mcphaup MJ, Wilson EM, Garbern JY, Merrf DE, Fischbeck KH. 1995. trinucleotide repeat in the transgenic mice. 9(february).

Boehning M, Dugast-Darzacq C, Rankovic M, Hansen AS, Yu T, Marie-Nelly H, McSwiggen DT, Kokic G, Dailey GM, Cramer P, et al. 2018. RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat Struct Mol Biol. 25(9):833–840. doi:10.1038/s41594-018-0112-y.

Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio A V., Li CH, Shrinivas K, Manteiga JC, Hannett NM, et al. 2018. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell.:1–14. doi:10.1016/j.cell.2018.10.042.

Bourguet W, Germain P, Gronemeyer H, Ad A-. 2000. Nuclear receptor ligand-binding domains : three-dimensional structures , molecular interactions and pharmacological implications. 21(October):1876–1882.

Boysen G, Barbieri CE, Prandi D, Blattner M, Chae S-S, Dahija A, Nataraj S, Huang D, Marotz C, Xu L, et al. 2015. SPOP mutation leads to genomic instability in prostate cancer. Elife. 4:1–18. doi:10.7554/elife.09207.

Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Jülicher F, Hyman AA, In. 2009. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science (80-). 324:1729–1732.

Brélivet Y, Rochel N, Moras D. 2012. Structural analysis of nuclear receptors: From isolated domains to integral proteins. Mol Cell Endocrinol. 348:466–473. doi:10.1016/j.mce.2011.08.015.

Brown CJ, Goss SJ, Lubahn TDB, Joseph DR, Wilson EM, French FS, Willard HF. 1989. Androgen Receptor Locus on the Human X Chromosome: Regional Localization to Xql 1-12 and Description of a DNA Polymorphism. Receptor.:264–269.

Burton ZF, Killeen M, Sopta M, Ortolan LG, Greenblatt J. 1988. RAP30/74: a general initiation factor that binds to RNA polymerase II. Mol Cell Biol. 8(4):1602–1613. doi:10.1128/mcb.8.4.1602.

Buzón V, Carbó LR, Estruch SB, Fletterick RJ, Estébanez-Perpiñ E. 2012. A conserved surface on the ligand binding domain of nuclear receptors for allosteric control. Mol Cell Endocrinol. 348(2):394–402. doi:10.1016/j.mce.2011.08.012.

Callewaert L, Van Tilborgh N, Claessens F. 2006. Interplay between two hormone-independent activation domains in the androgen receptor. Cancer Res. 66(1):543–553. doi:10.1158/0008-5472.CAN-05-2389.

Cato AC, Henderson D, Ponta H. 1987. The hormone response element of the mouse mammary tumour virus DNA mediates the progestin and androgen induction of transcription in the proviral long terminal repeat region. EMBO J. 6(2):363–8.

Centenera MM, Harris JM, Tilley WD, Butler LM, Roma D, Cancer M, Medicine D. 2008. The Contribution

of Different Androgen Receptor Domains to Receptor Dimerization and Signaling. 22(11):2373–2382. doi:10.1210/me.2008-0017.

Chan M, Chang MC, González R, Lategan B, Del Barco E, Vera-Badillo F, Quesada P, Goldstein R, Cruz I, Ocana A, et al. 2015. Outcomes of estrogen receptor negative and progesterone receptor positive breast cancer. PLoS One. 10(7):1–13. doi:10.1371/journal.pone.0132449.

Chan SC, Dehm SM. 2014. Constitutive activity of the androgen receptor. 1st ed. Elsevier Inc.

Chang C, Kokontis J, Liao S. 1988. Molecular cloning of human and rat complementary DNA encoding androgen receptors. Science (80-). 240(4850):324–326. doi:10.1126/science.3353726.

Chen EJ, Sowalsky AG, Gao S, Cai C, Voznesensky O, Schaefer R, Loda M, True LD, Ye H, Troncoso P, et al. 2015. Abiraterone treatment in castration-resistant prostate cancer selects for progesterone responsive mutant androgen receptors. Clin Cancer Res. 21(6):1273–1280. doi:10.1158/1078-0432.CCR-14-1220.

Chen W, Roeder RG. 2011. Mediator-dependent nuclear receptor function. Semin Cell Dev Biol. 22(7):749–758. doi:10.1016/j.semcdb.2011.07.026.

Chen Z, Zhang C, Wu D, Chen H, Rorick A, Zhang X, Wang Q. 2011. Phospho-MED1-enhanced UBE2C locus looping drives castration-resistant prostate cancer growth. EMBO J. 30(12):2405–2419. doi:10.1038/emboj.2011.154.

Chen ZA, Jawhari A, Fischer L, Buchen C, Tahir S, Kamenski T, Rasmussen M, Lariviere L, Bukowski-Wills JC, Nilges M, et al. 2010. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J. 29(4):717–726. doi:10.1038/emboj.2009.401.

Cheng HS, Lee JXT, Wahli W, Tan NS. 2019. Exploiting vulnerabilities of cancer by targeting nuclear receptors of stromal cells in tumor microenvironment. Mol Cancer. 18(1):51. doi:10.1186/s12943-019-0971-9.

Chin AF, Toptygin D, Elam WA, Schrank TP, Hilser VJ. 2016. Phosphorylation Increases Persistence Length and End-to-End Distance of a Segment of Tau Protein. Biophys J. 110(2):362–371. doi:10.1016/j.bpj.2015.12.013.

Chmelar R, Buchanan G, Need EF, Tilley W, Greenberg NM. 2007. Androgen receptor coregulators and their involvement in the development and progression of prostate cancer. Int J Cancer. 120(4):719–733. doi:10.1002/ijc.22365.

Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, et al. 2018. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science (80-). 361(6400). doi:10.1126/science.aar2555.

Choy WY, Forman-Kay JD. 2001. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. J Mol Biol. 308(5):1011–1032. doi:10.1006/jmbi.2001.4750.

Chung WH, Craighead JL, Chang WH, Ezeokonkwo C, Bareket-Samish A, Kornberg RD, Asturias FJ. 2003. RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. Mol Cell. 12(4):1003–1013. doi:10.1016/S1097-2765(03)00387-3.

Claessens F, Denayer S, Van Tilborgh N, Kerkhofs S, Helsen C, Haelens A. 2008. Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. Nucl Recept Signal. 6:e008. doi:10.1621/nrs.06008.

Cleutjens KBJM, Van Eekelen CCEM, Van Der Korput HAGM, Brinkmann AO, Trapman J. 1996. Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specific antigen promoter. J Biol Chem. 271(11):6379–6388. doi:10.1074/jbc.271.11.6379.

Cleutjens KBJM, van der Korput HAGM, van Eekelen CCEM, van Rooij HCJ, Faber PW, Trapman J. 2014. An Androgen Response Element in a Far Upstream Enhancer Region Is Essential for High, Androgen-Regulated Activity of the Prostate-Specific Antigen Promoter. Mol Endocrinol. 11(2):148–161. doi:10.1210/mend.11.2.9883.

Clever, U., Karlson P. 1960. Induction of puff changes in the salivary gland chromosomes of Chironomus tentans by ecdysone. Exp Cell Res. 20:623–6.

Conicella AE, Dignon GL, Zerze GH, Schmidt HB, D'Ordine AM, Kim YC, Rohatgi R, Ayala YM, Mittal J, Fawzi NL. 2019 May 18. TDP-43 α-helical structure tunes liquid-liquid phase separation and function. bioRxiv.:640615. doi:10.1101/640615.

BIBLIOGRAPHY

Crnalic S, Hörnberg E, Wikström P, Lerner UH, Tieva Å, Svensson O, Widmark A, Bergh A. 2010. Nuclear androgen receptor staining in bone metastases is related to a poor outcome in prostate cancer patients. Endocr Relat Cancer. 17(4):885–895. doi:10.1677/ERC-10-0059.

Dahal L, Shammas SL, Clarke J. 2017. Phosphorylation of the IDP KID Modulates Affinity for KIX by Increasing the Lifetime of the Complex. Biophys J. 113(12):2706–2712. doi:10.1016/j.bpj.2017.10.015.

Dalal K, Ban F, Li H, Morin H, Roshan-Moniri M, Tam KJ, Shepherd A, Sharma A, Peacock J, Carlson ML, et al. 2018. Selectively targeting the dimerization interface of human androgen receptor with small-molecules to treat castration-resistant prostate cancer. Cancer Lett. 437(August):35–43. doi:10.1016/j.canlet.2018.08.016.

Dehm Scott M, Regan KM, Schmidt LJ, Tindall DJ. 2007. Selective Role of an NH 2 -Terminal WxxLF Motif for Aberrant Androgen Receptor Activation in Androgen Depletion – Independent Prostate Cancer Cells. (20):10067–10078. doi:10.1158/0008-5472.CAN-07-1267.

Dehm Scott M., Regan KM, Schmidt LJ, Tindall DJ. 2007. Selective role of an NH2-terminal WxxLF motif for aberrant androgen receptor activation in androgen depletion-independent prostate cancer cells. Cancer Res. 67(20):10067–10077. doi:10.1158/0008-5472.CAN-07-1267.

Dhiman VK, Bolt MJ, White KP. 2018. Nuclear receptors in cancer - Uncovering new and evolving roles through genomic analysis. Nat Rev Genet. 19(3):160–174. doi:10.1038/nrg.2017.102.

Ding D, Xu L, Menon M, Reddy GPV, Barrack ER. 2005. Effect of GGC (Glycine) Repeat Length Polymorphism in the Human Androgen Receptor on Androgen Action. 139(April 2004):133–139. doi:10.1002/pros.20128.

Domain C, Conicella AE, Mittal J, Fawzi NL, Conicella AE, Zerze H, Mittal J, Fawzi NL. 2016. Article ALS Mutations Disrupt Phase Separation Mediated by a -Helical Structure in the TDP-43 Low-Complexity Article ALS Mutations Disrupt Phase Separation Mediated by a -Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. :1537–1549. doi:10.1016/j.str.2016.07.007.

Dubbink HJ, Hersmus R, Verma CS, Korput HAGMVANDER, Berrevoets CORA, Tol JVAN, K HAGM, T J, M ACJZ. 2004. Distinct Recognition Modes of FXXLF and LXXLL Motifs by the Androgen Receptor. 18(9):2132–2150. doi:10.1210/me.2003-0375.

Dunker a. K, Bondos SE, Huang F, Oldfield CJ. 2015. Intrinsically disordered proteins and multicellular organisms. Semin Cell Dev Biol. 37:44–55. doi:10.1016/j.semcdb.2014.09.025.

Dunker AK, Bondos SE, Huang F, Oldfield CJ. 2015. Intrinsically disordered proteins and multicellular organisms. Semin Cell Dev Biol. 37:44–55. doi:10.1016/j.semcdb.2014.09.025.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, et al. 2001. Intrinsically disordered protein. 3263(00):26–59.

Dunker AK, Obradovic Z. 2001. The protein trinity - Linking function and disorder. Nat Biotechnol. 19(9):805–806. doi:10.1038/nbt0901-805.

Dunker AK, Uversky VN. 2010. Drugs for “protein clouds”: Targeting intrinsically disordered transcription factors. Curr Opin Pharmacol. 10(6):782–788. doi:10.1016/j.coph.2010.09.005.

Dvir A, Conaway JW, Conaway RC. 2001. Mechanism of transcription initiation and promoter escape by RNA polymerase II. Curr Opin Genet Dev. 11(2):209–214. doi:10.1016/S0959-437X(00)00181-7.

Eftekharzadeh B, Banduseela VC, Chiesa G, Martínez-Cristóbal P, Rauch JN, Nath SR, Schwarz DMC, Shao H, Marin-Argany M, Di Sanza C, et al. 2019. Hsp70 and Hsp40 inhibit an inter-domain interaction necessary for transcriptional activity in the androgen receptor. Nat Commun. 10(1):3562. doi:10.1038/s41467-019-11594-y.

Elmendorf BJ, Shilatifard A, Yan Q, Conaway JW, Conaway RC. 2001. Transcription Factors TFIIF, ELL, and Elongin Negatively Regulate SII-induced Nascent Transcript Cleavage by Non-arrested RNA Polymerase II Elongation Intermediates. J Biol Chem. 276(25):23109–23114. doi:10.1074/jbc.M101445200.

Epstein JI, Amin MB, Beltran H, Lotan TL, Mosquera J-M, Reuter VE, Robinson BD. 2017. Proposed Morphologic Classification of Prostate Cancer With Neuroendocrine Differentiation. J Med Biol Eng. 37(2):171–180. doi:10.1007/s40846-016-0196-y.

Erkizan H V., Kong Y, Merchant M, Schlottmann S, Barber-Rotenberg JS, Yuan L, Abaan OD, Chou TH, Dakshanamurthy S, Brown ML, et al. 2009. A small molecule blocking oncogenic protein EWS-FLI1

interaction with RNA helicase A inhibits growth of Ewing's sarcoma. Nat Med. 15(7):750–756. doi:10.1038/nm.1983.

Errington WJ, Khan MQ, Bueler SA, Rubinstein JL, Chakrabartty A, Privé GG. 2012. Adaptor protein self-assembly drives the control of a cullin-RING ubiquitin ligase. Structure. 20(7):1141–1153. doi:10.1016/j.str.2012.04.009.

Escobedo A, Topal B, Kunze MBA, Aranda J, Chiesa G, Mungianu D, Bernardo-Seisdedos G, Eftekharzadeh B, Gairí M, Pierattelli R, et al. 2019. Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. Nat Commun. 10(1). doi:10.1038/s41467-019-09923-2.

Evans RM. 1988. The Steroid and Thyroid Hormone Receptor Superfamily. 240(4854):889–895.

Fei J, Jadaliha M, Harmon TS, Li ITS, Hua B, Hao Q, Holehouse AS, Reyer M, Sun Q, Freier SM, et al. 2017. Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. J Cell Sci. 130(24):4180–4192. doi:10.1242/jcs.206854.

Feng H, Jenkins LMM, Durell SR, Hayashi R, Mazur SJ, Cherry S, Tropea JE, Miller M, Wlodawer A, Appella E, et al. 2009. Structural Basis for p300 Taz2-p53 TAD1 Binding and Modulation by Phosphorylation. Structure. 17(2):202–210. doi:10.1016/j.str.2008.12.009.

Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki RW, Pappu RV, Brangwynne CP. 2016. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. Cell. 165(7):1686–1697. doi:10.1016/j.cell.2016.04.047.

Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, Gavin A, Visser O, Bray F. 2018. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer. 103:356–387. doi:10.1016/j.ejca.2018.07.005.

Fischbeck KH, Spada AR La, Wilson EM, Lubahn DB, Harding AE. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature. 352(6330):77–79.

Fischer E. 1894. Einfluss der configuration auf die wirkung der enzyme. Berichte der Dtsch Chem Gesellschaft. 27(3):2985–2993.

Fokkens M, Schrader T, Klärner FG. 2005. A molecular tweezer for lysine and arginine. J Am Chem Soc. 127(41):14415–14421. doi:10.1021/ja052806a.

Fratta P, Collins T, Pemble S, Nethisinghe S, Devoy A, Giunti P, Sweeney MG, Hanna MG, Fisher EMC. 2014. Sequencing analysis of the spinal bulbar muscular atrophy CAG expansion reveals absence of repeat interruptions. Neurobiol Aging. 35(2):443.e1-443.e3. doi:10.1016/j.neurobiolaging.2013.07.015.

Fu M, Wang C, Reutens AT, Wang J, Angeletti RH, Siconolfi-Baez L, Ogryzko V, Avantaggiati M-L, Pestell RG. 2000. p300 and P/CAF Acetylate the Androgen Receptor at sites governing hormone-dependent transactivation. J Biol Chem. 275(27):20853–20860. doi:10.1074/jbc.M000660200.

Fujii S, Kagechika H. 2019. Androgen receptor modulators: a review of recent patents and reports (2012-2018). Expert Opin Ther Pat. 29(6):1–15. doi:10.1080/13543776.2019.1618831.

Funk JD, Nedialkov YA, Xu D, Burton ZF. 2002. A key role for the α1 helix of human RAP74 in the initiation and elongation of RNA chains. J Biol Chem. 277(49):46998–47003. doi:10.1074/jbc.M206249200.

Gaiser F, Tan S, Richmond TJ. 2000. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. J Mol Biol. 302(5):1119–1127. doi:10.1006/jmbi.2000.4110.

Van Geersdaele LK, Stead MA, Harrison CM, Carr SB, Close HJ, Rosbrook GO, Connell SD, Wright SC. 2013. Structural basis of high-order oligomerization of the cullin-3 adaptor SPOP. Acta Crystallogr Sect D Biol Crystallogr. 69(9):1677–1684. doi:10.1107/S0907444913012687.

Gelmann EP. 2002. Molecular biology of the androgen receptor. J Clin Oncol. 20(13):3001–3015. doi:10.1200/JCO.2002.10.018.

Geng C, He B, Xu L, Barbieri CE, Kumar V, Anne S. 2013. Prostate cancer-associated mutations in speckle-type POZ protein (SPOP) regulate steroid receptor coactivator 3 protein turnover. 110(17). doi:10.1073/pnas.1304502110.

Ghazy MA, Brodie SA, Ammerman ML, Ziegler LM, Ponticelli AS. 2004. Amino Acid Substitutions in Yeast TFIIF Confer Upstream Shifts in Transcription Initiation and Altered Interaction with RNA Polymerase II. Mol Cell Biol. 24(24):10975–10985. doi:10.1128/mcb.24.24.10975-10985.2004.

BIBLIOGRAPHY

Gibson BA, Doolittle LK, Jensen LE, Gamarra N, Redding S, Rosen MK. 2019. Organization and Regulation of Chromatin by Liquid-Liquid Phase Separation. bioRxiv.:523662. doi:10.1101/523662.

Giovannucci E, Stampfer MJ, Brown M, Brufsky A, Talcott J, Hennekens CH, Kantoff PW, Stampfer J, Krithivas K, Brown M, et al. 1997. The CAG repeat within the androgen receptor gene and its. 94(6):2106–2110.

Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, Brenner JC, et al. 2012. The mutational landscape of lethal castration-resistant prostate cancer. Nature. 487(7406):239–243. doi:10.1038/nature11125.

Groft CM, Uljon SN, Wang R, Werner MH. 2002. Structural homology between the Rap30 DNA-binding domain and linker histone H5: Implications for preinitiation complex assembly. Proc Natl Acad Sci. 95(16):9117–9122. doi:10.1073/pnas.95.16.9117.

Guillén-Boixet J, Buzon V, Salvatella X, Méndez R. 2016. CPEB4 is regulated during cell cycle by ERK2/Cdk1-mediated phosphorylation and its assembly into liquid-like droplets. Elife. 5(NOVEMBER2016):1–26. doi:10.7554/eLife.19298.

Haelens A, Tanner T, Denayer S, Callewaert L, Claessens F. 2007. The hinge region regulates DNA binding, nuclear translocation, and transactivation of the androgen receptor. Cancer Res. 67(9):4514–4523. doi:10.1158/0008-5472.CAN-06-1701.

Hager GL, McNally JG, Misteli T. 2009. Transcription Dynamics. Mol Cell. 35(6):741–753. doi:10.1016/j.molcel.2009.09.005.

Hammoudeh DI, Follis AV, Prochownik E V., Metallo SJ. 2009. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. J Am Chem Soc. 131(21):7390–7401. doi:10.1021/ja900616b.

Hara T, Miyazaki J, Araki H, Yamaoka M, Kanzaki N, Kusaka M, Miyamoto M. 2003. Novel mutations of androgen receptor: a possible mechanism of bicalutamide withdrawal syndrome. Cancer Res. 63(1):149–53.

He B, Bai S, Hnat Andrew T, Kalman RI, Minges JT, Patterson C, Wilson EM. 2004. An Androgen Receptor NH 2 -terminal Conserved Motif Interacts with the COOH Terminus of the Hsp70-interacting Protein (CHIP)*. 279(29):30643–30653. doi:10.1074/jbc.M403117200.

He B, Bai S, Hnat Andrew T., Kalman RI, Minges JT, Patterson C, Wilson EM. 2004. An androgen receptor NH2-terminal conserved motif interacts with the COOH terminus of the Hsp70-interacting protein (CHIP). J Biol Chem. 279(29):30643–30653. doi:10.1074/jbc.M403117200.

He B, Gampe RT, Kole AJ, Hnat AT, Stanley TB, An G, Stewart EL, Kalman RI, Minges JT, Wilson EM. 2004. Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. Mol Cell. 16(3):425–438. doi:10.1016/j.molcel.2004.09.036.

He B, Kemppainen J a., Wilson EM. 2000. FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. J Biol Chem. 275(30):22986–22994. doi:10.1074/jbc.M002807200.

He B, Kemppainen JA, Voegel JJ, Gronemeyer H, Wilson EM. 1999. Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH2-terminal domain. J Biol Chem. 274(52):37219–37225. doi:10.1074/jbc.274.52.37219.

He Y, Fang J, Taatjes DJ, Nogales E. 2013. Structural visualization of key steps in human transcription initiation. Nature. 495(7442):481–486. doi:10.1038/nature11991.

Heemers H V., Tindall DJ. 2007a. Androgen receptor (AR) coregulators: A diversity of functions converging on and regulating the AR transcriptional complex. Endocr Rev. 28(7). doi:10.1210/er.2007-0019.

Heemers H V., Tindall DJ. 2007b. Androgen receptor (AR) coregulators: A diversity of functions converging on and regulating the AR transcriptional complex. Endocr Rev. 28(7):778–808. doi:10.1210/er.2007-0019.

Heinlein CA, Chang C. 2004. Androgen receptors in prostate cancer. Endocr Rev. 25(2):273–308. doi:10.1097/01.ju.0000075099.20662.7f.

Helsen C, Dubois V, Verfaillie A, Young J, Trekels M, Vancraenenbroeck R, De Maeyer M, Claessens F. 2012. Evidence for DNA-Binding Domain-Ligand-Binding Domain Communications in the Androgen Receptor. Mol Cell Biol. 32(15):3033–3043. doi:10.1128/mcb.00151-12.

Hennig S, Kong G, Mannen T, Sadowska A, Kobelke S, Blythe A, Knott GJ, Iyer SS, Ho D, Newcombe EA,

et al. 2015. Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. J Cell Biol. 210(4):529–539. doi:10.1083/jcb.201504117.

Hernandez-Munoz I, Lund AH, van der Stoop P, Boutsma E, Muijrers I, Verhoeven E, Nusinow DA, Panning B, Marahrens Y, van Lohuizen M. 2005. Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. Proc Natl Acad Sci. 102(21):7635–7640. doi:10.1073/pnas.0408918102.

Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. 2017. A Phase Separation Model for Transcriptional Control. Cell. 169(1):13–23. doi:10.1016/j.cell.2017.02.007.

Hodgson MC, Astapova I, Hollenberg AN, Balk SP. 2007. Activity of androgen receptor antagonist bicalutamide in prostate cancer cells is independent of NCoR and SMRT corepressors. Cancer Res. 67(17):8388–8395. doi:10.1158/0008-5472.CAN-07-0617.

Hong H, Kohli K, Trivedi A, Johnson DL, Stallcup MR. 2002. GRIP1, a novel mouse protein that serves as a transcriptional coactivator in yeast for the hormone binding domains of steroid receptors. Proc Natl Acad Sci. 93(10):4948–4952. doi:10.1073/pnas.93.10.4948.

Hsieh C-L, Fei T, Chen Y, Li T, Gao Y, Wang X, Sun T, Sweeney CJ, Lee G-SM, Chen S, et al. 2014. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proc Natl Acad Sci. 111(20):7319–7324. doi:10.1073/pnas.1324151111.

Huang P, Chandra V, Rastinejad F. 2010. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. Annu Rev Physiol. 72:247–272. doi:10.1146/annurev-physiol-021909-135917.

Huang W, Peng Y, Kiselar J, Zhao X, Albaqami A, Mendez D, Chen Y, Chakravarthy S, Gupta S, Ralston C, et al. 2018. Multidomain architecture of estrogen receptor reveals interfacial cross-talk between its DNA-binding and ligand-binding domains. Nat Commun. 9(1). doi:10.1038/s41467-018-06034-2.

Huggins C, Hodges C V. 1941. Studies of prostatic cancer: I. The effect of castration, of estrogen and of androgen injection on serum phosphatase in metastatic carcinoma of the prostate. Arch Surg. 43(2):209. doi:10.1001/archsurg.1941.01210140043004.

Hur E, Pfaff SJ, Payne ES, Grøn H, Buehrer BM, Fletterick RJ. 2004. Recognition and accommodation at the androgen receptor coactivator binding interface. PLoS Biol. 2(9):E274. doi:10.1371/journal.pbio.0020274.

Hyman AA, Weber CA, Jülicher F. 2014. Liquid-Liquid Phase Separation in Biology. Annu Rev Cell Dev Biol. 30(1):39–58. doi:10.1146/annurev-cellbio-100913-013325.

Ii P, Burke KA, Janke AM, Rhine CL, Fawzi NL. 2015. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Article Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. Mol Cell. 60(2):231–241. doi:10.1016/j.molcel.2015.09.006.

Jenster G et al. 1995. Identification of Two Transcription Activation Units in the N-terminal Domain of the Human Androgen Receptor. :7341–7346.

Jin F, Claessenss F, Fondell JD. 2012. Regulation of androgen receptor-dependent transcription by coactivator MED1 is mediated through a newly discovered noncanonical binding motif. J Biol Chem. 287(2):858–870. doi:10.1074/jbc.M111.304519.

Jin H-J, Zhao JC, Wu L, Kim J, Yu J. 2014. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. Nat Commun. 5(May):3972. doi:10.1038/ncomms4972.

Jishage M, Malik S, Wagner U, Uberheide B, Ishihama Y, Hu X, Chait BT, Gnatt A, Ren B, Roeder RG. 2012. Transcriptional regulation by pol II(G) involving mediator and competitive interactions of gdown1 and tfiif with pol II. Mol Cell. 45(1):51–63. doi:10.1016/j.molcel.2011.12.014.

Juzumiene D, Chang C, Fan D, Hartney T, Norris JD, McDonnell DP. 2005. Single-step purification of full-length human androgen receptor. Nucl Recept Signal. 3:e001. doi:10.1621/nrs.03001.

Kaiser TE, Intine R V., Dundr M. 2008. De Novo Formation of a Subnuclear Body. Science (80-). 322:1713–1717.

Kamada K, De Angelis J, Roeder RG, Burley SK. 2002. Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. Proc Natl Acad Sci. 98(6):3115–3120. doi:10.1073/pnas.051631098.

BIBLIOGRAPHY

Kamada K, Roeder RG, Burley SK. 2003. Molecular mechanism of recruitment of TFIIF- associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. Proc Natl Acad Sci. 100(5):2296–2299. doi:10.1073/pnas.262798199.

Kang Z, Jänne OA, Palvimo JJ. 2004. Coregulator Recruitment and Histone Modifications in Transcriptional Regulation by the Androgen Receptor. Mol Endocrinol. 18(11):2633–2648. doi:10.1210/me.2004-0245.

Khan S, Lingrel JB. 2010. Thematic Minireview Series on Nuclear Receptors in Biology and Diseases. J Biol Chem. 285(50):38741–38742. doi:10.1074/jbc.r110.196014.

Killeen MT, Greenblatt JF. 1992. The general transcription factor RAP30 binds to RNA polymerase II and prevents it from binding nonspecifically to DNA. Mol Cell Biol. 12(1):30–37. doi:10.1128/mcb.12.1.30.

Kim MS, Je EM, Oh JE, Yoo NJ, Lee SH. 2013. Mutational and Expressional Analyses Of SPOP, A Candidate Tumor Suppressor Gene, In Prostate, Gastric and Colorectal Cancers. Apmis. 121(7):626–633. doi:10.1111/apm.12030.

Kornberg R. 2007. The molecular basis of eukaryotic transcription (Nobel Lecture). Angew Chemie - Int Ed. 46(37):6956–6965. doi:10.1002/anie.200701832.

Kozlowski LP, Bujnicki JM. 2012. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins | Lukasz Kozlowski - Academia.edu.

Kroschwald S, Maharana S, Simon A. 2017 May 22. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. Matters. doi:10.19185/matters.201702000010.

Kumar R, Betney R, Li J, Thompson E Brad, McEwan IJ. 2004. Induced alpha-helix structure in AF1 of the androgen receptor upon binding transcription factor TFIIF. Biochemistry. 43(11):3008–3013. doi:10.1021/bi035934p.

Kumar R, Betney R, Li J, Thompson E. Brad, McEwan IJ. 2004. Induced α-Helix Structure in AF1 of the Androgen Receptor upon Binding Transcription Factor TFIIF. Biochemistry. 43(11):3008–3013. doi:10.1021/bi035934p.

Kumar R, Moure CM, Khan SH, Callaway C, Grimm SL, Goswami D, Griffin PR, Edwards DP. 2013. Regulation of the structurally dynamic N-terminal domain of progesterone receptor by protein-induced folding. J Biol Chem. 288(42):30285–99. doi:10.1074/jbc.M113.491787. [accessed 2019 Sep 7]. http://www.ncbi.nlm.nih.gov/pubmed/23995840.

Kumar R, Serrette JM, Khan SH, Miller AL, Thompson EB. 2007. Effects of different osmolytes on the induced folding of the N-terminal activation domain (AF1) of the glucocorticoid receptor. Arch Biochem Biophys. 465(2):452–460. doi:10.1016/j.abb.2007.06.019.

Kumar R, Thompson EB. 2005. Gene regulation by the glucocorticoid receptor: Structure:function relationship. J Steroid Biochem Mol Biol. 94(5):383–394. doi:10.1016/j.jsbmb.2004.12.046.

Kumar R, Volk DE, Li J, Lee JC, Gorenstein DG, Thompson EB. 2004. TATA box binding protein induces structure in the recombinant glucocorticoid receptor AF1 domain. Proc Natl Acad Sci U S A. 101(47):16425–16430. doi:10.1073/pnas.0407160101.

Kumar S, Tyagi RK. 2012. Androgen receptor association with mitotic chromatin - Analysis with introduced deletions and disease-inflicting mutations. FEBS J. 279(24):4598–4614. doi:10.1111/febs.12046.

Lagarde WH, Blackwelder AJ, Minges JT, Hnat AT, French FS, Wilson EM. 2012. Androgen Receptor Exon 1 Mutation Causes Androgen Insensitivity by Creating Phosphorylation Site and Inhibiting Melanoma Antigen-A11 Activation of NH 2 - and Carboxyl- terminal Interaction-dependent Transactivation *. 287(14):10905–10915. doi:10.1074/jbc.M111.336081.

Lambert M, Jambon S, Depauw S, David-Cordonnier M-H. 2018. Targeting Transcription Factors for Cancer Treatment. Molecules. 23(6):1479. doi:10.3390/molecules23061479.

Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ. 2017. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature. 547(7662):236–240. doi:10.1038/nature22822.

Lavery DN, McEwan IJ. 2008a. 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(11):3352–3359. doi:10.1021/bi702220p.

BIBLIOGRAPHY

Lavery DN, McEwan IJ. 2008b. 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(11):3352–3359. doi:10.1021/bi702220p.

Lebendiker M, Danieli T. 2014. Production of prone-to-aggregate proteins. FEBS Lett. 588(2):236–246. doi:10.1016/j.febslet.2013.10.044.

Lecker SH, Goldberg AL, Mitch WE. 2006. Protein Degradation by the Ubiquitin–Proteasome Pathway in Normal and Disease States. J Am Soc Nephrol. 17(7):1807–1819. doi:10.1681/ASN.2006010083.

Lee DK, Duan HO, Chang C. 2000. From Androgen Receptor to the General Transcription Factor TFIIH. Biochemistry. 275(13):9308–9313.

Lee DK, Duan HO, Chang C. 2001. Androgen Receptor Interacts with the Positive Elongation Factor P-TEFb and Enhances the Efficiency of Transcriptional Elongation. J Biol Chem. 276(13):9978–9984. doi:10.1074/jbc.M002285200.

Lee DK, Li M, Chang C. 2003. The second largest subunit of RNA polymerase II interacts with and enhances transactivation of androgen receptor. Biochem Biophys Res Commun. 302(1):162–169. doi:10.1016/S0006-291X(03)00126-8.

Li M, Miwa S, Kobayashi Y, Merry DE, Yamamoto M, Tanaka F, Doyu M, Hashizum Y, Fischbeck KH\\, Sobue G. 1998. Nuclear Inclusions of the Androgen ReceDtor Protein in Spinal and BGlbar Muscular Atrophy. :249–254.

Li Y, Chan SC, Brand LJ, Hwang TH, Silverstein KAT, Dehm SM. 2013. Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. Cancer Res. 73(2):483–489. doi:10.1158/0008-5472.CAN-12-3630.

Liao, S, Fang S. 1969. Receptor-proteims for androgens and the mode of action of androgens on gene transcription in ventral prostate. Vitam Horm. 27:17–90.

Liao G, Chen LY, Zhang A, Godavarthy A, Xia F, Ghosh JC, Li H, Chen JD. 2003. Regulation of androgen receptor activity by the nuclear receptor corepressor SMRT. J Biol Chem. 278(7):5052–5061. doi:10.1074/jbc.M206374200.

Liao M, Wilson EM. 2001. Production and Purification of Histidine-Tagged Dihydrotestosterone-Bound Full-Length Human Androgen Receptor. In: Steroid Receptor Methods. New Jersey: Humana Press. p. 67–79. [accessed 2019 Sep 5]. http://link.springer.com/10.1385/1-59259-115-9:67.

Liao S, Liang T, Shao T-C, Tymoczko JL. 1973. Androgen-Receptor Cycling in Prostate Cells. p. 232–240. [accessed 2019 Sep 5]. http://link.springer.com/10.1007/978-1-4684-3237-4_11.

Lin Y-H, Forman-Kay JD, Chan HS. 2018. Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates. Biochemistry.:acs.biochem.8b00058. doi:10.1021/acs.biochem.8b00058.

Liu C, Lou W, Zhu Y, Nadiminty N, Schwartz CT, Evans CP, Gao AC. 2014. Niclosamide inhibits androgen receptor variants expression and overcomes enzalutamide resistance in castration-resistant prostate cancer. Clin Cancer Res. 20(12):3198–3210. doi:10.1158/1078-0432.CCR-13-3296.

Liu G, Sprenger C, Sun S, Soriano Epilepsia K, Haugk K, Zhang X, Coleman I, Nelson PS, Plymate S. 2015. AR Variant ARv567es Induces Carcinogenesis in a Novel Transgenic Mouse Model of Prostate Cancer. Neoplasia. 15(9):1009-IN3. doi:10.1593/neo.13784.

Liu G, Sprenger C, Wu P-J, Sun S, Uo T, Haugk K, Epilepsia KS, Plymate S. 2015. MED1 mediates androgen receptor splice variant induced gene expression in the absence of ligand. Oncotarget. 6(1). doi:10.18632/oncotarget.2672.

Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. 2006. Intrinsic disorder in transcription factors. Biochemistry. 45(22):6873–6888. doi:10.1021/bi0602718.

Liu Z, Huang Y. 2014. Advantages of proteins being disordered. Protein Sci. 23(5):539–550. doi:10.1002/pro.2443.

Lu H, Yu D, Hansen AS, Ganguly S, Liu R, Heckert A, Darzacq X, Zhou Q. 2018. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. Nature. 1:1. doi:10.1038/s41586-018-0174-3.

Lu JR, Su TJ, Penfold J. 1999. Adsorption of serum albumins at the air/water interface. Langmuir.

15(20):6975–6983. doi:10.1021/la990131h.

Lubahn DB, Joseph DR, Sullivan PM, Willard HF, French FS, Wilson EM. 1988. Cloning of human androgen receptor complementary DNA and localization to the X chromosome. Science (80-). 240(4850):327–330. doi:10.1126/science.3353727.

Mainwaring, W.I.P., Mangan F. 1970. Advances in the Biosciences 7: Schering Workshop on Steroid Hormone "Receptors."

Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schiitq G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, et al. 1995. The nuclear receptor : the second decade.pdf. Cell. 83:835–839.

Marcelli M, Stenoien DL, Szafran AT, Simeoni S, Agoulnik IU, Weigel NL, Moran T, Mikic I, Price JH, Mancini MA. 2006. Quantifying effects of ligands on androgen receptor nuclear translocation, intranuclear dynamics, and solubility. J Cell Biochem. 98(4):770–788. doi:10.1002/jcb.20593.

Marshall TW, Link KA, Petre-Draviam CE, Knudsen KE. 2003. Differential requirement of SWI/SNF for androgen receptor activity. J Biol Chem. 278(33):30605–30613. doi:10.1074/jbc.M304582200.

Martin EW, Mittag T. 2018. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. Biochemistry. 57(17):2478–2487. doi:10.1021/acs.biochem.8b00008.

