



UNIVERSITAT DE
BARCELONA

Targeting the disordered N-terminal domain of androgen receptor

Novel inhibitors tackling biomolecular condensation to treat late stage prostate cancer

Marta Frigolé Vivas

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



TARGETING THE
DISORDERED
N-TERMINAL DOMAIN
OF ANDROGEN
RECEPTOR

NOVEL INHIBITORS
TACKLING
BIOMOLECULAR
CONDENSATION TO
TREAT LATE STAGE
PROSTATE CANCER

MARTA
FRIGOLÉ
VIVAS



UNIVERSITAT DE
BARCELONA

Targeting the disordered N-terminal domain of androgen receptor

Novel inhibitors tackling biomolecular condensation to treat late stage prostate cancer

Tesi doctoral realitzada en el Programa de Doctorat en Química Orgànica de la Universitat de Barcelona.

Memòria presentada per Marta Frigolé Vivas per optar al títol de doctora en Química Orgànica per la Universitat de Barcelona.

Marta Frigolé Vivas

Doctoranda

Xavier Salvatella Giralt

Director de la Tesi

-
ICREA
IRB Barcelona

Antoni Riera Escalé

Director i tutor de la Tesi

-
Universitat de Barcelona
IRB Barcelona



UNIVERSITAT DE
BARCELONA

Targeting the disordered N-terminal domain of androgen receptor

Novel inhibitors tackling biomolecular condensation to treat late stage prostate cancer

Marta Frigolé Vivas

2020



UNIVERSITAT DE
BARCELONA

This work was executed from september 2015 to july 2020 and supported by a “La Caixa” PhD fellowship from “la Caixa” - Fundació Bancària “la Caixa” (LCF/BQ/SO15/52260042) for the 2015-2019 period and by a Generalitat de Catalunya “grup consolidat” AGAUR grant (2017 SGR 324) for the 2019-2020 period.

The experimental work was carried out in the “Unitat de Recerca en Síntesi Asimètrica” and “Laboratory of Molecular Biophysics” laboratories at the IRB Barcelona.

ACKNOWLEDGEMENTS

The PhD has been a journey in which I have grown professionally and personally. It has taught me about science and myself, about competition and collaboration, about success and failure. I would like to thank all the people that have been part of this journey one way or another.

First, I would like to thank my PhD supervisors. Toni, thank you for welcoming me to the IRB and introducing me to this challenging project. I will always remember the day I first sat down at your office and you encouraged me to apply for the IRB fellowships. Also thank you for your time and advice whenever I have needed it and for keeping your door always open.

Xavier, thank you for allowing me the space, time and resources to deepen into a new field for me and learn all about proteins and cells with a chemistry mindset. We know that phase transitions, from soluble to demixed or chemistry to biophysics, are not easy processes. The learning curve has been exponential. Thank you for the time devoted to discussions and analysis, for allowing me to be independent, to push the project forward and to stay positive in those moments where no experiment made sense.

Second, I would like to thank all the people that have actively worked to advance this project onwards:

To Dr. Carolina Sánchez, for all your meticulous chemical synthesis and rigorous documentation. It is a real pleasure to work with you! All the analogues you have synthesised have been a real contribution to validating our hypothesis. Thank you!

To Dr. Alessandro Ruffoni. When you started we were completely in the dark. I will always remember when we realised that EPI-001 signals were decreasing with time and how this led to the “oligomeric state” hypothesis. Thank you for your ideas, always thinking out-of-the-box, for all the Italian coffee, for all the parties and chats. It was only one year in which we shared the project and the fume-hood, but what a year!

To Mateusz Biesaga, thank you for believing in this project and taking it to the next step. When you joined the project I felt for the first time what team-work was. I have been very lucky to meet my ideal partner in crime (science-wise). I enjoy our chats a lot, also the planning and the communication between us :).

To the two students I have had the honor to work with and mentor, Sílvia and Carla. Both of you worked on a project that is not even mentioned in this thesis, but not because of that is less important. Thank you for your time, energy and willingness to learn (and continue the scientific career!).

To all the collaborators I have had the pleasure to work with. Collaboration in science and in life is essential. Thank you to:

- Dr. Marianne Sadar and her team for the incredible effort in the biological evaluation of the CAM compounds.
- Dr. Karina Pombo-Garcia for all the effort you have put in the STED imaging and for making me feel at home during the days we spent together in Dresden.

- Dr. Tony Hymann and his lab for putting the EMBO practical course together where I did not only learn a lot but also meet great friends.
- Dr. Tanja Mittag for building a nice collaboration that will lead to important discoveries with regards to the implications of SPOP in the AR degradation pathway in the absence and presence of the CAM compounds.
- Dr. Isabelle Burn-Health for always welcoming me into the EBL and being available for discussions.
- Dr. Xavier Barril and Dr. Maciej Majewski for the hard work in the RAP74 project even if unfortunately it did not deliver any new potent inhibitor.

To the PhD committee, Dr. Javi Luque, Dr. Fèlix Urpí and Dr. Àngel Nebreda, for your time and advice, for believing in the project and encouraging me to develop it further.

To Poble studio and specially to Mònica Iglesias for the amazing thesis cover that captures the essence of this work and myself.

Third, to all the scientists that have taught me and helped me throughout the years. Having an interdisciplinary project is one thing. Being interdisciplinary is another. This has meant a lot of walking distance within the IRB from one lab to another but most importantly having learnt from you all. Thank you:

- Dr. Marga Gairí, for always being there for me, interested in the science but also genuinely supportive in all the aspects related to it.
- Dr. Jesús García, for all the time you have spent teaching me NMR and sitting with me in front of the spectrophotometer. Also for all your funny/historical stories during lunchtime.
- Dr. Mar Vilanova and the MS facility, for the enormous effort in analysing the samples, for being gentle in telling me that there was no reaction and for caring about the progression of the project.
- Dr. Lídia Bardia, Dr. Anna Lladó and Dr. Nikos Giakoumakis and the microscopy facility, for being so helpful with the FRAP, the macros and the settings. For all the stories and laughs next to the spinning disk, although I think the Zeiss is my favourite now ;)
- Vicky Ruiz-Calero, for receiving my petitions for aminoacid analysis with a big smile.
- Dr. Delia Zafra and Israel Ramos for believing in me and the project from the beginning.

Fourth, to the members of the lab with whom we have shared bench, victories and pains for five years.

- At LMB: Paula (me encanta estar contigo, mil gracias por todo tu tiempo, risas y amistad), Mireia (per divertir-nos, pels cafès i per les abraçades), Claudio di Sanza (for being brave enough to introduce me to cell biology), Víctor Buzón (per una amistat que travessa les fronteres del lab, per les abraçades i les discussions fins ben entrada la matinada), Jomi (per compartir el "pain" d'expressar i purificar EDM4, per les teves bromes que fan el dia més amè, per tots els dinars plegats), Isabel (per fer-me companyia a les 8 del matí, per la teva calma i serenitat), Ela (for being my

“big sister”, for the advise and desk-to-desk chats), Busra (for being so adventurous, always happy to have a tea), Albert (per ser un company de congrés genial, pel nostre viatge a Canadà que va ser brutal), Laura (for your work-hard/party-hard energy, for welcoming me to italy, for being an incredible housemate) and Jonathan (for your kindness and willingness to help).

- At URSA: Dan and Craig (for making chemistry life much more fun, for all the parties and bad jokes), Anna (per tenir sempre un moment per escoltar, per ensenyar-me balls i com maquillar-me), Enric, Salva and Edu (pel carinyu especial que ens tenim, per estar sempre disposats a ajudar-nos i fer unes risas).
- At EBL: To all of you and specially to Núria for all the cell biology advice and help with the Cary spectrophotometer.

Finalment, tinc la gran sort a la meva vida d'estimar i ser molt estimada. Tots vosaltres formeu un gran coixí flonjo i suau que em protegeix de les caigudes a la vegada que m'empeny gentilment a seguir endavant, a aprendre. Gràcies!

- Albert, quina sort tenir-te. Durant aquests 5 anys n'hem passat de tots colors però tu sempre hi has sigut. Gràcies pels consells, per fer-me riure, per les cites per Barcelona, per la cara d'alegria que fas quan entro per la porta del lab, per posar-me al dia de tots els cotilleos, per creure en mi. El més preuat que m'emporto del doctorat és un amic incondicional i això em fa molt feliç! Sempre trobarem un moment i un lloc per posar-nos al dia, explicar-nos coses i veure'ns, allà on sigui que ens porti la vida a partir d'ara.
- Nenes, sense amigues la vida no és tan divertida ni tan compartida. Gràcies per acompanyar-me tots aquests anys, perquè ja en fa molts que som inseparables! Estel, Olga i Mònica, la nostra amistat és un tresor que guardo ben protegit dins del meu cor.
- Anna i Maria (i els dos que falten i la gossa). De cada etapa de la vida he intentat endur-me'n un aprenentatge i una amistat. I quina amistat la del B! Juntes ens hem obert pas cap a la universitat i Barcelona, cap a noves aventures, i viatges, i amors, i desamors, i així us heu convertit en una part essencial de mi. I amb el temps la família ha crescut amb en Pol, en Llorenç i la Magrana. Gràcies per fer-me costat i per totes les paelles d'en Pol, escalades i aventures vàries que sempre prometen amb en Llorenç i per tots els fills de la Magrana que puguem algun dia adoptar.
- Magalí, perquè ja de ben petites hem tingut vides gairebé paral·leles. M'ha encantat retrobar-nos al doctorat i al tren, compartint històries i histèries, amistat i família.
- Núria, d'una companya de tren excel·lent a una amiga. Gràcies per fer els viatges més amens, per fer-me parlar amb tots els desconeguts del tren i per moltes escaldades!
- Coral, gràcies per estar disposada a fer una bona copa de vi cada vegada que se'm morien les cèl·lules. Qui ho hagués dit que, d'aquell pis de Sant Gervasi, en sorgiria una amistat tan bonica. Paula, gràcies per obrir-me un món nou de diversió i amistat a Barcelona. Les dues (i els gigis) ja sou família :D.
- Carles, gràcies per tots els cafès i consells, per fer temps sempre que l'he necessitat i per ser un company de congrés genial, sempre a punt per fer-ne alguna i riure.
- Jordi, gràcies per donar-me la oportunitat per descobrir. Descobrir altres ma-

neres de fer, altres relacions, altres esports, altres indrets, altres feines i, per tant, per descobrir-me a mi mateixa. Gràcies per fer-me reflexionar, per les infusions fins ben entrada la nit i per l'acompanyament. Per molts més cafès i aventures!

- Padrino i l'skimo team, gràcies per obrir-me al món de l'esquí de muntanya, per les rialles i les excursions (i totes les que vindran!).
- Gràcies també a tota la família que sempre porto al cor: padrina (per ser-hi sempre, per cuidar-me, t'estimo), Carla (per les teves abraçades, alegria i desimboltura que sempre em sorprèn d'alguna forma o altra), Eva (per tots els consells i guidance), Assumpció (per ensenyar-me que el món és molt gran, perquè em sento molt bé quan hi ets), Padrí, Tieta i cosines (per l'alegria i bogeria que aporteu, veient la vida en positiu).
- Avis de casa, avis del meu cor, per ser els meus millor amics, per ser incondicionals i per, encara ara, acompanyar-me en totes les aventures. Gràcies per cuidar-me, per estimar-me. Sóc avui el fruit del vostre amor.
- Papa i mama, quina sort la meva! Gràcies per donar-m'ho tot. Per ser-hi sempre, per donar-me suport, per animar-me a explorar, i, sobretot, per estimar-me tant. I van passant els anys i anem construint aquest petit nucli que som, tan nostre, tan comfortable, tan unit. No tinc més paraules que gràcies, us estimo!
- Xevi, compartir amb tu el camí de la vida és un regal que em fa feliç cada dia. Gràcies per estar sempre present, per les abraçades i per totes les aventures que hem tingut i les que ens esperen. Junts.

A tots i totes, moltes gràcies per ser i ser-hi!

Confía, al final todo sale bien.

- Argia de la Peña Lago

*Estimar és un lloc.
Perdura al fons de tot: d'allí venim.
I es el lloc on va quedant la vida.*

- Joan Margarit

The present thesis is dedicated to Xevi
and to my parents, Núria and Jordi

LIST OF RELEVANT ABBREVIATIONS

¹H,¹⁵N CP-HISQC	Cross-polarization assisted heteronuclear in-phase single-quantum correlation	LCST	Lower critical solution temperature
¹H,¹⁵N HSQC	Two-dimensional ¹ H, ¹⁵ N single quantum correlation	LLPS	Liquid-liquid phase separation
ADT	Androgen deprivation therapy	NLS	Nuclear localisation signal
AR	Androgen receptor	NMR	Nuclear magnetic resonance spectroscopy
AR-DBD	Androgen receptor DNA-binding domain	PC	Prostate cancer
AR-LBD	Androgen receptor ligand-binding domain	Pc	Partition coefficient
AR-NTD	Androgen receptor N-terminal domain	PLA	Proximity ligation assay
CD	Circular dichroism spectroscopy	PPI	Protein-protein interaction
CRPC	Castration-resistant prostate cancer	SAR	Structure-activity relationship
Csat	Saturation concentration	SDBB	Structure-based drug discovery
CSP	Chemical shift perturbation	STED	Stimulation emission depletion microscopy
DHT	Dihydrotestosterone	Tc	Cloud point temperature
FRAP	Fluorescence recovery after photobleaching	TFE	2,2,2-trifluoroethanol
ID	Intrinsic disorder	TSA	Turbidity shift assay
IDPs	Intrinsically disordered proteins	UCST	Upper critical solution temperature

TABLE OF CONTENTS

1. Introduction and objectives	21
1.1. Intrinsically disordered proteins	23
1.2. Androgen receptor and castration-resistant prostate cancer	25
First objective: Study of the reversible interaction of EPI-001 with its target by solution NMR spectroscopy	28
1.3. Intrinsically disordered proteins and biomolecular condensation	29
Second objective: Study of the role of EPI-001 as a modulator of androgen receptor condensation	32
Third objective: Design, synthesis and evaluation of novel EPI-001 analogues	33
Bibliography	34
2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy	39
2.1. Background	41
2.2. Experimental results	60
2.3. Discussion	79
Bibliography	81
3. EPI-001 as a modulator of androgen receptor condensation	87
3.1. Background	89
3.2. Experimental results	111
3.3. Discussion	125
Bibliography	127
4. Rational design, synthesis and biological evaluation of novel EPI-001 analogues	135
4.1. Background	137
4.2. Experimental results	138
4.3. Discussion	152
Bibliography	154
5. General discussion and conclusions	157
6. Experimental methods	165
Detailed index	197

1. INTRODUCTION AND OBJECTIVES

Drug discovery is the process by which new drugs are found and developed into the market ¹. Given the scope of this thesis, the term “drug” will refer to small molecules and the term “target” to proteins, otherwise noted.

Currently, most drug discovery strategies rely on the resolved three-dimensional structure of the target. Structure-based drug discovery (SBDD) ²⁻⁴ is advantageous as it accelerates and lowers the cost of the drug discovery process. However, the essential requisite for SBDD, the prior knowledge of the target protein structure, has narrowed the types of proteins in drug discovery programs to globular proteins ⁵.

It is estimated that 33-50% of our proteome is intrinsically disordered ^{6,7}. Intrinsically disordered proteins (IDPs) and protein regions (IDRs) do not spontaneously fold into a stable secondary and tertiary structure in their native state yet are functionally active ⁸⁻¹². IDPs have long been considered undruggable as their lack of a stable tertiary structure impedes the application of SBDD. Nevertheless, IDPs participate in many physiological processes ¹³, as well as in diseases ¹⁴. **Hence finding new approaches to target IDPs is crucial to expand the druggable proteome and fill unmet needs across diseases in which disorder plays a role.**

1.1. Intrinsically disordered proteins

The discovery and study of IDPs (and IDRs, which will be included in the term IDP for the rest of the thesis) in the late 1990s questioned the central dogma in molecular biology that structure defines function, exemplified by the “lock-and-key” model put forward by Emil Fisher in 1894 ¹⁵.

IDPs lack a defined secondary and tertiary structure. They have multiple minima in their free-energy landscape and can therefore populate several distinct conformations (from fully extended to completely collapsed). Their overall structure is better defined as a dynamic ensemble of many discrete conformational states, each of which is populated with a certain statistical weight. The structure of IDPs contrasts with that of globular proteins, which have a well-defined global minima in their free-energy landscape and their three-dimensional fold, therefore, fluctuates around one equilibrium position ¹⁶ (**Fig 1.1.1**).

1. Introduction and objectives

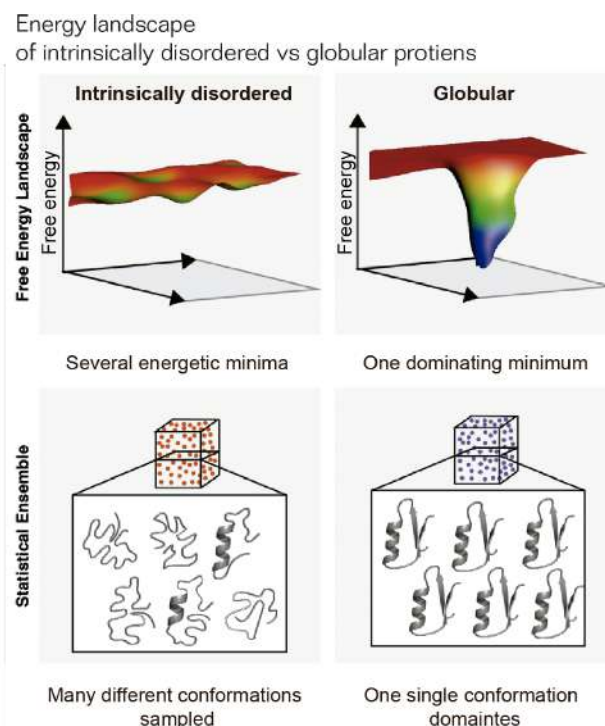


Fig. 1.1.1. Comparison of the energy landscape and conformational ensemble of IDPs and globular proteins. Adapted from Flock et al. 2014 ¹⁶.

Intrinsic disorder (ID) propensity is encoded in the primary sequence of proteins. Compared with folded proteins, IDPs have an overall low mean hydrophobicity and high net charge ^{9,17}. Protein folding is a spontaneous process. When the hydrophobic core of a protein collapses, there is a large release of water molecules, thereby increasing the entropy of the system. Once the protein is depleted of water, the enthalpy of the interactions between amino acid residues is enhanced (hydrogen bonds and van der Waals interactions) ¹⁸. The low hydrophobic content of IDPs leads to a weakened hydrophobic effect, thereby preventing their folding into a stable globular structure.

Globular proteins are involved mostly in processes requiring a large control of the three-dimensional structure, for instance enzymatic reactions. In contrast, IDPs have prevalent roles in regulation and signalling processes and in protein-protein interaction hubs ⁸. For example, ID is widespread in transcription and transcription regulation ¹⁹, particularly in the activation domain of transcription factors ²⁰. More recently, it has been demonstrated that IDPs are also involved in biomolecular condensation ²¹⁻²³, which will be further discussed in sections 1.3 and 3.1.

IDPs can: i) interact with their partners with both high specificity and low affinity, i.e. their interactions are highly specific and short-lived ²⁴, with some exceptions ²⁵; ii) rapidly respond to changes in the environment ^{26,27}; iii) bind to distinct partners through single and/or multivalent interactions ^{28,29} and iv) bind to multiple partners. In the bound state, IDPs can remain disordered or “fuzzy” ³⁰ or become structured ³¹. Most binding events are through short motifs embedded within larger disordered sequences (named molecular recognition features or short linear motifs, MoRFs). MoRFs are evolutionarily conserved and usually amphiphatic, which allows for the gain in secondary structure when binding ³².

To avoid promiscuous binding, IDPs are tightly regulated by post-transcriptional modifications (PTMs) and/or by ensuring appropriate protein levels during the required time. On the one hand, PTMs can change the energy landscape of IDPs, thus stabilising or destabilising secondary structures and causing a switch between disordered and ordered states³³. On the other hand, ensuring appropriate IDP levels is important to prevent non-specific interactions³⁴. The cellular concentration of IDPs is regulated at all steps of production from transcript clearance and translation rate to protein degradation³⁵, as exemplified by the observation that IDPs have shorter half-lives than structured proteins³⁶.

The misregulation of IDPs is associated with disease. In fact, the D² concept (disorder in disorders) was introduced to highlight the involvement of ID in numerous conditions¹⁴, including cancer, neurodegeneration, cardiovascular diseases and diabetes^{14,34,37}. For example, the disordered oncogene c-Myc and the cancer suppressor p53 have altered abundance in many types of cancers. Also, mutations or misregulation of some disordered proteins like amyloid β -peptide and α -synuclein promote aggregation in neurodegenerative conditions like Alzheimer's disease and Parkinson's disease, respectively.

IDPs are promising drug targets³⁸. They are very abundant in our proteome^{6,7}, are involved in many important biological functions, and are over-represented in major pathologies¹⁴.

Up to now, most successful strategies to target IDPs have avoided the ID nature of the target³⁷. Drugs have been directed to the enzymes that regulate IDPs, the ordered domain of a protein with large ID, or the globular binding partner. Only a few strategies are aimed to drug the ID domain directly. In fact, only two compounds targeting an IDP have reached clinical trials: MSI-1436, which targets the protein tyrosine phosphatase 1B for the treatment of obesity (Trodonquimine, NCT00509132) and cardiovascular diseases (Trodonquimine, NCT00005696), and **EPI-506, which targets the disordered N-terminal transactivation domain of the androgen receptor (AR-NTD) for the treatment of castration-resistant prostate cancer** (Ralaniten, NCT02606123).

1.2. Androgen receptor and castration-resistant prostate cancer

The androgen receptor (AR) is a hormone-activated transcription factor that regulates the expression of specific genes related to the development and maintenance of the male phenotype³⁹. The AR is a member of the steroid-activated nuclear receptors (NRs)⁴⁰. Other members include the estrogen receptor (ER), the progesterone receptor (PR), the glucocorticoid receptor (GR), the thyroid receptor (TR) and the mineralocorticoid receptor (MR).

1. Introduction and objectives

Like all NRs, the AR has the following domain architecture ⁴¹ (**Fig. 1.2.1**):

- *N-terminal intrinsically disordered transactivation domain (AR-NTD)*

The NTD is the least conserved domain among the NR family. The AR-NTD is the largest in size (558 amino acids, depending on the variation of several polymorphic regions) and the most important for transcriptional activity of the AR ⁴¹. The NTD contains the activation function 1 (AF-1). The AF-1 in AR bears the main transactivation function of the receptor, as opposed to the other members of the family, in which it is located in the AR-LBD (in the activation function 2, AF-2). The AF-1 can be functionally subdivided into two transactivation units (Tau-1 and Tau-5) ⁴².

The AR-NTD is the least characterised domain as it is intrinsically disordered. However, it has several MoRFs that are key for function, like the ²³LQNLF²⁷, the ¹⁷⁹LKDIL¹⁸³ and the ⁴³³WHTLF⁴³⁷ motifs ⁴¹. More specific sequence-structure-function details are described in section 2.1.3.

- *DNA-binding domain (AR-DBD)*

The DBD is the most conserved domain among NRs. For the AR, the DBD binds as a dimer to the androgen response elements (AREs) on the DNA ⁴¹.

- *Hinge region*

The hinge region serves as a short flexible linker between the DBD and the LBD. It contains the nuclear localization signal (NLS) ⁴¹.

- *C-terminal ligand (steroid)-binding domain (AR-LBD)*

The overall structure of the LBD is similar between the members of the family. Steroid binding, dihydrotestosterone (DHT) in the case of the AR, causes a conformational rearrangement, exposing the AF-2 surface ⁴¹. The AF-2 binds LXXLL motifs of co-activators ⁴³, although in the case of the AR, it has higher affinity for the ²³LQNLF²⁷ motif in the AR-NTD, establishing the N/C interaction ⁴⁴.

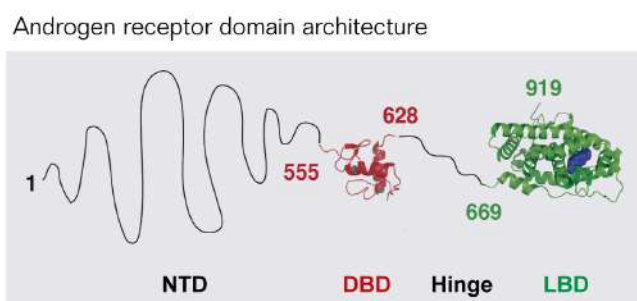


Fig. 1.2.1. Androgen receptor domain architecture: N-terminal intrinsically disordered transactivation domain (AR-NTD), the DNA-binding domain (AR-DBD), the hinge region and the ligand-binding domain (AR-LBD). Adapted from De Mol et al. 2016 ⁴⁵.

1. Introduction and objectives

-

The AR signalling pathway is depicted in Figure 1.2.2³⁹. Testosterone is produced in the testis and transported to the target tissues (e.g. the prostate). In prostate cells, testosterone is converted to dihydrotestosterone (DHT) by the enzyme 5 α -reductase. Unactivated AR is bound to chaperones in the cytosol. DHT binding to the AR-LBD causes dissociation of the chaperones, a structural rearrangement (the N/C interaction) and the translocation of the receptor into the nucleus. In the nucleus, the AR dimerises and binds to the AREs in the promoter region of genes such as the prostate-specific antigen (PSA) gene. Bound at the promoter, the AR recruits members of the basal transcription machinery and other coregulators to start the transcription of genes related to the development and maintenance of the male phenotype (Fig. 1.2.2).

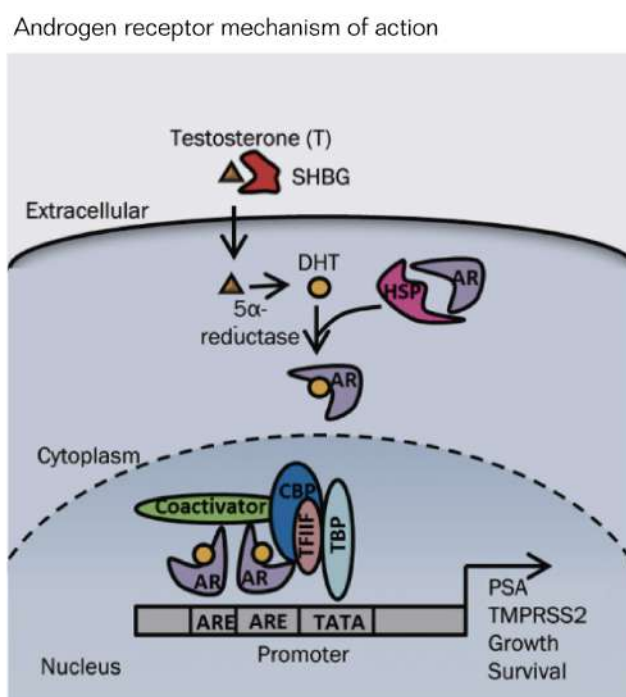


Fig.1.2.2. Mechanism of action of the androgen receptor. Adapted from Tan et al. 2015³⁹.

The AR is central in many diseases, including androgen-insensitivity syndrome, spinal and bulbar muscular atrophy, benign prostatic hyperplasia and prostate cancer (PC). Given the scope of this thesis, we will focus on PC.

PC is the second most commonly diagnosed cancer worldwide and the sixth leading cause of death by cancer in men⁴⁶. In 2018, it was the most common cancer diagnosed and the third cause of cancer death amongst men in Europe⁴⁷. Elevated PSA blood level is a biomarker for PC⁴⁸.

PC is initially localized in the prostate, hence prostatectomy and radiotherapy are the initial treatment options. PC tumour growth and proliferation depends on the transcriptional activity of the AR⁴⁹. For this reason, men with intermediate to high recurrence of PC and relapsed PC receive androgen deprivation therapy (ADT)⁵⁰. ADT is based on reducing androgen production. This is achieved by surgical or chemical castration or by blocking androgen binding to the AR-LBD with antiandrogens such as bicalutamide, enzalutamide and apalutamide. ADT is given alone or in combination with abiraterone (that inhibits *de novo* synthesis of androgen by inhibiting the enzyme

CYP17A1) or docetaxel (chemotherapy).

Unfortunately, most PC patients become refractory to these treatments and progress into castration-resistant prostate cancer (CRPC)⁵¹. Several resistance mechanisms include, among others, genomic amplification of the AR locus, emergence of AR splice variants lacking the AR-LBD, and gain-of-function mutations in the AR-LBD and in other factors like SPOP, MED12 or FOXA1^{52,53}.

CRPC currently has no cure and it is characterised by a very poor prognosis as it is responsible for the death of *ca* 35,000/40,000 American/European men each year according to GlobalData. Over 80% of CRPC cases are metastatic (mCRPC). Although there is no cure for CRPC, there are some therapeutic options: i) as this cancer still relies on the AR signaling axis, CRPC patients continue to receive ADT (enzalutamide and abiraterone) in combination with chemotherapy (docetaxel); ii) a second option is immunotherapy (spirucel-T) and, iii) in those cases with bone metastasis, treatment with a radioligand (radium 223 dichloride) is possible⁵⁰. **Finding new strategies to fight CRPC is of paramount importance.**

The AR-NTD is a promising and validated drug target for CRPC (see section 2.1.4). However, its ID nature has hampered the finding and progression of AR-NTD inhibitors into the market⁵⁴. The Sadar laboratory pioneered the discovery of AR-NTD inhibitors. EPI-001 was the first small molecule reported to inhibit the AR-NTD and provided a starting point to find new drugs that target the AR and thus fight CRPC^{55,56}. EPI-506, a pro-drug of EPI-001, entered clinical trials in 2015. Nevertheless, its mechanism of action was not fully understood. Given the expertise of Dr. Xavier Salvatella in AR biophysics and Dr. Antoni Riera in medicinal chemistry, some years ago their laboratories started a project to study the mechanism of action of EPI-001. Initial work by the lab showed that EPI-001 binds to the AR-NTD by means of nuclear magnetic resonance (NMR) experiments⁴⁵. Specifically, 10 molar equivalents of EPI-001 caused chemical shifts perturbations in the three helices of the Tau-5 region.

First objective: Study the reversible interaction of EPI-001 with its target by solution NMR spectroscopy

The first objective of the thesis was to continue the NMR studies on the interaction of EPI-001 and the AR-NTD and investigate whether structuration of the three helices in the Tau-5 could improve the binding affinity of EPI-001 for the AR-NTD. In this regard, using solution NMR, we aimed at characterising the chemical shift perturbations that EPI-001 caused on several constructs spanning the AR-NTD in the absence and presence of the helix-stabilising co-solvent 2,2,2-trifluoroethanol.

The work towards this objective is described in **Chapter 2** of this doctoral thesis.

During the course of this work we realised that the AR-NTD and the full length AR undergo biomolecular condensation⁵⁷. As EPI-001 interacts with the AR-NTD and the AR-NTD condensates, we hypothesised that the mechanism of action of EPI-001 could be related to its ability to modulate AR condensation.

1.3. Intrinsically disordered proteins and biomolecular condensates

The cytoplasm and nucleoplasm are very densely packed environments⁵⁸. To achieve spatiotemporal control, cells localise reactions in organelles²². Classical organelles are the Golgi apparatus and the Endoplasmic Reticulum. All these organelles are membrane-bound, i.e. they have a physical separation between the interior and the exterior of the organelle, and the partition of molecules is regulated by the membrane transport machinery. However, many other organelles lack a physical separation around them and yet are able to concentrate biochemical processes. Examples include the nucleoli (for ribosome biogenesis), the Cajal bodies (for assembly and or/modification of splicing machinery), the PML bodies (for transcriptional regulation and apoptosis signaling) and the stress granules (for mRNA storage upon stress)²² (**Fig 1.3.1**).

Representation of membrane-less organelles in eukaryotic cells

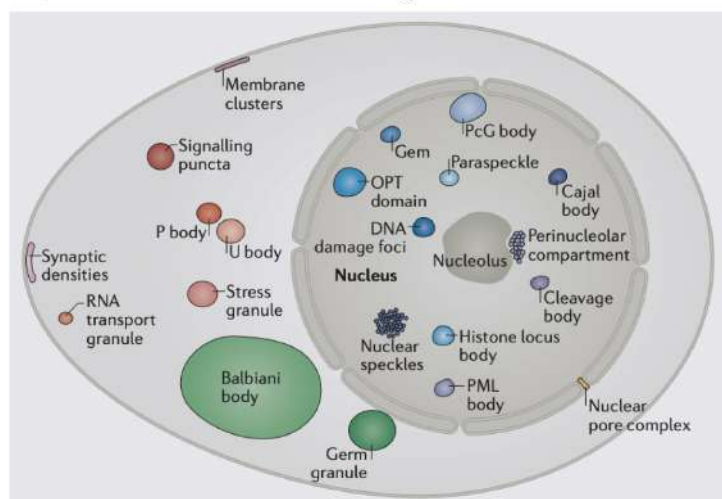


Fig. 1.3.1. Representation of the different membrane-less organelles found in eukaryotic cells. Adapted from Banani et al. 2017²².

While membrane-less compartments differ in composition, subcellular location and function, they all share the ability to concentrate biomolecules and are formed by phase transitions. For this reason they are termed biomolecular condensates²².

Phase transitions have been widely studied in physics but their application in living systems is an important emerging principle in biology⁵⁹. Liquid-liquid phase separation (LLPS) is one particular type of phase transition and the most physiologically relevant and studied so far in the condensates

1. Introduction and objectives

field⁶⁰. The first evidence that membrane-less organelles were liquid and formed by LLPS was reported in 2009 in a landmark paper by Dr. Clifford Branwynne, Dr. Tony Hyman and Dr. Frank Jülicher⁵⁹. They characterised P granules in the germline of *Caenorhabditis elegans* and demonstrated that they behaved like liquids: they were round and could drip, wet the surface and undergo fusion events.

In the cell, membraneless organelles allow the maintenance of concentration differences without the need for energy input. The chemical potential in both phases is equal, i.e. there is no net diffusive flux between the dense and light phase, but there is constant flux of individual molecules between them²².

The principles of LLPS have been widely studied for polymers in soft matter physics^{21,61}. LLPS is the process in which a polymer that is homogeneously distributed within a solvent spontaneously demixes into two coexisting liquid phases, namely a polymer-enriched phase (dense phase) and a polymer-depleted phase (light phase) (**Fig. 1.3.2a**). A negative free energy of demixing is achieved above a certain concentration (saturation concentration or C_{sat}) with a gain in enthalpy (when polymer-polymer contacts are preferred over polymer-solvent contacts)⁶¹ and/or in entropy (by releasing solvent molecules from the hydration shell into the light phase)⁶² (**Fig. 1.3.2b and c**). LLPS is extremely sensitive to changes in physical-chemical conditions of the environment⁶³. The C_{sat} can be affected by changes in temperature, ionic strength, pH, post-translational modifications, etc.⁶⁴.

The physics behind the LLPS of polymers can be explained by the Flory-Huggins theory. However, this theory does not account for several characteristics that biopolymers have with respect to polymers (like sequence effects, the fact that they are not at thermodynamic equilibrium, etc.)⁶⁵. For this reason, other mathematical models are needed in order to fully explain the behaviour of phase separating biopolymers⁶⁶.

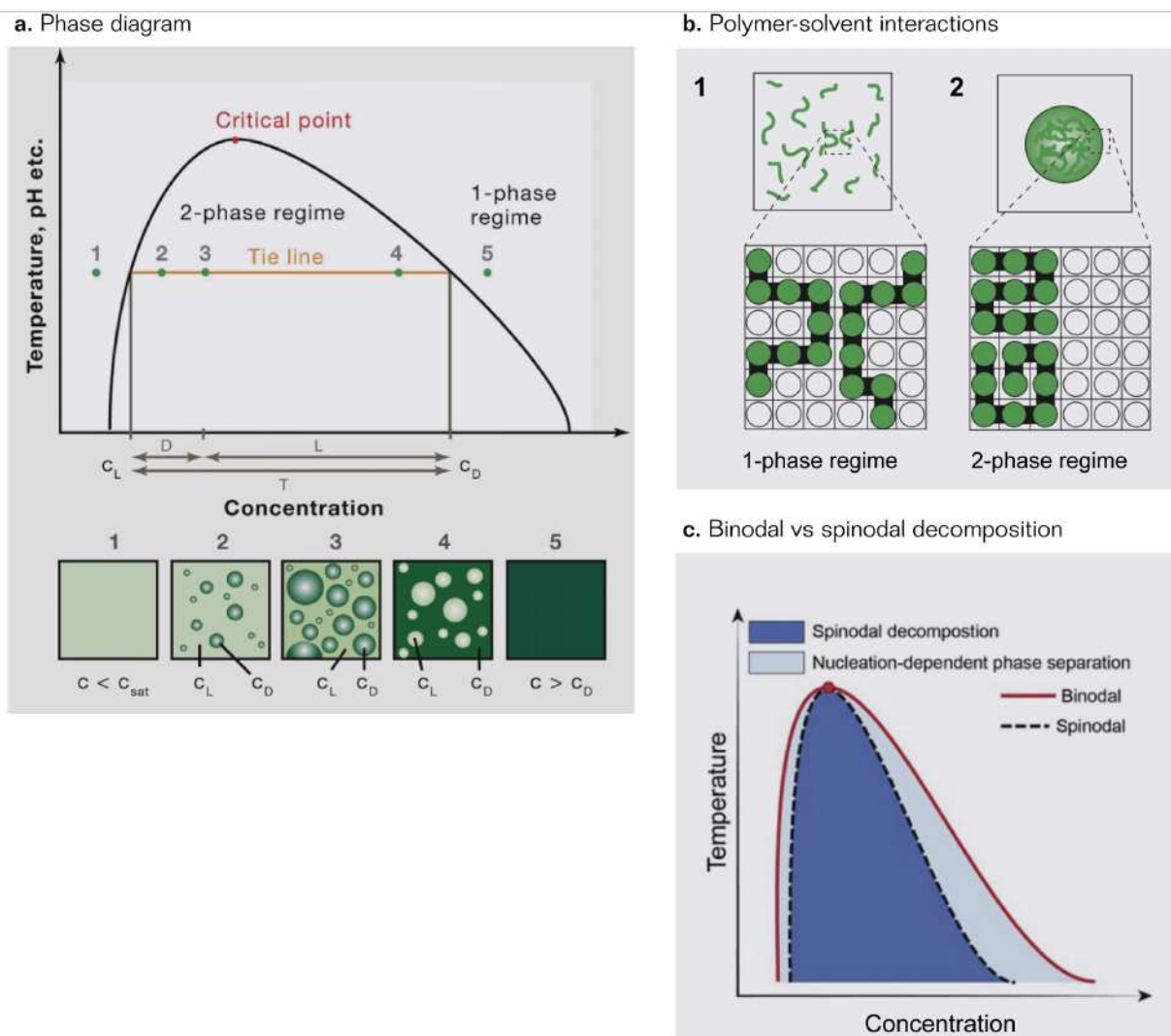


Fig. 1.3.2. a) The phase diagram is the representation of the one-phase and two-phase regimes depending on environmental conditions such as temperature, pH, ionic strength, etc. (black line). (1) At concentrations below C_{sat} , the system is in the one-phase regime. (2) At any condition within the two-phase regime, the system demixes into a light phase (with $c=c_L$) and a dense phase (with $c=c_D$). (2-4) At increasing concentrations, only the volume fractions of the two phases change relatively to each other but c_L and c_D remain constant. Adapted from Alberti et al. 2019⁶⁷. **b)** (1) Below the C_{sat} the polymer is homogeneously distributed in the solvent. (2) Above the C_{sat} the system demixes into two phases and polymer-polymer interactions are stronger. Adapted from Brangwynne et al. 2015⁶¹. **c)** There is a region of instability at the interface between the one-phase and two-phase regimes in which the system demixes when nucleated⁶⁸. At the interface, a metastable well-mixed state, albeit supersaturated, can persist even when phase separation is thermodynamically favourable. In this scenario, nucleation accelerates and determines whether phase separation takes place. Adapted from Posey et al. 2018⁶⁹.

Multivalency is the key feature for LLPS²³. Multivalent interactions (scaffold-scaffold and scaffold-client) allow the formation of networks, the basis for LLPS. Proteins can undergo LLPS when they contain several interaction modules with self-association affinity or affinity for a particular ligand⁶⁰. These interaction motifs can be structured domains connected by a disordered linker binding to a ligand⁷⁰ or motifs (e.g. “stickers”) embedded in disordered sequences (“e.g. spacers”)^{71,72}.

1. Introduction and objectives

IDPs play a central role in LLPS as by definition they are multivalent proteins that can simultaneously interact with other molecules or with themselves⁷³. Other biomolecules that can undergo LLPS are RNA and DNA, which contain several interaction motifs with other nucleic acid molecules or proteins⁶⁰. Given the scope of this thesis, we will focus on IDPs as biopolymers from now on.

Since 2009, many other studies characterising condensation have been published. Biomolecular condensates are usually heterogeneous in composition. For this reason, it is useful to apply the scaffold and client concept⁷⁴. Scaffold molecules are those that drive phase separation and clients are those that partition into pre-formed condensates.

Biomolecular condensates participate in many cellular functions⁷⁵ and need to be tightly controlled⁶⁸. Dysregulation in the assembly and disassembly of biomolecular condensates appears to be linked to various diseases including neurodegeneration^{76,77}, cancer⁷⁸ and infectious diseases^{79,80} although the exact pathomechanisms are still unclear (e.g. whether aberrant condensate formation and dissolution is the cause of the disease or its phenotype)^{80,81}.

Biomolecular condensates offer a new layer of regulation for IDPs. Therefore, modulating condensates with drugs is a potential opportunity to target currently undruggable proteins^{81,82} and contribute to drug pharmacodynamics⁸³. However, the druggability of biomolecular condensates is still an underexplored area of research.

Second objective: Study the role of EPI-001 as a modulator of androgen receptor condensation

The second objective was to examine whether the mechanism of action of EPI-001 is related to its ability to modulate AR condensation. In this regard, we studied the effects of reversible and irreversible binding of EPI-001 on AR condensation *in vitro* and in cells. This work is one of the first studies of how small molecules can inhibit a disordered target by modulating its condensate formation.

The work towards this objective is described in **Chapter 3** of this doctoral thesis.

The first two objectives of the thesis were related to the mechanistic role of EPI-001 inhibition and its connection with biomolecular condensation.

Third objective: Design, synthesis and evaluation of novel EPI-001 analogues

The third objective was to apply the acquired knowledge to design new and improved AR inhibitors. To the best of our knowledge, this work is one of the first examples of rational drug design for IDPs based on the modulation of biomolecular condensates.

The work towards this objective is described in **Chapter 4** of this doctoral thesis.

Bibliography

- (1) Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011**, *162* (6), 1239–1249.
- (2) Anderson, A. C. The Process of Structure-Based Drug Design. *Chem. Biol.* **2003**, *10* (9), 787–797.
- (3) Wang, X.; Song, K.; Li, L.; Chen, L. Structure-Based Drug Design Strategies and Challenges. *Curr. Top. Med. Chem.* **2018**, *18*, 998–1006.
- (4) Batool, M.; Ahmad, B.; Choi, S. A Structure-Based Drug Discovery Paradigm. *Int. J. Mol. Sci.* **2019**, *20* (2783), 1–18.
- (5) Uversky, V. N. Intrinsically Disordered Proteins and Novel Strategies for Drug Discovery. *Expert Opin. Drug Discov.* **2012**, *7*, 475–488.
- (6) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **2004**, *337* (3), 635–645.
- (7) Deiana, A.; Forcelloni, S.; Porrello, A.; Giansanti, A. New Classification of Intrinsic Disorder in the Human Proteome. *bioRxiv*, **2018**. Doi: <https://doi.org/10.1101/446351> [Preprint].
- (8) Wright, P. E.; Dyson, H. J. Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* **1999**, *293* (2), 321–331.
- (9) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why Are “Natively Unfolded” Proteins Unstructured Under Physiologic Conditions? *PROTEINS: Structure, Function, and Genetics* **2000**, *41*, 415–427.
- (10) Tompa, P. Intrinsically Unstructured Proteins. *Trends Biochem. Sci.* **2002**, *27* (10), 527–533.
- (11) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (3), 197–208.
- (12) Uversky, V. N.; Dunker, A. K. Understanding Protein Non-Folding. *Biochim. Biophys. Acta* **2010**, *1804* (6), 1231–1264.
- (13) Dunker, A. K.; Bondos, S. E.; Huang, F.; Oldfield, C. J. Intrinsically Disordered Proteins and Multicellular Organisms. *Semin. Cell Dev. Biol.* **2015**, *37*, 44–55.
- (14) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annu. Rev. Biophys.* **2008**, *37*, 215–246.
- (15) 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.; Obradovic, Z. Intrinsically Disordered Protein. *J. Mol. Graph. Model.* **2001**, *19* (1), 26–59.
- (16) Flock, T.; Weatheritt, R. J.; Latysheva, N. S.; Babu, M. M. Controlling Entropy to Tune the Functions of Intrinsically Disordered Regions. *Curr. Opin. Struct. Biol.* **2014**, *26*, 62–72.
- (17) Uversky, V. N. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Front. Phys.* **2019**, *7*, 1–18.
- (18) Tanford, C. The Hydrophobic Effect and the Organization of Living Matter. *Science* **1978**, *200* (4345), 1012–1018.
- (19) Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.; Obradovic, Z. Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions. *J. Proteome Res.* **2007**, *6* (5), 1882–1898.
- (20) Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. Intrinsic Disorder in Transcription Factors. *Biochemistry* **2006**, *45* (22), 6873–6888.
- (21) Hyman, A. A.; Weber, C. A.; Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **2014**, *30*, 39–58.

- (22) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular Condensates: Organizers of Cellular Biochemistry. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (5), 285–298.
- (23) Shin, Y.; Brangwynne, C. P. Liquid Phase Condensation in Cell Physiology and Disease. *Science* **2017**, *357* (6357), 1–11.
- (24) Dyson, H. J.; Wright, P. E. Coupling of Folding and Binding for Unstructured Proteins. *Curr. Opin. Struct. Biol.* **2002**, *12* (1), 54–60.
- (25) Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Kragelund, B. B.; Best, R. B.; Schuler, B. Extreme Disorder in an Ultrahigh-Affinity Protein Complex. *Nature* **2018**, *555* (7694), 61–66.
- (26) Uversky, V. N. Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding. *Protein J.* **2009**, *28*, 305–325.
- (27) Theillet, F.-X.; Binol, A.; Frembgen-kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J.; Elcock, A. H.; Gershenson, A.; Selenko, P.; Carolina, N.; Hill, C.; States, U. Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6661–6714.
- (28) van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114*, 6589–6631.
- (29) Weng, J.; Wang, W. Dynamic Multivalent Interactions of Intrinsically Disordered Proteins. *Curr. Opin. Struct. Biol.* **2019**, *62*, 9–13.
- (30) Tompa, P.; Fuxreiter, M. Fuzzy Complexes: Polymorphism and Structural Disorder in Protein-Protein Interactions. *Trends Biochem. Sci.* **2008**, *33* (1), 2–8.
- (31) Wright, P. E.; Dyson, H. J. Linking Folding and Binding. *Curr. Opin. Struct. Biol.* **2009**, *19* (1), 31–38.
- (32) Cumberworth, A.; Lamour, G.; Babu, M. M.; Gsponer, J. Promiscuity as a Functional Trait: Intrinsically Disordered Regions as Central Players of Interactomes. *Biochem. J.* **2013**, *454* (3), 361–369.
- (33) Bah, A.; Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-Translational Modifications. *J. Biol. Chem.* **2016**, *291* (13), 6696–6705.
- (34) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. Intrinsically Disordered Proteins: Regulation and Disease. *Curr. Opin. Struct. Biol.* **2011**, *21* (3), 432–440.
- (35) Gsponer, J.; Futschik, M. E.; Teichmann, S. A.; Babu, M. M. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science* **2008**, *322* (5906), 1365–1368.
- (36) van der Lee, R.; Lang, B.; Kruse, K.; Gsponer, J.; Sánchez de Groot, N.; Huynen, M. A.; Matouschek, A.; Fuxreiter, M.; Babu, M. M. Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution. *Cell Rep.* **2014**, *8* (6), 1832–1844.
- (37) Ruan, H.; Sun, Q.; Zhang, W.; Liu, Y.; Lai, L. Targeting Intrinsically Disordered Proteins at the Edge of Chaos. *Drug Discov. Today* **2019**, *24* (1), 217–227.
- (38) Metallo, S. J. Intrinsically Disordered Proteins Are Potential Drug Targets. *Curr. Opin. Chem. Biol.* **2010**, *14* (4), 481–488.
- (39) Tan, M. H. E.; Li, J.; Xu, H. E.; Melcher, K.; Yong, E.-L. Androgen Receptor: Structure, Role in Prostate Cancer and Drug Discovery. *Acta Pharmacol. Sin.* **2015**, *36* (1), 3–23.
- (40) McEwan, I. J. *Nuclear Receptors: One Big Family in The Nuclear Receptor Superfamily*; McEwan, I. J., Ed.; Humana press, 2009; Vol. 505, pp 3–18.
- (41) Claessens, F.; Denayer, S.; Van Tilborgh, N.; Kerkhofs, S.; Helsen, C.; Haelens, A. Diverse Roles of Androgen Receptor (AR) Domains in AR-Mediated Signaling. *Nucl. Recept. Signal.* **2008**, *6*, 1–13.

1. Introduction and objectives

- (42) Jenster, G. Identification of Two Transcription Activation Units in the N-Terminal Domain of the Human Androgen Receptor. *J. Biol. Chem.* **1995**, *270* (13), 7341–7346.
- (43) Centenera, M. M.; Harris, J. M.; Tilley, W. D.; Butler, L. M. Minireview: The Contribution of Different Androgen Receptor Domains to Receptor Dimerization and Signaling. *Mol. Endocrinol.* **2008**, *22* (11), 2373–2382.
- (44) van Royen, M. E.; van Cappellen, W. A.; de Vos, C.; Houtsmuller, A. B.; Trapman, J. Stepwise Androgen Receptor Dimerization. *J. Cell Sci.* **2012**, *125*, 1970–1979.
- (45) 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.; McEwan, I. J.; Riera, A.; Salvatella, X. EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. *ACS Chem. Biol.* **2016**, *11*, 2499–2505.
- (46) Culp, M. B.; Soerjomataram, I.; Efstathiou, J. A.; Bray, F.; Jemal, A. Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. *Eur. Urol.* **2020**, *77* (1), 38–52.
- (47) Ferlay, J.; Colombet, M.; Soerjomataram, I.; Dyba, T.; Randi, G.; Bettio, M.; Gavin, A.; Visser, O.; Bray, F. Cancer Incidence and Mortality Patterns in Europe: Estimates for 40 Countries and 25 Major Cancers in 2018. *Eur. J. Cancer* **2018**, *103*, 356–387.
- (48) National Cancer Institute. Prostate-Specific Antigen (PSA) Test <https://www.cancer.gov/types/prostate/psa-fact-sheet> (accessed Mar 2020).
- (49) Taylor, B. S.; Schultz, N.; Hieronymus, H.; Gopalan, A.; Xiao, Y.; Carver, B. S.; Arora, V. K.; Kaushik, P.; Cerami, E.; Reva, B.; Antipin, Y.; Mitsiades, N.; Landers, T.; Dolgalev, I.; Major, J. E.; Wilson, M.; Socci, N. D.; Lash, A. E.; Heguy, A.; Eastham, J. A.; Scher, H. I.; Reuter, V. E.; Scardino, P. T.; Sander, C.; Sawyers, C. L.; Gerald, W. L. Integrative Genomic Profiling of Human Prostate Cancer. *Cancer Cell* **2010**, *18* (1), 11–22.
- (50) National Cancer Institute. Hormone Therapy for Prostate Cancer <https://www.cancer.gov/types/prostate/prostate-hormone-therapy-fact-sheet#how-is-hormone-therapy-used-to-treat-prostate-cancer> (accessed Mar 2020).
- (51) Chan, S. C.; Dehm, S. M. Constitutive Activity of the Androgen Receptor. *Adv. Pharmacol.* **2014**, *70*, 327–366.
- (52) Chandrasekar, T.; Yang, J. C.; Gao, A. C.; Evans, C. P. Mechanisms of Resistance in Castration-Resistant Prostate Cancer (CRPC). *Transl. Androl. Urol.* **2015**, *4* (3), 365–380.
- (53) 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.; Beltran, H.; Abida, W.; Bradley, R. K.; Vinson, J.; Cao, X.; Vats, P.; Kunju, L. P.; Hussain, M.; Feng, F. Y.; Tomlins, S. A.; Cooney, K. A.; Smith, D. C.; Brennan, C.; Siddiqui, J.; Mehra, R.; Chen, Y.; Rathkopf, D. E.; Morris, M. J.; Solomon, S. B.; Durack, J. C.; Reuter, V. E.; Gopalan, A.; Gao, J.; Loda, M.; Lis, R. T.; Bowden, M.; Balk, S. P.; Gaviola, G.; Sougnez, C.; Gupta, M.; Yu, E. Y.; Mostaghel, E. A.; Cheng, H. H.; Mulcahy, H.; True, L. D.; Plymate, S. R.; Dvigne, H.; Ferraldeschi, R.; Flohr, P.; Miranda, S.; Zafeiriou, Z.; Tunariu, N.; Mateo, J.; Perez-Lopez, R.; Demichelis, F.; Robinson, B. D.; Schiffman, M.; Nanus, D. M.; Tagawa, S. T.; Sigaras, A.; Eng, K. W.; Elemento, O.; Sboner, A.; Heath, E. I.; Scher, H. I.; Pienta, K. J.; Kantoff, P.; de Bono, J. S.; Rubin, M. A.; Nelson, P. S.; Garraway, L. A.; Sawyers, C. L.; Chinnaiyan, A. M. Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell* **2015**, *161* (5), 1215–1228.
- (54) Sadar, M. D. Discovery of Drugs That Directly Target the Intrinsically Disordered Region of the Androgen Receptor. *Expert Opin. Drug Discov.* **2020**, *15* (5), 551–560.
- (55) 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.; Sadar, M. D. Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor. *Cancer Cell* **2010**, *17* (6), 535–546.
- (56) 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.; Sadar, M. D. An Androgen Receptor N-Terminal Domain Antagonist for Treating Prostate Cancer. *J.*

Clin. Invest. **2013**, *123* (7), 2948–2960.

- (57) Szulc, E. M. Structural Insights into “Acid Blobs and Negative Noodles” - the Androgen Receptor as a Case Study. Ph.D. Thesis, Universitat de Barcelona - IRB Barcelona, 2019.
- (58) McGuffee, S. R.; Elcock, A. H. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. *PLoS Comput. Biol.* **2010**, *6* (3), 1–18.
- (59) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/condensation. *Science* **2009**, *324* (5935), 1729–1732.
- (60) Alberti, S. Phase Separation in Biology. *Curr. Biol.* **2017**, *27*, 1097–1101.
- (61) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer Physics of Intracellular Phase Transitions. *Nat. Phys.* **2015**, *11*, 899.
- (62) Martin, E. W.; Mittag, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* **2018**, *57* (17), 2478–2487.
- (63) Alberti, S.; Hyman, A. A. Are Aberrant Phase Transitions a Driver of Cellular Aging? *Bioessays* **2016**, *38* (10), 959–968.
- (64) Uversky, V. N. Intrinsically Disordered Proteins in Overcrowded Milieu: Membrane-Less Organelles, Phase Separation, and Intrinsic Disorder. *Curr. Opin. Struct. Biol.* **2017**, *44*, 18–30.
- (65) Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S. Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates. *Biochemistry* **2018**, *57* (17), 2499–2508.
- (66) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-like Domains. *Science* **2020**, *367* (6478), 694–699.
- (67) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176* (3), 419–434.
- (68) Wheeler, R. J.; Hyman, A. A. Controlling Compartmentalization by Non-Membrane-Bound Organelles. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2018**, *373* (1747), 1–9.
- (69) Posey, A. E.; Holehouse, A. S.; Pappu, R. V. Phase Separation of Intrinsically Disordered Proteins. *Methods Enzymol.* **2018**, *611*, 1–30.
- (70) Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q.-X.; Nixon, B. T.; Rosen, M. K. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* **2012**, *483* (7389), 336–340.
- (71) Choi, J.-M.; Dar, F.; Pappu, R. V. LASSI: A Lattice Model for Simulating Phase Transitions of Multivalent Proteins. *PLoS Comput. Biol.* **2019**, *15* (10), 1–39.
- (72) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaître, R.; Pozniakovskiy, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 1–12.
- (73) Fung, H. Y. J.; Birol, M.; Rhoades, E. IDPs in Macromolecular Complexes: The Roles of Multivalent Interactions in Diverse Assemblies. *Curr. Opin. Struct. Biol.* **2018**, *49*, 36–43.
- (74) Banani, S. F.; Rice, A. M.; Peeples, W. B.; Lin, Y.; Jain, S.; Parker, R.; Rosen, M. K. Compositional Control of Phase-Separated Cellular Bodies. *Cell* **2016**, *166* (3), 651–663.
- (75) Alberti, S. The Wisdom of Crowds: Regulating Cell Function through Condensed States of Living Matter. *J. Cell Sci.* **2017**, *130* (17), 2789–2796.
- (76) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoyanov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M.; Pozniakovski, A.; Poser, I.; Maghelli, N.; Royer, L. A.; Weigert, M.; Myers, E. W.; Grill, S.; Drechsel, D.; Hyman, A. A.; Alberti, S. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162* (5), 1066–1077.

1. Introduction and objectives

- (77) Nedelsky, N. B.; Taylor, J. P. Bridging Biophysics and Neurology: Aberrant Phase Transitions in Neurodegenerative Disease. *Nat. Rev. Neurol.* **2019**, *15* (5), 272–286.
- (78) Bouchard, J. J.; Otero, J. H.; Scott, D. C.; Szulc, E.; Martin, E. W.; Sabri, N.; Granata, D.; Marzahn, M. R.; Lindorff-Larsen, K.; Salvatella, X.; Schulman, B. A.; Mittag, T. Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. *Mol. Cell* **2018**, *72* (1), 19–36.
- (79) Heinrich, B. S.; Maliga, Z.; Stein, D. A.; Hyman, A. A.; Whelan, S. P. J. Phase Transitions Drive the Formation of Vesicular Stomatitis Virus Replication Compartments. *MBio* **2018**, *9* (5), 1–10.
- (80) Alberti, S.; Dormann, D. Liquid–Liquid Phase Separation in Disease. *Annu. Rev. Genet.* **2019**, *53*, 171–194.
- (81) Wheeler, R. J. Therapeutics - How to Treat Phase Separation-Associated Diseases. *Emerg. Top. Life Sci.* **2020**, 1–12. Doi: <https://doi.org/10.1042/ETLS20190176>.
- (82) Mullard, A. Biomolecular Condensates Pique Drug Discovery Curiosity. *Nat. Rev. Drug Discov.* **2019**, *18*, 324–326.
- (83) Klein, I. A.; Boija, A.; Afeyan, L. K.; Hawken, S. W.; Fan, M.; Dall’Agnese, A.; Oksuz, O.; Henninger, J. E.; Shrinivas, K.; Sabari, B. R.; Sagi, I.; Clark, V. E.; Platt, J. M.; Kar, M.; McCall, P. M.; Zamudio, A. V.; Manteiga, J. C.; Coffey, E. L.; Li, C. H.; Hannett, N. M.; Guo, Y. E.; Decker, T.-M.; Lee, T. I.; Zhang, T.; Weng, J.-K.; Taatjes, D. J.; Chakraborty, A.; Sharp, P. A.; Chang, Y. T.; Hyman, A. A.; Gray, N. S.; Young, R. A. Partitioning of Cancer Therapeutics in Nuclear Condensates. *Science* **2020**, *368* (6497), 1386–1392.

2. REVERSIBLE INTERACTION OF EPI-001 WITH ITS TARGET BY SOLUTION NMR SPECTROSCOPY

2.1. Background

2.1.1. Intrinsically disordered proteins as drug targets

As mentioned in section 1.1, intrinsically disordered proteins (IDPs) are potentially important drug targets given the high prevalence of intrinsic disorder (ID) in the eukaryotic proteome and their association with pathology. However, their highly dynamic nature presents an inherent challenge for traditional drug discovery methods based on structure-based rational design¹. Nevertheless, a growing number of small molecule ligands being able to fine-tune IDPs function can be found in the literature. We have grouped them on the basis of their mechanism of inhibition into the following: 1) regulatory inhibition, 2) orthosteric inhibition, and 3) allosteric inhibition (**Fig. 2.1.1.1**).

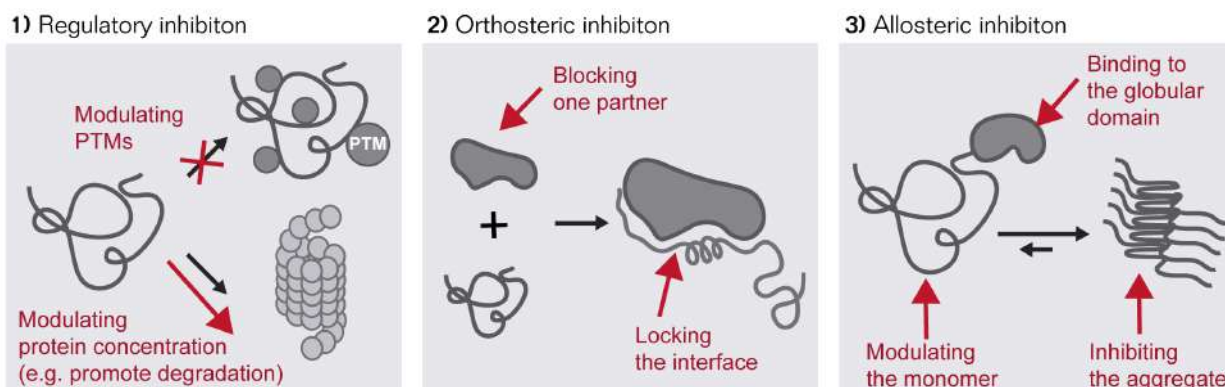


Fig. 2.1.1.1. Strategies for inhibiting IDPs: 1) regulatory inhibition, 2) orthosteric inhibition and 3) allosteric inhibition.

Strategies 1 and 2 target the machinery that regulates IDPs and the IDP binding partners, respectively. Structure-based approaches can be applied to strategies 1 and 2. The main focus of the thesis will be on strategy 3: targeting the IDP itself.

1. Regulatory inhibition

IDPs are tightly regulated by post-transcriptional modifications (PTMs)^{2,3} and through their protein concentrations⁴. Therefore, interfering with the regulatory machinery emerges as a successful strategy to inhibit an IDP.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

1.a. Modulating the protein post-transcriptional modifications

IDPs are modulated through a myriad of PTMs³. Hence inhibiting the enzymes responsible for PTMs is a plausible strategy to inhibit IDPs.

For example, the Tau protein is abnormally hyperphosphorylated in Alzheimer's disease⁵. Inhibition of the kinases responsible for Tau hyperphosphorylation, like the glycogen synthase kinase 3 β (GSK3 β), offers a new avenue for the treatment of tauopathies⁶. GSK3 β is a validated target for Alzheimer's disease and several potent selective inhibitors have been described. In particular, SAR502250 is able to reduce Tau aggregation, as well as alleviating the neuropsychiatric symptoms associated with the disease⁷.

1.b. Modulating the protein concentration

Chemical knockdown of a protein using target protein degraders, such as proteolysis targeting chimeras (PROTACs), is a recurrent strategy for drug discovery and an alternative to protein inhibition⁸. A PROTAC molecule is a small heterobifunctional molecule formed by tethering an E3 ligase ligand to a target protein binder through a linker. By bringing together the protein of interest and the E3 ligase, the target protein is polyubiquitinated and subsequently degraded by the proteasome⁹. PROTAC-like molecules targeting an IDP can fine-tune its concentration in cells.

As an example, Lu et al. designed a PROTAC molecule against Tau, inducing its proteasomal degradation¹⁰. Further studies must be carried out to evaluate the potential use of this molecule in the clinical setting.

Degradation of the protein target can also be accomplished through methods other than PROTACs. For example, Narayanan and co-workers developed the first androgen receptor (AR) degraders for the treatment of prostate cancer¹¹. Although the general structure of this molecule, named UT-155, does not resemble that of a PROTAC, it is able to downregulate AR levels.

2. Orthosteric inhibition

IDPs are central in protein-protein interaction (PPI) networks. PPIs between an IDP and its partner are the target of orthosteric inhibitors. Orthosteric inhibitors can be directed to:

2.a. One of the partners

Inhibitors can be designed to mimic the binding region of the IDP to the structured partner, outcompeting it¹².

A good example is p53-MDM2 inhibition. p53 is a tumour suppressor IDP that is inactivated by MDM2 binding. Several peptides¹³ and small molecules¹⁴, can break this interaction by mimicking the residues in p53 that bind to MDM2, thereby increasing the tumour suppression activity of p53.

-

Nevertheless, the partner does not need to be globular. This is the case of c-Myc, an oncogene that dimerises with its partner Max to bind to DNA through a basic helix-loop-helix zipper domain (bHLHZip). Soucek et al. designed a c-Myc bHLHZip mimic, named Omomyc, that can heterodimerise with wild type c-Myc, thus making it unavailable to interact with its partner Max¹⁵ and enhancing Myc-induced apoptosis as a cancer treatment^{16,17}.

In order to block a PPI, inhibitors can also be designed to bind to the monomeric IDP partner. For example, Erkizan et al. used surface plasmon resonance to screen compounds that bind EWS-FLI1¹⁸. EWS-FLI1 is an oncogenic fusion construct present in Ewing's sarcoma family of tumours. EWS-FLI1 binds to RNA helicase A (RHA) to exert its oncogenic function. Their lead compound, YK-4-279, blocks RHA binding to EWS-FLI1 and reduces the growth of the tumours in mouse models of the disease.

2.b. Blocking the PPI interface

Alternatively, the target can be the PPI interface. Compounds targeting the interface offer high selectivity as the pocket is formed only when both partners interact. However, it might be a challenge to define the interface in a PPI where one partner is disordered unless it gains structure upon binding.

For example, as depicted in section 2.a, p53 can be activated by blocking its binding to MDM2 with drugs. However, when MDM2 is inhibited, there is overexpression of MDMX. MDMX can also bind p53 and compensate for the inhibitory action of the aforementioned drugs. Graves et al. described a new class of small molecules able to drive homo- and/or heterodimerization of MDM2 and MDMX proteins¹⁹. The new dimeric structures with a small molecule core can no longer bind to p53 and the suppressor activity of p53 is restored.

Another example is the stabilisation of the cell-cycle phosphatase cell division cycle 25 C (Cdc25C) with a 14-3-3 adaptamer protein. Cdc25C undergoes a disorder-to-order transition when bound to 14-3-3. Ottmann and co-workers characterised a supramolecular ligand, named CLR01) able to stabilise this PPI²⁰. On the one hand, CLR01 binds to the complex interface, filling a gap between both proteins. On the other, it blocks the IDP in the favourable conformation for binding, hence reducing the entropic cost of binding.

3. Allosteric inhibition

The third strategy to inhibit an IDP is to bind to it and prevent its function. However, as mentioned before, directly targeting IDPs is a challenging process. To avoid binding to the IDP, many drugs target the globular part of a protein with ID (section 3.a). Another possibility is to target the aggregated form of the IDP in misfolding diseases, for which the structure may be elucidated (section 3.b). Section 3.c presents an underexplored approach that is the most relevant to this thesis, namely inhibiting the monomeric form of an IDP.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

3.a. The globular part of a protein with ID

Targeting the globular domain of a protein with ID is a valid strategy to inhibit an IDP while being able to apply structured-based drug discovery techniques.

