Structural and Functional Relationships of Retinoic Acid Binding Proteins and Synthetic Retinoid Derivatives

CHARLES WILLIAM ERNEST TOMLINSON

Structural and Functional Relationships of Retinoic Acid Binding Proteins and Synthetic Retinoid Derivatives

Charles William Ernest Tomlinson

Submitted for the Degree of Doctor of Philosophy (PhD)

Department Of Chemistry

Durham University

2020

Retinoic acid receptors (RARs) offer an unparalleled drug target for many underserved neurodegenerative and oncological pathologies due to their wide ranging, and potentially highly tuneable, influence over cell differentiation. This influence is based on the interaction of RAR and retinoid-X receptor (RXR) proteins in a heterodimer, which becomes an active transcription factor once retinoid ligands are bound, enabling binding at the retinoic acid response element (RARE) DNA sequence. Signalling output is influenced by the cellular retinoic acid binding proteins CRABPI and CRABPII, which mediate non-genomic signalling and perform nuclear import of retinoic acid, respectively. Current retinoid based treatments extend to minor modifications of all-*trans* retinoic acid, and a small number of synthetic derivatives with largely unknown modes of action. Recently developed synthetic retinoid derivatives, and their solvatochromic fluorescent counterparts, are the basis for the development of a tool kit for retinoid binding investigation.

A solvatochromic fluorescence assay for interpreting ligand affinity to retinoic acid binding proteins has been developed and validated, based on the activity of key synthetic retinoid DC271. The fluorescence displacement assay is backed up with rigorous structural evidence demonstrating ligand binding to CRABPII and CRABPI-L29C. The binding of synthetic retinoids to both CRAB and RAR proteins is characterised through use of the assay, and the resulting data reveals key chemical functionalities responsible for changes to binding affinities. Further work has aided in the validation and structural characterisation of a novel protein-protein interaction between CRABPII and cyclin D3, including key retinoic acid binding responses.

X-ray crystal structures of CRABPI-L29C with bound synthetic retinoid DC645 (2.4 Å), and with fatty acids MYR and TDA (1.6 Å) are reported. Co-crystal structures of nuclear transport protein CRABPII were determined with synthetic retinoids DC360 (1.8 Å), DC271 (1.5 Å), EC19 (2.8 Å), DC479 (1.8 Å) and DC645 (1.7 Å) in a variety of crystal forms. CRABPII mutant R29AK30A structural determination (1.35 Å) offers key insight into the folding of the protein, and its subsequent binding behaviour with retinoic acids in solution.

These studies have significantly improved the understanding of interactions between synthetic retinoids and cognate binding proteins, and will inform the design of future target-specific candidate molecules.

Acknowledgements

This work has been made possible by the support of many people, who deserve the thanks laid out here.

Firstly, it would not have been possible to achieve any of this without Ehmke Pohl. His constant support, advice, supervisorial-wisdom and willingness to help have been the drive behind many parts of this project. My second supervisor, Andy Whiting, has also been constant in his guidance and desire to see the results of the project.

Thanks also go to High Force Research, and particularly Roy Valentine, as our industrial partner and sponsor, and to the Medical Research Council for funding the project.

Enormous thanks to Ian Edwards for his excellent technical support, and general ability to keep everything and everyone running. Inhabitants of CG231, and Labs 229 and 234 have all been a constant presence and have offered vital advice throughout my experience; in particular, lab members past and present Becky, Dan, Katy, Emma, Kate and Stef, need thanks for their specific advice and in-it-together camaraderie. Katy especially deserves my thanks for all of her assistance, support, prodding, and timely proof reading help.

Finally thanks to my family and friends at home, whose unswerving support has made staying on for four-more-years possible.