Marzahn MR, Marada S, Lee J, Nourse A, Kenrick S, Zhao H, Ben-nissan G, Kolaitis R, Peters JL, Pounds S, et al. 2016. Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. 35(12):1254–1275.

Matias Pedro M, Donner P, Coelho R, Thomaz M, Peixoto C, Macedo S, Otto N, Joschko S, Scholz P, Wegg A, et al. 2000. Structural Evidence for Ligand Specificity in the Binding Domain of the Human Androgen Receptor. 275(34):26164–26171. doi:10.1074/jbc.M004571200.

Matias Pedro M., Donner P, Coelho R, Thomaz M, Peixoto C, Macedo S, Otto N, Joschko S, Scholz P, Wegg A, et al. 2000. Structural evidence for ligand specificity in the binding domain of the human androgen receptor: Implications for pathogenic gene mutations. J Biol Chem. 275(34):26164–26171. doi:10.1074/jbc.M004571200.

Mcewan IJ. 2009. The Nuclear Receptor Superfamily.

McEwan IJ, Gustafsson J-A. 2002. Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF. Proc Natl Acad Sci. 94(16):8485–8490. doi:10.1073/pnas.94.16.8485.

Mészáros B, Tompa P, Simon I, Dosztányi Z. 2007. Molecular Principles of the Interactions of Disordered Proteins. J Mol Biol. 372(2):549–561. doi:10.1016/j.jmb.2007.07.004.

Métivier R, Petit FG, Valotaire Y, Pakdel F. 2000. Function of N-Terminal Transactivation Domain of the Estrogen Receptor Requires a Potential α-Helical Structure and Is Negatively Regulated by the A Domain. Mol Endocrinol. 14(11):1849–1871. doi:10.1210/mend.14.11.0546. [accessed 2019 Sep 7]. https://academic.oup.com/mend/article-lookup/doi/10.1210/mend.14.11.0546.

Mirkin SM. 2007. Expandable DNA repeats and human disease. Nature. 447(7147):932–40. doi:10.1038/nature05977.

Mirsky AE, Pauling L. 1931. Chemistry: Mirsk Y and Pa Uling on the Structure of Native, Denatured, and Coagulated Proteins. Sci Pap Part I Cf H Poincar6, op cit. 83(3):179–184.

Mitrea DM, Cika JA, Stanley CB, Nourse A, Onuchic PL, Banerjee PR, Phillips AH, Park CG, Deniz AA, Kriwacki RW. 2018. Self-interaction of NPM1 modulates multiple mechanisms of liquid-liquid phase separation. Nat Commun. 9(1):1–13. doi:10.1038/s41467-018-03255-3.

Mitrea DM, Kriwacki RW. 2016. Phase separation in biology; functional organization of a higher order. Cell Commun Signal. 14(1):1. doi:10.1186/s12964-015-0125-7.

Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker a. K, Uversky VN. 2006. Analysis of Molecular Recognition Features (MoRFs). J Mol Biol. 362(5):1043–1059. doi:10.1016/j.jmb.2006.07.087.

Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. 2006. Analysis of Molecular Recognition Features (MoRFs). J Mol Biol. 362(5):1043–1059. doi:10.1016/j.jmb.2006.07.087.

De Mol E, Szulc E, Di Sanza C, Martínez-Cristóbal P, Bertoncini CW, Fenwick RB, Frigolé-Vivas M, Masín M, Hunter I, Buzón V, et al. 2017. Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators. Structure.:1–8.

doi:10.1016/j.str.2017.11.007.

Monahan Z, Ryan VH, Janke AM, Burke KA, Rhoads SN, Zerze GH, O'Meally R, Dignon GL, Conicella AE, Zheng W, et al. 2017. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. EMBO J. 36(20):2951–2967. doi:10.15252/embj.201696394.

Montgomery RB, Antonarakis ES, Hussain M, Fizazi K, Joshua AM, Attard G, Sadar M, Perabo F, Chi KN. 2019. A phase 1/2 open-label study of safety and antitumor activity of EPI-506, a novel AR N-terminal domain inhibitor, in men with metastatic castration-resistant prostate cancer (mCRPC) with progression after enzalutamide or abiraterone. J Clin Oncol. doi:10.1200/jco.2015.33.15_suppl.tps5072.

Morisaki T, Müller WG, Golob N, Mazza D, McNally JG. 2014. Single-molecule analysis of transcription factor binding at transcription sites in live cells. Nat Commun. 5. doi:10.1038/ncomms5456.

Muiznieks LD, Sharpe S, Pomès R, Keeley FW. 2018. Role of Liquid–Liquid Phase Separation in Assembly of Elastin and Other Extracellular Matrix Proteins. J Mol Biol. 430(23):4741–4753. doi:10.1016/j.jmb.2018.06.010.

Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, Azubel M, Elmlund D, Levi-Kalisman Y, Liu X, Gibbons BJ, et al. 2013. Architecture of an RNA polymerase II transcription pre-initiation complex. Science (80-). 342(6159). doi:10.1126/science.1238724.

Naar, Anders M, Lemon BD, Tjian R. 2001. Transcriptional coactivator complexes. :475–501.

Nadal M, Prekovic S, Gallastegui N, Helsen C, Abella M, Zielinska K, Gay M, Vilaseca M, Taulès M, Houtsmuller AB, et al. 2017. Structure of the homodimeric androgen receptor ligand-binding domain. Nat Commun. 8. doi:10.1038/ncomms14388.

Nagai Y, Kojima T, Muro Y, Hachiya T, Nishizawa Y, Wakabayashi T, Hagiwara M. 1997. Identification of a novel nuclear speckle-type protein, SPOP. FEBS Lett. 418(1–2):23–26. doi:10.1016/S0014-5793(97)01340-9.

Nair SJ, Yang L, Meluzzi D, Oh S, Yang F, Friedman M, Wang S, Suter T, Alshareedah I, Gamliel A, et al. 2019. Phase Separation of Ligand-Activa ted Enhancers Licenses Cooperative Chromosomal Enhancer Assembly. Nat Struct Mol Biol. 26(March). doi:10.1038/s41594-019-0190-5.

Nef P, Natl P, Sci A, Res CT, Vassar R, Ngai J, Axel R, Bozza TC, Kauer JS, Neurosci J, et al. 1999. Regulation of Transcription by a Protein Methyltransferase. Science (80-). 284(June):2174–2177.

Nguyen BD, Abbott KL, Potempa K, Kobor MS, Archambault J, Greenblatt J, Legault P, Omichinski JG. 2003. NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. Proc Natl Acad Sci. 100(10):5688–5693. doi:10.1073/pnas.1031524100.

Nishiyama T, Hashimoto Y, Kota T. 2015. The influence of androgen deprivation therapy on the bone loss in elderly patients with prostate cancer. Clin Cancer Res. 4(35):516–519.

Northrop BYJH. 1930. Crystalline pepsini. Isolation and tests of purity. (2):739–766.

Nott TJ, Craggs TD, Baldwin AJ. 2016. Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. Nat Chem. 8(6):569–575. doi:10.1038/nchem.2519.

Oldfield Christopher J., Dunker AK. 2014. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. Annu Rev Biochem. 83(1):553–584. doi:10.1146/annurev-biochem-072711-164947.

Oldfield Christopher J, Dunker AK. 2014. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. doi:10.1146/annurev-biochem-072711-164947.

Oparin A, Morgulis S. 1938. The Origin of Life. New York Macmillan.:270 pp.

Orr HT. 2012. Polyglutamine neurodegeneration: Expanded glutamines enhance native functions. Curr Opin Genet Dev. 22(3):251–255. doi:10.1016/j.gde.2012.01.001.

Overington JP, Al-Lazikani B, Hopkins AL. 2006. How many drug targets are there ? PubMed Commons. Nat Rev Drug Discov. 5(12):10. doi:10.1038/nrd2199.

Paakinaho V, Presman DM, Ball DA, Johnson TA, Schiltz RL, Levitt P, Mazza D, Morisaki T, Karpova TS, Hager GL. 2017. Single-molecule analysis of steroid receptor and cofactor action in living cells. Nat Commun. 8(May):1–14. doi:10.1038/ncomms15896.

BIBLIOGRAPHY

Paliouras M, Zaman N, Lumbroso R, Kapogeorgakis L, Beitel LK, Wang E, Trifiro M. 2011. Dynamic rewiring of the androgen receptor protein interaction network correlates with prostate cancer clinical outcomes. Integr Biol. 3(10):1020–1032. doi:10.1039/c1ib00038a.

Parker MG, Heery DM, Kalkhoven E, Hoare S. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature. 387(6634):733–736. doi:10.1038/42750.

Patel SS, Belmont BJ, Sante JM, Rexach MF. 2007. Natively Unfolded Nucleoporins Gate Protein Diffusion across the Nuclear Pore Complex. Cell. 129(1):83–96. doi:10.1016/j.cell.2007.01.044.

Paul B. Sigler. 1988. Acid blobs and Negative Noodles. Nature.:210–212.

Peng Y, Cao S, Kiselar J, Xiao X, Du Z, Hsieh A, Ko S, Chen Y, Agrawal P, Zheng W, et al. 2019. A Metastable Contact and Structural Disorder in the Estrogen Receptor Transactivation Domain. Structure. 27(2):229-240.e4. doi:10.1016/j.str.2018.10.026.

Perdigao N, Heinrich J, Stolte C, Sabir KS, Buckley MJ, Tabor B, Signal B, Gloss BS, Hammang CJ, Rost B, et al. 2015. Unexpected features of the dark proteome. Proc Natl Acad Sci. 112(52):15898–15903. doi:10.1073/pnas.1508380112.

Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera JM, Setlur S, Tchinda J, Tomlins SA, Hofer MD, Pienta KG, et al. 2006. TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. Cancer Res. 66(17):8337–8341. doi:10.1158/0008-5472.CAN-06-1482.

Plaschka C, Larivière L, Wenzeck L, Seizl M, Hemann M, Tegunov D, Petrotchenko E V., Borchers CH, Baumeister W, Herzog F, et al. 2015. Architecture of the RNA polymerase II-Mediator core initiation complex. Nature. 518(7539):376–380. doi:10.1038/nature14229.

Pomerantz MM, Li F, Takeda DY, Lenci R, Chonkar A, Chabot M, Cejas P, Vazquez F, Cook J, Shivdasani R a, et al. 2015. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat Genet. 47(11):1346–1351. doi:10.1038/ng.3419.

Prabhudesai S, Sinha S, Attar A, Kotagiri A, Fitzmaurice AG, Lakshmanan R, Ivanova MI, Loo JA, Klärner FG, Schrader T, et al. 2012. A Novel "Molecular Tweezer" Inhibitor of α-Synuclein Neurotoxicity in Vitro and in Vivo. Neurotherapeutics. 9(2):464–476. doi:10.1007/s13311-012-0105-1.

Presman DM, Ganguly S, Schiltz RL, Johnson TA, Karpova TS, Hager GL. 2016. DNA binding triggers tetramerization of the glucocorticoid receptor in live cells. Proc Natl Acad Sci U S A.:2–7. doi:10.1073/pnas.1606774113.

Rani PG, Ranish JA, Hahn S. 2004. RNA Polymerase II (Pol II)-TFIIF and Pol II-Mediator Complexes: the Major Stable Pol II Complexes and Their Activity in Transcription Initiation and Reinitiation. Mol Cell Biol. 24(4):1709–1720. doi:10.1128/mcb.24.4.1709-1720.2004.

Rastinejad F, Huang P, Chandra V, Khorasanizadeh S. 2013. Understanding nuclear receptor form and function using structural biology. J Mol Endocrinol. 51(3). doi:10.1530/JME-13-0173.

Reichheld SE, Muiznieks LD, Keeley FW, Sharpe S. 2017. Direct observation of structure and dynamics during phase separation of an elastomeric protein. Proc Natl Acad Sci. 114(22):E4408–E4415. doi:10.1073/pnas.1701877114.

Reid J, Kelly SM, Watt K, Price NC, McEwan IJ. 2002. Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions. J Biol Chem. 277(22):20079–20086. doi:10.1074/jbc.M201003200.

Reid J, Murray I, Watt K, Betney R, McEwan IJ. 2002a. The androgen receptor interacts with multiple regions of the large subunit of general transcription factor TFIIF. J Biol Chem. 277(43):41247–41253. doi:10.1074/jbc.M205220200.

Reid J, Murray I, Watt K, Betney R, McEwan IJ. 2002b. The androgen receptor interacts with multiple regions of the large subunit of general transcription factor TFIIF. J Biol Chem. 277(43):41247–41253. doi:10.1074/jbc.M205220200.

Reines D, Weiliky Conaway J, Conaway RC. 2013. The RNA polymerase II general elongation factors Daniel. Trends Biochem Sci. 71(2):233–236. doi:10.1038/mp.2011.182.doi.

Rhodes LE, Freeman BK, Auh S, Kokkinis AD, La Pean A, Chen C, Lehky TJ, Shrader JA, Levy EW, Harris-Love M, et al. 2012. Clinical features of spinal and bulbar muscular atrophy. BRAIN J Clin Neurol. 132(12):3242–3251. doi:10.1093/brain/awp258.

BIBLIOGRAPHY

Riback JA, Katanski CD, Kear-scott JL, Pilipenko E V, Rojek AE, Sosnick TR, Drummond DA, Riback JA, Katanski CD, Kear-scott JL, et al. 2017. Stress-Triggered Phase Separation Is an Adaptive , Evolutionarily Tuned Response Article Stress-Triggered Phase Separation Is an Adaptive , Evolutionarily Tuned Response. Cell. 168(6):1028-1040.e19. doi:10.1016/j.cell.2017.02.027.

Ribbeck K, Gorlich D. 2002. The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. 21(11):2664–2671. doi:10.1093/emboj/21.11.2664.

Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. 2015. Erratum: Integrative Clinical Genomics of Advanced Prostate Cancer. Cell. 162(2):454. doi:10.1016/j.cell.2015.06.053.

Rouleau N, Domanskyi A, Reeben M, Moilanen KH, Kang Z, Owen-Hughes T, Palvimo JJ, Janne OA. 2002. Novel ATPase of SNF2-like Protein Family Interacts with Androgen Receptor and Modulates Androgen-dependent Transcription. Mol Biol Cell. 13(6):158–168. doi:10.1091/mbc.01.

van Royen ME, van Cappellen W a., de Vos C, Houtsmuller a. B, Trapman J. 2012a. Stepwise androgen receptor dimerization. J Cell Sci. 125(8):1970–1979. doi:10.1242/jcs.096792.

van Royen ME, van Cappellen W a., de Vos C, Houtsmuller a. B, Trapman J. 2012b. Stepwise androgen receptor dimerization. J Cell Sci. 125(8):1970–1979. doi:10.1242/jcs.096792.

van Royen ME, Cunha SM, Brink MC, Mattern KA, Nigg AL, Dubbink HJ, Verschure PJ, Trapman J, Houtsmuller AB. 2007. Compartmentalization of androgen receptor protein-protein interactions in living cells. J Cell Biol. 177(1):63–72. doi:10.1083/jcb.200609178.

Ru SGD. 2014. Hsp90 interaction with clients. 40(2). doi:10.1016/j.tibs.2014.12.002.

Ruan H, Sun Q, Zhang W, Liu Y, Lai L. 2019. Targeting intrinsically disordered proteins at the edge of chaos. Drug Discov Today. 24(1):217–227. doi:10.1016/j.drudis.2018.09.017.

Sabari BR, Agnese AD, Boija A, Klein IA, Coffey L, Shrinivas K, Abraham BJ, Hannett NM, Zamudio A V, Manteiga JC, et al. 2018. Coactivator condensation at super - enhancers links phase separation and gene control - Science resubmission. doi:10.1126/SCIENCE.AAR3958.

Saitoh M, Takayanagi R, Goto K, Fukamizu A, Tomura A, Yanase T, Nawata H. 2002. The Presence of Both the Amino- and Carboxyl-Terminal Domains in the AR Is Essential for the Completion of a Transcriptionally Active Form with Coactivators and Intranuclear Compartmentalization Common to the Steroid Hormone Receptors: A Three-Dimensional. Mol Endocrinol. 16(4):694–706. doi:10.1210/me.16.4.694.

Schaufele F, Carbonell X, Guerbadot M, Borngraeber S, Chapman MS, Ma AAK, Miner JN, Diamond MI. 2005. The structural basis of androgen receptor activation: intramolecular and intermolecular amino-carboxy interactions. Proc Natl Acad Sci U S A. 102(28):9802–9807. doi:10.1073/pnas.0408819102.

Sever R, Glass CK. 2013. Signaling by nuclear receptors. Cold Spring Harb Perspect Biol. 5(3):1–4. doi:10.1101/cshperspect.a016709.

Shaffer PL, Jivan A, Dollins DE, Claessens F, Gewirth DT. 2004. Structural basis of androgen receptor binding to selective androgen response elements. Proc Natl Acad Sci U S A. 101(14):4758–4763. doi:10.1073/pnas.0401123101.

Shang Y, Myers M, Brown M. 2002. Formation of the androgen receptor transcription complex. Mol Cell. 9(3):601–10. doi:10.1016/S1097-2765(02)00471-9.

Shen TH, Lin HK, Scaglioni PP, Yung TM, Pandolfi PP. 2006. The Mechanisms of PML-Nuclear Body Formation. Mol Cell. 24(3):331–339. doi:10.1016/j.molcel.2006.09.013.

Shin Y, Brangwynne CP. 2017. Liquid phase condensation in cell physiology and disease. Science (80-). 357(6357):eaaf4382. doi:10.1126/science.aaf4382.

Shin Y, Chang YC, Lee DSW, Berry J, Sanders DW, Ronceray P, Wingreen NS, Haataja M, Brangwynne CP. 2018. Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. Cell. 175(6):1481-1491.e13. doi:10.1016/j.cell.2018.10.057.

Shliaha P V., Baird MA, Nielsen MM, Gorshkov V, Bowman AP, Kaszycki JL, Jensen ON, Shvartsburg AA. 2017. Characterization of Complete Histone Tail Proteoforms Using Differential Ion Mobility Spectrometry. Anal Chem. 89(10):5461–5466. doi:10.1021/acs.analchem.7b00379.

Shukla GC, Plaga AR, Shankar E, Gupta S. 2016. Androgen receptor-related diseases: What do we know?

Andrology. 4(3):366–381. doi:10.1111/andr.12167.

Skupien-Rabian B, Jankowska U, Swiderska B, Lukasiewicz S, Ryszawy D, Dziedzicka-Wasylewska M, Kedracka-Krok S. 2015. Proteomic and bioinformatic analysis of a nuclear intrinsically disordered proteome. J Proteomics. 130:76–84. doi:10.1016/j.jprot.2015.09.004.

Soutourina J. 2018. Transcription regulation by the Mediator complex. Nat Rev Mol Cell Biol. 19(4):262–274. doi:10.1038/nrm.2017.115.

La Spada A. 2017. Spinal and Bulbar Muscular Atrophy Summary Suggestive Findings. GeneReviews.:1–16.

La Spada AR, Roling DB, Harding AE, Warner CL, Spiegel R, Hausmanowa-petrusewicz I, Yee W, Fischbeck KH. 1992. binucleotide repeat in X-linked spinal and bulbar muscular atrophy. 2(december).

Staby L, O'Shea C, Willemoës M, Theisen F, Kragelund BB, Skriver K. 2017. Eukaryotic transcription factors: paradigms of protein intrinsic disorder. Biochem J. 474(15):2509–2532. doi:10.1042/BCJ20160631.

van der Steen T, Tindall DJ, Huang H. 2013. Posttranslational modification of the androgen receptor in prostate cancer. Int J Mol Sci. 14(7):14833–14859. doi:10.3390/ijms140714833.

Stelloo S, Bergman AM, Zwart W. 2019. Androgen receptor enhancer usage and the chromatin regulatory landscape in human prostate cancers. Endocr Relat Cancer. 26(5):R267–R285. doi:10.1530/erc-19-0032.

Strom AR, Emelyanov A V., Mir M, Fyodorov D V., Darzacq X, Karpen GH. 2017. Phase separation drives heterochromatin domain formation. Nature. 547(7662):241–245. doi:10.1038/nature22989.

Sun F, Chen HG, Li W, Yang X, Wang X, Jiang R, Guo Z, Chen H, Huang J, Borowsky AD, et al. 2014. Androgen receptor splice variant AR3 promotes prostate cancer via modulating expression of autocrine/paracrine factors. J Biol Chem. 289(3):1529–1539. doi:10.1074/jbc.M113.492140.

Sun M, Choueiri TK, Hamnvik OPR, Preston MA, De Velasco G, Jiang W, Loeb S, Nguyen PL, Trinh QD. 2016. Comparison of gonadotropin-releasing hormone agonists and orchiectomy: Effects of androgen-deprivation therapy. JAMA Oncol. 2(4):500–507. doi:10.1001/jamaoncol.2015.4917.

Takeda DY, Spisák S, Seo JH, Bell C, O'Connor E, Korthauer K, Ribli D, Csabai I, Solymosi N, Szállási Z, et al. 2018. A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer. Cell. 174(2):422-432.e13. doi:10.1016/j.cell.2018.05.037.

Tamae D, Mostaghel E, Montgomery B, Nelson PS, Balk SP, Kantoff PW, Taplin ME, Penning TM. 2015. The DHEA-sulfate depot following P450c17 inhibition supports the case for AKR1C3 inhibition in high risk localized and advanced castration resistant prostate cancer. Chem Biol Interact. 234:332–338. doi:10.1016/j.cbi.2014.12.012.

Tan J, Sharief Y, Hamil KG, Gregory CW, Zang D-Y, Sar M, Gumerlock PH, deVere White RW, Pretlow TG, Harris SE, et al. 2014. Dehydroepiandrosterone Activates Mutant Androgen Receptors Expressed in the Androgen-Dependent Human Prostate Cancer Xenograft CWR22 and LNCaP Cells. Mol Endocrinol. 11(4):450–459. doi:10.1210/mend.11.4.9906.

Tan ME, Li J, Xu HE, Melcher K, Yong E. 2015. Androgen receptor: structure, role in prostate cancer and drug discovery. Acta Pharmacol Sin. 36(1):3–23. doi:10.1038/aps.2014.18.

Tatarakis A, Behrouzi R, Moazed D. 2017. Evolving Models of Heterochromatin: From Foci to Liquid Droplets. Mol Cell. 67(5):725–727. doi:10.1016/j.molcel.2017.08.022.

Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, et al. 2010. Integrative Genomic Profiling of Human Prostate Cancer. Cancer Cell. 18(1):11–22. doi:10.1016/j.ccr.2010.05.026.

Theurillat J-PP, Namrata D. Udeshi 1*, Wesley J. Errington, 4 5*, Tanya Svinkina, 1 ylvan C. Baca, 2, 3 Marius Pop, 1, 2 3, Peter J. Wild, 6 Mirjam Blattner 7, Anna C. Groner, 3 Mark A. Rubin, 7, 8 Holger Moch, 6 Gilbert G. Privé, 4, 5 Steven A. Carr 1 Levi A. Garraway. 2014. Ubiquitylome analysis identifies dysregulation of effector substrates in SPOP-mutant prostate cancer. Science (80-). 1090(2006):3–8.

Tilley WD, Marcelli M, Wilson JD, McPhaul MJ. 1989. Characterization and expression of a cDNA encoding the human androgen receptor. Proc Natl Acad Sci U S A. 86(1):327–331. doi:10.1073/pnas.86.1.327.

Titus MA, Schell MJ, Lih FB, Tomer KB, Mohler JL. 2005. Testosterone and dihydrotestosterone tissue levels in recurrent prostate cancer. Clin Cancer Res. 11(13):4653–4657. doi:10.1158/1078-0432.CCR-05-

0525.

Tompa P. 2002. Intrinsically unstructured proteins. Trends Biochem Sci. 27(10):527–33.

Tompa P, Fuxreiter M. 2008. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci. 33(1):2–8. doi:10.1016/j.tibs.2007.10.003.

Tóth-Petróczy Á, Oldfield CJ, Simon I, Takagi Y, Dunker AK, Uversky VN, Fuxreiter M. 2008. Malleable machines in transcription regulation: The Mediator complex. PLoS Comput Biol. 4(12). doi:10.1371/journal.pcbi.1000243.

Trapman J, Klaassen P, Kuiper GGJM, van der Korput JAGM, Faber PW, van Rooij HCJ, van Kessel AG, Voorhorst MM, Mulder E, Brinkmann AO. 1988. Cloning, structure and expression of a cDNA encoding the human androgen receptor. Biochem Biophys Res Commun. 153(1):241–248. doi:10.1016/S0006-291X(88)81214-2.

Tsafou K, Tiwari PB, Forman-Kay JD, Metallo SJ, Toretsky JA. 2018. Targeting intrinsically disordered transcription factors: Changing the paradigm. J Mol Biol.:#pagerange#. doi:10.1016/j.jmb.2018.04.008.

Tsai KL, Tomomori-Sato C, Sato S, Conaway RC, Conaway JW, Asturias FJ. 2014. Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. Cell. 157(6):1430–1444. doi:10.1016/j.cell.2014.05.015.

Tveter, K.J., Unhjem, O., Attramadal, A., Aakvaag, A.,Hansson V. 1970. Advances in the Biosciences 7: Schering Workshop on Steroid Hormone “Receptors.”

Urbanucci A, Barfeld SJ, Kytölä V, Itkonen HM, Coleman IM, Vodák D, Sjöblom L, Sheng X, Tolonen T, Minner S, et al. 2017. Androgen Receptor Deregulation Drives Bromodomain-Mediated Chromatin Alterations in Prostate Cancer. Cell Rep. 19(10):2045–2059. doi:10.1016/j.celrep.2017.05.049.

Uversky VN. 2002. What does it mean to be natively unfolded? Eur J Biochem. 12(May 2001):2–12. doi:10.1110/ps.4210102.matic.

Uversky VN. 2009. Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding. Protein J. 28(7–8):305–325. doi:10.1007/s10930-009-9201-4.

Uversky VN. 2011. Intrinsically disordered proteins from A to Z. Int J Biochem Cell Biol. 43(8):1090–1103. doi:10.1016/j.biocel.2011.04.001.

Uversky VN. 2012. Intrinsically disordered proteins and novel strategies for drug discovery. Expert Opin Drug Discov. 7(6):475–488. doi:10.1517/17460441.2012.686489.

Uversky VN, Dunker AK. 2010. Understanding protein non-folding. Biochim Biophys Acta - Proteins Proteomics. 1804(6):1231–1264. doi:10.1016/j.bbapap.2010.01.017.

Uversky VN, Gillespie JR, Fink AL. 2000. Why are “natively unfolded” proteins unstructured under physiologic conditions? Proteins Struct Funct Genet. 41(3):415–427. doi:10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7.

Uversky VN, Oldfield CJ, Dunker AK. 2008. Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2 Concept . Annu Rev Biophys. 37(1):215–246. doi:10.1146/annurev.biophys.37.032807.125924.

Valdscholte J, Ris-Stalpers C, Kuiper GGJM, Jenster G, Berrevoets C, Claassen E, van Rooij HCJ, Trapman J, Brinkmann AO, Mulder E. 1990. A mutation in the ligand binding domain of the androgen receptor of human INCaP cells affects steroid binding characteristics and response to anti-androgens. Biochem Biophys Res Commun. 173(2):534–540. doi:10.1016/S0006-291X(05)80067-1.

Vijayvargia R, May MS, Fondell JD. 2007. A coregulatory role for the mediator complex in prostate cancer cell proliferation and gene expression. Cancer Res. 67(9):4034–4041. doi:10.1158/0008-5472.CAN-06-3039.

Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinanen R, Palmberg C, Palotie A, Tammela T, Isola J, Kallioniemi O. 1995. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. Nat Genet. 9:401–406.

Viswanathan SR, Ha G, Hoff AM, Wala JA, Carrot-Zhang J, Whelan CW, Haradhvala NJ, Freeman SS, Reed SC, Rhoades J, et al. 2018. Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. Cell. 174(2):433-447.e19. doi:10.1016/j.cell.2018.05.036.

BIBLIOGRAPHY

Wadosky KM, Koochekpour S. 2017. Androgen receptor splice variants and prostate cancer: From bench to bedside. Oncotarget. 8(11):18550–18576. doi:10.18632/oncotarget.14537.

Walker JM. The nuclear receptor superfamily : Method and protocol (springer Link) (E Book).

Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovsky A, Drechsel D, et al. 2018. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell. 174(3):688-699.e16. doi:10.1016/j.cell.2018.06.006.

Wang Q, Carroll JS, Brown M. 2005. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. Mol Cell. 19(5):631–642. doi:10.1016/j.molcel.2005.07.018.

Wang Q, Sharma D, Ren Y, Fondell JD. 2002. A coregulatory role for the TRAP-Mediator complex in androgen receptor-mediated gene expression. J Biol Chem. 277(45):42852–42858. doi:10.1074/jbc.M206061200.

Wang S, Zhao Y, Aguilar A, Bernard D, Yang C. 2017. Targeting the MDM2–p53 Protein–Protein Interaction for New Cancer Therapy: Progress and Challenges. Cold Spring Harb Perspect Med.

Wang Z, Wang Y, Zhang J, Hu Q, Zhi F, Zhang S, Mao D, Zhang Y, Liang H. 2017. Significance of the TMPRSS2: ERG gene fusion in prostate cancer. Mol Med Rep. 16(4):5450–5458. doi:10.3892/mmr.2017.7281.

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. J Mol Biol. 337(3):635–645. doi:10.1016/j.jmb.2004.02.002.

Watson PA, Arora VK, Sawyers CL. 2015. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. Nat Rev Cancer. 15(12):701–711. doi:10.1038/nrc4016.

Watson PA, Chen YF, Balbas MD, Wongvipat J, Socci ND, Viale A, Kim K, Sawyers CL. 2010. Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. Proc Natl Acad Sci U S A. 107(39):16759–65. doi:10.1073/pnas.1012443107.

van de Wijngaart DJ, Dubbink HJ, van Royen ME, Trapman J, Jenster G. 2012. Androgen receptor coregulators: Recruitment via the coactivator binding groove. Mol Cell Endocrinol. 352(1–2):57–69. doi:10.1016/j.mce.2011.08.007.

Van De Wijngaart DJ, Molier M, Lusher SJ, Hersmus R, Jenster G, Trapman J, Dubbink HJ. 2010. Systematic structure-function analysis of androgen receptor Leu 701 mutants explains the properties of the prostate cancer mutant L701H. J Biol Chem. 285(7):5097–5105. doi:10.1074/jbc.M109.039958.

Williams RM, Obradovi Z, Mathura V, Braun W, Garner EC, Young J, Takayama S, Brown CJ, Dunker a K. 2001. The protein non-folding problem: amino acid determinants of intrinsic order and disorder. Pac Symp Biocomput.:89–100.

Wilson, JD, Gloyna R. 1970. The intranuclear metabolism of testosterone in the accessory organs of reproduction. Recent Prog Horm Res. 26:309–36.

Wilson S, Qi J, Filipp F V. 2016. Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. Sci Rep. 6(August):1–15. doi:10.1038/srep32611.

Woychik NA, Hampsey M. 2002. The RNA polymerase II machinery: structure illuminates function. Cell. 108(4):453–63.

Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J Mol Biol. 293(2):321–331. doi:10.1006/jmbi.1999.3110.

Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z. 2007. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res. 6(5):1882–1898. doi:10.1021/pr060392u.

Xie YB, Sui YP, Shan LX, Palvimo JJ, Phillips DM, Jänne OA. 1992. Expression of androgen receptor in insect cells. Purification of the receptor and renaturation of its steroid- and DNA-binding functions. J Biol Chem. 267(7):4939–48. [accessed 2019 Sep 5]. http://www.ncbi.nlm.nih.gov/pubmed/1537872.

BIBLIOGRAPHY

Xu D, Zhan Y, Qi Y, Cao B, Bai S, Xu W, Gambhir SS, Lee P, Sartor O, Flemington EK, et al. 2015. Androgen Receptor Splice Variants Dimerize to Transactivate Target Genes. Cancer Res. 75(17):3663–3671. doi:10.1158/0008-5472.CAN-15-0381.

Yeh, S., Miyamoto, H., Shima, H., Chang C. 1998. From estrogen to androgen receptor : A new pathway for sex. Biochemistry. 95(May):5527–5532.

Yi P, Wang Z, Feng Q, Pintilie GD, Foulds CE, Lanz RB, Ludtke SJ, Schmid MF, Chiu W, O'Malley BW. 2015. Structure of a Biologically Active Estrogen Receptor-Coactivator Complex on DNA. Mol Cell. 57(6):1047–1058. doi:10.1016/j.molcel.2015.01.025.

Yin X, Giap C, Lazo JS, Prochownik E V. 2003. Low molecular weight inhibitors of Myc-Max interaction and function. Oncogene. 22(40):6151–6159. doi:10.1038/sj.onc.1206641.

Zhang C, Yan H, Burton ZF. 2003. Combinatorial Control of Human RNA Polymerase II (RNAP II) Pausing and Transcript Cleavage by Transcription Factor IIF, Hepatitis δ Antigen, and Stimulatory Factor II. J Biol Chem. 278(50):50101–50111. doi:10.1074/jbc.M307590200.

Zhao L, Zhou S, Gustafsson J-Å. 2019. Nuclear receptors: recent drug discovery for cancer therapies.

Zhao XY, Malloy PJ, Krishnan A V., Swami S, Navone NM, Peehl DM, Feldman D. 2000. Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. Nat Med. 6(6):703–706. doi:10.1038/76287.

Zhao Y, Wang Liguo, Ren S, Wang Lan, Blackburn PR, McNulty MS, Gao X, Qiao M, Vessella RL, Kohli M, et al. 2016. Activation of P-TEFb by Androgen Receptor-Regulated Enhancer RNAs in Castration-Resistant Prostate Cancer. Cell Rep. 15(3):599–610. doi:10.1016/j.celrep.2016.03.038.

Zhou XE, Suino-Powell K, Ludidi PL, McDonnell DP, Xu HE. 2010a. Expression, purification and primary crystallographic study of human androgen receptor in complex with DNA and coactivator motifs. Protein Expr Purif. 71(1):21–7. doi:10.1016/j.pep.2009.12.002.

Zhou XE, Suino-Powell K, Ludidi PL, McDonnell DP, Xu HE. 2010b. Expression, purification and primary crystallographic study of human androgen receptor in complex with DNA and coactivator motifs. Protein Expr Purif. 71(1):21–7. doi:10.1016/j.pep.2009.12.002. [accessed 2019 Sep 5]. http://www.ncbi.nlm.nih.gov/pubmed/19995608.

Zhu P, Baek SH, Bourk EM, Ohgi K a, Garcia-Bassets I, Sanjo H, Akira S, Kotol PF, Glass CK, Rosenfeld MG, et al. 2006. Macrophage/cancer cell interactions mediate hormone resistance by a nuclear receptor derepression pathway. Cell. 124(3):615–29. doi:10.1016/j.cell.2005.12.032.

Zhuang M, Calabrese MF, Liu J, Waddell MB, Nourse A, Hammel M, Miller DJ, Walden H, Duda DM, Seyedin SN, et al. 2009. Structures of SPOP-Substrate Complexes: Insights into Molecular Architectures of BTB-Cul3 Ubiquitin Ligases. Mol Cell. 36(1):39–50. doi:10.1016/j.molcel.2009.09.022.

9 SUPPLEMENTARY INFORMATION

9.1 EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor

EPI-001, a small molecule active against castration-resistant prostate cancer, targets transactivation unit 5 of the androgen receptor

Eva De Mol,[1] R. Bryn Fenwick,[1] Christopher T. W. Phang,[1] Victor Buzón,[2] Elzbieta Szulc,[1] Alex de la Fuente,[1] Albert Escobedo,[1] Jesús García,[1] Carlos W. Bertoncini,[1] Eva Estébanez-Perpiñá,[2] Iain J. McEwan,[3] Antoni Riera[1,4] and Xavier Salvatella[1,5,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain.

[2]Departament de Bioquímica i Biología Molecular, Universitat de Barcelona and Institute of Biomedicine of the University of Barcelona (IBUB), Baldiri Reixac 10, 08028 Barcelona, Spain.

[3]School of Medical Sciences, University of Aberdeen, IMS Building, Foresterhill, Aberdeen AB25 2ZD, Scotland, United Kingdom.

[4]Departament de Química Orgánica, Universitat de Barcelona, Martí i Franqués 1-11, 08028 Barcelona, Spain.

[5]ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain.