This has been the strategy of choice for targeting many transcription factors¹² such as the androgen receptor (AR)²¹. All AR inhibitors used in the clinic, namely antiandrogens, inhibit the structured AR hormone-binding domain by blocking the binding of the hormone to its pocket.

3.b. The aggregated IDP

IDPs are involved in numerous misfolding and aggregation diseases like Parkinson's (linked to α -synuclein aggregation), Alzheimer's (linked to Tau and A β) and Huntington's disease (linked to Huntingtin)²². Aggregation is a quasi-irreversible process in which soluble monomeric protein - usually disordered or unfolded- can self-associate into oligomers, which serve as nucleation points that lead to the final most thermodynamically stable state: the fibril or aggregate²³.

Treatments for these misfolding diseases can be guided to i) the monomer - although challenging as they are IDPs; ii) the aggregate - for which structures can be obtained to guide drug discovery and iii) the oligomeric intermediates - for which there is a growing evidence that they are the neurotoxic species²⁴. Various groups have attempted to modulate the aggregation process at different stages.

In the first example, Linse and co-workers targeted the secondary nucleation of A β ²⁵. This second nucleation takes place on the already formed fibrils and is highly effective in the production of toxic oligomers. To inhibit nucleation, they generated antibodies that slowed the aggregation kinetics of A β 42. Moreover, the fibrils formed had less tendency to grow when compared to the non-treated condition.

In the second example Zweckstetter and co-workers deviated the aggregation course of α -synuclein into non-toxic off-pathway oligomers²⁶. They described a small molecule, phthalocyanine tetrasulfonate, that polymerised into a superstructure that was able to stabilise α -synuclein helical oligomers, thereby preventing them from progressing into fibrils.

Tóth et al. followed a different strategy to target the same protein: α -synuclein²⁷. First, they characterised the monomeric α -synuclein ensemble from NMR data. They then used this ensemble to guide a computational screening, which yielded a compound, ELN484228, able to reverse the toxic effects of α -synuclein.

In the last example, Eisenberg and co-workers targeted the addition of monomers into the fibril²⁸. For this, they studied the key residues involved in Tau fibril formation by X-ray and micro-electron diffraction. They used this structure as a template to design peptides that disrupted the fibril interface.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

3.c. The monomeric form

To inhibit the monomeric form of an IDP, the small molecule ligand needs to bind to it. But how do small molecules bind IDPs?

Spontaneous binding between two molecules (e.g. a small molecule drug and a protein) occurs only if it is thermodynamically favourable (the overall Gibbs free energy of binding is negative, $\Delta G = G_{\text{final}} - G_{\text{initial}} < 0$)²⁹. ΔG is the sum of enthalpy (H) and entropy (S) contributions (Equation 1):

$$\text{Equation 1: } \Delta G = \Delta H - T\Delta S$$

ΔH is determined by the interatomic forces upon binding (gain or loss of molecular interactions like hydrogen bonds, electrostatic or van der Waals interactions). ΔS is the change in the number of different configurations that a system can have (change in the size of the conformational space of the protein, ligand and solvent molecules). ΔH and ΔS can contribute to the overall ΔG in four ways, depicted in Figure 3.1.1.2 (**Fig. 2.1.1.2**).

	$\Delta H < 0$	$\Delta H > 0$
$\Delta S < 0$	ii Enthalpically driven binding	i No binding
$\Delta S > 0$	iii Entropically and enthalpically favoured binding	iv Entropically driven binding

Fig. 2.1.1.2. Possible modes in which enthalpy and entropy contribute to the overall ΔG : (i) $\Delta H > 0$, $\Delta S < 0$; (ii) $\Delta H < 0$, $\Delta S < 0$; (iii) $\Delta H < 0$, $\Delta S > 0$; and (iv) $\Delta H > 0$, $\Delta S > 0$. Only modes yielding negative ΔG values (ii–iv) lead to binding. Adapted from Heller et al. 2015²⁹.

The binding of ligands to structured proteins is driven by the enthalpic contribution (modes ii and iii), which may come at an entropic cost (mode iii)²⁹. Structure-based drug discovery (SBDD) improves ligand affinity by increasing the number and strength of the interactions between the ligand and the target, compensating for possible entropic costs³⁰.

In the case of IDPs, the entropic cost of restricting the conformational ensemble to that for binding is very large. It can be compensated only when large interaction surfaces are present. Large surfaces allow a positive enthalpic contribution due to multiple interactions between the two molecules and/or a positive solvation entropy caused by the release of water molecules from the binding interface. A high entropic cost can be perceived negatively but it is important to highlight that the modulation of the entropic cost allows IDPs to change their energy landscapes, thus modulating their function. This can be achieved by intrinsic (e.g. PTMs) or extrinsic (e.g. drugs) factors³⁰.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

We now focus on how drugs can effectively modulate the energy landscape of IDPs to achieve inhibition. As depicted in Figure 2.1.1.3, the binding of a small molecule can cause a restriction of the energy landscape (disorder-to-order, scenario A), a minimal shift in the conformational ensemble (disorder-to-disorder or “fuzzy”, scenario B) or a re-weighting of the sampled populations (disorder-to-more-disorder, scenario C) (Fig. 2.1.1.3)^{29–31}.

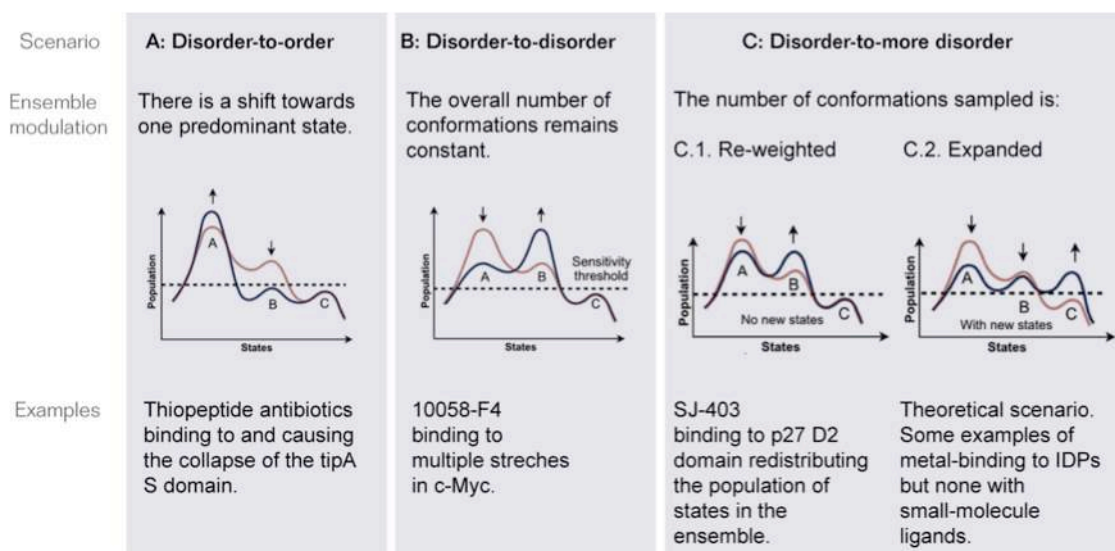


Fig 2.1.1.3. Consequences of small molecule binding to the energy landscape of an IDP. Examples illustrating each category are presented. Adapted from Flock et al. 2014; Heller et al. 2015 and Heller et al. 2018^{29–31}.

Examples of small molecules binding to an IDP in each category include:

3.c.1. Scenario A: Disorder-to-order

In the first scenario, the small molecule ligand causes a conformational shift towards one predominant state, resembling the disorder-to-order transitions of IDPs when binding to a structured partner³². As in the binding to a globular partner, the entropic cost of restricting the conformational ensemble must be compensated with an enthalpic gain from the drug-protein interactions and/or an entropic gain due to the release of water molecules from the binding interface²⁹.

For example, a bacterial transcriptional regulator, thiostrepton-induced protein A or tipA, causes multidrug resistance by binding to thiopeptide antibiotic drugs, preventing them from interacting with their expected target³³. Drug binding induces a large conformational rearrangement in tipA S domain, from a partially unfolded to a globular structure. This rearrangement upregulates the tipA gene, resulting in increased resistance.

3.c.2. Scenario B: Disorder-to-disorder

In the second scenario, the small molecule ligand binds to the IDP in a “fuzzy” way, either inhibiting the interaction with other partners or promoting “inactive” states. The conformational restrictions caused by binding are compensated by various mechanisms to yield an equivalent

number of conformational configurations in the apo and in the holo states. These mechanisms include³⁰:

- Structuration of only very short motifs (e.g. the molecular recognition features, MoRFs) which have a low impact on the entropy of the system.
- Binding to multiple sites. Multivalent-interactions allow for binding to multiple sites (e.g. multiple MoRFs) while the overall structure of the linkers is still disordered. Also, multivalent interactions increase the total number of configurations due to degenerate binding (“fuzzy”).
- Allosteric compensation. The binding of the ligand to one end causes, on another end, either the exposure of a binding site or a more flexible conformation.

A first example is the inhibition of c-Myc. As mentioned, c-Myc is an IDP that forms a helical dimer with its partner, Max. In 2003, Yin et al. discovered several small molecules able to disrupt the c-Myc-Max interaction using yeast-to-hybrid technology³⁴. Some years later, Metallo and co-workers extensively studied the interaction of these molecules with c-Myc using a variety of biophysical methods^{35,36}. They found that several molecules bound to different regions of the dimer interface, all causing local conformational changes that sequestered c-Myc in the disordered monomeric form, thereby preventing dimer formation. Jin et al. conducted computational studies of the same system to describe the molecular interactions between the protein and the ligands³⁷. They found that the same ligand was able to bind to different sites on the protein. In other words, the ligand was behaving like a “cloud” around a protein with no defined structure, also a “cloud”. The protein-cloud ligand-cloud theory put forward by Jin et al. proposed that IDPs contain short stretches of amino acids that can bind small molecules. Hence, by predicting the stretches of amino acids that are prone to ligand-binding, inhibitors could be designed.

A second example is the discovery of a compound able to inhibit the NURP-1 transcription regulator. NURP-1 forms a complex with MSLT-1, maintaining its IDP character (in a “fuzzy” manner). Neira et al. also searched for small molecules able to interact with the target in a “fuzzy” manner³⁸. They screened a library of compounds against the nuclear protein 1 (NUPR-1) by thermal shift assay with the aim to identify ligands that stabilised a region in NUPR-1 with higher secondary structure propensity. Their most active compound disrupted NUPR-1 interaction with male-specific-lethal protein 1 (MSL-1) *in vitro* and stopped tumour progression *in vivo*.

3.c.3. Scenario C: Disorder-to-more disorder

In the third scenario, the small molecule ligand causes a shift in the population or number of conformations sampled by the IDP^{29,31}. This shift can be due to:

- *Re-weighting of the previous existing conformations (in the apo form)*

In this case, stabilisation of an “inactive” state decreases the population of “active” conformations.

-

An example is the binding of SF-403 to the D2 domain of p27^{Kip1} (p27-D2) ³⁹. Ban et al. demonstrated that the binding of the small molecule SF-403 to p27-D2 caused a population shift, stabilising one of the 5 conformations already present in the apo state. The conformational ensemble was determined by using a combination of NMR-relaxation dispersion, small-angle X-ray scattering (SAXS) and mutagenesis experiments.

Another example is the inhibition of the intrinsically disordered C-terminal domain of PTP1B by MSI-1436. Many attempts to inhibit the PTP1B structured active site have resulted in compounds with poor drug-like properties. Krishnan et al. searched for allosteric inhibitors of PTP1B that bound to the disordered C-terminal domain ⁴⁰. Their lead compound, MSI-1436 binds to a region of the C-terminal domain with high helical propensity. This binding causes the formation of a secondary binding site close to the catalytic domain where a second molecule binds. This two-step allosteric inhibition induces a global conformational change that renders the protein inactive.

- *Increase in the number of possible states*

In this case, the “active” conformations are less populated due to an increase in the overall number of possible states.

The entropic expansion model is a theoretical scenario put forward by Vendruscolo and co-workers ²⁹. There is still no example of a small molecule ligand with this mechanism of action. However, Heller et al. enumerated examples of metal binding to IDPs that can be explained only when considering the entropic expansion theory. These examples include the binding of Ca²⁺ to phospholipase D bC2 and aC2, and the binding of Zn²⁺ to conantokin-G and T.

Several mechanisms can be drawn to explain how small molecules fine-tune IDP function although how a ligand can bind to an IDP with specificity and selectivity is still unanswered ⁴¹. From the protein point of view, given the conformational plasticity of IDPs and their MoRFs, the same IDP can bind to various molecules with similar affinity. From the small molecule point of view, a molecule that binds to an ID region of a protein can also bind to other ID regions of the same nature. Nevertheless, this promiscuous binding can be an advantage when searching for initial binding compounds that can be optimised to improve their affinity and specificity. In fact, this is the case for c-myc/Max inhibitors. Initial screening identified two compounds capable of inhibiting c-Myc but that also inhibited c-Jun, another oncogenic transcription factor. Follow-up screenings and optimizations yielded a compound that inhibited c-Myc/Max dimerization but not c-Jun ²⁴.

Another question worth tackling is: do small molecule drugs that bind to IDPs have common features that are distinct from compounds targeting ordered proteins? Ruan et al. compared a group of IDP-binding compounds and a group of approved drugs for globular proteins ⁴². They observed that small molecules binding to IDPs tend to be more hydrophobic and contain more rings, mostly aromatic, than those binding to globular proteins.

2.1.2. Methods for studying the interaction between small molecules and intrinsically disordered proteins

Drug discovery techniques are well established for ordered proteins but do not always apply to IDPs²⁴. X-ray crystallography is the gold standard for determining the interactions between small molecules and globular partners, although other biophysical techniques have also been implemented and reviewed elsewhere⁴³. In the case of IDPs, due to their conformational heterogeneity, we need to take into account the whole ensemble when determining their binding to small molecules. For this reason, solution-state techniques that do not require crystallization are the methods of choice²⁴. These can be ensemble-averaged techniques, like nuclear magnetic resonance (NMR), or single-molecule approaches, like single-molecule fluorescence resonance energy transfer (smFRET).

SBDD techniques can be applied to IDP drug discovery when targeting a globular domain of the IDP or a globular binding partner. However, approaches to optimise affinity are not so well established when both binding partners are disordered, or when the target is the aggregated state or the monomeric form. Techniques applicable in these cases are summarised in Table 2.1.2.1²⁴.

Table 2.1.2.1. Summary of techniques to study IDP-drug interactions. Adapted from Heller et al. 2017²⁴.

1. Inhibiting IDP-IDP interactions

Technique	Experimental measure	Major advantages and / or disadvantages
Fluorescence Resonance Energy Transfer (FRET)	Measures the efficiency of the energy transfer between a donor fluorophore and an acceptor fluorophore. The efficiency is strongly dependent on the distance between them, enabling quantification of molecular associations or conformational changes.	Labeling of the samples is required.
Yeast-two-Hybrid	Indirectly measures protein-protein interactions. One partner is fused to a DNA-binding domain and the other one to an activation domain of a transcription factor. When both partners are proximal, the DNA-binding domain and activation domain can form a complex and activate transcription. Small molecule blockers of the association alter the transcriptional output in a quantitative manner.	It is a cellular assay. The measurement of the interaction between the two partners is indirect.
Fluorescence Polarization (FP)	FP measures the effects of fluorescent molecules (e.g. peptide or protein) on polarised light. Free molecules in solution will depolarise	The effects might not be very large if the drug-IDP interaction is “fuzzy”. It

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

	the light due to their rapid tumbling. Instead, the polarization is maintained when they are bound to other molecules. FP quantifies the fraction of fluorescent ligands bound to another protein and its modulation by small molecules.	requires labeling of one partner.
--	--	-----------------------------------

2. Inhibiting the aggregated state of IDPs

Technique	Experimental measure	Major advantages and / or disadvantages
Solid-state Nuclear Magnetic Resonance	Since amyloid fibrils are solids formed by repetitive units, solid-state NMR experiments can be applied and the structure of the fibril can be elucidated.	
Kinetic measurements	The kinetics of aggregation can be monitored using amyloid-specific fluorescent dyes, such as thioflavin T (ThT). ThT fluorescence is enhanced and red-shifted upon binding to β -sheet rich structures. Other complementary biophysical techniques are transmission electron microscopy (TEM), atomic force microscopy (AFM), and Fourier transform infrared spectroscopy (FTIR).	The chemical kinetics theory of aggregation explains the non-linear combinations of microscopic events that lead to aggregation, namely primary nucleation, elongation, secondary nucleation and fragmentation. The model has been shown that therapeutic strategies against amyloid aggregation should not only abolish fibril formation but also target the toxic oligomeric species.

3. Inhibiting the monomeric form of IDPs

Technique	Experimental measure	Major advantages and / or disadvantages
Single Molecule Fluorescence Resonance Energy Transfer (smFRET)	Both donor and acceptor chromophores are placed on the same molecule.	Measurements are not ensemble-averaged. The sum of many measurements results in the distribution of conformations within a

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

		<p>given sample.</p> <p>It requires protein labeling.</p>
Small-Angle X-ray Scattering (SAXS)	It is based on the scattering of X-rays of a sample and detection of conformational changes upon binding to ligands.	<p>It is a label-free technique.</p> <p>It is a low-resolution technique and only large changes in the conformational ensemble are detected.</p>
Mass-spectrometry (MS)	A range of MS methods can monitor discrete conformers in a mixture (e.g. ion mobility mass spectrometry) and detect ligand binding (hydrogen/deuterium exchange MS).	Detects interactions in the gas phase. In this phase, the hydrophobic effect is weakened and charge-charge interactions are strengthened.
Thermal Denaturation (TD)	Monitors the denaturation of the protein target (either of the tryptophan residues or external dyes) upon increasing temperatures in the absence and presence of ligands.	It is non-quantitative as the degree of stabilisation is not directly proportional to the affinity of the ligand.
Isothermal Calorimetry (ITC)	Measures the heat absorbed or released when a binding partner is titrated into a solution containing the other binding partner. It is the best technique to measure the binding constant, the Gibbs free energy of binding, the enthalpy, the entropy and the stoichiometry of the interaction.	It can measure the contributions of enthalpy and entropy to binding, which, as mentioned in section 4.1.1.a, is key for small molecules interacting with IDPs.
Surface Plasmon Resonance (SPR)	Measures changes in the refractive index at the surface of a bio-functionalised gold-coated prism. Association and dissociation of biomolecules immobilised on a surface can be determined because changes in mass will influence the refractive index.	<p>It is a highly-sensitive, label-free technique.</p> <p>Controls should be used to detect non-specific binding to the chip.</p>
Circular Dichroism (CD)	Measures the overall secondary structure content of proteins. Different secondary structures absorb left- and right-handed circularly polarised light differently, giving rise to characteristic spectra for β -sheet, α -helix and random coil structures.	It is a low-resolution technique and only large changes in secondary structure are detected.
Molecular Dynamics (MD)	Provides accurate descriptions of protein ensembles and their changes upon ligand binding.	<p>Force fields still approximate and validation with experiments is needed.</p> <p>Usually, IDPs of interest</p>

		are large macromolecules in a complex environment, which might not always be simulatable.
Solution Nuclear Magnetic Resonance (NMR)	Monitors the binding between a disordered protein and a small molecule, from the point of view of the small molecule or the protein. Protein-NMR is very powerful as it detects the changes in the environment surrounding each amino acid. When a ligand binds to a protein the binding sites can be determined.	It is a very sensitive, high-resolution and quantitative technique. It can measure binding and also dynamics. In protein-NMR, previous assignment of the peaks is needed in order to determine ligand-binding sites.

Due to the scope of this thesis, we will focus on NMR techniques to study the interactions between small molecules and monomeric disordered proteins. NMR is our technique of choice since:

- IDP-drug interactions are weaker than traditional globular protein-drug interactions: NMR is a highly sensitive technique and millimolar (mM) affinities can be measured.
- IDP-drug interactions may be multivalent: protein-based NMR is a high-resolution technique that detects changes in the environment for each amino acid in the protein and can detect distinct binding sites with different affinities.
- IDP-drug interactions may be fuzzy in the bound state: in contrast to the large conformational changes that are detected by SAXS or FRET, NMR can detect small perturbations in the environment.

NMR experiments can be ligand-oriented or protein-oriented ⁴⁴:

a) Ligand-oriented experiments

The advantages are that there is no need for isotopic-labeling of the protein or previous assignment of the protein NMR signals (useful for high-molecular weight proteins and/or proteins with complex spectra).

Ligand-oriented experiments exploit the differences in the relaxation properties of a free versus bound ligand. A bound ligand, due to its lower tumbling, displays lower chemical shift values and broader signals in the ¹H spectra of the ligand. The differences between free and bound ligand are exploited in the saturation-transfer difference (STD) experiments. STD is based on transferring

-

magnetization from the macromolecular target to the ligand, which is only possible in the bound state ⁴⁵.

b) Protein-oriented experiments

Protein-oriented experiments are much more informative, although isotopic labeling and assignment of the NMR chemical shifts are a prerequisite.

A widely used experiment is the two-dimensional proton-nitrogen correlation (¹H,¹⁵N HSQC) ⁴⁴. ¹H,¹⁵N HSQC measures ¹H,¹⁵N spin pairs. Since each amino acid has at least one ¹H,¹⁵N pair, the HSQC experiment theoretically yields one signal per residue (named chemical shift) and two extra ones per glutamine and asparagine side-chain. Chemical shifts are a population-weighted average of all the conformations sampled within the millisecond time scale.

¹H,¹⁵N pairs are very sensitive to the chemical environment and their chemical shifts can be monitored upon addition of ligands ²⁴. The perturbation of each peak can be measured (chemical shift perturbations or CSPs) to determine ligand binding sites. Since each ¹H,¹⁵N pair will give rise to a signal depending on its environment, the NMR spectra for IDPs is usually much narrower on the ¹H dimension and with overlapping peaks compared to globular proteins. This is a consequence of intrinsic disorder: each amino acid type will sense similar environments. This can be partially solved by lowering the pH of the samples and the temperature of the experiments.

Furthermore, the high solvent exposure of IDPs contributes to a significant chemical exchange with bulk water, causing severe line-broadening and reduction of the ¹H,¹⁵N signal intensities ²⁴. The ¹H,¹⁵N signals can also be reduced due to formation of higher-order assemblies, for example oligomers, that are invisible to the NMR experiments, and to conformational fluctuations, whose rate is comparable to the timescale of the NMR experiment. This can be solved with instrumental advances like the use of cryo-probes and specialised pulse-sequences that do not rely on proton relaxation, like ¹³C,¹⁵N correlation spectra (CON) ⁴⁶. Another pulse sequence that we have implemented to attenuate the exchange with the solvent is the use of cross-polarization assisted heteronuclear in-phase single-quantum correlation (CP-HISQC) ⁴⁷. CP-HISQC uses cross-polarization to transfer the magnetization from ¹H to ¹⁵N and also proton-decoupling.

In addition to ¹H,¹⁵N HSQC spectra, other experiments can give insights into:

- *Secondary structure propensity*

The secondary structure propensity of an IDP can be calculated by comparing the chemical shifts to those of a random coil. Several algorithms, like the SSP or δ 2D, use a combination of proton (¹H), nitrogen (¹⁵N), carbonyl (¹³C') and carbon alpha (¹³Ca) chemical shifts to give a residue-specific secondary structure propensity ^{48,49}.

- *Ensemble conformation*

Nuclear Overhauser effects (NOEs) are very useful restraints when determining the tertiary contacts on a globular protein. However, for IDPs, NOEs cannot always be used to obtain inter-proton distances due to the high heterogeneity of the IDP ensembles. Other NMR parameters to obtain structural information at atomic-resolution are better suited. These include chemical shifts, hydrogen exchange rates⁵⁰, residual dipolar couplings (RDCs)⁵¹ and paramagnetic relaxation enhancements (PREs)⁵².

- *Protein dynamics*

NMR relaxation rates provide information about the amplitudes and time scales of internal motions. ¹⁵N backbone relaxation rates are among the most sensitive and widely used⁵³. Relaxation rates refer to the speed at which the magnetization of a spin returns to its initial value after the 90° pulse. The magnetization has two components: the transverse component (or R₂) and the longitudinal component (or R₁). Measurements of R₁ and R₂ suffer from systematic errors, mainly due to the pulse sequences used and the water saturation transfer. To minimise these errors, we used the R₁ pulse sequences previously optimised at the Universitat de Barcelona's NMR facility⁵⁴ and we extracted the R₂ values from R₁/R_{1ρ} measurements instead of directly determining R₂.

Of note, although it is not covered in this thesis, all the NMR parameters described are also used in screening campaigns for drug discovery⁵⁵ and as structural restraints for ensemble generation⁵⁶.

2.1.3. Androgen receptor N-terminal domain

The androgen receptor N-terminal domain (AR-NTD) (residues 1-558) is the largest domain in the androgen receptor (AR) and it harbours the main transactivation function of the receptor: the activation function 1 (AF-1) (see section 1.2)⁵⁷. The AF-1 (residues 141-494) binds to multiple co-activators and is regulated by a myriad of PTMs.

The AF-1 is structurally subdivided into two transactivation units (Tau-1 and Tau-5)⁵⁸. The Tau-1 (residues 102-371) carries most of the activity in the full-length AR in the presence of the hormone. In contrast, the Tau-5 (residues 361-537) is important for the transactivation function of AR at low concentrations of androgens (comparable to castrate levels), as well as for the AR splice variants lacking the ligand-binding domain found in castration-resistant prostate cancer patients⁵⁹ (**Fig. 2.1.3.1**).

Although the AR-NTD is intrinsically disordered, it contains several motifs of helical propensity. It is important to highlight the ²³LQNLF²⁷ (involved in the N/C interaction⁶⁰), the ¹⁷⁹LKDIL¹⁸³ (core of the Tau-1) and the ⁴³³WHTLF⁴³⁷ (in the Tau-5) (**Fig. 2.1.3.1**). Previous work from the Salvatella lab identified three helical regions in the Tau-5 (named R1, R2 and R3) by means of NMR⁶¹.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

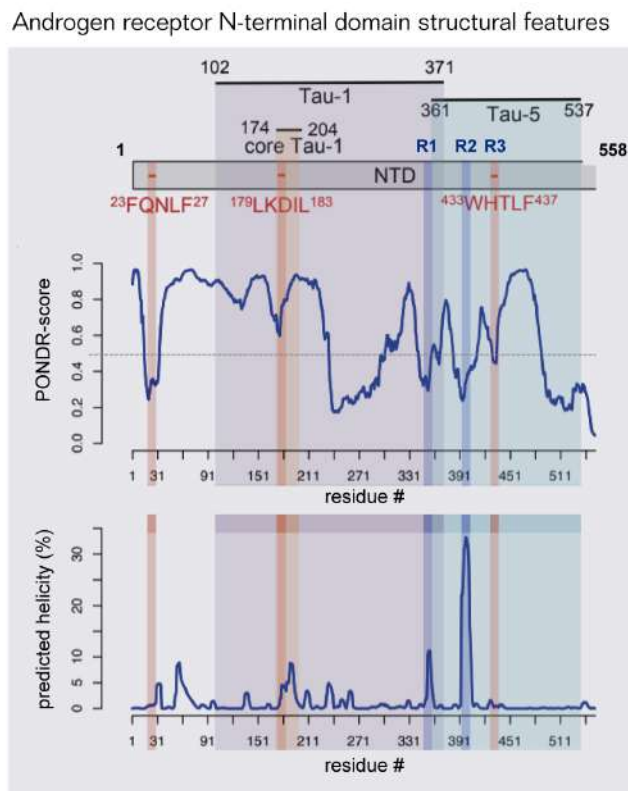


Fig. 2.1.3.1. Structural features and functional regions of the androgen receptor N-terminal domain. Above: plot of predicted degree of disorder by POND-R⁶² (POND-R-score higher than 0.5 indicates disorder whereas values lower than 0.5 indicate order). Below: plot of predicted helicity by Agadir⁶³. Adapted from De Mol Ph.D. Thesis, 2014⁶⁴.

2.1.4. EPI-001, the first inhibitor of the androgen receptor N-terminal domain to enter clinical trials

The AR-NTD is a promising drug target for CRPC because i) it gathers the main transcriptional activity of the AR at normal concentration (Tau-1) and low concentration (Tau-5) of hormone⁵⁹; ii) it is present in all forms of AR, including the constitutively active splice variants (SVs) that lack the ligand-binding domain (AR-LBD)⁶⁵ and iii) its low sequence homology to the other members of the steroid family of receptors makes it plausible for AR-NTD inhibitors to be highly specific⁶⁶. However, the intrinsically disordered nature of the AR-NTD has hampered the finding and progression of AR-NTD inhibitors into the market.

The Sadar lab has pioneered the discovery of AR-NTD inhibitors. The first *in vivo* evidence that inhibiting the AR-NTD resulted in a loss-of-function of the receptor was provided using decoy AR-NTD⁶⁷.

These results motivated them to perform a cellular phenotypic screening of several libraries. The assay consisted of the cellular expression of a fusion construct between the Gal4 DNA-binding domain and the AR-NTD. Binding of the construct to the Gal4 DNA response elements activated a luciferase gene expression. Hence compounds that inhibited the AR-NTD showed decreased

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

luciferase activity. A library containing natural product extracts provided *ca* 30 hits, clustered into different families. Three families of hits (Niphatenones, Sintokamides and EPIs) have been published so far and, interestingly, all of them are covalent inhibitors⁶⁶:

1. Niphatenones

Niphatenone B, the lead compound of this family, inhibits the full-length AR with an IC_{50} of 6 μ M and it is relatively active against splice variants⁶⁸ (**Fig. 2.1.4.1**). Niphatenones bind covalently to the AR AF-1. Further development has been stopped because they are not selective alkylators as they covalently bind to other macromolecules, like the glucocorticoid receptor and glutathione. Interestingly, (*S*)-niphatenone and (*R*)-niphatenone have similar IC_{50} values. Moreover, removal of the hydrophobic tail renders the compound inactive.

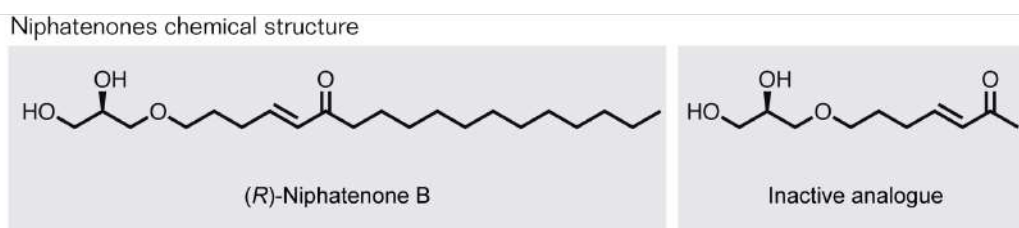


Fig. 2.1.4.1. Chemical structures of Niphatenone B and its inactive analogue.

2. Sintokamides

Sintokamide A was shown to bind to the AR AF-1 region, presumably in the Tau-1^{69,70} (**Fig. 2.1.4.2**). Sintokamide A is also a covalent inhibitor and blocks the activity of the full-length AR and its splice variants in cells and it is able to reduce tumour volume in a prostate cancer xenograft mouse model.

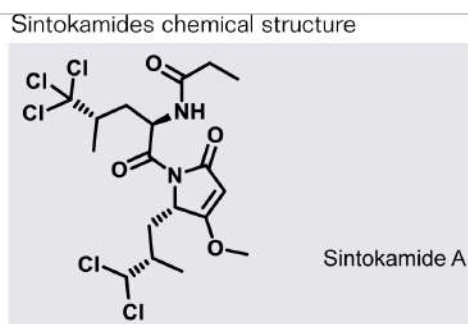


Fig. 2.1.4.2. Chemical structure of Sintokamide A.

3. EPI-001

EPI-001 is a mixture of 4 stereoisomers (EPI-002 to EPI-005) (**Fig. 2.1.4.3 a and b**)^{71,72}. EPI-002 is slightly more potent than the other stereoisomers. EPI-001 is also a covalent inhibitor and

substitution of the chlorohydrin warhead for a glycidol renders the compound inactive (BADGE) (**Fig. 2.1.4.3c**).

EPI inhibits the transcriptional activity of the full-length AR and its splice variants and blocks AR binding to its promoters and enhancers. Nevertheless, it is not very potent. EPI-001 has an IC_{50} *ca* 10 μ M and an optimal concentration of 25 μ M is required in *in vivo* assays^{66,71,72}. As compared to antiandrogens (see section 1.2), EPI compounds do not cause the translocation of the AR to the nucleus in the absence of the hormone. Moreover, EPI is specific for the AR as it does not inhibit closely-related receptors. *In vivo*, EPI reduced prostate tumour volume in a PC xenograft mouse model.

Given the activity of EPI compounds against AR splice variants, Imamura et al. developed ^{123}I -EPI-002, an imaging compound to identify patients with AR splice variant positive tumours (**Fig. 2.1.4.3d**). Interestingly, addition of an iodo atom in *ortho* position to the epichlorohydrin group of EPI-002 improved the activity of EPI-002 by one order of magnitude⁷³.

In 2015, ESSA Pharma, founded by Dr. Marianne Sadar and Dr. Raymond Andersen, started phase I/II clinical trials for a prodrug of EPI-002 (EPI-506 or ralaniten acetate, NCT02606123). Due to the low potency of EPI-506, high doses of the drug had to be administered, carrying side effects⁶⁶. Moreover, EPI-506 presented a poor pharmacokinetic profile and low metabolic stability⁷⁴. A second-generation of EPIs, EPI-7386, has entered phase I clinical trials in Q2 2020 (**Fig. 2.1.4.3e**)⁶⁶. The structure of a second generation compound has been made public (EPI-7170)⁷⁵.

EPIs chemical structure

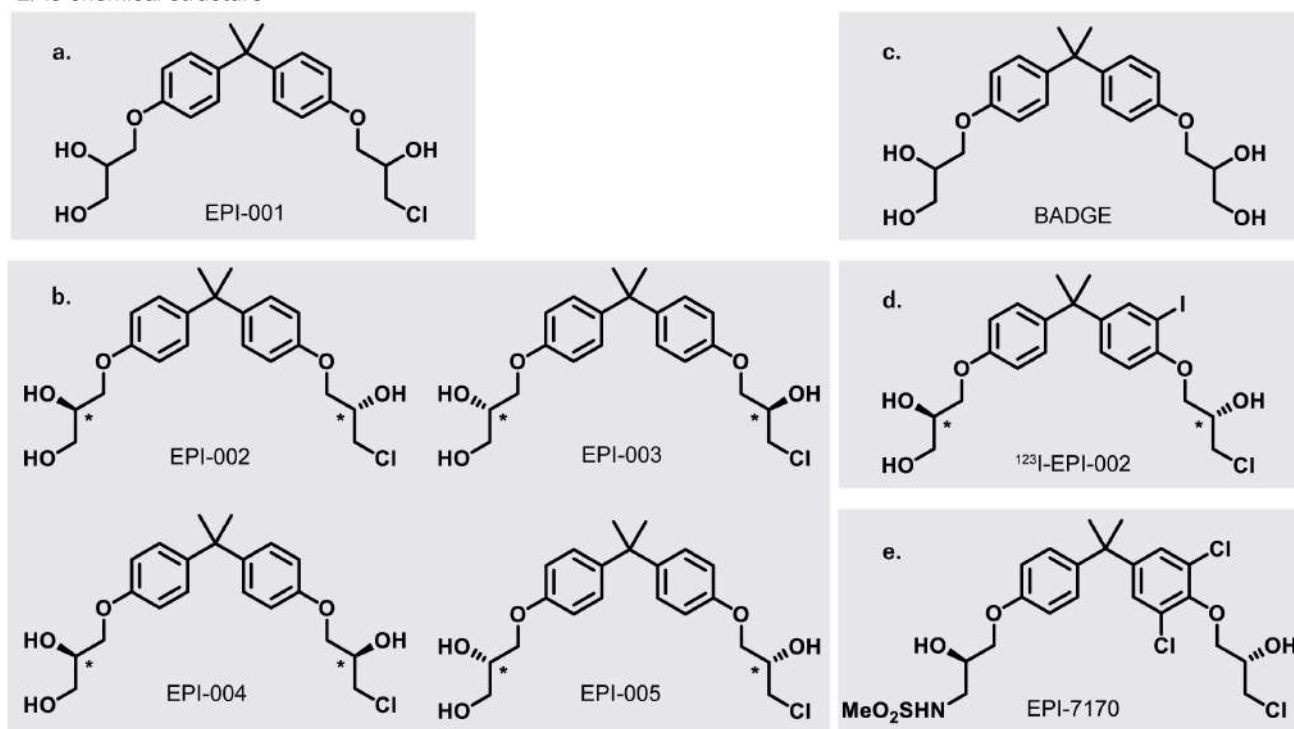


Fig. 2.1.4.3. Chemical structure of: **a)** EPI-001; **b)** EPI-001 stereoisomers; **c)** the inactive analogue BADGE; **d)** ^{123}I -EPI-002 and **e)** the second generation compound EPI-7170.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

EPI-001 has a proposed two-step mechanism of action (**Fig. 2.1.4.4**). It first interacts reversibly with the target and then the covalent reaction takes place. The AR-EPI adduct is responsible for the AR transcription inhibition⁷².

Proposed mechanism of action of EPI-001

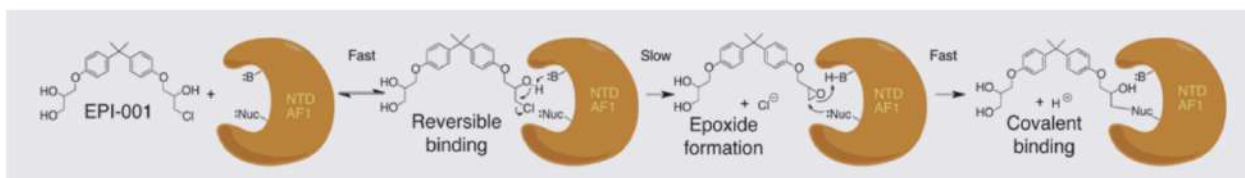


Fig. 2.1.4.4. Proposed mechanism of action of EPI-001. From Myung et al. 2013⁷².

Interaction of EPI-001 with the AF-1 was determined at low resolution *in vitro* by means of tryptophan fluorescence emission spectroscopy⁷¹ and in cells using AR pull-down experiments⁷².

High resolution data of the reversible interaction was acquired by NMR in the laboratory of Dr. Xavier Salvatella before I started my PhD⁶¹. De Mol et al. found that EPI-001 and all of its stereoisomers bound to the Tau-5 region in the AF-1 to the same extent. Specifically, 10 molar equivalents of EPI-001 caused small but systematic chemical shift perturbations (CSP) in the ¹⁵N dimension of the ¹H,¹⁵N spectrum of a construct spanning the AF-1 region (named AF-1*, residues 142-448). The CSP clustered in the R1, R2 and R3 regions within the Tau-5 (**Fig. 2.1.4.5a**). Interestingly, R1, R2 and R3 are three regions of sequence in the Tau-5 identified to be partially helical (**Fig. 2.1.4.5b**). EPI-001 did not cause CSP in the Tau-1 region, which also contains regions of helical propensity, nor in independent peptides spanning R1, R2 or R3. This last observation excluded a binding mechanism compared to the “protein cloud - ligand cloud” model for Myc, where several EPI-001 would bind to the AF-1* simultaneously, one to each helix (see section 2.1.1). EPI-001 affected the chemical shift of a large number of residues, more than of those expected for the binding of a small molecule. Given the fact that several EPI-001 molecules cannot bind simultaneously to three independent sites, those perturbations result from a combination of direct binding and indirect effects due to a conformational shift in the AF-1* ensemble upon EPI-001 binding. De Mol et al. did not further investigate whether this conformational shift was a conformational selection or a conformational expansion.

In any case, EPI-001 interacts with the conformations of the AF-1* ensemble in which R1, R2 and R3 adopt a partially helical structure. This observation led us to the first objective of this thesis: does structuration of R1, R2 and R3 helices improve the binding of EPI-001 to the AR-NTD?

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

Interaction of EPI-001 with the AF-1*

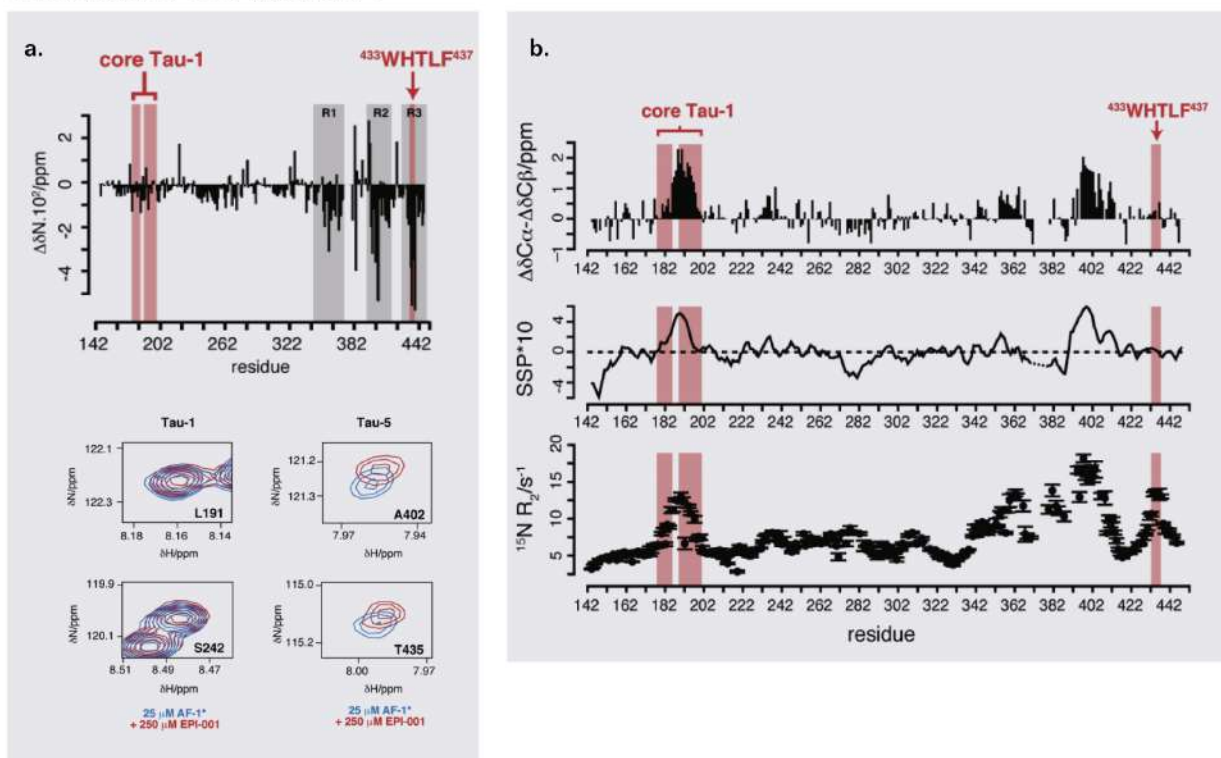


Fig. 2.1.4.5. a) EPI-001 interaction with the AF-1* by solution NMR spectroscopy. Above: 10 equivalents of EPI-001 cause chemical shift perturbations in R1, R2 and R3 but not in the Tau-1. Below: representative residues in the presence (red) and absence (blue) of EPI-001. **b)** Regions of helical propensity in the AF-1* experimentally determined by NMR. Above: difference between the secondary $C\alpha$ and $C\beta$ chemical shifts. Middle: secondary structure propensity of the residues of the AF-1* according to the SSP algorithm. SSP values are generated from backbone ($^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, $^{13}\text{C}'$, ^{15}N , ^1H) chemical shifts. SSP = 1 for fully helical and SSP = -1 for extended conformation. Below: ^{15}N transverse relaxation rates (R_2). Adapted from De Mol et al. 2016⁶¹.

2.2. Experimental results

The interaction between EPI-001 and the androgen receptor N-terminal domain (AR-NTD) is of very low affinity as demonstrated by De Mol et al.⁶¹. This is because the population of conformations of the AF-1 that EPI-001 interacts with, those with high helical propensity in the Tau-5 region, is of low percentage. We first hypothesised that increasing the helical propensity of the AF-1 would lead to a higher interaction affinity. To increase the helical content we used low amounts of the helical inducer 2,2,2-trifluoroethanol (TFE) reagent⁷⁶.

The constructs used to investigate the binding of EPI-001 to the AR-NTD were (**Fig. 2.2.1**):

- Tau-5*: a construct spanning the Tau-5 region, residues 330-448.
- AF-1*: a construct spanning the AF-1 region, residues 142-448.
- AR-NTD-L26P: a construct spanning the NTD, residues 1-558, bearing the L26P mutation (rationale in section 3.1.5).

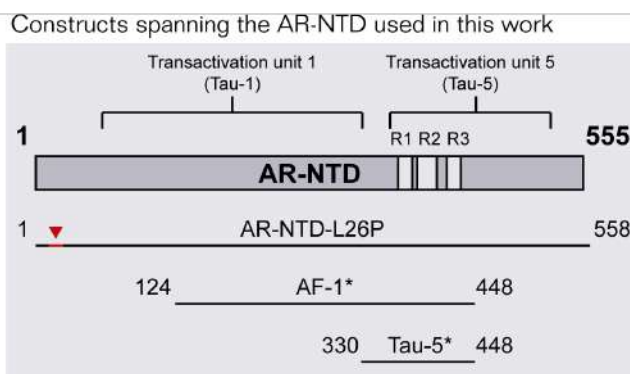


Fig.2.2.1. *In vitro* constructs used in this work spanning different regions of the AR-NTD: AR-NTD-L26P, AF-1* and Tau-5*. The AR-NTD has 2 transactivation units (Tau-1 and Tau-5). The red mark in the AR-NTD-L26P represents the L26P mutation. AF-1* and Tau-5* have already been described in the literature and contain three regions of helical propensity: R1, R2 and R3^{61,77}. AR-NTD-L26P was used in a previous PhD thesis of the group⁷⁸.

2.2.1. Trifluoroethanol increases the helical population in Tau-5* and AF-1*

2.2.1.1. Trifluoroethanol increases the helical population

Low amounts of TFE have been extensively used to induce helix formation⁷⁶. Indeed, titration of TFE into 16 μM Tau-5* increased the overall helical content of the Tau-5* as observed by circular dichroism (CD) spectroscopy (**Fig. 2.2.1.1**). This result relies on the fact that the molar residue ellipticity (MRE) at 220 nm presented a shoulder at 10% TFE that was absent at 0% TFE.

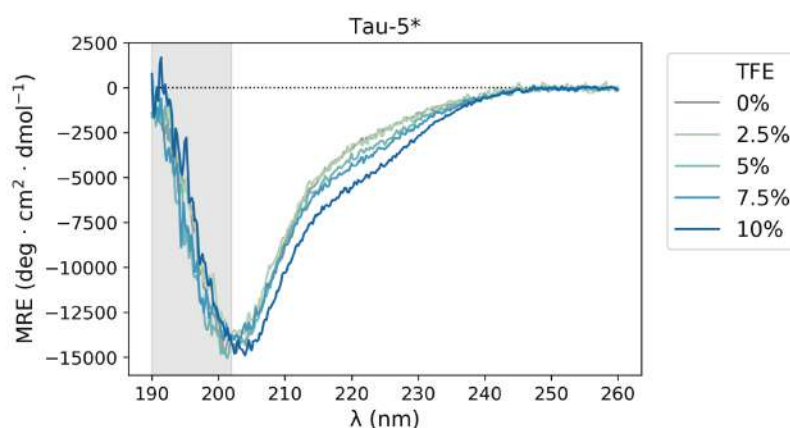


Fig. 2.2.1.1. CD spectra of 16 μM Tau-5* in 20 mM NaP buffer pH = 7.4 upon TFE titration. Shaded area highlights wavelength values for which high tension voltage (HT) and absorbance are out of range.

To obtain residue-specific information, we employed solution Nuclear Magnetic Resonance (NMR) spectroscopy. As mentioned in section 2.1.2, we have implemented the cross-polarization assisted heteronuclear in-phase single-quantum correlation (CP-HISQC) experiment⁴⁷. ^1H , ^{15}N CP-HISQC provides equivalent type of information to the standard ^1H , ^{15}N HSQC experiment. However, the ^1H , ^{15}N CP-HISQC pulse sequence is designed to be less sensitive to NH-solvent exchange. Therefore, the signals from the residues with fast-exchanging backbone amide groups that would not be detected using the conventional ^1H , ^{15}N HSQC experiment become observable.

We acquired the 2D ^1H , ^{15}N CP-HISQC and the 3D HNCA and HNCO spectra of the Tau-5* upon addition of TFE at increasing concentrations (**Fig. 2.2.1.2**). From the spectra we extracted the averaged ^1H and ^{15}N ($\Delta\delta_{\text{H,N}} = ((\Delta\delta_{\text{H}})^2 + (\Delta\delta_{\text{N}} / 5))^{1/2}$), the carbon alpha (Ca) and the carbonyl (C') chemical shifts, respectively. $\delta_{\text{H,N}}$ and C' but mainly Ca chemical shifts are very sensitive to the secondary structure. Together this data indicated that TFE addition increased the helical content of the Tau-5* in those regions with helical propensity (R1, R2 and R3) previously described⁶¹.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

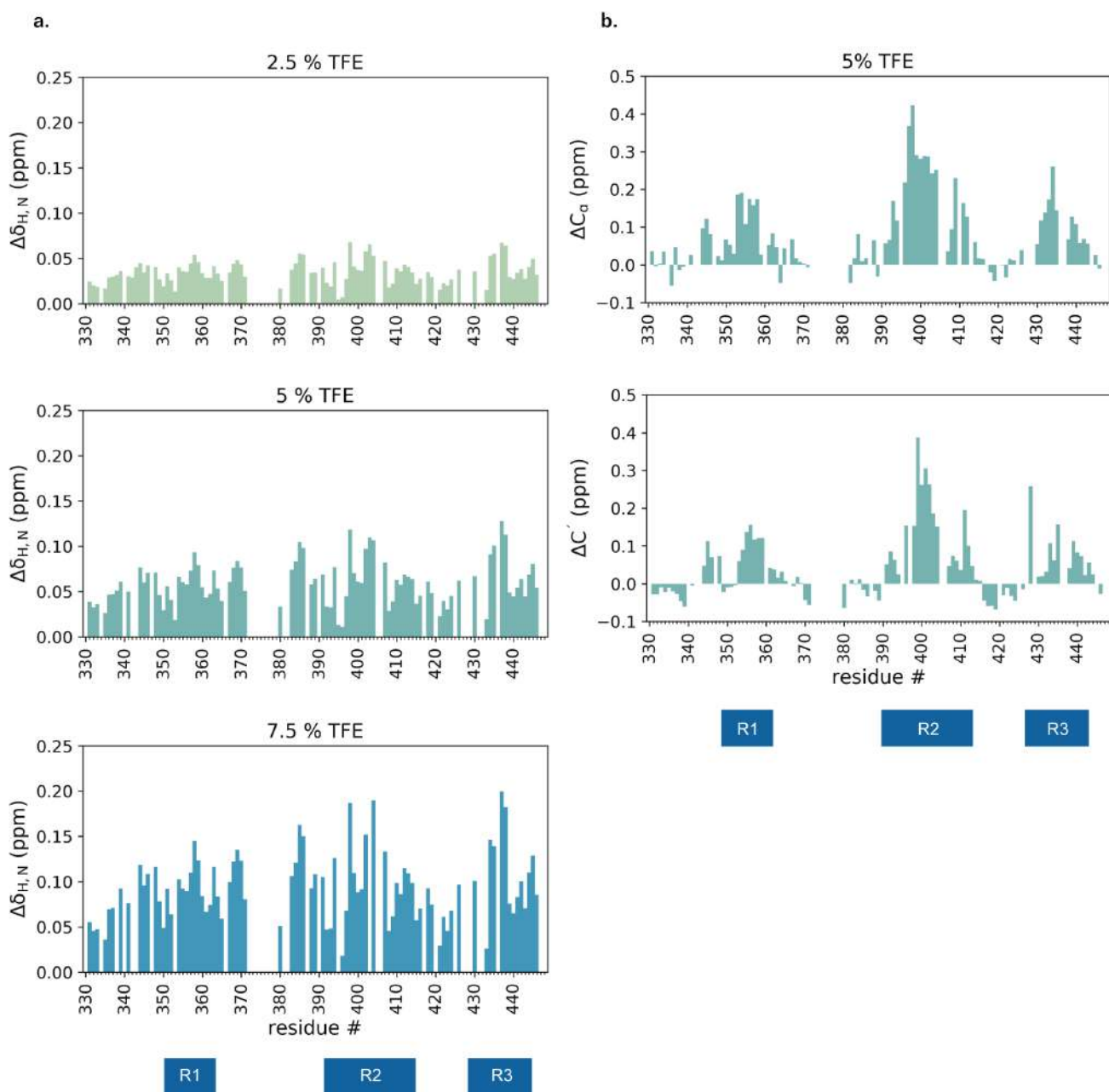


Fig. 2.2.1.2. a) Plot of the difference in the average ^1H and ^{15}N chemical shifts of $25\ \mu\text{M}$ ^{15}N Tau-5* in 20 mM NaP buffer, pH = 7.4 upon TFE addition (2.5, 5 and 7.5%). **b)** Plot of the difference in C_α and C' chemical shifts of $50\ \mu\text{M}$ $^{13}\text{C}, ^{15}\text{N}$ Tau-5* in 20 mM NaP buffer, pH = 7.4 upon addition of 5% TFE. Tau-5* regions with helical propensity are indicated as blue boxes as described in De Mol Ph.D. Thesis, 2014 ⁶⁴.

We next extended our findings to the AF-1 region. As for the Tau-5*, TFE increased the overall helical content of the AF-1* (**Fig. 2.2.1.3**). We observed that the effect of 10% TFE on the AF-1* was more pronounced than on the Tau-5*. This can be due to i) the contribution of the helical Tau-1 region, only present in the AF-1* construct or ii) a cooperative effect, which is enhanced in the larger AF-1* construct.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

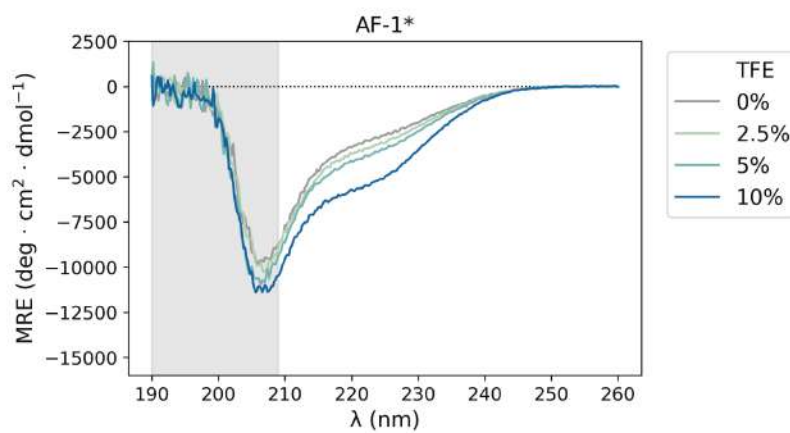
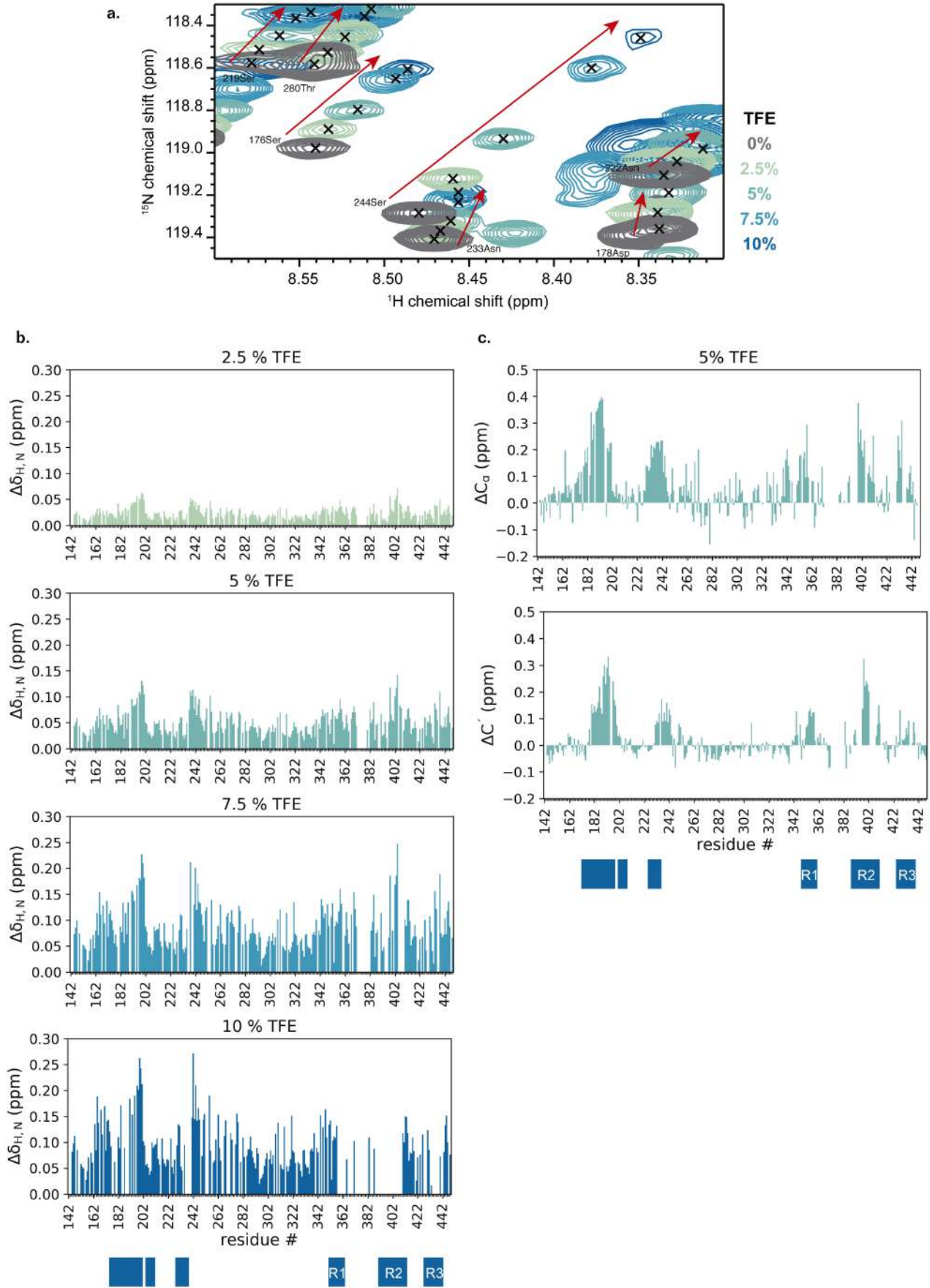


Fig. 2.2.1.3. CD spectra of 16 μM AF-1* in 20 mM NaP buffer pH = 7.4 upon TFE titration. Shaded area highlights wavelength values for which HT and absorbance are out of range.

The helical content was also particularly increased in those regions with helical propensity (**Fig. 2.2.1.4**).

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-



-

Fig. 2.2.1.4. **a)** Region of the ^1H , ^{15}N CP-HISQC NMR spectra (selected residues are indicated). **b)** Plot of the difference in averaged ^1H and ^{15}N chemical shifts of 25 μM ^{15}N AF-1* in 20 mM NaP buffer, pH = 7.4 upon TFE addition (2.5, 5, 7.5 and 10%). At 10% TFE many peaks were not detectable. **c)** Plot of the difference in Ca and C' chemical shifts of 50 μM ^{13}C , ^{15}N AF-1* in 20 mM NaP buffer, pH = 7.4 upon addition of 5% TFE addition. AF-1* regions with helical propensity are indicated as blue boxes as described in De Mol Ph.D. Thesis, 2014 ⁶⁴.

One key observation was that those residues showing increased helical content displayed broader peaks in the NMR spectra and the magnitude of broadening correlated with the increase in helicity (**Fig. 2.2.1.4a**). Peak broadening is a consequence of chemical exchange between two (or more) populations of the sample at a rate comparable to the NMR timescale. The chemical exchange with water was reduced by working at 4 °C and by using the CP-HISQC pulse sequence ⁴⁷. In this scenario, peak broadening can be due to i) exchange between different conformations of the monomer or ii) exchange with a higher order oligomer invisible to the NMR, since the sample only contained Tau-5* or only AF-1*. In other words, whether the protein was undergoing oligomerisation upon TFE addition.

2.2.1.2. Tau-5* and AF-1* are prone to oligomerise

To test the oligomerisation propensity of the Tau-5*, we obtained the ^1H , ^{15}N CP-HISQC spectra of 25 μM ^{15}N Tau-5* upon addition of non-labeled Tau-5* (**Fig. 2.2.1.5**). The chemical shifts of a monomeric protein should not be perturbed by an increase in concentration. In contrast, we observed concentration-dependent chemical shift perturbations (CSP). Moreover, CSPs were larger in those regions with helical propensity, especially R2 and R3.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

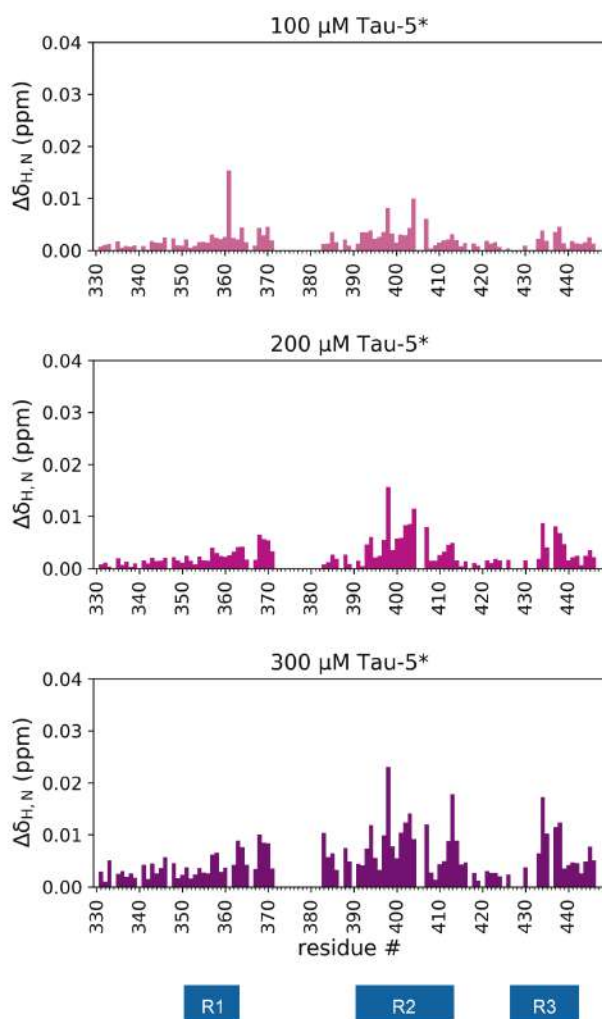


Fig. 2.2.1.5. Plot of the difference in average ^1H and ^{15}N chemical shifts of $25\ \mu\text{M}\ ^{15}\text{N}$ Tau-5* in 20 mM NaP buffer, pH = 7.4 upon addition of non-labeled Tau-5* to obtain a final sample concentration of: 100, 200 and 300 μM . Tau-5* regions with helical propensity are indicated as blue boxes as described in De Mol Ph.D. Thesis, 2014⁶⁴.

As the increase in concentration caused CSP in regions with helical propensity, we quantified the helical gain measuring the Ca and C' chemical shifts of a high concentration sample versus a low concentration sample. Indeed 300 μM Tau-5* (50 μM $^{13}\text{C},^{15}\text{N}$ + 250 μM non-labeled) had more helical content than 50 μM $^{13}\text{C},^{15}\text{N}$ Tau-5*, mainly in helix R₂ (**Fig. 2.2.1.6**). $^1\text{H},^{15}\text{N}$ peak broadening was very accused in residues located in the R3 helix. Therefore, the accuracy of the reported chemical shifts in this region is compromised.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

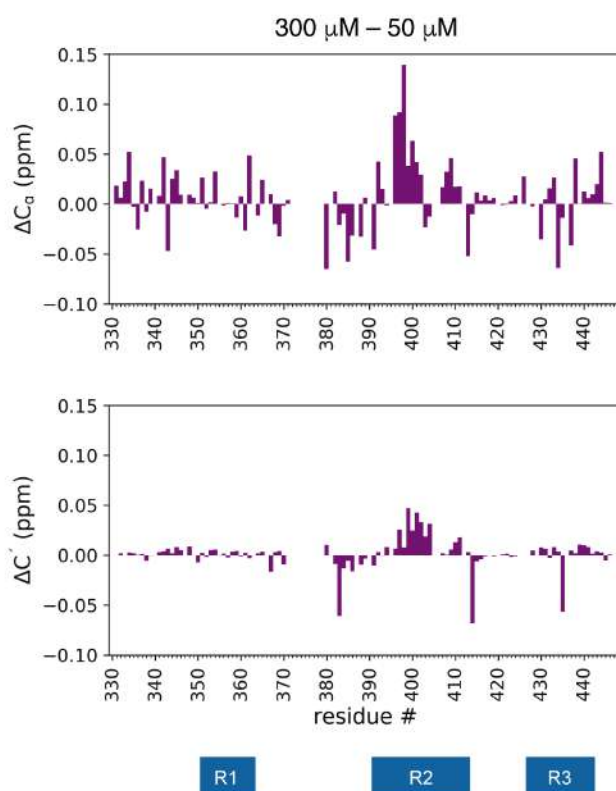


Fig. 2.2.1.6. Plot of the difference in Ca and C' chemical shifts of 300 μM Tau-5* (50 μM ^{13}C , ^{15}N + 250 μM non-labeled) versus 50 μM ^{13}C , ^{15}N Tau-5* in 20 mM NaP buffer, pH = 7.4. Tau-5* regions with helical propensity are indicated as blue boxes as described in De Mol Ph.D. Thesis, 2014 ⁶⁴.

The chemical exchange was further analysed by determining the ^{15}N transverse relaxation rates (R_2). The R_2 values were indirectly obtained from $R_1/R_{1\rho}$ measurements ⁷⁹. We observed that a 300 μM ^{15}N Tau-5* sample had increased R_2 relaxation rates compared to a 50 μM sample (**Fig. 2.2.1.7**). As expected, residues within predicted helical regions had higher R_2 values. Nevertheless, the increase in the R_2 upon concentration was more pronounced in those regions with helical propensity.