Contents

Abbreviations

ÄKTA	FPLC system (Cytiva)
ATP	Adenosine Triphosphate
ATRA	All-*Trans* Retinoic Acid
BCO	β-Carotene 15:15′ Monooxygenase
BL21(DE3)	*E. coli* lab strain (expression)
BLAST	Basic Local Alignment Search Tool
bp	Base Pair (DNA)
CAT	Chloramphenicol Acetyl Transferase
CBD	Cyclin Box Domain
CC 1/2	Correlation Coefficient ½ dataset
CCDC	Cambridge Crystallographic Data Centre
CCP4	Collaborative Computational Project 4
Coot	Crystallography Object-Oriented Toolkit
COS-1	Immortalised CV-1 monkey kidney cell line
hCRABPI	*Homo sapiens* Cellular Retinoic Acid Binding Protein I
hCRAPBII	*Homo sapiens* Cellular Retinoic Acid Binding Protein II
CRBPII	Cellular Retinol Binding Protein II
CV1	African Green Monkey Kidney (cell line)
CYP	Cytochrome P450 family Proteins
CYP26	Cytochrome P450 Protein 26
DBD	DNA Binding Domain
DC	Series of synthetic retinoids
DH5α	*E. coli* lab strain (plasmid storage)
DHA	Docosahexaenoic acid
DHQ	Dihydroquinoline
DIALS	Diffraction Integration for Advanced Light Sources (software)
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DR#	Direct Repeat (DNA element)
DTT	Dithiothreitol
EC50	Effective Concentration 50%
EDTA	Ethylenediaminetetraacetic acid
ERK1	Extracellular Signal Regulated Kinase 1 (signalling protein)
ESI	Electrospray Ionisation
Ex/Em	Excitation and Emission
FPLC	Fast Protein Liquid Chromatography
FRET	Forster Resonance Energy Transfer
GPCR	G-Protein Coupled Receptor
GST	Glutathione-S-Transferase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
His	Histidine
HP	High Purity (in reference to HisTrap®)
HPLC	High Performance Liquid Chromatography
HRV3C	Human Rhinovirus 3C protease
HT	High-throughput

HTRF	Homologous Time Resolved Fluorescence
IC50	Inhibitory Concentration 50%
IMAC	Immobilised Metal ion Affinity Chromatography
IPTG	Isopropyl ß-d-1-thiogalactopyranoside
IR#	Inverted Repeat (DNA element)
JA/JLA	J and J-lightweight; fixed angle rotors for Beckmann centrifuges
JCSG	Joint Consortium for Structural Genomics (crystallisation condition screen)
JEG-3	*Homo sapiens* choricocarcinoma (cell line)
LB	Lysogeny Broth
LBD	Ligand Binding Domain
LLG	Log-Likelihood gain
LRAT	Phosphatidylcholine-retinol O-Acetyl Transferase
MBP	Maltose Binding Protein
MS	Mass Spectrometry
MWCO	Molecular Weight Cut-off
MYR	Myristic acid
NBS	Non-binding surface, treatment for 96-well plates
NCBI	National Centre for Biotechnology Information
NCS	Non-crystallographic symmetry
NEB	New England Biolabs (supplier)
NICR	Northern Institute for Cancer Research, Newcastle, UK
NLS	Nuclear Localisation Sequence
NMR	Nuclear Magnetic Resonance
OD_{600}	Optical density 600 nm (measure of cell density)
OH	Hydroxide group
PA	Phytanic Acid
PACT	pH, Anion and Cation (crystallisation condition screen)
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank (pdbe.org)
PEG	Polyethylene Glycol
pET	Series of expression plasmids for use in *E. coli*
pGEX	Series of expression plasmids for use in *E. coli*
pOPINS3	Series of expression plasmids for use in *E. coli*
pJOE	Series of expression plasmids for use in *E. coli*
pUC	Series of plasmid cloning vectors for use in *E. coli*
QSAR	Quantitative Structure Activity Relationship
RA	Retinoid acid
RAMBA	Retinoic Acid Metabolism Blocking agent
RAR	Retinoic Acid Receptor (Protein)
RARE	Retinoid Acid Response Element (DNA element)
RMSD	Root Mean Squared Deviation
RNA	Ribonucleic acid
RXL	Arginine, X, Leucine motif for cyclin binding
RXR	Retinoid-X Receptor
RXRE	Retinoid-X Reponse Element
SDS	Sodium Dodecyl Sulphate (anionic surfactant)
SDS PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SEC	Size-exclusion chromatography
SF9	*Spodoptera frugiperda* 9 (cell line)
SG1	Shotgun 1 (crystallisation condition screen)
SHH	Sonic Hedgehog (protein)
SOC	Super Optimal broth with Catabolite repression
SPF	Sun Protection Factor
SUMO	Small, Ubiquitin like Modifier (protein)
TCEP	tris(2-carboxyethyl)phosphine, a reducing agent
TDA	Tridecanoic Acid
TEV	Tobacco Etch Virus protease
THQ	tetrahydroquinoline
TOF	Time-of-Flight (mass spectrometry)
TREp	Thyroid hormone Receptor pallindromic sequence (DNA element)
TTNN	Synthetic retinoid derivative
TTNPB	Synthetic retinoid derivative
UV	Ultraviolet (radiation)
WT	Wild type
XDS	X-ray Detector Software (crystallographic data processing software)
9/13-cis-RA	9/13-*cis* retinoic acid