Supplementary Information

Contents

1 Detailed experimental procedures: NMR

1.1 Triple resonance experiments for chemical shift assignment

To assign the backbone of AF-1* three-dimensional triple resonance experiments (HNCA and HN(CO)CA, HNCO and HN(CA)CO, CBCANH and CBCA(CO)NH) were acquired with standard Bruker pulse sequences on a Bruker DRX 800 MHz spectrometer equipped with a triple resonance TCI cryoprobe (AF-1*, AR 142-448) or on a Bruker Avance III UltraShield Plus 600 MHz spectrometer equipped with a triple resonance TCI cryoprobe (Tau-5*, AR 330-448, and AF-1*$_{265-340}$, AR 265-340). Spectra were acquired at 278 K of 340 µM AF-1* in 20 mM sodium phosphate buffer with 1 mM TCEP at pH 7.4.

1.2 [^{1}H,^{15}N]-HSQC experiments in the absence and presence of EPI-001 and related compounds

Sample preparation

A solution of 25 µM ^{15}N-labeled AF-1* protein was prepared in 20 mM sodium phosphate buffer at pH 7.4 containing 1 mM TCEP, 10% D_2O and 30 mM deuterated DSS (DSS-d_6), which was used for reference purposes. Aliquots were flash frozen in liquid nitrogen and stored at -20°C until they were used for NMR experiments. To measure a [^{1}H,^{15}N]-HSQC spectrum of 25 µM ^{15}N-labeled AF-1* an aliquot was thawed and 350µL was transferred to a D_2O-matched Shigemi tube. For the blank measurements, 0.5% (v/v) dioxane-d_8 was added to one of these aliquots after thawing. To add EPI-001, 1.75 µL of a 50 mM EPI-001 stock solution in 100% dioxane-d_8 was added to a protein aliquot. Samples of AF-1* and (2*R*,20*S*)-EPI-002, (2*S*,20*R*)-EPI-003, (2*R*,20*R*)-EPI-004 or (2*S*,20*S*)-EPI-005 were prepared in the same way.

NMR acquisition

The [^{1}H,^{15}N]-HSQC spectra of AF-1* in the absence or presence of EPI-001, (2*R*,20*S*)-EPI-002, (2*S*,20*R*)-EPI-003, (2*R*,20*R*)-EPI-004 or (2*S*,20*S*)-EPI-005 were acquired at 278 K on a Bruker DRX 800 MHz spectrometer equipped with a triple resonance TCI cryoprobe, using 2048 complex points in the direct dimension (^{1}H) and 512 increments in the indirect dimension (^{15}N). The number of scans was set to 16 and a sweep width (SW) of 8013 Hz (10 ppm) (^{1}H) and 1784 Hz (22 ppm) (^{15}N) was used. To allow complete equilibration of the sample temperature, the measurements were started 30 minutes after the samples were introduced into the NMR spectrometer that was already set to 278 K.

NMR processing

The spectra were processed using NMRDraw and NMRPipe [1]. Carrier frequencies in ppm were inserted with three decimals. The x-carrier frequency was determined by referencing to internal DSS-d_6. The DSS-d_6 frequency was obtained from a 1D experiment recorded immediately

before the 2D experiment. Indirect referencing in the ^{15}N dimension was performed by using the conversion factors from Wishart *et al* [2]. In the direct dimension (^{1}H) zero filling to the nearest power of two was applied, whereas in the indirect dimension (^{15}N) both forward-backward linear prediction and zero filling to the nearest power of two were used.

NMR data analysis

The processed spectra were analyzed in CcpNmr Analysis [3]. Non-overlapped peaks were picked using the automated peak picking routine in CcpNmr Analysis. Overlapped peaks were carefully picked manually.

The experimental reproducibility of the chemical shift measurements was determined by recording [^{1}H,^{15}N]-HSQC spectra of three individually prepared but otherwise identical NMR samples and comparing the chemical shifts, similar to Bruun *et al.* [4] The average standard deviation of the chemical shift measurements was ±0.5 ppb for ^{1}H and ±2.4 ppb for ^{15}N obtained as an average over all peaks in the spectra as shown in the following table.

Table 1: Digital resolution of the NMR measurements before and after processing of the data with relevant acquisition and processing parameters and the experimentally determined reproducibility of the chemical shift measurements.

	^{1}H	^{15}N
Sweep width (SW) in ppm	10	22
Number of points/increments	2048	512
Linear prediction (LP)	none	1024
Zero filling (ZF)	4096	2048
Digital resolution in ppb	4.88	43
Digital resolution after LP and ZF in ppb	2.44	10.7
Experimentally determined reproducibility in ppb	±0.5	±2.4

1.3 1D ^{1}H NMR experiments to monitor the effect of AF-1 and peptides R1, R2 and R3 on the EPI-001 resonances*

Sample preparation

To 350 µL aliquots of AF-1* samples the appropriate volume of a 50 mM EPI-001 stock solution in 100% dioxane-d_8 was added to obtain final concentration 250 µM EPI-001 in 0.5% (v/v)

dioxane-d_8. A blank sample was prepared without AF-1* by adding the same volume of 50 mM EPI-001 stock solution in 100% dioxane-d_8 to 350 µL 20 mM sodium phosphate buffer at pH 7.4, with 1 mM TCEP, 10% D_2O and 30 mM DSS-d_6.

Samples containing 100 µM EPI-001 and 25 µM R1 or R3 peptide were prepared by solubilizing lyophilized peptide aliquots directly in 350 µL 20 mM sodium phosphate buffer at pH 7.4, with 1 mM TCEP, 10% D_2O and 30 mM DSS-d_6 to final concentration 25 µM. 1.75 µL of a 20 mM EPI-001 stock solution in 100% dioxane-d_8 was added to these samples to obtain final concentration 100 µM EPI-001 in 0.5% (v/v) dioxane-d_8. A blank sample was prepared identically as described above. The R2 peptide only contains one charged residue and is 24 amino acids long. As expected, it was poorly soluble without the addition of a co-solvent. To obtain a soluble R2 peptide sample at 25 µM it was necessary to use 1% DMSO-d_6 as co-solvent.

The lyophilized R2 peptide was directly solubilized in 100% DMSO-d_6 and to an aliquot of this solution 20 mM sodium phosphate buffer at pH 7.4, containing 1 mM TCEP, 10% D_2O and 30 mM DSS-d_6 was added to obtain a final sample of 25 µM R2 peptide in buffer containing 1% (v/v) DMSO-d_6. To 350 µL of this sample, 1.75 µL of a 20 mM EPI-001 stock solution in 100% dioxane-d_8 was added. The final sample for NMR contained 25 µM R2 peptide, 100 µM EPI-001, 1% DMSO-d_6 and 0.5% dioxane-d_8 in 20 mM sodium phosphate buffer at pH 7.4, 1 mM TCEP, 10% D_2O and 30 mM DSS-d_6. For this sample, a blank sample was prepared with identical composition but without R2 peptide, i.e. 100 µM EPI-001 in 20 mM sodium phosphate buffer at pH 7.4 containing 1 mM TCEP, 10% D_2O, 1% DMSO-d_6, 0.5% dioxane-d_8 and 30 mM DSS-d_6.

NMR experiments

Spectra were acquired at 278 K on a Bruker Avance III UltraShield Plus 600 MHz spectrometer equipped with a triple resonance TCI cryoprobe. The SW was set to 11 ppm, the number of points to 32768 and the number of scans to 128. A relaxation delay of 7 s was used. The digital resolution was 0.20 Hz, i.e. 0.00034 ppm.

1.4 Measurement of ^{15}N R_2 relaxation rates

Backbone amide ^{15}N transverse relaxation (R_2) measurements were acquired at 278 K on a Bruker Avance 800 MHz spectrometer equipped with a cryoprobe using a sample of 250 mM ^{15}N-AF-1*. A series of ten 2D experiments with relaxation delays ranging from 20 to 200 ms (20, 35 (twice), 50, 75, 100, 125, 165 (twice), and 200 ms) were collected in a randomized order. A recycle delay between scans of 3 s was used. The processed spectra were analyzed in CcpNmr Analysis.

2 Detailed experimental procedures: synthesis of EPI-001 and stereoisomers.

2.1 Methods and materials for synthetic work

Non-aqueous reactions were carried out under nitrogen atmosphere. Dry tetrahydrofuran and dichloromethane were obtained using a Solvent Purification System (SPS). Other commercially available reagents and solvents were used without further purification. All reactions were monitored by TLC analysis using Merck 60 F_{254} silica gel on aluminum sheets. Silica gel chromatography was performed by using 35-70 mm silica or an automated chromatography system (Combiflash®, Teledyne Isco) with hexanes/ethyl acetate gradients as eluent unless noted otherwise.

NMR spectra were recorded at room temperature on a Varian Mercury 400 or a Bruker 300. ^{1}H and ^{13}C-NMR spectra were referenced to the residual peaks of the deuterated solvent. The following abbreviations were used to define the multiplicities: s, singlet; d, doublet; t, triplet; q, quadruplet; p, pentuplet; m, multiplet; br s, broad signal; apt, apparent. The chemical shifts (d) are expressed in ppm and the coupling constants (J), in hertz (Hz).

IR spectra were recorded in a Thermo Nicolet Nexus FT-IR apparatus, either by preparing a KBr pastille or by depositing a film of the product on a NaCl window. Absorptions are given in wavenumbers (cm^{-1}).

Optical rotations were measured at room temperature (25°C) using a Jasco P-2000 iRM-800 polarimeter. Concentration is expressed in g/100 mL and solvent is expressed for each case in brackets. The cell sized 10 cm long and had 1 ml of capacity. The λ of measure was 589 nm, which corresponds to a sodium lamp.

Mass spectrometry analysis was performed as high resolution ESI analysis in LTQ-FT Ultra (Thermo Scientific)

Chiral HPLC was done using a Chiralpak IA (col-HP-32) column in heptane:IPA:EtOH (50:25:25); flow rate 0.5 mL/min; Injection Volume 10 µL; wavelength detection 220 nm; Sample concentration 2 mg/mL

2.2 Synthesis description

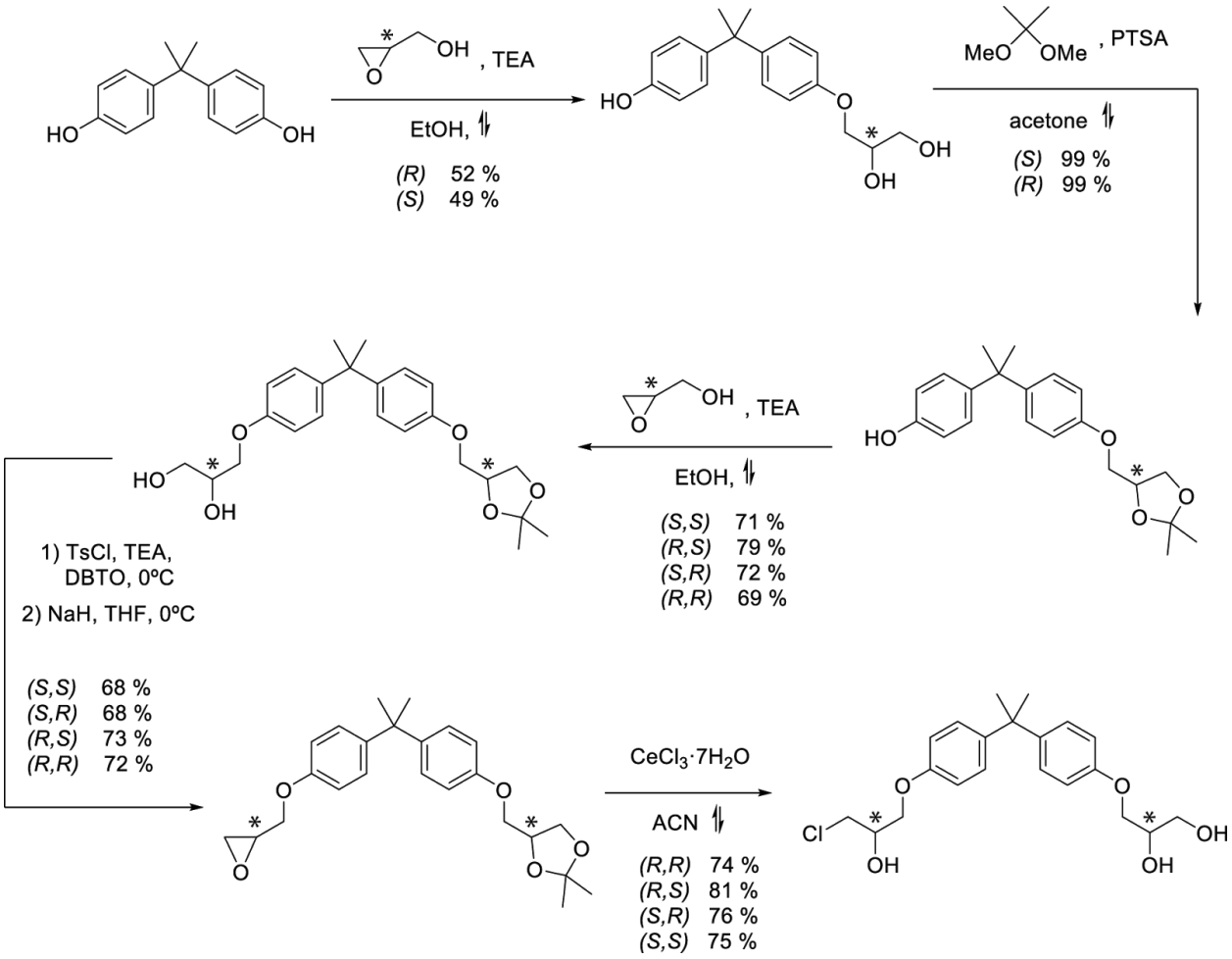


Figure S1: general synthetic scheme used in this work

General procedure 1A for diol formation

Starting material (1 eq) was dissolved in EtOH in a pressure tube. Then triethylamine (0.5 eq) and glycidol (1.2 eq) were added and the mixture was heated at reflux for 16h. After that the solvent was removed in vacuo and the obtained crude products purified by flash column chromatography on silica gel using hexane/ethyl acetate and increasing the polarity ratio from 60:40 to 0:100.

General procedure 1B for diol formation

Starting material (1 eq) and Cs_2CO_3 (1.1 eq) were placed in a pressure tube and dissolved in EtOH. Then glycidol (1.2 eq) was added and the mixture was heated at reflux for 16h. The

crude was treated with 5 mL HCl 1M, extracted with EtOAc (3x 5 mL), dried with $MgSO_4$ and purified by flash column chromatography on silica gel using hexane/ethyl acetate and increasing the polarity ratio from 60:40 to 0:100.

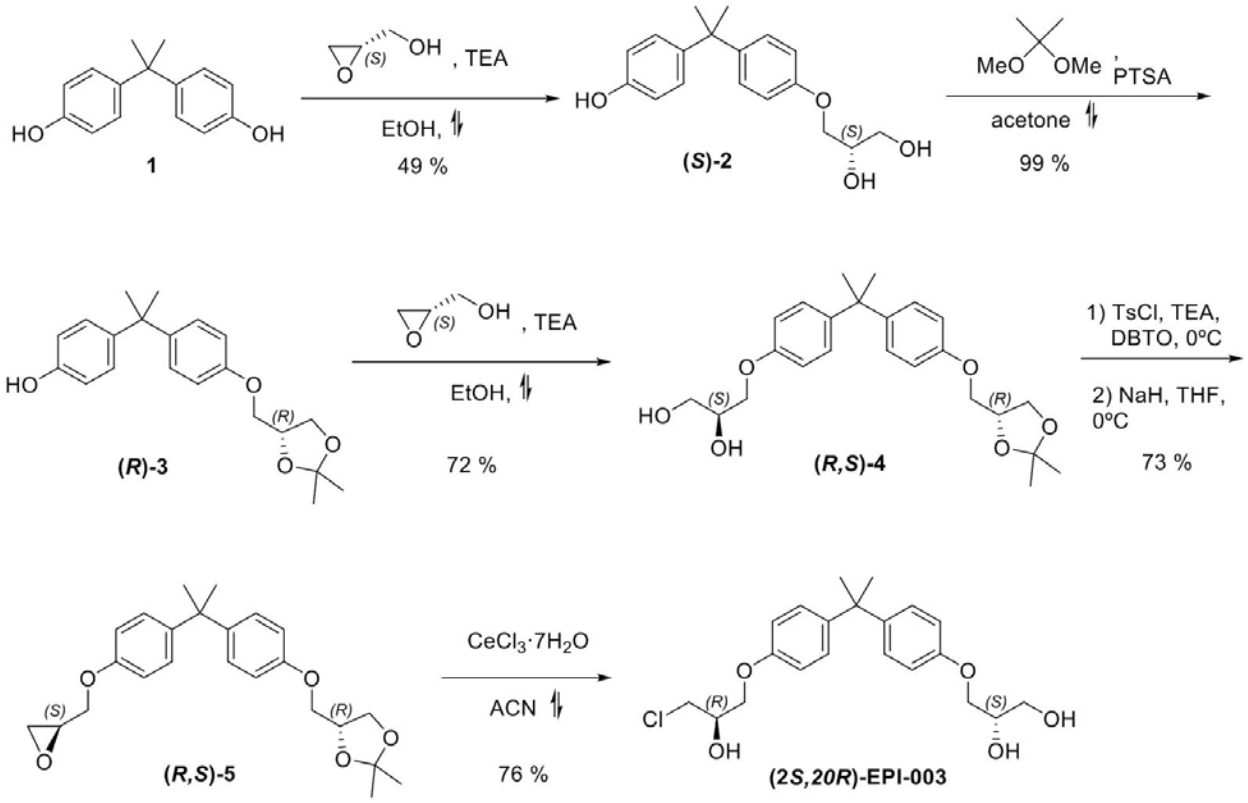
General procedure 2 for acetal formation

Starting material (1 eq) and PTSA (0.15 eq) were dissolved in acetone. Dimethoxypropane (2 eq) was added and the mixture stirred at reflux for 8h. The reaction was treated with aq $NaHCO_3$ sat. (5 mL) and organic solvent was removed in vacuo. The resulting aqueous phase was extracted with CH_2Cl_2 (3x5 mL), dried over $MgSO_4$ and solvents removed in vacuo. Compounds were used without further purification.

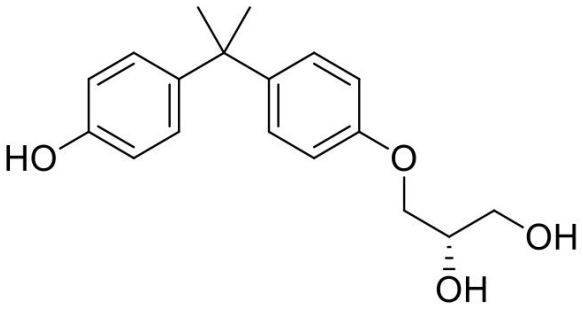
General procedure 3 for epoxide formation

To a solution of starting material (1 eq) and dibutyltin oxide (DBTO, 0.02 eq) in CH_2Cl_2 at 0ºC was added triethylamine (2.2 eq) and tosyl chloride (2.2 eq). The mixture was stirred at 0ºC until no staring material was observed by TLC. Aq. $NaHCO_{3\ sat}$ (5 mL) was added and the mixture was extracted with CH_2Cl_2 (3x5 mL) and dried over $MgSO_4$ to obtain a pale yellow oil.
The oil was dissolved in THF and added via cannula over a THF suspension of NaH (1.5 eq) at 0ºC. After 1h no starting material was observed by TLC. The reaction crude was treated with H_2O (10 mL). THF was removed in vacuo and the aqueous phase extracted with CH_2Cl_2 (3x5 mL), dried over $MgSO_4$ and purified by flash column chromatography on silica·TEA (2,5 % v/v) using hexane/ethyl acetate and increasing the polarity ratio from 100:0 to 70:30.
General procedure 4 for chlorohydrin formation and acetal removal
To a solution of starting material (1 eq) in acetonitrile was added $CeCl_3 \cdot 7H_2O$ (1.4 eq) and the mixture stirred for 18h at reflux. Solvent was removed in vacuo and the crude purified by flash column chromatography on silica gel using CH_2Cl_2/MeOH increasing the polarity ratio from 100:0 to 85:15. Chiral HPLC analysis showed ee > 99 % and dr 95 : 5.

2.3 Synthesis of (2*S*, 20*R*)-EPI-003



2.3.1 (*S*)-3-(4-(2-(4-hydroxyphenyl)propan-2-yl)phenoxy)propane-1,2-diol [**(*S*)-2**]



General procedure 1A was carried out with 1 (400 mg, 1.75 mmol), TEA (122 µL, 0.88 mmol), and (S)-glycidol (144 µL, 2.1 mmol) in EtOH (5.5 mL) to give **(*S*)-2** (256 mg, 49%) as a colorless oil.

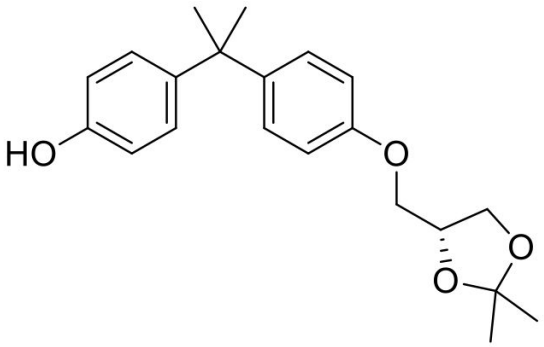
^{1}H-NMR (400 MHz, $CDCl_3$, δ/ppm): 7.15 (m, 2H), 7.09 (m, 2H), 6.81 (m, 2H), 6.72 (m, 2H), 4.09 (m, 1H), 4.03 (m, 1H), 4.01 (m, 1H), 3.84 (apt dd, J = 11.5, 3.5 Hz, 1H), 3.75 (apt dd, J = 11.5, 5.5 Hz, 1H), 1.62 (s, 6H).

^{13}C-NMR (100 MHz, $CDCl_3$, δ/ppm): 156.2 (C), 153.3 (C), 143.9 (C), 143.1 (C), 127.9 (CH), 127.8 (CH), 114.7 (CH), 113.9 (CH), 70.4 (CH), 69.2 (CH_2), 63.7 (CH_2), 41.7 (C), 31.0 (CH_3).

IR (film, v_{max} / cm^{-1}): 3353 (OH), 2962, 2936, 1510, 1246, 1179

HRMS (ES): calculated for $C_{18}H_{22}O_4Na$: 325.14103, found 325.14103

2.3.2 (*R*)-4-(2-(4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)propan-2-yl)phenol [**(*R*)-3**]



General procedure 2 was carried out with **(*S*)-2** (391 mg, 1.29 mmol), PTSA (37 mg, 0.19 mmol) and dimethoxypropane (0.32 µL, 2.58 mmol) in acetone (14 mL) to give **(*R*)-3** (441 mg, 99%) as a pale yellow oil.

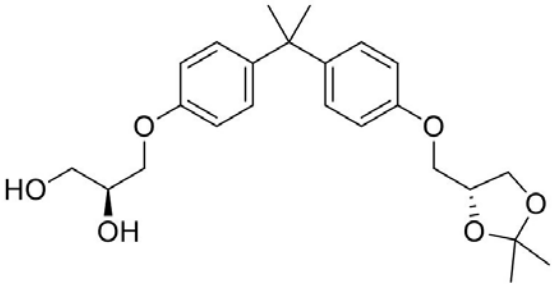
^{1}H-NMR (400 MHz, $CDCl_3$, δ/ppm): 7.13 (m, 2H), 7.08 (m, 2H), 6.81 (m, 2H), 6.73 (m, 2H), 4.48 (apt p, J = 6.0 Hz, 1H), 4.17 (apt dd, J = 8.5, 6.0 Hz, 1H), 4.05 (apt dd, J = 9.5, 6.0 Hz, 1H), 3.92 (m, 2H), 1.63 (s, 6H), 1.47 (s, 3H), 1.41 (s, 3H).

^{13}C-NMR (100 MHz, $CDCl_3$, δ/ppm): 156.3 (C), 153.4 (C), 143.6 (C), 143.1 (C), 127.8 (CH), 127.7 (CH), 114.7 (CH), 113.8 (CH), 109.8 (C), 74.0 (CH), 68.7 (CH_2), 66.9 (CH_2), 41.7 (C), 31.0 (CH_3), 26.8 (CH_3), 25.3 (CH_3).

IR (film, ν_{max} / cm^{-1}): 3393 (OH), 2962, 1510, 1233, 1045, 829

HRMS (ES): calculated for $C_{21}H_{27}O_4$: 343.19039, found 343.19073

2.3.3 (*S*)-3-(4-(2-(4-(((R)-2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)propan-2-yl)phenoxy)propane-1,2-diol [**(*R,S*)-4**]



General procedure 1A was carried out with **(*R*)-3** (233 mg, 0.68 mmol), TEA (47 µL, 0.34 mmol), and (S)-glycidol (60.5 µL, 0.88 mmol) in EtOH (3.5 mL) to give **(*R,S*)-4** (180 mg, 72%) as a pale yellow oil.

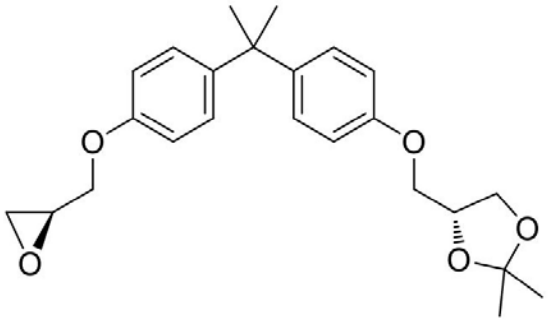
^{1}H-NMR (400 MHz, $CDCl_3$, δ/ppm): 7.11 (m, 4H), 6.80 (m, 4H), 4.45 (apt p, J = 6.0 Hz, 1H), 4.15 (apt dd, J = 8.5, 6.5 Hz, 1H), 4.08 (m, 1H), 4.04 – 3.99 (m, 3H), 3.89 (m, 2H), 3.83 (apt dd, J = 11.5, 4.0 Hz, 1H), 3.74 (apt dd, J = 11.5, 5.5 Hz, 1H), 3.17 (br, 1H), 2.74 (br, 1H), 1.62 (s, 6H), 1.45 (s, 3H), 1.39 (s, 3H).

^{13}C-NMR (100 MHz, $CDCl_3$, δ/ppm): 156.3 (C), 156.2 (C), 143.7 (C), 143.4 (C), 127.75 (CH), 127.70 (CH), 113.8 (CH), 109.7 (C), 74.0 (CH), 70.4 (CH), 69.1 (CH_2), 68.7 (CH_2), 66.9 (CH_2), 63.6 (CH_2), 41.7 (C), 31.0 (CH_3), 26.7 (CH_3), 25.3 (CH_3).

IR (film, ν_{max} / cm^{-1}): 3368 (OH), 2968, 2931, 1607, 1510, 1249, 1042, 828

HRMS (ES): calculated for $C_{24}H_{33}O_6$: 417.22717, found 417.22763

2.3.4
(*R*)-2,2-dimethyl-4-((4-(2-(4-((S)-oxiran-2-ylmethoxy)phenyl)propan-2-yl)phenoxy)methyl)-1,3-dioxolane [**(*R,S*)-5**]



General procedure 3 was carried out with **(*R,S*)-4** (200 mg, 0.48 mmol), DBTO (4 mg, 3%), TEA (84 µL, 0.6 mmol), NaH (18 mg, 0.72 mmol) and TsCl (98 mg, 0.51 mmol) in CH_2Cl_2 (4 mL) and THF (6 mL) to give **(*R,S*)-5** (139 mg, 73%) as a colorless oil.

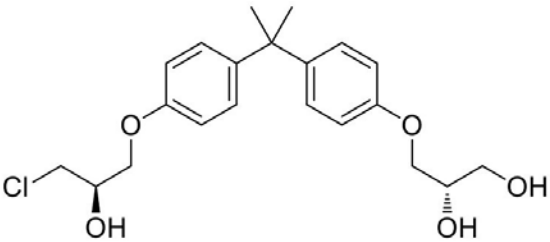
^{1}H-NMR (400 MHz, $CDCl_3$, δ/ppm): 7.13 (m, 4H), 6.82 (m, 4H), 4.46 (apt p, J = 6.0 Hz, 1H), 4.20 – 4.12 (m, 2H), 4.03 (apt dd, J = 9.5, 5.5 Hz, 1H), 3.94 (apt dd, J = 11.0, 5.5 Hz, 1H), 3.92 – 3.86 (m, 2H), 3.34 (m, , 1H), 2.89 (apt dd, J = 5.0, 4.0 Hz, 1H), 2.74 (apt dd, J = 5.0, 3.0 Hz, 1H), 1.63 (s, 6H), 1.46 (s, 3H), 1.40 (s, 3H).

^{13}C-NMR (100 MHz, $CDCl_3$, δ/ppm): 156.35 (C), 156.30 (C), 143.6 (C), 143.5 (C), 127.75 (CH), 127.70 (CH), 113.9 (CH), 113.8 (CH), 109.6 (C), 74.00 (CH), 68.7 (CH_2), 68.7 (CH_2), 66.9 (CH_2), 50.1 (CH), 44.7 (CH_2), 41.7 (C), 31.0 (CH_3), 26.8 (CH_3), 25.3 (CH_3).

IR (film, ν_{max} / cm^{-1}): 2966, 1607, 1509, 1248, 1183, 1038, 829

HRMS (ES): calculated for $C_{24}H_{31}O_5$: 399.21660, found 399.21776

2.3.5
(*S*)-3-(4-(2-(4-((R)-3-chloro-2-hydroxypropoxy)phenyl)propan-2-yl)phenoxy)propane-1,2-diol [**(2*S*,20*R*)-EPI-003**]



General procedure 4 was carried out with **(*R,S*)-5** (140 mg, 0.35 mmol) and $CeCl_3 \cdot 7H_2O$ (183 mg, 0.49 mmol) in ACN (6 mL) to give **(2*S*,20*R*)-EPI-003** (105 mg, 76%, ee > 99%, dr. 95 : 5) as colourless oil.

$[\alpha]^{20}_D$ = +2.9 (c=0.24, CH_3OH)

Chiral-HPLC (t_r, min): 16.6

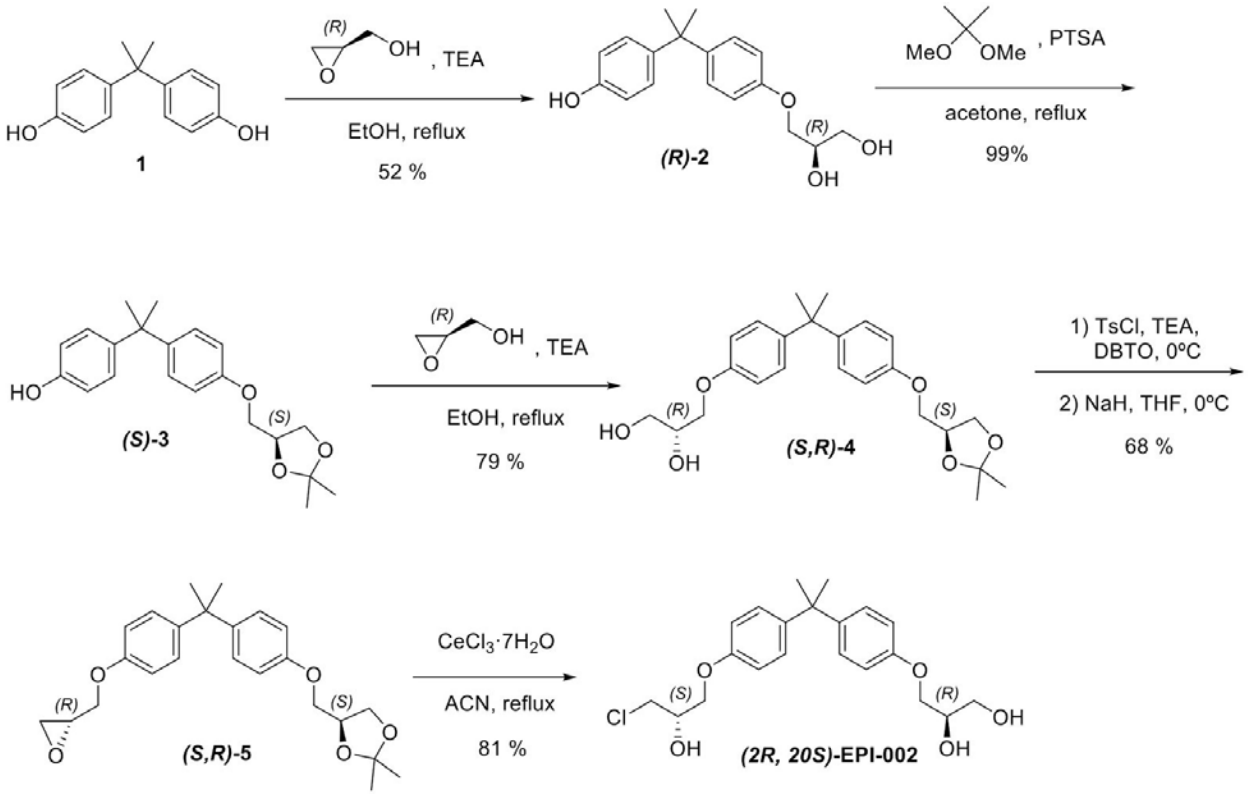
^{1}H-NMR (400 MHz, $CDCl_3$, δ/ppm): 7.13 (m, 4H), 6.81 (m, 4H), 4.20 (apt p, J = 5.0 Hz, 1H), 4.08 (m, 1H), 4.06 (apt dd, J = 5.0, 3.5 Hz, 2H), 4.02 (m, 2H), 3.82 (m, 1H), 3.78 (apt dd, J = 11.5, 5.5 Hz, 2H), 3.70 (apt dd, J = 11.5, 5.5 Hz, 1H), 2.90 (br, 1H), 2.79 (br, 1H), 2.41 (br, 1H), 1.63 (s, 6H).

^{13}C-NMR (100 MHz, $CDCl_3$, δ/ppm): 156.2 (C), 156.0 (C), 143.9 (C), 143.7 (C), 127.85 (CH), 127.80 (CH), 113.95 (CH), 113.90 (CH), 70.4 (CH), 69.9 (CH), 69.1 (CH_2), 68.4 (CH_2), 63.6 (CH_2), 45.9 (CH_2), 41.7 (C), 31.0 (CH_3).

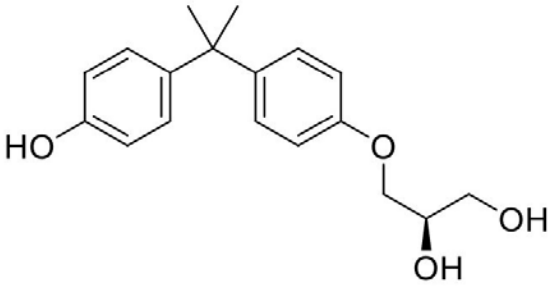
IR (film, ν_{max} / cm^{-1}): 3365 (OH), 2964, 1509, 1247, 1182, 1041, 828

HRMS (ES): calculated for $C_{21}H_{27}ClO_5Na$: 417.14392, found 417.14405

2.4 Synthesis of (2*R*, 20*S*)-EPI-002



2.4.1 (*R*)-3-(4-(2-(4-hydroxyphenyl)propan-2-yl)phenoxy)propane-1,2-diol [**(*R*)-2**]

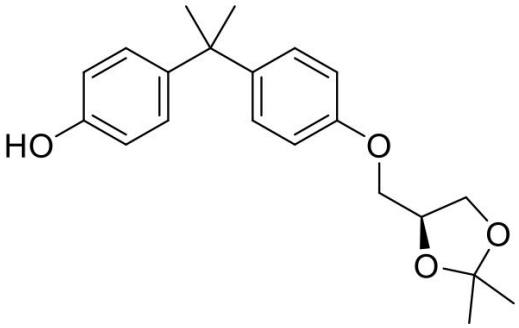


General procedure 1A was carried out with **1** (224 mg, 0.98 mmol), TEA (68 µL, 0.49 mmol), and (R)-glycidol (70.5 µL, 1.0 mmol) in EtOH (3 mL) to give **(*R*)-2** (152 mg, 52%) as a colorless oil.

General procedure 1B was carried out with 1 (298 mg, 1.3 mmol), Cs_2CO_3 (467 mg, 1.43 mmol), and (R)-glycidol (107 µL, 1.57 mmol) in MeOH (6 mL) to give **(*R*)-2** (193 mg, 48%) as a colorless oil.