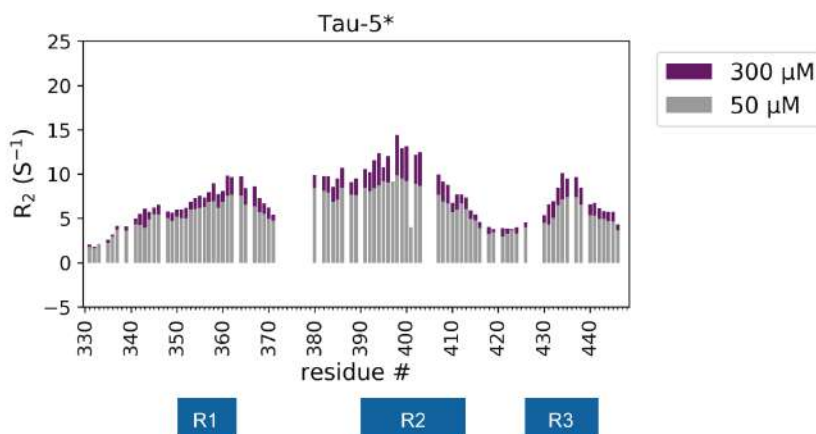


Fig. 2.2.1.7. Overlapping plots of the R_2 relaxation rates (extracted from $R_1/R_{1\rho}$ measurements) of a 50 μM and 300 μM ^{15}N Tau-5* sample in 20 mM NaP buffer, pH = 7.4. Tau-5* regions with helical propensity are indicated as blue boxes as described in De Mol Ph.D. Thesis, 2014 ⁶⁴.

-

We also measured the $R_1/R_{1\rho}$ relaxation rates of a $50\ \mu\text{M}$ ^{15}N Tau-5* sample upon addition of TFE. In this case we were only able to acquire the data for 0% and 2.5% of TFE as higher percentages caused sample evolution during data acquisition (*ca* 4 days). Addition of 2.5% TFE significantly increased the R_2 relaxation rates in helices R1, R2 and R3 (**Fig. 2.2.1.8**).

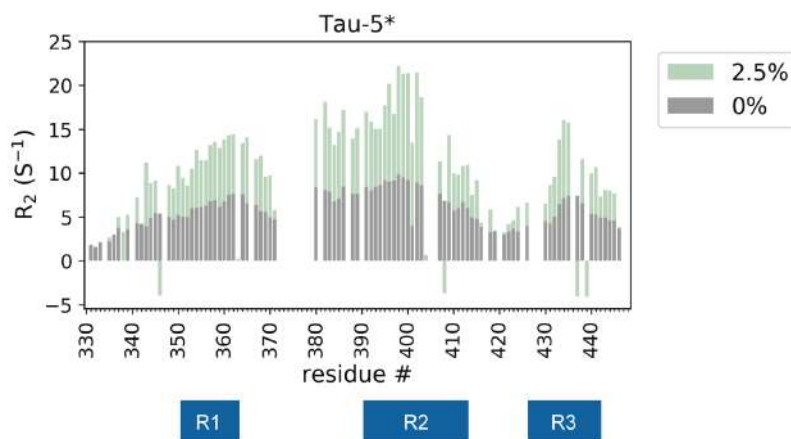


Fig. 2.2.1.8. Overlapping plots of the R_2 relaxation rates (extracted from $R_1/R_{1\rho}$ measurements) of a $50\ \mu\text{M}$ ^{15}N Tau-5* sample in the absence and presence of 2.5% TFE in 20 mM NaP buffer, pH = 7.4. Tau-5* regions with helical propensity are indicated as blue boxes as described in De Mol Ph.D. Thesis, 2014⁶⁴.

Together this data suggests that the Tau-5* is prone to oligomerise and does so through regions of helical propensity.

Oligomerisation, analogously to phase separation, depends on the valency of the protein of interest⁸⁰ and on the pattern of “stickers and spacers”⁸¹ (section 1.3). For this reason, we expected that the larger AF-1* construct would be more prone to oligomerise when compared to the Tau-5*.

The AF-1* also manifested concentration-dependent CSP (**Fig. 2.2.1.9**). It is important to highlight that the increase in the concentration of AF-1* caused perturbations in helical regions R2 and R3 in the Tau-5, as well as in residues 253-256 in the Tau-1, which are not predicted to be helical. This is in contrast to the effect of TFE, which only caused CSP in predicted helical regions (**Fig. 2.2.1.4**).

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

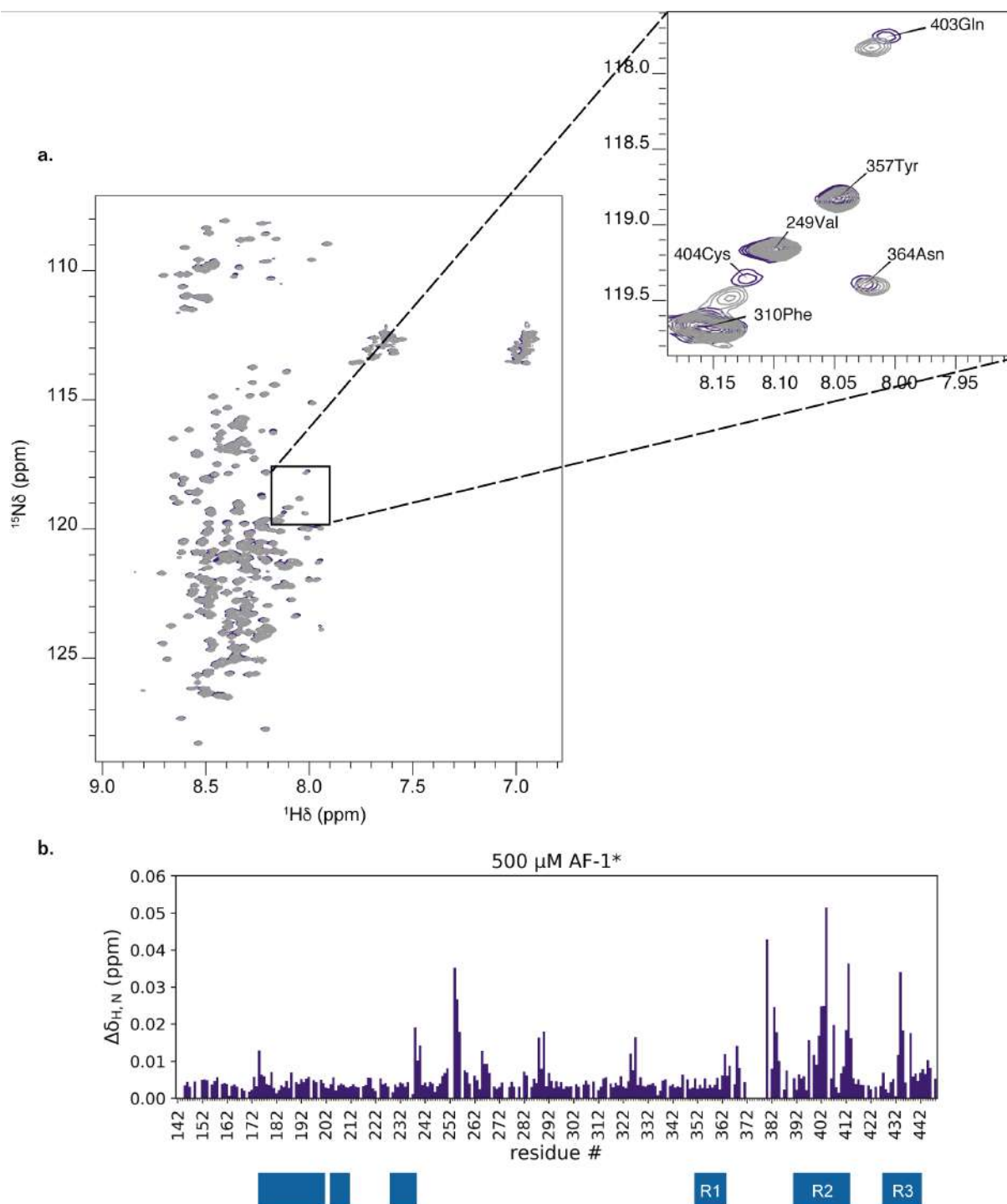


Fig. 2.2.1.9. a) Overlap of the ^1H , ^{15}N CP-HISQC NMR spectra of ^{15}N AF-1* at 50 μM (grey) and 500 μM (purple) and blow-up of a region. **b)** Plot of the difference in averaged ^1H and ^{15}N chemical shifts of 25 μM ^{15}N AF-1* in 20 mM NaP buffer, pH = 7.4 upon addition of non-labeled AF-1* to obtain a final sample concentration of 500 μM . AF-1* regions with helical propensity are indicated as blue boxes as described in De Mol Ph.D. Thesis, 2014⁶⁴.

From the spectra we can observe that the ^1H , ^{15}N peaks of residues exhibiting larger shifts upon concentration are also broader (**Fig. 2.2.1.9a**). To quantify this phenomena, we measured the $R_1/R_{1\rho}$ relaxation rates at 500 μM AF-1* (50 μM ^{15}N AF-1* + 450 μM non-labeled AF-1*) versus at 50 μM ^{15}N AF-1* (**Fig. 2.2.1.10**). As expected, residues within predicted helical regions had higher R_2

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

values. Noteworthy, upon concentration, residues in the Tau-5 region experienced higher R_2 relaxation rates compared to those in the Tau-1. Moreover, within the Tau-5, the increase in the transverse relaxation rates was more pronounced in the helical regions R1, R2 and R3.

From this data we can conclude that the increase in the transverse relaxation rates in specific regions upon concentration is due to an oligomerisation process associated with a gain in secondary structure. Upon concentration, the ensemble is shifted towards species with higher oligomeric propensity, with intermolecular interactions occurring mainly through the Tau-5 region, and/or with higher helical content in the Tau-5.

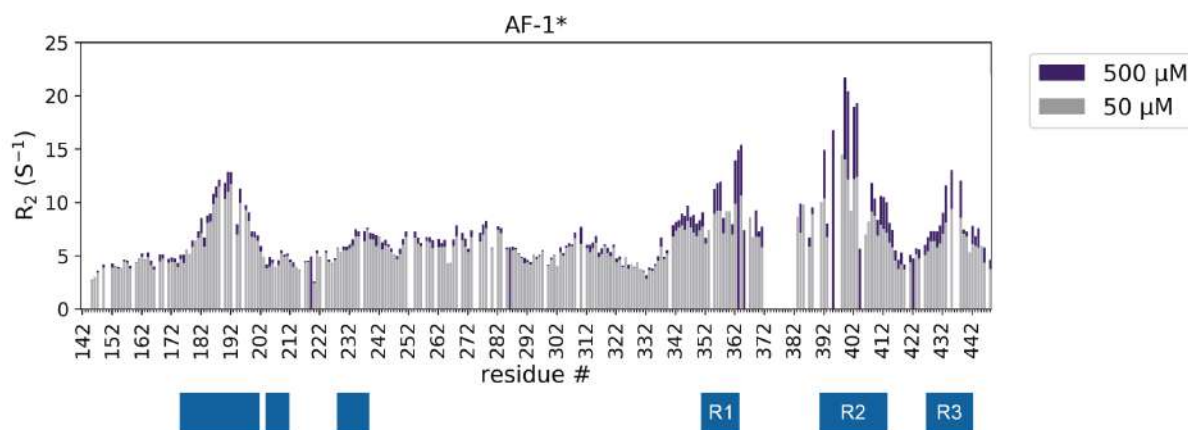


Fig. 2.2.1.10. Overlapping plots of the R_2 relaxation rates (extracted from $R_1/R_{1\rho}$ measurements) of a $50\ \mu\text{M}$ ^{15}N AF-1* and $500\ \mu\text{M}$ ($50\ \mu\text{M}$ ^{15}N + $450\ \mu\text{M}$ non-labeled) AF-1* sample in $20\ \text{mM}$ NaP buffer, $\text{pH} = 7.4$. AF-1* regions with helical propensity are indicated as blue boxes as described in De Mol Ph.D. Thesis, 2014

64

Finally, we also compared the $\text{C}\alpha$ and C' chemical shifts of a high concentration ($500\ \mu\text{M}$) versus a low concentration ($50\ \mu\text{M}$) AF-1* sample. The peak broadening at $500\ \mu\text{M}$ was very pronounced, which hindered the analysis. From this data we extracted that residues located in helix R2, like in the Tau-5* construct, but also in helix R3 displayed higher $\text{C}\alpha$ and C' chemical shifts at higher AF-1* concentration, indicating an increase in helicity upon concentration (**Fig. 2.2.1.11**).

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

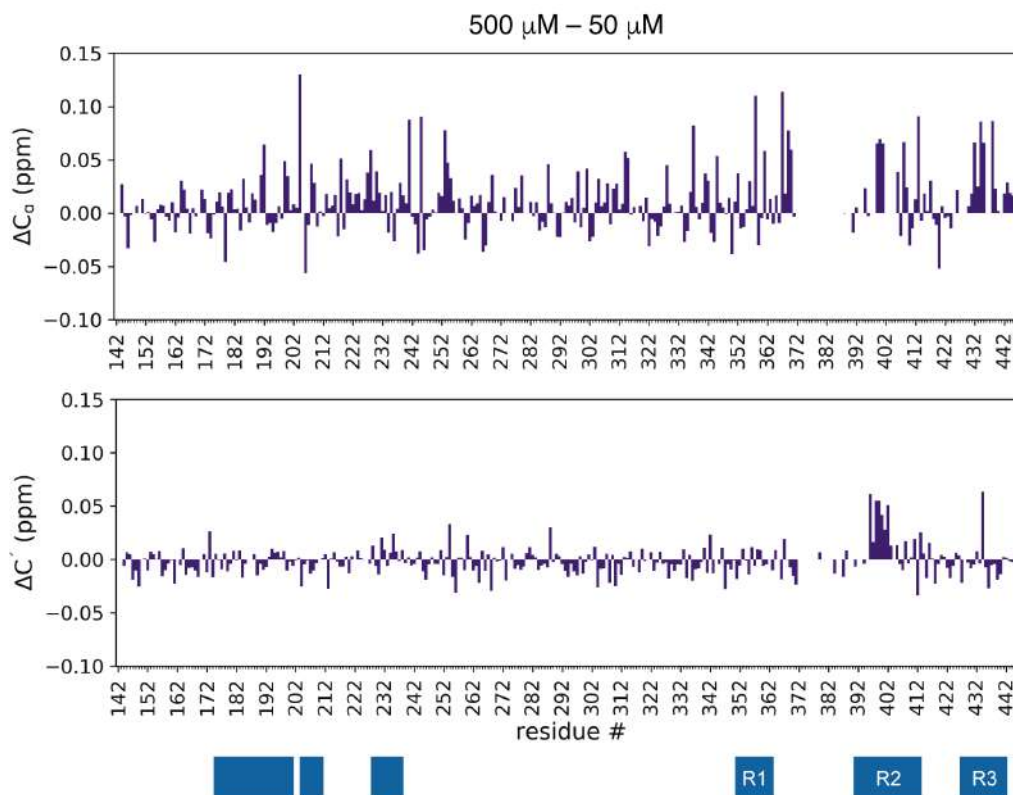


Fig. 2.2.1.11. Plot of the difference in Ca (top) and C' (bottom) chemical shifts of 500 μM ¹³C, ¹⁵N AF-1* (50 μM ¹⁵N + 450 μM non-labeled) versus 50 μM ¹³C, ¹⁵N AF-1* in 20 mM NaP buffer, pH = 7.4 upon addition of 5% TFE addition. AF-1* regions with helical propensity are indicated as blue boxes as described in De Mol Ph.D. Thesis, 2014 ⁶⁴.

In summary, the NMR data acquired indicates that the AF-1*, and particularly the Tau-5 region within the AF-1, is oligomerisation prone. Moreover, the oligomerisation is accompanied by a gain in helical content.

2.2.2. Interaction of EPI-001 with the AF-1* in the presence of TFE

We next determined whether the interaction affinity of EPI-001 for the AF-1 increased upon structuration of the construct by employing low amounts of TFE as in section 2.2.1.

First, we investigated the effect of TFE on the binding of EPI-001 to the AF-1* using ¹H NMR experiments based on the detection of the small-molecule. We recorded ¹H spectra of 250 μM EPI-001 in the presence of 25 μM AF-1* and increasing TFE percentages (**Fig. 2.2.2.1a**). We observed that the signals uniquely corresponding to EPI-001, such as EPI-001's aromatic ¹H, were broadening and shifting to lower chemical shift values upon TFE addition. This is indicative that structuration of the AF-1* leads to a stronger interaction with EPI-001. In contrast, the inactive

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

–

analogue BADGE experienced minor broadening and shifting of the signals upon TFE addition (**Fig. 2.2.2.1b**).

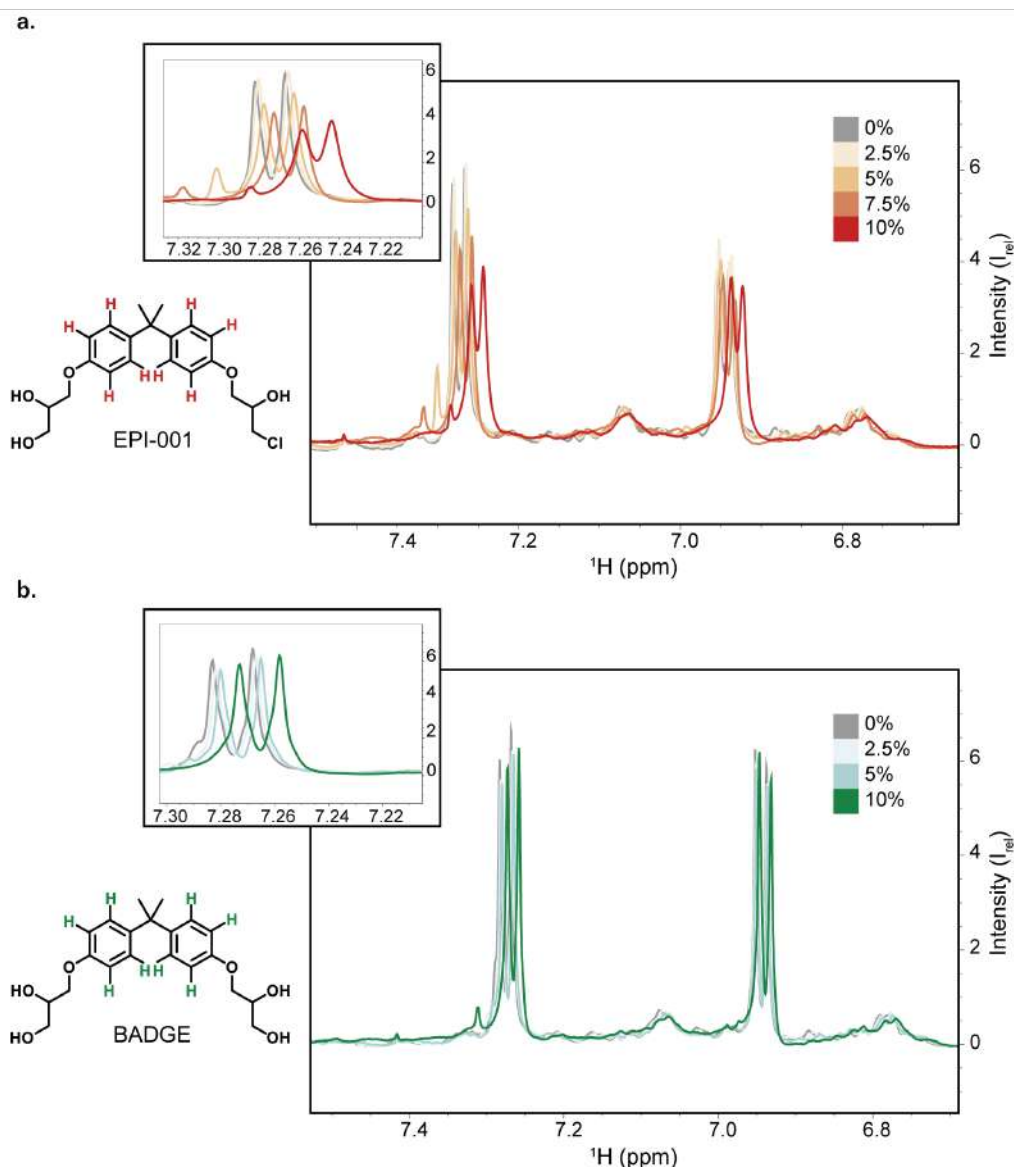


Fig. 2.2.2.1. Aromatic region in the 1D ^1H NMR spectra (and zoom-in from 7.22 to 7.32 ppm). **a)** The signals correspond to the ^1H highlighted in the chemical structure of 250 μM EPI-001 in the presence of 25 μM AF-1* and TFE (0, 2.5, 5, 7.5 and 10%). **b)** The signals correspond to the ^1H highlighted in the chemical structure of 250 μM BADGE in the presence of 25 μM AF-1* and TFE (0, 2.5, 5, and 10%).

However, we realised that for the sample with the highest TFE concentration (10%) EPI-001's signals were decreasing with time. We recorded the ^1H NMR spectra of this condition overtime and, indeed, EPI-001's signals decreased exponentially (**Fig. 2.2.2.2**). This phenomenon only occurred in the sample containing 10% TFE. This data led us to hypothesise that EPI-001 enhanced the AF-1* oligomerisation. Because the AF-1* oligomers are invisible to the NMR, EPI-001 bound to these oligomers suffers from severe peak broadening and, hence, its signals decrease in intensity. In contrast, BADGE signals remained constant in intensity over time under the same experimental

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

–

conditions (**Fig. 2.2.2.2**), thus suggesting that BADGE is not able to enhance AF-1* oligomerisation.

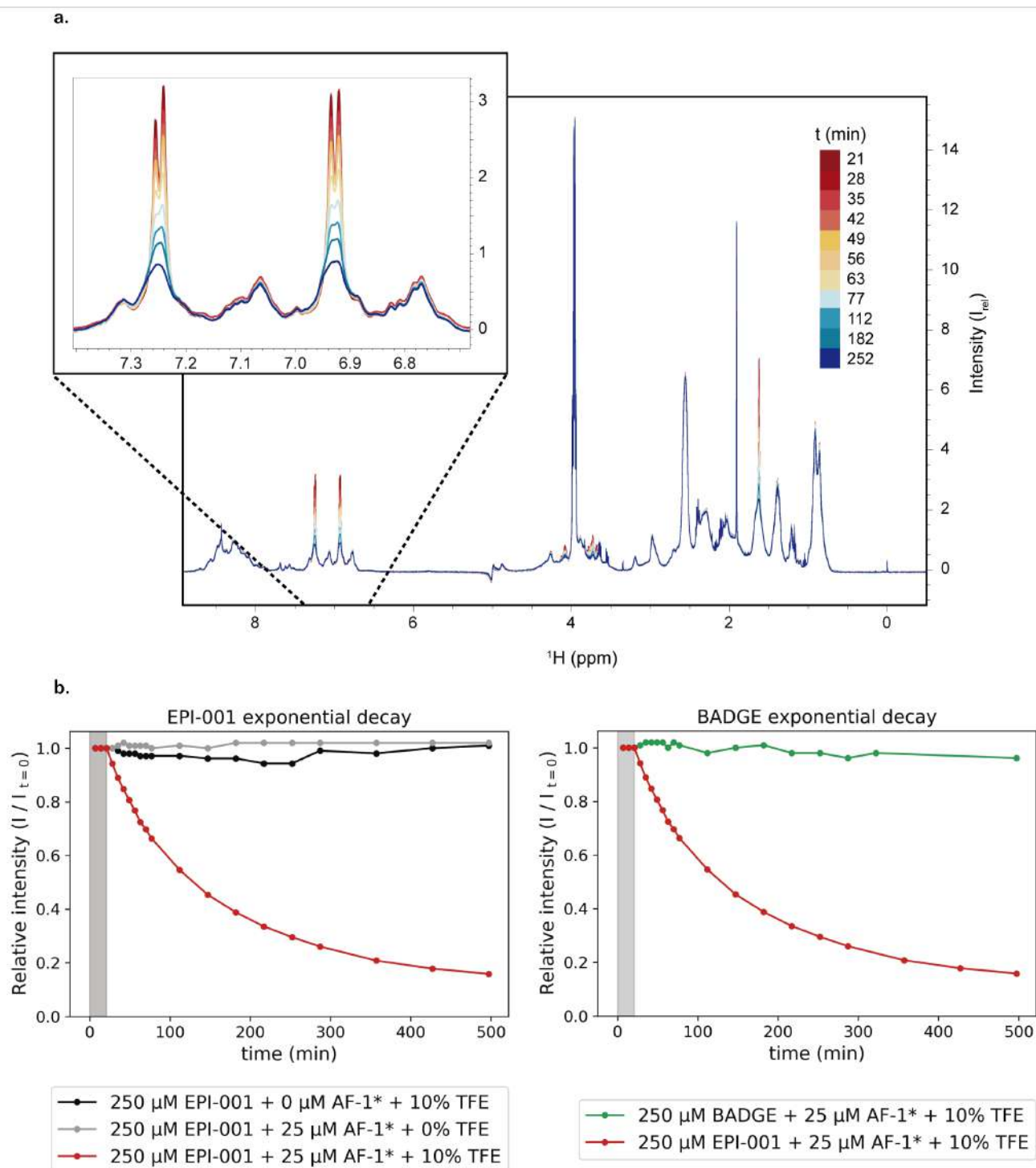


Fig. 2.2.2.2. a) Full spectra and zoom-in of the aromatic region of the ^1H NMR spectra of 250 μM EPI-001 in the presence of 25 μM AF-1* and 10% TFE overtime. **b)** Plot of the aromatic signal's intensity overtime (relative to $t = 0$ min) of EPI-001 in the presence of 10% TFE and 25 μM AF-1* (red), in the absence of TFE (grey), in the absence of AF-1* (black) and BADGE in the presence of 10% TFE and 25 μM AF-1* (green) in comparison to EPI-001. Shaded area indicates the time elapsed from sample preparation to first spectrum acquisition.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

We then analysed the interaction of EPI-001 with the AF-1* using protein-based ^1H , ^{15}N CP-HISQC experiments. We measured the perturbation caused by 10 equivalents (eq) of EPI-001 or BADGE to the chemical shifts of $25\ \mu\text{M}$ ^{15}N AF-1* at different TFE percentages (**Fig. 2.2.2.3**). At 0% TFE we were able to reproduce the previously published interaction data⁶¹. In the presence of TFE, addition of EPI-001 caused general chemical shift changes in the AF-1* ^1H , ^{15}N CP-HISQC spectra. BADGE also caused general perturbations to the AF-1* spectra but to a less extent. Noteworthy, addition of EPI-001 or BADGE to the spectra containing 10% TFE caused very severe peak broadening and many peaks could not be assigned.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

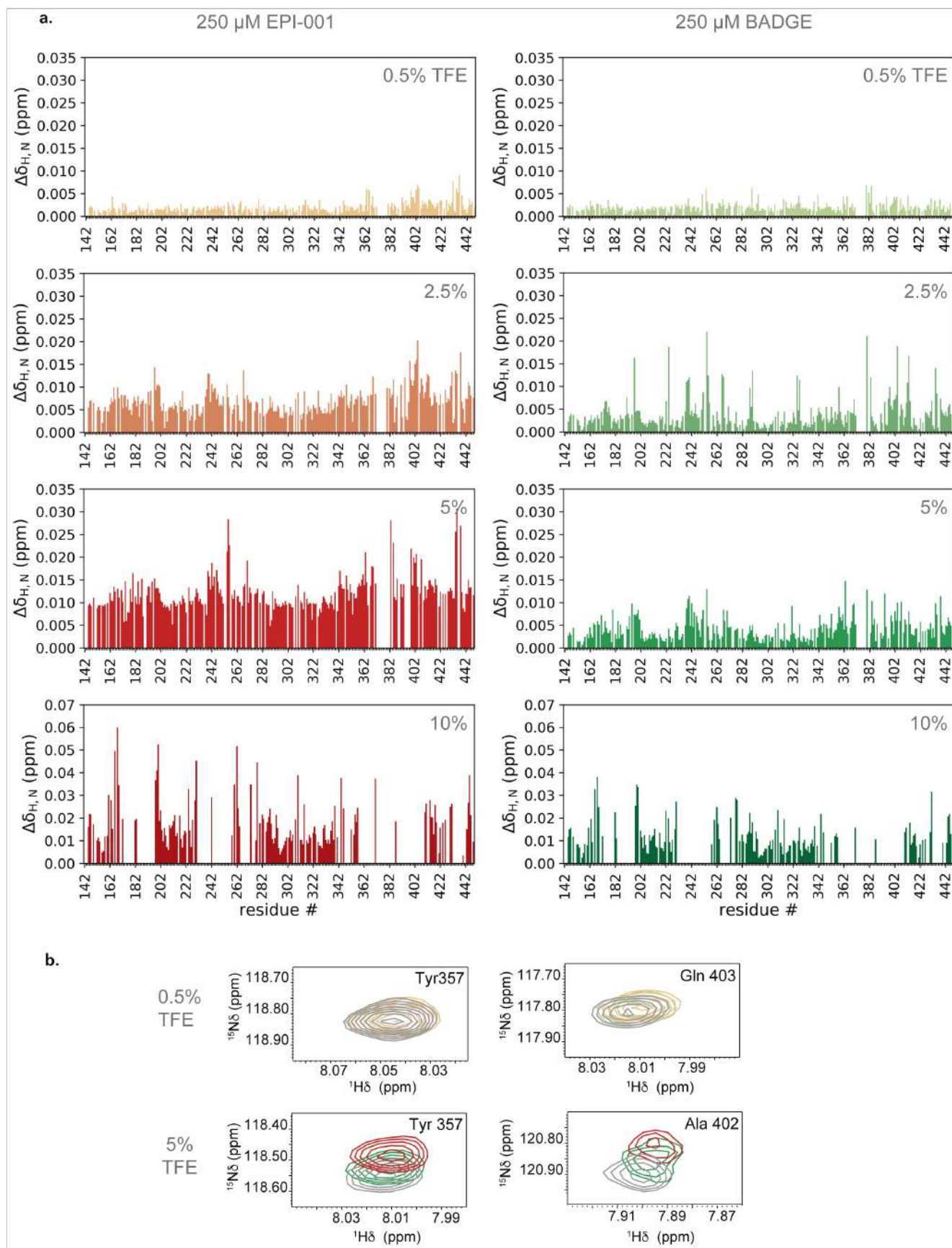


Fig. 2.2.2.3. a) Plot of the difference in average ^1H and ^{15}N chemical shifts of $25\ \mu\text{M}$ AF-1* in 20 mM NaP buffer, pH = 7.4 at 0, 2.5, 5 and 10% TFE in the absence (grey) and presence of $250\ \mu\text{M}$ EPI-001 (red) or $250\ \mu\text{M}$ BADGE (green). **b)** Blow-ups of selected residues at 0.5% and 5% TFE (grey) upon addition of EPI-001 (red) and BADGE (green).

In summary, this data suggests that EPI-001 enhances the AF-1* oligomerisation process. We were not able to discern between the direct and indirect chemical shifts caused by the addition of EPI-001 to the AF-1*. In other words, the ones arising from the small-molecule interaction (direct chemical shifts) from those of oligomerisation (indirect chemical shifts).

2.2.3. Interaction of EPI-001 with the AR-NTD

We also wanted to study the effect of EPI-001 to an even more oligomerisation prone construct, the AR-NTD-L26P (named AR-NTD for short and in consistency with section 3.2).

As for the AF-1*, the 1D ^1H spectrum of $250\ \mu\text{M}$ EPI-001 decreased in intensity in a sample containing $25\ \mu\text{M}$ AR-NTD and 10% TFE. This decay was much faster than the one observed when the AF-1* was present (**Fig. 2.2.3.1**). This indicates that the AR-NTD is more oligomerisation prone than the AF-1* and/or that EPI-001 interaction with the AR-NTD is stronger than with the AF-1*.

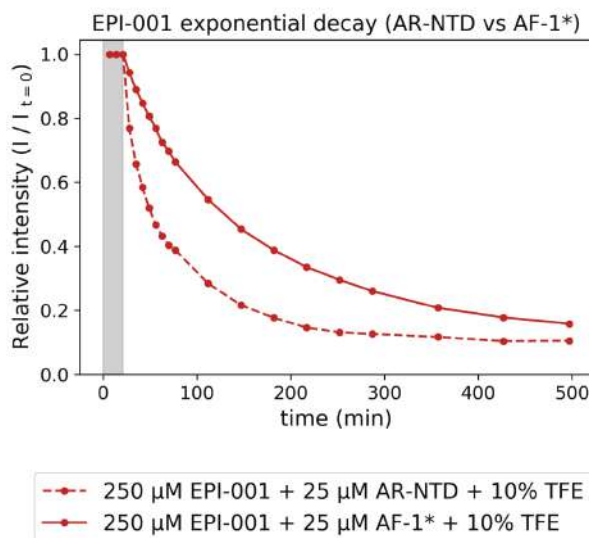


Fig. 2.2.3.1. Plot of the aromatic signal's intensity overtime (relative to $t = 0$ min) of EPI-001 in the presence of 10% TFE and $25\ \mu\text{M}$ AR-NTD (dotted line) or AF-1* (continuous line). Shaded area indicates the time elapsed from sample preparation to first spectrum acquisition.

During the course of this thesis we were able to assign the AR-NTD chemical shifts. This allowed us to map the interaction sites of EPI-001 onto the AR-NTD. However, what we observed was a general intensity decrease in the AR-NTD ^1H , ^{15}N CP-HISQC spectra upon addition of 10 eq of

-

EPI-001 (**Fig. 2.2.3.2a**). The decrease in intensity of the AR-NTD spectra was minimal upon addition of 10 eq of the inactive analogue BADGE (**Fig. 2.2.3.2b**). It is worth pointing out that the decrease in intensity is likely not due to experimental error. All samples were prepared from the same protein stock and following the same procedure. Spectral acquisition parameters were also validated to ensure no errors. Moreover, we can observe a stronger decrease in intensity for residues 1-80. This region contains the $^{23}\text{FQNLF}^{27}$ motif and the polyQ region which are aggregation prone ⁸²⁻⁸⁴. The Tau-5 region is also largely affected. Detectable residues in the Tau-5, which overlay with EPI-001's interaction motifs previously described ⁶¹, also experience a pronounced decrease in intensity. For these reasons and following the same rationale as for the AF-1*, we are confident that the intensity decrease upon EPI-001 addition is due to an enhancement of the AR-NTD oligomerisation process.

2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy

-

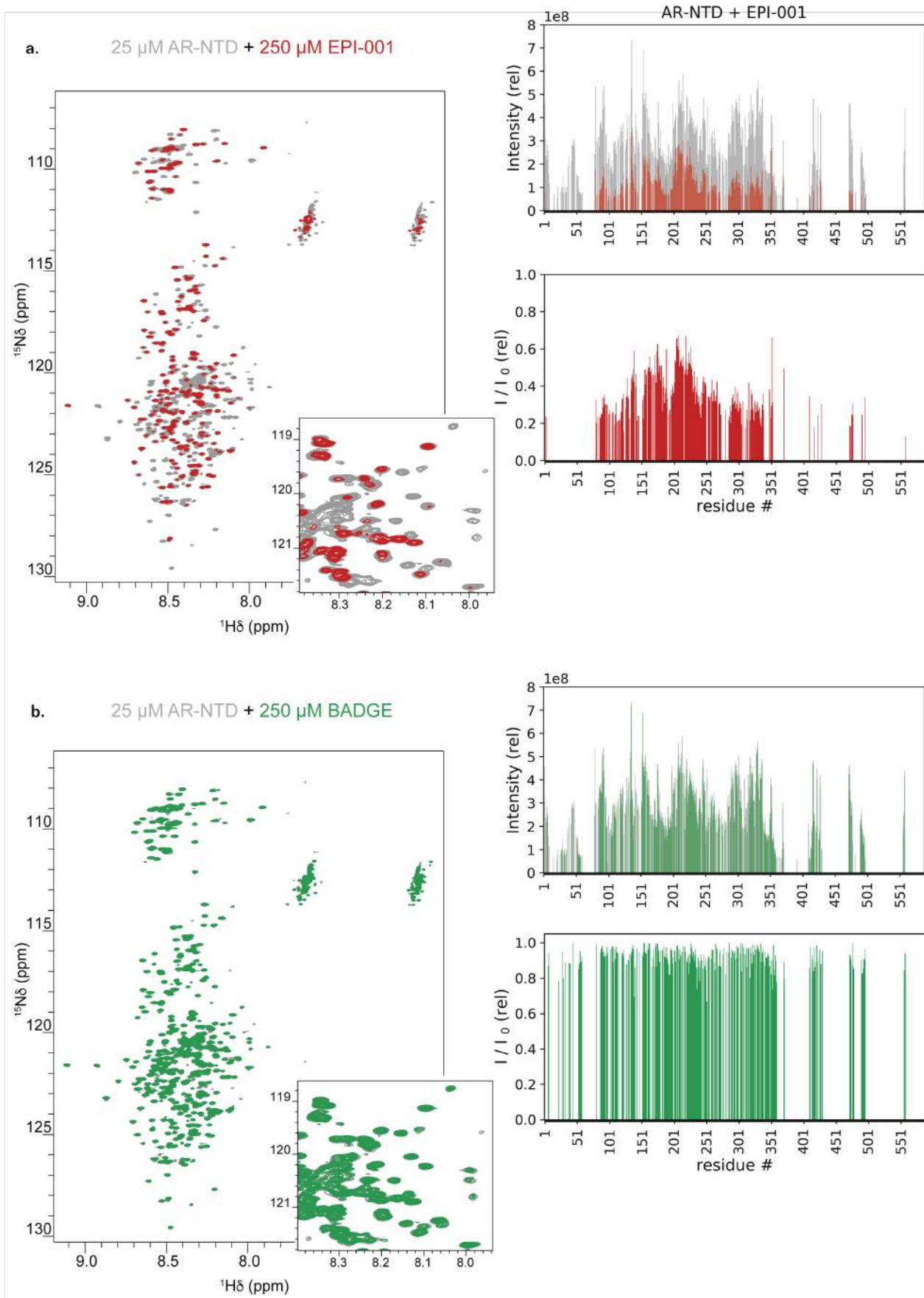


Fig. 2.2.3.2. a) ^1H , ^{15}N CP-HISQC spectra with blow-up of a region and plots of the intensity (top) and intensity changes (bottom) of the ^1H , ^{15}N CP-HISQC resonances of $25\ \mu\text{M}$ ^{15}N AR-NTD in $20\ \text{mM}$ NaP buffer, $\text{pH} = 7.4$ (grey) caused by addition of $250\ \mu\text{M}$ EPI-001 (red) or **b)** $250\ \mu\text{M}$ BADGE (green).

In summary, this data indicates that EPI-001 enhances the AR-NTD oligomerisation process. In contrast to the AF-1*, EPI-001 gives rise to a decrease in the AR-NTD signals even in the absence of TFE.

2.3. Discussion

Intrinsically disordered proteins (IDPs) are promising therapeutic targets as they are highly prevalent in the eukaryotic proteome and are associated with many diseases. However, the discovery of drugs targeting IDPs is very challenging as traditional structure-based methods cannot be applied due to their lack of a defined secondary and tertiary structure ¹.

Only two inhibitors targeting an IDP have reached clinical trials so far ⁶⁶. One of them, EPI-001, inhibits the disordered androgen receptor N-terminal domain (AR-NTD). Blocking the AR-NTD, specifically the Tau-5 region, is a promising strategy for the treatment of castration-resistant prostate cancer (CRPC) ^{71,85}.

These findings motivated the collaboration between the groups of Dr. Xavier Salvatella and Dr. Antoni Riera to study the interaction between the AR-NTD and EPI-001. Solution Nuclear Magnetic Resonance (NMR) is the methodology of choice to study IDPs as it is very sensitive, can detect mM affinity interactions, provides residue-specific resolution, is quantitative and can measure binding and also dynamics ²⁴.

Previous data from the group showed that EPI-001 caused chemical shift perturbations (CSP) in the three helical motifs within the Tau-5 region ⁶¹. Because EPI-001 did not cause CSP in other helical motifs, like the ones in the Tau-1, De Mol et al. concluded that EPI-001 interaction was specific to the Tau-5. They proposed a model in which EPI-001 binding to the Tau-5 depended on the ability of this domain to fold into an ensemble of conformations with high helical content and affinity for EPI-001. Moreover, EPI-001 binding caused CSP in a large number of residues, more than the ones expected from a small-molecule binding to a protein. This outcome was the consequence of the direct binding of EPI-001 as well as the indirect effects due to the population shift of the conformational ensemble.

These findings were the starting point of the present thesis. The initial plan was to increase the helical propensity of the constructs used by employing the helix inducer 2,2,2-trifluoroethanol (TFE) reagent as additive. The hypothesis was that if the helical conformations of the constructs were more populated, the affinity of EPI-001 for them would increase and the perturbations due to

-

direct binding would predominate over the indirect ones. However, we observed that structuration, promoted by TFE, enhanced the oligomerisation propensity of the constructs used. The oligomerisation process was further intensified with EPI-001 binding. Remarkably, we could not detect an increased oligomerisation propensity of the constructs upon addition of the biologically inactive analogue BADGE.

With this data in hand, we could classify EPI-001 to be a small-molecule inhibitor of an IDP with an allosteric inhibition mechanism (based on the classification proposed in section 2.1.1). EPI-001 causes a population shift or conformational selection towards conformations with helical content and oligomerisation propensity. However, we could not explore the direction of the conformational ensemble modulation (disorder-to-order, disorder-to-disorder or disorder-to-more-disorder) as we could not discern between the chemical shifts arising from the small-molecule interaction (direct chemical shifts) and those of structuration or oligomerisation (indirect chemical shifts).

In this chapter we studied the reversible interaction of EPI-001 with its target. Nevertheless, EPI-001 is a covalent inhibitor. In this sense, it would be interesting to study the effect of the covalent addition of EPI-001 to the spectrum of the AR-NTD and evaluate whether the oligomerisation propensity is more or less altered. Intensification of the oligomerisation effects would be a clear indication that covalent binding and oligomerisation enhancement are linked and related to EPI-001's mechanism of action.

The experiments conducted in chapter 2, together with other observations in the laboratory, led us to realise that the AF-1* and the AR-NTD underwent biomolecular condensation *in vitro*. As oligomerisation and condensation are tightly linked and as EPI-001 enhanced the oligomerisation propensity of the AR-NTD, we hypothesised that the mechanism of action of EPI-001 would be related to the modulation of the AR-NTD condensation process (chapter 3).

Bibliography

- (1) Uversky, V. N. Intrinsically Disordered Proteins and Novel Strategies for Drug Discovery. *Expert Opin. Drug Discov.* **2012**, *7*, 475–488.
- (2) Bah, A.; Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-Translational Modifications. *J. Biol. Chem.* **2016**, *291* (13), 6696–6705.
- (3) Darling, A. L.; Uversky, V. N. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front. Genet.* **2018**, *9* (158), 1–18.
- (4) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. Intrinsically Disordered Proteins: Regulation and Disease. *Curr. Opin. Struct. Biol.* **2011**, *21* (3), 432–440.
- (5) Martinelli, A. H. S.; Lopes, F. C.; John, E. B. O.; Carlini, C. R.; Ligabue-Braun, R. Modulation of Disordered Proteins with a Focus on Neurodegenerative Diseases and Other Pathologies. *Int. J. Mol. Sci.* **2019**, *20* (1322), 1–34.
- (6) Saraswati, A. P.; Ali Hussaini, S. M.; Krishna, N. H.; Babu, B. N.; Kamal, A. Glycogen Synthase Kinase-3 and Its Inhibitors: Potential Target for Various Therapeutic Conditions. *Eur. J. Med. Chem.* **2018**, *144*, 843–858.
- (7) Griebel, G.; Stemmelin, J.; Lopez-Grancha, M.; Boulay, D.; Boquet, G.; Slowinski, F.; Pichat, P.; Beeské, S.; Tanaka, S.; Mori, A.; Fujimura, M.; Eguchi, J. The Selective GSK3 Inhibitor, SAR502250, Displays Neuroprotective Activity and Attenuates Behavioral Impairments in Models of Neuropsychiatric Symptoms of Alzheimer’s Disease in Rodents. *Sci. Rep.* **2019**, *9* (18045), 1–15.
- (8) Bondeson, D. P.; Mares, A.; Smith, I. e. d.; Ko, E.; Campos, S.; Miah, A. h.; Mulholland, K. e.; Routly, N.; Buckley, D. l.; Gustafson, J. l.; Zinn, N.; Grandi, P.; Shimamura, S.; Bergamini, G.; Faelth-savitski, M.; Bantscheff, M.; Co, C.; Gordon, D. A.; Willard, R. R.; Flanagan, J. J.; Casillas, x. L. n.; Votta, B. J.; Den besten, W.; Famm, K.; Kruidenier, L.; Carter, P. s.; Harling, J. d.; Chur, I.; Crews, C. M. Catalytic in Vivo Protein Knockdown by Small- Molecule PROTAcS. *Nat. Chem. Biol.* **2015**, *11* (8), 611–617.
- (9) Winter, G. E.; Buckley, D. L.; Paulk, J.; Roberts, J. M.; Souza, A.; Dhe-Paganon, S.; Bradner, J. E. Phthalimide Conjugation as a Strategy for in Vivo Target Protein Degradation. *Science* **2015**, *348* (6241), 1376–1381.
- (10) Lu, M.; Liu, T.; Jiao, Q.; Ji, J.; Tao, M.; Liu, Y.; You, Q.; Jiang, Z. Discovery of a Keap1-Dependent Peptide PROTAC to Knockdown Tau by Ubiquitination-Proteasome Degradation Pathway. *Eur. J. Med. Chem.* **2018**, *146*, 251–259.
- (11) Ponnusamy, S.; Coss, C. C.; Thiyagarajan, T.; Watts, K.; Hwang, D.-J.; He, Y.; Selth, L. A.; McEwan, I. J.; Duke, C. B.; Pagadala, J.; Singh, G.; Wake, R. W.; Ledbetter, C.; Tilley, W. D.; Moldoveanu, T.; Dalton, J. T.; Miller, D. D.; Narayanan, R. Novel Selective Agents for the Degradation of Androgen Receptor Variants to Treat Castration-Resistant Prostate Cancer. *Cancer Res.* **2017**, *77* (22), 6282–6298.
- (12) Dunker, A. K.; Uversky, V. N. Drugs for “protein Clouds”: Targeting Intrinsically Disordered Transcription Factors. *Curr. Opin. Pharmacol.* **2010**, *10* (6), 782–788.
- (13) Chène, P.; Fuchs, J.; Bohn, J.; García-Echeverría, C.; Furet, P.; Fabbro, D. A Small Synthetic Peptide, Which Inhibits the p53-hdm2 Interaction, Stimulates the p53 Pathway in Tumour Cell Lines. *J. Mol. Biol.* **2000**, *299* (1), 245–253.
- (14) Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E. a. In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2. *Science* **2004**, *303* (5659), 844–848.
- (15) Soucek, L.; Helmer-Citterich, M.; Sacco, A.; Jucker, R.; Cesareni, G.; Nasi, S. Design and

Properties of a Myc Derivative That Efficiently Homodimerizes. *Oncogene* **1998**, *17*, 2463–2472.

- (16) Soucek, L.; Jucker, R.; Panacchia, L.; Ricordy, R.; Tato, F.; Nasi, S. Omomyc, a Potential Myc Dominant Negative, Enhances Myc-Induced Apoptosis. *Cancer Res.* **2002**, *62*, 3507–3510.
- (17) Annibali, D.; Whitfield, J. R.; Favuzzi, E.; Jauset, T.; Serrano, E.; Cuartas, I.; Redondo-Campos, S.; Folch, G.; González-Juncà, A.; Sodrì, N. M.; Massó-Vallés, D.; Beaulieu, M.-E.; Swigart, L. B.; Mc Gee, M. M.; Somma, M. P.; Nasi, S.; Seoane, J.; Evan, G. I.; Soucek, L. Myc Inhibition Is Effective against Glioma and Reveals a Role for Myc in Proficient Mitosis. *Nat. Commun.* **2014**, *5*, 1–11.
- (18) Erkizan, H. V.; Kong, Y.; Merchant, M.; Schlottmann, S.; Barber-Rotenberg, J. S.; Yuan, L.; Abaan, O. D.; Chou, T.-H.; Dakshanamurthy, S.; Brown, M. L.; Uren, A.; Toretsky, J. A. A Small Molecule Blocking Oncogenic Protein EWS-FLI1 Interaction with RNA Helicase A Inhibits Growth of Ewing's Sarcoma. *Nat. Med.* **2009**, *15* (7), 750–756.
- (19) Graves, B.; Thompson, T.; Xia, M.; Janson, C.; Lukacs, C.; Deo, D.; Di Lello, P.; Fry, D.; Garvie, C.; Huang, K.-S.; Gao, L.; Tovar, C.; Lovey, A.; Wanner, J.; Vassilev, L. T. Activation of the p53 Pathway by Small-Molecule-Induced MDM2 and MDMX Dimerization. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (29), 11788–11793.
- (20) Bier, D.; Mittal, S.; Bravo-Rodriguez, K.; Sowislok, A.; Guillory, X.; Briels, J.; Heid, C.; Bartel, M.; Wettig, B.; Brunsveld, L.; Sanchez-Garcia, E.; Schrader, T.; Ottmann, C. The Molecular Tweezer CLR01 Stabilizes a Disordered Protein-Protein Interface. *J. Am. Chem. Soc.* **2017**, *139* (45), 16256–16263.
- (21) Helsen, C.; Van den Broeck, T.; Voet, A.; Prekovic, S.; Van Poppel, H.; Joniau, S.; Claessens, F. Androgen Receptor Antagonists for Prostate Cancer Therapy. *Endocr. Relat. Cancer* **2014**, *21* (4), 105–118.
- (22) Babu, M. M. The Contribution of Intrinsically Disordered Regions to Protein Function, Cellular Complexity, and Human Disease. *Biochem. Soc. Trans.* **2016**, *44* (5), 1185–1200.
- (23) Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shamma, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M. Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *J. Am. Chem. Soc.* **2011**, *133* (36), 14160–14163.
- (24) Heller, G. T.; Aprile, F. A.; Vendruscolo, M. Methods of Probing the Interactions between Small Molecules and Disordered Proteins. *Cell. Mol. Life Sci.* **2017**, *74* (17), 3225–3243.
- (25) Munke, A.; Persson, J.; Weiffert, T.; De Genst, E.; Meisl, G.; Arosio, P.; Carnerup, A.; Dobson, C. M.; Vendruscolo, M.; Knowles, T. P. J.; Linse, S. Phage Display and Kinetic Selection of Antibodies That Specifically Inhibit Amyloid Self-Replication. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (25), 6444–6449.
- (26) Fonseca-Ornelas, L.; Schmidt, C.; Camacho-Zarco, A. R.; Fernandez, C. O.; Becker, S.; Zweckstetter, M. Small-Molecule-Induced Soluble Oligomers of α -Synuclein with Helical Structure. *Chemistry* **2017**, *23* (53), 13010–13014.
- (27) Tóth, G.; Gardai, S. J.; Zago, W.; Bertoncini, C. W.; Cremades, N.; Roy, S. L.; Tambe, M. A.; Rochet, J.-C.; Galvagnion, C.; Skibinski, G.; Finkbeiner, S.; Bova, M.; Regnstrom, K.; Chiou, S.-S.; Johnston, J.; Callaway, K.; Anderson, J. P.; Jobling, M. F.; Buell, A. K.; Yednock, T. A.; Knowles, T. P. J.; Vendruscolo, M.; Christodoulou, J.; Dobson, C. M.; Schenk, D.; McConlogue, L. Targeting the Intrinsically Disordered Structural Ensemble of α -Synuclein by Small Molecules as a Potential Therapeutic Strategy for Parkinson's Disease. *PLoS One* **2014**, *9* (2), 1–11.
- (28) Seidler, P. M.; Boyer, D. R.; Rodriguez, J. A.; Sawaya, M. R.; Cascio, D.; Murray, K.; Gonen, T.; Eisenberg, D. S. Structure-Based Inhibitors of Tau Aggregation. *Nat. Chem.* **2018**, *10* (2),

170–176.

- (29) Heller, G. T.; Sormanni, P.; Vendruscolo, M. Targeting Disordered Proteins with Small Molecules Using Entropy. *Trends Biochem. Sci.* **2015**, *40* (9), 491–496.
- (30) Flock, T.; Weatheritt, R. J.; Latysheva, N. S.; Babu, M. M. Controlling Entropy to Tune the Functions of Intrinsically Disordered Regions. *Curr. Opin. Struct. Biol.* **2014**, *26*, 62–72.
- (31) Heller, G. T.; Bonomi, M.; Vendruscolo, M. Structural Ensemble Modulation upon Small-Molecule Binding to Disordered Proteins. *J. Mol. Biol.* **2018**, *430*, 2288–2292.
- (32) Wright, P. E.; Dyson, H. J. Linking Folding and Binding. *Curr. Opin. Struct. Biol.* **2009**, *19* (1), 31–38.
- (33) Habazettl, J.; Allan, M.; Jensen, P. R.; Sass, H.-J.; Thompson, C. J.; Grzesiek, S. Structural Basis and Dynamics of Multidrug Recognition in a Minimal Bacterial Multidrug Resistance System. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (51), 5498–5507.
- (34) Yin, X.; Giap, C.; Lazo, J. S.; Prochownik, E. V. Low Molecular Weight Inhibitors of Myc-Max Interaction and Function. *Oncogene* **2003**, *22* (40), 6151–6159.
- (35) Follis, A. V.; Hammoudeh, D. I.; Wang, H.; Prochownik, E. V.; Metallo, S. J. Structural Rationale for the Coupled Binding and Unfolding of the c-Myc Oncoprotein by Small Molecules. *Chemistry and Biology* **2008**, *15* (11), 1149–1155.
- (36) Hammoudeh, D. I.; Follis, A. V.; Prochownik, E. V.; Metallo, S. J. Multiple Independent Binding Sites for Small-Molecule Inhibitors on the Oncoprotein c-Myc. *J. Am. Chem. Soc.* **2009**, *131*, 7390–7401.
- (37) Jin, F.; Yu, C.; Lai, L.; Liu, Z. Ligand Clouds around Protein Clouds: A Scenario of Ligand Binding with Intrinsically Disordered Proteins. *PLoS Comput. Biol.* **2013**, *9* (10), 1–11.
- (38) Neira, J. L.; Bintz, J.; Arruebo, M.; Rizzuti, B.; Bonacci, T.; Vega, S.; Lanas, A.; Velázquez-Campoy, A.; Iovanna, J. L.; Abián, O. Identification of a Drug Targeting an Intrinsically Disordered Protein Involved in Pancreatic Adenocarcinoma. *Sci. Rep.* **2017**, *7* (39732), 1–15.
- (39) Ban, D.; Iconaru, L. I.; Ramanathan, A.; Zuo, J.; Kriwacki, R. W. A Small Molecule Causes a Population Shift in the Conformational Landscape of an Intrinsically Disordered Protein. *J. Am. Chem. Soc.* **2017**, *139* (39), 13692–13700.
- (40) 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.; Tonks, N. K. Targeting the Disordered C Terminus of PTP1B with an Allosteric Inhibitor. *Nat. Chem. Biol.* **2014**, *10* (7), 558–566.
- (41) Metallo, S. J. Intrinsically Disordered Proteins Are Potential Drug Targets. *Curr. Opin. Chem. Biol.* **2010**, *14* (4), 481–488.
- (42) Ruan, H.; Sun, Q.; Zhang, W.; Liu, Y.; Lai, L. Targeting Intrinsically Disordered Proteins at the Edge of Chaos. *Drug Discov. Today* **2019**, *24* (1), 217–227.
- (43) Renaud, J.-P.; Chung, C.-W.; Danielson, U. H.; Egner, U.; Hennig, M.; Hubbard, R. E.; Nar, H. Biophysics in Drug Discovery: Impact, Challenges and Opportunities. *Nat. Rev. Drug Discov.* **2016**, *15* (10), 679–698.
- (44) Salvatella, X.; Giralto, E. NMR-Based Methods and Strategies for Drug Discovery. *Chem. Soc. Rev.* **2003**, *32* (6), 365–372.
- (45) Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88* (7), 990–994.
- (46) Hsu, S.-T. D.; Bertocini, C. W.; Dobson, C. M. Use of Protonless NMR Spectroscopy to Alleviate the Loss of Information Resulting from Exchange-Broadening. *J. Am. Chem. Soc.* **2009**, *131* (21), 7222–7223.
- (47) Yuwen, T.; Skrynnikov, N. R. CP-HISQC: A Better Version of HSQC Experiment for Intrinsically Disordered Proteins under Physiological Conditions. *J. Biomol. NMR* **2014**, *58*

- (3), 175–192.
- (48) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. Sensitivity of Secondary Structure Propensities to Sequence Differences between Alpha- and Gamma-Synuclein: Implications for Fibrillation. *Protein Sci.* **2006**, *15* (12), 2795–2804.
- (49) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* **2012**, *51* (11), 2224–2231.
- (50) Krishna, M. M. G.; Hoang, L.; Lin, Y.; Englander, S. W. Hydrogen Exchange Methods to Study Protein Folding. *Methods* **2004**, *34* (1), 51–64.
- (51) Jensen, M. R.; Markwick, P. R. L.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Bernadó, P.; Blackledge, M. Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure* **2009**, *17* (9), 1169–1185.
- (52) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132* (24), 8407–8418.
- (53) Kay, L. E.; Torchia, D. A.; Bax, A. Backbone Dynamics of Proteins as Studied by ¹⁵N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **1989**, *28* (23), 8972–8979.
- (54) Gairí, M.; Dyachenko, A.; González, M. T.; Feliz, M.; Pons, M.; Giralt, E. An Optimized Method for (¹⁵N R(1) Relaxation Rate Measurements in Non-Deuterated Proteins. *J. Biomol. NMR* **2015**, *62* (2), 209–220.
- (55) Lepre, C. A.; Moore, J. M.; Peng, J. W. Theory and Applications of NMR-Based Screening in Pharmaceutical Research. *Chem. Rev.* **2004**, *104* (8), 3641–3676.
- (56) Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous Determination of Protein Structure and Dynamics. *Nature* **2005**, *433* (7022), 128–132.
- (57) Claessens, F.; Denayer, S.; Van Tilborgh, N.; Kerkhofs, S.; Helsen, C.; Haelens, A. Diverse Roles of Androgen Receptor (AR) Domains in AR-Mediated Signaling. *Nucl. Recept. Signal.* **2008**, *6*, 1–13.
- (58) Jenster, G. Identification of Two Transcription Activation Units in the N-Terminal Domain of the Human Androgen Receptor. *J. Biol. Chem.* **1995**, *270* (13), 7341–7346.
- (59) Dehm, S. M.; Regan, K. M.; Schmidt, L. J.; Tindall, D. J. Selective Role of an NH₂-Terminal WxxLF Motif for Aberrant Androgen Receptor Activation in Androgen Depletion-Independent Prostate Cancer Cells. *Cancer Res.* **2007**, *67* (20), 10067–10077.
- (60) van de Wijngaart, D. J.; Dubbink, H. J.; van Royen, M. E.; Trapman, J.; Jenster, G. Androgen Receptor Coregulators: Recruitment via the Coactivator Binding Groove. *Mol. Cell. Endocrinol.* **2012**, *352*, 57–69.
- (61) 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.; McEwan, I. J.; Riera, A.; Salvatella, X. EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. *ACS Chem. Biol.* **2016**, *11*, 2499–2505.
- (62) Xue, B.; Dunbrack, R. L.; Williams, R. W.; Dunker, A. K.; Uversky, V. N. PONDR-FIT: A Meta-Predictor of Intrinsically Disordered Amino Acids. *Biochim. Biophys. Acta* **2010**, *1804* (4), 996–1010.
- (63) Muñoz, V.; Serrano, L. Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. *Nat. Struct. Mol. Biol.* **1994**, *1*, 399–409.
- (64) De Mol, E. Structure, Dynamics and Interactions of the N-Terminal Domain of the Androgen Receptor. Ph.D. Thesis, Universitat de Barcelona - IRB Barcelona, 2014.

- (65) Antonarakis, E. S.; Lu, C.; Wang, H.; Lubber, B.; Nakazawa, M.; Roeser, J. C.; Chen, Y.; Mohammad, T. A.; Chen, Y.; Fedor, H. L.; Lotan, T. L.; Zheng, Q.; De Marzo, A. M.; Isaacs, J. T.; Isaacs, W. B.; Nadal, R.; Paller, C. J.; Denmeade, S. R.; Carducci, M. A.; Eisenberger, M. A.; Luo, J. AR-V7 and Resistance to Enzalutamide and Abiraterone in Prostate Cancer. *N. Engl. J. Med.* **2014**, *371* (11), 1028–1038.
- (66) Sadar, M. D. Discovery of Drugs That Directly Target the Intrinsically Disordered Region of the Androgen Receptor. *Expert Opin. Drug Discov.* **2020**, *15* (5), 551–560.
- (67) Quayle, S. N.; Mawji, N. R.; Wang, J.; Sadar, M. D. Androgen Receptor Decoy Molecules Block the Growth of Prostate Cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (4), 1331–1336.
- (68) Banuelos, C. a.; Lal, A.; Tien, A. H.; Shah, N.; Yang, Y. C.; Mawji, N. R.; Meimetis, L. G.; Park, J.; Kunzhong, J.; Andersen, R. J.; Sadar, M. D. Characterization of Niphatenones That Inhibit Androgen Receptor N-Terminal Domain. *PLoS One* **2014**, *9* (9), 1–10.
- (69) Sadar, M. D.; Williams, D. E.; Mawji, N. R.; Patrick, B. O.; Wikanta, T.; Chasanah, E.; Irianto, H. E.; Soest, R. V.; Andersen, R. J. Sintokamides A to E, Chlorinated Peptides from the Sponge. *Org. Lett.* **2008**, *10*, 4947–4950.
- (70) Banuelos, C. A.; Tavakoli, I.; Tien, A. H.; Caley, D. P.; Mawji, N. R.; Li, Z.; Wang, J.; Yang, Y.-C.; Imamura, Y.; Yan, L.; Wen, J. G.; Andersen, R. J.; Sadar, M. D. Sintokamide A Is a Novel Antagonist of Androgen Receptor That Uniquely Binds Activation Function-1 in Its Amino-Terminal Domain. *J. Biol. Chem.* **2016**, *291*, 22231–22243.
- (71) 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.; Sadar, M. D. Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor. *Cancer Cell* **2010**, *17* (6), 535–546.
- (72) Myung, J. K.; Banuelos, 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.; Sadar, M. D. An Androgen Receptor N-Terminal Domain Antagonist for Treating Prostate Cancer. *J. Clin. Invest.* **2013**, *123* (7), 2948–2960.
- (73) Imamura, Y.; Tien, A. H.; Pan, J.; Leung, J. K.; Banuelos, C. A.; Jian, K.; Wang, J.; Mawji, N. R.; Fernandez, J. G.; Lin, K.-S.; Andersen, R. J.; Sadar, M. D. An Imaging Agent to Detect Androgen Receptor and Its Active Splice Variants in Prostate Cancer. *JCI Insight* **2016**, *1* (11), 1–15.
- (74) Obst, J. K.; Wang, J.; Jian, K.; Williams, D. E.; Tien, A. H.; Mawji, N.; Tam, T.; Yang, Y. C.; Andersen, R. J.; Chi, K. N.; Montgomery, B.; Sadar, M. D. Revealing Metabolic Liabilities of Ralaniten To Enhance Novel Androgen Receptor Targeted Therapies. *ACS Pharmacol. Transl. Sci.* **2019**, *2* (6), 453–467.
- (75) Le Moigne, R.; Mawji, N. R.; Adriana Banuelos, C.; Wang, J.; Jian, K.; Andersen, R. J.; Sadar, M. D.; Zhou, H.-J.; Virsik, P. A New Generation of N-Terminal Domain Androgen Receptor Inhibitors, with Improved Pharmaceutical Properties, in Castration Resistant Prostate Cancer Models. Presented at the AACR, March 19, 2019. Abstract Number:1292.
- (76) Luo, P.; Baldwin, R. L. Mechanism of Helix Induction by Trifluoroethanol: A Framework for Extrapolating the Helix-Forming Properties of Peptides from Trifluoroethanol/water Mixtures back to Water. *Biochemistry* **1997**, *36* (27), 8413–8421.
- (77) De Mol, E.; Szulc, E.; Di Sanza, C.; Martínez-Cristóbal, P.; Bertoncini, C. W.; Fenwick, R. B.; Frigolé-Vivas, M.; Masín, M.; Hunter, I.; Buzón, V.; Brun-Heath, I.; García, J.; De Fabritiis, G.; Estébanez-Perpiñá, E.; McEwan, I. J.; Nebreda, Á. R.; Salvatella, X. Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators. *Structure* **2018**, *26*, 1–8.
- (78) Szulc, E. M. Structural Insights into “Acid Blobs and Negative Noodles” - the Androgen Receptor as a Case Study. Ph.D. Thesis, Universitat de Barcelona - IRB Barcelona, 2019.

- (79) Palmer, A. G.; Massi, F. Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy. *Chem. Rev.* **2006**, *106* (5), 1700–1719.
- (80) Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q.-X.; Nixon, B. T.; Rosen, M. K. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* **2012**, *483* (7389), 336–340.
- (81) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-like Domains. *Science* **2020**, *367* (6478), 694–699.
- (82) Eftekharzadeh, B.; Banduseela, V. C.; Chiesa, G.; Martínez-Cristóbal, P.; Rauch, J. N.; Nath, S. R.; Schwarz, D. M. C.; Shao, H.; Marin-Argany, M.; Di Sanza, C.; Giorgetti, E.; Yu, Z.; Pierattelli, R.; Felli, I. C.; Brun-Heath, I.; García, J.; Nebreda, Á. R.; Gestwicki, J. E.; Lieberman, A. P.; Salvatella, X. Hsp70 and Hsp40 Inhibit an Inter-Domain Interaction Necessary for Transcriptional Activity in the Androgen Receptor. *Nat. Commun.* **2019**, *10* (3562), 1–14.
- (83) Davies, P.; Watt, K.; Kelly, S. M.; Clark, C.; Price, N. C.; McEwan, I. J. Consequences of Poly-Glutamine Repeat Length for the Conformation and Folding of the Androgen Receptor Amino-Terminal Domain. *J. Mol. Endocrinol.* **2008**, *41* (5), 301–314.
- (84) Eftekharzadeh, B.; Piai, A.; Chiesa, G.; Mungianu, D.; García, J.; Pierattelli, R.; Felli, I. C.; Salvatella, X. Sequence Context Influences the Structure and Aggregation Behavior of a PolyQ Tract. *Biophys. J.* **2016**, *110* (11), 2361–2366.
- (85) Antonarakis, E. S.; Chandhasin, C.; Osbourne, E.; Luo, J.; Sadar, M. D.; Perabo, F. Targeting the N-Terminal Domain of the Androgen Receptor: A New Approach for the Treatment of Advanced Prostate Cancer. *The Oncologist* **2016**, *21* (1), 1427–1435.

3. EPI-001 AS A MODULATOR OF ANDROGEN RECEPTOR CONDENSATION

3.1. Background

3.1.1. Intrinsically disordered proteins are biopolymers which can undergo condensation

As described in section 1.3, intrinsically disordered proteins (IDPs) play a central role in liquid-liquid phase separation (LLPS) as by definition are multivalent proteins that can interact simultaneously with other molecules and themselves¹.