Declaration and Statement of Copyright

The work in this thesis was carried out in the Department of Chemistry, Durham University, between October 2016 and December 2020. All work is the author's own, unless it is otherwise stated. This work has not previously been submitted for a degree at this, or any other institution.

Parts of this work have been published, described below:

Tomlinson, C. W. E., Chisholm, D. R., Valentine, R., Whiting, A. & Pohl, E. (2018), Novel Fluorescence Competition Assay for Retinoic Acid Binding Proteins. *ACS Medicinal Chemistry Letters* **9**, 1297-1300.

Chisholm, D. R., Tomlinson, C. W., Zhou, G.-L., Holden, C., Affleck, V., Lamb, R., Newling, K., Ashton, P., Valentine, R., Redfern, C., Erostyakgh, J., Makkaight, G., Ambler, C. A., Whiting, A. & Pohl, E. (2019), Fluorescent retinoic acid analogues as probes for biochemical and intracellular characterization of retinoid signalling pathways. *ACS chemical biology* **14**, 369-377.

Pohl, E. & Tomlinson, C. W. E. (2020). Retinoid Signalling Pathways. Chapter 7: Classical pathways of gene regulation by retinoids. *Methods in Enzymology*, pp. 151-173: Academic Press.

Tomlinson, C. W. E. & Whiting, A. (2020). Retinoid Signalling Pathways. Chapter 20: The development of methodologies for high-throughput retinoic acid binding assays in drug discovery and beyond. *Methods in Enzymology*, pp. 539-560: Academic Press.

Tomlinson, C.W.E., Cornish, K.A.S., Whiting, A. & Pohl, E. (2021). Structure–functional relationship of cellular retinoic acid-binding proteins I and II interacting with natural and synthetic ligands. *Acta Crystallographica* D77, 164-175.

1: Introduction

1.1. Retinoids

Retinoids are a class of lipid-soluble molecules related to, and derived from, vitamin A, which play a wide and active role within the chordate growth and development cycle. Vitamins, first termed 'vital-amines' before the discovery of non-amine based molecules, are essential metabolic compounds characterised by the inability of an organism to synthesise them; as such, they must be obtained through the diet. (Al Tanoury *et al.*, 2013) Malnutrition in vitamin A is closely tied to deficiencies in vision and growth, researched extensively as an outcome of world war rationing and economic depression. (Huxley *et al.*, 2000, Wolf, 2002) In *Homo sapiens*, sources of vitamin A, including animal retinyl esters and retinal as well as vegetable sourced pro-vitamin A (typically β-carotene), are absorbed by the gut epithelia and transported via the circulatory system to developing tissues. (D'Ambrosio *et al.*, 2011) Ultimately, all retinoid molecules were once derived from plant pro-vitamins, metabolised and exchanged through the food chain to higher organisms.