Spectroscopic data is consistent with **(*S*)-2.**

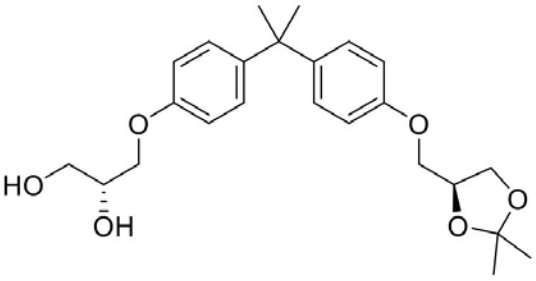
2.4.2 (*S*)-4-(2-(4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)propan-2-yl)phenol [**(*S*-3)**]



General procedure 2 was carried out with **(*R*)-2** (264 mg, 0.87 mmol), PTSA (25 mg, 0.13 mmol) and dimethoxypropane (220 µL, 1.74 mmol) in acetone (12 mL) to give **(*S*)-3** (44 mg, 99%) as a colorless oil.

Spectroscopic data is consistent with **(*R*)-3**.

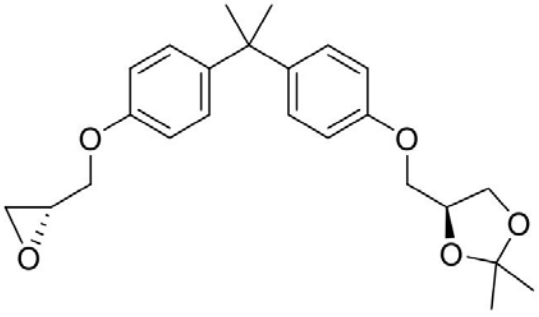
2.4.3
(*R*)-3-(4-(2-(4-(((S)-2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)propan-2-yl)phenoxy)propane-1,2-diol [**(*S,R*)-4**]



General procedure 1A was carried out with **(*S*)-3** (52 mg, 0.15 mmol), TEA (11 µL, 0.08 mmol), and (R)-glycidol (11 µL, 0.17 mmol) in EtOH (2 mL) to give **(*S,R*)-4** (49 mg, 79%) as pale colourless oil. General procedure 1B was carried out with **(*S*)-3** (76 mg, 0.22 mmol), Cs_2CO_3 (108 mg, 0.33 mmol), and (R)-glycidol (30 µL, 0.44 mmol) in MeOH (3 mL) to give **ent-4** (21 mg, 23%) as a colorless oil.

Spectroscopic data is consistent with **(*R,S*)-4**.

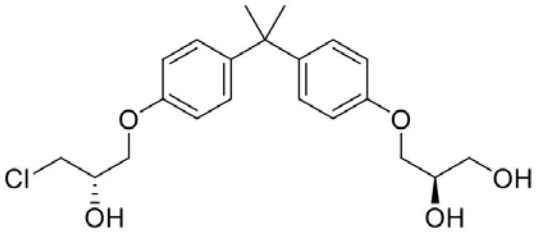
2.4.4
(*S*)-2,2-dimethyl-4-((4-(2-(4-((R)-oxiran-2-ylmethoxy)phenyl)propan-2-yl)phenoxy)methyl)-1,3-dioxolane [**(*S,R*)-5**]



General procedure 3 was carried out with **(*S,R*)-4** (205 mg, 0.49 mmol), DBTO (4 mg, 3%), TEA (76 µL, 1.1 mmol), TsCl (104 mg, 1.1 mmol) and NaH (18 mg, 0.72 mmol) in CH_2Cl_2 (5 mL) and THF (3 mL) to give **(*S,R*)-5** (134 mg, 68%) as a colorless oil.

Spectroscopic data is consistent with **(*R,S*)-5**.

2.4.5
(*R*)-3-(4-(2-(4-((S)-3-chloro-2-hydroxypropoxy)phenyl)propan-2-yl)phenoxy)propane-1,2-diol
[**(2*R*,20*S*)-EPI-002**]



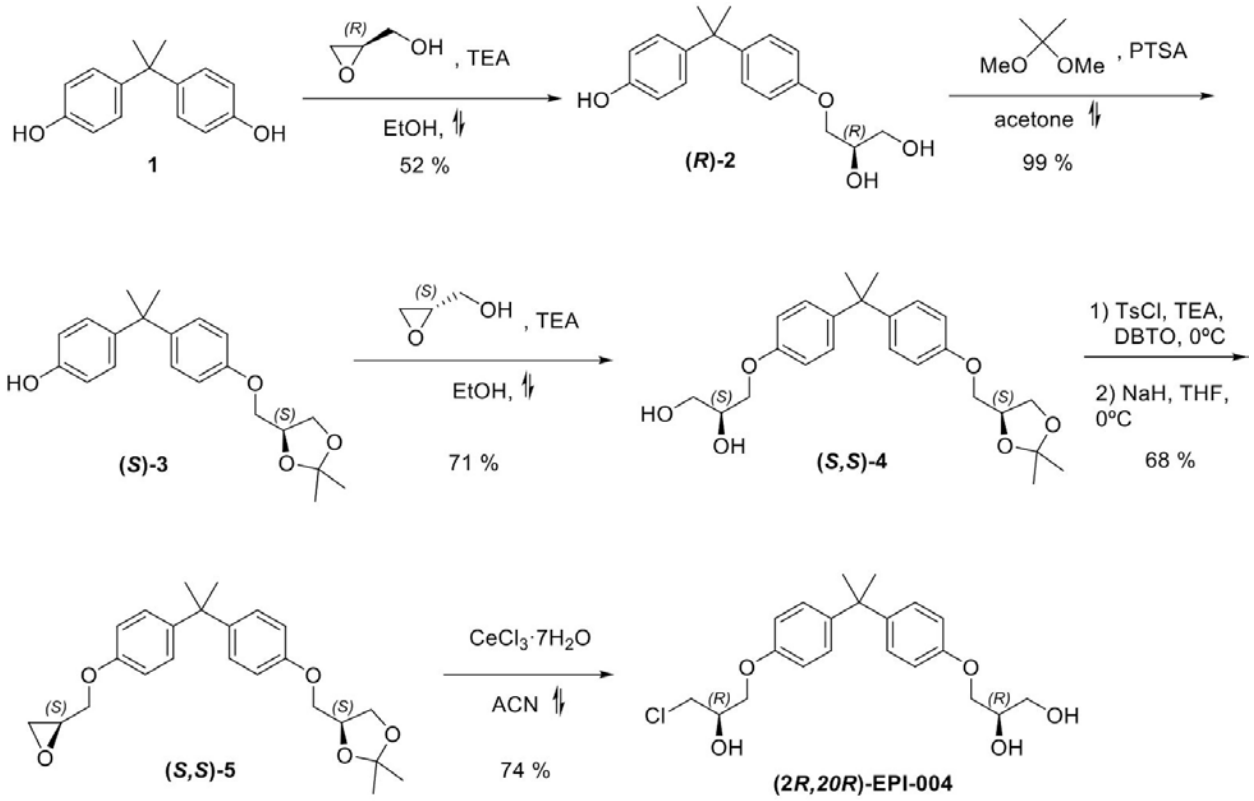
General procedure 4 was carried out with **(*S,R*)-5** (86 mg, 0.21 mmol) and $CeCl_3 \cdot 7H_2O$ (112 mg, 0.3 mmol) in ACN (3 mL) to give **(2*R*,20*S*)-EPI-002** (69 mg, 81%, ee > 99%, dr. 90 : 10) as a colorless oil.

$[\alpha]^{20}_D$ = -4.5 (c=0.21, CH_3OH)

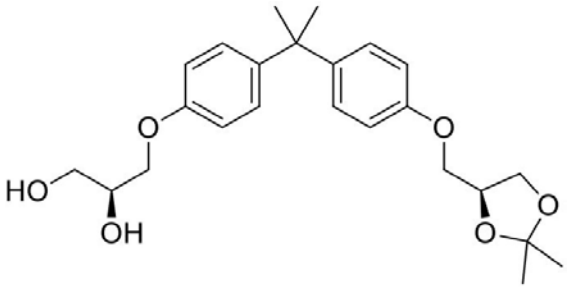
Chiral-HPLC (t_r , min): 24.2

Spectroscopic data is consistent with **(2*S*,20*R*)-EPI-003**.

2.5 Synthesis of (2*R*,20*R*)-EPI-004



2.5.1
(*S*)-3-(4-(2-(4-(((S)-2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)propan-2-yl)phenoxy)propane-1,2-diol [**(*S,S*)-4**]



General procedure 1A was carried out with **(*S*)-3** (121 mg, 0.35 mmol), TEA (24.5 µL, 0.18 mmol), and (S)-glycidol (31 µL, 0.46 mmol) in EtOH (2 mL) to give **(*S,S*)-4** (105 mg, 71%) as a pale yellow oil.

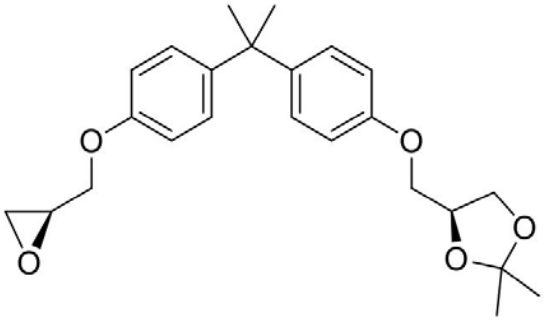
^{1}H-NMR (400 MHz, $CDCl_3$, δ/ppm): 7.13 (m, 4H), 6.81 (m, 4H), 4.45 (apt p, J = 6.0 Hz, 1H), 4.15 (apt dd, J = 8.5, 6.0 Hz, 1H), 4.08 (m, 1H), 4.06 – 4.01 (m, 3H), 3.90 (m, 2H), 3.83 (apt dd, J = 11.5, 4.0 Hz, 1H), 3.74 (apt dd, J = 11.5, 5.5 Hz, 1H), 2.65 (br, 1H), 2.10 (br, 1H), 1.63 (s, 6H), 1.45 (s, 3H), 1.39 (s, 3H).

^{13}C-NMR (100 MHz, $CDCl_3$, δ/ppm): 156.4 (C), 156.2 (C), 143.8 (C), 143.5 (C), 127.8 (CH), 127.7 (CH), 113.9 (CH), 109.7 (C), 74.0 (CH), 70.3 (CH), 69.2 (CH_2), 68.8 (CH_2), 66.9 (CH_2), 63.7 (CH_2), 41.7 (C), 31.0 (CH_3), 26.8 (CH_3), 25.4 (CH_3).

IR (film, v_{max} / cm^{-1}): 3404 (OH), 2975, 2932, 1510, 1249, 1045, 829

HRMS (ES): calculated for $C_{24}H_{33}O_6$: 417.22717, found 417.22689

2.5.2
(*S*)-2,2-dimethyl-4-((4-(2-(4-((S)-oxiran-2-ylmethoxy)phenyl)propan-2-yl)phenoxy)methyl)-1,3-dioxolane [**(*S,S*)-5**]



General procedure 3 was carried out with **(*S,S*)-4** (59 mg, 0.14 mmol), DBTO (2 mg, 5%), TEA (22 µL, 0.15 mmol), TsCl (30 mg, 0.15 mmol) and NaH (5 mg, 0.21 mmol) in CH_2Cl_2 (3 mL) and THF (2 mL) to give **(*S,S*)-5** (38 mg, 68%) as a colorless oil.

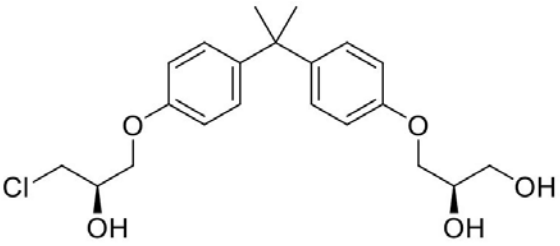
^{1}H-NMR (400 MHz, $CDCl_3$, δ/ppm): 7.12 (m, 4H), 6.81 (m, 4H), 4.46 (apt p, J = 6.0 Hz, 1H), 4.20 – 4.12 (m, 2H), 4.03 (apt dd, J = 9.5, 5.5 Hz, 1H), 3.94 (apt dd, J = 11.0, 5.5 Hz, 1H), 3.92 – 3.86 (m, 2H), 3.33 (m, 1H), 2.89 (apt dd, J = 5.0, 4.0 Hz, 1H), 2.74 (apt dd, J = 5.0, 3.0 Hz, 1H), 1.63 (s, 6H), 1.45 (s, 3H), 1.39 (s, 3H).

^{13}C-NMR (100 MHz, $CDCl_3$, δ/ppm): 156.4 (C), 156.3 (C), 143.6 (C), 143.5 (C), 127.75 (CH), 127.70 (CH), 113.9 (CH), 113.8 (CH), 109.7 (C), 74.00 (CH), 68.75 (CH_2), 68.70 (CH_2), 66.9 (CH_2), 50.2 (CH), 44.8 (CH_2), 41.7 (C), 31.0 (CH_3), 26.8 (CH_3), 25.3 (CH_3).

IR (film, v_{max} / cm^{-1}): 2968, 1508, 1248, 1183, 1038, 829

HRMS (ES): calculated for $C_{24}H_{31}O_5$: 399.21660, found 399.21672

2.5.3
(*R*)-3-(4-(2-(4-((R)-3-chloro-2-hydroxypropoxy)phenyl)propan-2-yl)phenoxy)propane-1,2-diol [**(2*R*,20*R*)-EPI-004**]



General procedure 4 was carried out with **(*S,S*)-5** (30 mg, 0.08 mmol) and $CeCl_3 \cdot 7H_2O$ (39 mg, 0.11 mmol) in ACN (3 mL) to give **(2*R*,20*R*)-EPI-004** (22 mg, 74%, ee > 99%, dr. 97 : 3) as a colorless oil.

$[\alpha]^{20}_D$ = -8.4 (c=0.21, CH_3OH)

Chiral-HPLC (t_r , min): 25.7

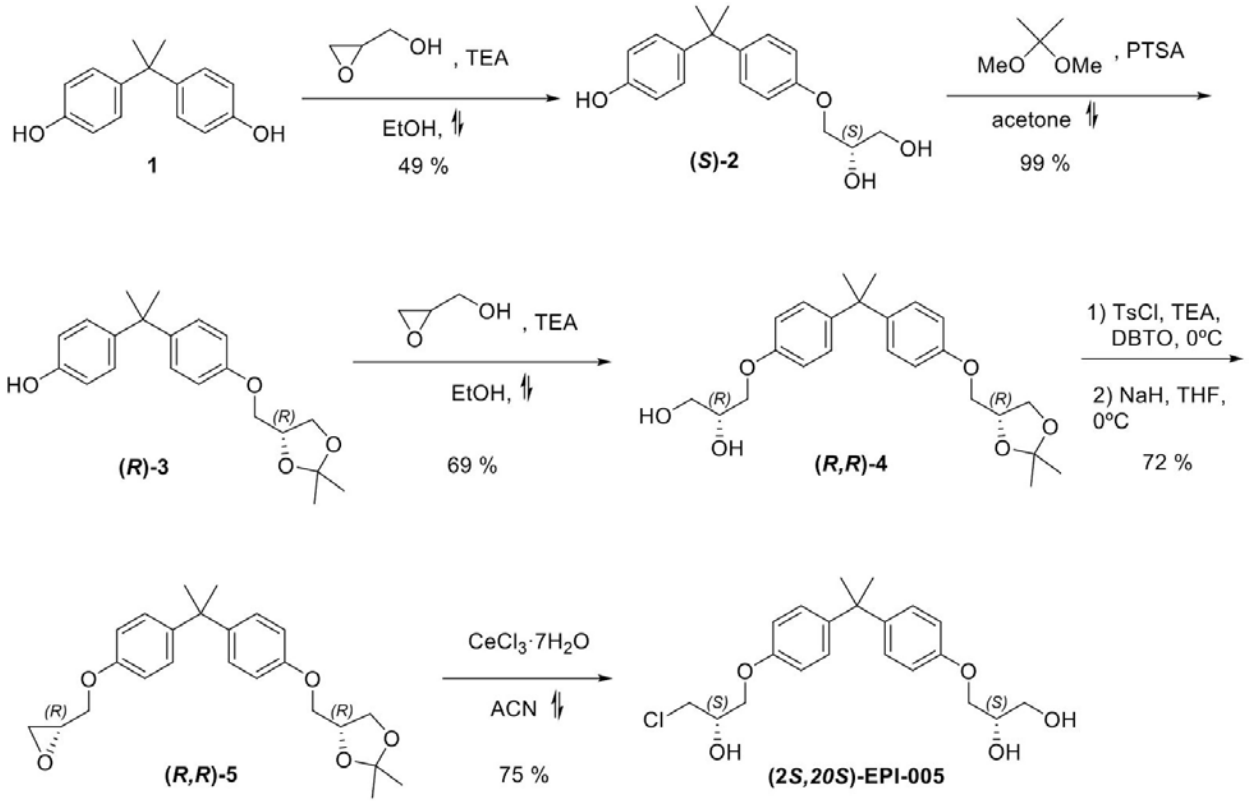
^{1}H-NMR (400 MHz, $CDCl_3$, δ/ppm): 7.13 (m, 4H), 6.81 (m, 4H), 4.20 (m, 1H), 4.08 (m, 1H), 4.06 (m, 2H), 4.02 (m, 2H), 3.82 (m, 1H), 3.78 (apt dd, J = 11.5, 5.5 Hz, 2H), 3.70 (apt dd, J = 11.5, 5.5 Hz, 1H), 2.64 (br, 1H), 2.58 (br, 1H), 2.11 (br, 1H), 1.63 (s, 6H).

^{13}C-NMR (100 MHz, $CDCl_3$, δ/ppm): 156.2 (C), 156.0 (C), 143.9 (C), 143.7 (C), 127.85 (CH), 127.80 (CH), 113.95 (CH), 113.90 (CH), 70.3 (CH), 69.9 (CH), 69.2 (CH_2), 68.4 (CH_2), 63.7 (CH_2), 45.9 (CH_2), 41.7 (C), 31.0 (CH_3).

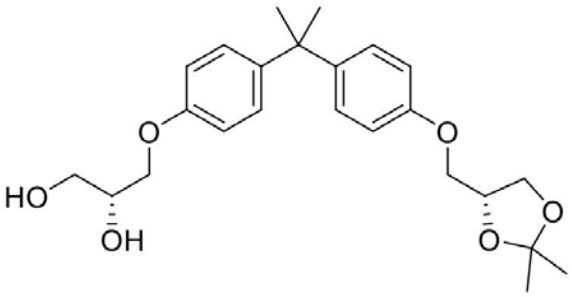
IR (film, ν_{max} / cm^{-1}): 3362 (OH), 2962, 1607, 1509, 1247, 1182, 1041, 828

HRMS *(ES)*: calculated for $C_{21}H_{27}ClO_5Na$: 417.14392, found 417.14424

2.6 Synthesis of (2*S*,20*S*)-EPI-005



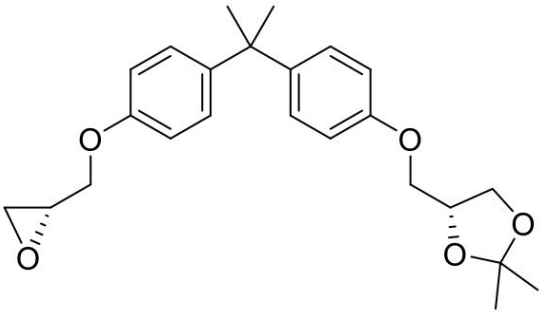
2.6.1
(*R*)-3-(4-(2-(4-(((R)-2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)propan-2-yl)phenoxy)propane-1,2-diol [**(*R,R*)-4**]



General procedure 1A was carried out with **(*R*)-3** (204 mg, 0.60 mmol), TEA (42 µL, 0.30 mmol), and (R)-glycidol (53 µL, 0.77 mmol) in EtOH (2 mL) to give **(*R,R*)-4** (172 mg, 69%) as pale yellow oil.

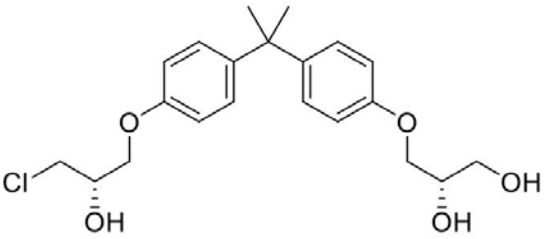
Spectroscopic data is coincident with **(*S,S*)-4**.

2.6.2
(*R*)-2,2-dimethyl-4-((4-(2-(4-((R)-oxiran-2-ylmethoxy)phenyl)propan-2-yl)phenoxy)methyl)-1,3-dioxolane [(R,R)-5]



General procedure 3 was carried out with **(*R,R*)-4** (174 mg, 0.42 mmol), DBTO (3.5 mg, 2%), TEA (76 µL, 0.54 mmol), TsCl (89 mg, 0.46 mmol) and NaH (13 mg, 0.49 mmol) in CH_2Cl_2 (5 mL) and THF (7 mL) to give **(*R,R*)-5** (139 mg, 72%) as a colorless oil. Spectroscopic data is coincident with **(*S,S*)-5**.

2.6.3
(*S*)-3-(4-(2-(4-((S)-3-chloro-2-hydroxypropoxy)phenyl)propan-2-yl)phenoxy)propane-1,2-diol [(2S,20S)-EPI-005]



General procedure 4 was carried out with **(*R,R*)-5** (122 mg, 0.31 mmol) and $CeCl_3 \cdot 7H_2O$ (160 mg, 0.43 mmol) in ACN (5 mL) to give **(2*S*,20*S*)-EPI-005** (90 mg, 75%, ee > 99%, dr. 95 : 5) as a colorless oil.

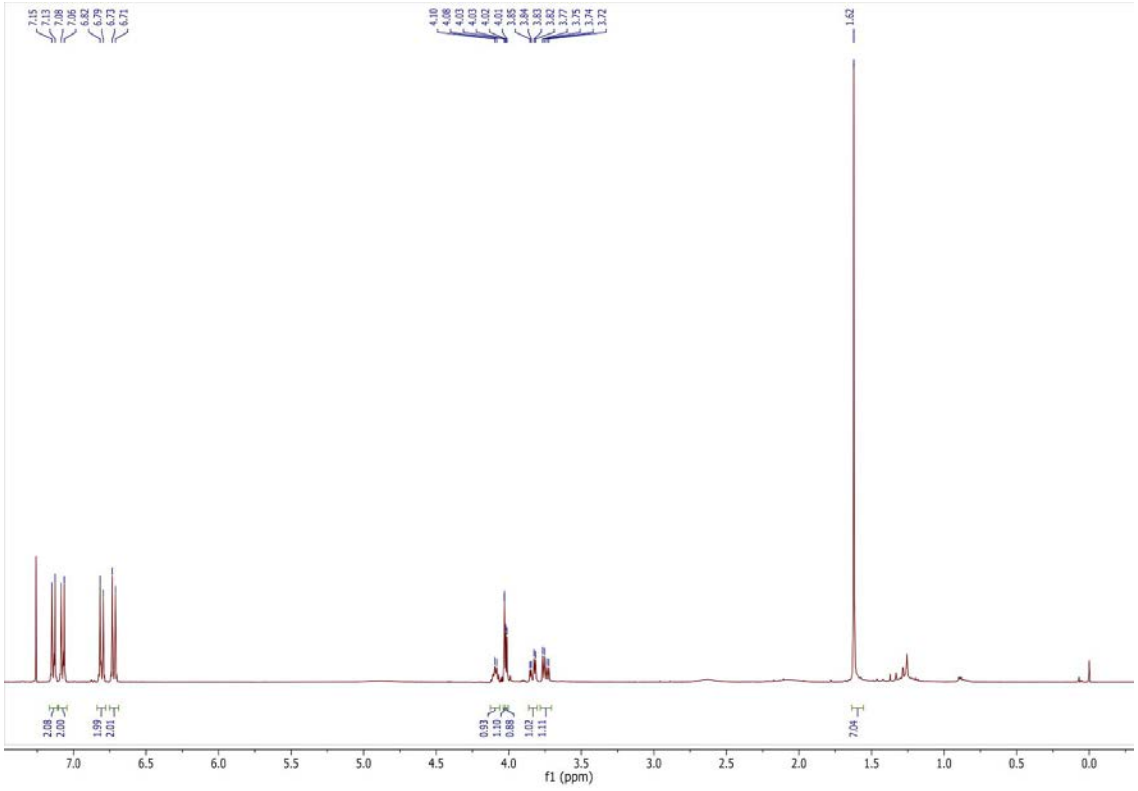
$[\alpha]^{20}_{D}$ = +4.3 (c=0.245, CH_3OH)

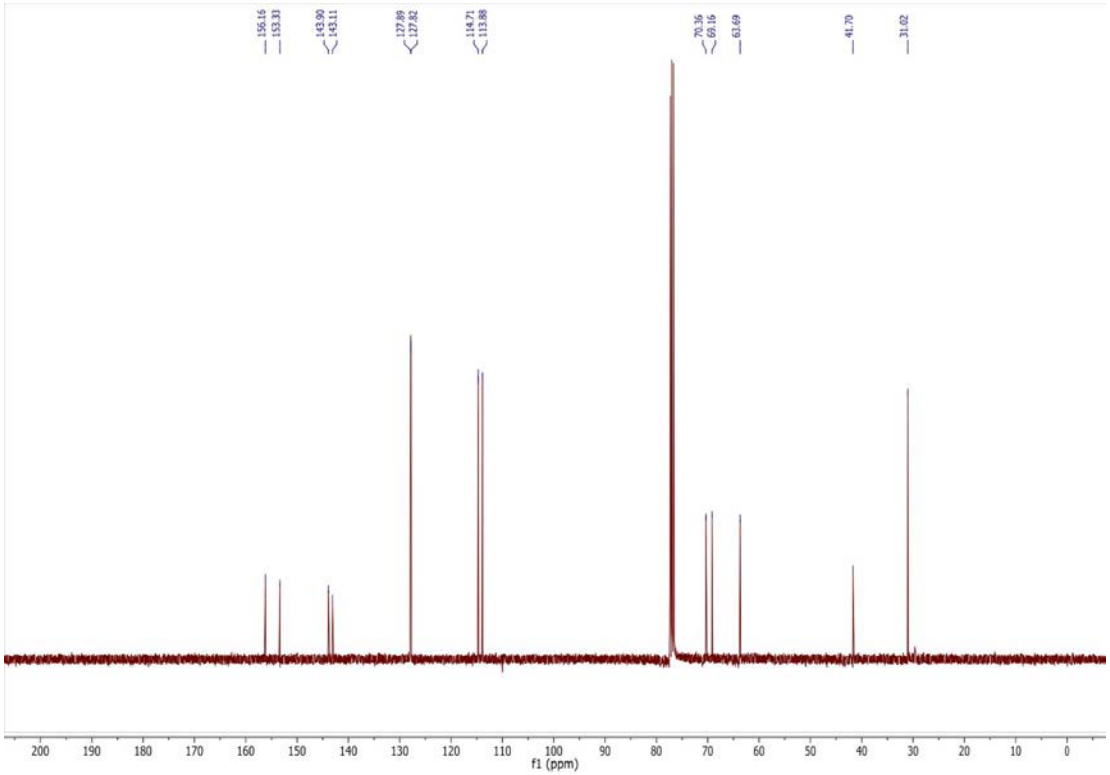
Chiral-HPLC (t_r , min): 15.6

Spectroscopic data is consistent with **(2*R*,20*R*)-EPI-005**.

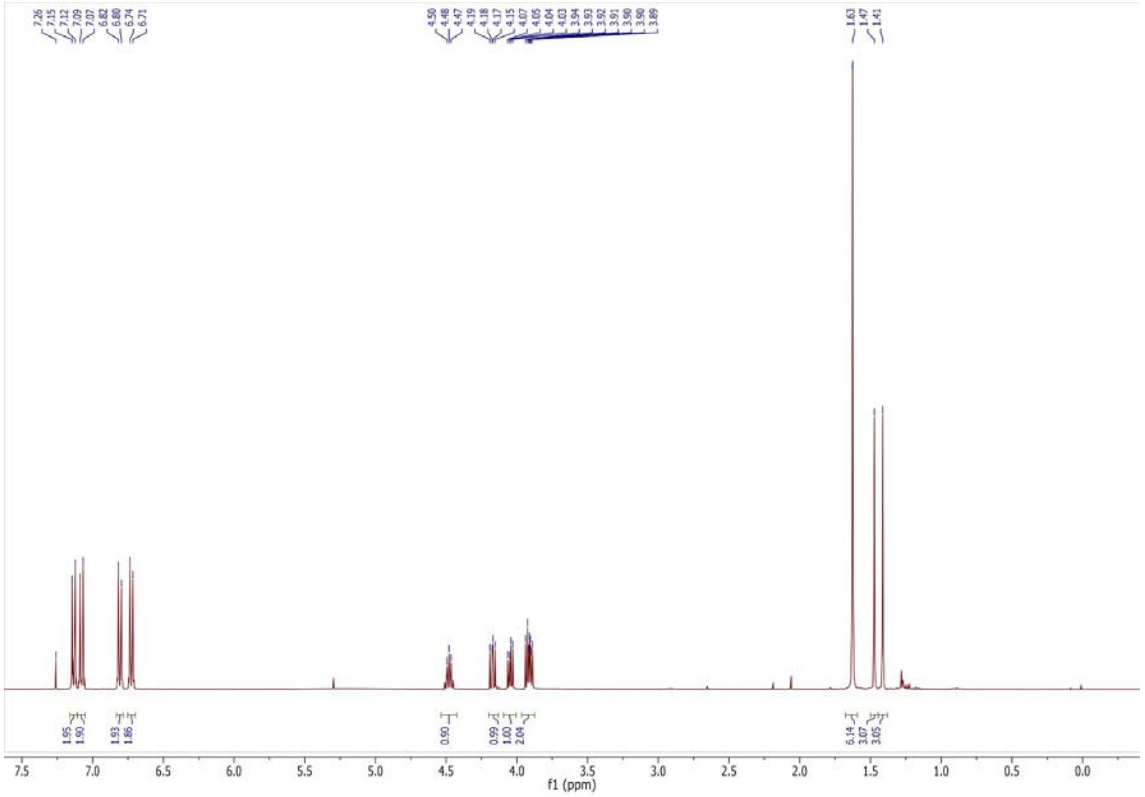
2.7 NMR spectra and HPLC

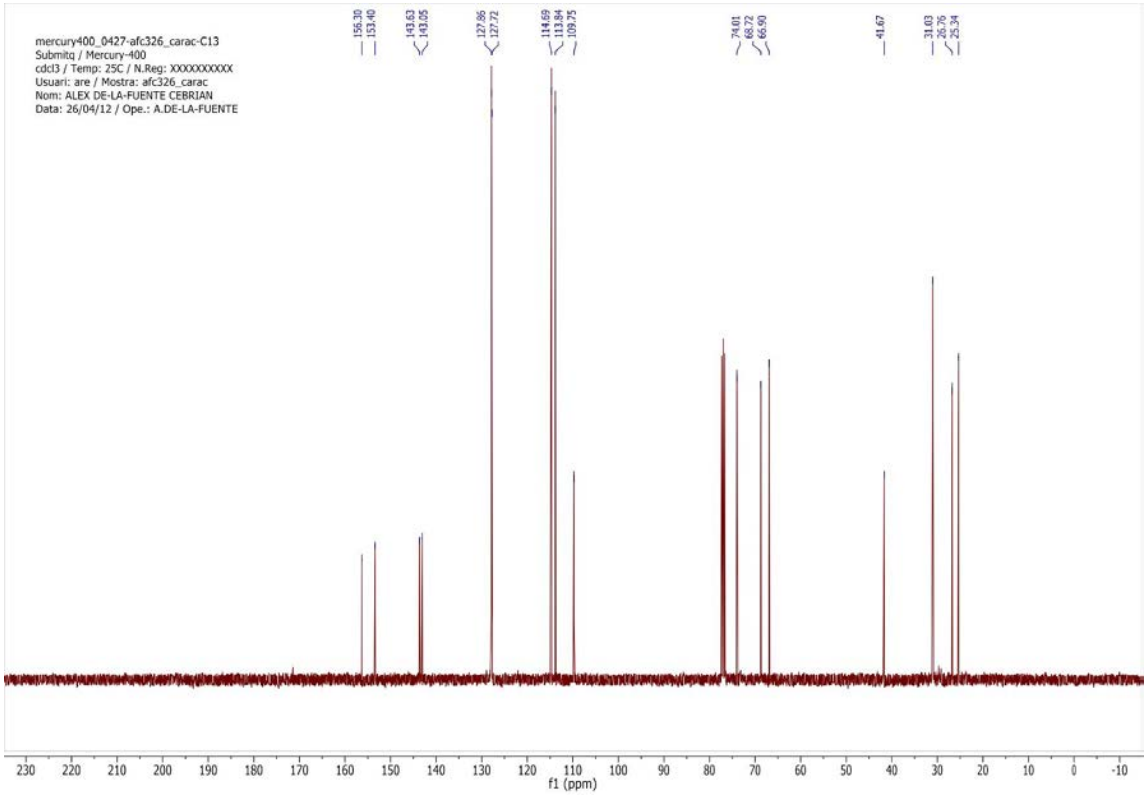
(*S*) or (*R*)-2



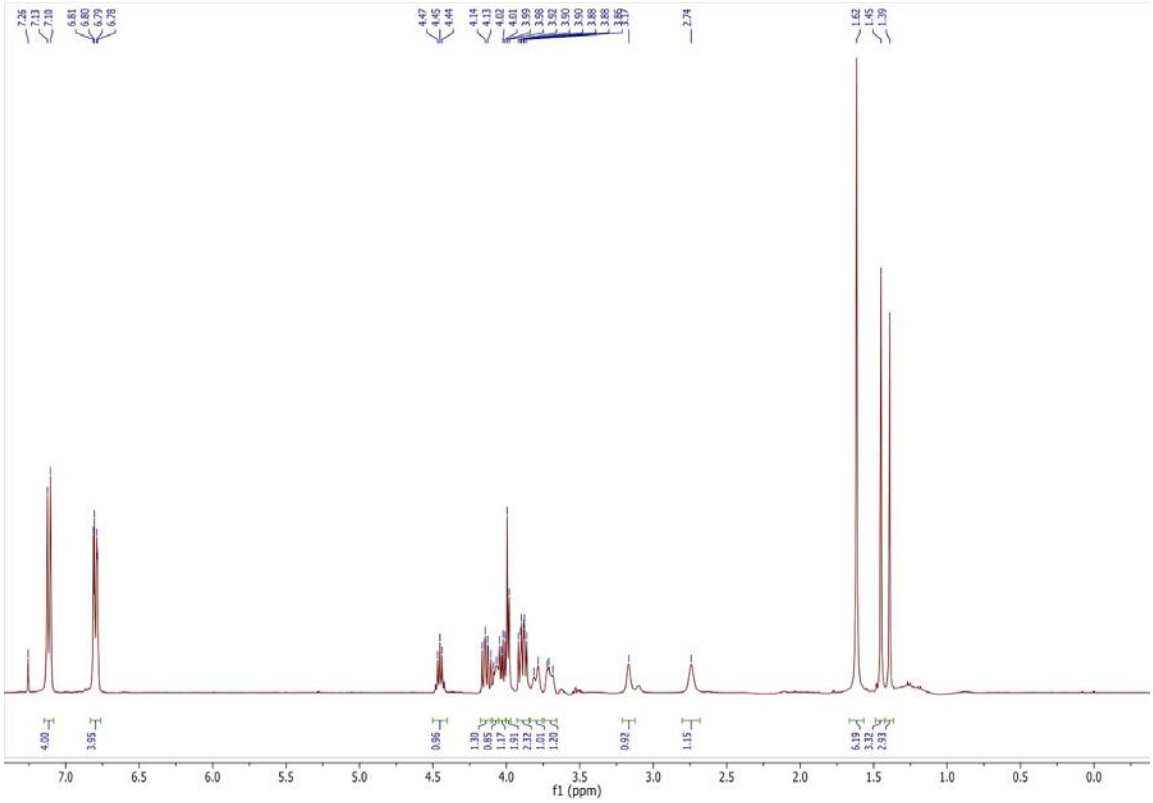


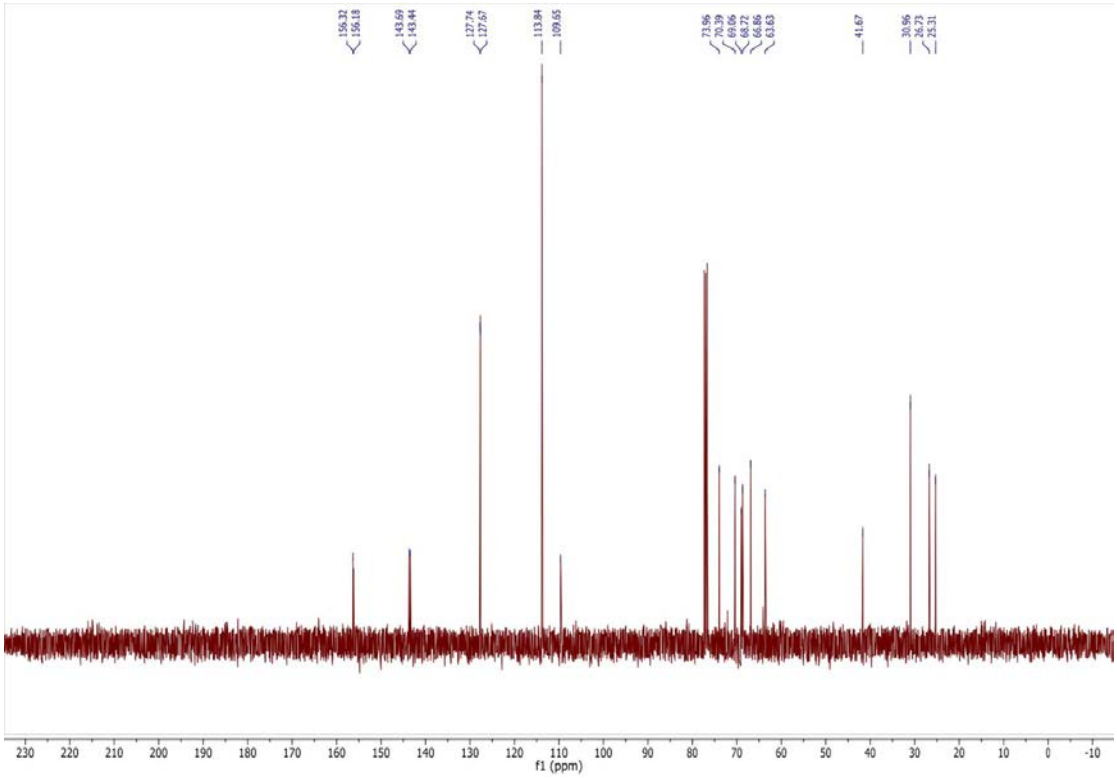
(*R*) or (*S*)-3


7.26
7.15
7.12
7.09
7.07
6.82
6.80
6.74
6.71
4.50
4.48
4.47
4.19
4.18
4.17
4.15
4.07
4.05
4.04
4.03
3.94
3.93
3.92
3.91
3.90
3.90
3.89
1.63
1.47
1.41
1.95
1.90
1.93
1.86
0.90
0.99
1.00
2.04
6.14
3.07
3.05
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