IDPs are heteropolymers and, therefore, their LLPS propensity is sequence and composition dependent². The composition also determines the type of transition an IDP will undergo: Lower Critical Solution Temperature (LCST) or Upper Critical Solution Temperature (UCST) (**Fig 3.1.1.1a**). LCST occurs upon temperature increase and is entropically favoured as it is driven by the release of water molecules from the hydration shell surrounding hydrophobic moieties. In contrast, UCST occurs upon temperature decrease and is enthalpically favoured. It is driven by molecular interactions, such as charge-charge, cation- π , dipole-dipole, hydrogen bonding and π - π interactions³ (**Fig. 3.1.1.1b**). For example, LLPS is driven by cation- π interactions (Arg-Tyr) in the Fused in Sarcoma (FUS) protein⁴ and by negative and positive charge patches within the Dead-box helicase 4 (DDX4) protein⁵. Whether an IDP will undergo UCST or LCST depends on the balance between hydrophobic and polar/charged amino acids² (**Fig. 3.1.1.1a**). Nevertheless, it is still not possible to predict the phase behaviour of a protein from its primary sequence.

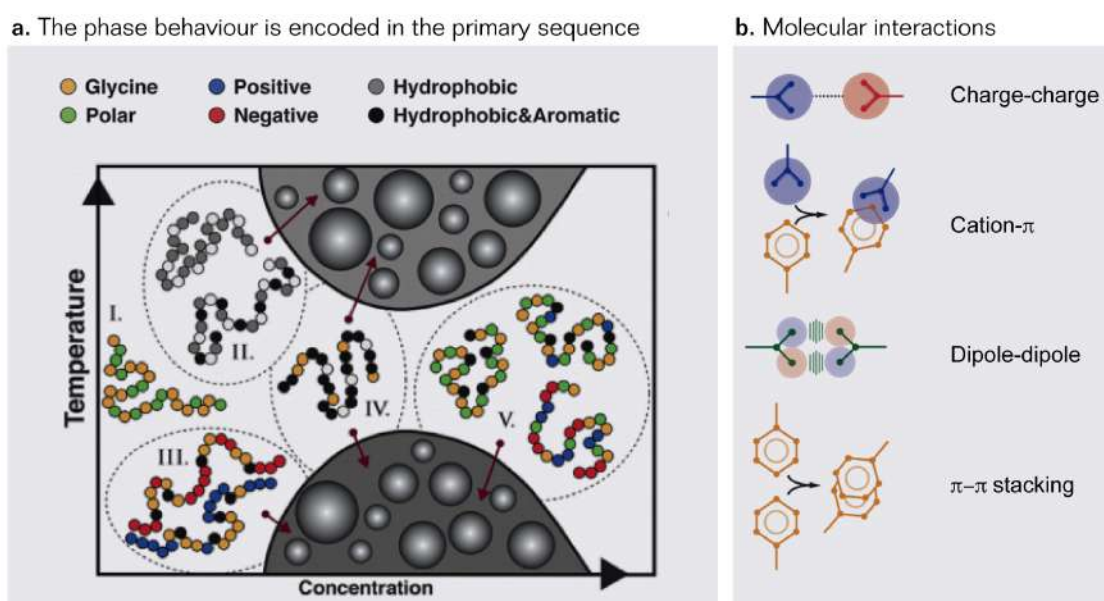


Fig. 3.1.1.1. a) The amino acid composition determines the UCST and LCST behaviour of proteins undergoing LLPS. (I) Purely polar IDPs are likely soluble over a wide range of temperatures. (II) IDPs with LCST behaviour are predominantly hydrophobic. (III) IDPs that undergo UCST are mixtures of oppositely charged amino acids or (V) IDPs with aromatic amino acids dispersed into a polar or charged sequence. (IV) If the hydrophobic amino acids are mainly aromatic can result in a LCST or UCST behaviour. Adapted from Martin and Mittag 2018 ². **b)** Molecular interactions that drive phase separation. Adapted from Brangwynne et al. 2015 ³.

Once the system has demixed (density transition), the dense phase can adopt various material states: condensates can harden into a percolated state (viscosity transition) or even turn into solids (connectivity transition) (**Fig. 3.1.1.2**) ^{6,7}. This process highly depends on the patterning of “stickers” and “spacers” throughout the sequence. For example, the spacer composition can determine whether the same multivalent domains will undergo condensation or not and the liquid or gel propensity of the resulting droplets ⁸. In particular for the FUS protein, Wang et al. determined that glycine residues in the spacer regions enhance the “liquidity” of the system whereas glutamine and serine promoted the hardening of FUS droplets ⁴. More recently, Martin et al. also described that the number and patterning of aromatic stickers influenced the gelation of the droplets: patchy patterning of aromatic residues led to aggregation whereas uniform distribution of the aromatic residues led to soluble states ⁹. Interestingly, the aromatic residues present in their studied protein were more uniformly spaced than 99.99% of all the randomly generated sequences, suggesting a clear bias in the patterning of aromatic residues in physiological IDPs to prevent aggregation.

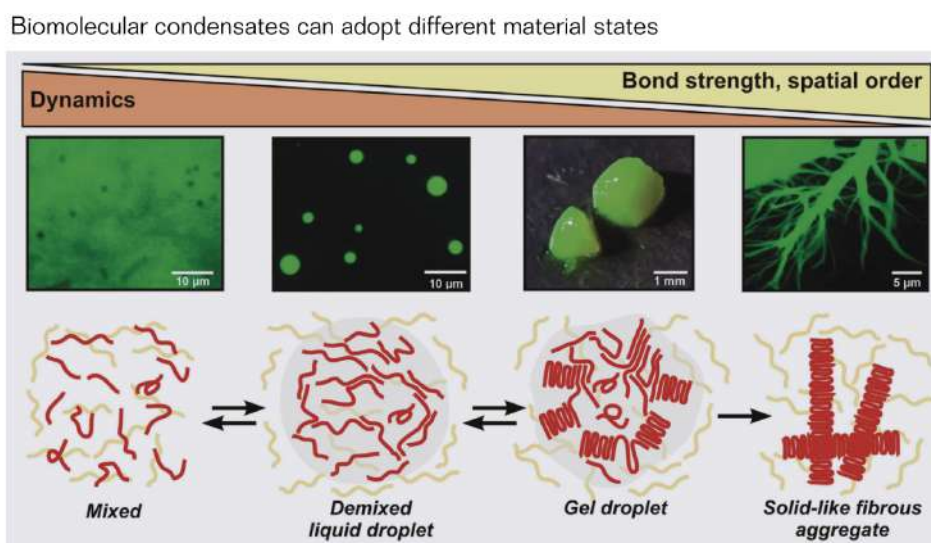


Fig. 3.1.1.2. Purified FUS protein in various material states: from liquid to fibrillar aggregates. Adapted from Alberti and Hyman 2016 ⁶.

One important question is which structure IDPs adopt in the dense phase. There are opposite views on this topic. One suggests that secondary structure elements within the IDP drive the LLPS process. In contrast, others argue that IDPs remain largely disordered within the dense liquid phase ¹⁰. On the one hand, there is evidence that transient reversible imperfect cross- β structures (named LARKS, low-complexity aromatic-rich kinked segment) form networks in FUS or hnRNPA1 ^{11,12}. However, it has not been proven yet whether these cross- β fibrils promote phase separation or are the consequence of a liquid-to-solid transition once the condensates are formed. α -helices have also

been implicated in promoting phase separation. Fawzy and co-workers determined that a helical stretch of 20 amino acids promoted phase separation in the TAR DNA-binding protein 43 protein (TDP-43)¹³. Bolognesi et al. came to the same conclusion using deep-mutagenesis studies¹⁴. Moreover, they showed how mutations that increased TDP-43 aggregation were less toxic than those increasing the “liquid-like” character of the assemblies formed. On the other hand, structures of several IDPs in the dense phase have been determined by solution Nuclear Magnetic Resonance (NMR) to remain largely disordered and dynamic for the same FUS protein¹⁵ and others, like DDX4¹⁶. The environment in the condensates is probably different from that of the bulk phase. This, together with the fact that IDPs ensembles are highly sensitive to changes in the environment, suggests that the overall ensemble of the IDP will vary in and out of the condensate. However, whether this shift is responsible for the LLPS of a protein is yet to be determined¹⁰.

3.1.2. Biomolecular condensation in physiology and disease

3.1.2.1. Function of biomolecular condensates

A collective behaviour for proteins was proposed many years ago but its significance has become evident after the introduction of LLPS in cell biology. One can envisage several mechanisms that would benefit from a collective behaviour. However, not many examples have shown that the phase-separation ability of a protein is functionally relevant. Some possible functions of condensates are listed below^{7,17–19} (**Fig. 3.1.2.[1-7]**):

1. Accelerating or repressing biochemical reactions

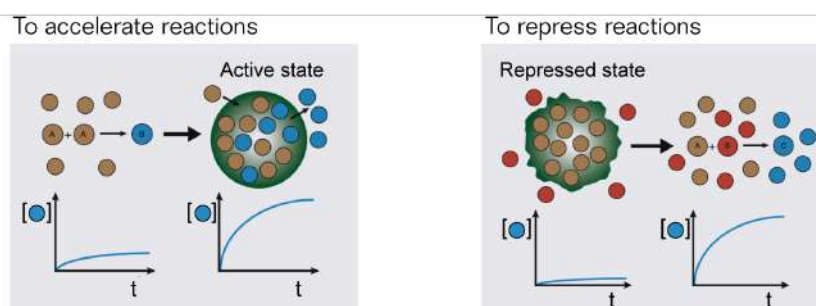


Fig. 3.1.2.1. Possible function of condensates: to accelerate or repress reactions. Adapted from Alberti 2017¹⁷.

1.a. Accelerating reactions

The reaction efficiency can be accelerated by partitioning the substrate of the reaction to the condensate that contains the required enzyme. The reaction can also be accelerated if the microenvironment of the dense phase is optimal for the reaction to occur.

3. EPI-001 as a modulator of androgen receptor condensation

–

For example, phase separation of miRISC proteins concentrated the target mRNA accelerating its deadenylation²⁰. Another example is the formation, upon stress, of nuclear proteolytic condensates, which promoted the proteasomal degradation of not properly assembled ribosomal proteins²¹. A third example showed how LLPS increased the dwell time of a protein complex between N-WASP and Arp2/3 which led to increased actin assembly²².

1.b. Repressing reactions

The sequestration of the substrate of the reaction in a condensed state or the evolution of the material properties into more solid-like states can slow down diffusion and serve as long-term reservoirs.

For example, the yeast Sup35 translation termination factor formed liquid condensates upon acidification (pH induced LLPS). These liquid condensates evolved into gels which sequestered Sup35. When the pH was restored, the gelified condensates dissolved and the protected Sup35 was available to restart translation²³.

2. Sensing and decision-making

LLPS can integrate small changes in the solution conditions resulting in fast cooperative responses. This can be used to sense environmental changes or to control the formation of condensates.

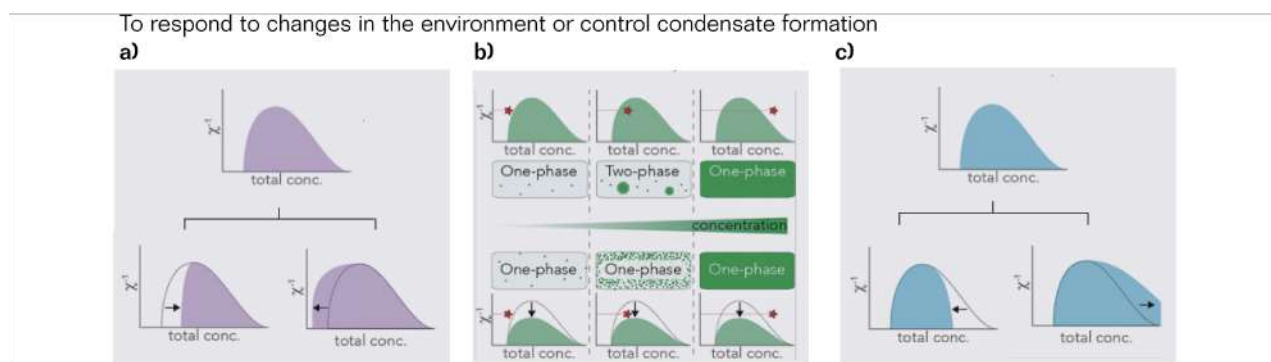


Fig. 3.1.2.2. Possible function of condensates: to respond to changes in the environment or control condensate formation. Adapted from Holehouse and Pappu 2018¹⁹.

The response can come in three no mutually-excluded ways:

2.a. Increasing or decreasing the C_{sat}

C_{sat} can be modulated by shifting the left-arm of the phase diagram. For example, Zhang et al. showed that the RNA-binding protein Whi3 C_{sat} was modulated by RNA²⁴.

2.b. Increasing or decreasing the critical point

The critical point can be modulated to determine whether a solution in a particular condition is demixed or not. This was the case for the post-transcriptional modifications in FUS²⁵ or CPE-binding protein 4 (CPEB4)²⁶.

3. EPI-001 as a modulator of androgen receptor condensation

-

2.c. Increasing or decreasing the concentration of the dense phase

The dense phase concentration can be modulated by shifting the right-arm of the phase diagram. For example, the presence of RNA shifted the high-concentration arm of the phase diagram of the DEAD-box helicase LAF-1 while the lower-concentration arm and the critical point remained unaffected²⁷.

3. Buffering the concentration of biomolecules

Demixing of a protein solution occurs when its concentration is raised above its C_{sat} . A further increase in the protein concentration results in a larger volume fraction of the dense phase but the concentrations of the dense and dilute phase remain constant¹⁸.

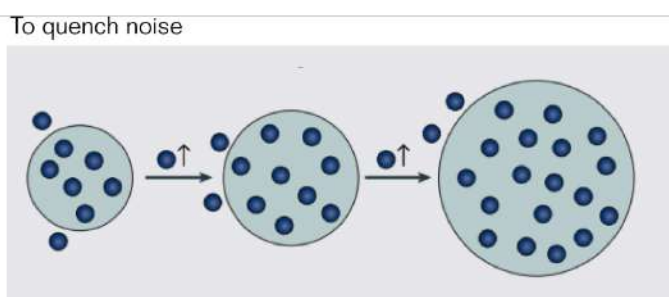


Fig. 3.1.2.3. Possible function of condensates: to quench noise. Adapted from Banani et al. 2017¹⁸.

This mechanism was proposed to be used by cells to quench the noise coming from variations in the cellular concentrations of biomolecules²⁸.

4. Localising molecules

To target molecules to pre-existing condensates
or to sequester them

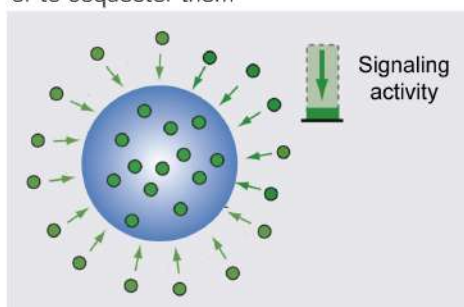


Fig. 3.1.2.4. Possible function of condensates: to target molecules to pre-existing condensates or to sequester them. Adapted from Shin and Brangwynne 2017²⁹.

4.a. To target molecules to specific pre-existing condensates

Condensation may be used to localise proteins in time and space according to the need. This is the case for the speckle-type POZ protein (SPOP). The substrates of SPOP triggered the protein phase separation and co-localisation into the nuclear assemblies³⁰.

3. EPI-001 as a modulator of androgen receptor condensation

–

4.b. To sequester them from freely diffusing

Condensation can be used to sequester factors not required for current cellular needs, preventing off-target effects. For example, upon stress, rapamycin complex 1 (TORC1) was sequestered into stress granules repressing TORC1 signalling³¹.

5. Inducing mechanical forces

It was proposed that the maturation of condensates into viscoelastic assemblies could deform the neighboring cellular structures, such as membranes³².

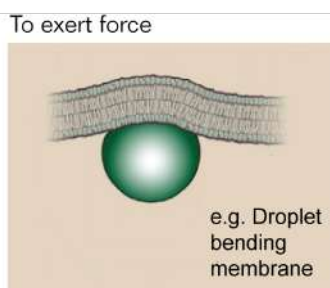


Fig. 3.1.2.5. Possible function of condensates: to exert mechanical forces. Adapted from Alberti et al. 2019⁷.

6. Filtering molecules

The dense phase consists of a highly interconnected mesh of biopolymers. The number and dynamics of these cross-links determine the pore size within a condensate that can be used as a mechanical filter to include or exclude certain macromolecules from a condensate.

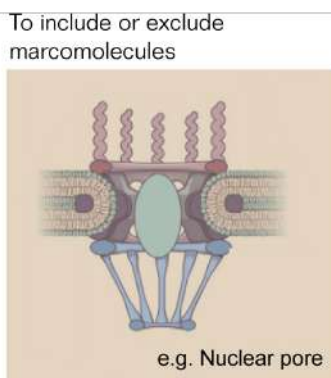


Fig. 3.1.2.6. Possible function of condensates: to filter macromolecules. Adapted from Alberti et al. 2019⁷.

For example, it was demonstrated that condensation of the disordered FG domains of the nuclear pore complex allowed selective transport across the nuclear pore³³. Another example showed how different types of DNA were predominantly excluded or absorbed in DDX4 condensates depending on their length³⁴. In a last example, Wei et al. identified LAF-1 droplets to have a characteristic mesh size within 3–8 nm, being permeable to small molecules and proteins up to a certain size and molecular weight²⁷.

7. Integrating and amplifying signals

Condensates can integrate and amplify signals. In this sense, the input can be a single component,

whereas the output signal can be determined by the collective properties of the condensate.

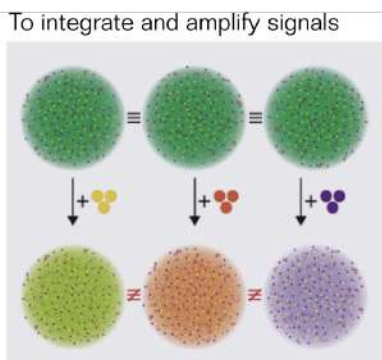


Fig. 3.1.2.7. Possible function of condensates: to integrate and amplify signals. Adapted from Holehouse and Pappu 2018 ¹⁹.

For example, within the transcription field, condensation of super-enhancers (SE) was put forward to explain all the features related to SE activity including the sensitivity of SE to perturbation and the ability to produce simultaneous activation at multiple genes among others ³⁵.

3.1.2.2. Regulation of biomolecular condensates

Biomolecular condensates participate in many cellular functions and they need to be tightly controlled. Condensate function is regulated by several means ³⁶:

1. Compositional control

One form of regulating condensate function is to control the composition of the condensate. The phase boundary is not a physical barrier but defines a sharp change in the environment properties. The initial condensate environment is defined by the scaffold and its hydration shell, whereas the cytoplasmic environment is defined by the cytoplasmic proteins. The relative solubility of clients between both environments (partition coefficient, P_c) regulates the composition within the condensate. The P_c is in turn governed by the relative scaffold/client stoichiometries and the strength of the scaffold-client interactions ^{22,37}. Therefore, changes in the physico-chemical properties of the client that affect its solubility in the condensate environment and/or changes in the scaffold-client affinities fine-tune the composition of condensates.

2. Formation/dissolution control

Condensation is a cooperative “all-or-nothing” process, hence changes in the physico-chemical properties of the cytoplasmic environment or the scaffolding molecule can have rapid effects on the formation and dissolution of condensates throughout the cell:

- Changes in the cytoplasm (protein concentration, temperature, molecular crowders or pH) can trigger the formation and dissolution of condensates by shifting the whole phase

–

diagram to a region where phase separation is promoted or inhibited. For example, acidification of the cytoplasm caused Sup35 to condensate ²³.

- Also, altering the physico-chemical properties of the scaffold protein by post-transcriptional modifications can modulate the formation of condensates. For example, phosphorylation of CPEB4 triggered condensate dissolution and protein activation ²⁶. Phosphorylation also caused the disassembly of FUS condensates ³⁸.
- Moreover, nucleation can also control condensate formation. Fully mixed solutions containing a high concentration of scaffold protein can be metastable and only condensate via a nucleation-dependent mechanism. For higher protein concentrations, in the region below the spinodal curve of the phase diagram, demixing occurs spontaneously ³⁹.
- Finally, formation and dissolution of condensates can be altered by small molecule solutes in the cell or drugs. For example high concentration of ATP was shown to inhibit the phase separation of several IDPs ⁴⁰. The modulation of condensation by small molecules is further discussed in section 3.1.3.

3. Identity control

Another layer of regulation needs to manage the fusion or separation of coexisting biomolecular condensates. The fact that two scaffold molecules are miscible or immiscible depends on the relative energetic benefit of cross- or self-interaction and also on the biophysical properties of the phases, particularly the droplet surface tension. Miscibility not only governs the coexistence of condensates but also the nesting of different condensates. One such example is the subcompartmentalization of the nucleoli ⁴¹.

3.1.2.3. Dysregulation of biomolecular condensates

Dysregulation in the assembly and disassembly of biomolecular condensates is linked to various diseases including neurodegeneration ^{6,42,43}, cancer ³⁰ and infectious diseases ^{44,45}. Theoretically, there are several mechanisms by which condensates can lead to the gain- or loss-of-function disease phenotypes (**Fig. 3.1.2.8**) ⁴⁵.

3. EPI-001 as a modulator of androgen receptor condensation

Theoretical mechanisms by which condensates lead to disease

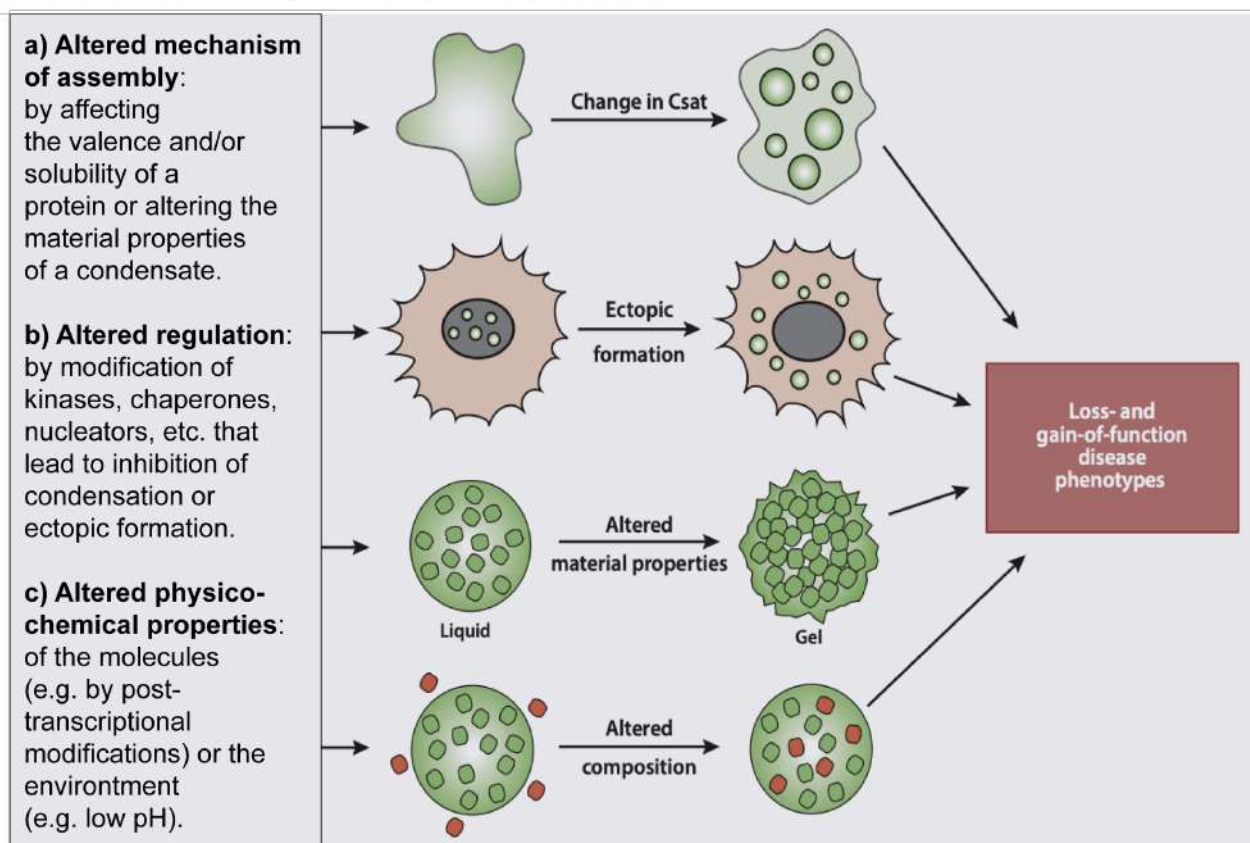


Fig. 3.1.2.8. Theoretical mechanisms by which condensates can lead to a gain- or loss-of-function disease phenotypes. Adapted from Alberti and Dormann 2019⁴⁵.

1. Condensates and neurodegeneration

A common feature in all neurodegenerative disorders is the presence of cytosolic or nuclear protein aggregates. Aberrant phase separation and irreversible liquid-to-solid transitions offer a unifying mechanism for the formation of these pathological aggregates^{43,45}.

A considerable amount of evidence has been accumulated in the recent years linking amyotrophic lateral sclerosis (ALS) and aberrant phase transitions of RNA-binding proteins such as FUS, TDP-43, HNRNPA1, HNRNPA2, and TIA1 in stress granules (SGs)^{42,46–50}. These proteins form liquid condensates that transition to a more solid-like state. Importantly, this transition is accelerated by ALS-associated mutations. Moreover, aberrant PTMs on some of these proteins are also associated with ALS, such as demethylation in FUS^{4,51,52}. As mentioned in section 1.3, FUS LLPS is driven by Arg-Tyr interactions⁴. Therefore, demethylation of FUS arginine residues increases the valency of arginine “stickers” promoting FUS LLPS and reduces the dynamism of the resulting condensates.

Other neurodegenerative diseases have also been linked to aberrant phase transitions. For example, Tau mutations and abnormal PTMs promoted Tau condensed state and its subsequent hardening and aggregation in Alzheimer’s disease models^{53,54}. In another example, polyQ-expanded huntingtin N-

terminal fragment formed liquid condensates that evolved into solid-like assemblies that resembled the pathological aggregates in Huntington's disease⁵⁵.

2. Condensates and cancer

Only one study has linked phase separation and cancer so far⁴⁵. SPOP is a substrate adaptor of the cullin 3 RING ubiquitin ligase (CRL3)³⁰. SPOP acts as a tumour suppressor protein as it promotes ubiquitination and subsequent proteasomal degradation of various proto-oncogenic proteins. Bouchard et al. demonstrated that SPOP condensation and localisation to nuclear membraneless organelles was promoted by self-association and the binding of its substrates. Cancer mutations in SPOP disrupted SPOP-substrate interactions, SPOP phase separation and localisation to membraneless organelles which led to the accumulation of oncogenic proteins.

Despite the limited evidence for the involvement of phase separation in cancer, other examples are yet to come such as from the transcription field. Transcription dysregulation is connected with cancer⁴⁵ and associated with phase separation^{56,57}. In particular, several transcription factors have been shown to phase separate⁵⁸. Thus, phase separation can provide a new angle on the understanding of the molecular mechanisms by which transcription can be dysregulated and lead to diseases such as cancer.

3.1.3. Biomolecular condensates as an opportunity to drug IDPs

Biomolecular condensates represent a unique opportunity for drug discovery^{59,60}. We have envisaged several strategies to target condensates and have classified them using the same rationale as for IDPs (section 2.1.1). In this logic, we have grouped them according to their target into 1) regulatory inhibition, 2) orthosteric inhibition, 3) allosteric inhibition and 4) colligative inhibition (**Fig 3.1.3.1**).

3. EPI-001 as a modulator of androgen receptor condensation

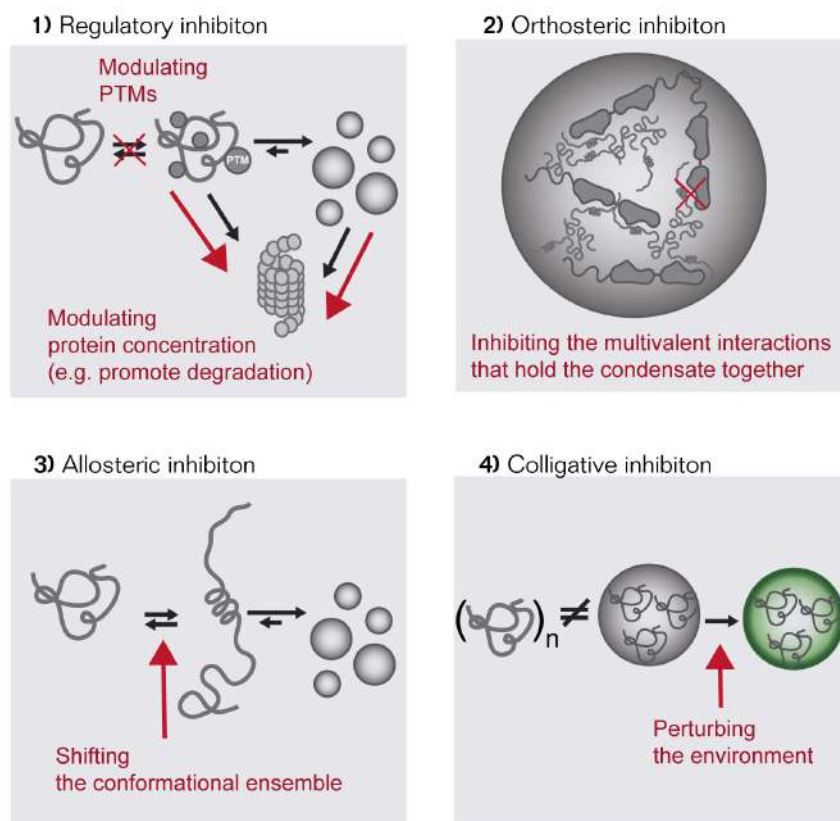


Fig. 3.1.3.1. Strategies to modulate condensates: 1) regulatory inhibition, 2) orthosteric inhibition, 3) allosteric inhibition and 4) colligative inhibition.

1. Regulatory inhibition

Similarly to IDPs, condensates are tightly regulated. Therefore, targeting the regulatory machinery is a promising therapeutic strategy to modulate condensates. Moreover, since these regulatory mechanisms are reversible, it may be possible to modulate condensation in both directions, favouring and disfavouring it. For instance:

- Inhibiting the enzymes responsible for post-translational modifications can alter the physico-chemical properties of either the scaffold or the client to promote or inhibit condensation^{26,61}.
- Stabilising and destabilising chaperones can promote the disassembly of condensates^{62,63}.
- Altering the relative composition of condensates can also disfavour condensation. The composition can be changed by degradation of one of its components, for instance using proteolysis targeting chimera (PROTACs) small molecules^{64,65}.

For example, inhibition of stress-granule regulatory mechanisms has led to the discovery of small molecule inhibitors in neurodegeneration⁶⁶ and also cancer⁶⁷.

2. Orthosteric inhibition

Biomolecular condensates are held together by multivalent interactions of the same molecule (scaffold-scaffold) or between different molecules (scaffold-client). Targeting the relevant interfaces appears as a plausible approach to modulate condensate formation and/or composition. For example, Fang et al. disclosed that compounds containing a planar moiety, such as mitoxantrone, intercalated the RNA impeding the RNA-protein interactions that promoted stress-granule (SG) formation⁶⁸.

It is important to highlight that condensates are formed by a network of interactions. In this regard it is important to either target the ones contributing the most to the stability of biomolecular condensates or target many of them at the same time, perhaps using a similar approach as the ligand-clouds strategy for IDPs (see section 2.1.1). For example, Yoshizawa et al. showed how the protein Karyopherin-b2 inhibited FUS condensation. By tightly binding to FUS nuclear localization signal and weakly binding to multiple regions across FUS, Karyopherin-b2 competed with FUS-FUS interactions that held the condensate together⁶⁹.

A final approach to inhibit the interactions that hold a condensate together could be based on the fact that scaffold-client interactions depend on the relative scaffold/client stoichiometries (concentration and valency)^{22,37}. Hence altering the stoichiometry or valency of the scaffolding macromolecules with small molecules could modulate condensate formation and composition. However, this is a speculative strategy as, to our knowledge, there is no example in the literature of a small molecule with this mechanism of action.

3. Allosteric inhibition

IDPs are allosteric proteins⁷⁰ and play a central role in biomolecular condensation¹⁸. Shifting the IDP conformational ensemble to states which are prone to demix could promote and stabilise condensation and, conversely, reaching states which are unlikely to demix could diminish and destabilise condensation.

Although there are several examples of small molecules shifting the conformational ensemble of IDPs (section 2.1.1), there is yet no example of a small molecule inhibiting condensates by shifting the conformational ensemble of the scaffolding IDP. Nevertheless, allosteric inhibition of condensation has been shown using proteins. Posey et al. demonstrated that profilin shifted the concentration of phase separation of huntingtin (Htt-NTFs) to higher values. Profilin stabilised Htt-NTFs monomeric forms that led to reduced Htt-NTF aggregation⁷¹.

4. Colligative inhibition

The fourth strategy that we propose only applies to condensates, as opposed to strategies 1 to 3 that also apply to IDPs (section 2.1.1). This strategy aims at targeting those higher-order properties that emerge from the collective behaviour of the proteins within the condensate: the colligative properties.

It is proposed that cells have the capacity to modulate condensates using biological small molecules, like metabolites³⁶. Therefore, extrinsic modulation of condensates using drugs appears to be a plausible mechanism to interfere with condensates that are in pathway to disease³⁶.

As mentioned in section 3.1.2.1 (Figure 3.1.2.2), condensates are very sensitive to small changes in solution conditions. We suspect that the direction and size of the environmental changes caused by small molecules will be similar to those caused by intrinsic factors (pH, temperature, metabolite concentration, etc.). In this sense, we suggest that an extrinsic change of the environment can affect the condensate in several ways:

- *Material properties*

On the one hand, aberrant liquid-to-solid transitions are proposed as the first-step towards aggregation of IDPs in several neurodegenerative diseases like ALS⁴² and Alzheimer's disease⁵⁴. Therefore, decreasing the viscosity and/or retarding the liquid-to-solid transition could be beneficial for these types of diseases.

On the other hand, the liquid-character of condensates is essential for diffusion-limited biochemical reactions to occur⁶. Macromolecules constituting condensates are in constant rearrangement within the condensate as well as in constant exchange with the bulk phase. In order to inhibit a certain biochemical reaction from happening, slowing down the condensate dynamics could be a good approach. In this sense, Guan et al. discovered that the estrogen receptor (ER) antagonist fulvestrant had superior anti-proliferative capacity compared to other ER antagonists as fulvestrant was able to impair ER mobility and, hence, lowered the accessibility of ER to chromatin⁷². In another work, Boija et al. described how phase separation of transcription factors, like ER, is linked to gene activation. We speculate whether fulvestrant's mode of action is to reduce the mobility of ER condensates.

- *Nucleation*

Many macromolecules can exist in a metastable supersaturated state and will only demix if nucleated. In this sense, a change in the environment can either promote or prevent nucleation, thereby promoting or preventing condensation.

3. EPI-001 as a modulator of androgen receptor condensation

–

- *The phase diagram: C_{sat} , critical point, dense phase concentration*

Finally, another possible consequence of changing the environment is the re-shaping of the phase diagram so that the protein of interest lies within the one-state or the two-state region without changing its concentration.

One aspect to take into account is that, as mentioned, not all condensates are of the same nature. The environment inside the condensate depends on the type of weak multivalent interactions that hold the condensate together². At the same time, partition of drugs inside a condensate will be influenced by the condensate's environment. In order for a small molecule to partition into the condensate needs to be able to mimic those interactions that hold the condensate together. For example:

- Condensates formed by hydrophobic interactions are, by definition, depleted in water. In this case, demixing is driven by a gain in entropy due to the release of water molecules from the hydration shell of hydrophobic residue side chains.

Water-depleted condensates can be physiologically modulated by high concentrations of surfactants (e.g. bile acids) and hydrotropes (e.g. ATP⁴⁰). As small molecule drugs are relatively hydrophobic molecules, we envision that it will be feasible for them to partition in hydrophobic condensates.

- Condensates formed by polar sequences (embedded with hydrophobic and aromatic residues) contain higher amounts of water molecules than those formed exclusively by hydrophobic interactions. For example, it has been estimated that FUS condensates contain *ca* 65% of water molecules¹⁵. Moreover, in this type of condensates, the free energy of demixing is governed by the enthalpic contribution. Given these two premises, we envisage that small molecules that partition into these condensates should be less hydrophobic and, if they bind to a protein, their binding needs to be of higher affinity to compete with the condensate-forming interactions.

In this category we also include condensates that are held together by hydrogen bonding interactions. In this case chaotropes, which reduce the effective strength of hydrogen bonding, were proposed to prevent condensate formation³⁶. Biological chaotropes include urea and alcohols.

- Condensates formed by charge-charge interactions will be susceptible to ions and charged small molecules. Ions may promote condensation by shielding charge-charge repulsion or formation of salt bridges or may prevent it by reducing the number of “effective” charge patches.

For example, the polycationic metabolites spermine and spermidine interact with nucleic acids to allow condensation⁷³. We envision that directing small molecules to this type of condensates will be more challenging. Compared to the poly-ionic nature of spermine and spermidine, drug-like small molecules may not be charged enough to affect this type of

3. EPI-001 as a modulator of androgen receptor condensation

-

condensates. Noteworthy, *ca* 78% of current drugs are ionizable but predominantly contain one single acidic or basic group⁷⁴.

There are currently only two examples in the literature studying the druggability of condensates^{75,76}. In the first example, Wheeler et al. conducted a high-throughput screening in order to find drugs with the capacity to dissolve SGs containing FUS-GFP and attenuated ALS disease progression. Their lead compound was lipoamide. Lipoamide showed high selectivity for SG over other membraneless organelles. Wheeler et al. suggested that lipoamide affected the physical chemistry of stress granules making the environment more “dissolving”, although they could not categorically identify lipoamide’s cellular target. Lipoamide showed promising results in motor neurons and in relevant *in vivo* models of ALS disease.

From a drug discovery point of view, targeting the colligative properties of condensates is a new mechanism of action and perhaps drugs affecting condensates will not always follow the pre-established drug discovery rules. In this sense, it is very interesting to point out some findings that Wheeler et al. encountered and could not be unambiguously addressed as they candidly state in the preprint⁷⁵. First, they showed divergent effects of lipoamide on the droplets *in vitro* and in cells. *In vitro*, lipoamide caused the formation of bigger and more liquid droplets and, in cells, it caused the dissolution of SG. Also, they found that the concentration of lipoamide in the cell was much larger than the one they achieved *in vitro*. Another fact was that (*R*)-lipoic acid and (*S*)-lipoic acid had comparable EC₅₀, which is surprising as in structure-based drug discovery usually one compound of the enantiomeric pair exhibits larger biological activity than the other one.

In the second example, Klein et al. studied whether approved drugs could selectively partition into specific nuclear condensates and whether the selective partitioning influenced the pharmacodynamic properties of the drugs⁷⁶. First, they determined that several antineoplastic drugs preferentially partitioned *in vitro* into the same nuclear condensates as their protein target. Then, they focused on MED1 condensates. They evaluated a library of 81 small molecules and showed that compounds that contained aromatic rings preferentially partitioned into MED1 condensates. Interestingly, MED1 IDR is enriched in aromatic amino acids in comparison to the other proteins they studied. Finally, to gain insight into the influence of condensates to the efficacy of the drugs tested, they demonstrated that DNA-platination by Cisplatin was more efficient at DNA sites where MED1 condensates were present.

This work highlights the importance of taking biomolecular condensates into account in the drug discovery process. Klein et al. suggest that biomolecular condensates might help to increase the efficacy of drugs by concentrating them into the condensates that bear their target and also might help to explain resistance to drugs⁷⁶. Moreover, they also describe new assays to explore the partitioning of drugs into several nuclear condensates based on fluorescent versions of the drugs. We believe that new assays such as the ones described by Klein et al. will be essential to further investigate the role of condensates in drug pharmacodynamics.

In summary, biomolecular condensates emerge as a new opportunity to drug currently undruggable targets and treat unmet diseases⁷⁵ as well as a new layer of regulation that can influence drug

pharmacodynamics and target engagement⁷⁶. Recently, two companies have been created to target biomolecular condensation: Dewpoint Therapeutics and FAZE Therapeutics, focusing on cardiovascular diseases and neurodegeneration, respectively. Also, two publications have highlighted this topic in perspective articles^{59,60}. These articles lay out a series of questions that arise when thinking of drugging condensates: “Will small molecules with this mechanism of action meet the established drug bioavailability rules? Will modulation of condensates lead to an efficacious treatments? How will selectivity for one specific condensate be achieved? Will small molecules traffic in and out these structures, and will this impact exposure?” We believe that the answer to all of these questions will come along with more studies and examples of drugs targeting condensates.

3.1.4. Biomolecular condensates and transcription

Several recent pieces of evidence have proposed biomolecular condensation to be the organising mechanism necessary for productive transcription. Transcription activation requires the interplay between many different actors and many of them have been linked to condensation.

For example, this is the case for the RNA-Polymerase II (RNA-polII). Cisse et al. showed that the RNA-polII formed dynamic clusters in the nucleus of living cells⁷⁷. Moreover, several publications demonstrated that the RNA-polII condensation was driven by its disordered C-terminal domain (polII-CTD)⁷⁸⁻⁸⁰. In addition, the polII-CTD hyperphosphorylation status defined the transition from initiation condensates to elongation/splicing-factor condensates. Also, the RNA-polII cluster life-time was correlated with the mRNA output⁸¹. Finally, the Mediator coactivator complex clusters were shown to colocalise with the RNA-polII clusters in living cells⁸².

Another body of evidence indicates that many transcription factors also condensate. Boija et al. established that the disordered activation domains in transcription factors, such as the ones from OCT4, c-Myc, p53 and estrogen receptor (ER) among others, co-phase separate with the Mediator subunit MED1⁵⁸. It is particularly relevant for the work in this thesis that ER underwent condensation. ER, as well as androgen receptor, is a hormone-activated transcription factor. ER is activated by estrogens. Boija et al. revealed that estrogens enhanced ER colocalisation in MED1 droplets by increasing the ER-MED1 interaction affinity⁵⁸. Additionally, in another publication Nair et al. determined that estrogens promoted the assembly of estrogen-responsive enhancers. They contained high levels of enhancer RNA and recruited an ER α -dependent megadalton-scale protein complex, named MegaTrans. The MegaTrans enhancers were shown to be formed by LLPS⁸³.

Linking RNA-polII and transcription factors, Wei et al. demonstrated that polII-CTD clusters formed in certain genomic loci and nucleated condensation of a transcription associated factor, the TAF15. TAF15 condensation further recruited the RNA-polII, which appeared as a positive feedback loop to increase the transcriptional output⁸⁴.

Finally, super-enhancers (SEs) have also been proposed to form through phase transitions. SEs are clusters of enhancers which concentrate a particularly high density of interacting factors and drive higher levels of transcription when compared to typical enhancers. The features related to SE formation and dissolution (sharp formation and dissolution, high sensitivity to perturbation, able to activate multiple genes at the same time, etc.) led Hnisz et al. to postulate that SEs were formed by biomolecular condensation³⁵. Sabari et al. validated this perspective experimentally which, together with the work from Boija et al., led to the current model of “transcriptional condensates”^{57,58} (**Fig. 3.1.4.1**). Moreover, Shrinivas et al. demonstrated that the features of the DNA binding site are the ones driving condensation in specific loci⁸⁵. Nevertheless, the fact that these clusters are in reality phase-separated compartments has been questioned by Chong et al., who described the recruitment of RNA-polIII into foci formed by transcription factors that did not resemble condensates⁸⁶.

Proposed model for transcriptional condensates

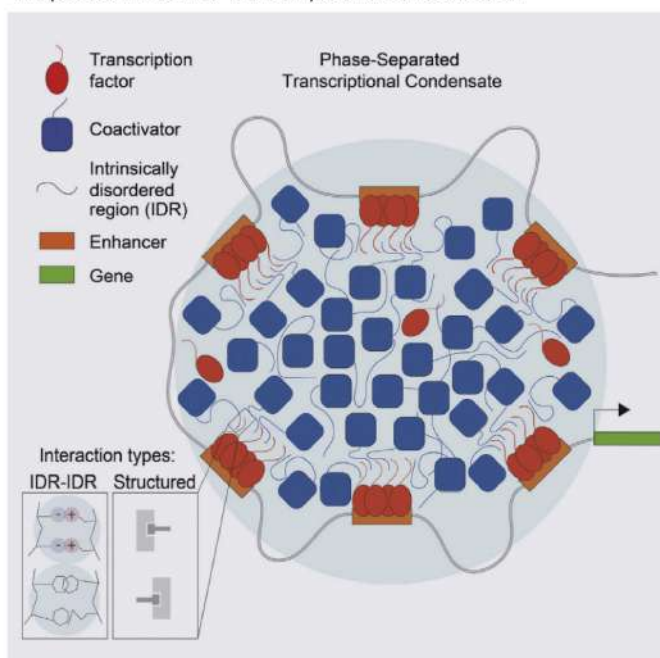


Fig. 3.1.4.1. SE condensates facilitate the compartmentalization and concentration of transcriptional components at specific genes through the condensation capability of IDRs in transcription factors, coactivators and RNA/DNA. From Boija et al. 2018⁵⁸.

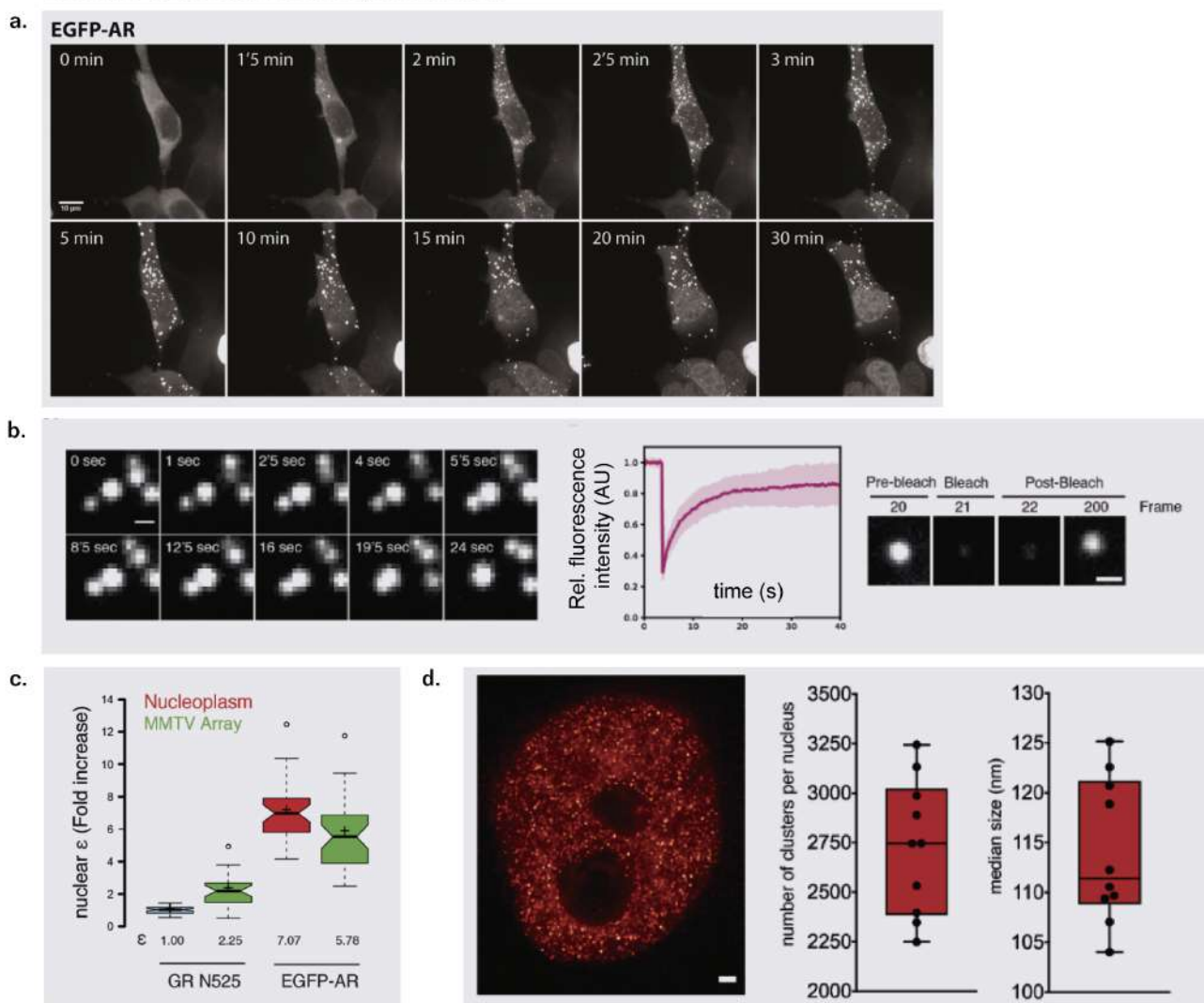
3.1.5. Androgen receptor undergoes biomolecular condensation

During the course of this thesis, in the group of Dr. Xavier Salvatella we realised that the androgen receptor N-terminal domain (AR-NTD) underwent biomolecular condensation *in vitro* and in cells. This was further investigated *in vitro* by Dr. Elzbieta Szulc as part of her PhD thesis^{30,87}, in cells by Dr. Paula Martínez-Cristóbal and using superresolution STED microscopy by myself (unpublished data). In this section we summarise some of the key findings that aided us to connect the mode of action of EPI-001 with its ability to modulate AR biomolecular condensation (section 3.2).

3. EPI-001 as a modulator of androgen receptor condensation

In cells, androgen receptor (AR) underwent condensation upon hormone (dihydrotestosterone, DHT) binding (**Fig. 3.1.5.1a**). Overexpressed eGFP-AR was kept soluble in the cytoplasm in several cell lines. Addition of 1 nM DHT nucleated eGFP-AR condensation in the cytosol. AR condensates had liquid-like properties as they were spherical, fused and relaxed into a sphere within seconds and were highly dynamic according to fluorescence-recovery after photobleaching (FRAP) experiments (**Fig. 3.1.5.1b**). 30-60 min after hormone addition, the eGFP-AR translocated to the nucleus, in agreement with the AR classical pathway⁸⁸. In lower overexpression systems, AR did not condensate in the cytosol. Nevertheless, in both scenarios it had a speckled-type distribution in the nucleus^{89,90}. The nuclear clusters were calculated to be at least hexamers by the number and brightness analysis performed in collaboration with the Hager laboratory⁹¹ (**Fig. 3.1.5.1c**). Moreover, we determined that the nuclear clusters were of 105-125 nm in diameter using ultra-resolution STED microscopy in collaboration with the Hyman and the Honigmann laboratories (**Fig. 3.1.5.1d**). The fact that i) the AR underwent condensation upon DHT addition and ii) that DHT addition caused the interaction between the ²³FQNL²⁷ and the AR ligand-binding domain (AR-LBD) suggested that the N/C interaction plays a role in the AR condensation.

AR forms condensates in cells upon activation



–

Fig. 3.1.5.1. a) Time-resolved fluorescence microscopy of eGFP-AR upon 1 nM DHT activation in PC3 cells. **b)** Left: fusion event of two droplets of eGFP-AR in PC3 cells. Right: FRAP of eGFP-AR droplets. **c)** Number and brightness calculation of the oligomerization state of eGFP-AR in the nucleoplasm (red) and upon DNA binding (green) in the cell line 3617. Glucocorticoid receptor mutant (GR N525) is used as reference. **d)** Left: representative STED image of a HeLa cell nucleus stably transfected with AR-eGFP upon 4 h 1 nM DHT exposure, scale bar = 1 μm . Right: number and size quantification of clusters from 10 independent nuclei. Adapted from Szulc Ph.D. Thesis, 2019⁸⁷.

Purified AR-NTD (AR₁₋₅₅₈) underwent LCST type phase separation *in vitro* (**Fig. 3.1.5.2a**). AR-NTD condensation is driven by hydrophobic and aromatic interactions, in agreement with the literature². This was determined by mapping the dependence of the AR-NTD condensation to the ionic strength (NaCl concentration) of the buffer (**Fig. 3.1.5.2b**). To determine which regions of the AR-NTD are involved in the oligomerisation and phase separation process, solution nuclear magnetic resonance (NMR) spectroscopy of the light phase was used (**Fig. 3.1.5.2c**). The ¹H,¹⁵N HSQC spectrum of the AR-NTD was acquired at increasing protein concentrations. The regions of the AR-NTD with a concentration-dependent decrease in signal intensity had an increased ¹⁵N transverse relaxation rate. This indicated that concentration led to more transient inter-residue long-range interactions. Particularly, the most affected regions corresponded to the following motifs: the ²³FQNLF²⁷ motif, the helical ⁵⁴LLLL⁵⁷ sequence preceding the polyQ tract, the two partially helical motifs in Tau-5 region (R1 and R2) and the highly hydrophobic and aromatic region after the polyG tract. The polyQ tract and the R3 helical region in the Tau-5 could not be detected, see section 2.1.3 and 2.1.4.

3. EPI-001 as a modulator of androgen receptor condensation

AR undergoes LCST-type phase behaviour *in vitro*

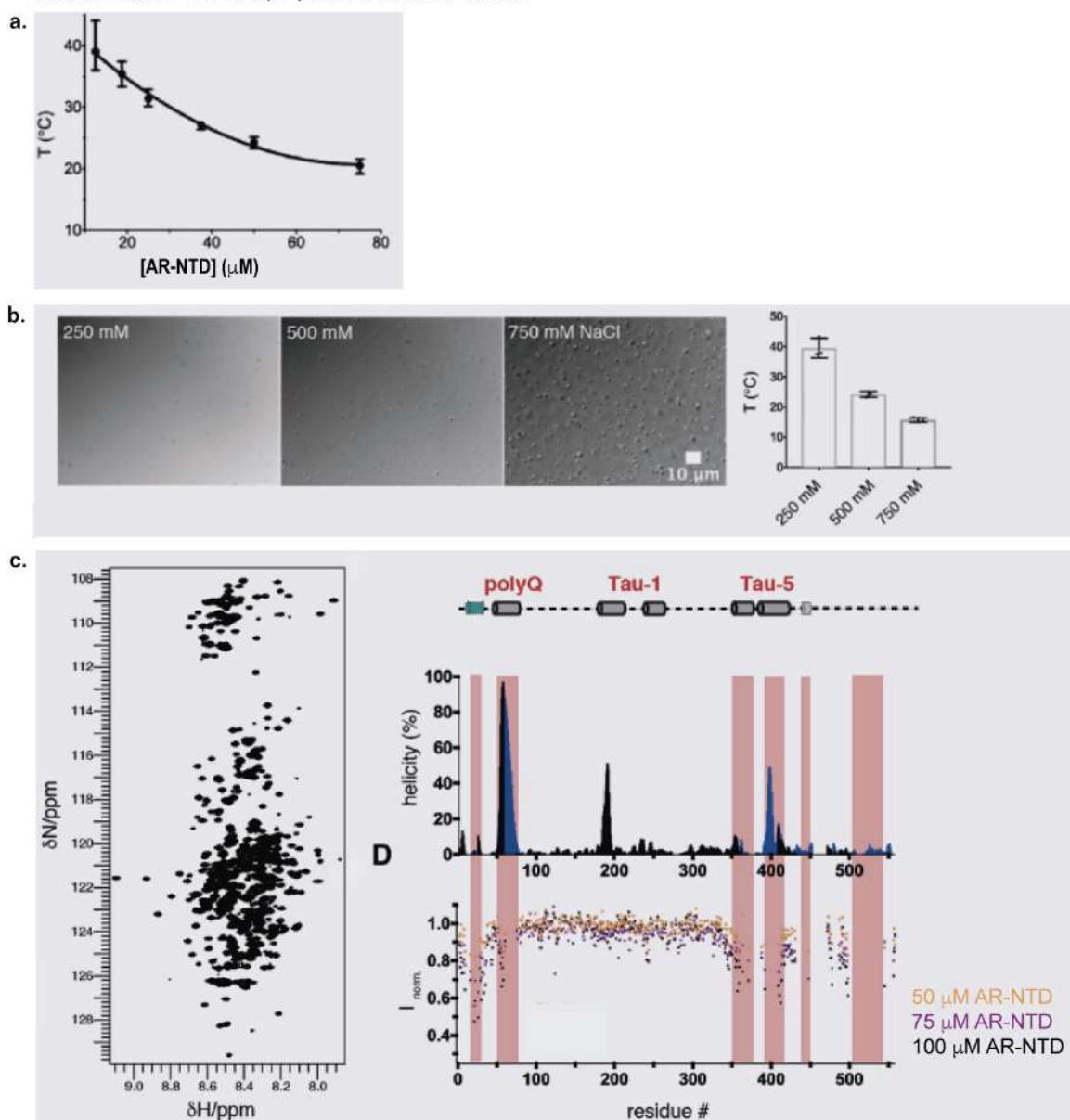


Fig. 3.1.5.2. a) Phase diagram of AR-NTD condensation at 500 mM NaCl obtained by apparent absorbance measurements of the cloud point temperature (T_c) by means of the Turbidity Shift Assay (TSA, explained in detail in section 3.2). **b)** Left: 50 μM of AR-NTD at three ionic strengths (250 mM, 500 mM and 750 mM NaCl) observed by differential interference contrast (DIC) microscopy. Right: T_c measurement by means of the TSA. **c)** Left: ^1H , ^{15}N HSQC spectrum of 25 μM ^{15}N AR-NTD. Right: plot of the residue-specific helicity using the $\delta 2\text{D}$ algorithm⁹² and loss of peak intensity in the ^1H , ^{15}N HSQC spectrum upon AR-NTD concentration (50 μM , 75 μM and 100 μM) normalised to their intensity at 25 μM . Shaded areas indicate complete loss of signal upon concentration. Adapted from Szulc Ph.D. Thesis, 2019⁸⁷.

However, we observed that the AR-NTD droplets underwent rapid maturation *in vitro* (within minutes) as fusion events were arrested and “worm-like” strings of droplets formed (**Fig. 3.1.5.3a**). Dr. Szulc demonstrated that this liquid-to-solid transition was due to the aggregation prone $^{23}\text{FQNLF}^{27}$ motif⁹³. Indeed, when $^{23}\text{FQNLF}^{27}$ fiber formation was perturbed by a Leu to Pro mutation ($^{23}\text{FQNPF}^{27}$), the AR-NTD-L26P droplets maintained a liquid-character throughout the duration of the experiments (**Fig. 3.1.5.3b**). The mutation also affected the condensation propensity

3. EPI-001 as a modulator of androgen receptor condensation

of the AR-NTD, as the AR-NTD-L26P mutant had lower propensity to phase separate (**Fig. 3.1.5.3c**).

AR-NTD maturation is prevented by the L26P mutation

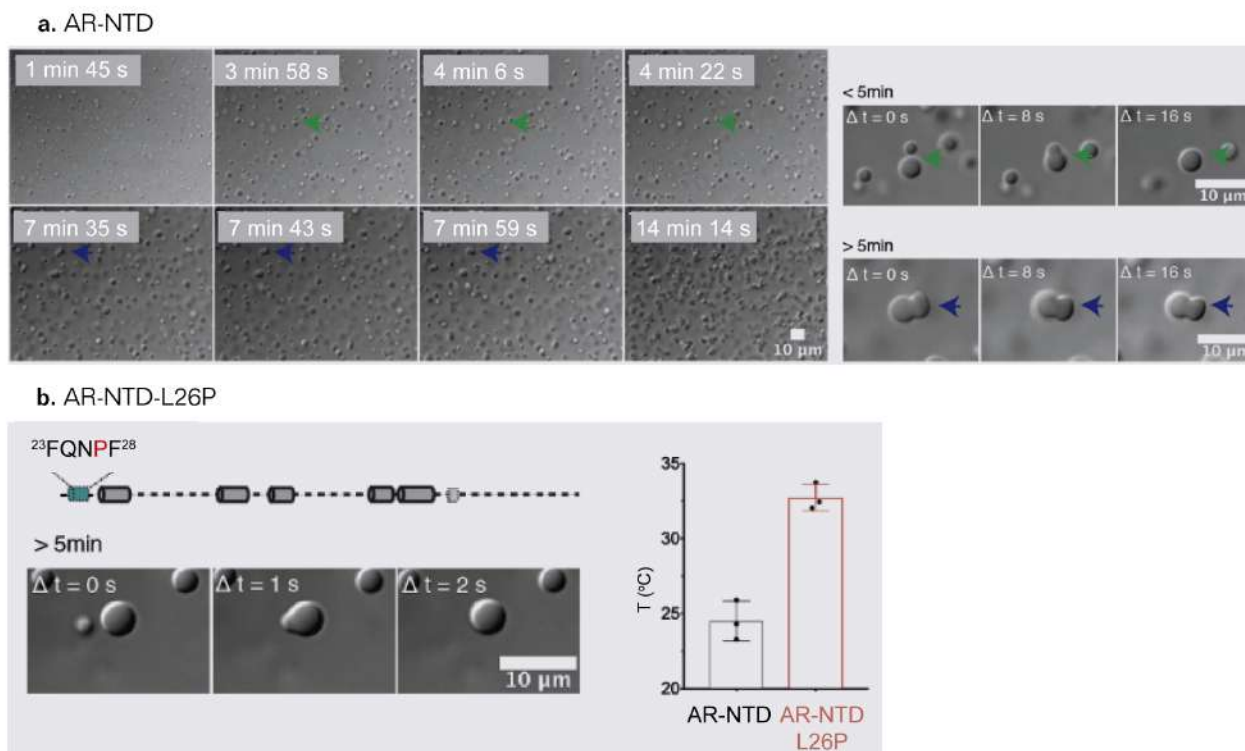


Fig. 3.1.5.3. a) DIC images at different times of a 50 μ M AR-NTD sample at 500 mM NaCl. Zoom-in on the right. **b)** AR-NTD-L26P mutant fusion events of a sample incubated for more than 5 min. **c)** T_c measurement of 50 μ M AR-NTD versus 50 μ M AR-NTD-L26P in 500 mM NaCl by means of the TSA assay. Adapted from Szulc Ph.D. Thesis, 2019⁸⁷.

Finally, since most of the sticky motifs in the AR-NTD that had been determined by NMR were predicted to be partially helical, Dr. Szulc investigated the importance of condensation for the secondary structure propensity of these motifs. This was motivated by the hypothesis that the dry environment in the AR-NTD condensates promoted a helical gain in the motifs important for the AR function and its protein-protein interactions⁹⁴. Indeed, addition of 2,2,2-trifluoroethanol (TFE) to the AR-NTD-L26P increased its helical population (**Fig. 3.1.5.4a**) and its propensity to condensate (**Fig. 3.1.5.4b**). Moreover, additional helix-breaking mutations in the AR-NTD-L26P helices (see section 2.1.3 and 2.1.4) resulted in a decreased propensity to phase separate. Interestingly, perturbing the three helical motifs in the Tau-5 (R1, R2 and R3) had the largest effect on condensation, suggesting a substantial contribution of the helicity in the Tau-5 region to the condensation of the AR-NTD (**Fig. 3.1.5.4c**). These observations led to a “folding upon condensation” model in which condensation would facilitate recruitment of the transcription machinery and co-activators to the AR transcriptional condensates.

3. EPI-001 as a modulator of androgen receptor condensation

AR-NTD acquires secondary structure upon condensation

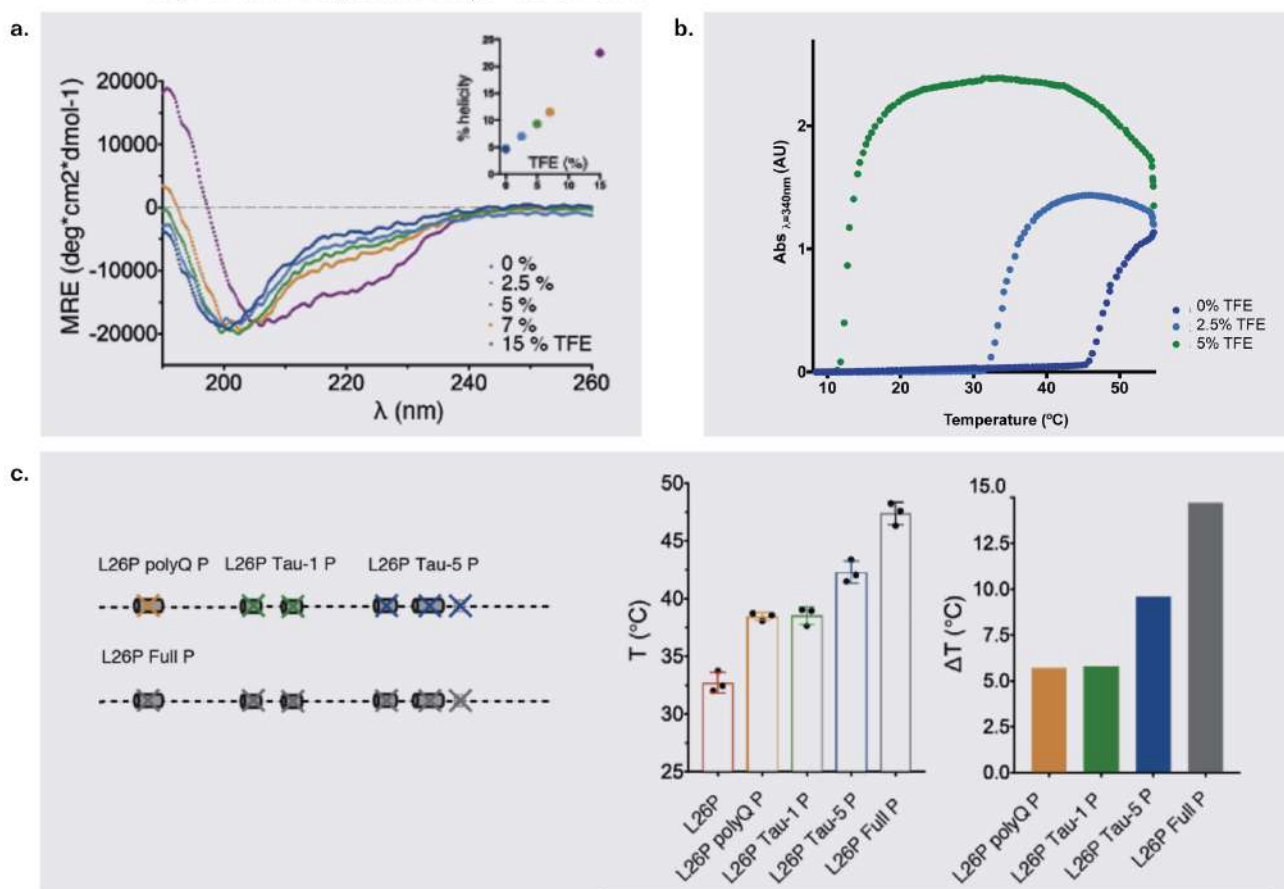


Fig. 3.1.5.4. **a)** Circular dichroism (CD) spectra of AR-NTD-L26P upon TFE titration at 5 °C and % of helicity extracted from the CD plot (insert). **b)** Apparent absorbance (at $\lambda = 340$ nm) measurement at increasing temperatures measurement upon TFE titration. **c)** Left: design of helix-breaking mutants. Right: their effect on the T_c value measured by TSA. Adapted from Szulc Ph.D. Thesis, 2019⁸⁷.

3.2. Experimental Results

As described in section 3.1.5, several observations led us to realise that the androgen receptor (AR) underwent biomolecular condensation in cells and that the disordered androgen receptor N-terminal domain (AR-NTD) was sufficient to recapitulate the process *in vitro*.