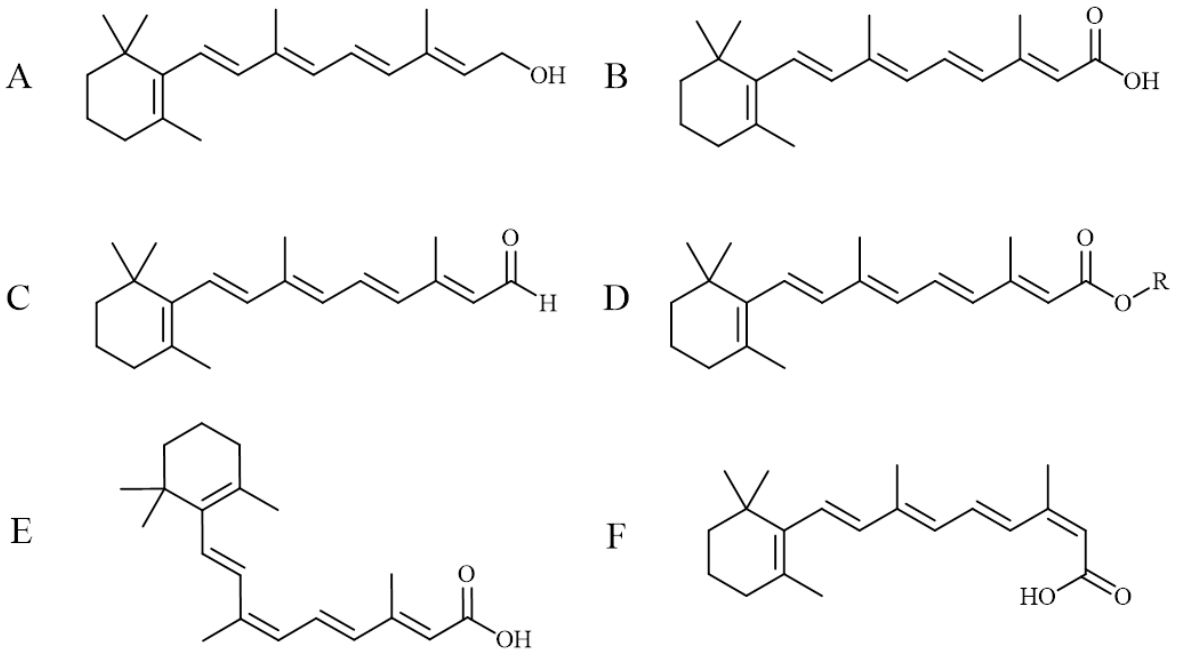


Figure 1.1: Chemical structures of naturally occurring vitamin A precursors and derivatives: A - retinol (vitamin A), B - all-*trans* retinoic acid, C - retinal (retinaldehyde), D - retinyl ester (generic), E - 9-*cis* retinoic acid, F - 13-*cis* retinoic acid. Structures prepared in ChemDraw (PerkinElmer Informatics).

Retinoids and natural derivatives are typified by a β-ionone ring and long polyene chain of nine carbon atoms (Figure 1.1). There are two broad pathways of retinoid activity: 11-*cis* retinal, isomerised from dietary all-*trans* retinal, is the visual pigment relied upon for light detection in the eye, and all-*trans* retinoic acid (ATRA - an oxidised derivative of retinal) is a potent controller of nuclear expression, exerting a powerful control over the development and maintenance of

differentiated tissues. Over five hundred retinoic acid responsive genes have been identified, exemplifying the wide-reaching influence of RA signalling. (Balmer & Blomhoff, 2002)