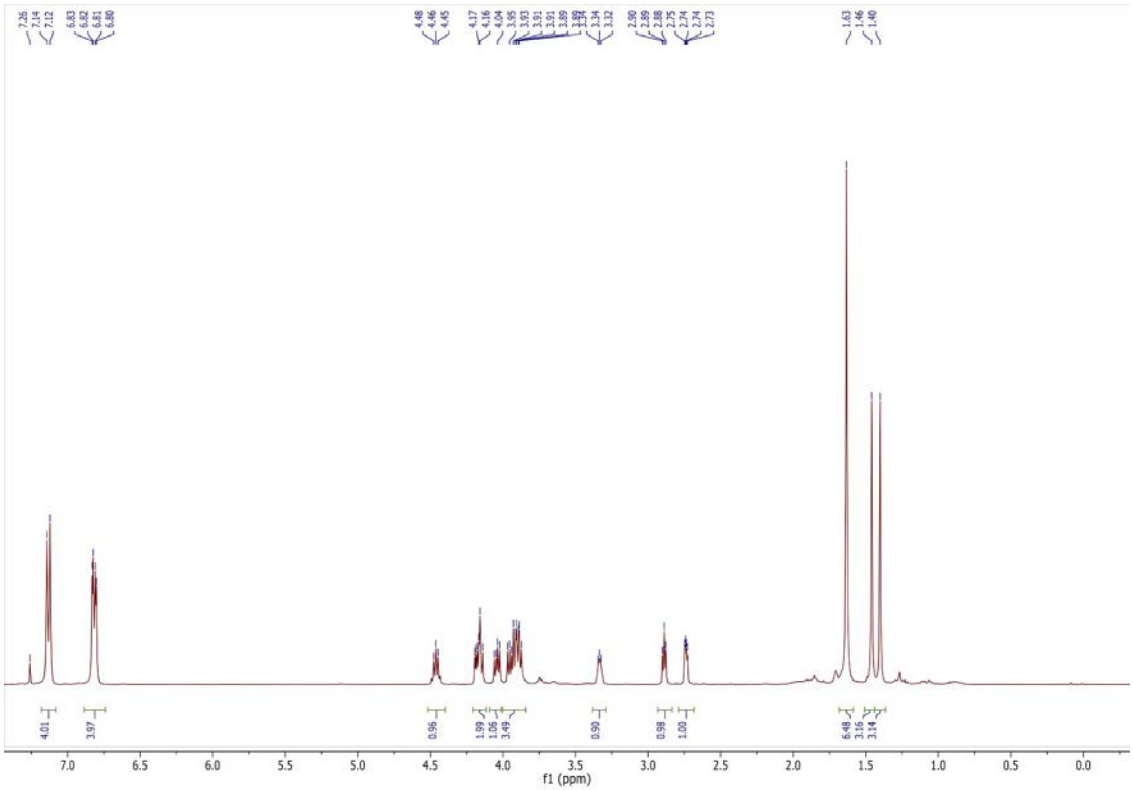

mercury400_0427-afc326_carac-C13
Submitq / Mercury-400
cdcl3 / Temp: 25C / N.Reg: XXXXXXXXXX
Usuari: are / Mostra: afc326_carac
Nom: ALEX DE-LA-FUENTE CEBRIAN
Data: 26/04/12 / Ope.: A.DE-LA-FUENTE
156.30
153.40
143.63
143.05
127.86
127.72
114.69
113.84
109.75
74.01
68.72
66.90
41.67
31.63
26.76
25.34
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

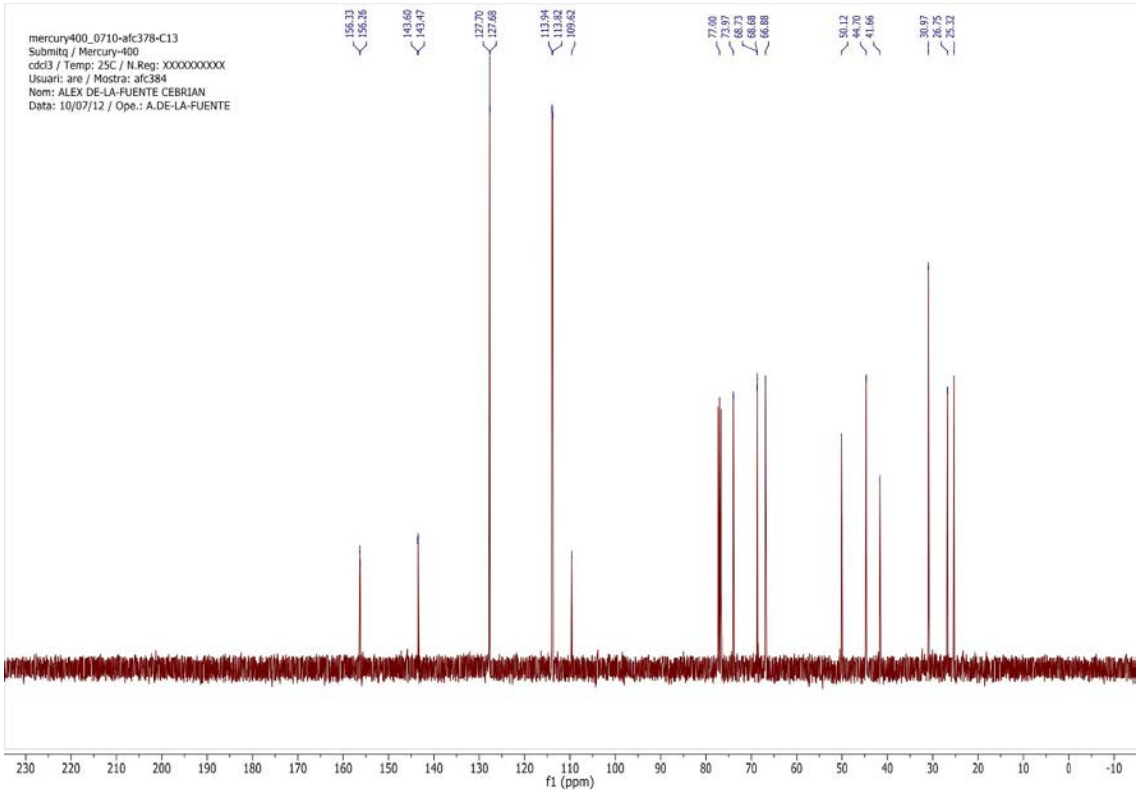
(*R*,*S*) or (*S*,*R*)-4


7.26
7.13
7.10
6.81
6.80
6.79
6.78
4.47
4.45
4.44
4.14
4.13
4.02
4.01
3.99
3.98
3.92
3.90
3.90
3.88
3.86
3.79
2.74
1.62
1.45
1.39
4.00
3.95
0.96
1.30
0.85
1.17
1.91
2.32
1.01
1.20
0.92
1.15
6.19
3.32
2.93
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

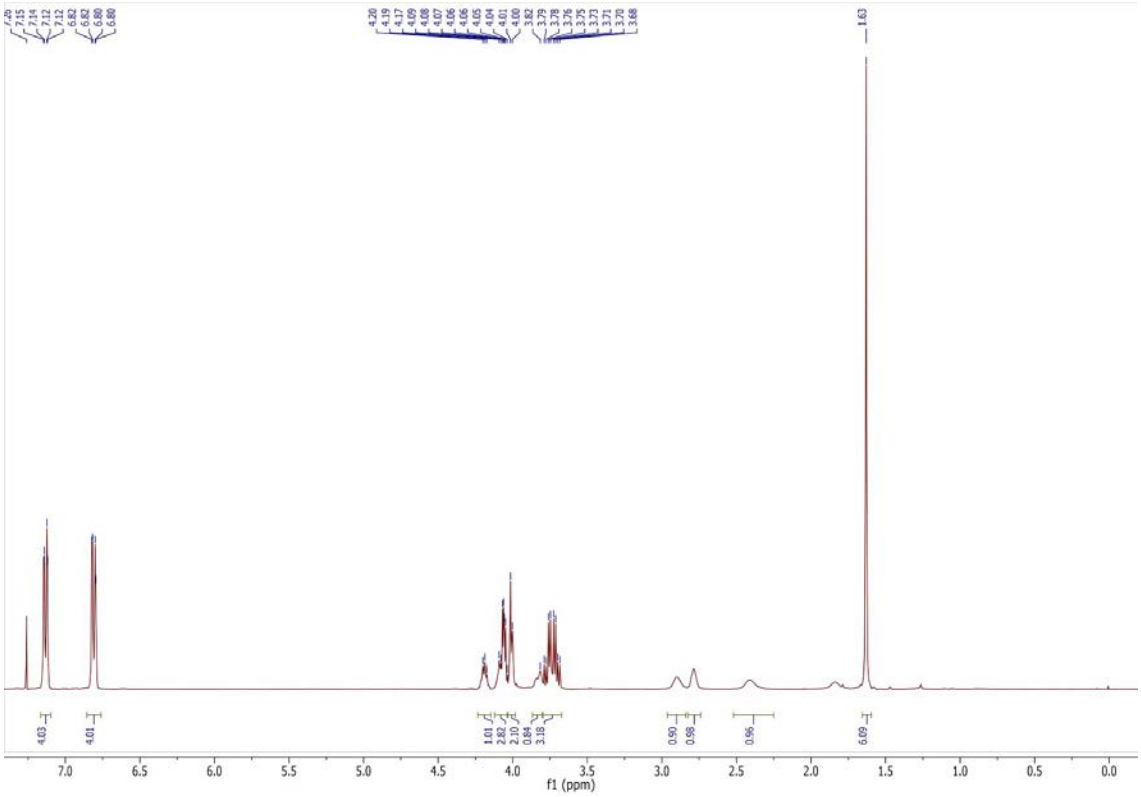

156.32
156.18
143.69
143.44
127.74
127.67
113.84
109.65
71.96
70.39
69.86
69.27
66.86
63.63
41.67
30.96
26.73
25.31
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

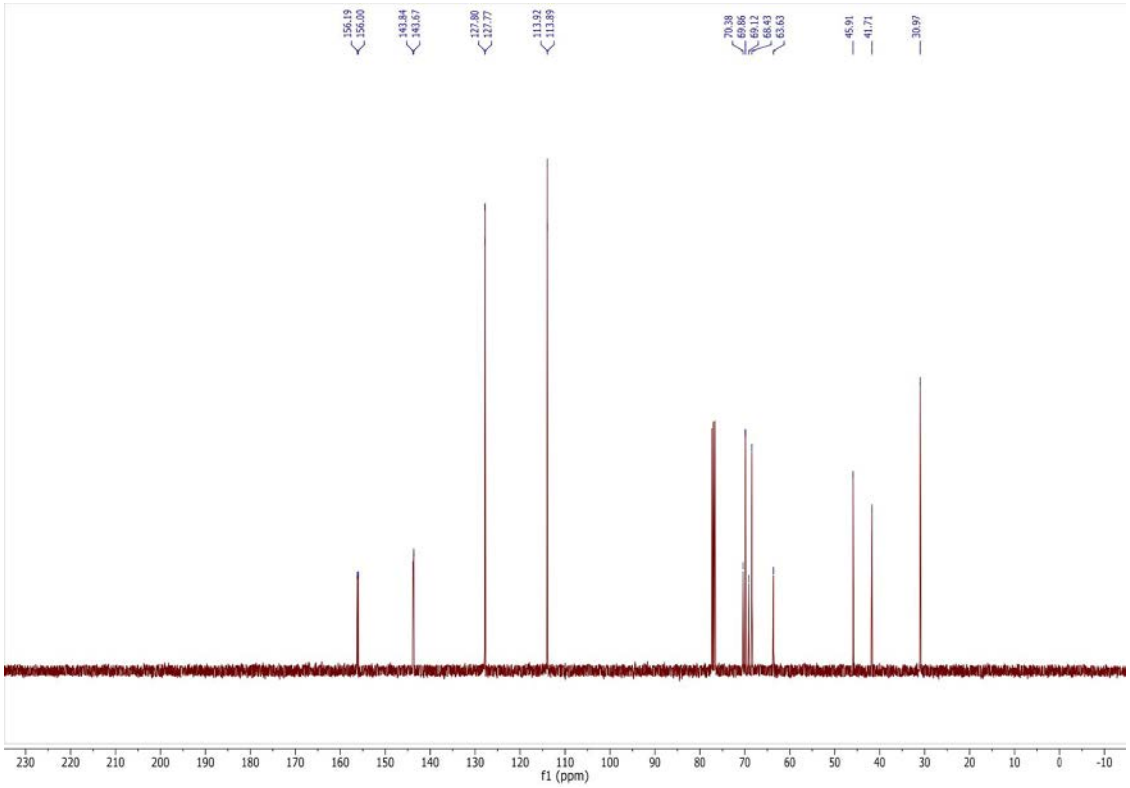
(*R*,*S*) or (*S*,*R*)-5


7.26
7.14
7.12
6.83
6.82
6.81
6.80
4.48
4.46
4.45
4.17
4.16
4.04
3.95
3.93
3.91
3.91
3.89
3.34
3.32
2.90
2.89
2.88
2.75
2.74
2.74
2.73
1.63
1.46
1.40
4.01
3.97
0.96
1.99
1.05
3.49
0.99
0.98
1.00
6.48
3.16
3.14
f1 (ppm)

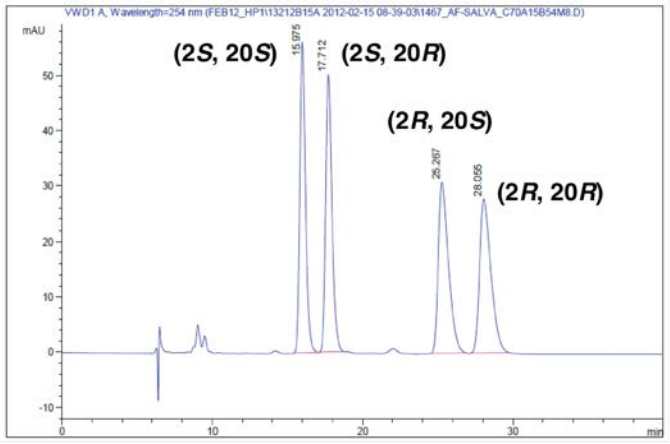

mercury400_0710-afc378-C13
Submitq / Mercury-400
cdcl3 / Temp: 25C / N.Reg: XXXXXXXXXX
Usuari: are / Mostra: afc384
Nom: ALEX DE-LA-FUENTE CEBRIAN
Data: 10/07/12 / Ope.: A.DE-LA-FUENTE
156.33
156.26
143.60
143.47
127.70
127.68
113.94
113.82
109.62
77.00
73.97
68.73
68.68
66.88
50.12
44.70
41.66
30.97
26.75
25.32
f1 (ppm)

(2*S*,20*R*)-EPI-003 or (2*R*,20*S*)-EPI-002

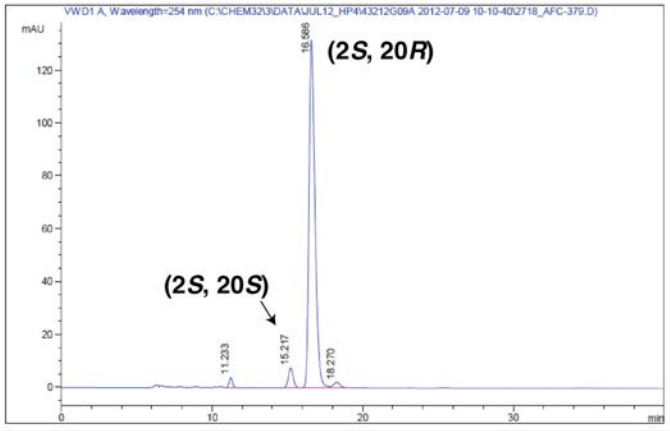




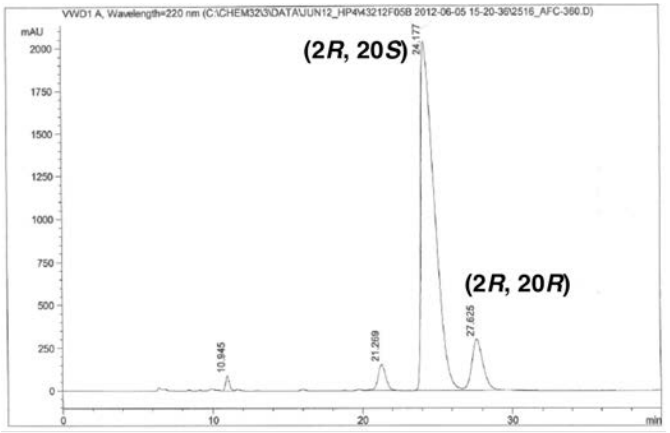
HPLC EP1-001



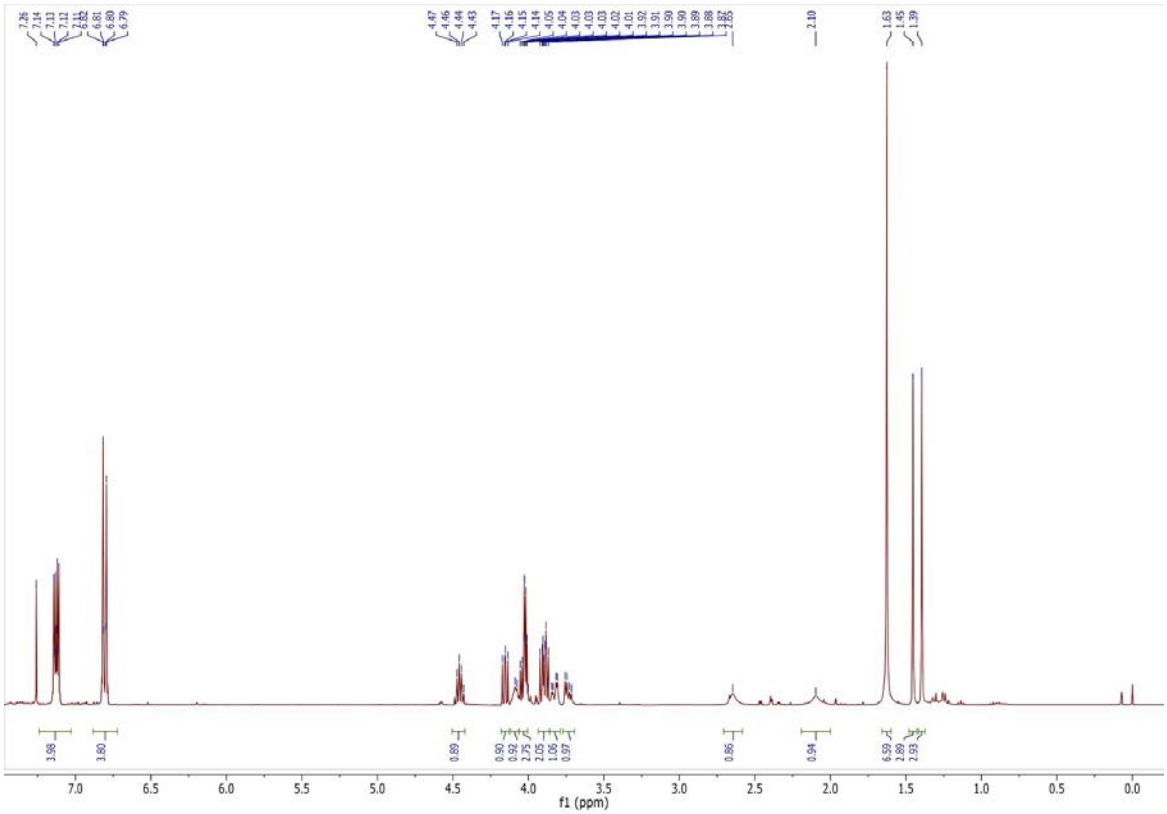
HPLC (2*S*,20*R*)-EPI-003

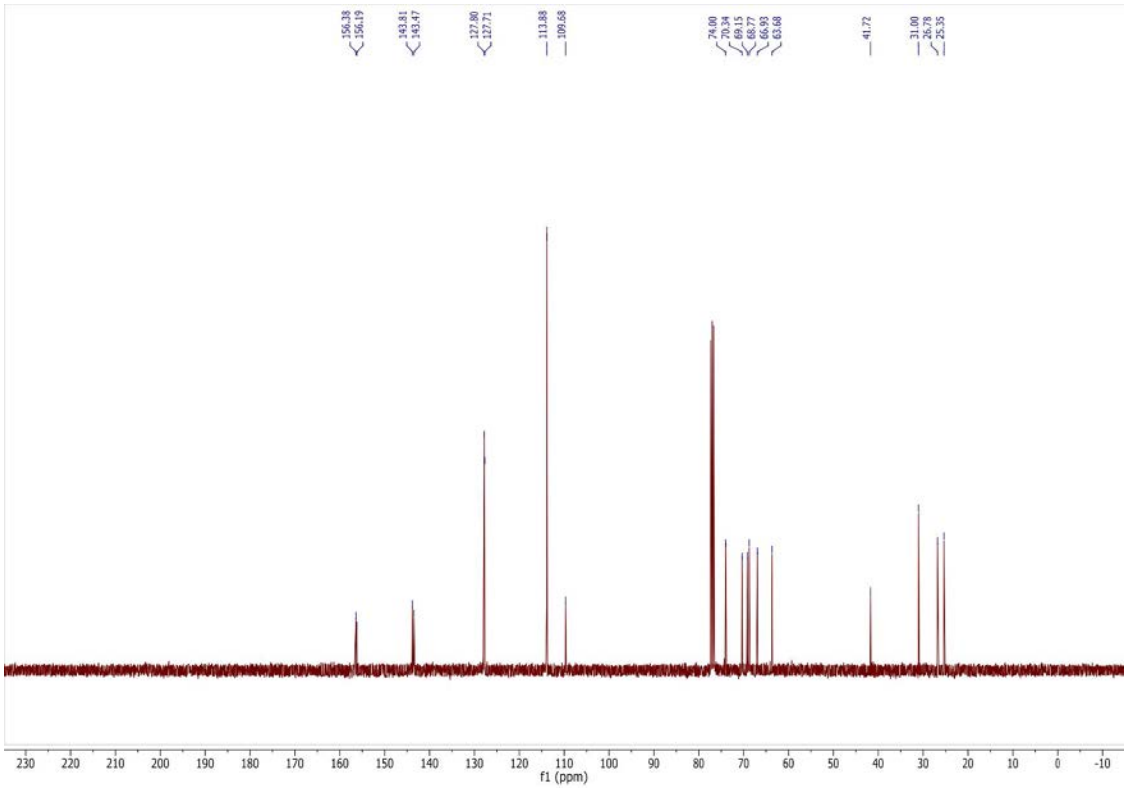


HPLC (2*R*-20*S*)-EPI-002

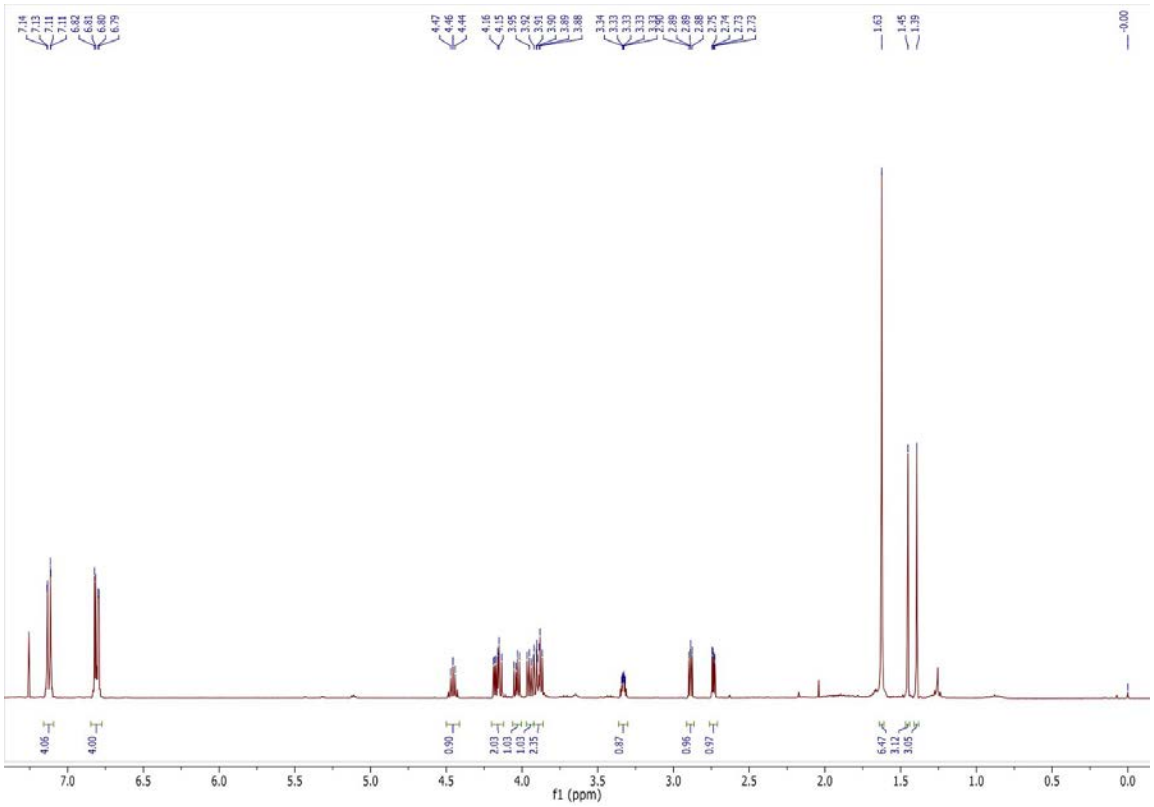


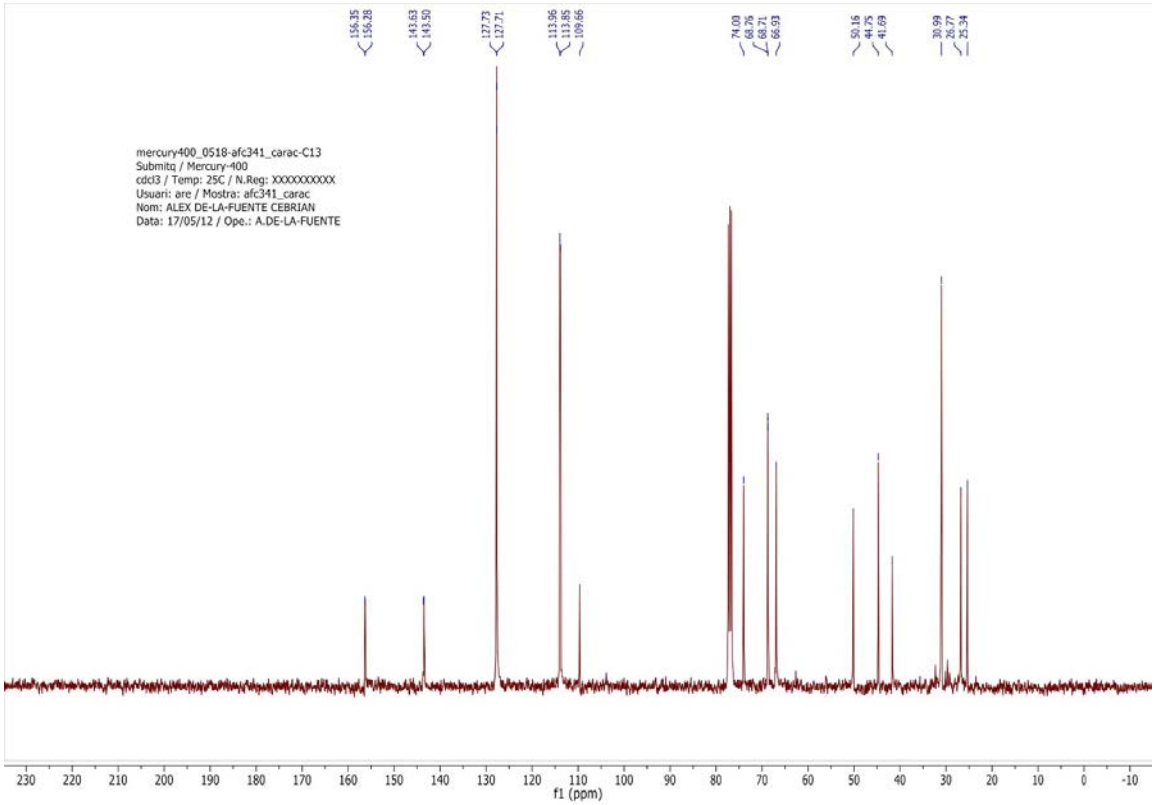
(*S*,*S*) or (*R*,*R*)-4


7.26
7.14
7.13
7.12
6.82
6.81
6.80
6.79
4.47
4.46
4.44
4.43
4.17
4.16
4.15
4.14
4.05
4.04
4.03
4.03
4.02
4.01
3.92
3.91
3.90
3.90
3.89
3.88
2.10
1.63
1.45
1.39
3.98
3.80
0.89
0.90
0.92
2.75
2.05
1.06
0.97
0.86
0.94
6.59
3.89
2.93
f1 (ppm)

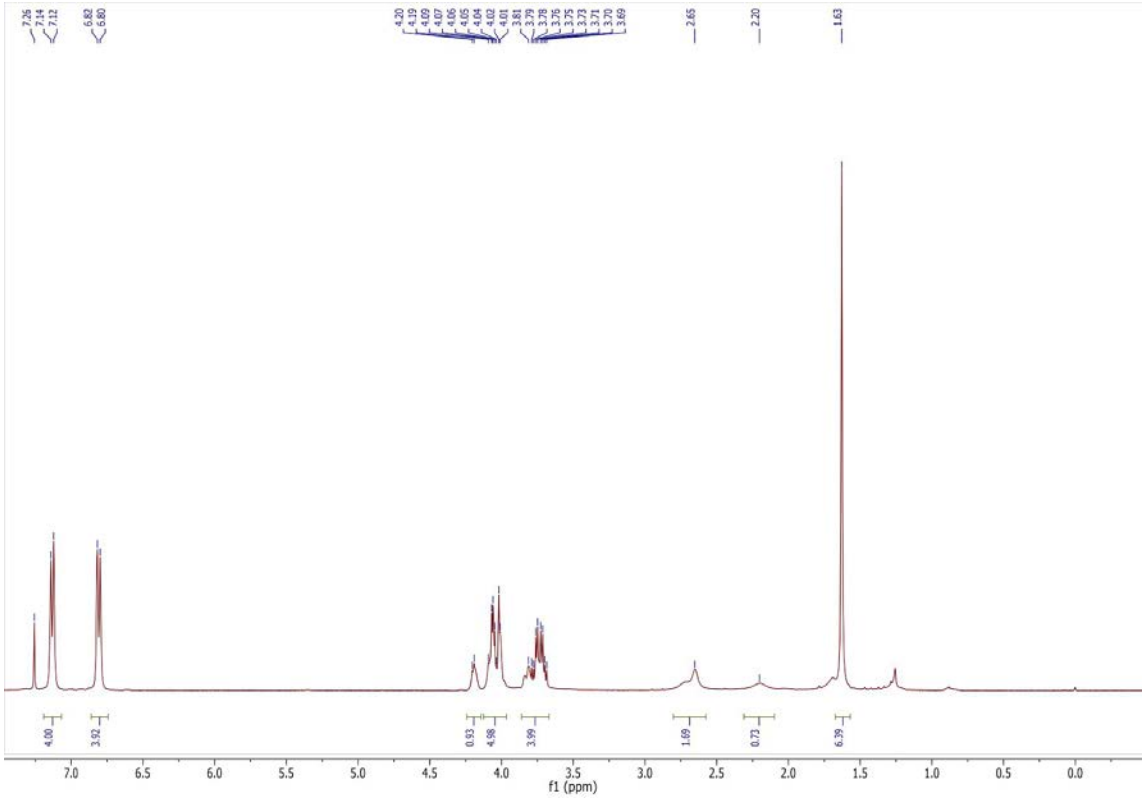

156.38
156.19
143.81
143.47
127.80
127.71
113.88
109.68
74.00
70.34
69.15
68.77
66.93
63.68
41.72
31.00
26.78
25.35
f1 (ppm)

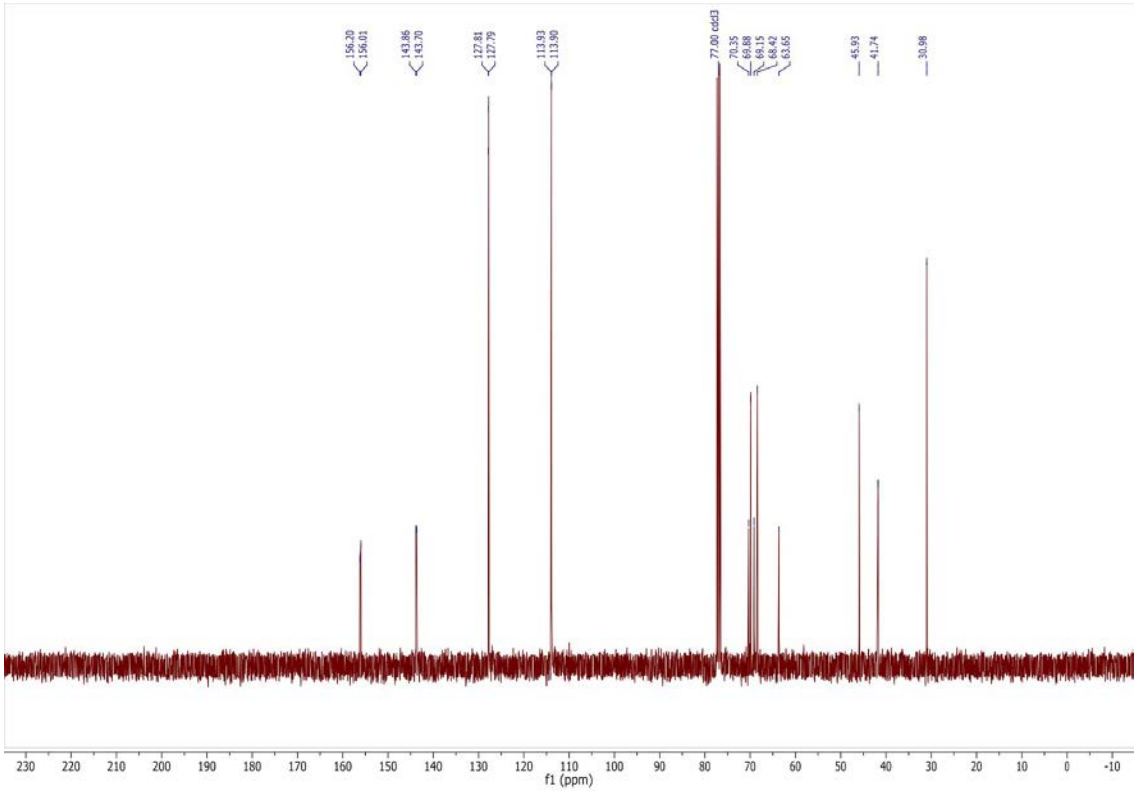
(*S*,*S*) or (*R*,*R*)-5


f1 (ppm)

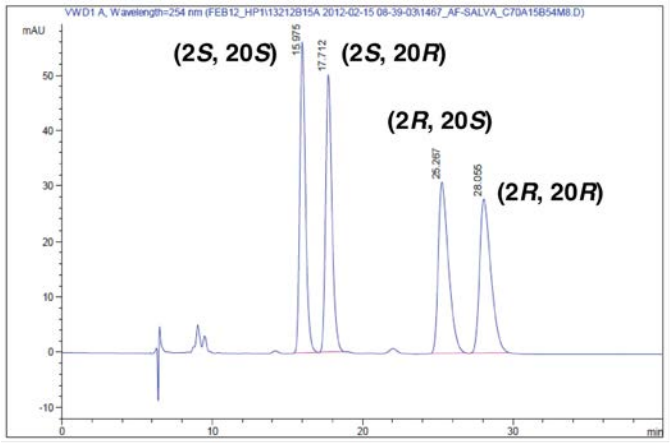

mercury400_0518-afc341_carac-C13
Submitq / Mercury-400
cdcl3 / Temp: 25C / N.Reg: XXXXXXXXXX
Usuari: are / Mostra: afc341_carac
Nom: ALEX DE-LA-FUENTE CEBRIAN
Data: 17/05/12 / Ope.: A.DE-LA-FUENTE
f1 (ppm)

(2*R*,20*R*)-EPI-004 or (2*S*,20*S*)-EPI-005

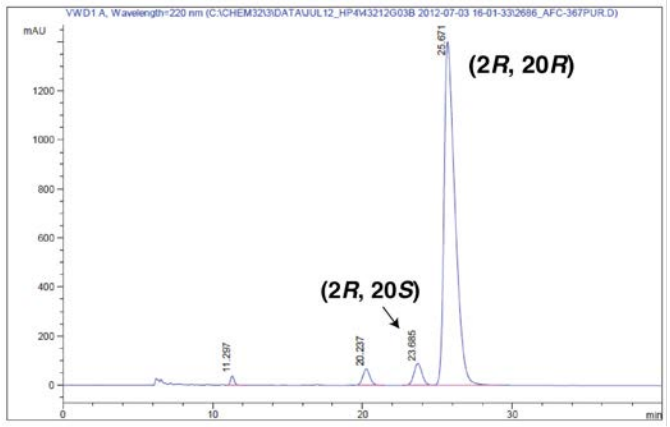




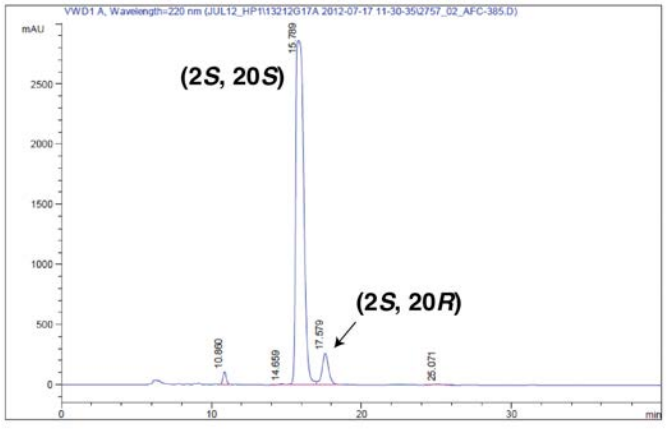
HPLC EP1-001



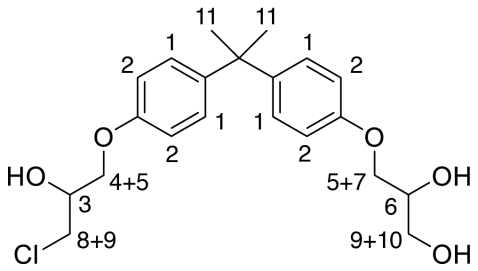
HPLC (2*R*, 20*R*)- EPI-004



HPLC (2*S*, 20*S*)- EPI-005



2.7.1 NMR assignment of EPI-001



^{1}H NMR (800 MHz, Deuterium Oxide) δ = 7.32 – 7.25 (m, 4H, H-1), 6.99 – 6.92 (m, 4H, H-2), 4.28 – 4.23 (m, 1H, H-3), 4.16 (dd, J=10.3, 4.4, 1H, H-4), 4.13 – 4.09 (m, 2H, H-5), 4.08 – 4.04 (m, 1H, H-6), 4.01 (dd, J=10.1, 6.6, 1H, H-7), 3.81 (dd, J=11.7, 4.5, 1H, H-8), 3.77 – 3.72 (m, 2H, H-9), 3.67 (dd, J=11.8, 6.4, 1H, H-10), 1.65 (s, 6H, H-11).