Since the AR-NTD is key for the condensation of the receptor and EPI-001 binds to the AR-NTD⁹⁵, we hypothesised that the mode of action of EPI-001 involved the modulation of the ability of the AR to condensate. From the experiments in section 2.2.3, we deduced that EPI-001 binding enhanced the AR-NTD oligomerisation process. This is because EPI-001 promotes and stabilises those conformations prone to self-associate. If we translate these findings to the condensation field, EPI-001 could be able to lower the saturation concentration (C_{sat}) of the AR-NTD by stabilising the condensed phase and/or the conformations prone to condensation.

All the experiments included in this chapter have been carried out using the AR-NTD-L26P construct (see section 3.1.5). AR-NTD-L26P will be named AR-NTD for short and in consistency with section 2.2.

3.2.1. *In vitro* modulation of the AR-NTD condensation by EPI-001

3.2.1.1. EPI-001 reversible binding promotes condensation of the AR-NTD

First, we determined whether EPI-001 had an effect on the condensation of the AR-NTD *in vitro*. We imaged the AR-NTD at increasing concentrations in conditions for phase separation (10% ficoll and 150 mM NaCl) in the absence and presence of 75 μM EPI-001 (**3.2.1.1a**). In the conditions tested without EPI-001, the AR-NTD formed droplets at around 7.5 μM . In the presence of EPI-001, the AR-NTD already demixed at 2 μM . This implies that EPI-001 promotes the condensation of the AR-NTD. To evaluate whether EPI-001 could modulate the condensation of any protein of interest, we run the same experiment using as a model the Fused in Sarcoma (FUS) protein. We observed that EPI-001 did not influence FUS ability to condensate (**3.2.1.1b**). This experiment was conducted in collaboration with Dr. Edgar Boczek at the laboratory of Dr. Tony Hyman at the Max Plan Institute for Cell Biology and Genetics (MPI-CBG) during a short-stay in summer 2019. To unequivocally conclude on specificity, we are planning to check the effect of EPI-001 on a panel of known condensating proteins. However, we can already hint that EPI-001 will, at least, be partially selective as it does not modulate FUS condensates.

3. EPI-001 as a modulator of androgen receptor condensation

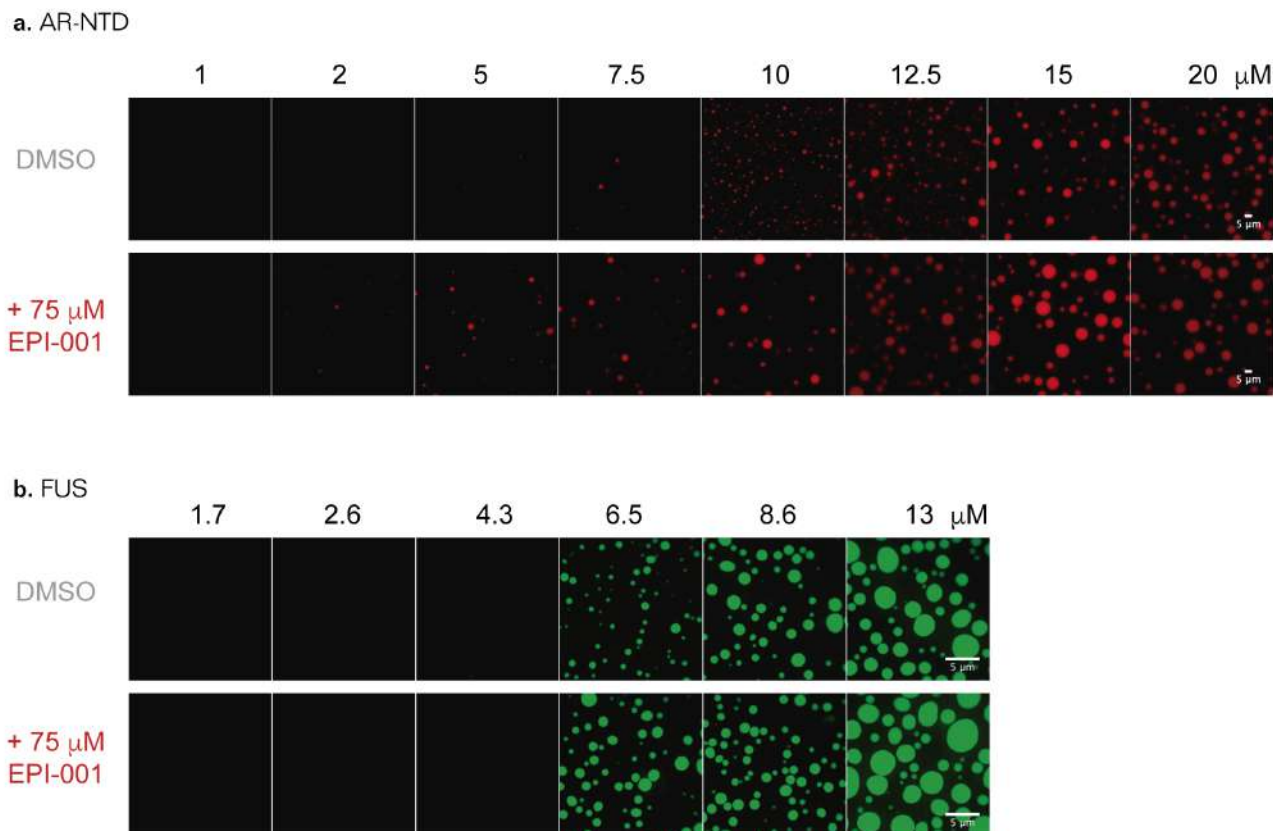


Fig. 3.2.1.1. a) Confocal fluorescence microscopy images of the AR-NTD at different concentrations (indicated in the figure) in 20 mM Hepes 7.4, 150 mM NaCl, 10% ficoll, 2 mM TCEP buffer in the absence (top) and presence (bottom) of 75 μM EPI-001. 1% of the AR-NTD is rhodamine-labelled in all samples. Scale bar, 5 μm. **b)** Confocal fluorescence microscopy images of FUS-GFP at different concentrations (indicated in the figure) in 25 mM Tris-HCl 7.4, 150 mM KCl, 2.5% glycerol and 0.5 mM DTT buffer in the absence (top) and presence (bottom) of 75 μM EPI-001. Scale bar, 5 μm.

To quantify the effects of EPI-001 on the AR-NTD phase separation, we implemented a temperature-dependent turbidity shift assay (TSA). In this assay we measure the sample turbidity as a reporter of condensation and determine the temperature at which a given sample demixes, the cloud point temperature (T_c). When droplets are formed, they scatter light and cause an abrupt increase in turbidity. Therefore, values above 0 absorbance units (AU) in the TSA assay indicate condensation. We measure turbidity as the apparent absorbance at a wavelength in which the absorption of the other components in the sample is minimal, in our case 350 nm. Addition of 5 equivalents (eq) of EPI-001 to a 30 μM AR-NTD and 500 mM NaCl sample lowered the T_c of the AR-NTD by 10 °C (**Fig. 3.2.1.2a**). This is consistent with the imaging data (**3.2.1.1**) and consolidates the finding that EPI-001 promotes the AR-NTD condensation. In contrast, the inactive analogue BADGE was only able to lower the AR-NTD T_c by 3 °C (**Fig. 3.2.1.2a**). We run the TSA assay at various protein concentrations to build the left-arm of the phase diagram (**Fig. 3.2.1.2b**). We observed that EPI-001, but not BADGE, largely enhanced the AR-NTD condensation at all concentrations tested.

Interestingly, in the context of another work, Posey et al. used the polyphasic linkage framework to describe the influence of ligands to the phase separation of a protein ⁷¹. Translating their conclusions to our case study, the fact that EPI-001 lowers the AR-NTD T_c is equivalent to

3. EPI-001 as a modulator of androgen receptor condensation

establishing that EPI-001 preferentially partitions into the condensed phase of the AR-NTD. If the compound would interact with both phases equally, the T_c would not be affected, as in the case of BADGE.

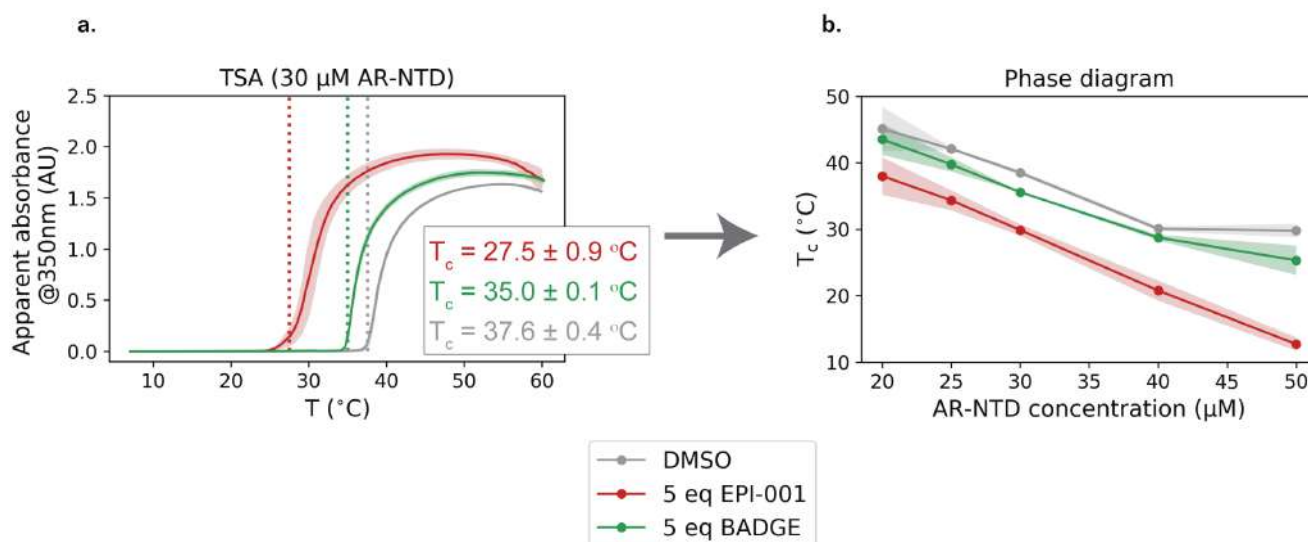


Fig. 3.2.1.2. a) Apparent absorbance (at $\lambda = 350$ nm) measurement at increasing temperatures of $30 \mu\text{M}$ AR-NTD in 20 mM Hepes 7.4 , 500 mM NaCl, 2 mM TCEP buffer in the absence (grey) and presence of 5 eq of EPI-001 (red) or BADGE (green). Dashed lines indicate condensation onset and calculated T_c values are depicted in the insert. **b)** T_c obtained using the TSA assay of AR-NTD at different concentrations in 20 mM Hepes 7.4 , 500 mM NaCl, 2 mM TCEP buffer in the absence (grey) and presence of 5 eq of EPI-001 (red) or BADGE (green). Mean values of 3 independent measurements and the standard deviation are plotted.

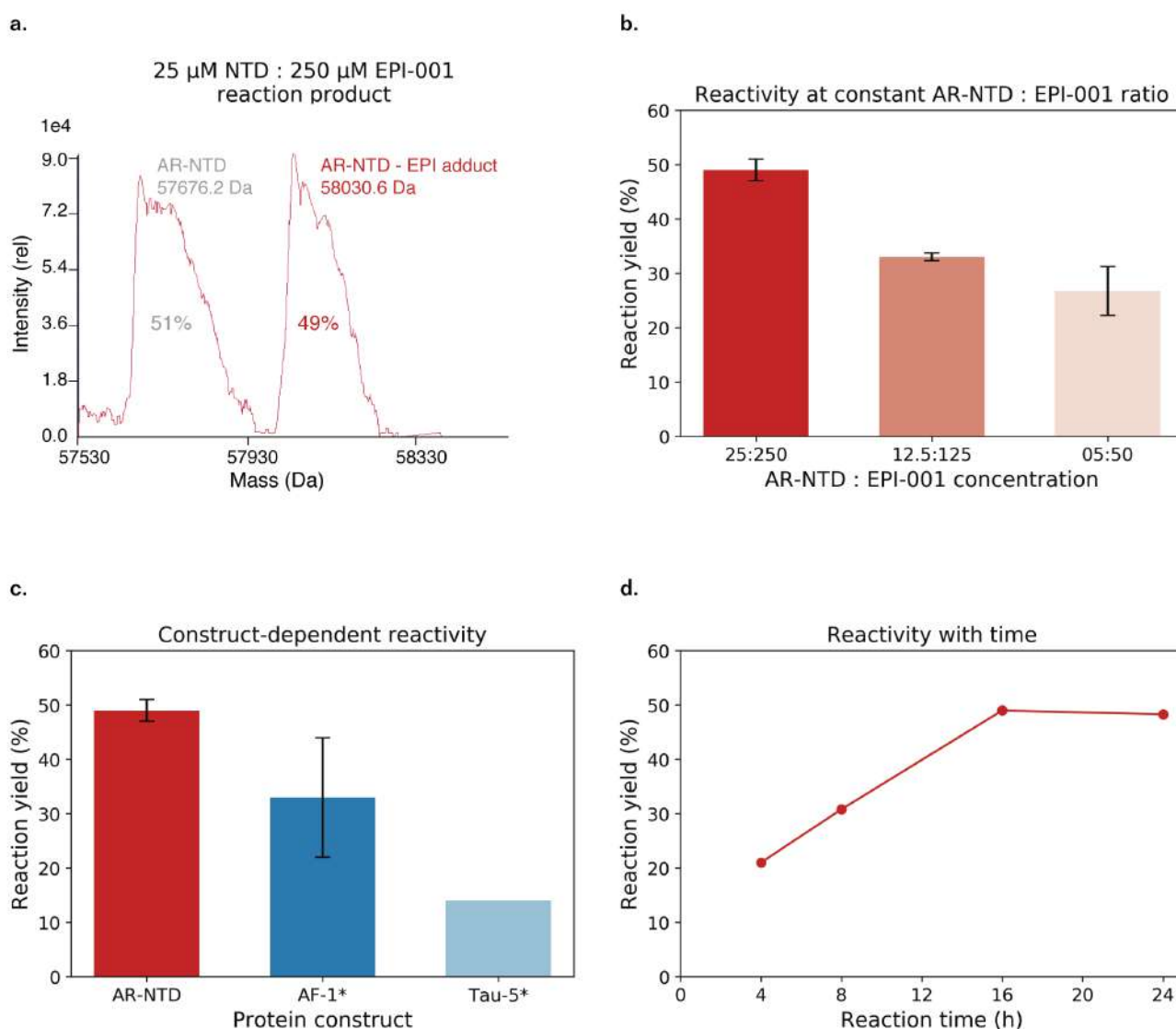
3.2.1.2. EPI-001 covalently reacts with the AR-NTD

EPI-001 was proposed to have a two-step mechanism of action⁹⁶. First, it interacts reversibly with the target, as confirmed in section 2.2. Second, EPI-001's electrophilic chlorohydrin warhead covalently reacts with a nucleophilic side chain of the AR-NTD.

In order to recapitulate the covalent binding *in vitro* (named epilation for short), we incubated 10 eq of EPI-001 with $25 \mu\text{M}$ AR-NTD overnight at 37 °C and $\text{pH} = 8.0$. We could not detect covalent addition at lower temperatures nor pH (data not shown). The reaction yielded $49 \pm 2\%$ of the AR-NTD - EPI adduct, measured by liquid chromatography coupled to intact mass spectrometry (LC-MS) (**Fig. 3.2.1.3a**). Intact LC-MS is able to detect the wild-type AR-NTD intact mass and differentiate it from that of the AR-NTD - EPI adduct. Interestingly, we always observed the addition of 1 EPI-001 per AR-NTD molecule under these conditions. Two sets of evidence indicate that the limiting reagent in the reaction is the oligomeric population of the AR-NTD. First, by maintaining the protein:inhibitor ratio constant but altering the absolute concentrations, we observed a concentration-dependent increase in the reaction yield (**Fig. 3.2.1.3b**). Second, as EPI-001 was shown to reversibly interact with the Tau-5 region⁹⁵, we determined the reactivity of EPI-001 towards constructs shorter than the AR-NTD but containing the Tau-5 region (AF-1* and Tau-5*, defined in section 2.2). Noteworthy, the oligomerisation propensity of these constructs

3. EPI-001 as a modulator of androgen receptor condensation

increased with their length (Tau-5* < AF-1* < AR-NTD)⁹⁷. Under the same reaction conditions, we observed, an increased reaction yield with increasing construct length: 14% for Tau-5*, 33% for AF-1* and 49% for AR-NTD (**Fig. 3.2.1.3c**). These results suggest that reactivity of EPI-001 depends on the propensity of the AR-NTD constructs to oligomerise. Considering oligomerisation as the starting point for condensation, and as condensation also depends on the ionic strength, we investigated the reactivity of EPI-001 at low salt (no condensation) versus high salt (condensation) conditions. However, the protein in the dense phase aggregated at incubation times necessary for EPI-001 reaction to take place (> 16 h) (**Fig. 3.2.1.3d**) and we were not able to resuspend the aggregate back in solution even when employing harsh methods like the use of urea or cyanoguanidine, which hampered the MS analysis.



3. EPI-001 as a modulator of androgen receptor condensation

–

Fig. 3.2.1.3. a) Representative spectra of the reaction product of 25 μM AR-NTD + 250 μM EPI-001 overnight at 37 $^{\circ}\text{C}$ and pH = 8.0 analysed by LC-MS. **b)** Reaction yield quantified by LC-MS at constant AR-NTD : EPI-001 ratio (1:10) overnight at 37 $^{\circ}\text{C}$ and pH = 8.0 but different absolute concentrations (indicated in the figure). Mean values of 2 independent measurements and the standard deviation are plotted. **c)** Reaction yield quantified by LC-MS 25 μM AR-NTD, AF-1* or Tau-5* + 250 μM EPI-001 overnight at 37 $^{\circ}\text{C}$ and pH = 8.0. Mean values of 3 independent measurements and the standard deviation are plotted. **d)** Reaction yield of 25 μM AR-NTD + 250 μM EPI-001 at different time points (4, 8, 16, 24 h) at 37 $^{\circ}\text{C}$ and pH = 8.0 quantified by LC-MS.

Finally, we wanted to determine which residue or residues of the AR-NTD were reacting with EPI-001. For this, we first incubated 25 μM AR-NTD with 250 μM EPI-001 overnight at 37 $^{\circ}\text{C}$ and pH = 8.0. Then, we digested the product of the reaction and quantified the amount of EPI-001 per residue in the AR-NTD. We observed that the residue side chain acting as a nucleophile was that of cysteines. In our experimental conditions, even though the AR-NTD has 11 potential nucleophilic cysteines, EPI-001 preferentially reacted with Cys267 and Cys519. The reactivity of Cys240 was less reliable as the different technical replicas gave variable results (**Fig. 3.2.1.4**). Additional experiments should be carried out to discern which is the property shared among these cysteines that allows them to be good nucleophiles for EPI-001. Moreover, it would be interesting to understand whether the preferential binding for Cys267 and Cys519 is altered under conditions leading to condensation.

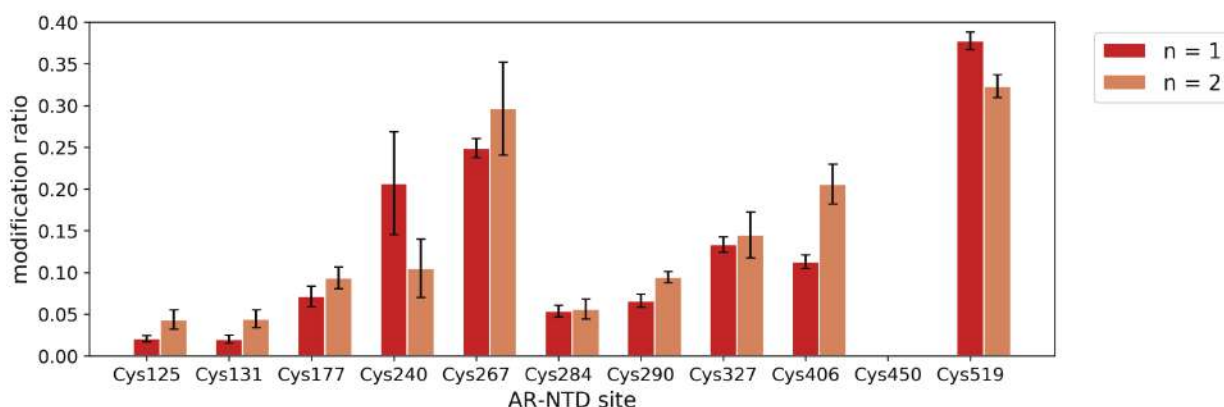


Fig. 3.2.1.4. EPI-001 modification ratio (r) for cysteine site (N) in the AR-NTD ($r = N_{\text{modified}} / (N_{\text{modified}} + N_{\text{no-modified}})$). Mean values and standard deviation of 3 technical replicas per independent measurement ($n=1$ and $n=2$) are plotted.

3.2.1.3. Epilation effect on AR-NTD condensation

To determine the effect of epilation on the AR-NTD condensation propensity, we measured the T_c of AR-NTD samples at increasing epilation percentage. We observed that i) epilation greatly enhanced the AR-NTD condensation propensity (**Fig. 3.2.1.5a**) and ii) the condensation was linearly promoted with epilation percentage (**Fig. 3.2.1.5b**). It is important to point out the difference between EPI-001 binding reversibly and irreversibly on the enhancement of the AR-NTD condensation. Addition of 5 eq of EPI-001 lowered the AR-NTD T_c by 10 $^{\circ}\text{C}$, whereas only

3. EPI-001 as a modulator of androgen receptor condensation

0.4 eq were required to achieve the same shift when EPI-001 was covalently bound. The effect of the epilation was confirmed by fluorescence microscopy imaging. 0% and 50% epilated samples were imaged at increasing sodium chloride salt (NaCl) concentration (**Fig. 3.2.1.5c**). We observed that 0% epilated samples formed droplets at 750 mM NaCl, whereas 50% epilated samples already formed them at 550 mM NaCl. We rationalise this finding by the fact that hydrophobic and aromatic residues are essential for the AR-NTD to condensate as it follows an LCST-type behaviour (section 3.1.5). Therefore, *decorating* the AR-NTD with hydrophobic and aromatic EPI-001 molecules increases the overall propensity of the protein to condensate.

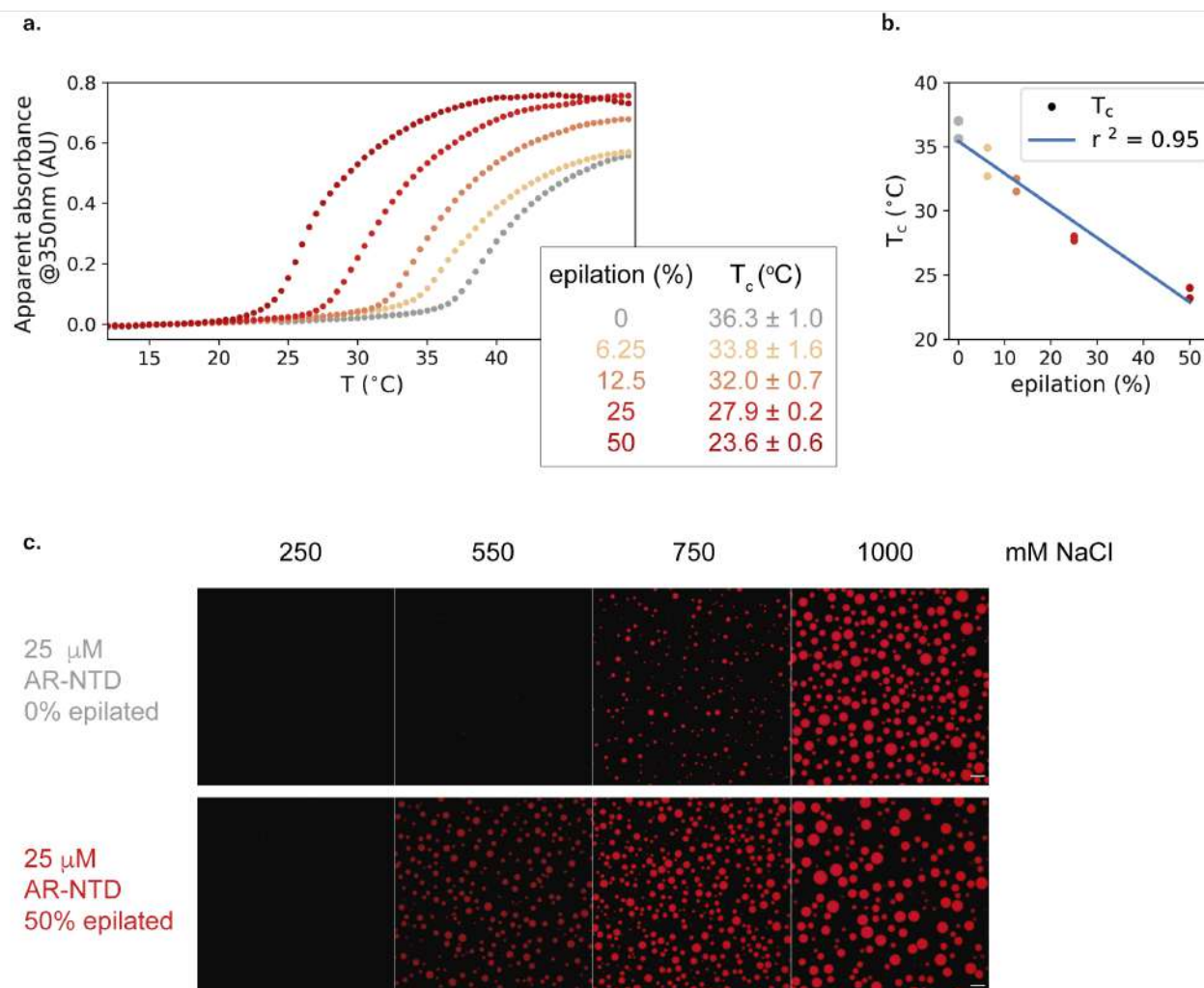


Fig. 3.2.1.5. a) Apparent absorbance (at $\lambda = 350$ nm) measurement at increasing temperatures of 25 μ M AR-NTD in 20 mM NaP 7.4, 500 mM NaCl, 2 mM TCEP buffer with different percentages of epilation. Calculated T_c values (insert). Mean values and the standard deviation of 2 independent measurements are considered, for clarity only $n=1$ is represented. **b)** Correlation between T_c and epilation percentage. **c)** Confocal fluorescence microscopy images of 25 μ M AR-NTD 0% and 50% epilated in 20 mM NaP 7.4, 2 mM TCEP buffer at different NaCl concentrations (indicated in the figure). 1% of the AR-NTD is rhodamine-labelled in all samples. Scale bar, 10 μ m.

To determine whether enhancement of the condensation process influenced the AR-NTD material properties of the condensed phase we performed fluorescence recovery after photobleaching (FRAP) experiments. *In vitro*, we can observe how epilation does not affect the fluorescence recovery of the AR-NTD (**Fig. 3.2.1.6**). Remarkably, the mobile fraction of wild-type and epilated

samples was around 50%. We suspect this is due to the long time required to prepare the epilated AR-NTD (see section 6: experimental methods). During this time, the AR-NTD evolves and hardens even if centrifugation is carried out prior to the FRAP sample preparation to remove any possible aggregates.

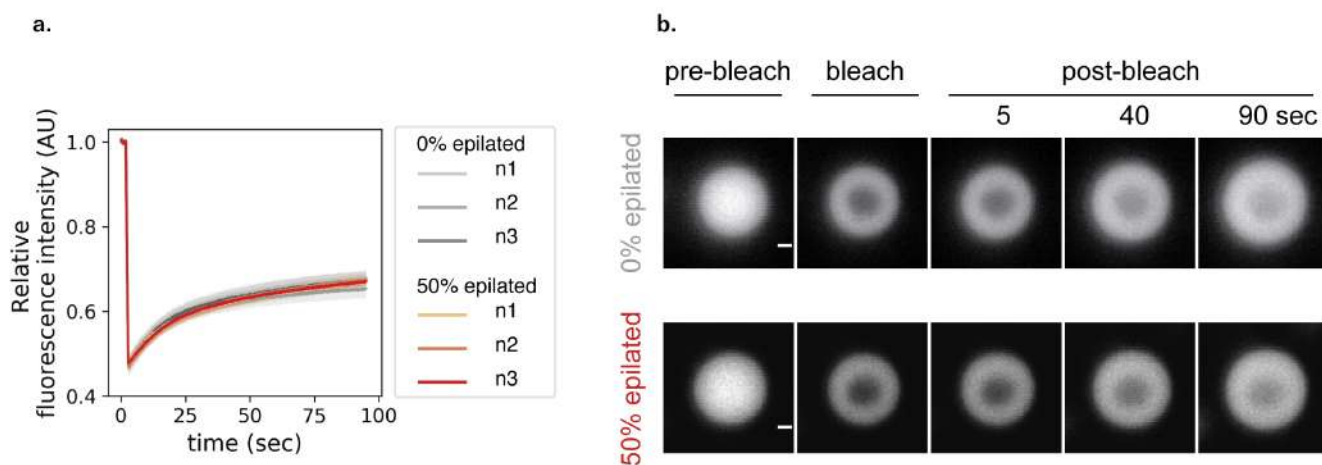


Fig. 3.2.1.6. a) Average relative fluorescence intensity curve of 0% (grey) and 50% (red) epilated AR-NTD droplets (25 μ M in 20 mM NaP 7.4, 500 mM NaCl, 2 mM TCEP buffer) as a function of time upon photobleaching. 3 independent experiments are plotted. The line represents the mean and the shaded area represents the standard deviation of the FRAP measurements (10 droplets per experiment). **b)** Representative images of the FRAP measurement of 0% (above) and 50% (below) epilated AR-NTD droplets. Scale bar, 1 μ m.

In summary, EPI-001 partitions into the droplets and stabilises the AR-NTD condensates *in vitro*. EPI-001 covalent addition is a crucial step for the promotion and stabilisation of the AR-NTD droplets without affecting its material properties. Altogether, this data indicates that EPI-001 mode of action includes the modulation of the AR-NTD condensation propensity.

3.2.2. In cell modulation of AR-NTD condensation by EPI-001

Previous work from our laboratory (section 3.1.5), demonstrated that the AR undergoes LLPS in cells when it is activated by the hormone. We found that the condensates of the AR have liquid-like properties, are spherical, fuse and are highly dynamic as shown by FRAP (unpublished data).

To investigate the effect of EPI-001 on the condensation propensity of the full length AR in cells, we expressed eGFP-tagged full length AR (eGFP-AR)⁹⁸ in a prostate cancer cell line (PC3) and we monitored the formation and growth of eGFP-AR cytosolic droplets upon 1 nM dihydrotestosterone (DHT) addition. However, AR droplet formation and growth competed with the translocation of AR to the nucleus, which hampered the quantitative analysis (**Fig. 3.2.2.1a**). For this reason, we generated a mutant form, the eGFP-AR- Δ NLS, bearing a deletion of the nuclear localisation signal

(NLS)⁹⁹. eGFP-AR- Δ NLS also condensates upon hormone activation but does not translocate to the nucleus (**Fig. 3.2.2.1b**).

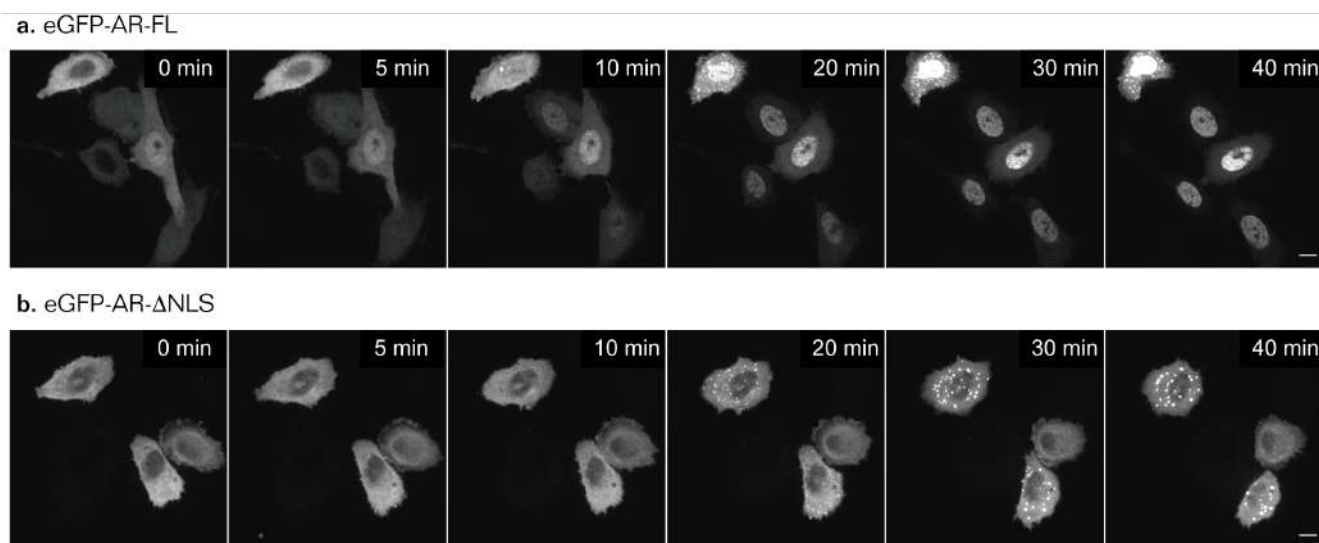


Fig. 3.2.2.1. a) Time-resolved fluorescence microscopy of eGFP-AR upon 1 nM DHT activation in PC3 cells. Scale bar, 10 μ m. **b)** Time-resolved fluorescence microscopy of eGFP-AR- Δ NLS upon 1 nM DHT activation in PC3 cells. Scale bar, 10 μ m.

PC3 cells were transiently transfected with eGFP-AR- Δ NLS and the size and number of droplets per cell were monitored upon activation with 1 nM DHT ($t = 0$ min). Prior to activation, cells were treated for 1 h with vehicle (DMSO) or 25 μ M EPI-001. EPI-001-treated cells showed a comparable density of droplets but they were of larger size (**Fig. 3.2.2.2**). This indicates that EPI-001 does not promote AR nucleation but stabilises the already-formed droplets. If EPI-001 would nucleate the AR, we would observe an increase in the number of droplets as there would be more nucleation points. Moreover, those droplets would be of smaller size as the protein in the dense phase would have to be distributed among all the droplets. However, what we observe is the opposite. AR droplets are of larger size because EPI-001 stabilises the condensed phase⁷¹.

3. EPI-001 as a modulator of androgen receptor condensation

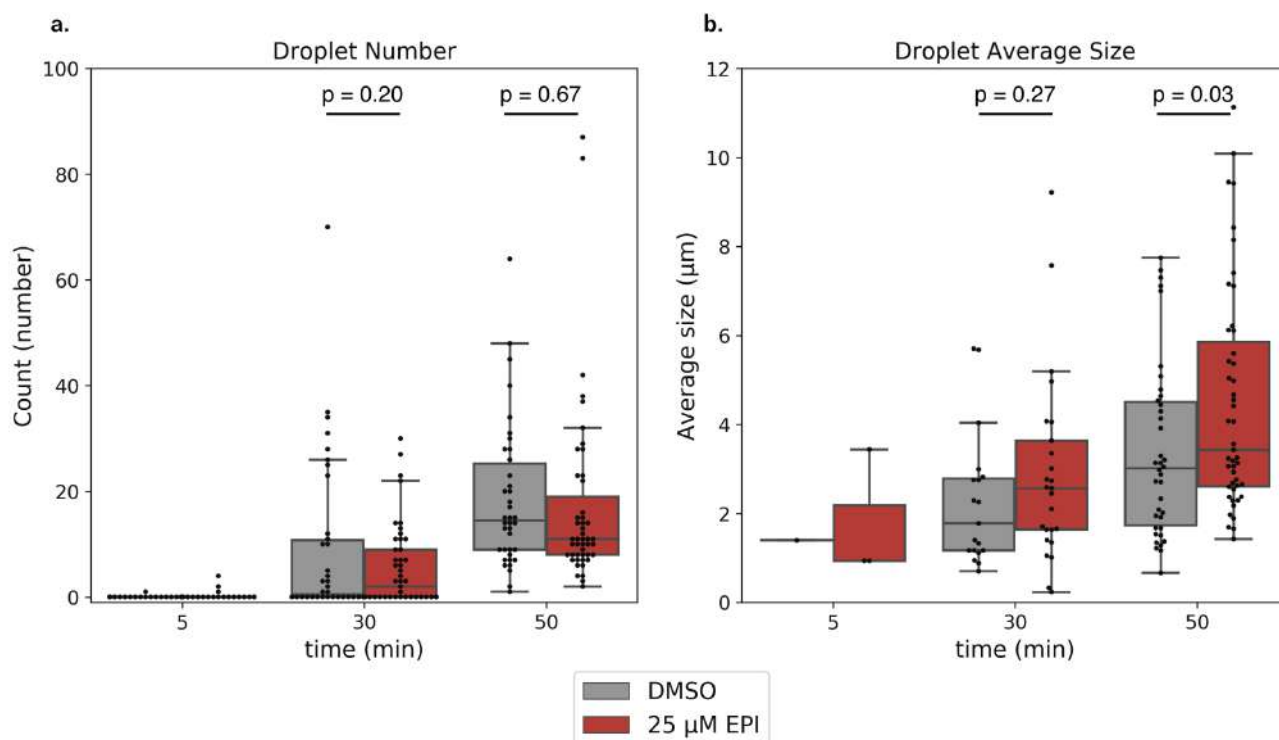


Fig. 3.2.2.2. PC3 cells transiently transfected with eGFP-AR-ΔNLS were treated for 1 h with vehicle (DMSO, grey) or 25 μM EPI-001 (red). eGFP-AR-ΔNLS was activated by addition of 1 nM DHT ($t = 0$ min) and the number and average size of droplets was quantified after 5, 30 and 50 min. **a)** Plot of the droplet number per cell over time. **b)** Plot of the droplet average size per cell over time. Values of 3 independent experiments are plotted. Significance tested with an independent two samples t -test.

As mentioned, the AR serves its function in the nucleus where it also forms clusters (section 3.1.5). For this reason, we studied the effect of EPI-001 in the AR nuclear clusters. To image the AR nuclear clusters we employed super-resolution stimulation emission depletion (STED) microscopy as the size of the clusters is below the diffraction limit. These experiments were carried out in collaboration with Dr. Karina Pombo-Garcia in Dr. A. Honigmann's laboratory at the MPI-CBG. For these experiments, we shifted to a cellular system with lower AR levels by using HeLa cells stably expressing AR-eGFP. EPI-001-treated cells showed a comparable number of clusters per nucleus to that of untreated cells. Nevertheless, the clusters in EPI-001-treated nuclei were of larger size (**Fig. 3.2.2.3**). These experiments are consistent with the ones measuring cytosolic droplets (**Fig. 3.2.2.2**).

3. EPI-001 as a modulator of androgen receptor condensation

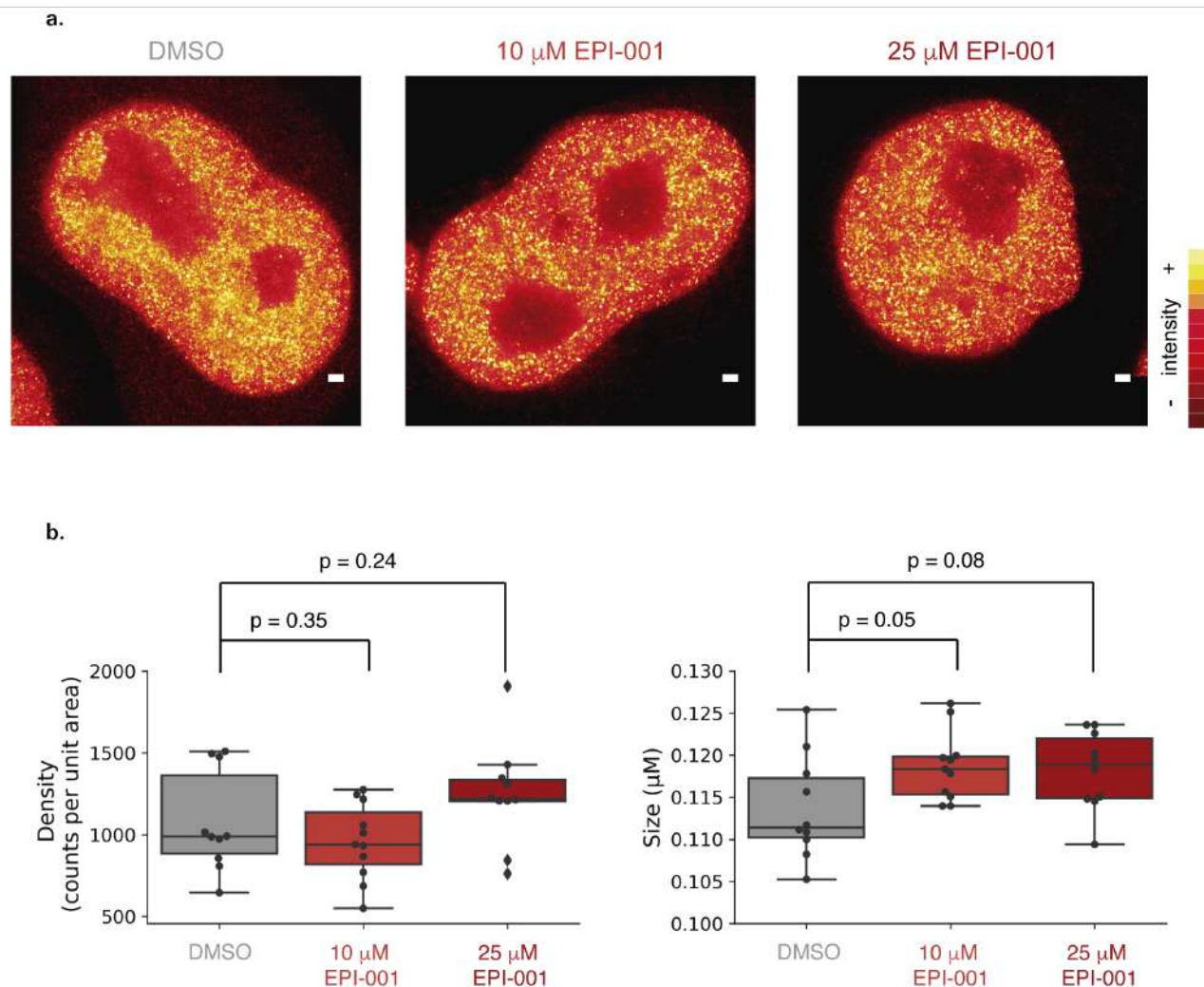


Fig. 3.2.2.3. a) Representative STED images of a HeLa cell nucleus stably transfected with AR-eGFP upon 4 h 1 nM DHT activation pre-treated with vehicle (DMSO), 10 μM EPI-001 or 25 μM EPI-001 for 1 h. Scale bar, 1 μm . **b)** Quantification of the density and size of clusters per nuclei of AR-eGFP HeLa cells upon 4 h 1 nM DHT activation. Cells were treated for 1 h prior to activation with vehicle (DMSO, grey), 10 μM EPI-001 (light red) or 25 μM EPI-001 (dark red). Values from 10 independent nuclei are plotted. Significance tested with an independent two samples *t*-test.

Finally, we examined the cellular dynamics of the receptor upon treatment with EPI-001. We performed FRAP experiments of nuclear AR-eGFP stably expressed in HeLa cells treated with vehicle (DMSO) or 25 μM EPI-001 for 1 h prior to the addition of 1 nM DHT for 4 h. As for the *in vitro* data acquired (Fig. 3.2.1.6), we could not observe significant differences between the dynamics of the receptor in the nucleus of EPI-001 treated cells compared to that of non-treated cells (Fig. 3.2.2.4).

3. EPI-001 as a modulator of androgen receptor condensation

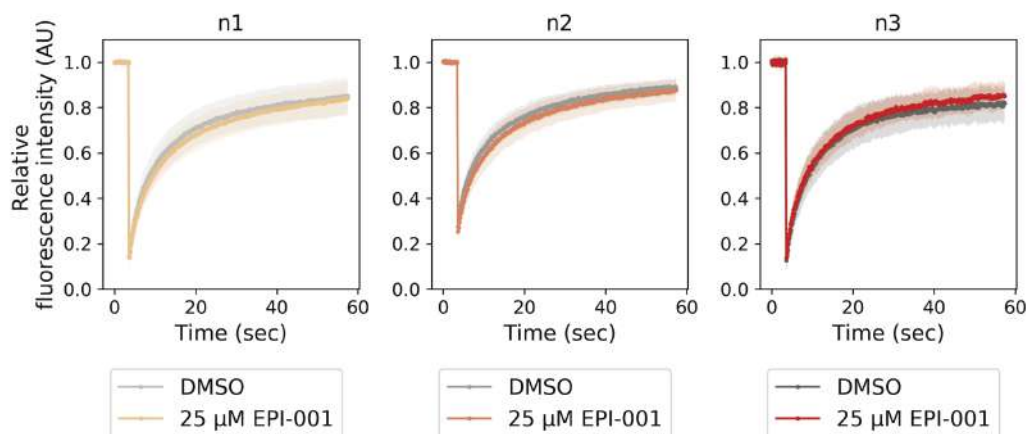


Fig. 3.2.2.4. Average relative fluorescence intensity curve as a function of time upon photobleaching of a 15x15 pixel nuclear region of HeLa cells stably expressing AR-eGFP treated with vehicle (DMSO, grey) or 25 μM EPI-001 (red) for 1 h prior to the addition of 1 nM DHT for 4 h. 3 independent experiments (n1, n2 and n3) are plotted separately for clarity. The line represents the mean and the shaded area represents the standard deviation of the FRAP measurements (at least 15 nuclei per experiment).

In summary, EPI-001 does not induce AR nucleation but stabilises the already-formed AR droplets without affecting its material properties in cells. These results are in agreement with the *in vitro* condensation data on the AR-NTD.

3.2.3. Consequences of modulating AR-NTD condensation with EPI-001 to AR function

EPI-001 i) inhibits the AR function, as reported in the literature^{96,100}, and ii) stabilises the AR-NTD condensates, as indicated in the previous sections of this chapter. However, it is difficult to rationalise the link between stabilising the AR-NTD condensation and inhibiting AR function as it is still not well established what is the role of condensation in the AR pathway.

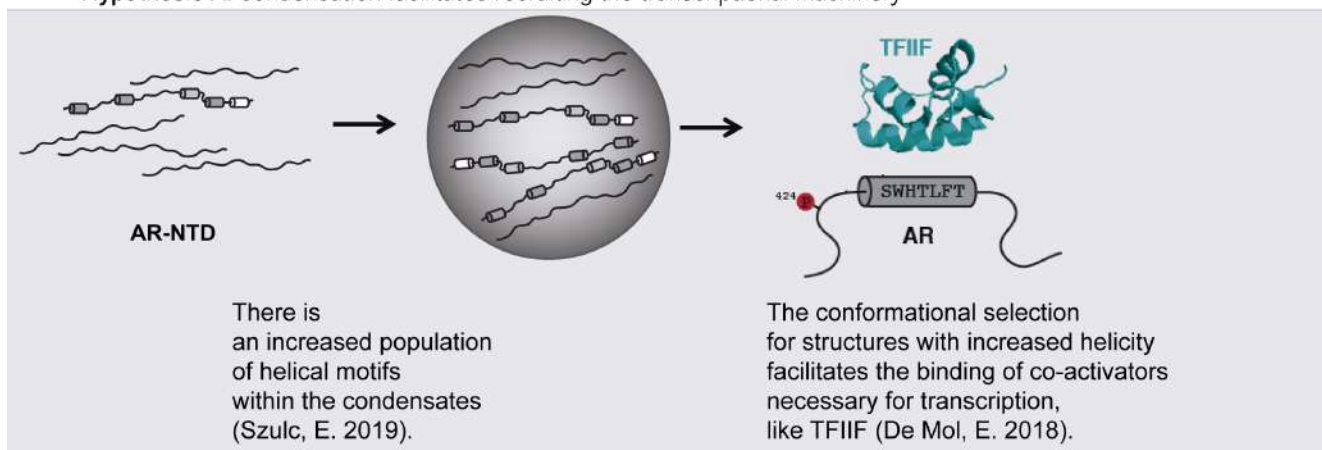
In this sense, many hypothetical explanations can be drawn to link both observations. We have experimentally explored two scenarios, as depicted in Figure 3.2.3.1.

3. EPI-001 as a modulator of androgen receptor condensation

-

Possible functions of AR condensates

· **Hypothesis A:** condensation facilitates recruiting the transcriptional machinery



· **Hypothesis B:** condensation is a buffering system

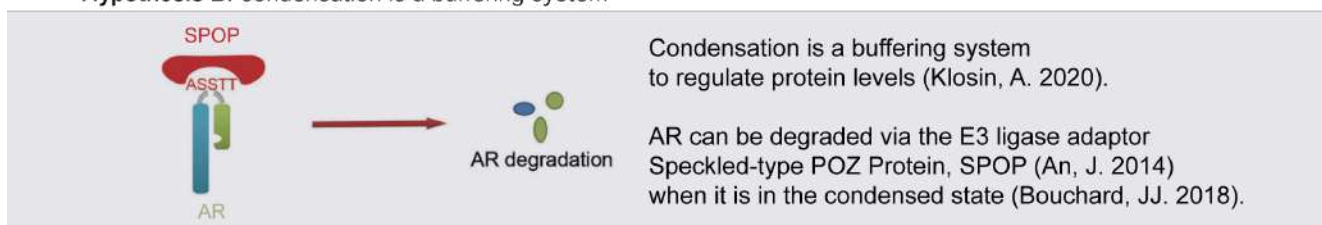


Fig. 3.2.3.1. Possible functions of AR condensates. **a)** Condensation facilitates recruiting the transcriptional machinery. Adapted from De Mol et al. 2018 and Szulc Ph.D. Thesis, 2019^{87,94}. **b)** Condensation is a buffering system. Adapted from An et al. 2014; Bouchard et al. 2018 and Klosin et al. 2020^{28,30,101}.

3.2.3.1. Condensation facilitates recruiting the transcriptional machinery

In scenario A, transcriptional condensates bring together all the components necessary for transcription, such as the RNA-polymerase II (RNA polII), the mediator complex, transcription factors and coactivators^{56,58}. Moreover, in the condensed state, the AR-NTD population of structured conformations is increased, thereby facilitating the interaction between the AR-NTD and the general transcriptional machinery necessary for transcription initiation^{87,94}. In this sense, modulation of the AR-NTD condensation by EPI-001 might hamper these interactions leading to a reduced transcriptional activity of AR.

We have interrogated the effect of EPI-001 on the interaction between the AR and the RNA-polIII. For this, we exploited the proximity-ligation assay (PLA), an immunofluorescence-based technique that allows the detection of proteins that are in close proximity (< 40 nm) inside cells^{94,102}. We detected the interaction between the AR and the endogenous RNA-polIII in the presence of 1 nM DHT in HeLa cells stably expressing AR-eGFP. We observed that addition of 10 or 25 μ M of EPI-001 for 1 h prior to DHT activation for 4 h reduced the number of AR - RNA-polIII interactions (**Fig. 3.2.3.2**).

3. EPI-001 as a modulator of androgen receptor condensation

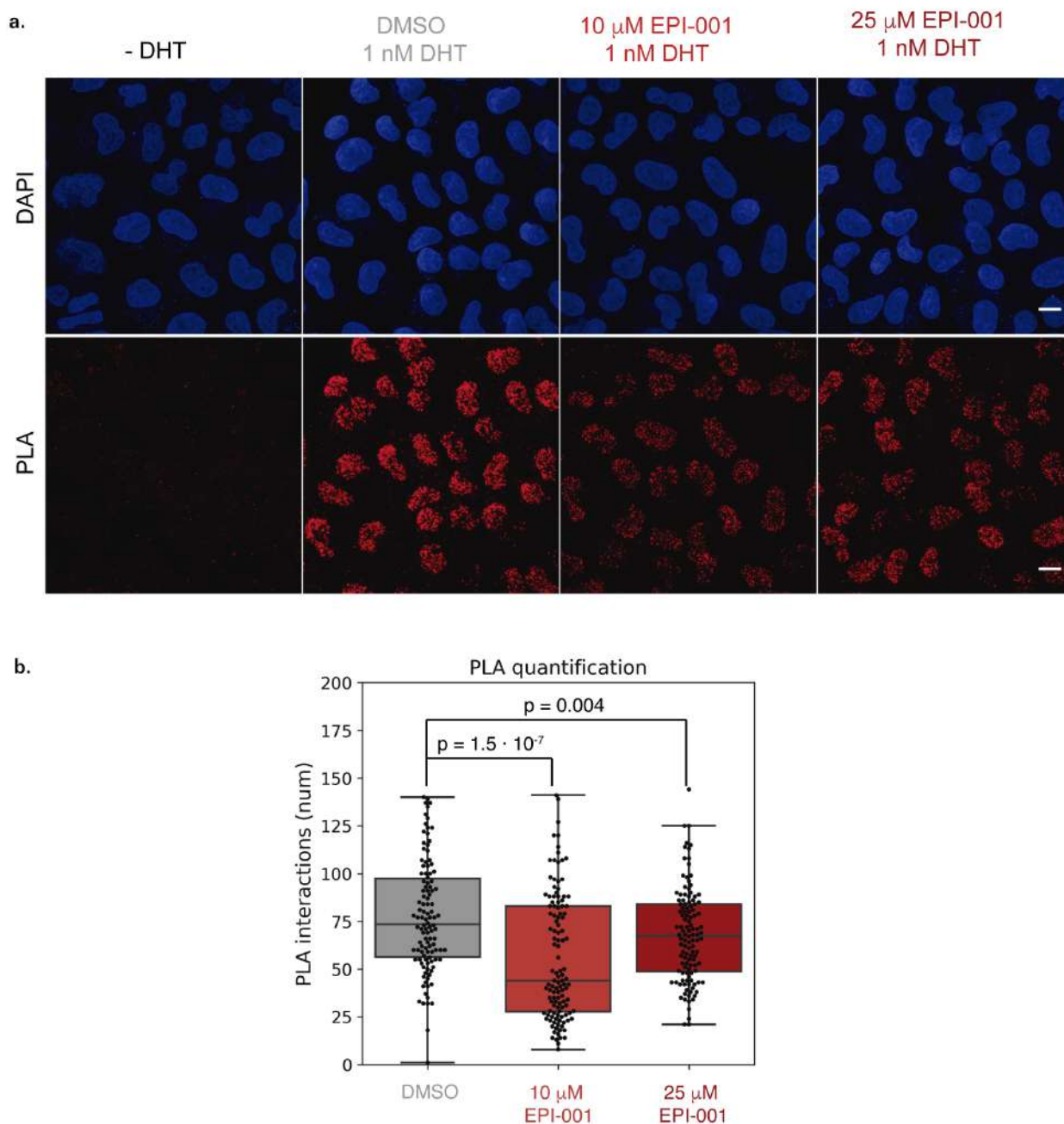


Fig. 3.2.3.2. a) Representative images of PLA interactions between AR and RNA-polIII in HeLa cells stably expressing AR-eGFP. Cells were treated with vehicle (DMSO), 10 μ M EPI-001 or 25 μ M EPI-001 for 1 h prior to activation with 1 nM DHT for 4 h. As a control, in the absence of activation no PLA signal is detected. DAPI indicates the location of nuclei. Scale bar, 10 μ m. **b)** Quantification of the PLA interactions per nucleus between AR and RNA-polIII of DMSO (grey), 10 μ M EPI-001 (light red) or 25 μ M EPI-001 (dark red) treated cells for 1 h prior to 4 h 1 nM DHT activation. Error bars represent the mean and standard deviation of 2 independent experiments (at least 50 nuclei per experiment). Significance tested with an independent two samples *t*-test.

3.2.3.2. Condensation is a buffering system

In scenario B, condensation could be involved in the nuclear degradation pathway of the AR via the speckle-type POZ protein (SPOP). SPOP is an E3 ligase adaptor that colocalises with the AR in the condensed state, thereby facilitating the AR clearance^{30,101}. In this sense, promoting the condensation of the AR might lead to increased AR degradation, thereby reducing AR transcriptional activity.

Preliminary experiments were conducted to understand whether EPI-001 promoted the AR degradation. For this purpose, we examined the AR protein expression levels in total homogenates of AR-eGFP stably expressed HeLa cells treated with increasing concentrations of EPI-001 for 1 h prior to 4h incubation with DHT. Western blot analysis (WB) revealed no significant decay of AR levels in the conditions tested (Fig. 3.2.3.3).

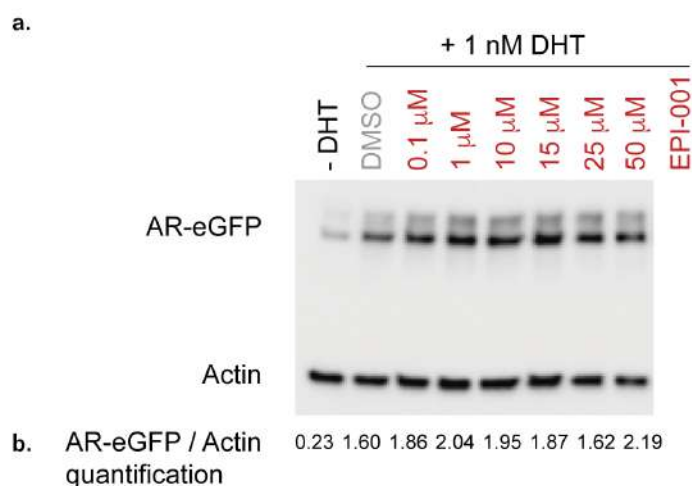


Fig. 3.2.3.3. a) WB analysis to assess AR-eGFP protein expression levels in HeLa cells stably expressing AR-eGFP treated with vehicle (DMSO) or increasing concentrations of EPI-001 (indicated in the figure) for 1 h prior to activation with 1 nM DHT for 4 h. β -actin levels are determined as loading control. **b)** Quantification of AR-eGFP levels normalised to those of actin.

In summary, our results indicate that the reduction in transcriptional activity might be explained by a decrease in the interactions between the AR and the RNA-polIII. AR degradation induced by EPI-001 does not seem to be implicated in the inhibition of the AR activity. Nevertheless, further studies will be carried out to fully establish or discard this hypothesis.

3.3. Discussion

Biomolecular condensates are involved in the spatiotemporal control of cellular processes¹⁸. They are formed by phase transitions, the most studied case being liquid-liquid phase separation (LLPS)¹⁷. LLPS is the process in which a biopolymer that is homogeneously distributed within a solvent demixes into two coexisting liquid phases, a polymer-enriched phase (dense phase) and a polymer-depleted phase (light phase). LLPS is driven by multivalent interactions between the biopolymers. Intrinsically disordered proteins (IDPs) are key players in biomolecular condensation as by definition are multivalent proteins that can simultaneously interact with other molecules and themselves¹.

Biomolecular condensates participate in many cellular functions^{7,17-19} and are linked to various diseases including neurodegeneration^{42,43}, cancer³⁰ and aging⁶. Biomolecular condensation offers a new layer of regulation for IDPs and, hence, a potential new opportunity to target currently undruggable proteins^{59,60}. However, this is still an underexplored area of research.

During the course of this thesis we realised that the androgen receptor (AR) underwent condensation upon hormone binding⁸⁷. Moreover, the disordered androgen receptor N-terminal domain (AR-NTD) was sufficient to recapitulate the condensation process *in vitro*^{30,87}. Purified AR-NTD underwent LCST type phase separation. Since the AR-NTD demixed and EPI-001 interacted with the AR-NTD, we hypothesised whether the mode of action of EPI-001 would be related to its ability to modulate the AR-NTD condensation propensity.

We observed that EPI-001 stabilised the AR-NTD condensates *in vitro* and in cells. *In vitro*, stabilisation of the condensates was larger when EPI-001 was covalently bound to the AR-NTD than when it was reversibly interacting with it. Moreover, covalent binding of EPI-001 did not affect the material properties of the condensates.

In section 3.1.3, we put forward four different mechanisms by which, theoretically, a small-molecule can modulate condensation: 1) regulatory inhibition, 2) orthosteric inhibition, 3) allosteric inhibition and 4) colligative inhibition. The experiments carried out in this thesis indicate that EPI-001 acts as both an allosteric and a colligative inhibitor. This is depicted in the following working model (**Fig. 3.3.1**). First, EPI-001 preferentially partitions into the condensed phase of the AR-NTD as it lowers the AR-NTD cloud point temperature. According to the polyphasic linkage framework⁷¹, stabilisation is equivalent to partitioning into the condensed phase. Moreover, EPI-001 reversibly interacts with the AR-NTD shifting its ensemble towards oligomerisation prone conformations and enhancing its condensation process. Then, EPI-001 covalently modifies the AR-NTD. EPI-001-modified AR-NTD (epilated AR-NTD) demixes more readily than the wild-type AR-NTD. In other words, epilation permanently modifies the AR-NTD causing a shift in the phase boundary so that the epilated AR-NTD demixes in conditions in which the wild-type AR-NTD does not.

3. EPI-001 as a modulator of androgen receptor condensation

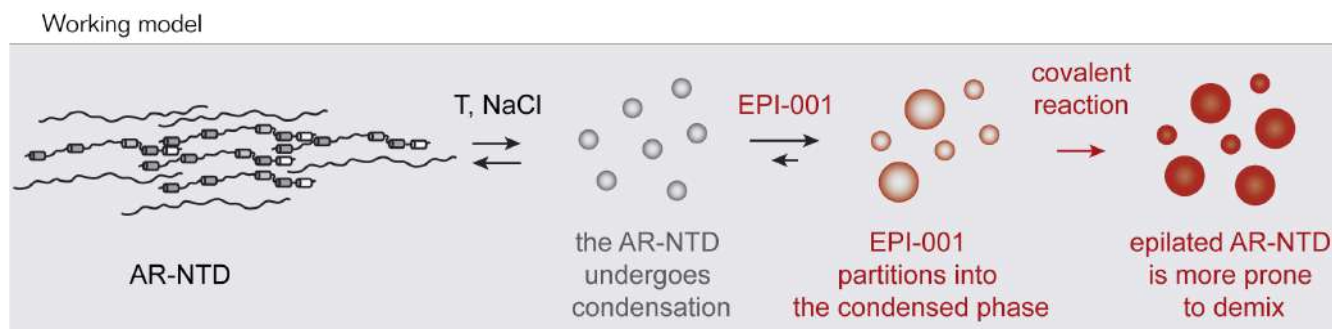


Fig. 3.3.1. Current working model for EPI-001 mechanism of action related to its ability to stabilise AR condensation.

Functionally, EPI-001 inhibits the AR transcriptional activity as described in the literature^{96,100}. The link between inhibition of the AR transcriptional activity and the stabilisation of the AR condensation by EPI-001 is difficult to establish. This is mainly because it is yet unclear which is the functional role of the condensates in the AR pathway. If AR condensates were to be the active state of the protein, then hampering the AR condensation would seem the appropriate strategy to follow. However, what we observe is the opposite, that EPI-001 promotes the AR condensation in cells. In this sense, we started to investigate two scenarios from the many hypothetical explanations that can be put forward: a) AR condensates which have been epilated are no longer functional, e.g. in the condensates AR is not capable to participate in essential protein-protein interactions; and b) AR condensates are on pathway for nuclear degradation, hence stabilising them accelerates the AR degradation. Preliminary experiments indicate scenario A to be the more plausible although further work is needed to understand the functional consequences of EPI-001 on the AR condensates.

Further work should also be aimed at exploring the covalent reaction. In this sense, it would be important to detect the covalent addition of EPI-001 to the AR-NTD in cells and determine which are the reactive residue or residues. This would allow us to understand whether the *in vitro* recapitulation of the reaction is valid. It would also be interesting to find the conditions for the reaction to happen within the condensed state and observe whether the reaction selectivity and/or overall yield is altered.

Finally, it would also be interesting to explore the effect of EPI-001 on the full length AR *in vitro*. Dr. Mireia Pessarrodona in the laboratory of Dr. Xavier Salvatella put a large effort in trying to express and purify the full length AR-eGFP in insect and also mammalian cells without much success. Perhaps this should be carried out in collaboration with other laboratories.

The experiments conducted in chapter 3 represent, to our knowledge, one of the first studies addressing how small-molecules can inhibit a disordered target via modulating its condensate formation. As the final objective for this thesis we wanted to exploit these findings to rationally design and synthesise improved EPI-001 analogues based on their ability to modulate the AR-NTD condensation propensity (chapter 4).

Bibliography

- (1) Fung, H. Y. J.; Birol, M.; Rhoades, E. IDPs in Macromolecular Complexes: The Roles of Multivalent Interactions in Diverse Assemblies. *Curr. Opin. Struct. Biol.* **2018**, *49*, 36–43.
- (2) Martin, E. W.; Mittag, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* **2018**, *57* (17), 2478–2487.
- (3) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer Physics of Intracellular Phase Transitions. *Nat. Phys.* **2015**, *11*, 899.
- (4) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaître, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 1–12.
- (5) 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.; Baldwin, A. J. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **2015**, *57* (5), 936–947.
- (6) Alberti, S.; Hyman, A. A. Are Aberrant Phase Transitions a Driver of Cellular Aging? *Bioessays* **2016**, *38* (10), 959–968.
- (7) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176* (3), 419–434.
- (8) Harmon, T. S.; Holehouse, A. S.; Rosen, M. K.; Pappu, R. V. Intrinsically Disordered Linkers Determine the Interplay between Phase Separation and Gelation in Multivalent Proteins. *Elife* **2017**, *6*, 1–31.
- (9) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-like Domains. *Science* **2020**, *367* (6478), 694–699.
- (10) Peran, I.; Mittag, T. Molecular Structure in Biomolecular Condensates. *Curr. Opin. Struct. Biol.* **2019**, *60*, 17–26.
- (11) Murray, D. T.; Kato, M.; Lin, Y.; Thurber, K. R.; Hung, I.; McKnight, S. L.; Tycko, R. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **2017**, *171* (3), 615–627.
- (12) Gui, X.; Luo, F.; Li, Y.; Zhou, H.; Qin, Z.; Liu, Z.; Gu, J.; Xie, M.; Zhao, K.; Dai, B.; Shin, W. S.; He, J.; He, L.; Jiang, L.; Zhao, M.; Sun, B.; Li, X.; Liu, C.; Li, D. Structural Basis for Reversible Amyloids of hnRNPA1 Elucidates Their Role in Stress Granule Assembly. *Nat. Commun.* **2019**, *10* (2006), 1–12.
- (13) Conicella, A. E.; Zerze, G. H.; Mittal, J.; Fawzi, N. L. ALS Mutations Disrupt Phase Separation Mediated by α -Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* **2016**, *24* (9), 1537–1549.
- (14) Bolognesi, B.; Faure, A. J.; Seuma, M.; Schmiedel, J. M.; Tartaglia, G. G.; Lehner, B. The Mutational Landscape of a Prion-like Domain. *Nat. Commun.* **2019**, *10* (4162), 1–12.
- (15) Murthy, A. C.; Dignon, G. L.; Kan, Y.; Zerze, G. H.; Parekh, S. H.; Mittal, J.; Fawzi, N. L. Molecular Interactions Underlying Liquid-Liquid Phase Separation of the FUS Low-Complexity Domain. *Nat. Struct. Mol. Biol.* **2019**, *26* (7), 637–648.
- (16) Brady, J. P.; Farber, P. J.; Sekhar, A.; Lin, Y.-H.; Huang, R.; Bah, A.; Nott, T. J.; Chan, H. S.; Baldwin, A. J.; Forman-Kay, J. D.; Kay, L. E. Structural and Hydrodynamic Properties of an Intrinsically Disordered Region of a Germ Cell-Specific Protein on Phase Separation. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (39), 8194–8203.
- (17) Alberti, S. Phase Separation in Biology. *Curr. Biol.* **2017**, *27*, 1097–1101.