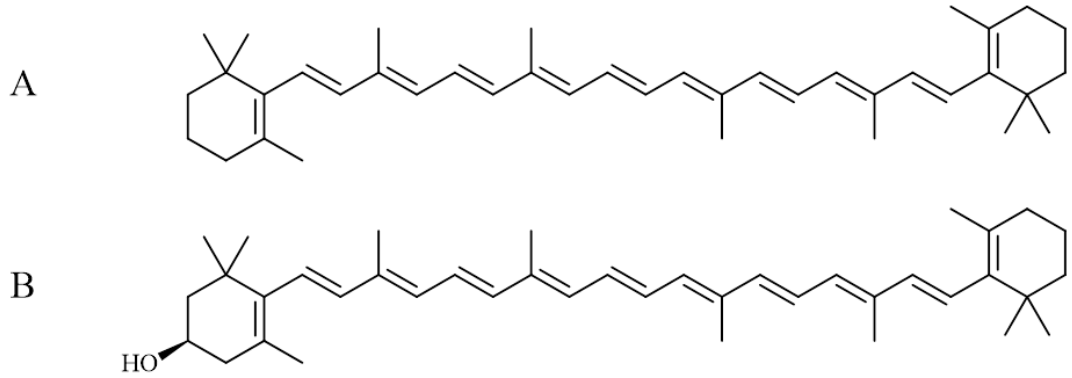


Figure 1.2: Chemical structures of naturally occurring pro-vitamin A: A - β-carotene, H - β-cryptoxanthin (asymmetric precursor). Structures prepared in ChemDraw® (PerkinElmer Infomatics).

Metabolic synthesis of all-*trans* retinoic acid (Figure 1.1 - B) is carried out by oxidation of retinal (retinaldehyde (Figure 1.1- C)) or retinol (Figure 1.1 - A), derived enzymatically from dietary pro-vitamins (β-carotene (Figure 1.2 - A) and β-cryptoxanthin (Figure 1.2 - B)) or directly from retinal (Figure 1.1 - C) or retinyl esters (Figure 1.1 - D) contained in animal products. (D'Ambrosio *et al.*, 2011) Synthesis of retinal from dietary pro-vitamin β-carotene is done by symmetrical, oxidative cleavage at the 15:15′ ethylene bond position. β-carotene 15:15′ monooxygenase (BCO) catalyses this process with great specificity to form two retinal molecules. This cleavage process is localised to the duodenum and liver, the primary vitamin A absorption and storage sites, respectively. (Devery & Milborrow, 1994, Lindqvist & Andersson, 2002, Lakshman, 2004) BCO is also catalytically active with β-cryptoxanthin, albeit with a lower activity, and demonstrates a general selectivity for pro-vitamin A molecules with at least one intact β-ionone ring. (Lindqvist & Andersson, 2002) β-cryptoxanthin is a pro-vitamin A found in many fruits and vegetables, with a bioavailability rivalling that of β-carotene. The greater hydrophilicity afforded by its polar hydroxyl group, and hence greater solubility and accessibility from micelles, increases its availability and processing versus β-carotene. (Burri *et al.*, 2016) A broad secondary path for retinal/retinol synthesis also exists, based on asymmetric cleavage of pro-vitamin A molecules to produce non-retinal products, which, after further processing, can be truncated to form retinal or retinoic acid. (von Lintig, 2012, Burri *et al.*, 2016)

Retinoid storage typically occurs as the incorporation of retinyl esters (Figure 1.1 - D) into lipid droplets localised in the hepatic stellate cells. (Blaner *et al.*, 2009) The esters are prepared by reduction of retinal to retinol by aldehyde reductase activities in the presence of cellular retinol binding protein (CRBPII), and further esterification by phosphatidylcholine-retinol O-acetyl transferase (LRAT). (Ong, 1993, Blaner *et al.*, 2009, D'Ambrosio *et al.*, 2011) The process is

reversed when dietary pro-vitamins are not available, and retinoid stores must be mobilised to maintain homeostasis.

One property shared by all natural retinoid derivatives is that of isomerisation. *Cis-trans* isomerisation can occur along the polyene chain to form 13-*cis*, 11-*cis* and 9-*cis* derivatives, as well as less common multi-*cis* isomers. (Bull *et al.*, 2019) Each isomer can have a distinct biological effect; in the case of retinal a very specific isomerisation is the key to the visual process. (von Lintig, 2012, Murayama *et al.*, 1997) Elsewhere, isomerisation can be burdensome and disrupt the activity of a species in solution. This can be particularly egregious under laboratory conditions, where determining and maintaining the concentration of a single retinoid isomer is a considerable task. (Murayama *et al.*, 1997)