3 - Comparison of the backbone chemical shifts of Tau-5*, AF-1*$_{265\text{-}340}$ and AF-1*

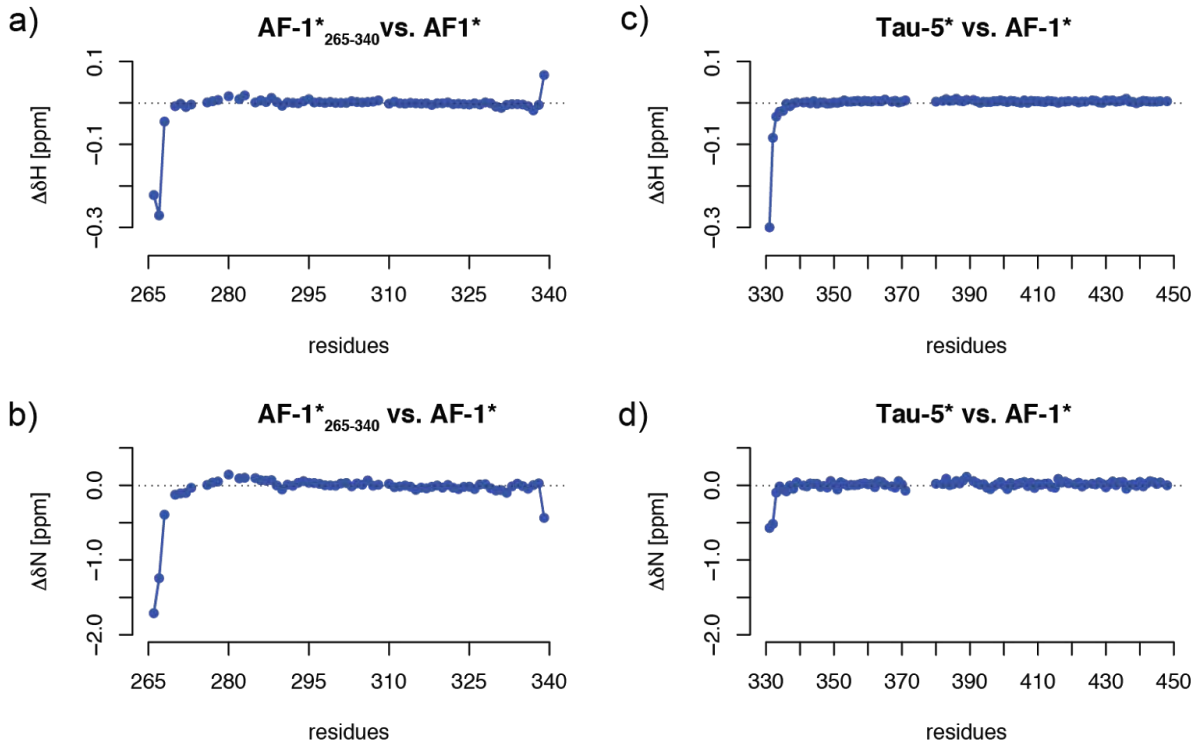


Figure S2: Difference in ^{1}H (a and c) and ^{15}N (b and d) chemical shift between the resonances corresponding to the residues of the NTD in AF-1*$_{265\text{-}340}$ and AF-1* (a and b) and in Tau-5* and AF-1*(c and d).

4 - Sequences of peptides R1, R2 and R3

peptide R1 (AR 341-371): Ac-STLSLYKSGALDEAAAYQSRDYYNFPLALAG-NH_2
peptide R2 (AR 391-414): Ac-LDYGSAWAAAAAQCRYGDLASLHG-NH_2
peptide R3 (AR 426-446): Ac-SAAASSSWHTLFTAEEGQLYG-NH_2

5 - Effect of AF-1* and peptides R1, R2 and R3 on the resonances of EPI-001.

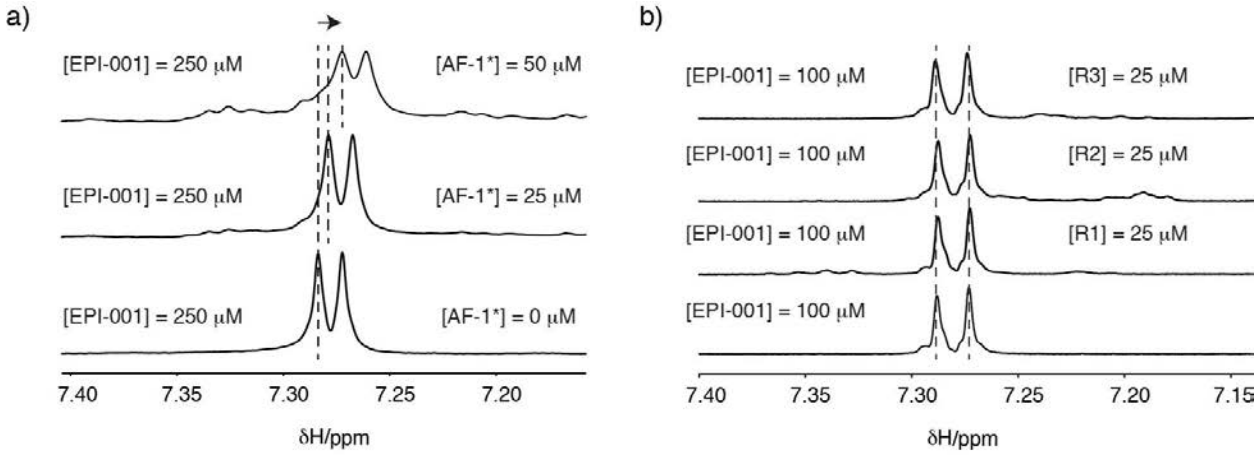


Figure S3: Effect of AF-1* (a) and peptides R1, R2 and R3 (b) on the resonances of EPI-001

6 - Comparison of the changes in the resonances of AF-1* caused by EPI-002, EPI-003, EPI-004 and EPI-005.

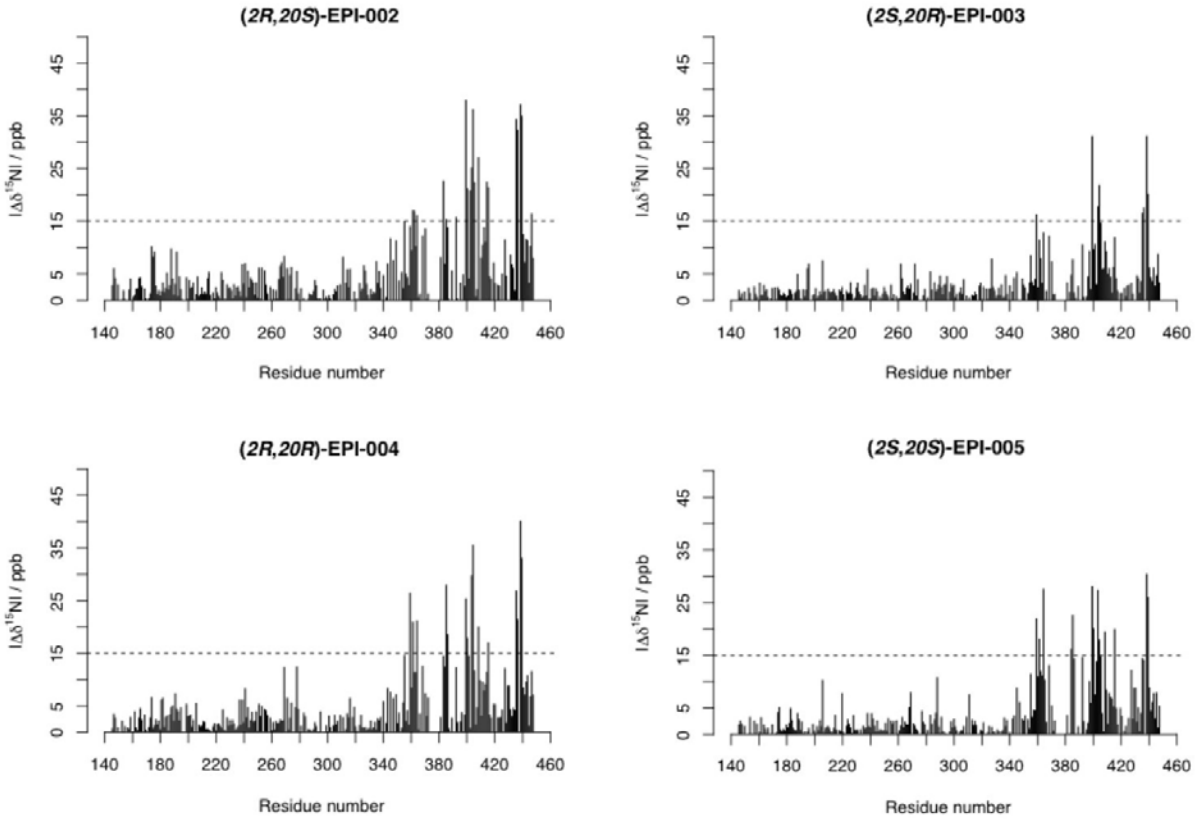


Figure S4: Absolute value of the changes in ^{15}N chemical shift caused by the four stereoisomers of EPI-001 on AF-1*.

7. Confirmation of reversible nature of the interaction by mass spectrometry analysis

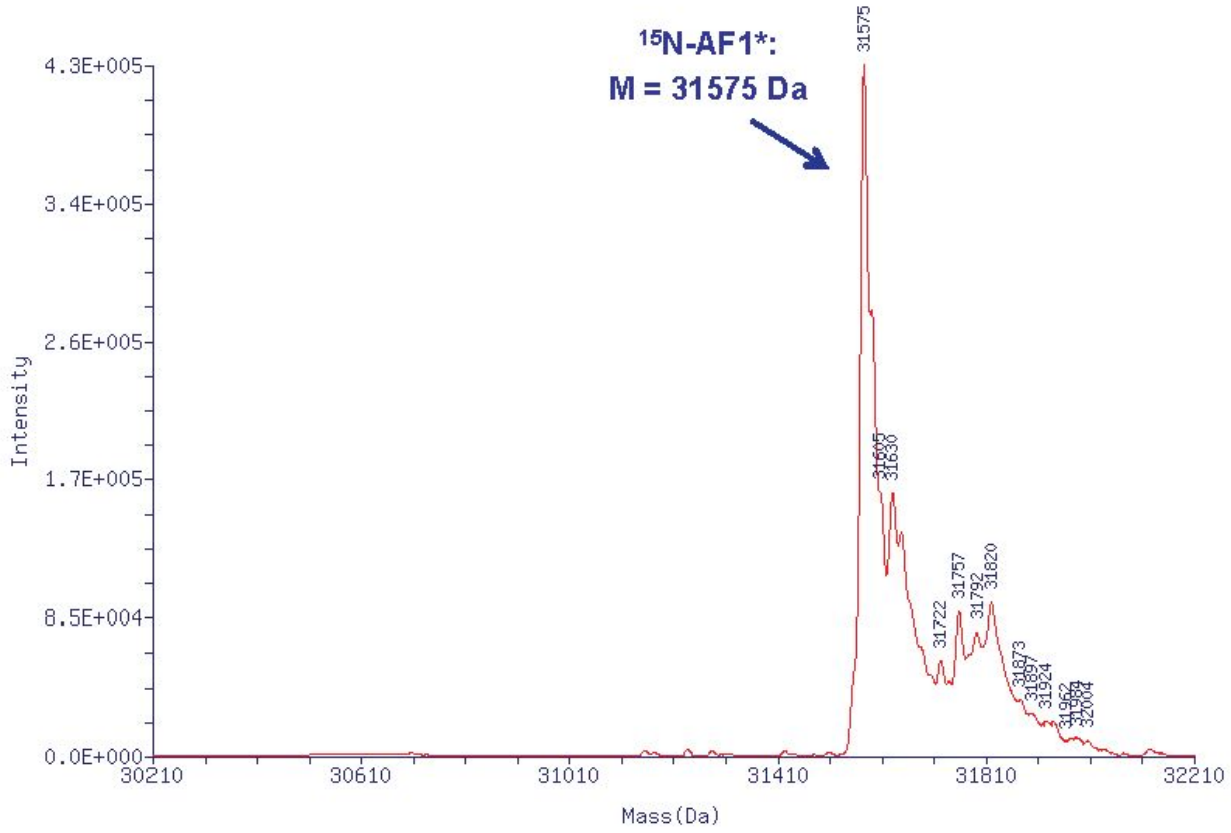


Figure S5: MS analysis of ^{15}N-labeled AF-1* and 10 molar equivalents EPI-001 after 4h of incubation at 278K, confirming that the small molecule has not reacted with the protein.

8. References

(1) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.
(2) Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR 5*, 67–81.
(3) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins 59*, 687–696.
(4) Bruun, S. W., Iešmantavičius, V., Danielsson, J., and Poulsen, F. M. (2010) Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein. *Proc. Natl. Acad. Sci. U. S. A. 107*, 13306–13311.

9.2 Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators

SUPPLEMENTARY INFORMATION

Structure, Volume *26*

Supplemental Information

Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators

Eva De Mol, Elzbieta Szulc, Claudio Di Sanza, Paula Martínez-Cristóbal, Carlos W. Bertoncini, R. Bryn Fenwick, Marta Frigolé-Vivas, Marianela Masín, Irene Hunter, Víctor Buzón, Isabelle Brun-Heath, Jesús García, Gianni De Fabritiis, Eva Estébanez-Perpiñá, Iain J. McEwan, Ángel R. Nebreda, and Xavier Salvatella

Supplemental items

Regulation of androgen receptor activity by transient interactions of its transactivation domain with general transcription regulators

Eva De Mol[1], Elzbieta Szulc[1], Claudio Di Sanza[1], Paula Martínez-Cristóbal[1], Carlos W. Bertoncini[1], R. Bryn Fenwick[1], Marta Frigolé-Vivas[1], Marianela Masín[1], Irene Hunter[2], Víctor Buzón[1], Isabelle Brun-Heath[1], Jesús García[1], Gianni De Fabritiis[3,4], Eva Estébanez-Perpiñá[5], Iain J. McEwan[2], Ángel R. Nebreda[1,4] and Xavier Salvatella[1,4,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain.
[2]Institute of Medical Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, IMS Building, Foresterhill, Aberdeen AB25 2ZD, Scotland, United Kingdom.
[3]Computational Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), Doctor Aiguader 88, 08003 Barcelona, Spain.
[4]ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain.
[5]Departament de Bioquímica i Biomedicina Molecular, Universitat de Barcelona and Institute of Biomedicine of the University of Barcelona (IBUB), Baldiri Reixac 15-21, 08028 Barcelona, Spain.

*Lead Contact: E-mail: xavier.salvatella@irbbarcelona.org, phone: +34934020459

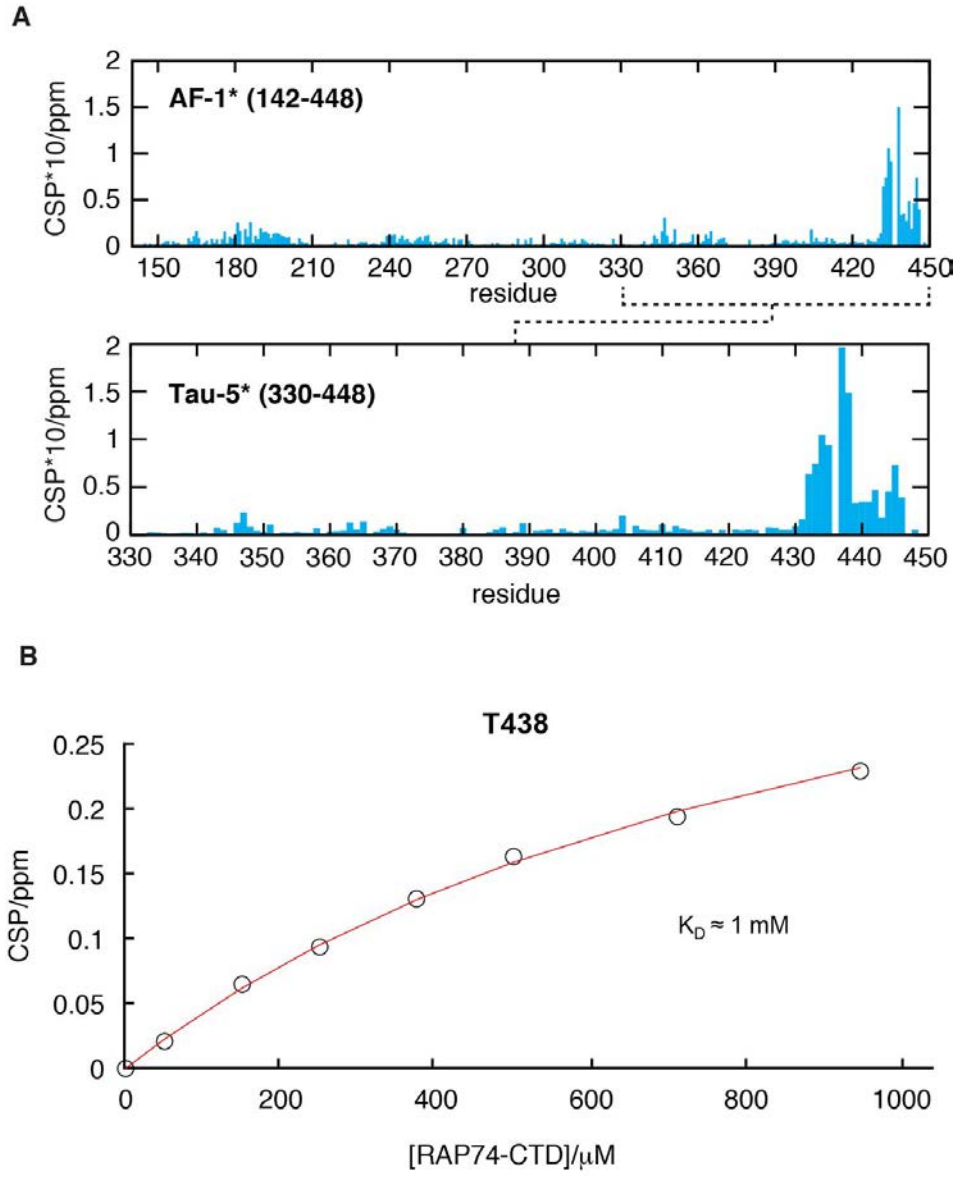


Figure S1, related to Figure 1: (A) Plot of the CSPs (CSP= $\sqrt{\Delta\delta_H{}^2 + (\Delta\delta_N/5)^2}$) observed in 50 μM AF-1* and in 50 μM Tau-5* after addition of 500 μM RAP74-CTD. (B) Plot of the CSPs observed for T438 of Tau-5* during a titration with RAP74-CTD and fit to estimate the K_D.

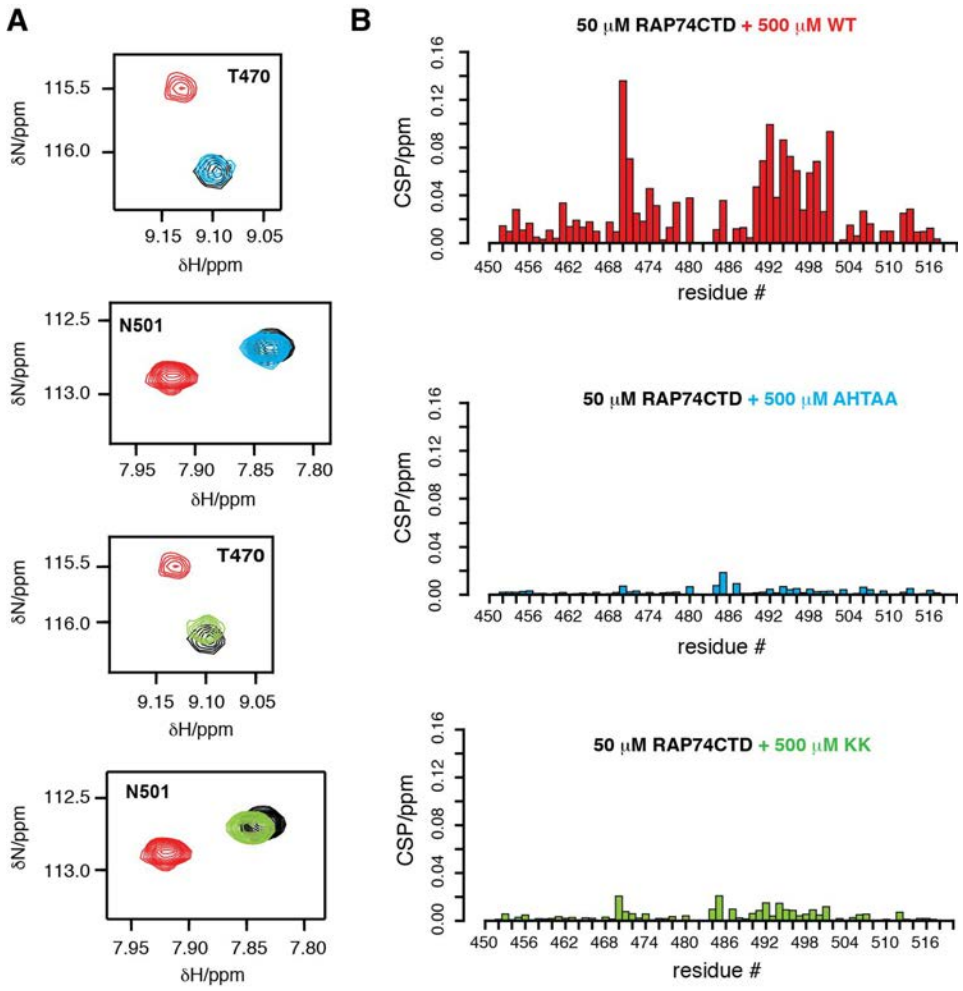


Figure S2. Chemical shift perturbations (CSPs) caused by peptides AHTAA and KK on RAP74-CTD, related to Figure 2. (A) Regions of the $^{1}H,^{15}N$-HSQC spectrum of RAP74-CTD (black) illustrating the CSPs caused by peptides WT (red), AHTAA (blue) and KK (green) (see Fig. 3A) on the resonances of T470 and N501 (B) Plot of the CSPs caused by peptides WT, AHTAA and KK as a function of residue number.

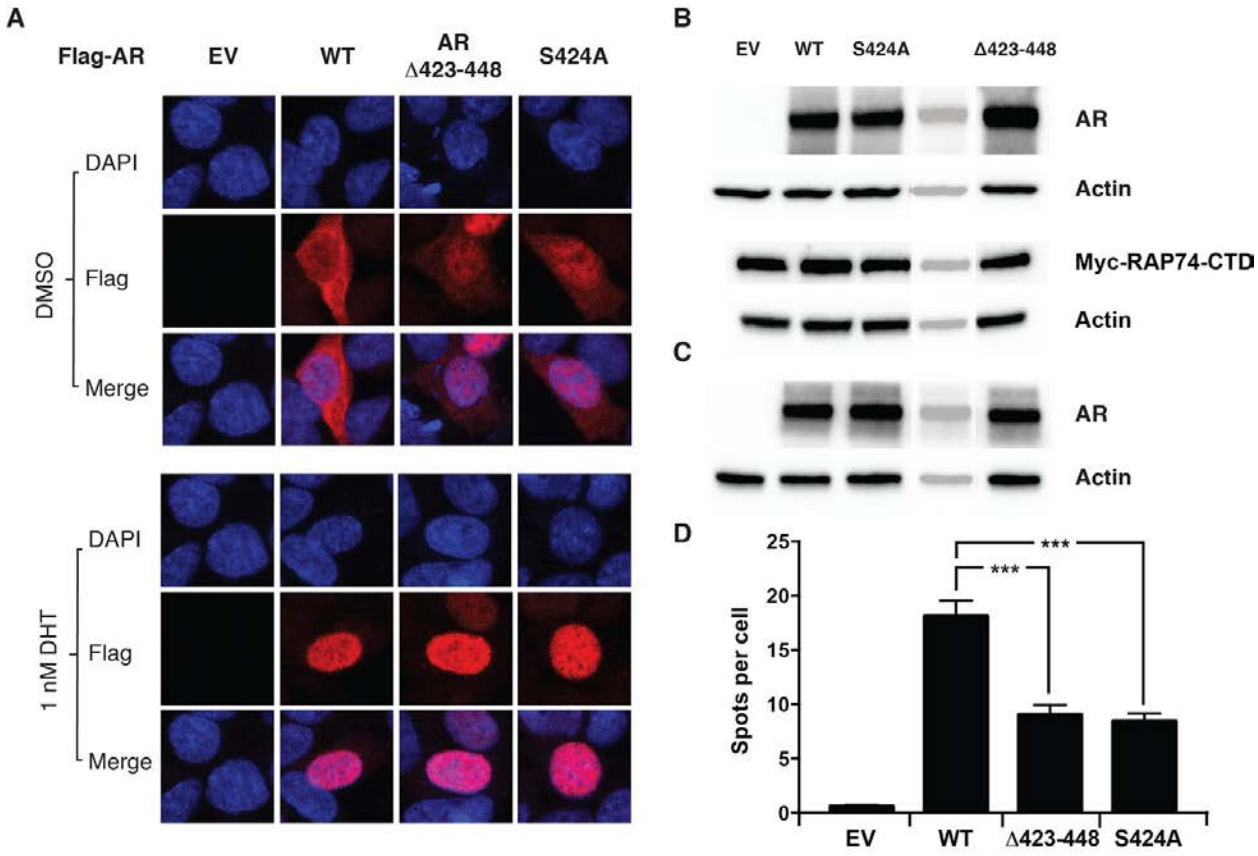


Figure S3, related to Figure 3: (A) Subcellular localization of AR in transfected cells. HEK293T cells were transfected with plasmids expressing an empty vector (EV), WT or mutant Flag-AR (red) and were analyzed by immunofluorescence to determine the cellular localization of AR proteins in the absence and in the presence of 1nM DHT. DAPI-stained nuclei, where DAPI stands for 2-(4-amidinophenyl)-1H -indole-6-carboxamidine, are shown in blue. Note the nuclear localization of all AR proteins upon activation by DHT. (B) Expression levels of AR protein and RAP74-CTD in transfected cells used for the experiments shown in Figure 3F: HEK293T cells were co-transfected with plasmids expressing the indicated Flag-tagged AR proteins, as well as Myc-RAP74-CTD, treated with 1 nM DHT and analyzed by Western blotting. (C) Expression levels of AR protein in transfected cells used for the experiments shown in

Figure 3G: HEK293T cells were transfected with plasmids expressing the indicated Flag-tagged AR proteins treated with 1 nM DHT and analyzed by Western blotting. In panels B and C actin was used as a loading control and the shaded lane corresponds to a mutant not related to the current study. (D) Plot of the average number of spots per cell observed in the PLA experiment reported in Figure 3F in HEK293T cells transfected with EV, WT or mutant Flag-AR, where the error bars represent the standard error and *** indicates $p < 0.001$.

9.3 Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments

SUPPLEMENTARY INFORMATION

Molecular Cell, Volume *72*

Supplemental Information

Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments

Jill J. Bouchard, Joel H. Otero, Daniel C. Scott, Elzbieta Szulc, Erik W. Martin, Nafiseh Sabri, Daniele Granata, Melissa R. Marzahn, Kresten Lindorff-Larsen, Xavier Salvatella, Brenda A. Schulman, and Tanja Mittag

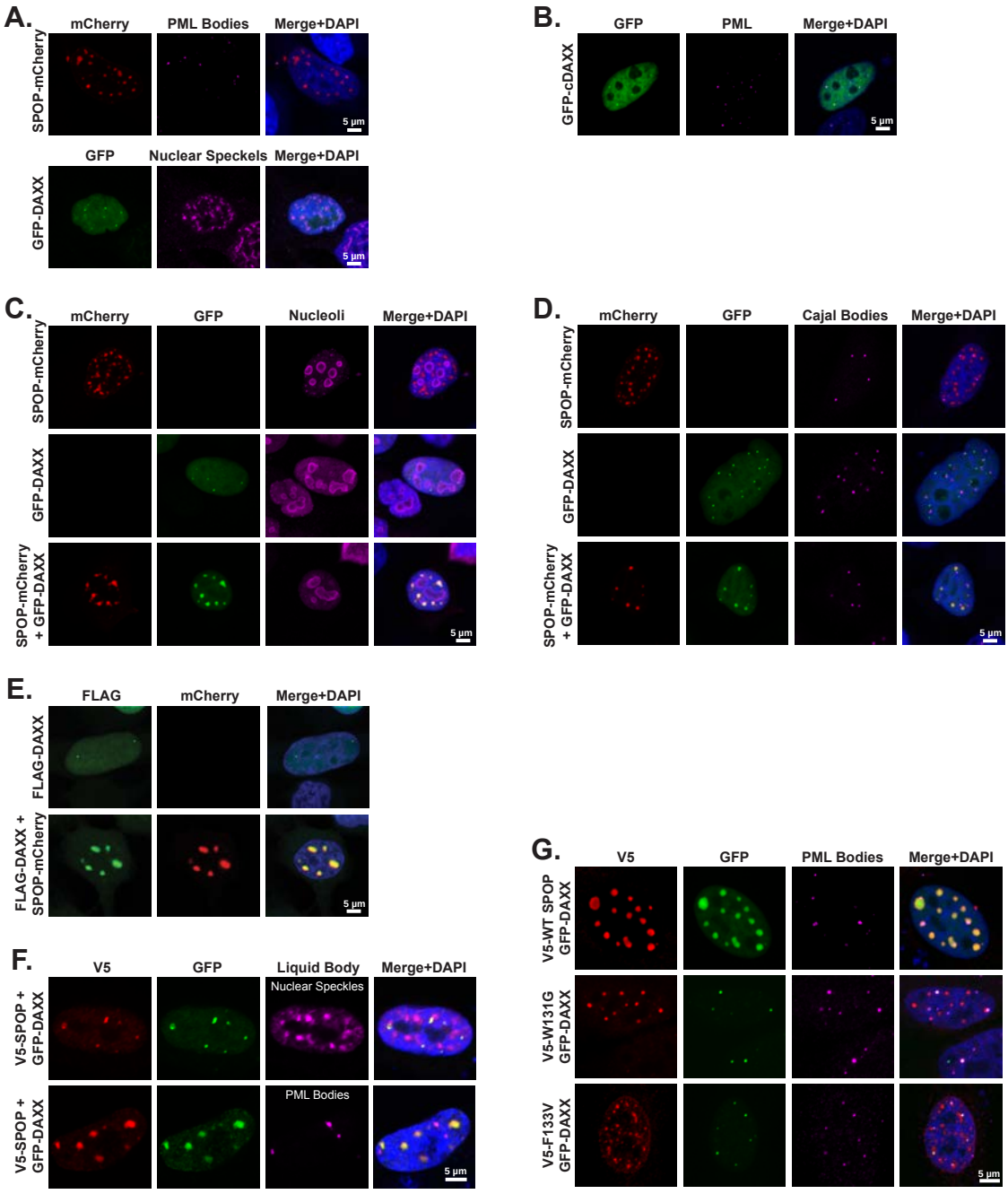


Figure S1: SPOP and DAXX co-localize in bodies that are neither nuclear speckles, PML bodies, nucleoli nor Cajal bodies, independent of expression tag, Related to Figure 1. (A) SPOP does not localize to PML bodies, and DAXX does not localize to nuclear speckles. HeLa cells were transfected with the indicated constructs (left column) and analyzed by confocal microscopy. mCherry (red) and GFP fluorescence (green) were observed for SPOP and DAXX, while SC-35 and PML (both magenta) were used as markers for nuclear speckles and PML bodies, respectively, and detected by IF. DAPI (blue) marks the nucleus. **(B)** The C-terminal intrinsically disordered region of DAXX, cDAXX, shows the same behavior in cells as DAXX. HeLa cells were transfected and analyzed as in (A). **(C)-(D)** SPOP and DAXX do not localize to nucleoli or Cajal bodies. HeLa cells were transfected and analyzed as in (A). B23 (C) and coilin (D), both magenta, were used as markers for nucleoli and Cajal bodies, respectively, and detected by IF. **(E)** Epitope tagged FLAG-DAXX co-localizes with SPOP-mCherry. HeLa cells were transfected and analyzed as in (A). FLAG-

DAXX (green) was detected by IF. **(F)** SPOP and DAXX localize to SPOP/DAXX bodies in PC-3 cells. PC-3 cells were transfected with the indicated GFP-DAXX and V5-SPOP and analyzed by confocal microscopy. V5-SPOP (red) and GFP fluorescence (green) were observed for SPOP and DAXX, while SC-35 and PML (both magenta) were used as markers for nuclear speckles and PML bodies, respectively, and detected by IF. **(G)** SPOP cancer mutants fail to localize to SPOP/DAXX bodies. HeLa cells were transfected and analyzed as in (A). GFP-DAXX colocalized with PML bodies in the presence of mCherry-SPOP cancer mutants.