- (18) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular Condensates: Organizers of Cellular Biochemistry. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (5), 285–298.
- (19) Holehouse, A. S.; Pappu, R. V. Functional Implications of Intracellular Phase Transitions. *Biochemistry* **2018**, *57* (17), 2415–2423.
- (20) Sheu-Gruttadauria, J.; MacRae, I. J. Phase Transitions in the Assembly and Function of Human miRISC. *Cell* **2018**, *173* (4), 946–957.
- (21) Yasuda, S.; Tsuchiya, H.; Kaiho, A.; Guo, Q.; Ikeuchi, K.; Endo, A.; Arai, N.; Ohtake, F.; Murata, S.; Inada, T.; Baumeister, W.; Fernández-Busnadiego, R.; Tanaka, K.; Saeki, Y. Stress- and Ubiquitylation-Dependent Phase Separation of the Proteasome. *Nature* **2020**, *578*, 296–300.
- (22) Case, L. B.; Zhang, X.; Ditlev, J. A.; Rosen, M. K. Stoichiometry Controls Activity of Phase-Separated Clusters of Actin Signaling Proteins. *Science* **2019**, *363* (6431), 1093–1097.
- (23) Franzmann, T. M.; Jahnel, M.; Pozniakovsky, A.; Mahamid, J.; Holehouse, A. S.; Nüske, E.; Richter, D.; Baumeister, W.; Grill, S. W.; Pappu, R. V.; Hyman, A. A.; Alberti, S. Phase Separation of a Yeast Prion Protein Promotes Cellular Fitness. *Science* **2018**, *359* (6371), 1–8.
- (24) Zhang, H.; Elbaum-Garfinkle, S.; Langdon, E. M.; Taylor, N.; Occhipinti, P.; Bridges, A. A.; Brangwynne, C. P.; Gladfelter, A. S. RNA Controls PolyQ Protein Phase Transitions. *Mol. Cell* **2015**, *60* (2), 220–230.
- (25) 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.; Best, R. B.; Cole, R. N.; Mittal, J.; Shewmaker, F.; Fawzi, N. L. Phosphorylation of the FUS Low-Complexity Domain Disrupts Phase Separation, Aggregation, and Toxicity. *EMBO J.* **2017**, *36* (20), 2951–2967.
- (26) Guillén-Boixet, J.; Buzon, V.; Salvatella, X.; Méndez, R. CPEB4 Is Regulated during Cell Cycle by ERK2/Cdk1-Mediated Phosphorylation and Its Assembly into Liquid-like Droplets. *Elife* **2016**, *5*, 1–26.
- (27) Wei, M.-T.; Elbaum-Garfinkle, S.; Holehouse, A. S.; Chen, C. C.-H.; Feric, M.; Arnold, C. B.; Priestley, R. D.; Pappu, R. V.; Brangwynne, C. P. Phase Behaviour of Disordered Proteins Underlying Low Density and High Permeability of Liquid Organelles. *Nat. Chem.* **2017**, *9* (11), 1118–1125.
- (28) Klosin, A.; Oltsch, F.; Harmon, T.; Honigmann, A.; Jülicher, F.; Hyman, A. A.; Zechner, C. Phase Separation Provides a Mechanism to Reduce Noise in Cells. *Science* **2020**, *367* (6476), 464–468.
- (29) Shin, Y.; Brangwynne, C. P. Liquid Phase Condensation in Cell Physiology and Disease. *Science* **2017**, *357* (6357), 1–11.
- (30) Bouchard, J. J.; Otero, J. H.; Scott, D. C.; Szulc, E.; Martin, E. W.; Sabri, N.; Granata, D.; Marzahn, M. R.; Lindorff-Larsen, K.; Salvatella, X.; Schulman, B. A.; Mittag, T. Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. *Mol. Cell* **2018**, *72* (1), 19–36.
- (31) Wippich, F.; Bodenmiller, B.; Trajkovska, M. G.; Wanka, S.; Aebersold, R.; Pelkmans, L. Dual Specificity Kinase DYRK3 Couples Stress Granule Condensation/dissolution to mTORC1 Signaling. *Cell* **2013**, *152* (4), 791–805.
- (32) Bergeron-Sandoval, L.-P.; Michnick, S. W. Mechanics, Structure and Function of Biopolymer Condensates. *J. Mol. Biol.* **2018**, *430* (23), 4754–4761.
- (33) Schmidt, H. B.; Görlich, D. Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. *Trends Biochem. Sci.* **2016**, *41* (1), 46–61.
- (34) Nott, T. J.; Craggs, T. D.; Baldwin, A. J. Membraneless Organelles Can Melt Nucleic Acid Duplexes and Act as Biomolecular Filters. *Nat. Chem.* **2016**, *8* (6), 569–575.
- (35) Hnisz, D.; Shrinivas, K.; Young, R. A.; Chakraborty, A. K.; Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **2017**, *169* (1), 13–23.
- (36) Wheeler, R. J.; Hyman, A. A. Controlling Compartmentalization by Non-Membrane-Bound Organelles. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2018**, *373* (1747), 1–9.

- (37) Banani, S. F.; Rice, A. M.; Peeples, W. B.; Lin, Y.; Jain, S.; Parker, R.; Rosen, M. K. Compositional Control of Phase-Separated Cellular Bodies. *Cell* **2016**, *166* (3), 651–663.
- (38) Lin, Y.; Currie, S. L.; Rosen, M. K. Intrinsically Disordered Sequences Enable Modulation of Protein Phase Separation through Distributed Tyrosine Motifs. *J. Biol. Chem.* **2017**, *292* (46), 19110–19120.
- (39) Posey, A. E.; Holehouse, A. S.; Pappu, R. V. Phase Separation of Intrinsically Disordered Proteins. *Methods Enzymol.* **2018**, *611*, 1–30.
- (40) Patel, A.; Malinowska, L.; Saha, S.; Wang, J.; Alberti, S.; Krishnan, Y.; Hyman, A. A. ATP as a Biological Hydrotrope. *Science* **2017**, *356* (6339), 753–756.
- (41) Feric, M.; Vaidya, N.; Harmon, T. S.; Mitrea, D. M.; Zhu, L.; Richardson, T. M.; Kriwacki, R. W.; Pappu, R. V.; Brangwynne, C. P. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **2016**, *165* (7), 1686–1697.
- (42) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoyanov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M.; Pozniakovski, A.; Poser, I.; Maghelli, N.; Royer, L. A.; Weigert, M.; Myers, E. W.; Grill, S.; Drechsel, D.; Hyman, A. A.; Alberti, S. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162* (5), 1066–1077.
- (43) Nedelsky, N. B.; Taylor, J. P. Bridging Biophysics and Neurology: Aberrant Phase Transitions in Neurodegenerative Disease. *Nat. Rev. Neurol.* **2019**, *15* (5), 272–286.
- (44) Heinrich, B. S.; Maliga, Z.; Stein, D. A.; Hyman, A. A.; Whelan, S. P. J. Phase Transitions Drive the Formation of Vesicular Stomatitis Virus Replication Compartments. *MBio* **2018**, *9* (5), 1–10.
- (45) Alberti, S.; Dormann, D. Liquid–Liquid Phase Separation in Disease. *Annu. Rev. Genet.* **2019**, *53*, 171–194.
- (46) Alami, N. H.; Smith, R. B.; Carrasco, M. A.; Williams, L. A.; Winborn, C. S.; Han, S. S. W.; Kiskinis, E.; Winborn, B.; Freibaum, B. D.; Kanagaraj, A.; Clare, A. J.; Badders, N. M.; Bilican, B.; Chaum, E.; Chandran, S.; Shaw, C. E.; Eggan, K. C.; Maniatis, T.; Paul Taylor, J. Axonal Transport of TDP-43 mRNA Granules Is Impaired by ALS-Causing Mutations. *Neuron* **2014**, *81* (3), 536–543.
- (47) Gopal, P. P.; Nirschl, J. J.; Klinman, E.; Holzbaur, E. L. F. Amyotrophic Lateral Sclerosis-Linked Mutations Increase the Viscosity of Liquid-like TDP-43 RNP Granules in Neurons. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (12), 2466–2475.
- (48) Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A. P.; Kim, H. J.; Mittag, T.; Taylor, J. P. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **2015**, *163* (1), 123–133.
- (49) Kim, H. J.; Kim, N. C.; Wang, Y.-D.; Scarborough, E. A.; Moore, J.; Diaz, Z.; MacLea, K. S.; Freibaum, B.; Li, S.; Molliex, A.; Kanagaraj, A. P.; Carter, R.; Boylan, K. B.; Wojtas, A. M.; Rademakers, R.; Pinkus, J. L.; Greenberg, S. A.; Trojanowski, J. Q.; Traynor, B. J.; Smith, B. N.; Topp, S.; Gkazi, A.-S.; Miller, J.; Shaw, C. E.; Kottlors, M.; Kirschner, J.; Pestronk, A.; Li, Y. R.; Ford, A. F.; Gitler, A. D.; Benatar, M.; King, O. D.; Kimonis, V. E.; Ross, E. D.; Weihl, C. C.; Shorter, J.; Taylor, J. P. Mutations in Prion-like Domains in hnRNPA2B1 and hnRNPA1 Cause Multisystem Proteinopathy and ALS. *Nature* **2013**, *495* (7442), 467–473.
- (50) Mackenzie, I. R.; Nicholson, A. M.; Sarkar, M.; Messing, J.; Purice, M. D.; Pottier, C.; Annu, K.; Baker, M.; Perkerson, R. B.; Kurti, A.; Matchett, B. J.; Mittag, T.; Temirov, J.; Hsiung, G.-Y. R.; Krieger, C.; Murray, M. E.; Kato, M.; Fryer, J. D.; Petrucelli, L.; Zinman, L.; Weintraub, S.; Mesulam, M.; Keith, J.; Zivkovic, S. A.; Hirsch-Reinshagen, V.; Roos, R. P.; Züchner, S.; Graff-Radford, N. R.; Petersen, R. C.; Caselli, R. J.; Wszolek, Z. K.; Finger, E.; Lippa, C.; Lacomis, D.; Stewart, H.; Dickson, D. W.; Kim, H. J.; Rogaeva, E.; Bigio, E.; Boylan, K. B.; Taylor, J. P.; Rademakers, R. TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics. *Neuron* **2017**, *95* (4), 808–816.

- (51) Hofweber, M.; Hutten, S.; Bourgeois, B.; Spreitzer, E.; Niedner-Boblenz, A.; Schifferer, M.; Ruepp, M.; Simons, M.; Niessing, D.; Madl, T.; Dormann, D. Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* **2018**, *173* (3), 706–719.
- (52) Qamar, S.; Wang, G.; Randle, S. J.; Ruggeri, F. S.; Varela, J. A.; Lin, J. Q.; Phillips, E. C.; Miyashita, A.; Williams, D.; Ströhl, F.; Meadows, W.; Ferry, R.; Dardov, V. J.; Tartaglia, G. G.; Farrer, L. A.; Kaminski Schierle, G. S.; Kaminski, C. F.; Holt, C. E.; Fraser, P. E.; Schmitt-Ulms, G.; Klenerman, D.; Knowles, T.; Vendruscolo, M.; St George-Hyslop, P. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation- π Interactions. *Cell* **2018**, *173* (3), 720–734.
- (53) Ambadipudi, S.; Biernat, J.; Riedel, D.; Mandelkow, E.; Zweckstetter, M. Liquid-Liquid Phase Separation of the Microtubule-Binding Repeats of the Alzheimer-Related Protein Tau. *Nat. Commun.* **2017**, *8* (275), 1–13.
- (54) Wegmann, S.; Eftekhazadeh, B.; Tepper, K.; Zoltowska, K. M.; Bennett, R. E.; Dujardin, S.; Laskowski, P. R.; MacKenzie, D.; Kamath, T.; Commins, C.; Vanderburg, C.; Roe, A. D.; Fan, Z.; Molliex, A. M.; Hernandez-Vega, A.; Muller, D.; Hyman, A. A.; Mandelkow, E.; Taylor, J. P.; Hyman, B. T. Tau Protein Liquid-Liquid Phase Separation Can Initiate Tau Aggregation. *EMBO J.* **2018**, *37* (7), 1–21.
- (55) Peskett, T. R.; Rau, F.; O’Driscoll, J.; Patani, R.; Lowe, A. R.; Saibil, H. R. A Liquid to Solid Phase Transition Underlying Pathological Huntingtin Exon1 Aggregation. *Mol. Cell* **2018**, *70* (4), 588–601.
- (56) Aaron, P. J.; Kingston, R. E. Dynamic Condensates Activate Transcription. *Science* **2018**, *361* (6400), 329–330.
- (57) Sabari, B. R.; Dall’Agnese, A.; Boija, A.; Klein, I. A.; Coffey, E. L.; Shrinivas, K.; Abraham, B. J.; Hannett, N. M.; Zamudio, A. V.; Manteiga, J. C.; Li, C. H.; Guo, Y. E.; Day, D. S.; Schuijers, J.; Vasile, E.; Malik, S.; Hnisz, D.; Lee, T. I.; Cisse, I. I.; Roeder, R. G.; Sharp, P. A.; Chakraborty, A. K.; Young, R. A. Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control. *Science* **2018**, *361*, 1–11.
- (58) Boija, A.; Klein, I. A.; Sabari, B. R.; Dall’Agnese, A.; Coffey, E. L.; Zamudio, A. V.; Li, C. H.; Shrinivas, K.; Manteiga, J. C.; Hannett, N. M.; Abraham, B. J.; Afeyan, L. K.; Guo, Y. E.; Rimel, J. K.; Fant, C. B.; Schuijers, J.; Lee, T. I.; Taatjes, D. J.; Young, R. A. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **2018**, *175*, 1–14.
- (59) Mullard, A. Biomolecular Condensates Pique Drug Discovery Curiosity. *Nat. Rev. Drug Discov.* **2019**, *18*, 324–326.
- (60) Wheeler, R. J. Therapeutics - How to Treat Phase Separation-Associated Diseases. *Emerg Top Life Sci* **2020**, 1–12. Doi: <https://doi.org/10.1042/ETLS20190176>.
- (61) Ryan, V. H.; Fawzi, N. L. Physiological, Pathological, and Targetable Membraneless Organelles in Neurons. *Trends Neurosci.* **2019**, *42* (10), 693–708.
- (62) Mateju, D.; Franzmann, T. M.; Patel, A.; Kopach, A.; Boczek, E. E.; Maharana, S.; Lee, H. O.; Carra, S.; Hyman, A. A.; Alberti, S. An Aberrant Phase Transition of Stress Granules Triggered by Misfolded Protein and Prevented by Chaperone Function. *EMBO J.* **2017**, *36* (12), 1669–1687.
- (63) Liu, Z.; Zhang, S.; Gu, J.; Tong, Y.; Li, Y.; Gui, X.; Long, H.; Wang, C.; Zhao, C.; Lu, J.; He, L.; Li, Y.; Liu, Z.; Li, D.; Liu, C. Hsp27 Chaperones FUS Phase Separation under the Modulation of Stress-Induced Phosphorylation. *Nat. Struct. Mol. Biol.* **2020**, *27* (4), 363–372.
- (64) Burslem, G. M.; Crews, C. M. Small-Molecule Modulation of Protein Homeostasis. *Chem. Rev.* **2017**, *117* (17), 11269–11301.
- (65) Raina, K.; Crews, C. M. Targeted Protein Knockdown Using Small Molecule Degraders. *Curr. Opin. Chem. Biol.* **2017**, *39*, 46–53.
- (66) Wolozin, B.; Ivanov, P. Stress Granules and Neurodegeneration. *Nat. Rev. Neurosci.* **2019**, *20*

- (11), 649–666.
- (67) Gao, X.; Jiang, L.; Gong, Y.; Chen, X.; Ying, M.; Zhu, H.; He, Q.; Yang, B.; Cao, J. Stress Granule: A Promising Target for Cancer Treatment. *Br. J. Pharmacol.* **2019**, *176* (23), 4421–4433.
- (68) Fang, M. Y.; Markmiller, S.; Vu, A. Q.; Javaherian, A.; Dowdle, W. E.; Jolivet, P.; Bushway, P. J.; Castello, N. A.; Baral, A.; Chan, M. Y.; Linsley, J. W.; Linsley, D.; Mercola, M.; Finkbeiner, S.; Lecuyer, E.; Lewcock, J. W.; Yeo, G. W. Small-Molecule Modulation of TDP-43 Recruitment to Stress Granules Prevents Persistent TDP-43 Accumulation in ALS/FTD. *Neuron* **2019**, *103* (5), 802–819.
- (69) Yoshizawa, T.; Ali, R.; Jiou, J.; Fung, H. Y. J.; Burke, K. A.; Kim, S. J.; Lin, Y.; Peeples, W. B.; Saltzberg, D.; Soniat, M.; Baumhardt, J. M.; Oldenbourg, R.; Sali, A.; Fawzi, N. L.; Rosen, M. K.; Chook, Y. M. Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. *Cell* **2018**, *173* (3), 693–705.
- (70) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508* (7496), 331–339.
- (71) Posey, A. E.; Ruff, K. M.; Harmon, T. S.; Crick, S. L.; Li, A.; Diamond, M. I.; Pappu, R. V. Profilin Reduces Aggregation and Phase Separation of Huntingtin N-Terminal Fragments by Preferentially Binding to Soluble Monomers and Oligomers. *J. Biol. Chem.* **2018**, *293* (10), 3734–3746.
- (72) Guan, J.; Zhou, W.; Hafner, M.; Blake, R. A.; Chalouni, C.; Chen, I. P.; De Bruyn, T.; Giltneane, J. M.; Hartman, S. J.; Heidersbach, A.; Houtman, R.; Ingalla, E.; Kategaya, L.; Kleinheinz, T.; Li, J.; Martin, S. E.; Modrusan, Z.; Nannini, M.; Oeh, J.; Ubhayakar, S.; Wang, X.; Wertz, I. E.; Young, A.; Yu, M.; Sampath, D.; Hager, J. H.; Friedman, L. S.; Daemen, A.; Metcalfe, C. Therapeutic Ligands Antagonize Estrogen Receptor Function by Impairing Its Mobility. *Cell* **2019**, *178* (4), 949–963.
- (73) Jain, A.; Vale, R. D. RNA Phase Transitions in Repeat Expansion Disorders. *Nature* **2017**, *546* (7657), 243–247.
- (74) Manallack, D. T.; Prankerd, R. J.; Yuriev, E.; Oprea, T. I.; Chalmers, D. K. The Significance of Acid/base Properties in Drug Discovery. *Chem. Soc. Rev.* **2013**, *42* (2), 485–496.
- (75) Wheeler, R. J.; Lee, H. O.; Poser, I.; Pal, A.; Doleman, T.; Kishigami, S.; Kour, S.; Anderson, E. N.; Marrone, L.; Murthy, A. C.; Janel, M.; Zhang, X.; Boczek, E.; Fritsch, A.; Fawzi, N. L.; Sternecker, J.; Pandey, U.; David, D. C.; Davis, B. G.; Baldwin, A. J.; Hermann, A.; Bickle, M.; Alberti, S.; Hyman, A. A. Small Molecules for Modulating Protein Driven Liquid-Liquid Phase Separation in Treating Neurodegenerative Disease. *bioRxiv*, **2019**. Doi: <https://doi.org/10.1101/721001> [Preprint].
- (76) Klein, I. A.; Bojja, A.; Afeyan, L. K.; Hawken, S. W.; Fan, M.; Dall’Agnese, A.; Oksuz, O.; Henninger, J. E.; Shrinivas, K.; Sabari, B. R.; Sagi, I.; Clark, V. E.; Platt, J. M.; Kar, M.; McCall, P. M.; Zamudio, A. V.; Manteiga, J. C.; Coffey, E. L.; Li, C. H.; Hannett, N. M.; Guo, Y. E.; Decker, T.-M.; Lee, T. I.; Zhang, T.; Weng, J.-K.; Taatjes, D. J.; Chakraborty, A.; Sharp, P. A.; Chang, Y. T.; Hyman, A. A.; Gray, N. S.; Young, R. A. Partitioning of Cancer Therapeutics in Nuclear Condensates. *Science* **2020**, *368* (6497), 1386–1392.
- (77) Cisse, I. I.; Izeddin, I.; Causse, S. Z.; Boudarene, L.; Senecal, A.; Muresan, L.; Dugast-Darzacq, C.; Hajj, B.; Dahan, M.; Darzacq, X. Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science* **2013**, *341* (6146), 664–667.
- (78) Lu, H.; Yu, D.; Hansen, A. S.; Ganguly, S.; Liu, R.; Heckert, A.; Darzacq, X.; Zhou, Q. Phase-Separation Mechanism for C-Terminal Hyperphosphorylation of RNA Polymerase II. *Nature* **2018**, *558* (7709), 318–323.
- (79) Boehning, M.; Dugast-Darzacq, C.; Rankovic, M.; Hansen, A. S.; Yu, T.; Marie-Nelly, H.; McSwiggen, D. T.; Kokic, G.; Dailey, G. M.; Cramer, P.; Darzacq, X.; Zweckstetter, M. RNA Polymerase II Clustering through Carboxy-Terminal Domain Phase Separation. *Nat. Struct. Mol. Biol.* **2018**, *25*, 833–840.

- (80) Guo, Y. E.; Manteiga, J. C.; Henninger, J. E.; Sabari, B. R.; Dall’Agnese, A.; Hannett, N. M.; Spille, J.-H.; Afeyan, L. K.; Zamudio, A. V.; Shrinivas, K.; Abraham, B. J.; Boija, A.; Decker, T.-M.; Rimel, J. K.; Fant, C. B.; Lee, T. I.; Cisse, I. I.; Sharp, P. A.; Taatjes, D. J.; Young, R. A. Pol II Phosphorylation Regulates a Switch between Transcriptional and Splicing Condensates. *Nature* **2019**, *572* (7770), 543–548.
- (81) Cho, W.-K.; Jayanth, N.; English, B. P.; Inoue, T.; Andrews, J. O.; Conway, W.; Grimm, J. B.; Spille, J.-H.; Lavis, L. D.; Lionnet, T.; Cisse, I. I. RNA Polymerase II Cluster Dynamics Predict mRNA Output in Living Cells. *Elife* **2016**, *5*, 1–31.
- (82) Cho, W.-K.; Spille, J.-H.; Hecht, M.; Lee, C.; Li, C.; Grube, V.; Cisse, I. I. Mediator and RNA Polymerase II Clusters Associate in Transcription-Dependent Condensates. *Science* **2018**, *361* (6400), 412–415.
- (83) Nair, S. J.; Yang, L.; Meluzzi, D.; Oh, S.; Yang, F.; Friedman, M. J.; Wang, S.; Suter, T.; Alshareedah, I.; Gamliel, A.; Ma, Q.; Zhang, J.; Hu, Y.; Tan, Y.; Ohgi, K. A.; Jayani, R. S.; Banerjee, P. R.; Aggarwal, A. K.; Rosenfeld, M. G. Phase Separation of Ligand-Activated Enhancers Licenses Cooperative Chromosomal Enhancer Assembly. *Nat. Struct. Mol. Biol.* **2019**, *26* (3), 193–203.
- (84) Wei, M.-T.; Chang, Y.-C.; Shimobayashi, S. F.; Shin, Y.; Brangwynne, C. P. Nucleated Transcriptional Condensates Amplify Gene Expression. *bioRxiv*, **2019**. Doi: <https://doi.org/10.1101/737387> [Preprint].
- (85) Shrinivas, K.; Sabari, B. R.; Coffey, E. L.; Klein, I. A.; Boija, A.; Zamudio, A. V.; Schuijers, J.; Hannett, N. M.; Sharp, P. A.; Young, R. A.; Chakraborty, A. K. Enhancer Features That Drive Formation of Transcriptional Condensates. *Mol. Cell* **2019**, *75* (3), 549–561.
- (86) Chong, S.; Dugast-Darzacq, C.; Liu, Z.; Dong, P.; Dailey, G. M.; Cattoglio, C.; Heckert, A.; Banala, S.; Lavis, L.; Darzacq, X.; Tjian, R. Imaging Dynamic and Selective Low-Complexity Domain Interactions That Control Gene Transcription. *Science* **2018**, *361* (6400), 1–9.
- (87) Szulc, E. M. Structural Insights into “Acid Blobs and Negative Noodles” - the Androgen Receptor as a Case Study. Ph. D. Thesis, Universitat de Barcelona - IRB Barcelona, 2019.
- (88) Feldman, B. J.; Feldman, D. The Development of Androgen-Independent Prostate Cancer. *Nat. Rev. Cancer* **2001**, *1* (1), 34–45.
- (89) van Royen, M. E.; van de Wijngaart, D. J.; Cunha, S. M.; Trapman, J.; Houtsmuller, A. B. A Multi-Parameter Imaging Assay Identifies Different Stages of Ligand-Induced Androgen Receptor Activation. *Cytometry A* **2013**, *83* (9), 806–817.
- (90) van Royen, M. E.; van Cappellen, W. A.; Geverts, B.; Schmidt, T.; Houtsmuller, A. B.; Schaaf, M. J. M. Androgen Receptor Complexes Probe DNA for Recognition Sequences by Short Random Interactions. *J. Cell Sci.* **2014**, *127*, 1406–1416.
- (91) Presman, D. M.; Ganguly, S.; Schiltz, R. L.; Johnson, T. A.; Karpova, T. S.; Hager, G. L. DNA Binding Triggers Tetramerization of the Glucocorticoid Receptor in Live Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (29), 8236–8241.
- (92) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* **2012**, *51* (11), 2224–2231.
- (93) Eftekharzadeh, B.; Banduseela, V. C.; Chiesa, G.; Martínez-Cristóbal, P.; Rauch, J. N.; Nath, S. R.; Schwarz, D. M. C.; Shao, H.; Marin-Argany, M.; Di Sanza, C.; Giorgetti, E.; Yu, Z.; Pierattelli, R.; Felli, I. C.; Brun-Heath, I.; García, J.; Nebreda, Á. R.; Gestwicki, J. E.; Lieberman, A. P.; Salvatella, X. Hsp70 and Hsp40 Inhibit an Inter-Domain Interaction Necessary for Transcriptional Activity in the Androgen Receptor. *Nat. Commun.* **2019**, *10* (3562), 1–14.
- (94) De Mol, E.; Szulc, E.; Di Sanza, C.; Martínez-Cristóbal, P.; Bertoncini, C. W.; Fenwick, R. B.; Frigolé-Vivas, M.; Masín, M.; Hunter, I.; Buzón, V.; Brun-Heath, I.; García, J.; De Fabritiis, G.; Estébanez-Perpiñá, E.; McEwan, I. J.; Nebreda, Á. R.; Salvatella, X. Regulation

of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators. *Structure* **2018**, *26*, 1–8.

- (95) 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.; McEwan, I. J.; Riera, A.; Salvatella, X. EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. *ACS Chem. Biol.* **2016**, *11*, 2499–2505.
- (96) Myung, J. K.; Banuelos, 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.; Sadar, M. D. An Androgen Receptor N-Terminal Domain Antagonist for Treating Prostate Cancer. *J. Clin. Invest.* **2013**, *123* (7), 2948–2960.
- (97) De Mol, E. Structure, Dynamics and Interactions of the N-Terminal Domain of the Androgen Receptor. Ph. D. Thesis, Universitat de Barcelona - IRB Barcelona, 2014.
- (98) Stenoien, D. L.; Cummings, C. J.; Adams, H. P.; Mancini, M. G.; Patel, K.; DeMartino, G. N.; Marcelli, M.; Weigel, N. L.; Mancini, M. A. 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.* **1999**, *8* (5), 731–741.
- (99) Kumar, S.; Tyagi, R. K. Androgen Receptor Association with Mitotic Chromatin - Analysis with Introduced Deletions and Disease-Inflicting Mutations. *FEBS J.* **2012**, *279* (24), 4598–4614.
- (100) 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.; Sadar, M. D. Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor. *Cancer Cell* **2010**, *17* (6), 535–546.
- (101) An, J.; Wang, C.; Deng, Y.; Yu, L.; Huang, H. Destruction of Full-Length Androgen Receptor by Wild-Type SPOP, but Not Prostate-Cancer-Associated Mutants. *Cell Rep.* **2014**, *6* (4), 657–669.
- (102) Söderberg, O.; Gullberg, M.; Jarvius, M.; Ridderstråle, K.; Leuchowius, K.-J.; Jarvius, J.; Wester, K.; Hydbring, P.; Bahram, F.; Larsson, L.-G.; Landegren, U. Direct Observation of Individual Endogenous Protein Complexes in Situ by Proximity Ligation. *Nat. Methods* **2006**, *3* (12), 995–1000.

4. RATIONAL DESIGN, SYNTHESIS AND BIOLOGICAL EVALUATION OF NOVEL EPI-001 ANALOGUES

4.1. Background

4.1.1. CRPC treatments currently in the pipeline

Several new pipeline agents for castration-resistant prostate cancer (CRPC) are expected to be approved in the coming years. Finding new CRPC therapies is of high interest because CRPC is a clear unmet medical need and because prostate cancer (PC) and CRPC market size is very large (section 1.2).

Since 2010, new treatments for CRPC, enzalutamide (Xtandi) and abiraterone (Zytiga), have modestly improved the overall survival rates but the disease remains incurable.

New treatment options include PARP inhibitors (e.g. Lynparza), immunotherapies (e.g. Keytruda), AKT inhibitors (e.g. ipatasertib) and one radioligand (177Lu-PSMA-617) among others. These therapies target other pathways not related to the androgen receptor (AR) axis. Nonetheless, AR is still the main driver for PC and CRPC.

All current therapies directed towards the AR are antiandrogens, which target the AR ligand-binding domain (AR-LBD). However, most patients treated with antiandrogens inevitably develop resistances, like the emergence of AR splice variants (SVs) lacking the ligand-binding domain. SVs are present in *ca* 30% of the 5-year diagnosed prevalent cases of CRPC in the 8 major pharmaceutical markets.

Other approaches targeting AR include:

4.1.1.1. Degradation of AR

Oncternal Therapeutics (previously GTx pharma) will soon initiate Phase I clinical trials of its lead selective androgen receptor degrader small molecule (SARD). SARD molecules were first discovered by the Narayanan laboratory ¹. SARDs bind to the AR-LBD and AR-NTD and cause AR degradation. Other other companies also attempt to find AR degraders. Arvinas is developing ARV-110, a small molecule degrader based on the proteolysis targeting chimera technology that degrades the AR for the treatment of mCRPC ². ARV-110 is currently in Phase I/II clinical trials (NCT03888612).

Of notice, there is still no approved therapy based on protein degradation.

4.1.1.2. Targeting the N-terminal domain of AR (AR-NTD)

ESSA Pharma, founded by Dr. Marianne Sadar and Dr. Raymond Andersen, pioneered the development of AR-NTD inhibitors based on one chemical series (EPI-001 series) obtained by phenotypic screening ³. ESSA Pharma's first lead compound, EPI-506, underwent Phase I/II clinical trials that were stopped due to metabolic instability (Ralaniten, NCT02606123). A second lead EPI-7386, which is also based on EPI-001, entered clinical trials in Q2 2020 ⁴. Despite this fact, EPI-7386's potency is still low, *ca* 500 nM, making it necessary to find improved AR-NTD inhibitors.

Inhibiting the AR-NTD benefits a broader group of patients because it is present in all forms of AR, including the AR-SVs.

Our approach to target the AR-NTD is different from the aforementioned as it is based on the modulation of biomolecular condensation, a novel and unique way to find AR-NTD inhibitors and to rationally explore a different chemical space. However, it also presents some uncertainties as it is a disruptive new mechanism of action. For example (i) there are only two publications so far of small molecules targeting condensates ^{5,6} and (ii) perhaps not all traditional drug discovery rules and methods will be applicable.

To our knowledge this is the first application of rational drug design to target biomolecular condensates (section 4.2).

4.2. Experimental Results

Our aim was to build on the acquired knowledge on the ability of EPI-001 to enhance AR condensation (section 3.2 and 3.3) to find new and improved inhibitors that target the androgen receptor N-terminal domain (AR-NTD) for the treatment of castration-resistant prostate cancer (CRPC).

In order to achieve our aim, we hypothesised that the new inhibitors should have two abilities:

1. Inhibitors should be able to partition into the proteinaceous hydrophobic dense phase

The AR-NTD condensation follows a lower critical solution temperature (LCST) behaviour, which is driven by the entropic gain upon water release when non-polar motifs are desolvated. Therefore, AR-NTD condensates are depleted in water (e.g. confer a more hydrophobic environment than the bulk solution outside the condensate). We hypothesised

-

that the more hydrophobic the compound, the more it would partition into the condensates and the more it would inhibit the AR-NTD. This would be in line with some findings described in the literature (section 2.1.4) such as:

- a) Niphatenone B presents an IC_{50} of 6 μ M whereas the derivative without the long alkane chain is inactive (**Fig. 2.1.4.1**)⁷.
- b) Introduction of an iodo atom in *ortho* position to the epichlorohydrin group of EPI-002 improved the activity of EPI-002 by one order of magnitude (**Fig. 2.1.4.3**)⁸.

2. Inhibitors should be able to promote AR condensation, probably by irreversibly modifying the AR-NTD

The second assumption relies, first, on the fact that 3 out of the 4 families of compounds described in the literature to inhibit the AR transcriptional activity by binding to the AR-NTD contain a reactive warhead^{3,9-11}. Second, it relies on our finding that covalently attached EPI-001 enhances the AR-NTD condensation to a greater extent than through non-covalent binding. For these reasons, we hypothesised that a covalent warhead is required.

In this sense, EPI-001 would be one of a kind as i) it is hydrophobic enough to partition into the hydrophobic dense phase and ii) it can react with the AR-NTD promoting the ability of the AR-NTD to condensate. This would clearly differentiate EPI-001 from its inactive analogue BADGE, which is less hydrophobic and does not bear the electrophilic epichlorohydrin warhead.

4.2.1. Design and synthesis of novel EPI-001 analogues

4.2.1.1. Optimisation of the synthetic route

To explore the structure-activity relationship (SAR) around the EPI-001 structure we had to implement a more versatile chemical route than the one described for EPI-001¹². For this we took advantage of an ongoing project in the laboratory of Dr. Antoni Riera which explored the effect of the EPI-001 linker (**Fig. 4.2.1.1a**). The aim was to expand the linker from 1-atom to 2-atoms. This allowed to, on the one hand, escape EPI-001 intellectual property protection and, on the other, check the best geometry between the two nearly symmetrical parts of the EPI-001 molecule: linear, *cis* conformation, *trans* conformation and flexible (**Fig. 4.2.1.1b**).

4. Rational design, synthesis and biological evaluation of novel EPI-001 analogues

-

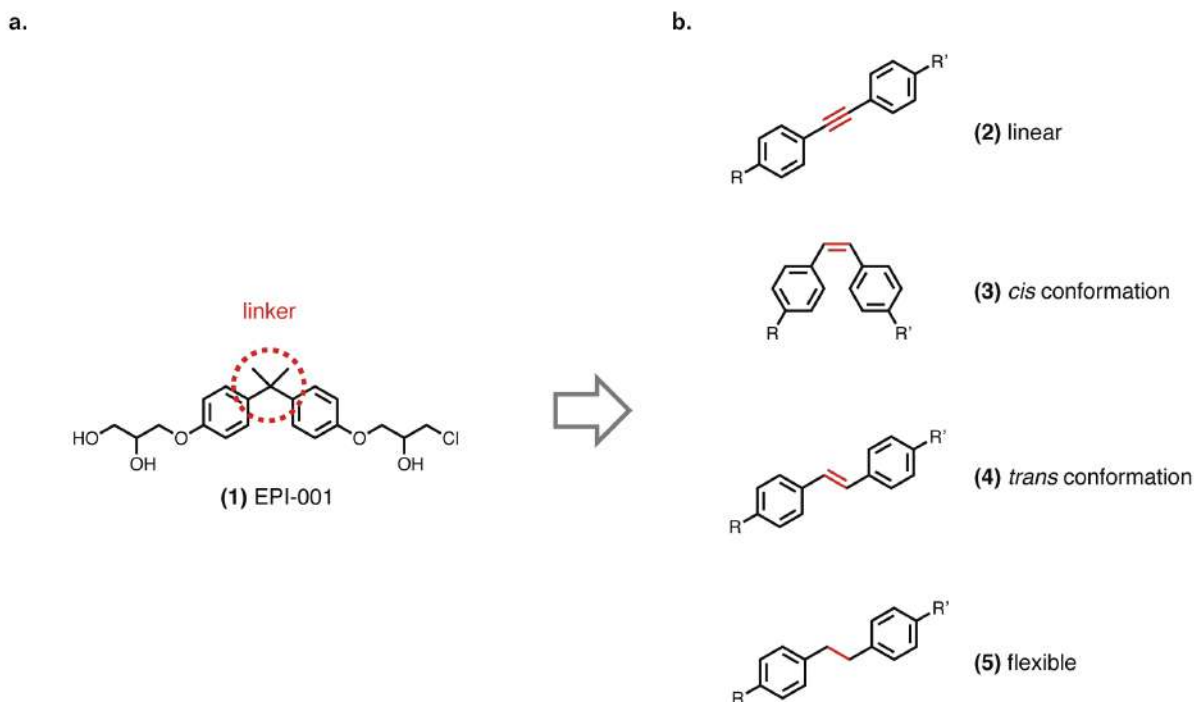


Fig. 4.2.1.1. a) EPI-001 chemical structure. **b)** Linker-expanded analogues.

The initial route was designed to access an advanced intermediate (**10**) that is common to all the linker-expanded analogues (**Fig. 4.2.1.2**). However, the route presented several drawbacks: 1) the first reaction of the route, the glycidol addition, was not selective and resulted in a poor 33% yield. Moreover, the need for column purification hampered the scale-up of the route; 2) the key intermediate (**10**) was achieved by means of a Colvin rearrangement, which requires the use of the very dangerous (trimethylsilyl)diazomethane reagent; 3) the core structure (the kind of aromatic rings and its substituents) was set from the start of the route. This hampered the synthesis of a large family of analogues as each variation of the core to enable SAR studies required to perform all steps of the route; and 4) the route was not stereospecific.

4. Rational design, synthesis and biological evaluation of novel EPI-001 analogues

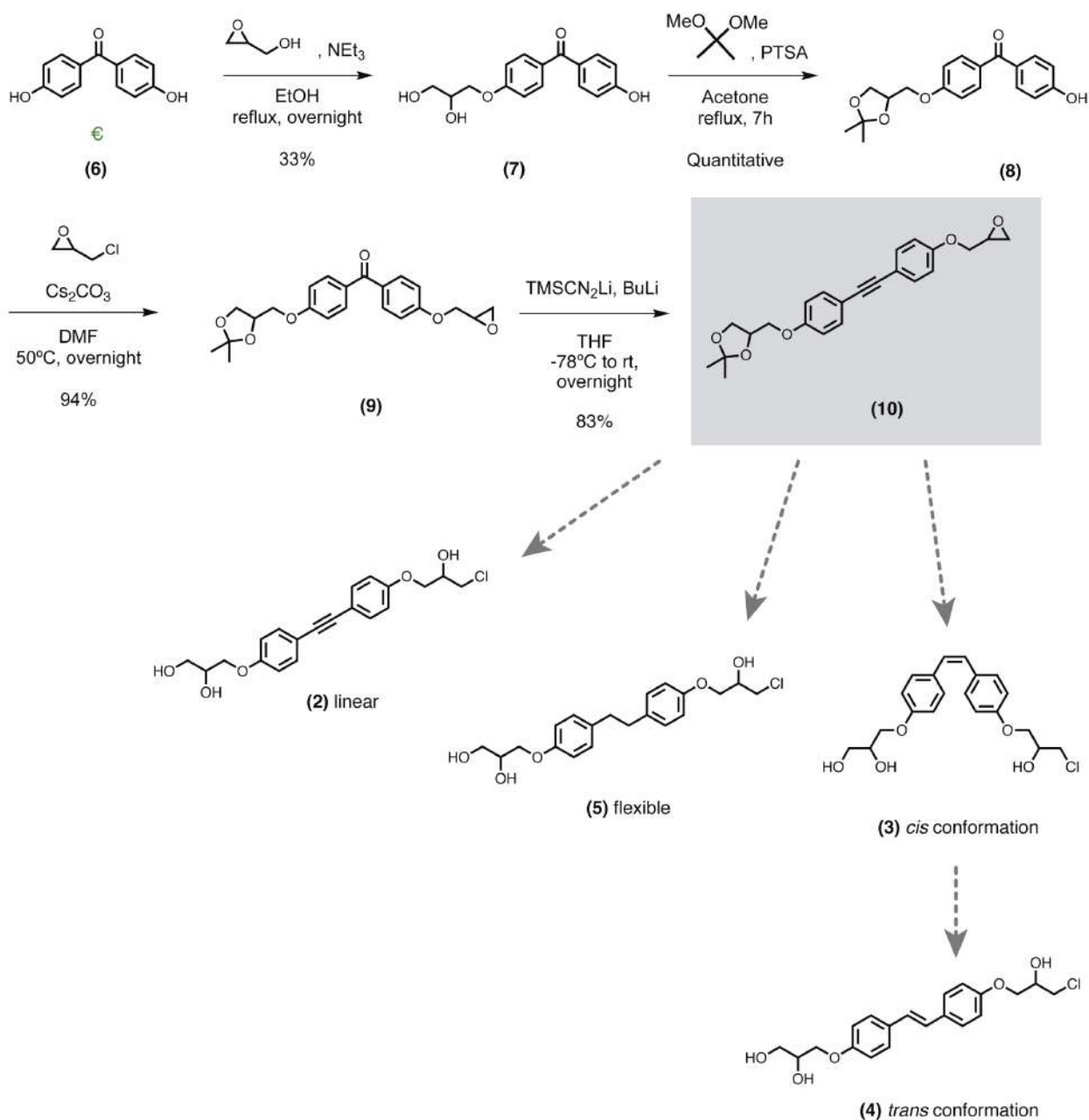


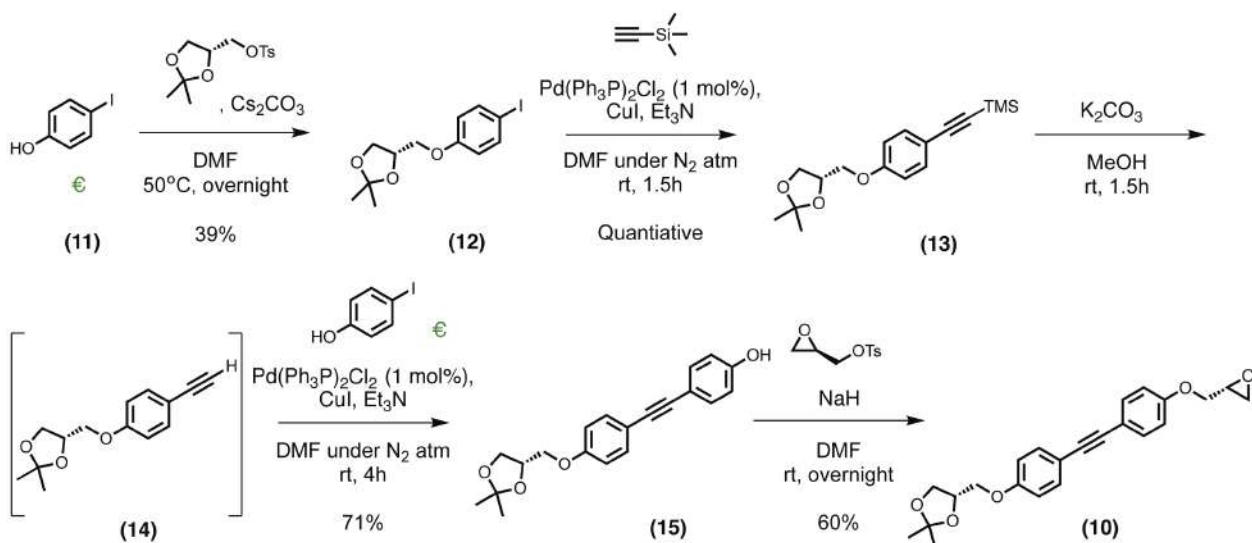
Fig. 4.2.1.2. Initial synthetic route. The commercial diol **(6)** was treated with glycidol to give the monoalkylated product **(7)** in 33% yield and protected to give the product **(8)** quantitatively. Treatment of **(8)** with epichlorohydrin yielded epoxide **(9)** in 94% yield. The linker expansion was then accomplished through the Colvin rearrangement of **(9)** to give the key alkyne intermediate **(10)** in 83% yield.

€ indicates commercially available.

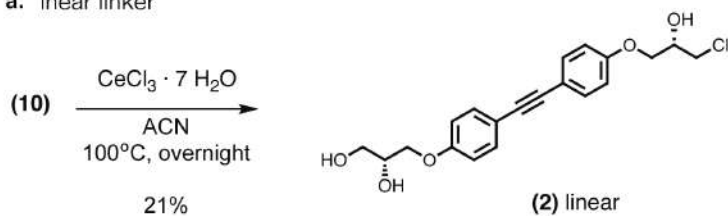
The drawbacks from the established synthetic route motivated us to optimise it before building the library of analogues (**Fig. 4.2.1.3**). The new strategy also aimed at synthesising an advanced intermediate from which to derive all the four linker-expanded analogues. Moreover, the new route was highly modular, stereospecific and didn't involve the use of hazardous reagents. Since (2R,20S)-EPI (named EPI-002) is reported to be slightly more active than the other EPI-001 stereoisomers¹³, we synthesised all the analogues with this stereochemistry.

4. Rational design, synthesis and biological evaluation of novel EPI-001 analogues

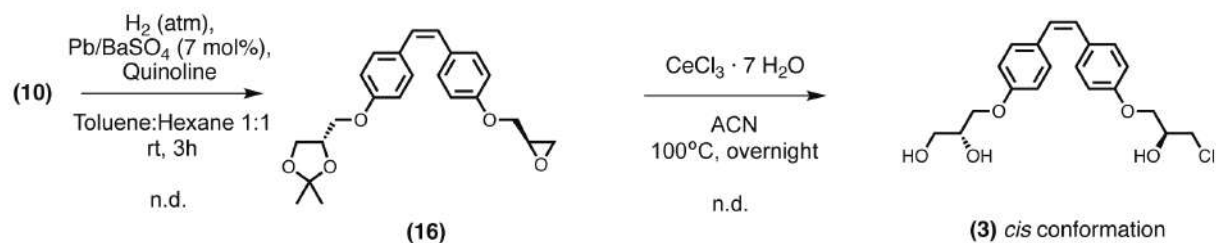
-



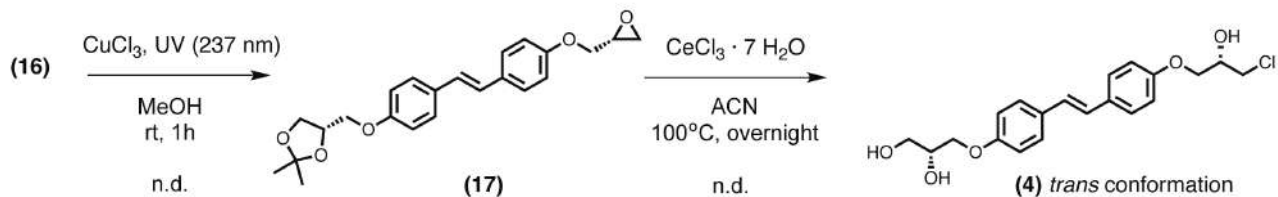
a. linear linker



b. cis linker



c. trans linker



4. Rational design, synthesis and biological evaluation of novel EPI-001 analogues

d. flexible linker

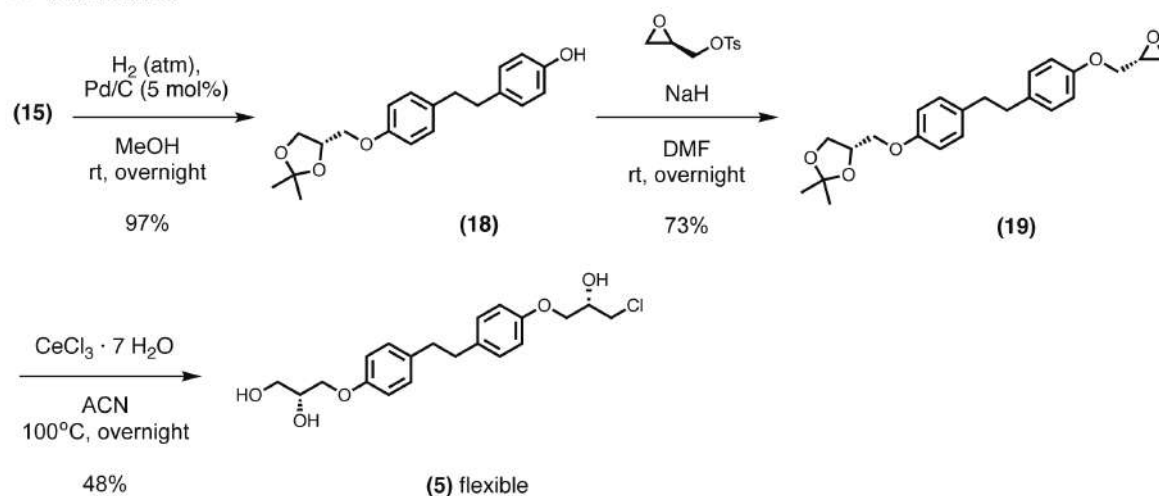


Fig. 4.2.1.3. Optimised synthetic route. The commercial 4-iodophenol (**11**) was protected to give the product (**12**) in 39% yield. The 2-carbon linker was installed incorporating trimethylsilylacetylene via a Sonogashira reaction to yield product (**13**) quantitatively. The deprotected product of (**13**) was used without isolation in a second Sonogashira reaction to introduce the second aromatic ring with an overall 71% yield. Treatment of (**15**) with tosyl glycidol afforded intermediate (**10**) in 60% yield. **a**) The linear-linker analogue (**2**) was obtained through acetal deprotection and epoxide opening of (**10**) using CeCl₃ under thermal conditions. **b**) The *cis*-linker analogue was achieved by a Lindlar hydrogenation of (**10**). Subsequent acetal deprotection and epoxide opening of (**16**) achieved the final product (**3**). **c**) The *trans*-linker analogue (**4**) was obtained by isomerisation of (**16**) and subsequent acetal deprotection and epoxide opening. **d**) The flexible-linker analogue was obtained by total hydrogenation of intermediate (**15**) in 97% yield. Treatment of (**18**) with tosyl glycidol furnished (**19**) in 73% yield. Acetal deprotection and epoxide opening of (**19**) produced the final product (**5**) in 48% yield.

n.d. indicates non-determined yield for the reactions in which the product was majoritary but not pure.

€ indicates commercially available.

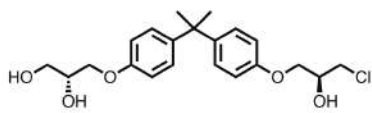
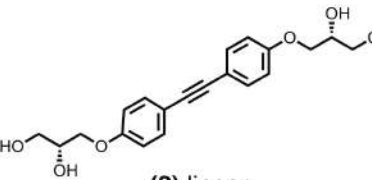
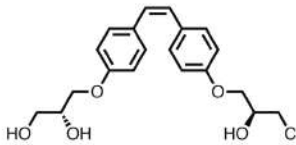
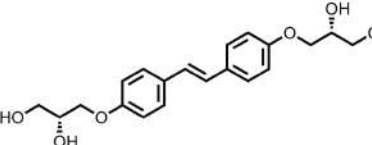
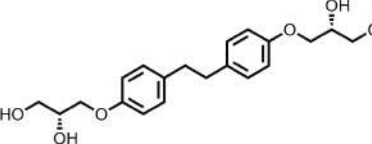
The four linker-expanded analogues were first re-synthesised to stereoselectively provide the 2R,20S stereoisomer. This work was accomplished with the help of Dr. Carolina Sánchez in the laboratory of Dr. Antoni Riera. Then, the compounds (**2**, **3**, **4** and **5**) were biologically evaluated in collaboration with the laboratory of Dr. Marianne Sadar at the University of British Columbia. The triple bond analogue was selected as the core scaffold due to its promising results in the PSA-luciferase transcriptional activity assay and the AR-dependent proliferation assay (in LNCaP cell line) (**Table 4.2.1.1**). In similar experiments, the IC₅₀ values corresponding to EPI-002 were 7.40 ± 1.46 μM in the PSA-luciferase assay and 10 μM in the AR-dependent proliferation assay¹³.

4. Rational design, synthesis and biological evaluation of novel EPI-001 analogues

-

Table. 4.2.1.1. Biological evaluation of the linker-expanded analogues.

(*) indicates that the values are not head-to-head but extracted from the literature ¹³.

	AR transcriptional activity IC ₅₀ (PSA-luc) (μM)	AR-dependent proliferation IC ₅₀ (LNCaP cell line) (μM)
 <p>(1') EPI-002</p>	7.40 ± 1.46 *	ca 10 *
 <p>(2) linear</p>	6.79 ± 0.57	4.11 ± 1.43
 <p>(3) cis conformation</p>	6.15 ± 2.41	7.82
 <p>(4) trans conformation</p>	14.74 ± 3.02	22.48
 <p>(5) flexible</p>	1.49 ± 1.62	19.75

4.2.1.2. CAM family of analogues

Once the synthetic route was optimised, we designed 8 linear EPI-001 analogues (named CAMs, **Fig.4.2.1.4**) to interrogate i) whether variations in CAM's hydrophobicity affected the ability of the compounds to modulate the AR-NTD condensation propensity and ii) whether the degree of modulation of the AR-NTD condensation was correlated with inhibition of the AR activity in cells. To achieve differences in hydrophobicity among the family of compounds, a substituent was introduced in *ortho* position to the epichlorohydrin group. This position had already been described to tolerate substituents such as an iodo atom ⁸. To evaluate the effect of the position of the

-

substituent, CAM5 and CAM6 were also prepared. The synthesis of all CAM compounds was conducted with the help of Dr. Carolina Sánchez. Finally, their hydrophobicity was experimentally determined by HPLC (Chrom logD) (**Fig. 4.2.1.4**).

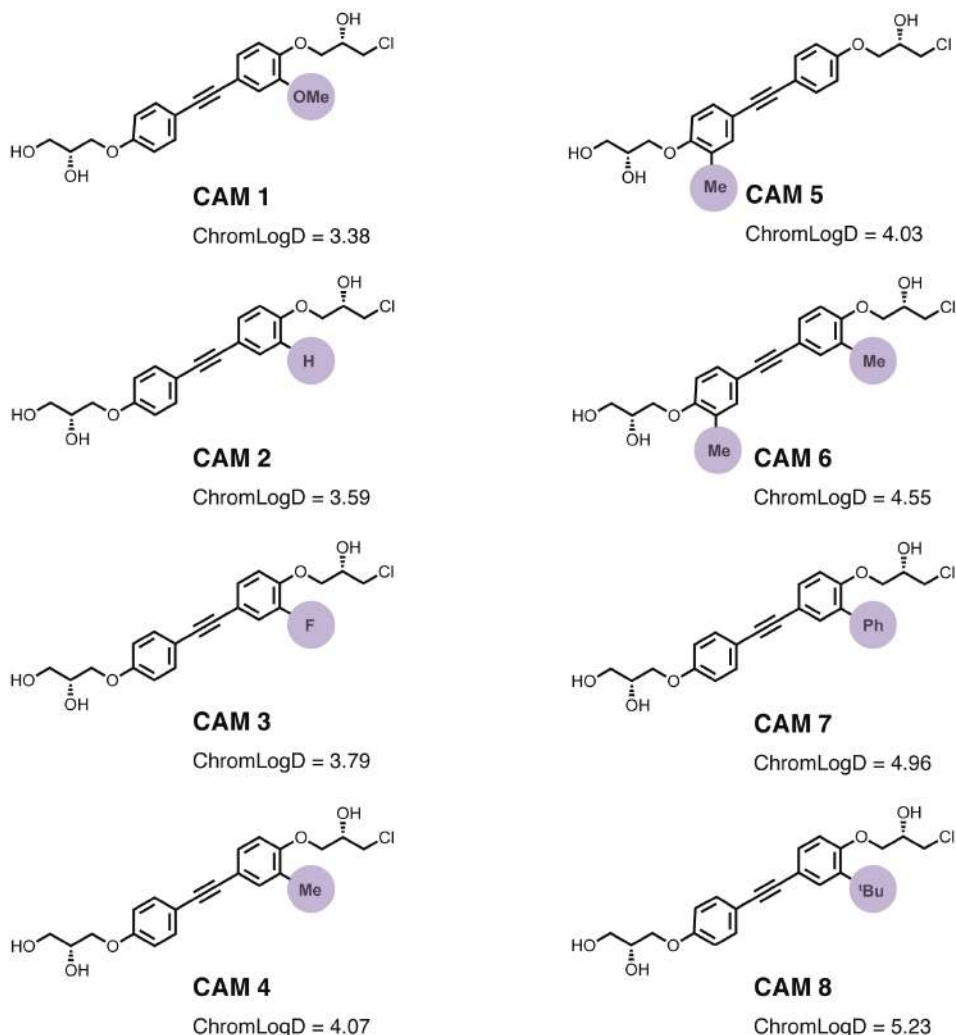


Fig. 4.2.1.4. Chemical structure of the CAM family of compounds with their hydrophobicity indicated as Chrom logD value. Hydrophobicity was experimentally measured by HPLC (high Chrom logD values indicate high hydrophobicity).

In summary, we rationally designed and synthesised a family of 8 compounds that enabled SAR studies around the structure of the triple-bond linker analogue.

4.2.2. Biological evaluation of the CAM family of analogues

In order to quantitatively rank the CAM compounds with regards to their efficacy to enhance the AR-NTD condensation *in vitro* and compare it to that of EPI-001, we determined their half-

-

maximal effective concentration (EC_{50}) using the thermal shift assay (TSA, section 3.2) (**Fig. 4.2.2.1a**). We measured the cloud point temperature (T_c) of samples at constant protein concentration (25 μM) and varying compound equivalents. Then, we calculated the percentage of shift caused by the compound in comparison to the reference shift, the one caused by 200 μM EPI-001. Next, we plotted the normalised shift versus the compound concentration. Finally, we extrapolated the compound concentration that achieved 50% of the maximum activity, the EC_{50} . We observed a correlation between the hydrophobicity and the extent to which a compound was able to modulate the AR-NTD condensation. In this regard, the more hydrophobic the compound, the lower the EC_{50} in the TSA assay. Therefore, the more hydrophobic the compound, the more it partitioned into the AR-NTD condensates (**Fig. 4.2.2.1b**). Of note, EC_{50} values should be determined in three independent replicas to be unequivocally conclusive.

a.

Compound	EC_{50} (μM)	Chrom logD
CAM1	123.1	3.38
CAM2	222.1	3.59
CAM3	64.1	3.79
CAM4	48.2	4.07
CAM5	89.3	4.03
CAM6	67.2	4.55
CAM7	43.6	4.96
CAM8	73.1	5.23

b.

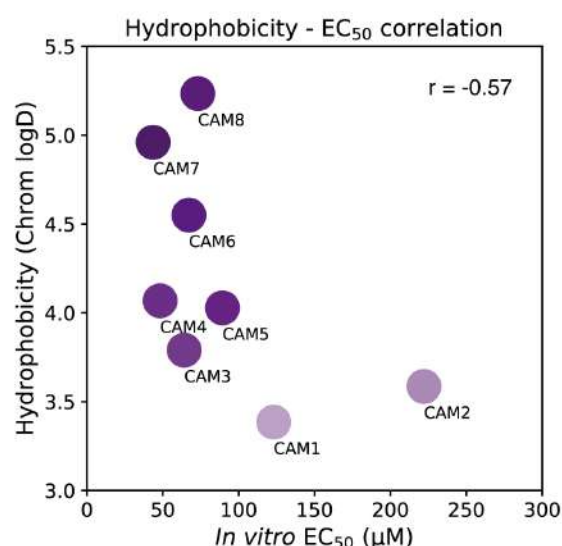


Fig. 4.2.2.1. a) EC_{50} and Chrom logD (hydrophobicity) values obtained for the CAM family of compounds.

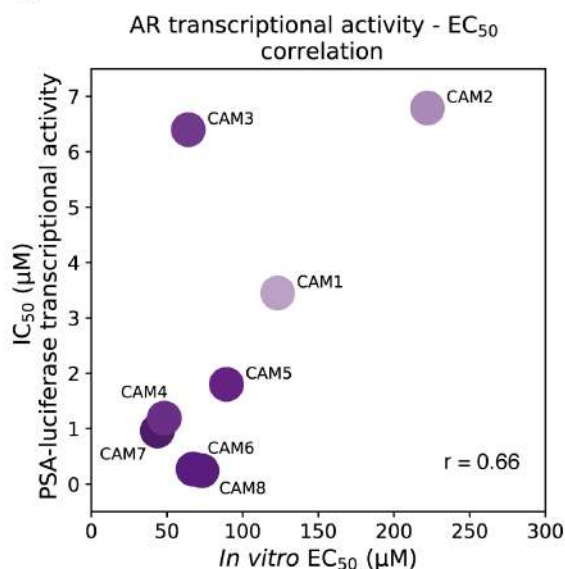
b) Pearson correlation (r) between hydrophobicity and EC_{50} for the CAM family of compounds. $r = 1$ indicates perfect positive correlation, $r = 0$ indicates no correlation, $r = -1$ indicates perfect negative correlation.

The EC_{50} data aligned with the in cell activity of these compounds. 3 compounds achieved cellular IC_{50} s one order of magnitude lower than those of EPI-002 published in the literature (IC_{50} PSA-luc = $7.40 \pm 1.46 \mu\text{M}$, IC_{50} AR-dependent proliferation = $10 \mu\text{M}$)¹³ (**Fig. 4.2.2.2**). CAM8 was one of the most lipophilic and potent inhibitors. However, due to its low solubility, CAM7 was selected as the lead compound. The cellular assays were conducted in collaboration with the laboratory of Dr. Marianne Sadar.

a.

Compound	EC ₅₀ (μM)	AR transcriptional activity IC ₅₀ (μM)	AR-dependent proliferation IC ₅₀ (μM)	AR-independent proliferation IC ₅₀ (μM)
CAM1	123.1	3.45 ± 0.55	9.97 ± 1.40	No toxicity
CAM2	222.1	6.79 ± 0.57	4.11 ± 1.43	No toxicity
CAM3	64.1	6.40 ± 0.26	5.57 ± 3.22	No toxicity
CAM4	48.2	1.19 ± 0.68	1.44 ± 0.50	Toxic at > 32.17
CAM5	89.3	1.80 ± 0.88	1.61	Toxic at > 14.97
CAM6	67.2	0.27 ± 0.07	Not tested	Not tested
CAM7	43.6	0.96 ± 0.08	1.22 ± 0.18	Toxic at > 8.28
CAM8	73.1	0.24 ± 0.04	0.91 ± 0.09	Toxic at > 8.28

b.



c.

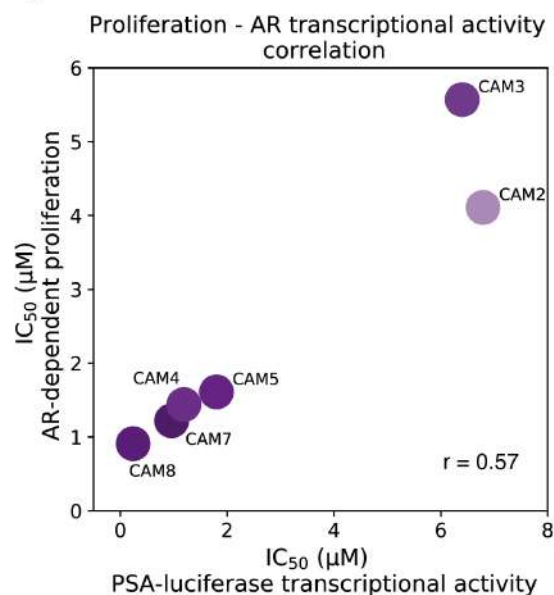


Fig. 4.2.2.2. a) CAMs *in vitro* EC₅₀ and cellular potency in the AR transcriptional assay (PSA-luciferase assay) and in the AR-dependent proliferation assay (in LNCaP cell line). Cytotoxicity was measured by means of the AR-independent proliferation assay (in PC3 cell line). Pearson correlation (r) b) between EC₅₀ and cellular potency for the CAM family of compounds and c) between cellular potency measured in the PSA-luciferase transcriptional activity assay and in the AR-dependent proliferation assay. $r = 1$ indicates perfect positive correlation, $r = 0$ indicates no correlation, $r = -1$ indicates perfect negative correlation.

Noteworthy, the site where the hydrophobic substituent was introduced did not have a substantial effect on the EC₅₀. In this regard, CAM4, bearing a methyl substituent in ring 1, and CAM5, bearing a methyl substituent in ring 2, had EC₅₀ in the same order of magnitude. Only when a methyl substituent was placed on both rings, such as in CAM6, there was a gain in hydrophobicity and increased cellular potency (Fig. 4.2.2.3).

4. Rational design, synthesis and biological evaluation of novel EPI-001 analogues

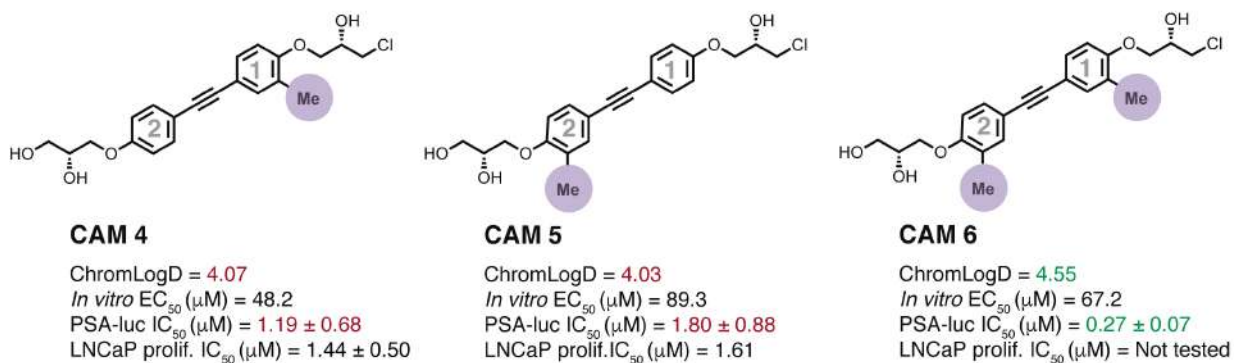


Fig. 4.2.2.3. Chemical structure of CAM4, CAM5 and CAM6 and their hydrophobicity, *in vitro* EC₅₀ and cellular potency in the AR transcriptional assay (PSA-luciferase assay) and in the AR-dependent proliferation assay (LNCaP cell line). Numbers 1 and 2 are placed to differentiate between the two aromatic rings.