1.1.1. Synthetic retinoid derivatives

The long, conjugated, polyene tail of ATRA is prone to photoisomerisation under natural light conditions and in solution. (Murayama *et al.*, 1997) The resulting mixture of ATRA, 9-*cis*-RA and 13-*trans*-RA is far less potent as a biological agent because the altered binding properties of each isomer reduce the overall activity of the mixture. Measurements dependant on the activity of a single species, especially quantisation of cell differentiation, can therefore be unreliable and not reflective of true signalling outputs. (Murayama *et al.*, 1997) To prevent this phenomenon, synthetic retinoids have been developed to mimic the shape and activity of ATRA - without its predilection to isomerisation. (Christie *et al.*, 2008, Dawson *et al.*, 1983, Strickland *et al.*, 1983) By maintaining stability in solution, and *in vivo*, these molecules can be used in place of ATRA in biological assays and in other situations where the light-based degradation of all-*trans*-RA becomes an issue.

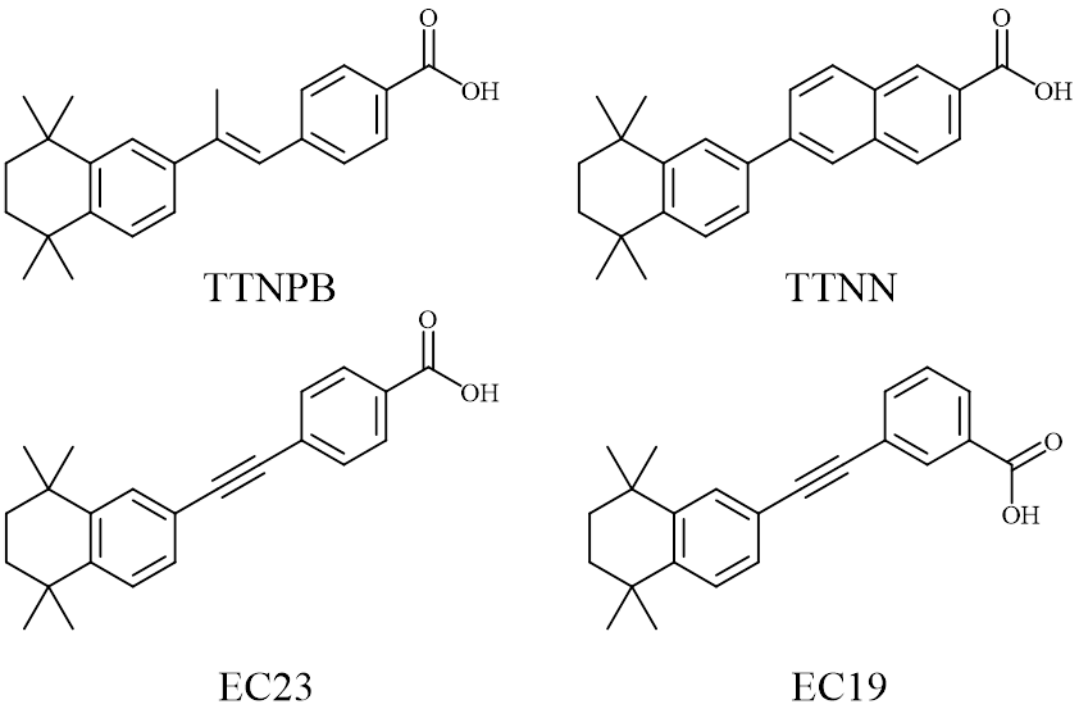


Figure 1.3: Chemical structures of synthetic retinoid derivatives TTNPB (Strickland *et al.*, 1983), TTNN (Dawson *et al.*, 1983), EC23, and EC19 (Christie *et al.*, 2008). Structures prepared in ChemDraw® (PerkinElmer Infomatics).