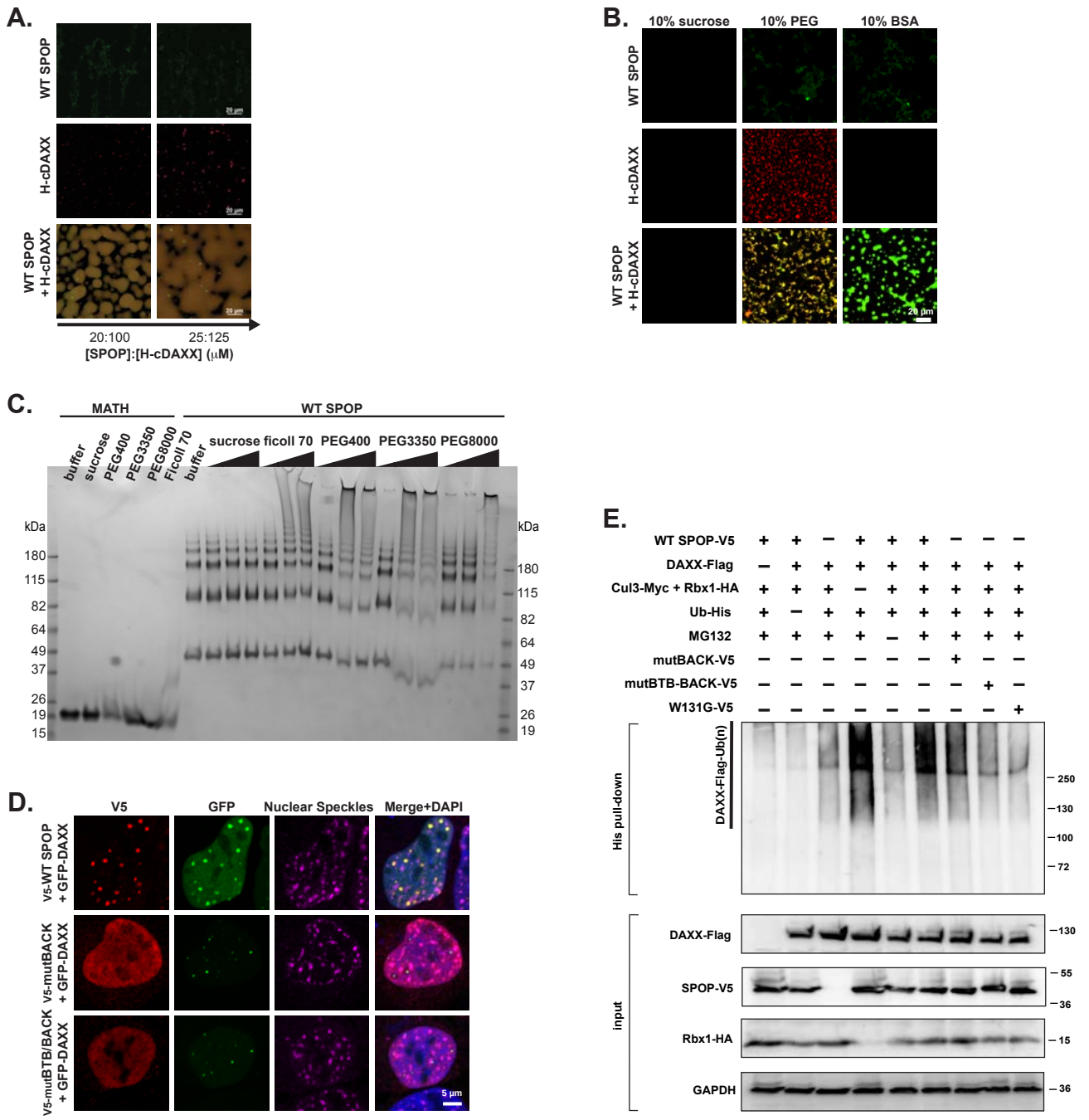


Figure S2: Molecular crowding enhances DAXX and SPOP/DAXX phase separation and SPOP self-assembly, Related to Figure 2. (A) H-cDAXX phase separates at high concentrations, and phase separation is enhanced by SPOP. Fluorescence microscopy images of WT SPOP (green) alone, H-cDAXX (red) alone, and SPOP + H-cDAXX at a constant ratio of 1 SPOP : 5 DAXX as in Fig. 2a, but at higher concentrations. **(B)** Phase separation of SPOP and DAXX is largely independent of crowding agent. Fluorescence microscopy images as in (A) of 15 μM WT SPOP (green) and 75 μM H-cDAXX (red) in the presence of different crowders or sucrose (as a control for viscosity). **(C)** SPOP forms large oligomers in molecular crowders, but increased solvent viscosity is not sufficient. In vitro cross-linking assay where an excess of amine cross-linker BS^3 was added to 15μM MATH or WT SPOP samples containing 1, 7.5, and 20% w/v of the indicated crowders or sucrose. **(D)**

SPOP multivalency is required for SPOP/DAXX co-localization in cells. HeLa cells were transfected with GFP-DAXX and WT V5-SPOP or mutants mutBACK or mutBTB/BACK and analyzed by confocal microscopy. GFP fluorescence was observed for DAXX, while V5-SPOP (red), and nuclear speckles (magenta) were detected by IF. **(E)** SPOP mutants are defective at DAXX ubiquitination in cells*.* Western Blots showing Flag-DAXX ubiquitination in HEK293T cells that were transfected with the constructs indicated above. 24 h post-transfection, cells were incubated with MG132 or DMSO at 20 μM for 4 hours. Cells were lysed under denaturing conditions, His_6-Ubiquitin was pulled down and the resulting proteins were analyzed by immunoblotting with anti-Flag antibody. Input materials were also checked by immunoblotting using anti-Flag, anti-V5, anti-HA, and anti-GAPDH antibodies (loading control).

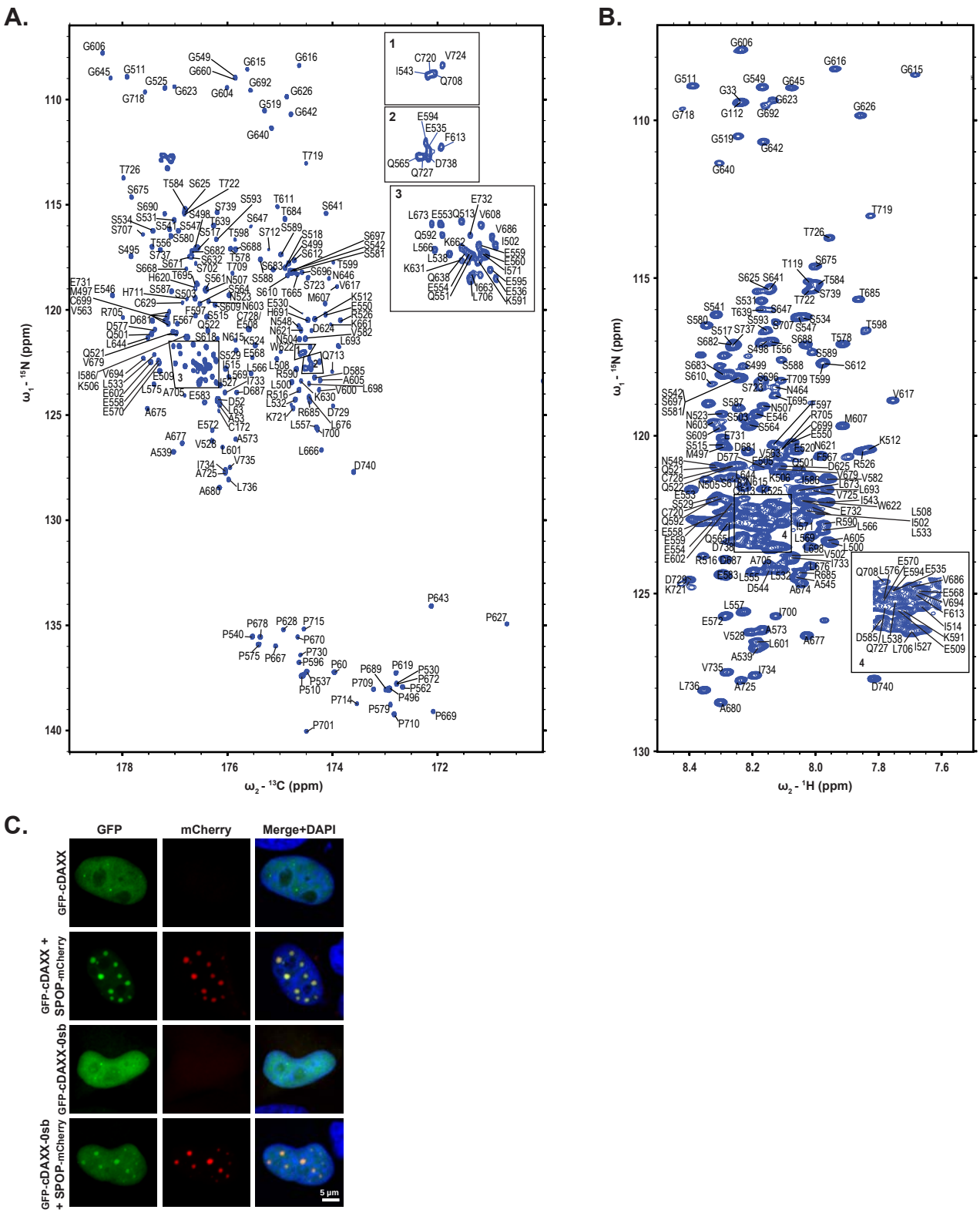


Figure S3: cDAXX is intrinsically disordered and localizes to PML bodies, Related to Figure 3. (A) ^{15}N,^{13}C CON NMR spectrum at 600 MHz and 25 °C and **(B)** ^{1}H, ^{15}N HSQC spectrum of cDAXX at 800 MHz and 25 °C with resonance assignments. **(C)** cDAXX-0sb localizes to PML bodies when not expressed with SPOP. HeLa cells were transfected with the indicated constructs (left column) and analyzed by confocal microscopy. GFP (green) and mCherry (red) fluorescence were observed for cDAXX and SPOP, respectively.

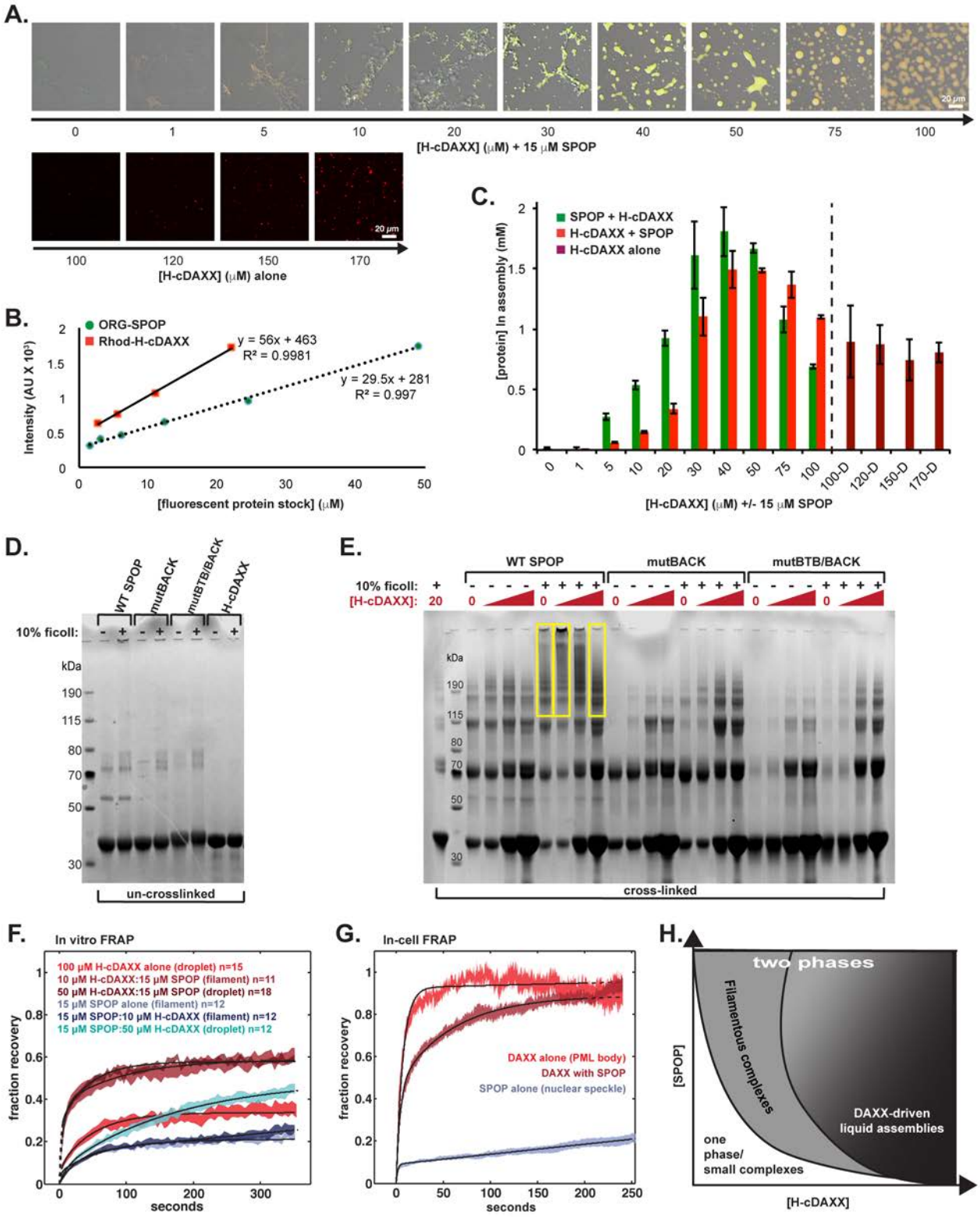


Figure S4: Filamentous assemblies and droplets are formed by different molecular processes, Related to Figure 4. (A-C) Quantification of protein concentration in droplets shown in samples. **(A)** Representative fluorescence microscopy/DIC images of 15 μM SPOP and H-cDAXX at concentrations indicated (top), or H-cDAXX only (bottom). All samples contain 10% w/v ficoll 70, 500 nM ORG-SPOP and 1/60 (but not below 500 nM) Rhodamine-H-cDAXX. Red and green fluorescence intensities were quantified from triplicate images. **(B)** Standard curve relating fluorescence intensity to fluorescent protein concentration. The

equations shown were used together with the ratio of labeled protein : unlabeled protein to calculate the protein concentrations in (C). **(C)** At low SPOP/ H-cDAXX molar ratios, LLPS is driven by H-cDAXX; the filamentous assemblies at high SPOP/ H-cDAXX molar ratios are not driven by LLPS. Quantification of protein concentration in mesoscale assemblies in (A) shows two slopes. Error bars represent the SD from three images. Dashed vertical line separates protein concentrations determined from the two sets of data: left corresponds to samples from (A), top, and right corresponds to H-cDAXX alone samples (A, bottom). **(D)** Control gel of 15 μM SPOP variants and 20 μM H-cDAXX +/- ficoll 70. **(E)** In vitro cross-linking of SPOP variants (15 μM) in the absence and presence (1, 20, and 50) μM of H-cDAXX and the absence and presence of 10% w/v ficoll 70. In the presence of ficoll, WT SPOP forms large oligomers (left yellow box); at high SPOP/ H-cDAXX molar ratios the oligomers become larger (middle yellow box), but low ratios reduce the complex size (right yellow box). We presume that at high ratios, bridging of different SPOP oligomers becomes favored over stabilization of large SPOP oligomers. (See schematic in Fig. 4D). **(F)** In vitro FRAP shows SPOP becomes more mobile in droplets versus filaments while DAXX maintains high mobility in both types of assemblies. FRAP measurements of samples as indicated. All samples contain 4% w/v ficoll 70 and 200 nM ORG-SPOP or 100 nM Rhodamine-H-cDAXX; the protein listed first in the legend was labeled and measured. The colored areas are average values ± SEM. The solid lines are best fits to a biexponential recovery. (For fit values see Table S4.) **(G)** In cell FRAP is consistent with mobilites seen in vitro. **(H)** Simplified phase diagram for SPOP/H-cDAXX phase separation. H-cDAXX alone can undergo LLPS at high concentrations (intersection of dark shaded area with x-axis). SPOP enhances H-cDAXX LLPS (dark shaded area), and forms large filamentous assemblies with H-cDAXX (light shaded area). The physical properties of the SPOP/H-cDAXX assemblies retain characteristics of both species along the phase boundary (gradient shading).

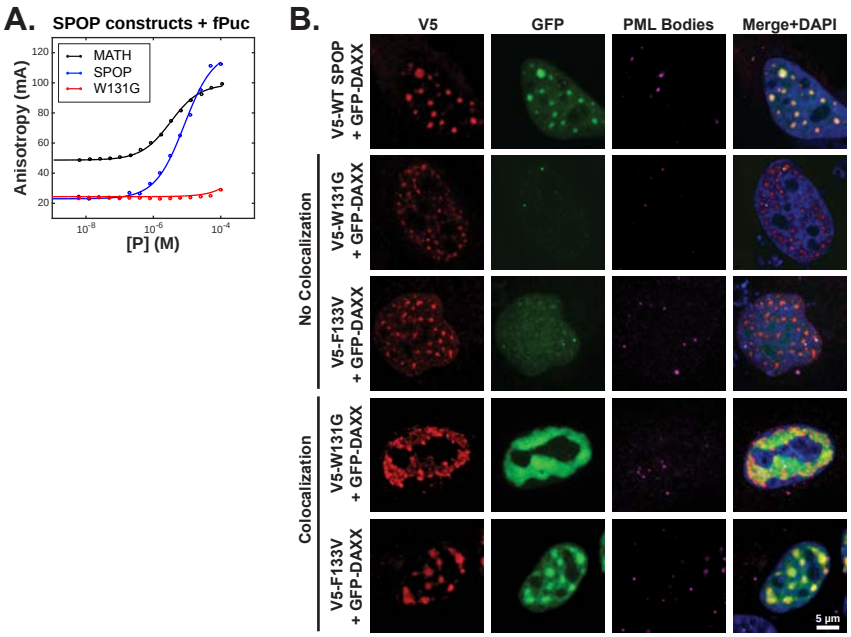


Figure S5: SPOP cancer mutants are defective for binding substrate in vitro and co-localization with DAXX in cells, Related to Figure 5. (A) The MATH domain and SPOP bind a standard SB motif-containing peptide with micromolar affinity, while SPOP W131G nearly abrogates binding. Representative fluorescence anisotropy binding isotherms for the indicated SPOP constructs into fluorescein-Puc$^{91\text{-}106}$ (fPuc). (For K_D values resulting from fits of this data, see Table S5.) **(B)** SPOP cancer mutants are defective at co-localization with DAXX in cells. HeLa cells were transfected with GFP-DAXX and V5-SPOP WT or mutants F133V or W131G and analyzed by confocal microscopy. GFP fluorescence was observed for DAXX, while V5 (red) and PML (magenta) were used as markers for SPOP and PML bodies, respectively, and detected by IF. DAPI (blue) marks the nucleus. Cells with SPOP/DAXX co-localization or lack-thereof are indicated.

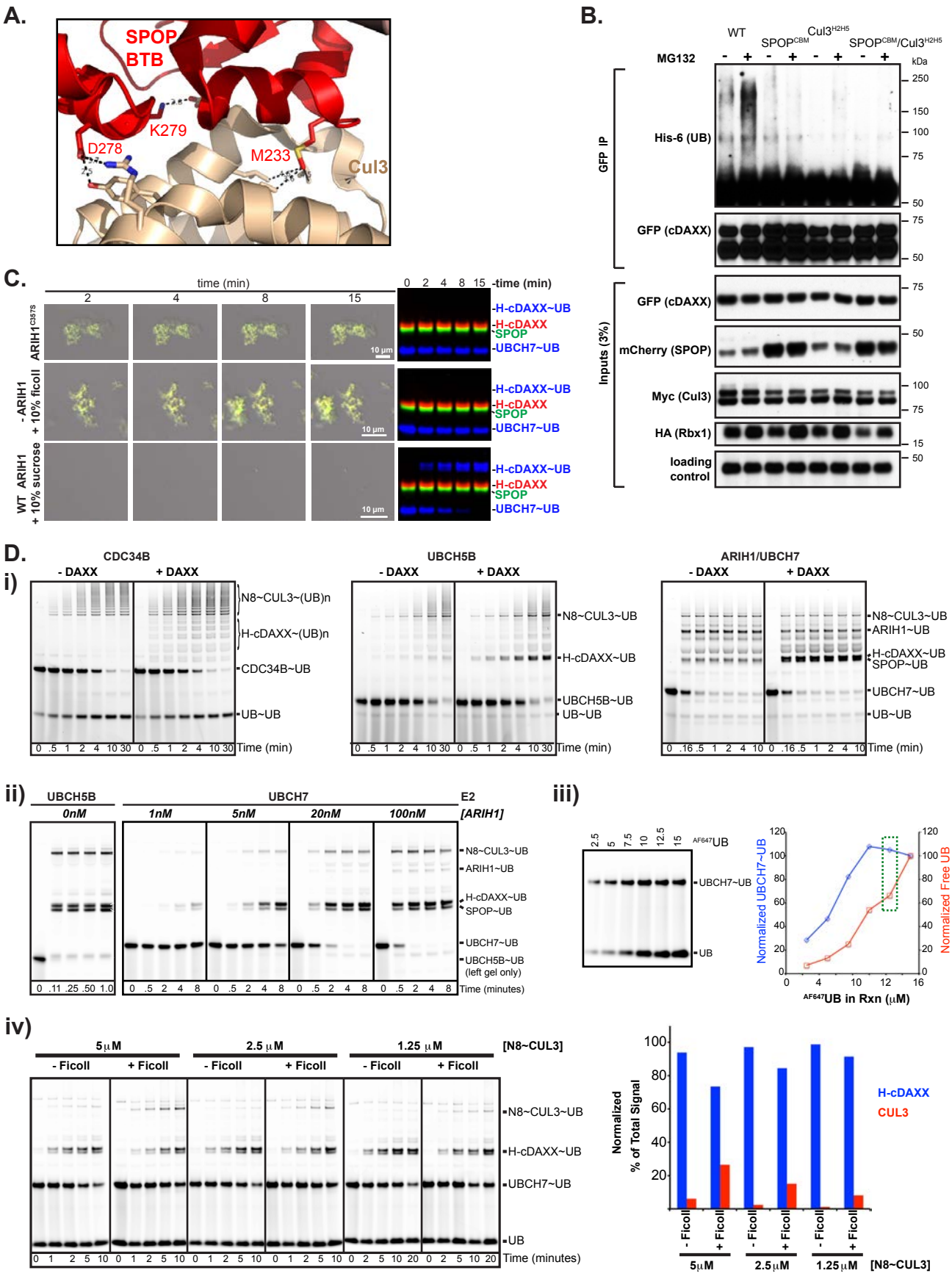


Figure S6: Interaction of Cul3 with SPOP and ARIH1 is critical for ubiquitination, Related to Figure 6. (A) SPOP mutant D278A/K279A/M233E disrupts binding to Cul3.

Respective residues are indicated in stick representation on the ribbon diagram of the SPOP-Cul3 structure, PDB ID 4EOZ (Errington et al., 2012). **(B)** DAXX ubiquitination in cells depends on $CRL3^{SPOP}$. Western Blots showing GFP-cDAXX ubiquitination in HEK293T cells that were transfected with the constructs indicated on top. 24 h post-transfection, cells were incubated with MG132 or DMSO at 20 μM for 4 hours. Cells were lysed, GFP-cDAXX was pulled down and the resulting proteins were analyzed by immunoblotting with anti-GFP and anti-His_6 antibody. Input materials were also checked by immunoblotting using anti-GFP, anti-V5, anti-Myc, anti-HA, and anti-GRP170 antibodies (loading control). **(C)** In the absence of ARIH1, *UB does not accumulate in SPOP/H-cDAXX assemblies *in vitro.* Fluorescence microscopy images merged with DIC showing the time-course of *in vitro* ubiquitination assays (described in Fig. 6F,G) containing 5 μM SPOP (green), 20 μM H-cDaxx (red), 1.25 μM N8~Cul3/Rbx1, 20 nM ARIH1 as indicated on the left, and 1.5 μM UbcH7~*UB (blue). All reactions contain either 10% ficoll 70 or sucrose, and 500 nM ORG-SPOP and Rhodamine-H-cDAXX; *UB denotes stoichiometrically labeled Alexa647-Ubiquitin. **(D)** UBCH7 has the strongest activity towards H-cDAXX, with a rate controlled by the amount of ARIH1. Ubiquitination experiments to optimize (i) the E2 conjugating enzymes with most activity towards H-cDAXX and not to Cul3, (ii) the amount of ARIH1 to slow the reaction to a measurable rate for microscopic analysis, (iii) the complete loading of UBCH7~*UB with minimal free *UB, (iv) the amount of N8~Cul3 which maximizes H-cDAXX~UB transfer and minimizes N8~Cul3-*UB in assemblies. Reaction (i) contains 40 nM SPOP, 5 μM H-cDaxx, 40 nM N8~Cul3/Rbx1, 400 nM E2, and 300 nM ARIH1 where indicated. Reactions (ii) and (iv) contain 5 μM SPOP, 20 μM H-cDaxx, 5 μM N8~Cul3/Rbx1 (unless otherwise indicated), 400 nM E2, 5 nM ARIH1 (unless otherwise indicated). The optimized parameters were not independent, and a compromise was used for final assays. Since the microscopic studies were conducted in HEPES buffer instead of Tris and necessitated double the UbcH7~*UB to visualize the signal, ARIH1 was quadrupuled to maintain kinetics.

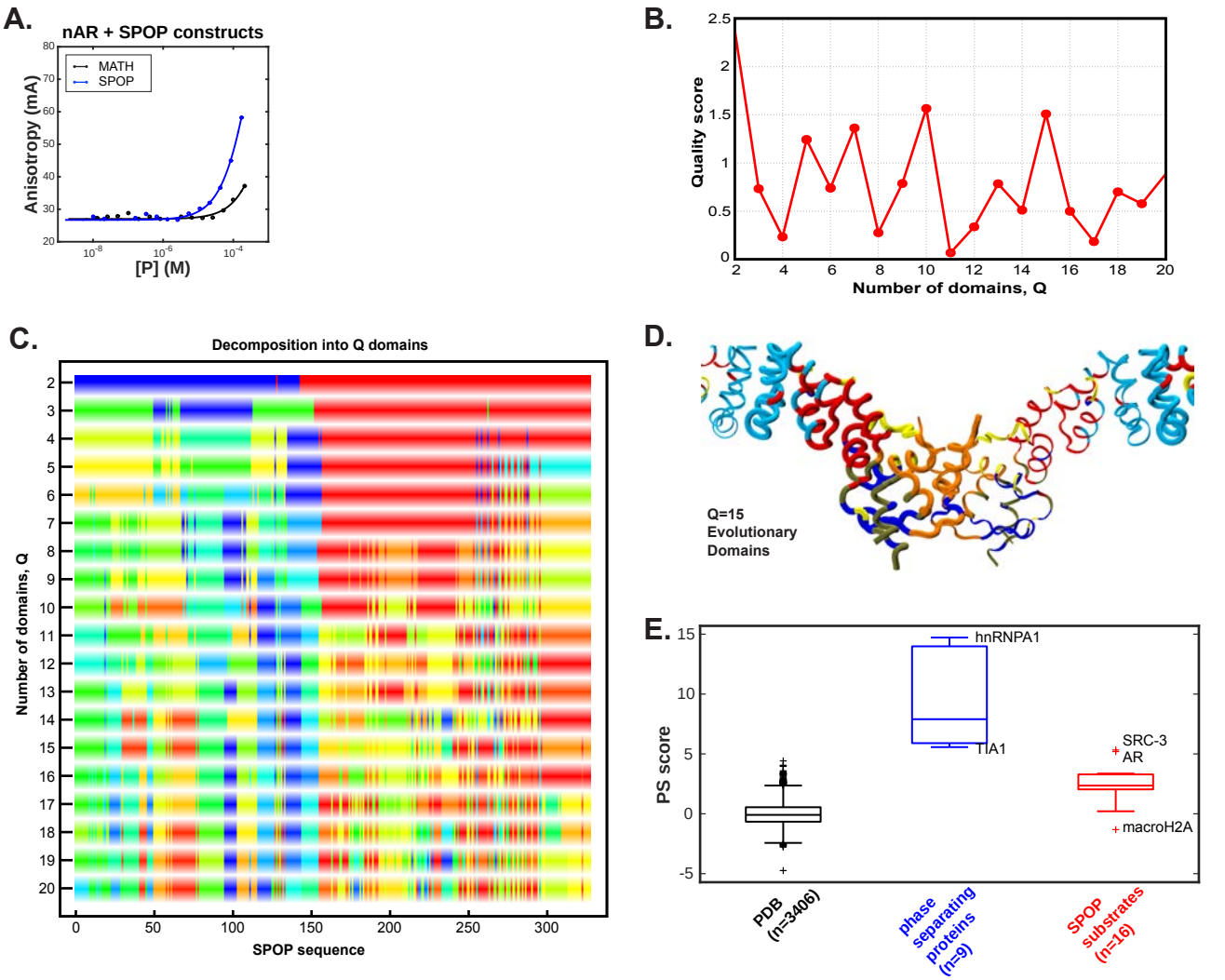


Figure S7: Clustering of co-evolutionary couplings into evolutionary domains supports coevolution of BTB and BACK domain residues across interfaces instead of with residues in the core of the domains, Related to Figure 7. (A) Direct FA binding isotherms for $SPOP^{MATH}$ and $SPOP^{28-359}$ into full-length Rhodamine-nAR. Symbols are experimental data points; continuous lines are non-linear least-squares fits (Roehrl et al., 2004). All measurements were conducted in triplicate. **(B)** Quality score showing the most meaningful subdivisions in number of domains Q. Higher quality scores better fit the model to the co-evolutionary couplings (Granata et al., 2017). **(C)** Representation of possible subdivisions on the primary structure of SPOP. **(D)** Evolutionary domains (Q=15) of the BTB/BACK unit represented on the oligomer structure model: interfacial residues coevolve with residues across the interface rather than with the core of the domains, in the BTB domain (in orange) and the BACK domain (in cyan). **(E)** Phase separation propensity scores (PS scores) (Vernon et al., 2018) for a non-redundant set of proteins from the PDB (black), a set of proteins with an experimentally demonstrated high phase separation propensity (hnRNPA1, hnRNPA2B1, FUS, EWS1, DDX3X, DDX4, TDP-43, EIF4H, TIA1; with PS score from high to low; blue; the predictor was trained on these proteins), and for known SPOP substrates (SRC-3, AR, BMI1, BRD4, Gli3, Gli2, BRD2, TRIM24, DAXX, ER, BRD3, ERG, PDX1, DEK, BRMS1, macroH2A; red). Each box shows the interquartile range, the central line as the median, the whiskers extend to 1.5 x the interquartile range, and any outliers from a normal distribution are plotted individually. Half of the SPOP substrates have PS scores that fall

outside the maximum range (99.3% of values if normally distributed) of PDB proteins and are on the order of PDB outliers. Three SPOP substrates, macroH2A (-1.32), BRMS1 (0.21) and DEK (0.33), rank below the 90th percentile of PDB protein PS scores. The predictor captures phase separation of disordered sequences via pi-pi contacts, not phase separation via multivalent domain/motif interactions.

Table S1. Sequences of cDAXX, cDAXX-0sb, Related to Figure 1.

Construct	Sequence
cDAXX	SPMSSLQISN EKNLEPGKQ**I** **SRSS**GEQQNK GRIVSPSLLS EEPLAPSSID AESNGEQPEE LTLEEESPVS QLFELEIEAL PLDTPSSVET D**ISSSR**KQSE EPFTTVLENG AGM**VSSTS**FN GGVSPHNWGD SGPPCKKSRK EKKQTGSGPL GNSYVERQRS VHEKNGKKIC TLPSPPSPLA SLAPV**ADSST** RVDSPSHGL**V** **TSSL**CIPSPA RLSQTPHSQP PRPGTCKTSV ATQCDPEEII VLSDSD
cDAXX-0sb	SPMSSLQISN EKNLEPGKQ**S** **PRSI**GEQQNK GRIVSPSLLS EEPLAPSSID AESNGEQPEE LTLEEESPVS QLFELEIEAL PLDTPSSVET D**SPRSI**KQSE EPFTTVLENG AGM**TPTVS**FN GGVSPHNWGD SGPPCKKSRK EKKQTGSGPL GNSYVERQRS VHEKNGKKIC TLPSPPSPLA SLAPV**SPDSA** RVDSPSHGL**S** **PVSL**CIPSPA RLSQTPHSQP PRPGTCKTSV ATQCDPEEII VLSDSD

Table S2. Dissociation constants of peptides containing cDAXX SB motifs or cDAXX-0sb mutated motifs binding to MATH, Related to Figure 3.

Peptide	Ac-Sequence-COOH [a]	K_D ± S.D. (mM) [b]
SB1	PGKQ**ISRSS**GEQQ	5 ± 2
Mut1	PGKQ**SPRSI**GEQQ	1000 ± 1000
SB2	VETD**ISSSR**KQSE	0.3 ± 0.1
Mut2	VETD**SPRSI**KQSE	nb [c]
SB3	GAGM**VSSTS**FNGG	0.03 ± 0.01
Mut3	GAGM**TPTVS**FNGG	nb [c]
SB4	LAPV**ADSST**RVDS	0.08 ± 0.03
Mut4	LAPV**SPDSA**RVDS	2 ± 2
SB5	SHGL**VTSSL**CIPS	0.13 ± 0.01
Mut5	SHGL**SPVSL**CIPS	nb [c]
tag-SB	HHHH**LESTS**LYKK	0.16 ± 0.07

[a] Bold portions in the sequences denote predicted SB motifs and the scrambled motifs to reduce binding.
[b] Errors represent standard deviations from triplicate competition FA experiments.
[c] No binding detected.

Table S3. Dissociation constants of cDAXX constructs binding to SPOP^{MATH} and SPOP, Related to Figure 3.

DAXX construct	MATH K_D ± S.D. (μM) [a]	SPOP K_D ± S.D. (μM) [a]
H-cDAXX	60 ± 5	1.2 ± 0.1
cDAXX	38 ± 6	1.7 ± 0.2
cDAXX-0sb	hundreds [b]	hundreds [b]

[a] Errors represent standard deviations from triplicate direct binding FA experiments.
[b] Binding too weak to fit reliable K_D, but experiment carried out in triplicate.

Table S4. FRAP recovery rates of SPOP and DAXX in vitro and in cells, Related to Figure S4.

condition	t1	t2	mobile fraction
in vitro			
100 μM DAXX / 0 μM SPOP	0.4 ± 0.12	49.0 ± 1.2	0.33
10 μM DAXX / 15 μM SPOP	4.9 ± 0.15	64.8 ± 3.4	0.58
~50 μM DAXX / 15 μM SPOP	1.7 ± 0.10	41.0 ± 1.1	0.58
15 μM SPOP / 0 μM DAXX	1.3 ± 0.25	60.0 ± 2.0	0.21
15 μM SPOP / 10 μM DAXX	0.7 ± 0.29	56.6 ± 1.1	0.22
15 μM SPOP / 50 μM DAXX	7.5 ± 0.70	165 ± 5.6	0.47
in vivo			
SPOP (in nuclear speckles)	1.6 ± 0.10	2026 ± 1000	0.20
DAXX (in PML bodies)	3.2 ± 0.27	18.5 ± 1.4	0.95
DAXX (in DAXX/SPOP bodies)	3.8 ± 0.11	53.5 ± 1.2	0.89

Table S5. Dissociation constants of SPOP constructs binding to fPuc, Related to Figure S5.