Finally, we wanted to determine the *in vivo* efficacy of the lead CAM compound, CAM7, in a CRPC relevant animal model, also in collaboration with Dr. Marianne Sadar's laboratory. Unfortunately, *in vivo* experiments had to be stopped due to the poor solubility of CAM7 in the required experimental conditions. The back-up lead compound, CAM6, will be tested in Q3 2020.

In summary, stabilisation of the AR-NTD condensates improves the inhibitor potency. By applying these rational criteria we have been able to improve the potency of EPI-001 10-fold.

4.2.3. Consequences of modulating AR-NTD condensation with CAMs to AR function

Since several CAM compounds outperformed EPI-001 in the *in vitro* TSA and cellular assays, we wanted to determine whether they also had a more pronounced effect in the other cellular assays previously established to test AR condensation (section 3.2.2 and 3.2.3). These experiments were conducted with CAM7 or CAM6 depending on which was the lead compound at the time of the experiment.

First, we studied the effect of CAM7 in the AR nuclear clusters employing super-resolution stimulation emission depletion (STED) microscopy. 1 μM CAM7 promoted the formation of significantly larger AR nuclear clusters than DMSO or 10 μM EPI-001 (**Fig. 4.2.3.1**).

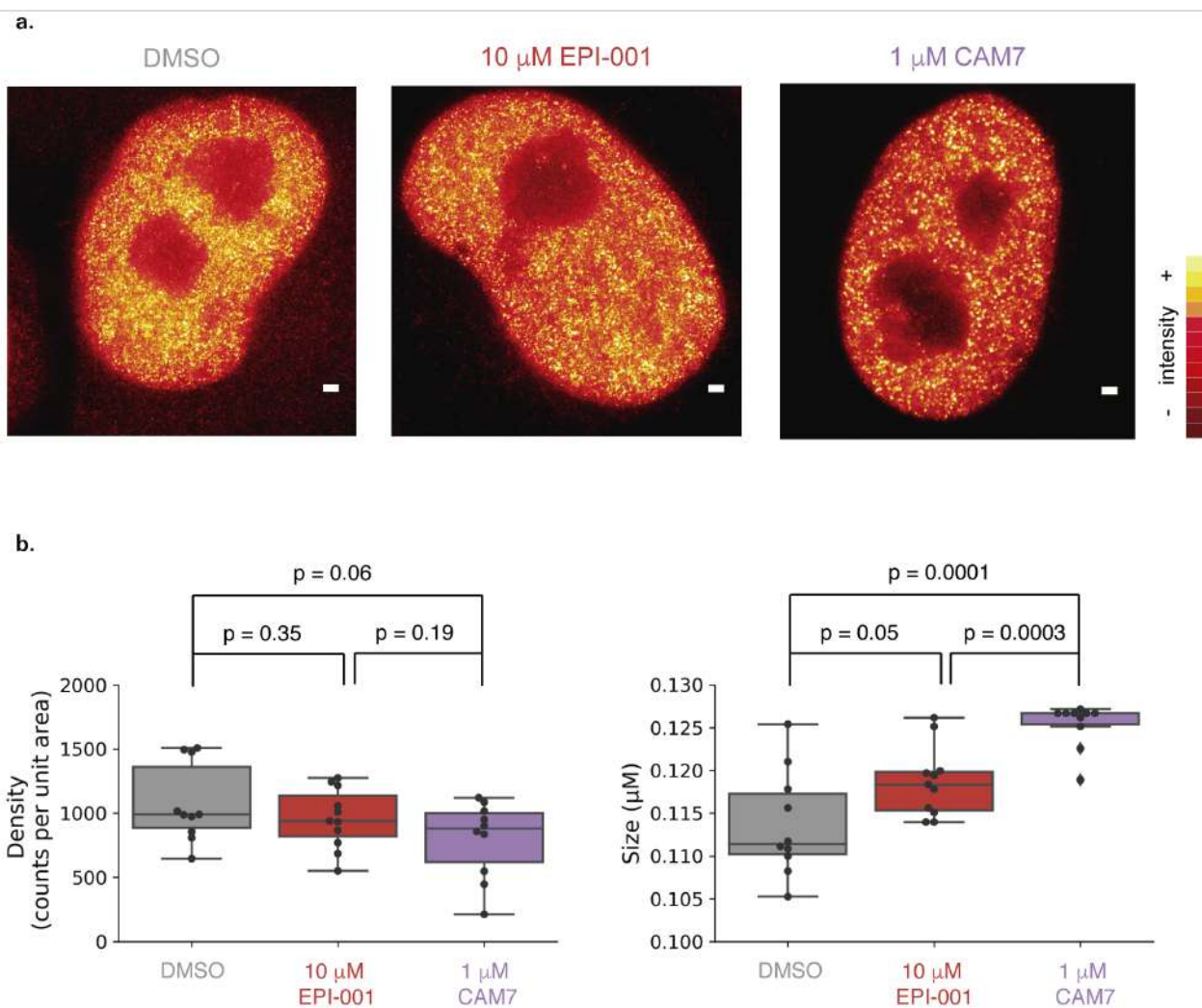


Fig. 4.2.3.1. a) Representative STED images of a HeLa cell nucleus stably transfected with AR-eGFP upon 4 h activation with 1 nM DHT pre-treated with vehicle (DMSO), 10 μM EPI-001 or 1 μM CAM7 for 1 h. Scale bar, 1 μm. **b)** Quantification of the density and size of clusters per nuclei of AR-eGFP HeLa cells upon 4 h with 1 nM DHT activation. Prior to activation, cells were treated for 1 h with vehicle (DMSO, grey), 10 μM EPI-001 (red) or 1 μM CAM7 (purple). Values from 10 independent nuclei are plotted. Significance tested with an independent two samples *t*-test.

Second, we wanted to better understand the inhibitory effect of CAMs to AR function. As in the case for EPI-001, we explored two scenarios (section 3.2.3):

4.2.3.1. Condensation facilitates recruiting the transcriptional machinery

We interrogated whether CAM6 was able to disrupt the interaction between the AR and the RNA-polymerase II (RNA-polII) making use of the established proximity-ligation assay (PLA)^{14,15}. We detected the interaction between the AR and the endogenous RNA-polII in the presence of 1 nM DHT in HeLa cells stably expressing AR-eGFP. We observed that the addition of 1 or 5 μM of CAM6 for 1 h prior to DHT activation for 4 h reduced the number of AR - RNA-polII interactions

-

(Fig. 4.2.3.2). Moreover, 1 μM CAM6 reduced the number of PLA interactions to the same extent as 10 μM EPI-001.

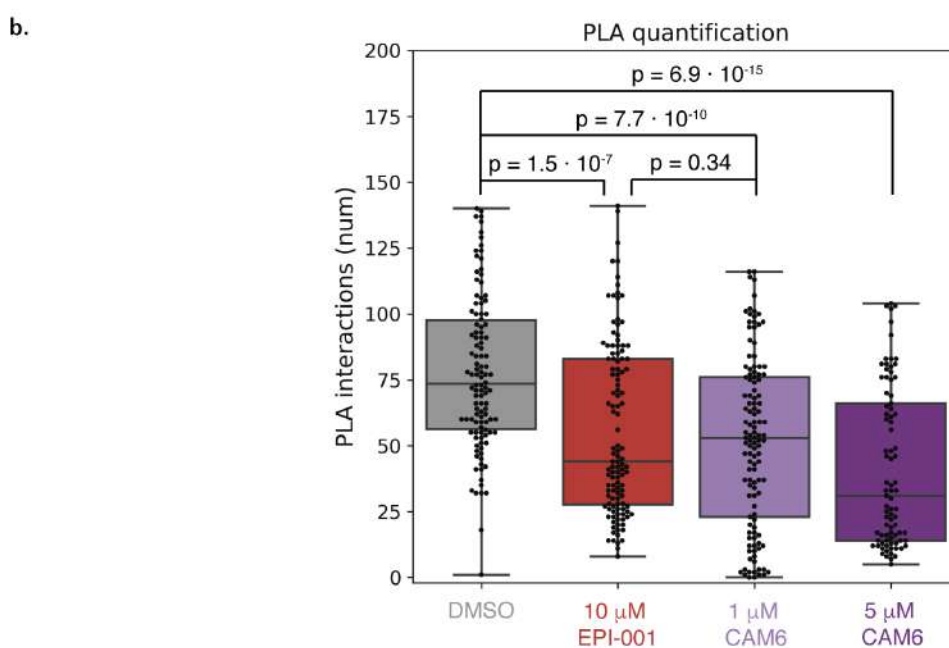
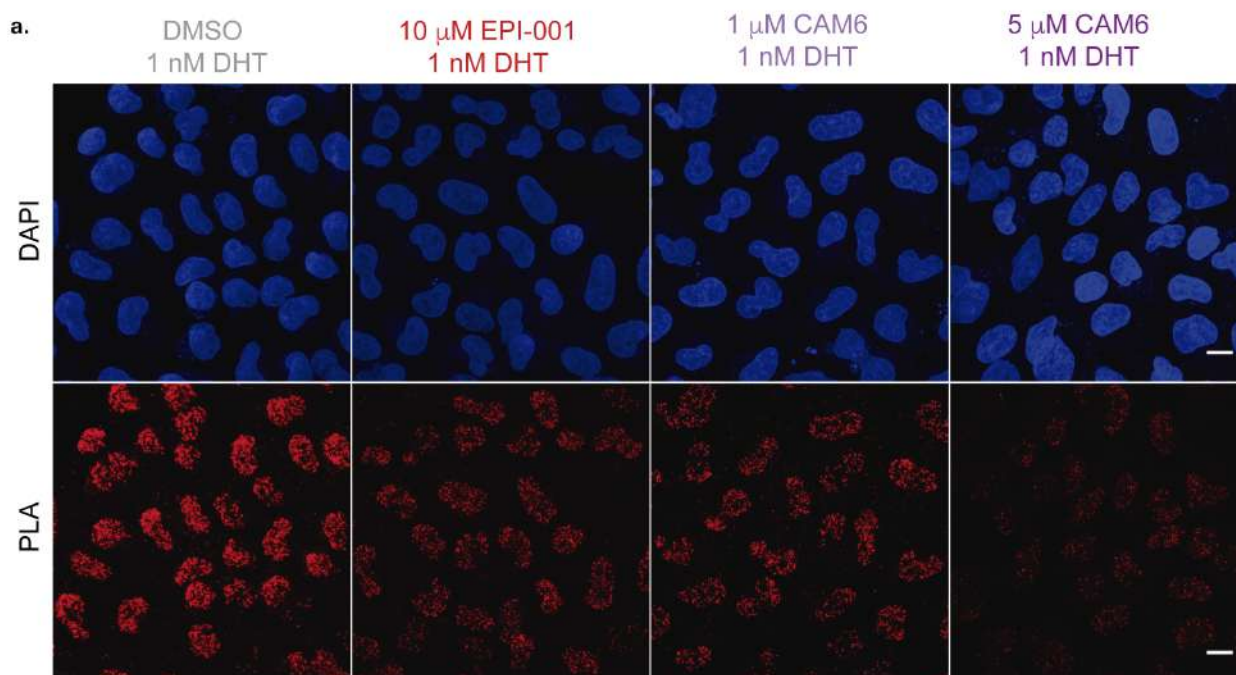


Fig. 4.2.3.2. s) Representative images of PLA interactions between AR and RNA-polII in HeLa cells stably expressing AR-eGFP. Cells were treated with vehicle (DMSO), 1 μM CAM6 or 5 μM CAM6 for 1 h prior to activation with 1 nM DHT for 4 h. 10 μM EPI-001 condition is depicted for comparison. As a control, in the absence of activation no PLA signal is detected (not shown). DAPI indicates the location of nuclei. Scale bar, 10 μm . **b)** Quantification of the PLA interactions per nucleus between AR and RNA-polII of DMSO (grey), 10 μM EPI-001 (red), 1 μM CAM6 (light purple) or 5 μM CAM6 (dark purple) treated cells for 1 h prior to 4 h 1 nM DHT activation. Error bars represent the mean and standard deviation of 2 independent experiments (at least 50 nuclei per experiment). Significance tested with an independent two samples *t*-test.

-

4.2.3.2. Condensation is a buffering system

We investigated whether condensate stabilisation by CAM7 led to a higher degree of AR degradation. As for EPI-001, we examined the AR protein expression levels in total homogenates of HeLa cells stably expressing AR-eGFP, treated with increasing concentrations of EPI-001 or CAM7 for 1 h prior to activation with DHT for extra 4 h. Western blot analysis (WB) revealed no significant decay of AR levels in all EPI-001 treated cells. In contrast, CAM7 promoted AR degradation at concentrations equal or higher than 10 μM (Fig. 4.2.3.3).

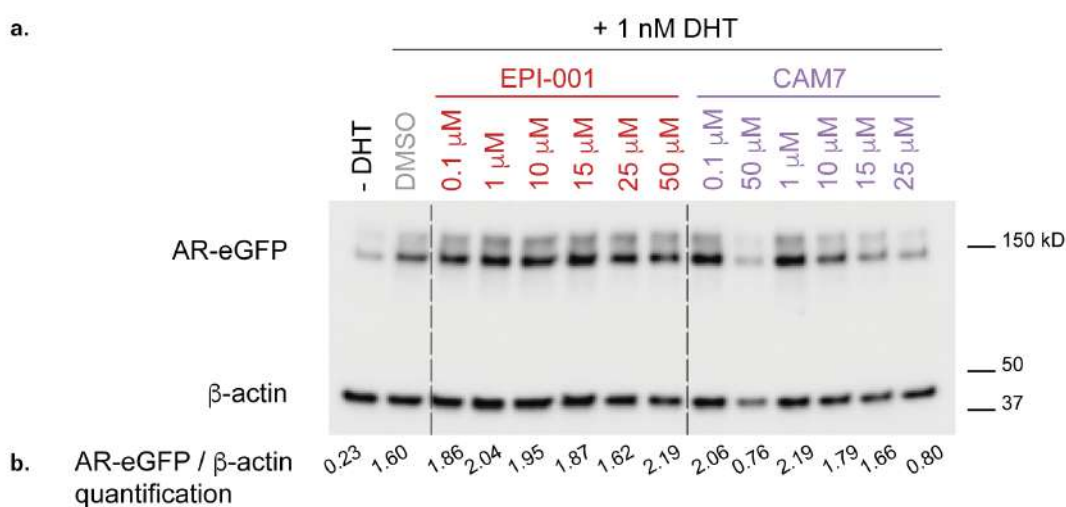


Fig. 4.2.3.3. a) WB analysis to assess AR-eGFP protein expression levels in HeLa cells stably expressing AR-eGFP treated with vehicle (DMSO), increasing concentrations of EPI-001 or CAM7 (indicated in the figure) for 1 h prior to activation with 1 nM DHT for 4 h. β -actin levels are determined as loading control. Dashed lines are drawn for visual guidance. **b)** Quantification of AR-eGFP expression levels normalised to those of actin.

Further experiments are being pursued in collaboration with Dr. Tanja Mittag laboratory at Saint Jude Children's Research Hospital to determine whether AR degradation promoted by CAM7 is SPOP-mediated.

In summary, our lead CAM compounds perform better than EPI-001 in all tested assays: they promote larger nuclear cluster formation, hamper AR - RNA-pollII interaction to a greater extent and are able to degrade the AR. It is yet to be investigated whether there is causality between any of these observations.

4.3. Discussion

Biomolecular condensates have great potential as new targets for therapeutics^{16,17}. However, still only few examples appear in the literature demonstrating the feasibility of targeting condensates.

We have built on our understanding of the mode of action of EPI-001 (section 3.2 and 3.3) to synthesise improved inhibitors of the androgen receptor N-terminal domain (AR-NTD) based on their ability to modulate the AR-NTD biomolecular condensation. The hypothesis was that, given the same ability to covalently react with the AR-NTD, compounds that partitioned better into the condensed phase would inhibit the AR activity to a greater extent. Biophysically, a compound that partitions better to the condensates also stabilises the condensed phase¹⁸.

We rationally designed and synthesised a family of compounds (CAM) to explore the SAR around EPI-001 structure. First, a linker expansion strategy was pursued to explore a free to operate chemical space. Moreover, the chemical route was optimised to access all the analogues with ease.

CAM compounds were designed with varying hydrophobicity. The idea being that, as the AR-NTD condensates confer a water-depleted environment, the more hydrophobic the compound, the more it would partition into the condensates. The data acquired indicates that the extent of condensate enhancement correlated with the in cell inhibitory activity. Noteworthy, by applying this rational approach we achieved compounds that are 10-fold more potent than the parent EPI-002 inhibitor in the AR transcriptional activity assay.

Targeting condensates is a new framework in which the established medicinal chemistry rules might not always hold true. For example, we can already envision that in the CAM family of compounds, where CAM4 (bearing the methyl substituent on ring 1) and CAM5 (bearing it on ring 2) have comparable cellular potencies.

Further SAR exploration should be carried out to fully understand which parameter or parameters govern the partitioning of a compound into the AR-NTD condensates and the link between partitioning and inhibitory activity.

Further work should also be centered in understanding the cellular consequences of the CAM compounds on the AR condensation and transcriptional activity. In this regard, it would be interesting to explore the inhibition of other interactions, not just between the AR and RNA-polII, such as interactions with the mediator complex and others (e.g. with SPOP). Additionally, the outcome of the compound-mediated degradation experiments, which are being carried out in the laboratory of Dr. Tanja Mittag, will be very valuable to assure or discard the implication of SPOP in the AR-degradation pathway.

Finally, it would be desirable to explore the effect of “epilating” the AR-NTD with more potent compounds, such as CAM7 or CAM6, on the condensation propensity and material properties of

-

the AR-NTD. Perhaps the correlation between the condensation enhancement due to the covalent addition and the cellular inhibitory activity might be larger than between the partitioning of the compounds and their cellular inhibitory activity.

Targeting condensates with small-molecules is a disruptive mechanism of action to treat “undrugged” targets, such as intrinsically disordered proteins, and “undrugged” diseases. To our knowledge the work conducted in this thesis represents one of the first examples of rational drug design based on the modulation of biomolecular condensates.

Bibliography

- (1) Ponnusamy, S.; Coss, C. C.; Thiyagarajan, T.; Watts, K.; Hwang, D.-J.; He, Y.; Selth, L. A.; McEwan, I. J.; Duke, C. B.; Pagadala, J.; Singh, G.; Wake, R. W.; Ledbetter, C.; Tilley, W. D.; Moldoveanu, T.; Dalton, J. T.; Miller, D. D.; Narayanan, R. Novel Selective Agents for the Degradation of Androgen Receptor Variants to Treat Castration-Resistant Prostate Cancer. *Cancer Res.* **2017**, *77* (22), 6282–6298.
- (2) Salami, J.; Alabi, S.; Willard, R. R.; Vitale, N. J.; Wang, J.; Dong, H.; Jin, M.; McDonnell, D. P.; Crew, A. P.; Neklesa, T. K.; Crews, C. M. Androgen Receptor Degradation by the Proteolysistargeting Chimera ARCC-4 Outperforms Enzalutamide in Cellular Models of Prostate Cancer Drug Resistance. *Comms. Biol.* **2018**, *1* (100), 1–9.
- (3) 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.; Sadar, M. D. Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor. *Cancer Cell* **2010**, *17* (6), 535–546.
- (4) Sadar, M. D. Discovery of Drugs That Directly Target the Intrinsically Disordered Region of the Androgen Receptor. *Expert Opin. Drug Discov.* **2020**, *15* (5), 551–560.
- (5) Wheeler, R. J.; Lee, H. O.; Poser, I.; Pal, A.; Doleman, T.; Kishigami, S.; Kour, S.; Anderson, E. N.; Marrone, L.; Murthy, A. C.; Jahnel, M.; Zhang, X.; Boczek, E.; Fritsch, A.; Fawzi, N. L.; Sternecker, J.; Pandey, U.; David, D. C.; Davis, B. G.; Baldwin, A. J.; Hermann, A.; Bickle, M.; Alberti, S.; Hyman, A. A. Small Molecules for Modulating Protein Driven Liquid-Liquid Phase Separation in Treating Neurodegenerative Disease. *bioRxiv*, **2019**. Doi: <https://doi.org/10.1101/721001> [Preprint].
- (6) Klein, I. A.; Boija, A.; Afeyan, L. K.; Hawken, S. W.; Fan, M.; Dall'Agnesse, A.; Oksuz, O.; Henninger, J. E.; Shrinivas, K.; Sabari, B. R.; Sagi, I.; Clark, V. E.; Platt, J. M.; Kar, M.; McCall, P. M.; Zamudio, A. V.; Manteiga, J. C.; Coffey, E. L.; Li, C. H.; Hannett, N. M.; Guo, Y. E.; Decker, T.-M.; Lee, T. I.; Zhang, T.; Weng, J.-K.; Taatjes, D. J.; Chakraborty, A.; Sharp, P. A.; Chang, Y. T.; Hyman, A. A.; Gray, N. S.; Young, R. A. Partitioning of Cancer Therapeutics in Nuclear Condensates. *Science* **2020**, *368* (6497), 1386–1392.
- (7) Meimetis, L. G.; Williams, D. E.; Mawji, N. R.; Banuelos, C. A.; Lal, A. A.; Park, J. J.; Tien, A. H.; Fernandez, J. G.; de Voogd, N. J.; Sadar, M. D.; Andersen, R. J. Niphatenones, Glycerol Ethers from the Sponge Niphates Digitalis Block Androgen Receptor Transcriptional Activity in Prostate Cancer Cells: Structure Elucidation, Synthesis, and Biological Activity. *J. Med. Chem.* **2012**, *55* (1), 503–514.
- (8) Imamura, Y.; Tien, A. H.; Pan, J.; Leung, J. K.; Banuelos, C. A.; Jian, K.; Wang, J.; Mawji, N. R.; Fernandez, J. G.; Lin, K.-S.; Andersen, R. J.; Sadar, M. D. An Imaging Agent to Detect Androgen Receptor and Its Active Splice Variants in Prostate Cancer. *JCI Insight* **2016**, *1* (11), 1–15.
- (9) Sadar, M. D.; Williams, D. E.; Mawji, N. R.; Patrick, B. O.; Wikanta, T.; Chasanah, E.; Irianto, H. E.; Soest, R. V.; Andersen, R. J. Sintokamides A to E, Chlorinated Peptides from the Sponge. *Org. Lett.* **2008**, *10*, 4947–4950.
- (10) Banuelos, C. a.; Lal, A.; Tien, A. H.; Shah, N.; Yang, Y. C.; Mawji, N. R.; Meimetis, L. G.; Park, J.; Kunzhong, J.; Andersen, R. J.; Sadar, M. D. Characterization of Niphatenones That Inhibit Androgen Receptor N-Terminal Domain. *PLoS One* **2014**, *9* (9), 1–10.
- (11) Peng, S.; Wang, J.; Chen, H.; Hu, P.; He, X.-L.; He, Y.; Wang, M.; Tang, W.; He, Q.; Wang, Y.-Y.; Xie, J.; Guo, D.; Ren, S.; Liu, M.; Qiu, W.-W.; Yi, Z. Regression of Castration-Resistant Prostate Cancer by a Novel Compound QW07 Targeting Androgen Receptor N-Terminal Domain. *Cell Biol. Toxicol.* **2020**. Doi: <https://doi.org/10.1007/s10565-020-09511-x>

-

[ahead of print].

- (12) 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.; McEwan, I. J.; Riera, A.; Salvatella, X. EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. *ACS Chem. Biol.* **2016**, *11*, 2499–2505.
- (13) Myung, J. K.; Banuelos, 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.; Sadar, M. D. An Androgen Receptor N-Terminal Domain Antagonist for Treating Prostate Cancer. *J. Clin. Invest.* **2013**, *123* (7), 2948–2960.
- (14) Söderberg, O.; Gullberg, M.; Jarvius, M.; Ridderstråle, K.; Leuchowius, K.-J.; Jarvius, J.; Wester, K.; Hydbring, P.; Bahram, F.; Larsson, L.-G.; Landegren, U. Direct Observation of Individual Endogenous Protein Complexes in Situ by Proximity Ligation. *Nat. Methods* **2006**, *3* (12), 995–1000.
- (15) De Mol, E.; Szulc, E.; Di Sanza, C.; Martínez-Cristóbal, P.; Bertoncini, C. W.; Fenwick, R. B.; Frigolé-Vivas, M.; Masín, M.; Hunter, I.; Buzón, V.; Brun-Heath, I.; García, J.; De Fabritiis, G.; Estébanez-Perpiñá, E.; McEwan, I. J.; Nebreda, Á. R.; Salvatella, X. Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators. *Structure* **2018**, *26*, 1–8.
- (16) Mullard, A. Biomolecular Condensates Pique Drug Discovery Curiosity. *Nat. Rev. Drug Discov.* **2019**, *18*, 324–326.
- (17) Wheeler, R. J. Therapeutics - How to Treat Phase Separation-Associated Diseases. *Emerg. Top. Life Sci.* **2020**, 1–12. Doi: <https://doi.org/10.1042/ETLS20190176>.
- (18) Posey, A. E.; Ruff, K. M.; Harmon, T. S.; Crick, S. L.; Li, A.; Diamond, M. I.; Pappu, R. V. Profilin Reduces Aggregation and Phase Separation of Huntingtin N-Terminal Fragments by Preferentially Binding to Soluble Monomers and Oligomers. *J. Biol. Chem.* **2018**, *293* (10), 3734–3746.

5. GENERAL DISCUSSION AND CONCLUSIONS

Intrinsically disordered proteins (IDPs) are considered undruggable targets¹. Their lack of a defined secondary and tertiary structure has hampered the drug discovery process, as conventional structure-based techniques cannot be used. However, IDPs are very abundant in the human proteome and are directly associated with many diseases. **Therefore, enabling new drug discovery approaches targeting IDPs is of high importance as it can yield new therapies for current unmet clinical needs.**

IDPs have been shown to play a primordial role in the formation of membraneless organelles. **The new conceptual framework of condensates offers a disruptive new opportunity to drug IDPs. Although this approach has high potential, the druggability of condensates has not been clinically validated nor the focus of many published studies²⁻⁵.** We believe that there are several concerns that the field needs to tackle before this approach becomes a reality for patients. Can condensates be targeted selectively? Are currently experimental methods valid to screen compounds targeting condensates or should other methods be implemented? Are current libraries enriched or depleted with compounds able to modulate condensation? Many concerns question the pre-established drug discovery methods and rules that will only be resolved with time and peer-reviewed studies and patents. Nevertheless, in the last year, two companies were created with the aim of targeting condensates.

To the best of our knowledge, **this work is a pioneer study on the modulation of biomolecular condensates with small molecule inhibitors.** Moreover, the new family of modulators described herein is the first example of a structure-activity relationship study based on the ability of the small molecules to modulate condensation.

We have based our studies on the disordered N-terminal domain of the androgen receptor (AR-NTD). The AR-NTD is a good model as i) the androgen receptor (AR) is a validated therapeutic target for castration-resistant prostate cancer (CRPC), an unmet clinical need⁶; ii) the AR undergoes condensation upon activation in cells and the AR-NTD is sufficient to recapitulate the process *in vitro*⁷ and iii) there is a known small molecule inhibitor of the AR-NTD, named EPI-001, that inhibits AR activity and reached phase I/II clinical trials for the treatment of CRPC^{8,9}. EPI-001 has been used as a reference compound in this thesis.

The starting point of this work was the previous study from De Mol et al. in the laboratories of Dr. Xavier Salvatella and Dr. Antoni Riera⁹. They concluded that EPI-001 interacted with the Tau-5 region of the activation function 1 (AF-1) in the AR-NTD, in particular with the conformations of the AF-1 ensemble that adopt a partial helical structure. The interaction of EPI-001 with the AF-1 was of very low affinity. They suggested this was because the population of helical conformations in the AF-1 ensemble, those that EPI-001 interacted with, was very small.

In chapter 2, we used nuclear magnetic resonance (NMR) spectroscopy to further study the reversible interaction of EPI-001 with its target. The initial plan was to use low amounts of the 2,2,2-trifluoroethanol (TFE) reagent as additive in the NMR samples to increase the helical population of the AF-1 and, therefore, increase the affinity of EPI-001 for the construct. As expected, addition of TFE increased the helical population of all the constructs used. However,

what we observed was that TFE also promoted the oligomerisation of the AF-1. The oligomerisation was further enhanced by the addition of EPI-001. In addition, the construct spanning the whole AR-NTD was even more oligomerisation prone than the AF-1. At that time we did not have any proof or idea that the AR-NTD underwent biomolecular condensation. The propensity of the AF-1 and the AR-NTD to oligomerise, together with other findings in the laboratory, led us to investigate the role of condensates in the AR molecular pathway and in the mechanism of action of EPI-001.

Dr. Elzbieta Szulc (X. Salvatella's group, IRB Barcelona) and Dr. Paula Martínez-Cristóbal (X. Salvatella's group, IRB Barcelona) investigated the role of condensates in the AR molecular pathway⁷. They described that the AR underwent condensation in cells upon activation by the hormone and that the AR-NTD was sufficient to recapitulate the demixing *in vitro*. In chapter 3, we studied the effect of EPI-001 on the AR condensation propensity. We observed that EPI-001 stabilised the AR-NTD condensates *in vitro* and the AR clusters in cells. Noteworthy, EPI-001 did not promote AR nucleation in cells but stabilised the already-formed clusters. We also examined the effect of the covalent addition of EPI-001 to the AR-NTD (named epilation for short) on the AR-NTD condensation propensity. We observed that epilation of the AR-NTD stabilised the AR-NTD condensates to a greater extent than the reversible binding. These findings led us to propose the following working hypothesis: In a first step, EPI-001 partitions into the condensed phase of the AR-NTD and also reversibly interacts with the AR-NTD increasing the population of oligomerisation prone conformations. In a second step, EPI-001 covalently binds to the AR-NTD. Epilated AR-NTD demixes more readily than the wild-type, being able to demix in conditions in which the wild-type AR-NTD is not. Epilated AR-NTD is not capable of interacting with the transcriptional machinery, which leads to the decrease in the AR transcriptional activity upon EPI-001 treatment. We also explored the possibility that the epilated AR-NTD was more prone to be degraded than the wild-type AR. Further experiments are being carried out in this direction.

According to the working hypothesis, inhibitors should be able to partition into the proteinaceous hydrophobic dense phase and promote AR condensation, probably by irreversibly modifying the AR-NTD. Based on this knowledge, in chapter 4 we decided to explore the first assumption, the correlation between the ability of new EPI-001 analogues to partition into the condensates with their inhibitory activity in cells. We kept the irreversible addition as a constant by maintaining the epichlorohydrin moiety in all the compounds tested. To build the new analogues, we first expanded the linker in EPI-001 from 1-atom to 2-atoms. From the 4 possibilities envisaged, the triple-bond linker was chosen for its high synthetic accessibility and promising cellular activity. Then, we designed and synthesised 8 analogues (named CAM) varying their ability to stabilise the AR-NTD condensates. We observed a semi-quantitative correlation between the ability of the compounds to stabilise the AR-NTD condensation and their inhibition of the AR transcriptional activity. We believe that more analogues should be tested to conclude on the degree of correlation. In any case, by applying this rational approach, we achieved compounds one order of magnitude more potent than EPI-001 within a small family of analogues. Our lead CAM compounds outperformed EPI-001 in all the assays tested.

The present thesis has integrated many different techniques and fields, from protein NMR and biophysics, to chemistry and cellular biology to shed light into the question: can we target biomolecular condensates? Overall, **we believe that two contributions have been crucial for the success and originality of the work. First, the introduction of the biomolecular condensates field to the druggability of IDPs.** As detailed in section 3.1.3, we have put forward four strategies to modulate condensates. In our case study, the inhibition of the AR-NTD by EPI-001, we concluded that EPI-001 acts by a mixture of allosteric and colligative inhibition mechanisms. **Second, the integration of all the mechanistic data, from biophysical and cellular experiments, to yield a family of new and improved AR inhibitors.** This step forward has not been taken in any published study so far in the biomolecular condensates field. In this regard, the help of Dr. Carolina Sánchez (A. Riera's group, IRB Barcelona) in the synthesis of the molecules once the route had been optimised and the collaboration with the laboratory of Dr. Marianne Sadar (Michael Smith Genome Sciences Centre, University of British Columbia, Canada) for the biological evaluation of the compounds have been key to accelerate the project. We believe that working at the intersection between biophysics, cell biology and chemistry will be decisive for the success of targeting biomolecular condensates.

Finally, we consider that there are still some open questions in this work. For example, it would be important to further clarify why stabilisation of the AR condensation leads to the inhibition of the AR transcriptional activity. It would also be important to explore the effects of EPI-001 and CAMs on the nuclear AR clusters at endogenous protein levels. Condensation is a concentration-dependent process so it would be desirable to work at endogenous levels of AR. With regards to the new family of compounds, we are exploring whether other scaffolds different from EPI-001 can inhibit the AR activity through the same mechanism of action. Finally, in a broader sense, the **main open question is whether the findings presented in this thesis can be extended and expanded to drug other condensates (and IDPs) than those of AR.**

In conclusion, the data presented in this work indicates that:

- 1- The AR-NTD and the AF-1* are oligomerisation prone, particularly through the Tau-5 region.
- 2- Oligomerisation is associated with a gain in helical content.
- 3- EPI-001, but not BADGE, enhances the AR-NTD and AF-1* oligomerisation process.
- 4- EPI-001 partitions into the AR-NTD droplets and enhances their formation *in vitro*.
- 5- Covalent addition of EPI-001 to the AR-NTD has a more profound effect on the AR-NTD condensate stabilisation than reversible binding of EPI-001 to the AR-NTD.
- 6- In cells, EPI-001 does not induce AR nucleation but stabilises the already formed AR cytosolic and nuclear clusters.

5. General discussion and conclusions

–

7- We have optimised the synthetic route to access the linker-expanded EPI-001 analogues, from which the compound with the triple-bond linker was chosen as the most promising one for further derivatisation.

8- We have rationally designed and synthesised a family of 8 compounds (CAM family) containing the triple-bond linker based on their ability to modulate AR condensation.

9- The lead CAM compounds are one order of magnitude more potent than EPI-001 in the transcription inhibition and AR-dependent proliferation assays and outperform EPI-001 in all the condensation-assays tested.

10- EPI-001 and CAM inhibition activity could be explained by a decrease in the number of interactions between the AR and the RNA-polIII and also by an increased degradation of the AR.

Bibliography

- (1) Uversky, V. N. Intrinsically Disordered Proteins and Novel Strategies for Drug Discovery. *Expert Opin. Drug Discov.* **2012**, *7*, 475–488.
- (2) Mullard, A. Biomolecular Condensates Pique Drug Discovery Curiosity. *Nat. Rev. Drug Discov.* **2019**, *18*, 324–326.
- (3) Wheeler, R. J.; Lee, H. O.; Poser, I.; Pal, A.; Doleman, T.; Kishigami, S.; Kour, S.; Anderson, E. N.; Marrone, L.; Murthy, A. C.; Jahnel, M.; Zhang, X.; Boczek, E.; Fritsch, A.; Fawzi, N. L.; Sternecker, J.; Pandey, U.; David, D. C.; Davis, B. G.; Baldwin, A. J.; Hermann, A.; Bickle, M.; Alberti, S.; Hyman, A. A. Small Molecules for Modulating Protein Driven Liquid-Liquid Phase Separation in Treating Neurodegenerative Disease. *bioRxiv*, **2019**. Doi: <https://doi.org/10.1101/721001> [Preprint].
- (4) Wheeler, R. J. Therapeutics - How to Treat Phase Separation-Associated Diseases. *Emerg. Top. Life Sci.* **2020**, 1–12. Doi: 10.1042/ETLS20190176.
- (5) Klein, I. A.; Boija, A.; Afeyan, L. K.; Hawken, S. W.; Fan, M.; Dall’Agnese, A.; Oksuz, O.; Henninger, J. E.; Shrinivas, K.; Sabari, B. R.; Sagi, I.; Clark, V. E.; Platt, J. M.; Kar, M.; McCall, P. M.; Zamudio, A. V.; Manteiga, J. C.; Coffey, E. L.; Li, C. H.; Hannett, N. M.; Guo, Y. E.; Decker, T.-M.; Lee, T. I.; Zhang, T.; Weng, J.-K.; Taatjes, D. J.; Chakraborty, A.; Sharp, P. A.; Chang, Y. T.; Hyman, A. A.; Gray, N. S.; Young, R. A. Partitioning of Cancer Therapeutics in Nuclear Condensates. *Science* **2020**, *368* (6497), 1386–1392.
- (6) Sadar, M. D. Discovery of Drugs That Directly Target the Intrinsically Disordered Region of the Androgen Receptor. *Expert Opin. Drug Discov.* **2020**, *15* (5), 551–560.
- (7) Szulc, E. M. Structural Insights into “Acid Blobs and Negative Noodles” - the Androgen Receptor as a Case Study. Ph. D. Thesis, Universitat de Barcelona - IRB Barcelona, 2019.
- (8) 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.; Sadar, M. D. Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor. *Cancer Cell* **2010**, *17* (6), 535–546.
- (9) 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.; McEwan, I. J.; Riera, A.; Salvatella, X. EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. *ACS Chem. Biol.* **2016**, *11*, 2499–2505.

6. EXPERIMENTAL METHODS

6.1. Protein expression and purification

6.1.1. Androgen receptor N-terminal domain (AR-NTD) constructs

Three AR-NTD constructs were used in this thesis: Tau-5* (330-448), AF-1* (142-448) and AR-NTD (1-558, bearing the L26P mutation). All three of them were expressed as His₆-tag fusion in *E. coli* Rosetta (DE3) competent cells from Novagen (Merck) and purified from inclusion bodies as previously described¹⁻³. Noteworthy, the AR-NTD was stored at -80 °C in 8 M urea to avoid aggregation. The final size exclusion chromatography of the AR-NTD protocol for buffer exchange and concentration to the required concentration was performed before each experiment. ¹⁵N and ¹³C,¹⁵N enriched samples obtained by expression in M9 minimal medium containing ¹³C-glucose and ¹⁵NH₄Cl as sole nitrogen and carbon sources, obtained from Cambridge Isotope Laboratories, Inc.

6.1.2. Fused in Sarcoma (FUS) protein

GFP-tagged FUS was used from Dr. Tony Hyman laboratory's general stock. GFP-FUS had been expressed and purified as previously described⁴.

6.2. Nuclear Magnetic Resonance (NMR)

6.2.1. Assignment of the constructs

The assignment of the Tau-5* and the AF-1* constructs were previously described^{1,2}. The assignment of the AR-NTD was conducted in the laboratory of molecular biophysics by Dr. Jesús García and published in the PhD thesis of Dr. Elzbieta Szulc⁵.

6.2.2. Sample preparation and NMR acquisition

Samples were prepared at the indicated concentration in 20 mM sodium phosphate buffer at pH 7.4 containing 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) as reducing agent, 10% D₂O, deuterated 2,2,2-trifluoroethanol (TFE-d₃) when indicated and 10 μM deuterated N,N'-Dicyclohexylcarbodiimide (DSS-d₆) for reference purposes. EPI-001 and BADGE were added when required from a concentrated stock solution in TFE-d₃ to achieve, in the final sample, a concentration of TFE of 0.5%. The blank measurements were conducted in the presence of 0.5%

6. Experimental methods

-

TFE-d₃. All experiments were acquired with standard Bruker pulse sequences on 600 and 800 MHz Bruker Advance spectrometers equipped with a cryoprobe at 278 K. All reagents were acquired from Sigma-Aldrich (Merck) unless otherwise noted.

6.3. Circular Dichroism (CD)

CD spectra were acquired on 16 μ M samples in 20 mM sodium phosphate buffer at pH 7.4 containing 1 mM TCEP and TFE (Alfa Aesar) at the indicated concentration in a Jasco 815 UV spectropolarimeter at 278 K with a 1 mm optical path cuvette.

6.4. Microscopy of protein samples

AR-NTD samples were prepared in PCR low binding 200 μ L tubes (Eppendorf) on ice and condensation was initiated by increasing the sample temperature to room temperature. The proteins were mixed with the indicated buffer containing NaCl and 10% Ficoll 70 when necessary. EPI-001 was added when required from a concentrated stock solution in TFE-d₃ to achieve, in the final sample, a concentration of TFE of 0.5%. 1% of the protein was labeled with 10X rhodamine Red-X succinimidyl ester, 5 isomer (Thermo Fisher). 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). Coverslips were coated with polyethylene glycol (PEG) using a standard protocol ⁶. Fluorescence confocal images were recorded using a Zeiss LSM780 confocal microscope system with a Plan ApoChromat 63x 1.4 oil objective. Samples were equilibrated at room temperature for 30 minutes before imaging to allow sedimentation of the droplets on the coverslip.

In the case of FUS, samples were prepared at room temperature and condensation was initiated by diluting the NaCl concentration of the final sample with buffer. Samples were transferred into a 384-well non-binding microplate (Greiner Bio-One) and imaged using an IX71/IX81 inverted Spinning Disc Microscopes with an Andor Neo sCMOS/Andor Clara CCD camera and an UPlanSApo 60x oil-immersion objective or 100x oil-immersion objective (Olympus).

6.5. Turbidity shift assay

Samples were prepared in eppendorf tubes on ice in 20 mM Hepes or NaP buffer, pH 7.4, 2 mM TCEP with the indicated concentration of NaCl. The optical density of the sample (at 350 nm wavelength) was monitored as a function of temperature on a Cary 100 ultraviolet-visible spectrophotometer equipped with a multicell thermoelectric temperature controller at a heating rate

6. Experimental methods

–

of $1\text{ }^{\circ}\text{C}\cdot\text{min}^{-1}$. The condensation onset (cloud point temperature) was determined as the temperature of abrupt increase in the turbidity of the sample.

6.6. Mass spectrometry (MS)

The MS analysis was carried out by the IRB Barcelona Mass Spectrometry and Proteomics Core Facility.

6.6.1. Liquid chromatography coupled to MS

Samples were diluted in 3% acetonitrile / 1% formic acid aqueous solution to 5 μM of AR-NTD, 1 μM of AF-1* and 1 μM of Tau-5*. The proteins were loaded and separated in the LC-MS system with the following set up:

The AR-NTD samples were loaded to a phenyl analytical column (BioSuite pPhenyl 1000, 2.0×75 mm, 10 μm , Waters Corp.) with a 90 min run, using two consecutive steps with linear gradients from 5% to 80% of B in 60min, from 80% to 90% of B in 9 min, followed by isocratic elution at 90 % of B in 11 min and stabilization to initial conditions (A= 0.1% formic acid in water, B= 0.1% formic acid in acetonitrile).

The AF-1* and Tau-5* samples were loaded to a 180 $\mu\text{m} \times 2$ cm C18 Symmetry trap column (Waters Corp.) at a flow rate of 15 $\mu\text{L}/\text{min}$ using a nanoAcquity Ultra Performance LCTM chromatographic system (Waters Corp.). Proteins were separated using a C18 analytical column (BEH130TM C18 75 $\mu\text{m} \times 25$ cm, 1.7 μm , Waters Corp.) with a 80 min run, using two consecutive steps with linear gradients from 10% to 85% of B in 60 min (for the AF-1*) and from 5% to 85% of B in 60 min (for the Tau-5*), followed by isocratic elution at 85 % of B in 5 min and stabilization to initial conditions (A= 0.1% formic acid in water, B= 0.1% formic acid in acetonitrile).

The column outlets were directly connected to an Advion TriVersa NanoMate (Advion) fitted on an LTQ-FT Ultra mass spectrometer (ThermoFisher Scientific). Survey MS1 scans were acquired in the FT with the resolution (defined at 400 m/z) set to 100,000. The ion count target value was 1,000,000. Spray voltage in the NanoMate source was set to 1.70 kV. Capillary voltage and tube lens on the LTQ-FT were tuned to 35 V and 100 V. The spectrometer was working in positive polarity mode. At least one blank run before the analysis was performed in order to ensure the absence of cross contamination from previous samples.

6. Experimental methods

-

6.6.2. Trypsin digestion and MS analysis

6.6.2.1. Sample digestion

Sample was tryptic digested directly in solution. Briefly, 72.13 μg of sample were reduced with DTT 2 mM for 1 h at rt, carbamidomethylated for 30 min in the dark at rt with IAA 5 mM, DTT was added to a final concentration 2 mM to consume any unreacted IAA and finally, sample was digested with trypsin (2%w) in 50 mM ammonium bicarbonate at 37 °C overnight. Digestion was stopped by addition of formic acid to a final concentration of 1%. Samples were cleaned up through C18 tips (polyLC C18 tips) and peptides were eluted with 80% acetonitrile, 1% trifluoroacetic acid. Samples were evaporated to dryness, reconstituted in 60 μL (for n=2) or 70 μL (for n=1) and diluted 1:5 (for n=2) or 1:8 (for n=1) with 3% acetonitrile, 1% formic acid aqueous solution for MS analysis.

6.6.2.2. MS analysis

The nano-LC-MS/MS set up was as follows. Digested peptides were diluted in 3% acetonitrile / 1% formic acid. Samples were loaded to a 300 μm \times 5 mm PepMap100, 5 μm , 100 \AA , C18 μ -precolumn (ThermoFisher Scientific) at a flow rate of 15 $\mu\text{L}/\text{min}$ using a Thermo Scientific Dionex Ultimate 3000 chromatographic system (ThermoFisher Scientific). Peptides were separated using a C18 analytical column NanoEase MZ HSS T3 column (75 μm \times 250 mm, 1.8 μm , 100 \AA) (Waters) with a 90 min run, comprising three consecutive steps with linear gradients from 3 to 35% B in 60 min, from 35 to 50% B in 5 min, and from 50 % to 85 % B in 2 min, followed by isocratic elution at 85 % B in 5 min and stabilization to initial conditions (A= 0.1% formic acid in water, B= 0.1% formic acid in acetonitrile). The column outlet was directly connected to an Advion TriVersa NanoMate (Advion) fitted on an Orbitrap Fusion Lumos™ Tribrid (ThermoFisher Scientific). The mass spectrometer was operated in a data-dependent acquisition (DDA) mode. Survey MS scans were acquired in the Orbitrap with the resolution (defined at 200 m/z) set to 120,000. The lock mass was user-defined at 445.12 m/z in each Orbitrap scan. The top speed (most intense) ions per scan were fragmented by HCD and detected in the orbitrap. The ion count target value was 400,000 and 10,000 for the survey scan and for the MS/MS scan respectively. Target ions already selected for MS/MS were dynamically excluded for 15 s. Spray voltage in the NanoMate source was set to 1.60 kV. RF lens were tuned to 30%. The Minimum signal required to trigger MS to MS/MS switch was set to 5,000. The spectrometer was working in positive polarity mode and single charge state precursors were rejected for fragmentation.

A database search was performed with Proteome Discoverer software v2.3.0.480 (ThermoFisher Scientific) using Sequest HT and Amanda search engine with ptmRS validator, SwissProt Human canonical released 2019_05, contaminants database and user protein manually introduced. Search was run against targeted and decoy databases to determine the false discovery rate (FDR). Search parameters included trypsin enzyme specificity, allowing for two missed cleavage sites, oxidation in M, EPI in C, carbamidomethylation in C and acetylation in protein N-terminus as dynamic

6. Experimental methods

-

modifications. Peptide mass tolerance was 10 ppm and the MS/MS tolerance was 0.02 Da. Peptides with a q-value lower than 0.01 were considered as positive identifications with a high confidence level. The protein FDR validator node was used to estimate the number of falsely identified proteins among all the identified proteins. Proteins with an FDR lower than 0.01 were considered with high confidence, proteins with FDR between 0.01 and 0.05 as medium confidence level and higher than 0.05 as low confidence.

The ptmRS node was used to provide a confidence measure for the localization of EPI in the peptide sequences identified with this modification. Peptide spectrum matches (PSM) were considered for relative quantification. Ratios of the peptides with or without EPI were calculated and EPI sites relative abundance was determined.

On the other hand, MS/MS spectra were searched against the Swissprot Human release 2019_05 and contaminants database and user proteins manually introduced using MaxQuant v1.6.6.0 with andromeda search engine ⁷. Searches were run against targeted and decoy databases. Search parameters included trypsin enzyme specificity, allowing for two missed cleavage sites, oxidation in methionine, EPI in cysteine, carbamidomethylation in cysteine and acetylation in protein N-terminus as dynamic modifications. Peptide mass tolerance was 20 ppm and the MS/MS tolerance was 20 ppm and the minimal peptide length was 6 amino acids. Peptide and protein identifications were filtered at a false discovery rate (FDR) of 1 % based on the number of hits against the reversed sequence database.

We computed the EPI ratio (r) for each EPI site within AR-NTD protein taking into account three search nodes: Andromeda, from Max Quant (MQ) software, Amanda and Sequest, from Proteome Discoverer software (PD). For each EPI site, we counted the number of EPI-modified (N_{Mod}) specific peptide spectrum matches (PSMs) and the number of non-EPI (N_{NoMod}) PSMs, from which we then computed r as follows: $r = N_{Mod} / (N_{Mod} + N_{NoMod})$.

In this calculation, we only considered those EPI sites with localization probability higher than 75%. EPI peptides containing EPI sites with localization probability $< 75\%$ were reported in an additional table. For the final output, we combined all node results by choosing the node with the maximum number of PSM for each EPI site.

6.6.3. Epilated AR-NTD production

1-2 mL of 25 μ M AR-NTD were incubated with 0 (non-epilated) or 250 μ M EPI-001 (epilated) overnight at 37 °C and pH=8.0 in 20 mM phosphate buffer and 2 mM TCEP. The products of the reaction were dialysed 3 \times 2 h using a Pur-A-Lyzer Midi Dialysis Kit 3.5 KDa (Sigma-Aldrich Merck) against fresh 20 mM phosphate buffer and 2 mM TCEP to remove non-reacted EPI-001. Then, the proteins were concentrated using an Amicon Ultra-15 3 KDa MWCO (Merck) to ca $\frac{1}{4}$ of the initial volume, to reach an estimated concentration of 100 μ M. Then, the protein was aliquoted and freezed at -80 °C until needed. The protein concentration was measured using Reversed-phase

HPLC amino acid analysis ⁸ by the University of Barcelona unit of separative techniques and peptide synthesis.

6.7. Fluorescence recovery after photobleaching (FRAP) of protein samples

The protein samples for FRAP were prepared as in section 6.4. FRAP experiments were performed using a Zeiss LSM780 confocal microscope system with a Plan ApoChromat 63x 1.4 oil objective. The measurements were conducted within 30 min to 1 h from sample preparation and the bleached region was around 30% of the droplet diameter in the middle plane section. Droplets from different conditions were selected to be of the same diameter. 3 independent FRAP measurements were carried out for each condition (10 droplets per condition). The exported recovery curves were averaged and normalised using the EasyFrap software ⁹.

6.8. Cell culture

HeLa cells stably expressing AR-eGFP were a gift from the laboratory of Dr. Maria Pennuto and were cultured in DMEM medium containing 4.5 g·L⁻¹ glucose (ThermoFisher Scientific) supplemented with 10% (v/v) charcoal stripped FBS (Life) and antibiotics. PC3 cells were obtained from ATCC (CRL-1435) and were cultured in RPMI medium containing 4.5 g·L⁻¹ glucose (ThermoFisher Scientific) supplemented with 10% (v/v) charcoal stripped FBS (Life) and antibiotics. Cells were cultured in a humidified atmosphere containing 5% CO₂ at 37 °C.

6.9. Time-resolved live-cell imaging

PC3 cells were seeded in collagen I-coated μ -slide 4-well glass-bottom plates (Ibidi 80426) at 60% confluency 24 h before transfection. Then, 170 ng of expression vectors (eGFP-tagged full length AR or eGFP-tagged AR- Δ NLS) were transfected per well using a ratio of 1 mg of DNA to 3 ml of polyethylenimine (PEI) (Polysciences). 4 h post-transfection, the medium was changed to RPMI medium supplemented with 10% charcoal stripped FBS and cells were cultured for 16 h before treatment and imaging. Before imaging, cells were treated with either DMSO or the indicated compound for 1 h. Then, cells were imaged in 3D for 1 min every 15 sec to acquire a baseline readout of AR expression. Immediately after, dihydrotestosterone (DHT) was added to a final concentration of 1 nM and cells were imaged for 1 h every 15 sec. Time-lapse images were performed using a Spinning Disk confocal microscope (Olympus IX81) equipped with an iXon EMCCD Andor DU-897 camera and a Yokogawa CSU-XI scan head with a Plan ApoN 60x/1.42 oil objective. A stable temperature (37 °C) and CO₂ concentration (5%) were maintained during imaging in an incubation chamber (EMBL, Germany). eGFP was excited with a 488 nm laser and

6. Experimental methods

–

Z-stack images were acquired every 0.45 μm step size. Time-resolved images were compiled, processed and edited with Fiji (ImageJ) software.

6.10. Stimulated emission depletion (STED) super-resolution microscopy on fixed cells

AR-eGFP HeLa stable cells were cultured on a glass bottomed cellview cell culture dish (Greiner bio-one) in DMEM media provided with 10% (v/v) charcoal stripped FBS (Life) for 24h prior addition of the indicated compound for 1 h and subsequent addition of 1 nM DHT for 4 extra h. Then, cells were fixed with PBS containing 4% paraformaldehyde for 15 min. After washing with PBS, cells quenched with 300mM Glycine and 0.3% Triton in PBS and permeabilised with 0.5% Triton in PBS. Afterwards, cells were blocked with 2% BSA in PBS. Finally, AR-eGFP was stained with GFP-Booster Abberior AS 635P (ThermoFisher Scientific) 1:50 at 4 °C overnight.

STED imaging on fixed and live samples was performed using an Abberior 3D-2 Color-STED system (Abberior Instruments) with a 100x/1.4 NA oil or a 60x/1.2 NA water objectives (Olympus). Excitation of the Abberior Star Red probe was performed at 640 nm ¹⁰The depletion laser 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 ¹¹, as previously described ¹²⁻¹⁴. 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, i.e. objects were recognised by fitting the image intensity with a sum of powered Lorentzian functions. 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.

6.11. FRAP assay in live cells

HeLa cells stably expressing AR-eGFP were plated on μ -slide 4-well glass-bottom plates (Ibidi) in DMEM medium provided with 10% (v/v) charcoal-stripped serum (Life) at 80% confluency and cultured for 48 h. On the day of the experiment, cells were treated for 1 h with DMSO or the indicated compound prior to the addition of 1 nM DHT for 4 extra h. FRAP data for each condition was acquired over the course of 30 min (no trend was observed between FRAP data acquired at the beginning versus the end of the acquisition time). FRAP measurements were performed on a Spinning Disk confocal microscope (Olympus IX81) equipped with an iXon EMCCD Andor DU-897 camera and a Yokogawa CSU-XI scan head with a UPlanSApo 100x/1.40 oil objective. A stable temperature (37 °C) and CO₂ concentration (5%) were maintained during imaging in an

6. Experimental methods

-

incubation chamber (EMBL, Germany). Pre-bleaching (20 images) and fluorescence recovery images (200 images) of eGFP-AR were acquired with the same 488 nm laser power with an exposure time of 180 msec in a 15x15 pixel square format within a HeLa nucleus. Bleaching was conducted for 50 ms at 60% intensity of the 488 nm laser power. Mean intensity measurements were quantified in three different regions of interest: the bleached region, the background region outside the cell and the nucleus of the cell. Fiji (ImageJ) was used to measure the fluorescence intensities. Exported recovery curves were normalised using the EasyFrap software⁹.

6.12. Proximity-ligation assay (PLA)

HeLa cells stably expressing AR-eGFP were plated on 12 mm-diameter coverslips in 6-well plates (ThermoFisher Scientific) in DMEM medium provided with 10% (v/v) charcoal-stripped serum (Life) and cultured for 48 h before treatment. Cells were then treated with the indicated compound for 1 h and incubated with 1 nM DHT for an additional 4 hours. After that, cells were fixed in PBS containing 4% paraformaldehyde (EMS) for 10 minutes, subsequently washed in PBS and permeabilised with methanol for 7 minutes. Coverslips were then blocked in the manufacturer's blocking solution for 1 h at rt, and incubated with antibodies against Androgen Receptor (rabbit-anti-AR 441, sc-7305, Santa Cruz Biotechnology, 1:300) and RNA-polymerase II (mouse-anti-RNA-polIII, 8WG16, abcam, 1:300). Cells were subsequently incubated with Duolink II PLA probes (Sigma-Aldrich Merck) and stained according to the manufacturer's protocol. Cells were analyzed with a 63x objective lens on a Leica SP5 or Zeiss LSM780 confocal microscopes.

6.13. Western Blot (WB)

HeLa cells stably expressing AR-eGFP (a generous gift from the laboratory of Dr. Maria Pennuto) were plated on 6-well plates (ThermoFisher Scientific) in DMEM media provided with 10% charcoal-stripped serum (Life) and cultured for 48 h before treatment. Cells were then treated with the indicated compound for 1 h and incubated with 1 nM DHT for an additional 4 hours. After that, cells were washed with PBS and lysed in passive lysis buffer (Promega). Total cellular lysate was centrifuged at 15000 g to separate soluble and pellet fractions. Total protein was quantified using the Bradford protein assay (Bio-Rad). A constant protein concentration was loaded and resolved in mini-Protean TGX 4-20% acrylamide gel (Bio-Rad) and blotted onto a PVDF membrane (Bio-Rad). Protein levels were assessed by means of the following antibodies: anti- β -actin-HRP (ab8224, abcam, 1:10000), rabbit-anti-Androgen Receptor 441 (sc-7305, Santa Cruz Biotechnology, 1:1000), mouse-anti-RNA-polymerase II (mouse-anti-RNA-polIII, 8WG16, abcam, 1:1000) and anti-mouse HRP-conjugated secondary antibody (ab20272, abcam, 1:5000). Protein bands were quantified using Fiji (ImageJ) software.

6.14. Chrom logD determination

Reversed-phase liquid chromatography (Chrom logD) values were experimentally determined as a measure of hydrophobicity of the CAM family of compounds. The experimental evaluation was subcontracted to Fidelta, Ltd. Chrom logD values were calculated from the following equation: $\text{Chrom logD} = 0.0857 \cdot \text{CHI} - 2$. Chromatographic Hydrophobic Index (CHI) values were determined from gradient retention times at pH = 7.4. Chromatograms were obtained using an Agilent 1100 Series HPLC instrument equipped with diode array detector (DAD) coupled with Micromass Quattro API mass spectrometer. A Luna C18 column (50 mm 3 mm i.d., 5 mm particle size, 100 Å) from Phenomenex was used with a 5 min run, at 1 mL/min flow rate using the following gradient: 0-3 min with linear gradient from 0 to 100% of B; 3-3.5 min in 100% of B; 3.5-3.7 min from 100 to 0% of B; and 3.7-5 min re-equilibration with 100% of A. A = 50 mM ammonium acetate in water, B = acetonitrile. Data acquisition and processing were performed with MassLynx software version 4.1.

6.15. Biological evaluation

6.15.1. Transcriptional activity assay (PSA-luciferase assay)

The PSA-luciferase assay was performed by the laboratory of Dr. Marianne Sadar as previously described¹⁵. Transcriptional activities were determined in transient transfection assays using the PSA(6.1kb)-luciferase reporter. Transfected cells were incubated for 24 h in the absence and presence of 1 nM R1881 (the synthetic androgen methyltrienolone) with or without inhibitors. Luciferase activity was measured using the Steady Glo Luciferase assay buffer (Promega) with CLIPR (Molecular Devices). Luciferase activities were normalised to protein concentrations of the samples.

6.15.2. Proliferation assay

The AR-dependent and AR-independent proliferation assays were performed by the laboratory of Dr. Marianne Sadar as previously described¹⁵. PC3 and LNCaP cells were plated in 96-well plates in respective media plus 0.5% FBS. The next day, PC3 cells were treated with vehicle or inhibitor for 2 days, and LNCaP cells were pretreated with vehicle or inhibitor for 1 hour before treating with 0.1 nM R1881 for 3 days. Cell viability was measured using alamarBlue Cell Viability Assay (Invitrogen) following the manufacturer's protocol.

6.16. Chemical synthesis

6.16.1. Relevant chemical abbreviations

ACN Acetonitrile	MeOH Methanol
atm Atmosphere	PTSA p-Toluenesulfonic acid
BuLi Butyllithium	rt Room temperature
DMF Dimethylformamide	THF Tetrahydrofuran
EtOH Ethanol	UV Ultraviolet light
DCM Dichloromethane	CV Column volumes

6.16.2. General methods and materials for synthetic work

Commercially available reagents and solvents were used without further purification unless otherwise noted. Anhydrous and degassed THF was taken from a solvent purification system (SPS PS-MD-3). All reactions that required anhydrous conditions were performed under a dry nitrogen atmosphere. All reactions were monitored by thin layer chromatography (TLC) using silica gel on aluminium sheets (Merck). Solvents were evaporated under reduced pressure with a rotary evaporator. Silica gel chromatography was performed using an automated chromatography system (PuriFlash® 430, Interchim).

^1H and ^{13}C NMR spectra were recorded at room temperature on a Varian Mercury 400 MHz spectrometer from the *Centres Científics i Tecnològics* of the University of Barcelona. Chemical shifts (δ) were referenced to internal solvent resonances and reported relative to tetramethylsilane (TMS). The coupling constants (J) are reported in Hertz (Hz). The following abbreviations are used to define multiplicities: s (singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublets), ddd (doublet of doublet of doublets), ddt (doublet of doublet of triplets), dddd (doublet of doublet of doublet of doublets), dq (doublet of quartets), m (multiplet), quint (quintuplet), bs (broad signal).

IR spectra were recorded in a Thermo Nicolet 6700 FT-IR spectrometer using an ATR system, KBr discs or NaCl discs (Film) from the *Centres Científics i Tecnològics* or the Organic Chemistry Department of the University of Barcelona. The use of ATR, KBr discs or NaCl discs is indicated for each compound.

High resolution mass spectra (HRMS) were recorded in an LC/MSD-TOF G1969A (Agilent Technologies) from the *Centres Científics i Tecnològics* of the University of Barcelona.

6. Experimental methods

-

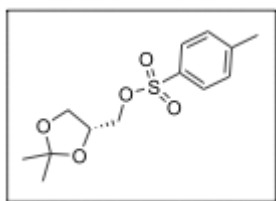
Melting points (Mp) were determined using a Büchi melting point apparatus and were not corrected.

Optical rotations ($[\alpha]_D$) were measured at rt (25 °C) using a Jasco P-2000 iRM800 polarimeter. Concentration is expressed in g/100 mL. The cell sized 10 cm long and had 1 mL of capacity. The measure of λ was 589 nm, which corresponds to a sodium lamp.

6.16.3. Synthetic procedures and characterisation

6.16.3.1. Synthesis of the starting materials

(R)-(2,2-dimethyl-1,3-dioxolan-4-yl)methyl 4-methylbenzenesulfonate



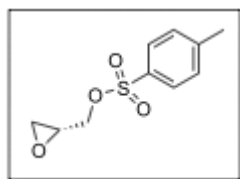
Crystallisation and purification of tosyl chloride:

10 g of tosyl chloride were dissolved in a minimum volume of chloroform (*ca* 25 mL), filtered and diluted with five volumes of petroleum ether (*ca* 125 mL). The tosyl acid and other impurities precipitated, the solution was filtered, clarified with charcoal and concentrated to obtain a white solid.

(*S*)-(+)-1,2-isopropylidene glycidol (0.47 mL, 3.78 mmol) was dissolved in pyridine (6 mL) under N₂ atmosphere. Then, the solution was cooled down in an ice bath and 4-DMAP (4.6 mg, 0.04 mmol) was added followed by the addition of tosyl chloride (1.08 g, 5.68 mmol) in small portions over 10 min. The reaction was allowed to warm up to rt and stirred overnight. The pyridine was evaporated and the resulting paste was diluted EtOAc (10 mL) and washed with H₂O (2×10 mL), 1 M HCl at 0 °C (5 mL), saturated aqueous solution of NaHCO₃ (10 mL) and H₂O (10 mL). The organic phase was dried over anhydrous MgSO₄, filtered and evaporated. The crude was purified by column chromatography on 25 g silica column equilibrated with 2% Et₃N in hexane and eluted with a gradient of 0-30% EtOAc in hexane over 10 CV. Compound-containing fractions were evaporated to obtain 1.08 g (100% yield) of the product.

The analytical data for this compound was in agreement with the reported characterisation ¹⁶.

(R)-oxiran-2-ylmethyl 4-methylbenzenesulfonate



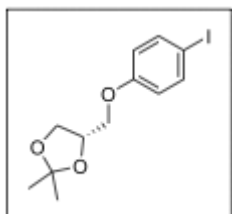
Tosyl chloride (5.43 g, 28.48 mmol) and DMAP (66 mg, 0.54 mmol) were dissolved in DCM (25 mL) and cooled down in an ice bath. Et₃N (4.2 mL, 29.83 mmol) and a solution of (*S*)-glycidol (1.8 mL, 27.12 mmol) were added dropwise. The mixture was allowed to warm up to rt and stirred for 3 h while a white precipitate formed. The suspension was diluted with DCM (15 mL) and washed with 5% w/v aq. sol. K₂CO₃ (40 mL), 1 M HCl (40 mL) and H₂O (40 mL). The organic phase was dried over anhydrous MgSO₄, filtered and evaporated. The crude was purified by column chromatography (120 g column, eluted with a gradient of 0-30% EtOAc in hexane) to obtain 3.94 g (64% yield) of the desired product as a colourless oil.

6. Experimental methods

The analytical data for this compound was in agreement with the reported characterisation ¹⁶.

6.16.3.2. Synthesis of the linker-expanded analogues

(S)-4-((4-iodophenoxy)methyl)-2,2-dimethyl-1,3-dioxolane



A suspension of Cs₂CO₃ (10.06 g, 30.88 mmol) and 4-iodophenol (4.53 g, 20.56 mmol) in 40 mL of anh. DMF was prepared under N₂ atm. and heated up to 80 °C for 20 min before dropwise addition of a solution of (*R*)-(2,2-dimethyl-1,3-dioxolan-4-yl)methyl 4-methylbenzenesulfonate (1.36 g, 10.29 mmol) in dry DMF (10 mL) over the course of 4 h. Then, the reaction was cooled to 50 °C, stirred overnight and finally quenched by addition of 150 mL

of H₂O. The resulting aqueous layer was extracted with 150 mL of EtOAc and this organic layer was washed with 1 M NaOH (100 mL), 40% NaOH (100 mL) and 10% v/w CuSO₄ (2×100 mL). The organic layer was dried over anh. MgSO₄, filtered and evaporated to obtain 1.33 g (39% yield) of the desired product.

$[\alpha]_D = +6.95$.

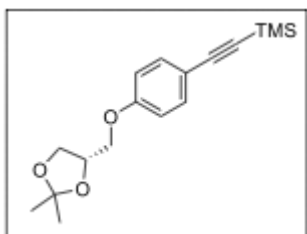
IR (Film): $\nu_{\max} = 2985, 2933, 1585, 1486, 1371, 1243, 1055 \text{ cm}^{-1}$.

¹H NMR (400 MHz, CDCl₃) δ 7.58 - 7.52 (m, 2H), 6.73 - 6.65 (m, 2H), 4.50 - 4.42 (m, 1H), 4.16 (dd, *J* = 8.5, 6.4 Hz, 1H), 4.01 (dd, *J* = 9.5, 5.4 Hz, 1H), 3.90 (dd, *J* = 9.8, 5.8 Hz, 1H), 3.88 (dd, *J* = 8.6, 5.8 Hz, 1H), 1.45 (bs, 3H), 1.40 (bs, 3H) ppm.