SPOP construct	K_D ± S.D. (μM) [a]
MATH	3.9 ± 0.4
SPOP	10 [b]
W131G	nb [c]

[a] Errors represent standard deviations from triplicate direct binding FA experiments.
[b] Value from single replicate, in agreement with (Marzahn et al. 2016).
[c] No binding detected.

Table S6. Number of predicted SB motifs in known SPOP substrates, Related to Figure 7.

Substrate	# SB motifs [a]	Disease Relevance or Function
Androgen receptor	9	Prostate cancer
BMI1	4	X-inactivation
BRMS1	1	Breast, bladder & skin cancers
DAXX	8	Prostate, ovarian, breast, stomach, cancers…
DEK	4	Breast, skin, prostate, colon cancers…
ERG	7	Prostate cancer
Puc/DUSP6/7	7/5/4	Kidney cancer
Estrogen Receptor	4	Breast & endometrial cancers
Ci/Gli2/3	6/10/22	Hedgehog signaling →Brain cancer
MacroH2A	2	X-inactivation
PDX-1	1	Type II diabetes
PTEN	2	Kidney cancer
SRC-3	17	Breast cancer
Trim24	13	Prostate cancer

[a] One mismatch from the consensus SB motif (Φ-Π-S-S/T-S/T, where Φ is a nonpolar and Π is a polar residue) allowed.

Table S7. Taxonomy of SPOP sequences in multiple sequence alignment, Related to Figure 7.

Taxonomy	**# sequences**
Plants	1,742
Metazoa	929
Fungi	156
Other	62
Unknown	232
Total	3,121
Total after trimming[a]	~2,600

[a] Sequences with similarity >90% were removed for the coevolution analysis.

Table S8. SPOP is highly conserved compared to SPOPL[a], Related to Figure 7.

Gene	Frequency of top missense SNP	SNP frequency $>10^{-4}$	SNP frequency $> 10^{-5}$
SPOP	4.7×10^{-5}	0	16
SPOPL	397.9×10^{-5}	6	66

[a] From the gnomAD database (Lek et al., 2016).

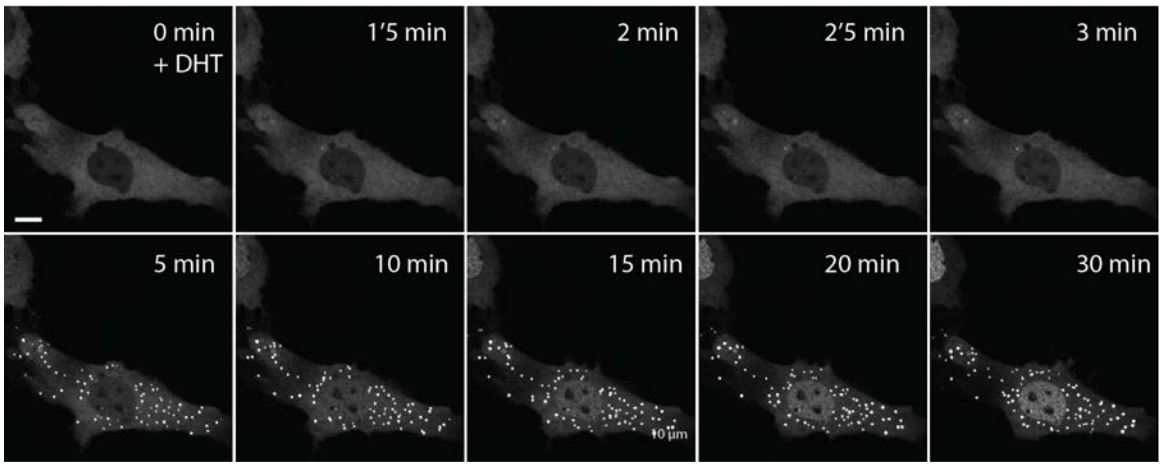
9.4 Hormone binding causes the condensation of androgen receptor that mediates folding of activation function motifs - *manuscript in preparation*

SUPPLEMENTARY INFORMATION

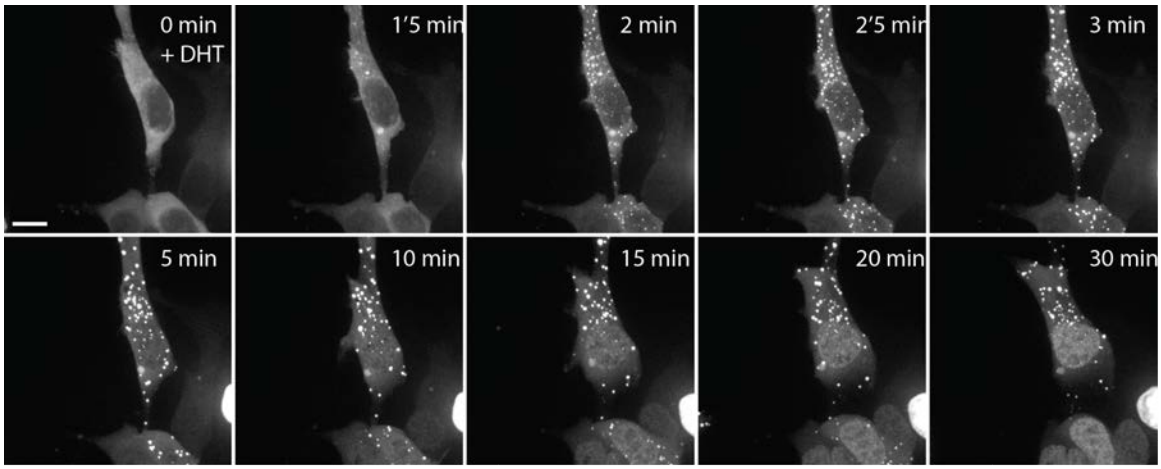
Hormone binding causes the condensation of androgen receptor that mediates folding of activation function motifs

Supplementary Material

A)

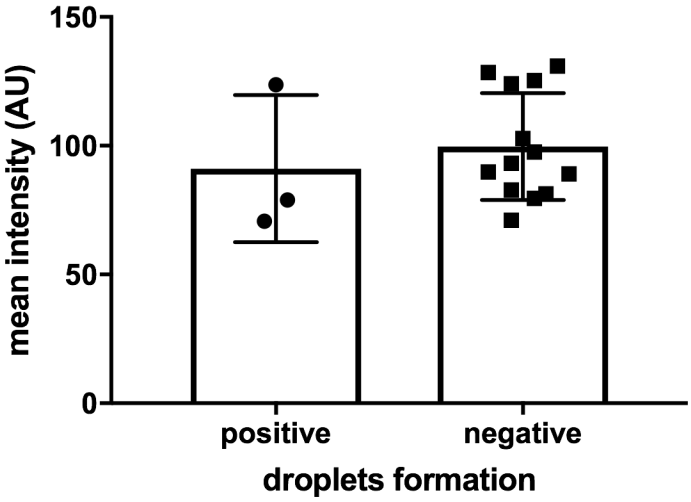


B)



S1. Time resolved fluorescence microscopy of EGFP-AR

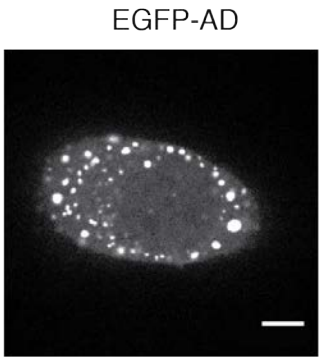
EGFP-AR forms droplet-like condensates in the cytosol upon 1nM DHT treatment in A) 3617 and B) HEK293T cells. Scale bar 10 µm. See also movies M2 and M3.



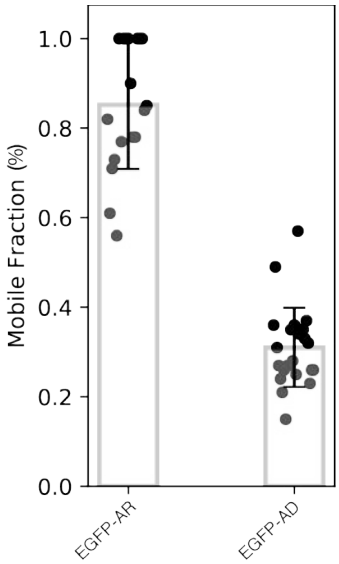
S2. Cytoplasmic droplets formation versus total concentration of the protein

Quantification of the mean fluorescence intensity of cytosolic EGFP-AR in PC3 cells does not correlate with the droplets formation after DHT addition (n=3 independent experiments). Each point in the plot signifies an individual cell. Error bars indicate the SD.

A)



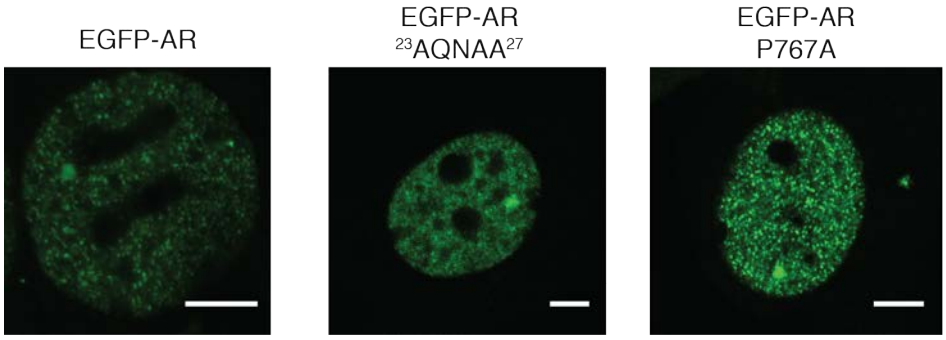
B)



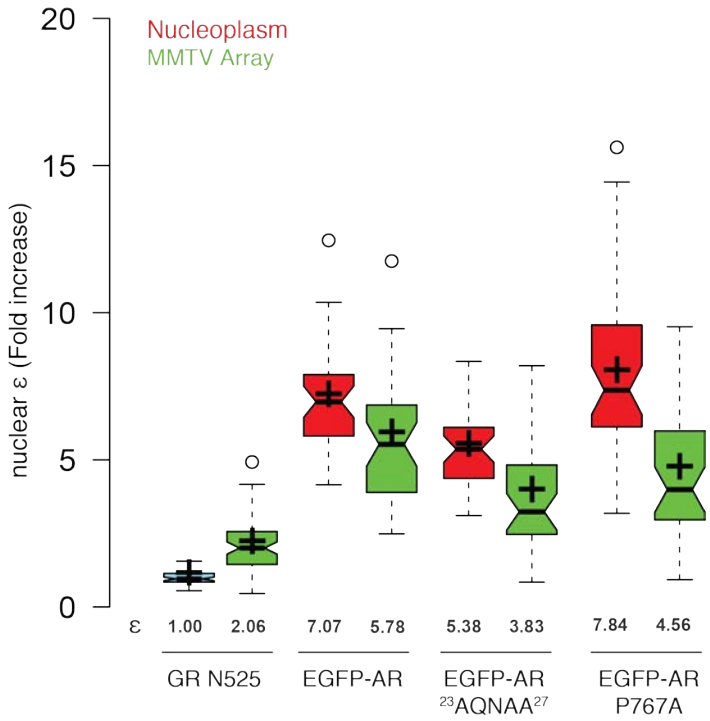
S3. EGFP-AD cytoplasmic aggregation

A) Representative fluorescence microscopy image from PC3 cells transiently transfected with the EGFP-AD show aggregation of the protein. Scale bar 10 µm. B) Comparison of the mobile fraction of the EGFP-AR and EGFP-AD from FRAP experiments. The graph reveals an immobile fraction of EGFP-AD molecules of circa 0.7. Error bars represent S.D. of n=24 cells.

A)

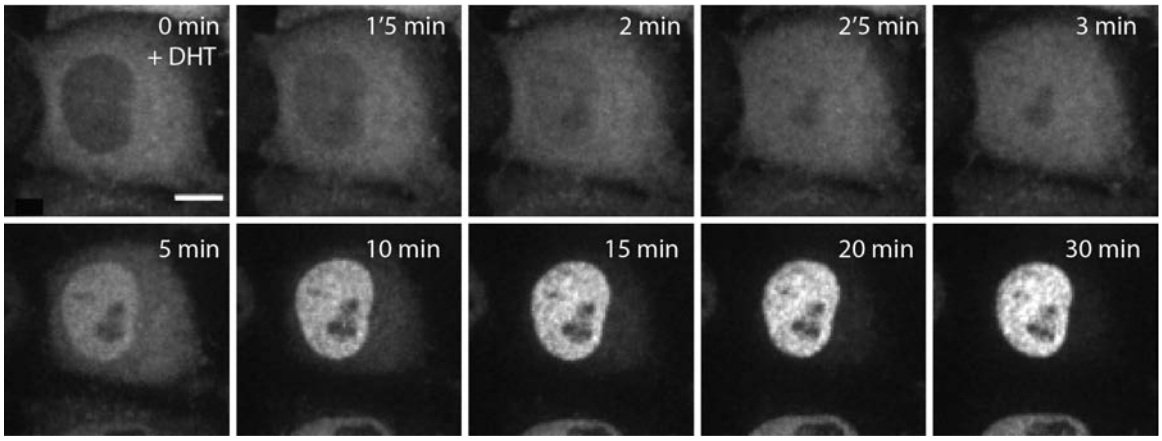


B)



S4. Oligomeric state of the EGFP-AR variants: EGFP-AR ^{23}AQNAA^{27} and EGFP-AR P767A in 3617 cells

A) Representative fluorescence images of 3617 cell nuclei transiently expressing the EGFP-AR, the N/C via ^{23}FQNLF^{27} deficient mutant (EGFP-AR ^{23}AQNAA^{27}) and the LBD dimerization deficient mutant (EGFP-AR P767A) (Nadal et al. 2017) show speckled-like distribution of the protein variants. Scale bar 5 uM. B) N&B assay of the EGFP-AR ^{23}AQNAA^{27} and EGFP-AR P767A in the nucleoplasm (red) and upon DNA binding (green) show high oligomeric species formation of the protein in the nucleoplasm and upon DNA binding. The figure shows the fold increase of the nuclear ε relative to the control. Glucocorticoid receptor mutant (GR N525) and EGFP-AR were used as a reference. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5-fold the interquartile range from the 25th and 75th percentiles, with outliers represented by dots; and crosses represent sample means (n=20).

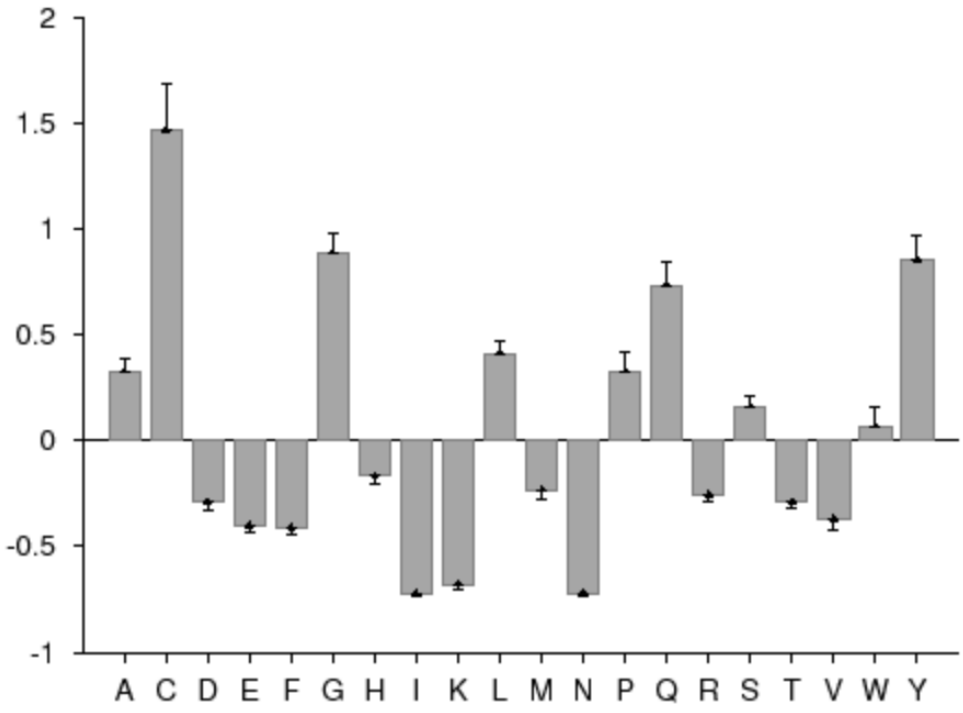


S5. Time resolved fluorescence microscopy of AR-EGFP in HeLa cells

AR-EGFP stably transfected HeLa cells do not show AR condensation upon treatment with 1nM DHT (at min 0). Scale bar 10 µm.

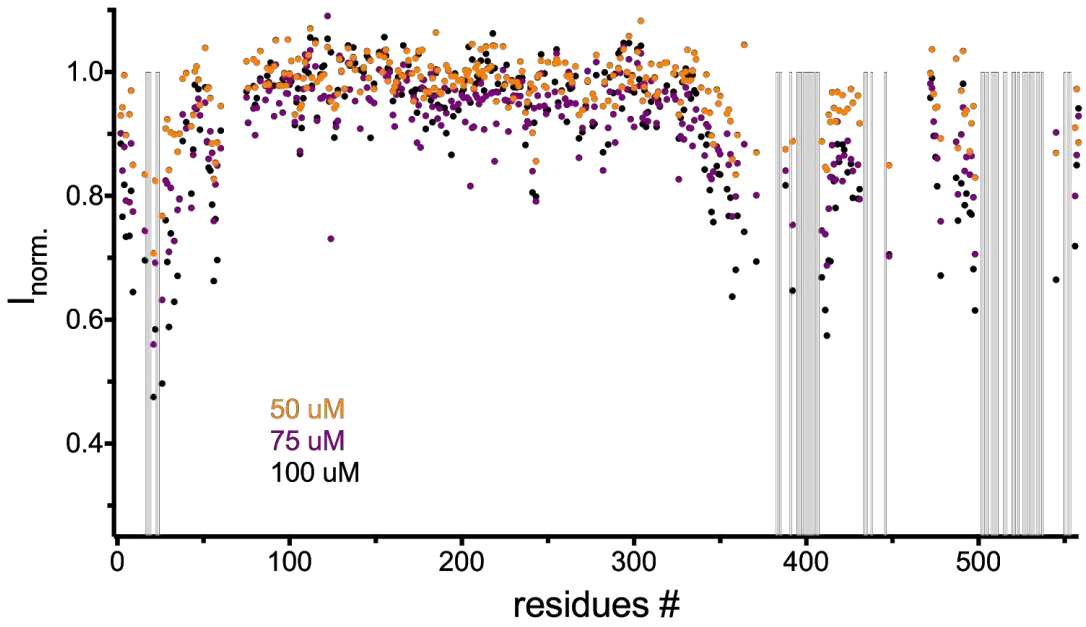
S6. Sequence of the AR AD (AR_{1-558})

Hydrophobic amino acids: Alanine (A), Leucine (L), Isoleucine (I), Valine (V), Methionine (M) in yellow, aromatic amino acids: Tyrosine (Y), Phenylalanine (F) and Histidine (H) in blue, small flexible amino acids: Proline (P), Glycine (G) in green. Red dashed squares indicate regions involved in long-range intermolecular interactions of the AD derived from a concentration dependent loss in peak intensities in the 1H-15N HSQC spectra of the AD construct (Fig. 2D and S8).



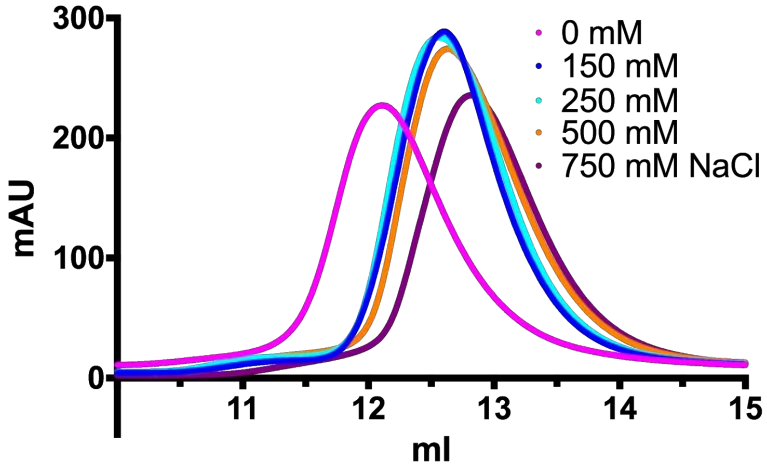
S7. The AR AD sequence amino acid profile

Plot of amino acid enrichment and depletion in comparison to sequences deposited in the DisProt3.4 database (Vacic et al. 2007).



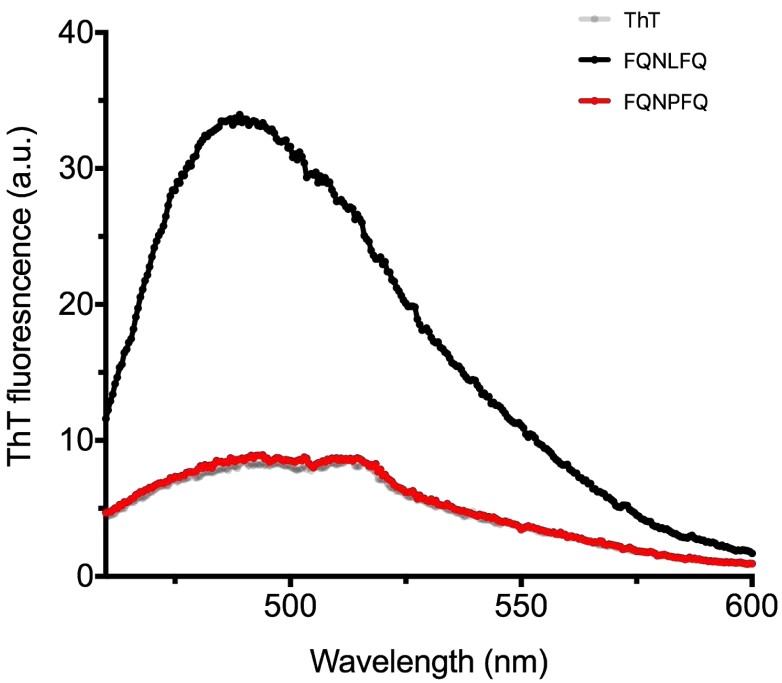
S8. Residue specific peak intensity loss in the AD 1H-15N HSQC spectra upon concentration

Residue specific loss of the peaks intensity in the 1H-15N HSQC spectra upon concentration of the AD construct (50 µM, 75 µM, 100 µM) normalized by their intensity at 25 µM. Grey bars represent peaks that are not observable in the 1H-15N HSQC spectrum in the context of the 100 uM AD compared to the 1H-15N HSQC spectra of the subdomains of the AD: $AD_{1\text{-}155}$, $AD_{142\text{-}448}$, $AD_{441\text{-}558}$ at comparable concentrations.



S9. Size exclusion chromatography (SEC) profile of the AD construct at different ionic strengths

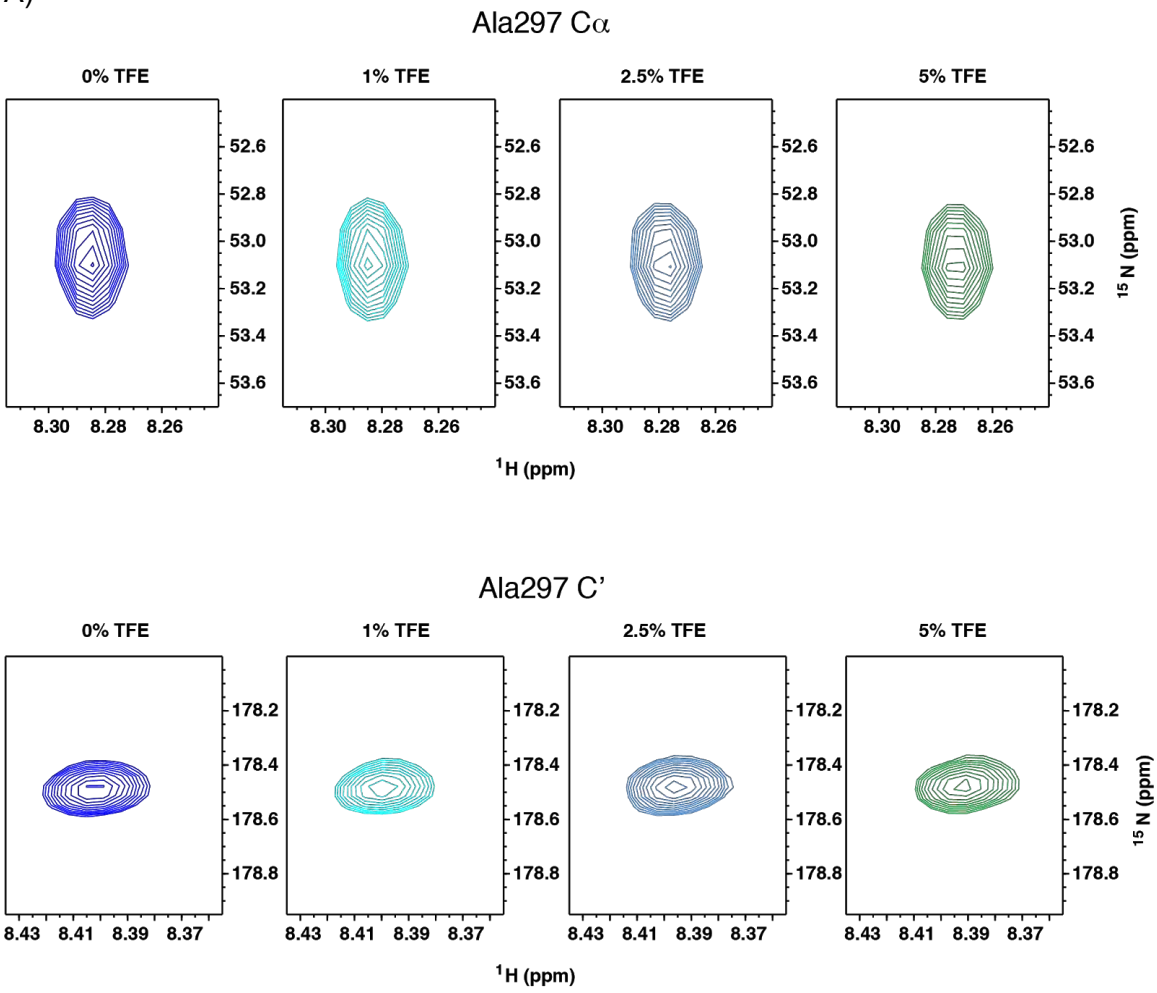
The AD construct shows elution at higher volumes from the SEC column upon increase of the ionic strength (from 0 to 750 mM NaCl). Column used - Superdex 200 Increase 10/300 GL.



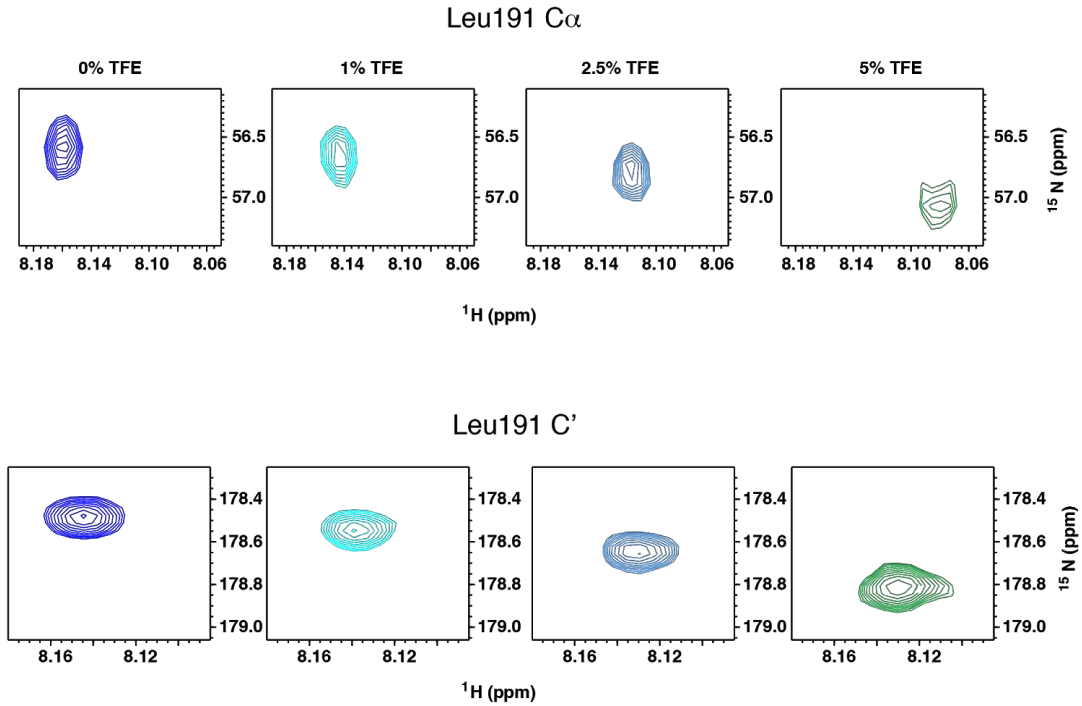
S10. FQNLFQ and FQNPFQ peptides Thioflavin T binding

Thioflavin T (ThT) binding of the FQNLFQ (black line) and FQNPFQ (red line) peptides after an overnight incubation shows abrogation of peptide aggregation upon introduction of L26P mutation.

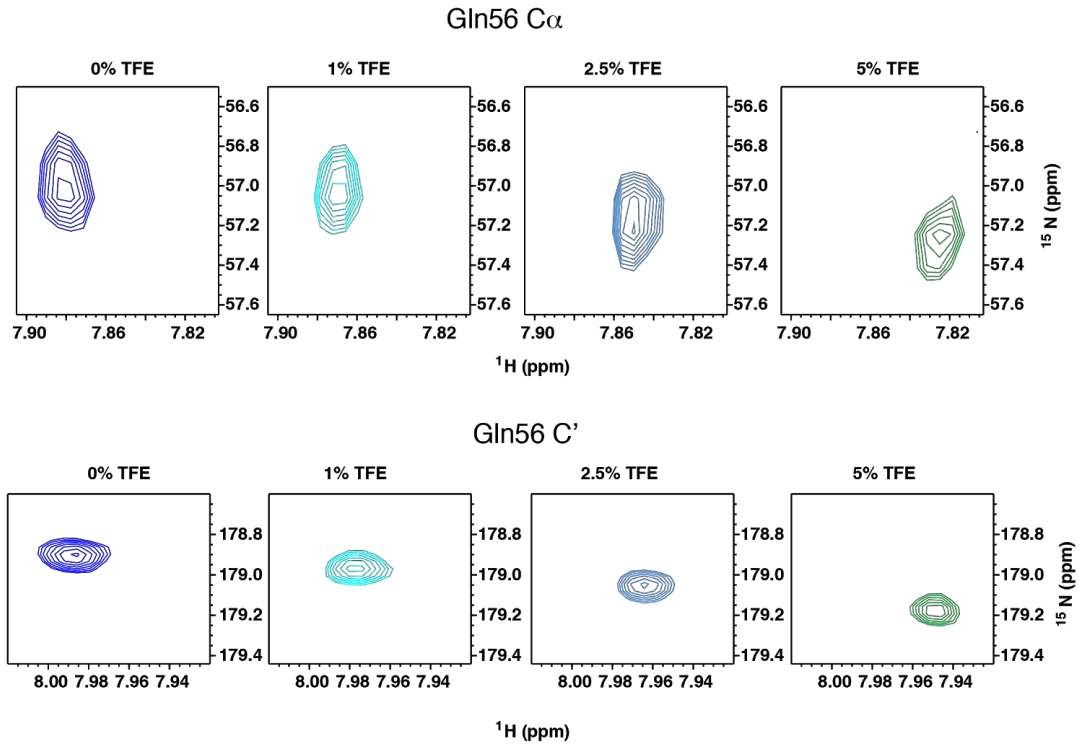
A)


Ala297 Cα
0% TFE
1% TFE
2.5% TFE
5% TFE
52.6
52.8
53.0
53.2
53.4
53.6
8.30 8.28 8.26
15 N (ppm)
1H (ppm)
Ala297 C'
178.2
178.4
178.6
178.8
8.43 8.41 8.39 8.37

B)

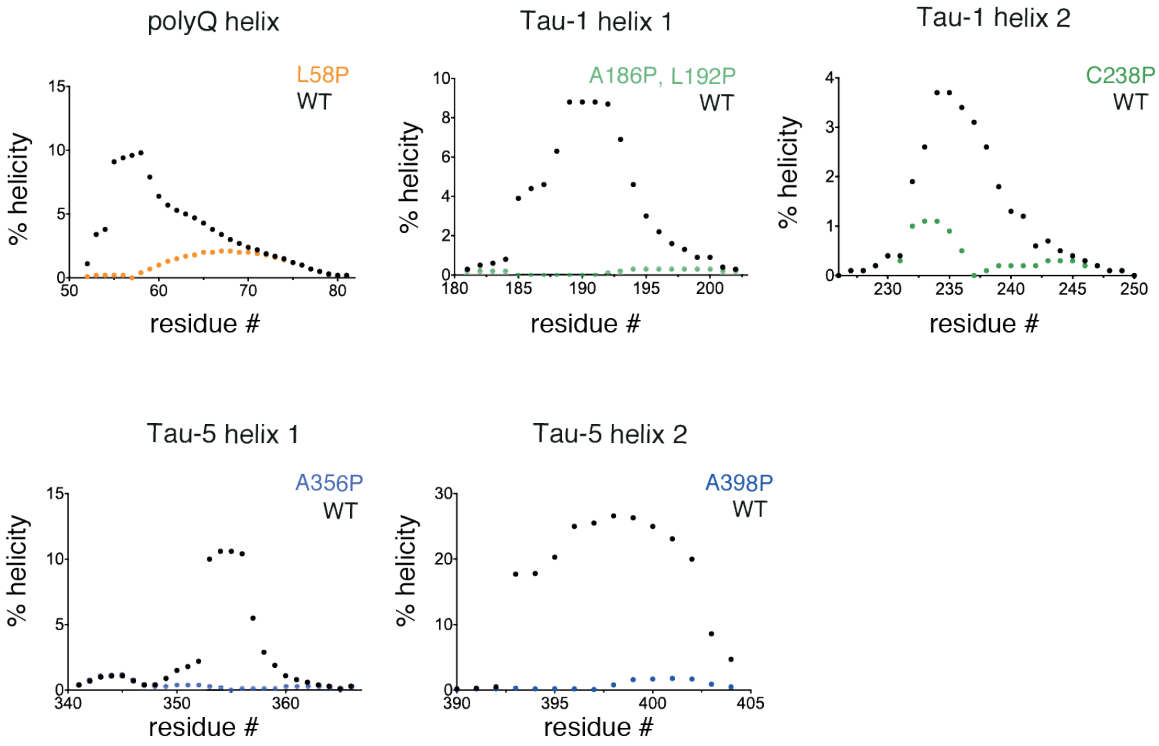

Leu191 Cα
0% TFE
1% TFE
2.5% TFE
5% TFE
56.5
57.0
8.18 8.14 8.10 8.06
15 N (ppm)
1H (ppm)
Leu191 C'
178.4
178.6
178.8
179.0
8.16 8.12
15 N (ppm)
1H (ppm)

C)


Gln56 Cα
0% TFE
1% TFE
2.5% TFE
5% TFE
56.6
56.8
57.0
57.2
57.4
57.6
7.90 7.86 7.82
15 N (ppm)
1H (ppm)
Gln56 C'
178.8
179.0
179.2
179.4
8.00 7.98 7.96 7.94
15 N (ppm)
1H (ppm)

S11. Expanded regions of the HNCA (top) and the The HNCO (bottom) spectra showing the cross-peaks of the Cα and C' resonances of residues A) Ala297, B) Leu191, and C) Gln56

The secondary structure of the residue A) Ala 297 does not change in the TFE concentrations used. The helicity of the residues B) Leu 191, C) Gln 56 increases upon TFE addition.



S12. Helicity of the indicated P mutants of the AD

Prediction of the helicity profile of the activation motifs by Agadir software (Muñoz and Serrano 1997).