¹³C NMR (101 MHz, CDCl₃) δ 158.6, 138.4, 117.1, 110.0, 83.4, 74.0, 69.0, 66.9, 26.9, 25.5 ppm.

HRMS (ESI): *m/z* [M + H]⁺ calculated for C₁₂H₁₆O₃I: 335.0139; found: 335.0136.

(S)-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)ethynyl)trimethylsilane



(*S*)-4-((4-iodophenoxy)methyl)-2,2-dimethyl-1,3-dioxolane was dissolved in 25 mL of anh. DMF under N₂ atm in a Schlenk flask and CuI (21 mg, 0.11 mmol), Pd(Ph₃P)₂Cl₂ (78 mg, 0.11 mmol), trimethylsilylacetylene (0.46 mL, 3.41 mmol) and 16 mL of Et₃N were added to the solution under N₂ atm. The reaction was stirred for 1.5 h (the solution turns black), diluted with EtOAc (50 mL) and washed with brine (100 mL). The aqueous layer was extracted with EtOAc (2×150

mL) and the combined organic layers were dried over anh. MgSO₄, filtered and concentrated under vacuum. The crude was purified by column chromatography (40 g silica column equilibrated with 2% Et₃N in hexane and eluted with a gradient of 0-30% EtOAc in hexane) to obtain 451 mg (100% yield) of the expected product as a brown oil.

6. Experimental methods

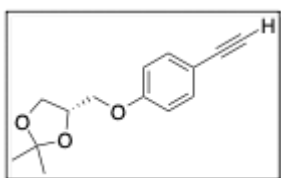
IR (ATR-FTIR): ν_{\max} = 2956, 2360, 2155, 1505, 838 cm^{-1} .

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.42 - 7.36 (m, 2H), 6.86 - 6.79 (m, 2H), 4.51 - 4.41 (m, 1H), 4.15 (dd, J = 8.5, 6.4 Hz, 1H), 4.04 (dd, J = 9.5, 5.4 Hz, 1H), 3.97 - 3.85 (m, 2H), 1.45 (bs, 3H), 1.40 (bs, 3H), 0.23 (s, 9H) ppm.

$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 158.8, 133.6, 115.9, 114.5, 109.9, 105.2, 92.8, 74.5, 68.9, 66.9, 26.9, 25.5, 0.2 ppm.

HRMS (ESI): m/z [$\text{M} + \text{H}$] $^+$ calculated for $\text{C}_{17}\text{H}_{24}\text{O}_3\text{Si}$: 304.1495; found: 305.1568.

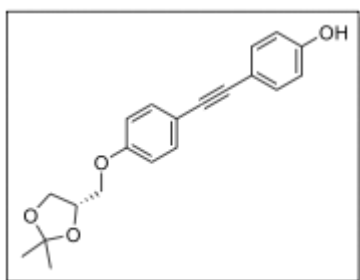
(S)-4-((4-ethynylphenoxy)methyl)-2,2-dimethyl-1,3-dioxolane



Potassium carbonate (107 mg, 0.78 mmol) was added to a solution of (S)-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)ethynyl)trimethylsilane (197 mg, 0.65 mmol) in MeOH (2 mL). The solution was stirred at rt for 1.5 h. The reaction mixture was partitioned between DCM (10 mL) and H_2O (10 mL) and the aqueous phase was extracted with DCM (2 \times 10 mL). The combined organic phases were dried over anh. MgSO_4 ,

filtered and evaporated to obtain the expected product (197 mg, 100% yield) as a brown oil. The product was used without further purification.

(S)-4-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)ethynyl)phenol



(S)-4-((4-ethynylphenoxy)methyl)-2,2-dimethyl-1,3-dioxolane (1.22 g, 5.23 mmol) was dissolved in DMF (60 mL) under N_2 atmosphere and CuI (50 mg, 0.26 mmol), $\text{Pd}(\text{Ph}_3\text{P})_2\text{Cl}_2$ (188 mg, 0.26 mmol), 4-iodophenol (1.28 g, 5.76 mmol) and 38 mL of Et_3N were added. The resultant solution was stirred at rt for 4 h. Then, the reaction mixture was diluted with EtOAc (150 mL) and washed with brine (150 mL) and 10% v/w CuSO_4 (150 mL). The aqueous layer was extracted with EtOAc (2 \times 150 mL) and the combined organic phases were

dried over anh. MgSO_4 , filtered and evaporated. The crude was purified by column chromatography on 80 g silica column equilibrated with 2% Et_3N in hexane and eluted with a gradient of 0-50% EtOAc in hexane. Compound-containing fractions were evaporated to obtain 1.21 g (71% yield) of the desired product as an orange solid.

Mp = 130 - 131 $^\circ\text{C}$.

IR (ATR-FTIR): ν_{\max} = 3305, 3216, 2913, 1728, 1177 cm^{-1} .

6. Experimental methods

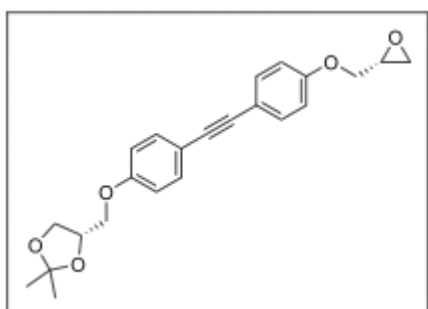
-

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.47 - 7.36 (m, 4H), 6.90 - 6.83 (m, 2H), 6.82 - 6.76 (m, 2H), 4.52 - 4.47 (m, 1H), 4.18 (dd, $J = 8.5, 6.4$ Hz, 1H), 4.07 (dd, $J = 9.6, 5.4$ Hz, 1H), 3.99 - 3.89 (m, 2H), 1.48 (bs, 3H), 1.42 (bs, 3H) ppm.

$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 158.3, 155.6, 133.1, 132.9, 116.2, 115.7, 115.5, 114.5, 109.9, 88.0, 87.7, 73.9, 68.7, 66.7, 26.8, 25.4 ppm.

HRMS (ESI) : m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{20}\text{H}_{20}\text{O}_4$: 324.1362; found: 325.1438.

(S)-2,2-dimethyl-4-(((4-(((R)-oxiran-2-yl)methoxy)phenyl)ethynyl)phenoxy)methyl)-1,3-dioxolane



A suspension of NaH (60% dispersion in oil, 74 mg, 1.84 mmol) in 0.5 mL of anh. DMF was prepared and cooled down to 0 °C. A solution of (S)-4-(((4-(((R)-oxiran-2-yl)methoxy)phenyl)ethynyl)phenol) (215 mg, 0.92 mmol) in 1 mL of anh. DMF was added dropwise and the resulting mixture was stirred at rt for 15 min. Then, a solution of (R)-oxiran-2-ylmethyl 4-methylbenzenesulfonate (419 mg, 1.84 mmol) in 1 mL of anh. DMF was added to the previous suspension and the mixture was stirred at rt overnight. The reaction was quenched by addition of 20 mL of NH_4Cl saturated solution and the resulting aqueous layer was extracted with EtOAc (3×20 mL). The combined organic extracts were washed with 10% v/w CuSO_4 (2×20 mL), dried over anh. MgSO_4 , filtered and evaporated. The crude was used in the following step without further purification (60% yield).

$M_p = 90 - 91$ °C.

$[\alpha]_D = +6.46$.

IR (KBr) : $\nu_{\text{max}} = 2923, 1605, 1516, 1242, 836$ cm^{-1} .

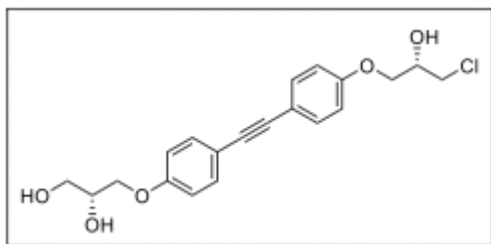
$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.47 - 7.40 (m, 4H), 6.91 - 6.84 (m, 4H), 4.48 (quint, $J = 5.8$ Hz, 1H), 4.24 (dd, $J = 11.0, 3.1$ Hz, 1H), 4.17 (dd, $J = 8.5, 6.4$ Hz, 1H), 4.07 (dd, $J = 9.5, 5.4$ Hz, 1H), 4.00 - 3.87 (m, 3H), 3.36 (ddt, $J = 5.7, 4.1, 2.8$ Hz, 1H), 2.92 (dd, $J = 4.9, 4.1$ Hz, 1H), 2.77 (dd, $J = 4.9, 2.6$ Hz, 1H), 1.47 (bs, 3H), 1.41 (bs, 3H) ppm.

$^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 158.4, 158.3, 132.9, 132.8, 116.3, 116.1, 114.6, 114.5, 109.8, 88.0, 87.9, 73.9, 68.8, 68.7, 66.8, 50.0, 44.7, 27.8, 23.4 ppm.

HRMS (ESI) : m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{23}\text{H}_{25}\text{O}_5$ 381.1697; found 381.1697.

6. Experimental methods

(R)-3-(4-((4-((S)-3-chloro-2-hydroxypropoxy)phenyl)ethynyl)phenoxy)propane-1,2-diol



(S)-2,2-dimethyl-4-((4-((R)-oxiran-2-yl)methoxy)phenyl)ethynyl)phenoxy)methyl)-1,3-dioxolane (142 mg, 0.13 mmol) was dissolved in ACN (10 mL). Then, $\text{CeCl}_3 \cdot 7\text{H}_2\text{O}$ (346 mg, 0.93 mmol) was added and the mixture refluxed at 100 °C overnight. The reaction mixture was allowed to cool down to rt and evaporated.

The paste was dissolved in MeOH, filtered through a Celite® pad and evaporated. The crude was purified by column chromatography (12 g silica column, eluted with a gradient of 30-100% EtOAc in hexane). Compound-containing fractions were evaporated to obtain the expected product (30 mg, 21 % yield) as a white solid.

$M_p = 141 - 142$ °C.

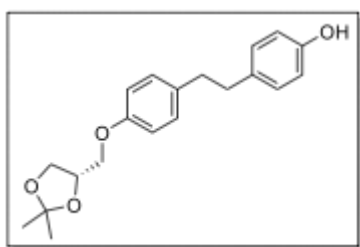
IR (KBr): $\nu_{\text{max}} = 3351, 2924, 1517, 1248, 1037, 832$ cm^{-1} .

¹H NMR (400 MHz, CD₃OD) δ 7.46 - 7.39 (m, 4H), 7.01 - 6.94 (m, 4H), 4.19 - 4.05 (m, 4H), 4.04 - 3.95 (m, 2H), 3.79 (dd, $J = 11.3, 4.9$ Hz, 1H), 3.76 - 3.63 (m, 3H) ppm.

¹³C NMR (100 MHz, CD₃OD) δ 160.3, 160.0, 113.8, 113.7, 117.5, 117.2, 115.8, 115.7, 88.8, 88.7, 71.7, 70.9, 70.4, 70.2, 64.1, 46.6 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{20}\text{H}_{22}\text{ClO}_5$ 377.1150; found 377.1149.

(S)-4-(4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenethyl)phenol



A solution of (S)-4-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)ethynyl)phenol (103 mg, 0.32 mmol) in 8 mL of MeOH and 1.5 mL of DCM and 34 mg of Pd over C were added to this solution. The mixture was purged with N_2 and with H_2 and stirred at rt overnight. The resulting suspension was then filtered through a Celite® pad, washed with DCM and evaporated to obtain 101 mg (97% yield) of the expected product.

$M_p = 116 - 118$ °C.

IR (ATR-FTIR): $\nu_{\text{max}} = 2990, 2923, 2847, 1611, 1510, 1452, 1370, 1241, 1212, 1050, 1031, 823$ cm^{-1} .

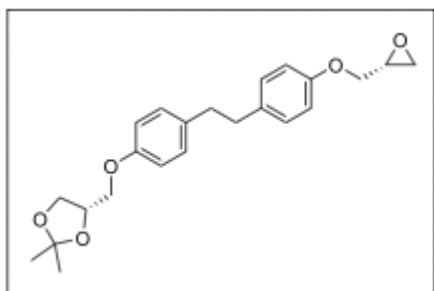
¹H NMR (400 MHz, CD₃OD) δ 7.05 - 7.00 (m, 2H), 6.96 - 6.91 (m, 2H), 6.85 - 6.78 (m, 2H), 6.68 - 6.64 (m, 2H), 4.45 - 4.39 (m, 1H), 4.13 (dd, $J = 8.4, 6.5$ Hz, 1H), 3.95 (dd, $J = 5.4, 4.3$ Hz, 2H), 3.84 (dd, $J = 8.4, 6.2$ Hz, 1H), 3.66 (qd, $J = 11.3, 5.2$ Hz, 1H), 2.80 - 2.71 (m, 4H), 1.41 (s, 3H), 1.36 (s, 3H) ppm.

6. Experimental methods

^{13}C NMR (101 MHz, CD_3OD) δ 156.9, 153.9, 153.9, 134.6, 134.58, 134.0, 134.0, 129.7, 129.5, 115.3, 114.5, 109.9, 109.9, 74.2, 69.0, 67.0, 37.4, 37.4, 26.9, 25.5 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{20}\text{H}_{25}\text{O}_4$: 329.1747; found: 329.1756.

(S)-2,2-dimethyl-4-((4-(4-(((R)-oxiran-2-yl)methoxy)phenethyl)phenoxy)methyl)-1,3-dioxolane



A suspension of NaH (60% dispersion in oil, 25 mg, 0.61 mmol) in 0.5 mL of anh. DMF was prepared and cooled down to 0 °C. A solution of (S)-4-(4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenethyl)phenol (101 mg, 0.31 mmol) in 1 mL of anh. DMF was added dropwise and the resulting mixture was stirred at rt for 15 min. Then, a solution of (R)-oxiran-2-ylmethyl 4-methylbenzenesulfonate (140 mg, 0.61 mmol) in 1 mL of anh. DMF was added to the previous suspension and the

mixture was stirred at rt overnight. The reaction was quenched by addition of 20 mL of NH_4Cl saturated solution and the resulting aqueous layer was extracted with EtOAc (3×20 mL). The combined organic extracts were washed with 10% v/w CuSO_4 (2×20 mL), dried over anh. MgSO_4 , filtered and evaporated. The crude was purified by column chromatography (12 g silica column equilibrated with 2% Et_3N in hexane and eluted with a gradient of 0-30% EtOAc in hexane) to obtain 86 mg (73% yield) of the expected product.

$M_p = 73 - 74$ °C.

IR (ATR-FTIR): $\nu_{\text{max}} = 2987, 2932, 1611, 1513, 1251$ cm^{-1} .

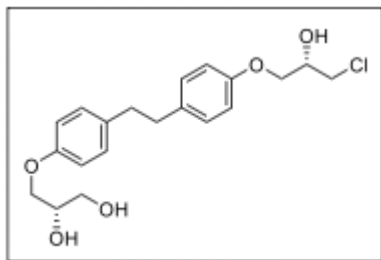
^1H NMR (400 MHz, CDCl_3) δ 7.08 - 7.02 (m, 4H), 6.87 - 6.78 (m, 4H), 4.47 (quint, $J = 6.0$ Hz, 1H), 4.21 - 4.13 (m, 2H), 4.04 (dd, $J = 9.5, 5.4$ Hz, 1H), 3.97 - 3.86 (m, 4H), 3.35 (dddd, $J = 5.8, 4.1, 3.2, 2.6$ Hz, 1H), 2.90 (dd, $J = 5.0, 4.1$ Hz, 1H), 2.82 (s, 4H), 2.75 (dd, $J = 5.0, 2.6$ Hz, 1H), 1.47 (bs, 3H), 1.41 (bs, 3H) ppm.

^{13}C NMR (100 MHz, CDCl_3) δ 157.0, 156.9, 134.6, 134.5, 129.6, 129.5, 114.6, 114.5, 109.9, 74.2, 69.0, 69.0, 67.1, 50.4, 44.9, 37.4, 37.4, 27.0, 25.5 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{23}\text{H}_{29}\text{O}_5$ 385.2010; found 385.2020.

6. Experimental methods

(R)-3-(4-(4-((S)-3-chloro-2-hydroxypropoxy)phenethyl)phenoxy)propane-1,2-diol



The starting material ((R)-3-(4-(4-((S)-3-chloro-2-hydroxypropoxy)phenethyl)phenoxy)propane-1,2-diol, 75 mg, 0.19 mmol) was dissolved in 5 mL of ACN. Then, CeCl₃·7H₂O (182 mg, 0.49 mmol) was added and the mixture refluxed at 100 °C overnight. The reaction mixture was allowed to cool down to rt and evaporated. The paste was dissolved in MeOH, filtered through a Celite® pad and evaporated. The crude was purified by column chromatography (4 g silica column, eluted with a gradient of 50-

100% EtOAc in hexane). Compound-containing fractions were evaporated to obtain the expected product (36 mg, 48 % yield) as a white solid.

Mp = 133 - 134 °C.

$[\alpha]_D = +9.28$.

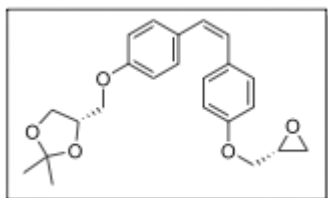
IR (ATR-FTIR): $\nu_{\max} = 3338, 3029, 2921, 2844, 1608, 1510, 1240, 1036, 824, 810 \text{ cm}^{-1}$.

¹H NMR (400 MHz, CDCl₃) δ 7.06 - 6.98 (m, 4H), 6.82 - 6.76 (m, 4H), 4.10 (quint, *J* = 5.2 Hz, 1H), 3.97 - 4.04 (m, 3H), 3.89 - 3.95 (m, 2H), 3.75 (dd, *J* = 11.3, 4.9 Hz, 1H), 3.60 - 3.70 (m, 3H), 2.79 (s, 4H) ppm.

¹³C NMR (100 MHz, CDCl₃) δ 158.6, 158.3, 135.7, 135.4, 130.5, 130.4, 115.4, 115.3, 71.9, 71.0, 70.3, 70.1, 64.2, 38.4, 38.4 ppm.

HRMS (ESI): *m/z* [M + H]⁺ calculated for C₂₀H₂₆ClO₅ 381.1463; found 381.1462.

(S)-2,2-dimethyl-4-((4-((Z)-4-(((R)-oxiran-2-yl)methoxy)styryl)phenoxy)methyl)-1,3-dioxolane



In a 25 mL round bottom flask with a magnetic stirrer, (S)-2,2-dimethyl-4-((4-((4-(((R)-oxiran-2-yl)methoxy)phenyl)ethynyl)phenoxy)methyl)-1,3-dioxolane (350 mg, 0.92 mmol) was dissolved in a 1:1 mixture of hexanes and toluene (8 mL). Quinoline (0.11 mL, 0.92 mmol) and Pd/BaSO₄ (7 mol%) were added to the starting material, and the suspension was put under H₂ (balloon). Stirring was kept for 3 h, and the crude was then filtered through a Celite® pad. The solvent was removed under vacuum and the product purified by column chromatography (12 g silica column, equilibrated with 2% Et₃N in hexane and eluted with a gradient of 0-100% EtOAc in hexane). The product was isolated as a mixture (Z isomer and starting epoxide) that was not further purified. The reaction wasn't reproducible.

6. Experimental methods

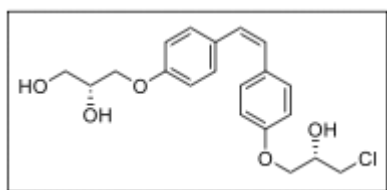
-

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.21 - 7.14 (m, 4H), 6.80 - 6.72 (m, 4H), 6.44 (s, 2H), 4.47 (quint, $J = 5.9$ Hz, 1H), 4.23 - 4.12 (m, 2H), 4.04 (dd, $J = 9.5, 5.4$ Hz, 1H), 3.97 - 3.87 (m, 3H), 3.34 (m, 1H), 2.90 (dd, $J = 5.0, 4.1$ Hz, 1H), 2.75 (dd, $J = 4.9, 2.6$ Hz, 1H), 1.46 (s, 3H), 1.40 (s, 3H) ppm.

$^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 157.5, 157.4, 130.5, 130.4, 130.1, 130.0, 128.5, 128.4, 114.3, 114.2, 109.7, 74.0, 68.7, 68.6, 66.8, 50.1, 44.7, 26.8, 25.4 ppm.

HRMS (ESI) : m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{23}\text{H}_{27}\text{O}_5$ 383.1859; found 383.1858.

(R)-3-(4-((Z)-4-((S)-3-chloro-2-hydroxypropoxy)styryl)phenoxy)propane-1,2-diol



(Z)-2,2-dimethyl-4-((4-(4-(oxiran-2-yl)methoxy)styryl)phenoxy)methyl)-1,3-dioxolane (73 mg, 0.19 mmol) was dissolved in 5 mL of ACN and $\text{CeCl}_3 \cdot 7\text{H}_2\text{O}$ (179 mg, 0.48 mmol) was added. The suspension was refluxed at 100 °C overnight. After this time, the mixture was concentrated, redissolved in MeOH and filtered through a Celite® pad. The solvent was removed under vacuum

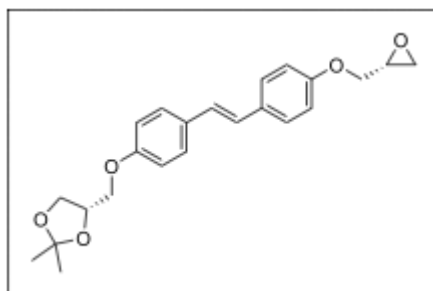
and purified by column chromatography (eluted with a gradient of 0-10% MeOH in DCM). The final product was isolated as a white solid (36 mg, 48% yield) that turned out to be a mixture of isomers (75 % of *Z* and 25 % of *E*).

$^1\text{H NMR}$ (400 MHz, CD_3OD) δ 7.21 - 7.14 (m, 4H), 6.85 - 6.81 (m, 4H), 6.46 (s, 2H) 4.09 - 4.16 (m, 1H), 4.02 - 4.07 (m, 3H), 3.94 - 3.99 (m, 2H), 3.77 (dd, $J = 11.3, 5.0$ Hz, 1H), 3.66 - 3.71 (m, 3H) ppm.

$^{13}\text{C NMR}$ (100 MHz, CD_3OD) δ 46.7, 64.2, 70.0, 70.3, 71.0, 71.8, 115.2, 115.3, 129.4, 129.6, 131.0, 131.1, 131.5, 131.8, 159.1, 159.4 ppm.

HRMS (ESI) : m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{20}\text{H}_{24}\text{ClO}_5$ 379.1312; found $[\text{M} + \text{H}]^+$ 379.1310.

(S)-2,2-dimethyl-4-((4-((E)-4-(((R)-oxiran-2-yl)methoxy)styryl)phenoxy)methyl)-1,3-dioxolane



A mixture of the *Z* and the *E* isomers and the totally reduced product (37 mg) was dissolved in 3 mL of MeOH and irradiated with a set of 237 nm wavelength lamps in the Rayonet[□] reactor in the presence of CuCl (3 mg) in a quartz flask. After 1 h, the solvent was concentrated and the crude was purified by column chromatography (4 g silica column, equilibrated with 2% Et_3N in hexane and eluted with a gradient of 0-30% EtOAc in hexane) to obtain 15 mg of the *E* isomer

(together with the totally reduced product).

6. Experimental methods

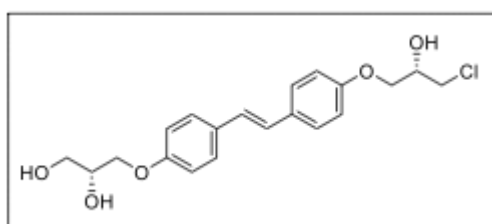
-

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.44 - 7.38 (m, 4H), 6.92 (s, 2H), 6.92 - 6.88 (m, 4H), 4.49 (quint, $J = 5.9$ Hz, 1H), 4.27 - 4.14 (m, 2H), 4.08 (dd, $J = 9.5, 5.4$ Hz, 1H), 4.01 - 3.88 (m, 3H), 3.39 - 3.33 (m, 1H), 2.91 (dd, $J = 4.9, 4.1$ Hz, 1H), 2.77 (dd, $J = 4.9, 2.7$ Hz, 1H), 1.47 (d, $J = 0.7$ Hz, 3H), 1.41 (d, $J = 0.7$ Hz, 3H) ppm.

$^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 158.2, 158.1, 131.1, 131.0, 127.6, 127.6, 126.5, 126.4, 115.0, 114.9, 109.9, 74.2, 69.0, 69.0, 67.0, 50.3, 44.9, 27.0, 25.5 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{23}\text{H}_{27}\text{O}_5$ 383.1853; found 383.1850.

(R)-3-(4-((E)-4-((S)-3-chloro-2-hydroxypropoxy)styryl)phenoxy)propane-1,2-diol



(S)-2,2-dimethyl-4-((4-((E)-4-((R)-oxiran-2-yl)methoxy)styryl)phenoxy)methyl)-1,3-dioxolane (15.3 mg, 0.04 mmol) was dissolved in 1 mL of acetonitrile. Then, $\text{CeCl}_3 \cdot 7\text{H}_2\text{O}$ (37.7 mg, 0.1 mmol) was added. The suspension was refluxed at 100 °C overnight. After this time, the mixture was concentrated, redissolved in MeOH

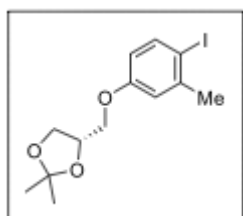
and filtered through a Celite® pad. The solvent was removed under vacuum and purified by column chromatography (eluted with a gradient of 0-10% MeOH in DCM) to obtain the final product (7.2 mg, 48% yield).

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.32 - 7.26 (m, 4H), 6.83 - 6.73 (m, 6H), 3.95 (dd, $J = 5.3, 1.1$ Hz, 2H), 3.67 - 3.59 (m, 2H), 3.58 - 3.52 (m, 2H), 3.22 (dq, $J = 3.2, 1.5$ Hz, 4H) ppm.

$^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 158.09, 157.82, 130.87, 130.64, 127.33, 126.24, 126.05, 114.67, 114.61, 70.32, 69.47, 68.97, 68.69, 63.29, 45.70, 29.55 ppm.

6.16.3.3. Synthesis of CAM family of compounds

(S)-4-((4-iodo-3-methylphenoxy)methyl)-2,2-dimethyl-1,3-dioxolane



A mixture of 4-iodo-2-methylphenol (1.05 g, 4.47 mmol) and Cs_2CO_3 (2.92 g, 8.95 mmol) in 10 mL of anh. DMF was prepared under N_2 atm. and heated up to 80 °C for 15 min. Then, a solution of (R)-(2,2-dimethyl-1,3-dioxolan-4-yl)methyl 4-methylbenzenesulfonate (1.18 g, 8.95 mmol) in dry DMF (5 mL) was added dropwise. The reaction was stirred overnight at 80 °C and quenched by addition of 50 mL of a saturated KHCO_3 solution. The resulting aqueous layer was extracted with EtOAc (3×50 mL). The extracted organic layers were concentrated and washed with 10% v/w CuSO_4 (2×20 mL). The organic layer was dried over anh. MgSO_4 , filtered and evaporated. The resulting crude was purified by column chromatography (25 g column, eluted with a gradient of 0-100% EtOAc in hexane) to obtain 1.36 g (87% yield) of the desired product as a yellow oil.

6. Experimental methods

$[\alpha]_D = +16.57$.

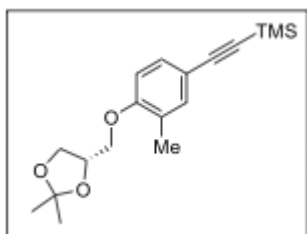
IR (Film): $\nu_{\max} = 2985, 2933, 2980, 1588, 1490, 1246, 1055, 842 \text{ cm}^{-1}$.

$^1\text{H NMR (400 MHz, CDCl}_3)$ δ 7.45 - 7.39 (m, 2H), 6.60 - 6.56 (m, 1H), 4.46 (qd, $J = 6.1, 4.7 \text{ Hz}$, 1H), 4.16 (dd, $J = 8.4, 6.3 \text{ Hz}$, 1H), 4.03 (dd, $J = 9.6, 4.7 \text{ Hz}$, 1H), 3.97 - 3.89 (m, 2H), 2.17 (s, 2H), 1.46 (d, $J = 0.7 \text{ Hz}$, 3H), 1.40 (d, $J = 0.7 \text{ Hz}$, 3H) ppm.

$^{13}\text{C NMR (101 MHz, CDCl}_3)$ δ 156.7, 139.3, 135.6, 129.8, 113.4, 109.80, 83.3, 74.1, 68.7, 66.9, 26.9, 25.5, 16.0 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{13}\text{H}_{18}\text{O}_3\text{I}$: 349.0295; found: 349.0297.

(S)-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)-3-methylphenyl)ethynyl)trimethylsilane



A suspension of the starting material ((S)-4-((4-iodo-3-methylphenoxy)methyl)-2,2-dimethyl-1,3-dioxolane, 357 mg, 1.02 mmol), CuI (2.0 mg, 0.01 mmol) and Pd(Ph₃P)₂Cl₂ (7.2 mg, 0.01 mmol) was prepared in 3 mL of anh. THF. Then, trimethylsilylacetylene (0.22 mL, 1.54 mmol) and 3 mL of Et₃N were added and the resulting mixture was stirred for 2.5 h. The solution was filtered through a Celite® pad and concentrated under vacuum. The resulting crude was purified by column

chromatography (25 g column, eluted with a gradient of 0-100% EtOAc in hexane) to accomplish 324 mg (99% yield) of the desired compound as a yellow oil.

$[\alpha]_D = +19.68$.

IR (Film): $\nu_{\max} = 2986, 2958, 2897, 2149, 1501, 1229, 843 \text{ cm}^{-1}$.

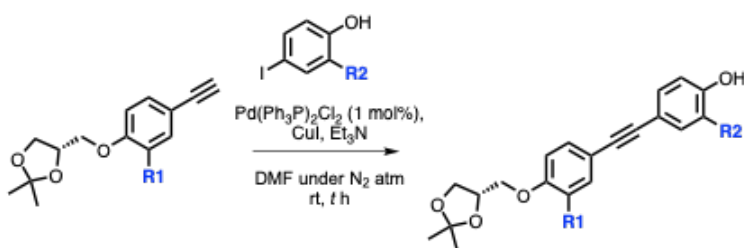
$^1\text{H NMR (400 MHz, CDCl}_3)$ δ 7.29 - 7.25 (m, 2H), 6.74 - 6.70 (m, 1H), 4.47 (qd, $J = 6.1, 4.6 \text{ Hz}$, 1H), 4.16 (dd, $J = 8.4, 6.3 \text{ Hz}$, 1H), 4.07 (dd, $J = 9.6, 4.6 \text{ Hz}$, 1H), 3.98 - 3.92 (m, 2H), 2.18 (s, 2H), 1.46 (d, $J = 0.7 \text{ Hz}$, 3H), 1.40 (d, $J = 0.7 \text{ Hz}$, 3H), 0.23 (s, 9H) ppm.

$^{13}\text{C NMR (101 MHz, CDCl}_3)$ δ 157.1, 134.5, 131.1, 127.0, 115.4, 110.8, 109.8, 105.5, 92.4, 74.1, 68.6, 67.0, 26.9, 25.6, 16.1, 0.2 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{18}\text{H}_{27}\text{O}_3\text{Si}$: 319.1724; found: 319.1718.

6. Experimental methods

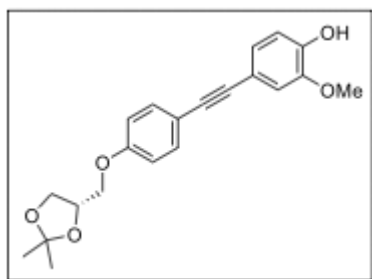
6.16.3.3.1. General procedure P1 for Sonogashira coupling:



R1	R2	time (h)	yield (%)
H	OMe	5	75
H	F	6	79
H	Me	5	85
H	Ph	18	78
H	^t Bu	18	71
Me	H	16	87
Me	Me	2.5	87

The starting building block (1 eq.) was dissolved in a mixture of THF and Et₃N 1:1 (0.3 M) under N₂ atm. Then, CuI (0.02 eq.), Pd(PPh₃)₂Cl₂ (0.02 eq.) and the corresponding substituted iodophenol (1 eq.) were added. The resultant solution was stirred at rt for the required time. After completion, the reaction mixture was partitioned between H₂O and DCM and the aqueous phase was extracted with DCM. The combined organic phases were dried over anh. MgSO₄, filtered and concentrated under vacuum. The crude was purified by column chromatography (0-50% or 0-100% EtOAc in hexane) to yield the expected product.

(*S*)-4-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)ethynyl)-2-methoxyphenol



Mp = 124 - 125 °C.

$[\alpha]_D = +7.14$.

¹H NMR (400 MHz, CDCl₃) δ 7.46 - 7.40 (m, 2H), 7.07 (dd, *J* = 8.2, 1.8 Hz, 1H), 7.01 (d, *J* = 1.8 Hz, 1H), 6.88 (dd, *J* = 8.5, 1.7 Hz, 3H), 5.72 (s, 1H), 4.48 (quint, *J* = 5.9 Hz, 1H), 4.17 (dd, *J* = 8.5, 6.4 Hz, 1H), 4.07 (dd, *J* = 9.5, 5.4 Hz, 1H), 3.96 (dd, *J* = 9.5, 5.9 Hz, 1H),

3.91 (dd, *J* = 8.5, 5.8 Hz, 1H), 3.91 (s, 3H), 3.88 (m, 4H), 1.47 (bs, 3H), 1.41 (bs, 3H) ppm.

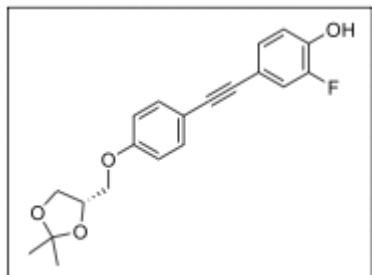
¹³C NMR (100 MHz, CDCl₃) δ 146.3, 146.2, 133.0, 125.6, 116.3, 115.2, 114.7, 114.7, 113.8, 110.0, 88.5, 87.5, 74.1, 68.9, 66.9, 56.1, 26.9, 25.5 ppm.

IR (ATR-FTIR): ν_{\max} = 3355, 2977, 2939, 2913, 2857, 1603, 1517, 1251, 1229, 1014, 825 cm⁻¹.

HRMS (ESI): *m/z* [M + H]⁺ calculated for C₂₃H₂₉O₅ 385.2010; found 385.2007.

6. Experimental methods

(S)-4-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)ethynyl)-2-fluorophenol



Mp = 128 - 129 °C.

$[\alpha]_D = +10.52$.

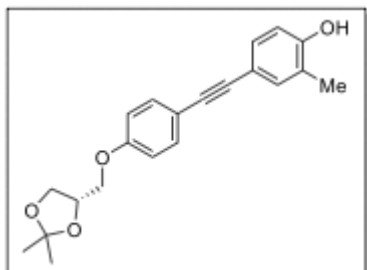
IR (ATR-FTIR): $\nu_{\max} = 3205, 2990, 2930, 2896, 2870, 1611, 1577, 1519, 1326, 1062, 827 \text{ cm}^{-1}$.

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.46 - 7.40 (m, 2H), 7.23 (dd, $J = 11.1, 1.9 \text{ Hz}$, 1H), 7.19 (ddd, $J = 8.4, 2.0, 1.1 \text{ Hz}$, 1H), 6.99 - 6.92 (m, 1H), 6.91 - 6.85 (m, 2H), 5.41 (d, $J = 3.9 \text{ Hz}$, 1H), 4.49 (quint, $J = 5.8 \text{ Hz}$, 1H), 4.18 (dd, $J = 8.5, 6.4 \text{ Hz}$, 1H), 4.07 (dd, $J = 9.5, 5.4 \text{ Hz}$, 1H), 3.96 (dd, $J = 9.5, 5.8 \text{ Hz}$, 1H), 3.91 (dd, $J = 8.5, 5.8 \text{ Hz}$, 1H), 1.47 (bs, 3H), 1.41 (bs, 3H) ppm.

$^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 158.7, 151.8, 149.4, 144.2, 144.1, 133.1, 128.7, 128.7, 118.8, 118.6, 117.5, 117.5, 116.2, 116.1, 115.9, 114.7, 110.1, 88.4, 87.2, 87.2, 74.1, 68.9, 66.9, 26.9, 25.5 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{20}\text{H}_{20}\text{O}_4\text{F}$ 343.1340; found 343.1347.

(S)-4-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)ethynyl)-2-methylphenol



Mp = 124 - 125 °C.

$[\alpha]_D = +8.97$.

IR (ATR-FTIR): $\nu_{\max} = 3385, 2969, 2926, 2891, 1607, 1512, 1049, 835, 823 \text{ cm}^{-1}$.

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.45 - 7.38 (m, 2H), 7.30 (dd, $J = 2.1, 0.9 \text{ Hz}$, 1H), 7.24 (ddd, $J = 8.2, 2.1, 0.6 \text{ Hz}$, 1H), 6.89 - 6.83 (m, 2H), 6.72 (d, $J = 8.2 \text{ Hz}$, 1H), 5.01 (s, 1H), 4.49 (quint, $J = 5.8 \text{ Hz}$, 1H), 4.17 (dd, $J = 8.5, 6.4 \text{ Hz}$, 1H), 4.06 (dd, $J = 9.5, 5.4 \text{ Hz}$, 1H), 3.96 (dd, $J = 9.5, 5.8 \text{ Hz}$, 1H), 3.91 (dd, $J = 8.5, 5.8 \text{ Hz}$, 1H), 2.24 (s, 3H), 1.47 (bs, 3H), 1.41 (bs, 3H) ppm.

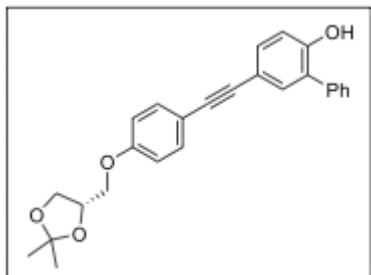
$^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 158.4, 154.1, 134.4, 133.0, 130.7, 116.5, 115.8, 115.1, 114.7, 110.0, 88.4, 87.6, 74.1, 68.9, 66.9, 26.9, 25.5, 15.7 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{21}\text{H}_{23}\text{O}_4$ 339.1591; found 339.1587.

6. Experimental methods

-

(S)-5-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)ethynyl)-[1,1'-biphenyl]-2-ol



Mp = 101 - 102 °C.

$[\alpha]_D = +5.98$.

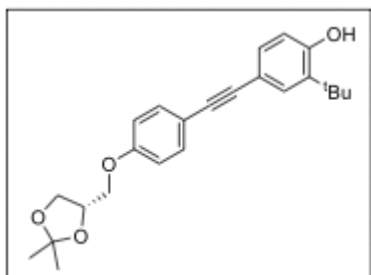
IR (ATR-FTIR): $\nu_{\max} = 3312, 2990, 2926, 2866, 1607, 1508, 1272, 1224, 1056, 1036, 822, 698 \text{ cm}^{-1}$.

$^1\text{H NMR (400 MHz, CDCl}_3)$ δ 7.54 - 7.39 (m, 9H), 6.97 - 6.94 (m, 1H), 6.90 - 6.85 (m, 2H), 5.34 (s, 1H), 4.48 (quint, $J = 5.8 \text{ Hz}$, 1H), 4.17 (dd, $J = 8.5, 6.4 \text{ Hz}$, 1H), 4.07 (dd, $J = 9.6, 5.4 \text{ Hz}$, 1H), 3.96 (dd, $J = 9.5, 5.9 \text{ Hz}$, 1H), 3.90 (dd, $J = 8.5, 5.8 \text{ Hz}$, 1H), 1.47 (bs, 3H), 1.41 (bs, 3H) ppm.

$^{13}\text{C NMR (100 MHz, CDCl}_3)$ δ 152.8, 152.8, 136.6, 133.7, 132.9, 132.3, 129.2, 129.1, 128.6, 128.5, 128.0, 116.3, 116.2, 115.9, 114.6, 109.9, 88.1, 88.0, 68.7, 66.7, 26.8, 25.4 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{26}\text{H}_{25}\text{O}_4$ 401.1747; found 401.1743.

(S)-2-(tert-butyl)-4-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)phenyl)ethynyl)phenol



Mp = 111 - 113 °C.

$[\alpha]_D = +6.26$.

IR (ATR-FTIR): $\nu_{\max} = 3373, 2960, 2922, 2887, 2866, 1606, 1510, 1404, 1368, 1243, 1202, 1053, 828 \text{ cm}^{-1}$.

$^1\text{H NMR (400 MHz, CDCl}_3)$ δ 7.48 - 7.40 (m, 3H), 7.23 (dd, $J = 8.1, 2.0 \text{ Hz}$, 1H), 6.90 - 6.84 (m, 2H), 6.63 (d, $J = 8.2 \text{ Hz}$, 1H), 4.49 (quint, $J = 5.8 \text{ Hz}$, 1H), 4.18 (dd, $J = 8.5, 6.4 \text{ Hz}$, 1H), 4.07 (dd, $J = 9.5, 5.4 \text{ Hz}$, 1H), 3.96 (dd, $J = 9.6, 5.9 \text{ Hz}$, 1H), 3.91 (dd, $J = 8.5, 5.8 \text{ Hz}$, 1H), 1.47 (s, 3H), 1.41 (s, 12H) ppm.

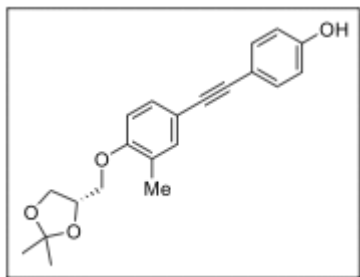
$^{13}\text{C NMR (100 MHz, CDCl}_3)$ δ 158.4, 154.5, 136.5, 133.0, 130.9, 130.5, 116.8, 116.5, 115.5, 114.7, 110.0, 88.8, 87.4, 74.1, 68.9, 66.9, 34.8, 29.6, 26.9, 25.5 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{24}\text{H}_{29}\text{O}_4$ 381.2060; found 381.2060.

6. Experimental methods

-

(S)-4-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)-3-methylphenyl)ethynyl)phenol



Mp = 134 - 136 °C.

$[\alpha]_D = +19.28$.

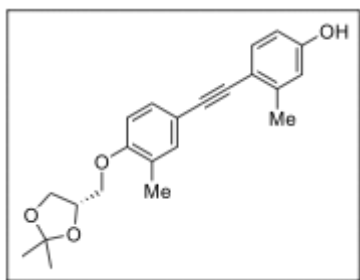
IR (Film): $\nu_{\max} = 3298, 2089, 1608, 1513, 1238, 1041, 835, 805 \text{ cm}^{-1}$.

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.39 - 7.34 (m, 2H), 7.35 - 7.28 (m, 2H), 6.80 - 6.77 (m, 2H), 6.77 - 6.73 (m, 1H), 6.03 (s, 1H), 4.51 (qd, $J = 6.0, 4.7 \text{ Hz}$, 1H), 4.19 (dd, $J = 8.5, 6.4 \text{ Hz}$, 1H), 4.08 (dd, $J = 9.7, 4.7 \text{ Hz}$, 1H), 4.02 - 3.94 (m, 2H), 2.20 (s, 2H), 1.49 (d, $J = 0.8 \text{ Hz}$, 3H), 1.43 (d, $J = 0.8 \text{ Hz}$, 3H) ppm.

$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 156.7, 155.9, 133.9, 133.2, 130.5, 127.1, 115.9, 115.8, 115.6, 111.0, 110.0, 88.1, 87.9, 74.2, 68.5, 66.8, 26.8, 25.5, 16.2 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{21}\text{H}_{22}\text{O}_4$: 338.1513; found: 338.1506.

(S)-4-((4-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)-3-methylphenyl)ethynyl)-3-methylphenol



Mp = 107 - 109 °C.

$[\alpha]_D (\text{MeOH}) = +23.23$.

IR (Film): $\nu_{\max} = 2985, 2933, 1585, 1486, 1371, 1243, 1055 \text{ cm}^{-1}$.

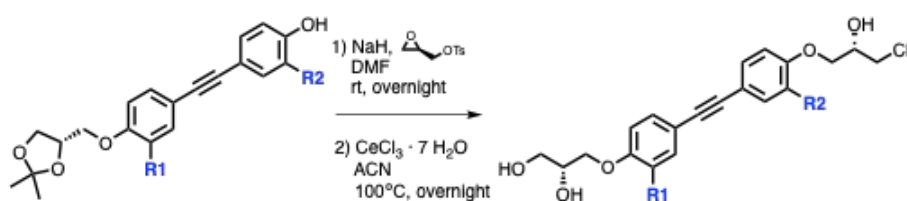
$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.32 - 7.28 (m, 3H), 7.23 (dd, $J = 8.3, 2.1 \text{ Hz}$, 1H), 6.79 - 6.75 (m, 1H), 6.72 (d, $J = 8.2 \text{ Hz}$, 1H), 4.89 (s, 1H), 4.49 (qd, $J = 6.1, 4.6 \text{ Hz}$, 1H), 4.18 (dd, $J = 8.4, 6.3 \text{ Hz}$, 1H), 4.09 (dd, $J = 9.6, 4.6 \text{ Hz}$, 1H), 3.98 (dd, $J = 6.0, 1.3 \text{ Hz}$, 1H), 3.96 (d, $J = 5.9 \text{ Hz}$, 1H), 2.24 (s, 3H), 2.21 (s, 3H), 1.47 (s, 3H), 1.41 (s, 3H) ppm.

$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 156.7, 154.0, 134.4, 134.0, 130.7, 130.4, 127.2, 124.1, 115.9, 115.1, 111.0, 109.9, 88.0, 87.9, 77.5, 77.2, 76.8, 74.2, 74.2, 69.0, 67.0, 67.0, 26.9, 25.6, 25.6, 16.2, 15.7 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{21}\text{H}_{28}\text{O}_5\text{Cl}$: 395.1620; found: 395.1617.

6. Experimental methods

6.16.3.3.2. General procedure P2 for epoxide formation and opening:

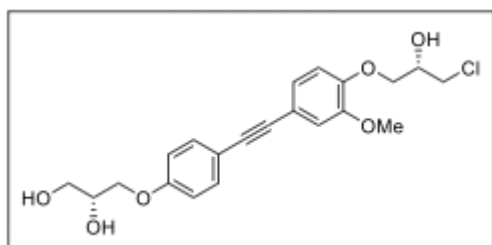


R1	R2	overall yield (%)
H	OMe	34
H	F	39
H	Me	52
H	Ph	42
H	^t Bu	59
Me	H	64
Me	Me	81

A suspension of NaH (60% dispersion in oil, 2 eq.) in DMF (0.6 M) was prepared and cooled down to 0 °C. A solution of the corresponding phenol (1 eq., 0.3 M) in anh. DMF was added dropwise and the resulting mixture was stirred at rt for 15 min. Then, a solution of (*R*)-oxiran-2-ylmethyl 4-methylbenzenesulfonate (1.2 eq., 0.3 M) in anh. DMF was added to the previous suspension and the mixture was stirred at rt overnight. The reaction was quenched by addition of NH_4Cl saturated solution and the resulting aqueous layer was extracted with EtOAc, dried over anh. MgSO_4 , filtered and evaporated. The crude was used in the following step without further purification.

The corresponding protected epoxide (1 eq.) was dissolved in ACN (0.04 M) and $\text{CeCl}_3 \cdot 7\text{H}_2\text{O}$ (2.5 eq.) was added. The suspension was refluxed at 100 °C overnight. After this time, the mixture was concentrated, redissolved in MeOH and filtered through a Celite® pad. The solvent was removed under vacuum and the crude was purified by column chromatography (50-100% EtOAc in hexane or 0%-20% MeOH in DCM) to obtain the final product.

(*R*)-3-(4-(((4-((*S*)-3-chloro-2-hydroxypropoxy)-3-methoxyphenyl)ethynyl)phenoxy)propane-1,2-diol



$M_p = 134 - 135 \text{ }^\circ\text{C}$.

$[\alpha]_D = -2.99$.

IR (ATR-FTIR): $\nu_{\text{max}} = 3317, 2922, 2857, 1606, 1517, 1246, 1221, 1026, 833 \text{ cm}^{-1}$.

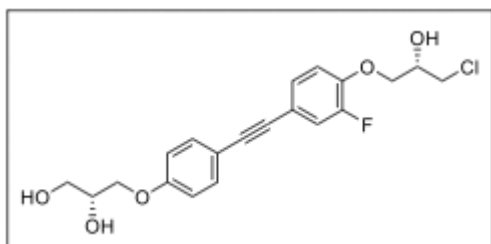
$^1\text{H NMR}$ (400 MHz, CD_3OD) δ 7.46 - 7.40 (m, 2H), 7.11 - 7.05 (m, 2H), 7.01 - 6.95 (m, 3H), 4.20 - 4.07 (m, 4H), 4.04 - 3.97 (m, 2H), 3.88 (s, 3H), 3.82 (dd, $J = 11.3, 4.8 \text{ Hz}$, 1H), 3.77 - 3.66 (m, 3H) ppm.

6. Experimental methods

^{13}C NMR (100 MHz, CD_3OD) δ 160.4, 150.7, 149.9, 133.8, 125.8, 118.2, 117.1, 116.1, 115.8, 115.0, 88.8, 88.8, 71.7, 71.4, 71.0, 70.4, 64.1, 56.6, 46.8 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{21}\text{H}_{24}\text{O}_6\text{Cl}$ 407.1256; found 407.1256.

(*R*)-3-(4-((4-((*S*)-3-chloro-2-hydroxypropoxy)-3-fluorophenyl)ethynyl)phenoxy)propane-1,2-diol



M_p = 116 - 118 °C.

$[\alpha]_D = -4.72$.

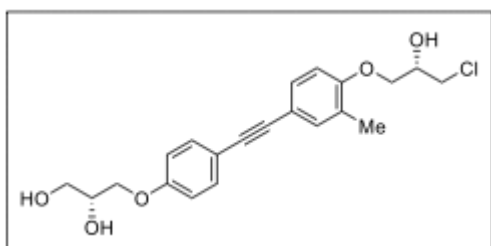
IR (ATR-FTIR): ν_{max} = 3376, 3243, 2917, 2848, 2509, 2410, 1519, 1321, 1296, 1270, 1244, 1014 cm^{-1} .

^1H NMR (400 MHz, CD_3OD) δ 7.47 - 7.42 (m, 2H), 7.29 - 7.21 (m, 2H), 7.12 (t, J = 8.7 Hz, 1H), 7.00 - 6.95 (m, 2H), 4.21 - 4.14 (m, 3H), 4.14 - 4.07 (m, 1H), 4.01 - 3.96 (m, 2H), 3.80 (dd, J = 11.2, 4.7 Hz, 1H), 3.74 - 3.65 (m, 3H) ppm.

^{13}C NMR (100 MHz, CD_3OD) δ 160.6, 154.6, 152.1, 148.4, 148.3, 133.9, 129.1, 129.1, 119.9, 119.7, 118.2, 118.1, 116.7, 116.3, 116.3, 115.8, 115.8, 89.7, 87.6, 87.5, 71.7, 71.4, 70.8, 70.4, 64.1, 46.6 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{20}\text{H}_{21}\text{O}_5\text{ClF}$ 395.1056; found 395.1057.

(*R*)-3-(4-((4-((*S*)-3-chloro-2-hydroxypropoxy)-3-methylphenyl)ethynyl)phenoxy)propane-1,2-diol



M_p = 89 - 90 °C.

$[\alpha]_D = -1.53$.

IR (ATR-FTIR): ν_{max} = 3274, 2921, 2853, 1606, 1510, 1456, 1238, 1109, 1037, 833, 813 cm^{-1} .

^1H NMR (400 MHz, CD_3OD) δ 7.44 - 7.39 (m, 2H), 7.32 - 7.29 (m, 1H), 7.28 (bs, 1H), 6.99 - 6.94 (m, 2H), 6.91 (d, J = 8.4 Hz, 1H), 4.19 (quint, J = 5.2 Hz, 1H), 4.13 - 4.07 (m, 3H), 4.05 - 3.96 (m, 2H), 3.83 (dd, J = 11.3, 4.9 Hz, 1H), 3.74 (dd, J = 11.3, 5.5 Hz, 1H), 3.69 (dd, J = 7.2, 5.1 Hz, 1H), 2.24 (s, 3H) ppm.

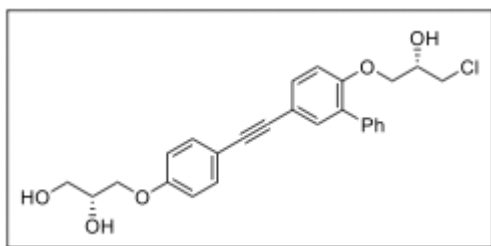
^{13}C NMR (100 MHz, CD_3OD) δ 188.3, 186.1, 162.7, 161.9, 159.5, 156.3, 145.4, 145.1, 143.9, 140.3, 117.1, 116.6, 99.8, 99.0, 98.5, 98.2, 92.2, 75.0, 44.4 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{21}\text{H}_{24}\text{O}_5\text{Cl}$ 391.1307; found 391.1306.

6. Experimental methods

-

(R)-3-(4-(((6-((S)-3-chloro-2-hydroxypropoxy)-[1,1'-biphenyl]-3-yl)ethynyl)phenoxy)propane-1,2-diol



Mp = 89 - 91 °C.

$[\alpha]_D = -0.77$.

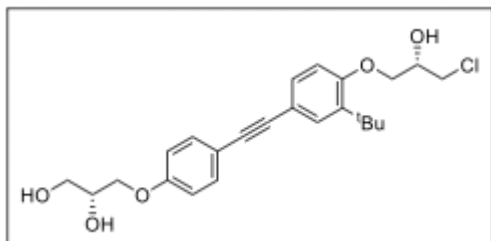
IR (ATR-FTIR): ν_{\max} = 3321, 2922, 2853, 1616, 1509, 1238, 1031, 1001, 768, 699 cm^{-1} .

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.56 - 7.52 (m, 2H), 7.49 - 7.39 (m, 7H), 7.07 (d, J = 8.5 Hz, 1H), 7.01 - 6.94 (m, 2H), 4.14 - 4.06 (m, 4H), 4.05 - 3.99 (m, 2H), 3.71 (dd, J = 7.3, 5.0 Hz, 2H), 3.64 (dd, J = 11.2, 4.4 Hz, 1H), 3.56 (dd, J = 11.4, 5.0 Hz, 1H) ppm.

$^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 160.1, 156.6, 138.9, 134.7, 133.8, 132.9, 132.6, 130.5, 128.9, 128.2, 117.7, 117.0, 115.6, 113.8, 89.1, 88.6, 71.6, 70.5, 70.3, 70.2, 64.0, 46.8 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{26}\text{H}_{26}\text{O}_5\text{Cl}$ 453.1463; found 453.1477.

(R)-3-(4-(((3-(tert-butyl)-4-((S)-3-chloro-2-hydroxypropoxy)phenyl)ethynyl)phenoxy)propane-1,2-diol



Mp = 116 - 118 °C.

$[\alpha]_D = -3.14$.

IR (ATR-FTIR): ν_{\max} = 3327, 2948, 2857, 1606, 1511, 1241, 1228, 1036, 811 cm^{-1} .

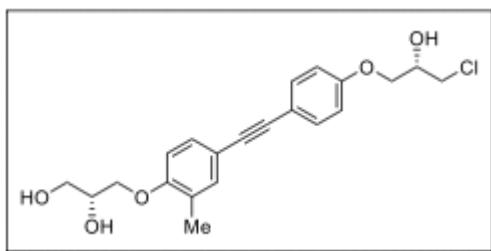
$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.48 - 7.39 (m, 3H), 7.34 (dd, J = 8.4, 2.1 Hz, 1H), 6.99 - 6.95 (m, 3H), 4.26 (quint, J = 5.2 Hz, 1H), 4.20 - 4.07 (m, 3H), 4.05 - 3.98 (m, 2H), 3.84 (dd, J = 11.3, 4.9 Hz, 1H), 3.77 (dd, J = 11.3, 5.5 Hz, 1H), 3.74 - 3.64 (m, 2H), 1.44 (s, 9H) ppm.

$^{13}\text{C NMR}$ (100 MHz, CDCl_3) δ 160.1, 158.6, 139.4, 133.7, 131.5, 130.9, 117.3, 116.9, 115.7, 113.3, 89.3, 88.4, 71.6, 70.9, 70.3, 70.1, 64.1, 47.2, 35.7, 30.3 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{24}\text{H}_{30}\text{O}_5\text{Cl}$ 433.1776; found 433.1778.

6. Experimental methods

(*R*)-3-(4-((4-((*S*)-3-chloro-2-hydroxypropoxy)phenyl)ethynyl)-2-methylphenoxy)propane-1,2-diol



Mp = 135 - 137 °C.

$[\alpha]_D = +0.54$.

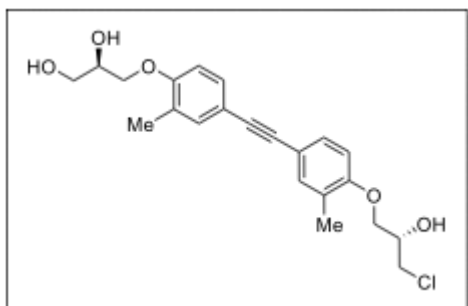
IR (Film): $\nu_{\max} = 3315, 2922, 2873, 2199, 1605, 1508, 1239, 1032, 808 \text{ cm}^{-1}$.

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 7.46 - 7.38 (m, 2H), 7.33 - 7.25 (m, 2H), 6.99 - 6.94 (m, 2H), 6.90 (d, $J = 8.4 \text{ Hz}$, 1H), 4.19 - 3.99 (m, 6H), 3.82 - 3.67 (m, 4H), 2.24 (s, 3H) ppm.

$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 159.9, 158.4, 134.5, 133.8, 131.4, 128.2, 117.6, 116.7, 115.8, 112.1, 89.1, 88.3, 71.8, 70.9, 70.3, 70.2, 64.2, 46.7, 16.2 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{21}\text{H}_{24}\text{O}_5\text{Cl}$ 391.1307; found: 391.1306.

(*R*)-3-(4-((4-((*S*)-3-chloro-2-hydroxypropoxy)-3-methylphenyl)ethynyl)-2-methylphenoxy)propane-1,2-diol



$^1\text{H NMR}$ (400 MHz, CD_3OD) δ 7.30 - 7.21 (m, 4H), 6.87 (d, $J = 8.3 \text{ Hz}$, 2H), 4.21 - 4.12 (m, 1H), 4.10 - 3.97 (m, 5H), 3.82 - 3.64 (m, 4H), 2.21 (s, 6H) ppm.

$^{13}\text{C NMR}$ (101 MHz, CD_3OD) δ 158.4, 158.0, 134.52, 134.5, 131.4, 128.2, 117.2, 116.9, 112.2, 112.1, 88.8, 88.6, 71.8, 71.0, 70.3, 70.2, 64.2, 46.9, 16.2, 16.2 ppm.

HRMS (ESI): m/z $[\text{M} + \text{H}]^+$ calculated for $\text{C}_{21}\text{H}_{28}\text{O}_5\text{Cl}$ 395.1620; found: 395.1617.

The final product was isolated as a slightly impure white solid.

Bibliography

- (1) 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.; McEwan, I. J.; Riera, A.; Salvatella, X. EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. *ACS Chem. Biol.* **2016**, *11*, 2499–2505.
- (2) De Mol, E.; Szulc, E.; Di Sanza, C.; Martínez-Cristóbal, P.; Bertoncini, C. W.; Fenwick, R. B.; Frigolé-Vivas, M.; Masín, M.; Hunter, I.; Buzón, V.; Brun-Heath, I.; García, J.; De Fabritiis, G.; Estébanez-Perpiñá, E.; McEwan, I. J.; Nebreda, Á. R.; Salvatella, X. Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators. *Structure* **2018**, *26*, 1–8.
- (3) Bouchard, J. J.; Otero, J. H.; Scott, D. C.; Szulc, E.; Martin, E. W.; Sabri, N.; Granata, D.; Marzahn, M. R.; Lindorff-Larsen, K.; Salvatella, X.; Schulman, B. A.; Mittag, T. Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. *Mol. Cell* **2018**, *72* (1), 19–36.
- (4) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaître, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 1–12.
- (5) Szulc, E. M. Structural Insights into “Acid Blobs and Negative Noodles” - the Androgen Receptor as a Case Study. Ph. D. Thesis, Universitat de Barcelona - IRB Barcelona, 2019.
- (6) Alberti, S.; Saha, S.; Woodruff, J. B.; Franzmann, T. M.; Wang, J.; Hyman, A. A. A User’s Guide for Phase Separation Assays with Purified Proteins. *J. Mol. Biol.* **2018**, *430* (23), 4806–4820.
- (7) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372.
- (8) Bartolomeo, M. P.; Maisano, F. Validation of a Reversed-Phase HPLC Method for Quantitative Amino Acid Analysis. *J. Biomol. Tech.* **2006**, *17* (2), 131–137.
- (9) Rapsomaniki, M. A.; Kotsantis, P.; Symeonidou, I.-E.; Giakoumakis, N.-N.; Taraviras, S.; Lygerou, Z. easyFRAP: An Interactive, Easy-to-Use Tool for Qualitative and Quantitative Analysis of FRAP Data. *Bioinformatics* **2012**, *28* (13), 1800–1801.
- (10) Göttfert, F.; Wurm, C. A.; Mueller, V.; Berning, S.; Cordes, V. C.; Honigmann, A.; Hell, S. W. Coaligned Dual-Channel STED Nanoscopy and Molecular Diffusion Analysis at 20 Nm Resolution. *Biophys. J.* **2013**, *105* (1), L01–L03.
- (11) Morales-Navarrete, H.; Segovia-Miranda, F.; Klukowski, P.; Meyer, K.; Nonaka, H.; Marsico, G.; Chernykh, M.; Kalaidzidis, A.; Zerial, M.; Kalaidzidis, Y. A Versatile Pipeline for the Multi-Scale Digital Reconstruction and Quantitative Analysis of 3D Tissue Architecture. *Elife* **2015**, *4*, 1–29.
- (12) Rink, J.; Ghigo, E.; Kalaidzidis, Y.; Zerial, M. Rab Conversion as a Mechanism of Progression from Early to Late Endosomes. *Cell* **2005**, *122* (5), 735–749.
- (13) Murray, D. H.; Jahnel, M.; Lauer, J.; Avellaneda, M. J.; Brouilly, N.; Cezanne, A.; Morales-Navarrete, H.; Perini, E. D.; Ferguson, C.; Lupas, A. N.; Kalaidzidis, Y.; Parton, R. G.; Grill, S. W.; Zerial, M. An Endosomal Tether Undergoes an Entropic Collapse to Bring Vesicles Together. *Nature* **2016**, *537* (7618), 107–111.
- (14) Senigagliaesi, B.; Penzo, C.; Severino, L. U.; Maraschini, R.; Petrosino, S.; Morales-Navarrete, H.; Pobega, E.; Ambrosetti, E.; Parisse, P.; Pegoraro, S.; Manfioletti, G.; Casalis, L.; Sgarra,

6. Experimental methods

–

- R. The High Mobility Group A1 (HMGA1) Chromatin Architectural Factor Modulates Nuclear Stiffness in Breast Cancer Cells. *Int. J. Mol. Sci.* **2019**, *20* (2733), 1–17.
- (15) 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.; Sadar, M. D. Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor. *Cancer Cell* **2010**, *17* (6), 535–546.
- (16) Imamura, Y.; Tien, A. H.; Pan, J.; Leung, J. K.; Banuelos, C. A.; Jian, K.; Wang, J.; Mawji, N. R.; Fernandez, J. G.; Lin, K.-S.; Andersen, R. J.; Sadar, M. D. An Imaging Agent to Detect Androgen Receptor and Its Active Splice Variants in Prostate Cancer. *JCI Insight* **2016**, *1* (11), 1–15.

DETAILED INDEX

1. Introduction and objectives	21
1.1. Intrinsically disordered proteins	23
1.2. Androgen receptor and castration-resistant prostate cancer	25
First objective: Study of the reversible interaction of EPI-001 with its target by solution NMR spectroscopy	28
1.3. Intrinsically disordered proteins and biomolecular condensation	29
Second objective: Study of the role of EPI-001 as a modulator of androgen receptor condensation	32
Third objective: Design, synthesis and evaluation of novel EPI-001 analogues	33
Bibliography	34
2. Reversible interaction of EPI-001 with its target by solution NMR spectroscopy	39
2.1. Background	41
2.1.1. Intrinsically disordered proteins as drug targets	41
2.1.2. Methods for studying the interaction between small molecules and intrinsically disordered proteins	49
2.1.3. Androgen receptor N-terminal domain	54
2.1.4. EPI-001, the first inhibitor of the androgen receptor N-terminal domain to enter clinical trials	55
2.2. Experimental results	60
2.2.1. Trifluoroethanol increases the helical population in Tau-5* and AF-1*	61
2.2.1.1. Trifluoroethanol increases the helical population	61
2.2.1.2. Tau-5* and AF-1* are prone to oligomerise	65
2.2.2. Interaction of EPI-001 with the AF-1* in the presence of TFE	71
2.2.3. Interaction of EPI-001 with the AR-NTD	76
2.3. Discussion	79
Bibliography	81
3. EPI-001 as a modulator of androgen receptor condensation	87
3.1. Background	89
3.1.1. Intrinsically disordered proteins are biopolymers which can undergo condensation	89
3.1.2. Biomolecular condensation in physiology and disease	91
3.1.2.1. Function of biomolecular condensates	91

3.1.2.2. Regulation of biomolecular condensates	95
3.1.2.3. Dysregulation of biomolecular condensates	96
3.1.3. Biomolecular condensates as an opportunity to drug IDPs	98
3.1.4. Biomolecular condensates and transcription	104
3.1.5. Androgen receptor undergoes biomolecular condensation	105
3.2. Experimental results	111
3.2.1. In vitro modulation	111
3.2.1.1. EPI-001 reversible binding promotes condensation of the AR-NTD	111
3.2.1.2. EPI-001 covalently reacts with the AR-NTD	113
3.2.1.3. Epilation effect on AR-NTD condensation	115
3.2.2. In cell modulation of AR-NTD condensation by EPI-001	117
3.2.3. Consequences of modulating AR-NTD condensation with EPI-001 to AR function	121
3.2.3.1. Condensation facilitates recruiting the transcriptional machinery	122
3.2.3.2. Condensation is a buffering system	124
3.3. Discussion	125
Bibliography	127
4. Rational design, synthesis and biological evaluation of novel EPI-001 analogues	135
4.1. Background	137
4.1.1. CRPC treatments currently in the pipeline	137
4.1.1.1. Degradation of AR	137
4.1.1.2. Targeting the N-terminal domain of AR (AR-NTD)	138
4.2. Experimental results	138
4.2.1. Design and synthesis of novel EPI-001 analogues	139
4.2.1.1. Optimisation of the synthetic route	139
4.2.1.2. CAM family of analogues	144
4.2.2. Biological evaluation of the CAM family of analogues	145
4.2.3. Consequences of modulating AR-NTD condensation with CAMs to AR function	148
4.2.3.1. Condensation facilitates recruiting the transcriptional machinery	149
4.2.3.2. Condensation is a buffering system	151
4.3. Discussion	152
Bibliography	154
5. General discussion and conclusions	157
6. Experimental methods	165
Detailed index	197