Replacement of the polyene chain using aromatic ring structures has generated synthetic derivatives of ATRA that vastly reduce the opportunity for photoisomerisation. TTNPB (Strickland *et al.*, 1983), TTNN (Dawson *et al.*, 1983) and EC23 (Christie *et al.*, 2008) are prominent examples of compounds designed to have the signalling ability of ATRA but with overall greater stability and usability (Figure 1.3). Each replaces the polyene chain of ATRA with varied acid-aromatic groups, and the β-ionone ring with a tetrahydronapthalene derived 'tail' section. (Strickland *et al.*, 1983, Dawson *et al.*, 1983, Christie *et al.*, 2008) Once the retinoid signalling ability of a molecule is quantified (see 1.4), synthetic derivatives can be updated and modified to increase their binding specificity to a particular protein. (Barnard *et al.*, 2009, Chisholm & Whiting, 2020) EC19 is a variation on EC23 designed to mimic the overall structure of 13-*cis*-RA by shifting the carboxylic acid group from the *para*- to *meta*- position, relative to the alkyne linker. EC23 induces neural differentiation of pluripotent stem cells in the same manner as ATRA, whilst EC19 generates epithelial-like cells in the same model. (Christie *et al.*, 2008) As with the native ATRA isomers, small differences in structure can have large implications for protein binding and downstream signalling. By utilising stabilised retinoid derivatives, which are less prone to isomerisation between active forms, targeted control of signalling may be achieved once specificity is established.

1.2. Retinoid signalling

1.2.1. Cellular retinoic acid binding proteins

The *Homo sapiens* cellular retinoic acid binding proteins, hCRABPI and hCRABPII, are intracellular lipid binding proteins associated with retinoic acid trafficking in the cytosol. (Budhu & Noy, 2002) The two share a 77 % sequence identity (86 % positive identity) and an orthogonal β-sheet fold capped with a pair of anti-parallel α-helices, formed from residues 15-35 part-way into the sequence (Figure 1.4 - A). (Kleywegt *et al*., 1994, Altschul *et al*., 1990, UniProt Consortium, 2017) This structure gives rise to a hydrophobic ligand binding pocket into which ATRA is incorporated (Figure 1.4 - B). (Kleywegt *et al.*, 1994) Wide differences in expression patterns suggest diverse roles for the proteins, with CRABPI being expressed in most adult cells, and CRABPII being limited to the skin and reproductive organs. (Blomhoff, 1994, Budhu & Noy, 2002) Upon binding to retinoic acid, general rigidity of the proteins (observed by X-ray crystallography and NMR) is increased by the fixing of unstructured loops into α-helix and β-sheet, though no large-scale structural rearrangement is observed. (Chen *et al*., 1998, Wang & Yan, 1999, Vaezeslami *et al*., 2006)

CRABPII is the key mediator of ATRA transport in mammalian cells and is demonstrably capable of nuclear localisation to deliver bound retinoid ligand to the nucleus. Passage through the nuclear pore is mediated by the formation of a 'non-canonical nuclear localisation sequence' (NLS) composed of three residues - K20, R29, and K30 - separated in the primary sequence but spatially related by the fold of the helical cap when the protein is ligand bound (Figure 1.4 - A). (Sessler & Noy, 2005) Entry into the nucleus allows CRABPII to exchange retinoic acid with the RAR/RXR nuclear receptors through a direct contact mechanism centred on residues Q74, P80 and K101. (Delva *et al.*, 1999, Budhu *et al*., 2001) The retinoic acid binding pocket contains, at its most buried end, a trio of residues (CRABPI - R111, R131, Y133; CRABPII - R111, R132, Y134) that are responsible for binding to ATRA via hydrogen bonding to the carboxylic acid moiety in concert with a conserved water molecule observed in all structures with similarly binding ligands. (Kleywegt *et al.*, 1994) The pocket is otherwise lined with hydrophobic residues creating a lipophilic environment for the unsaturated, fatty-acid-like portion of ATRA (Figure 1.4 - B). This hydrophobic contribution is considerable, as even when key interacting residues are mutated, general binding efficiency is maintained. (Vasileiou *et al.*, 2009) When the ligand binding site is unoccupied, the structure of the fold is maintained by a network of water molecules hydrogen bonded to backbone carbonyls. (Vaezeslami *et al.*, 2008)

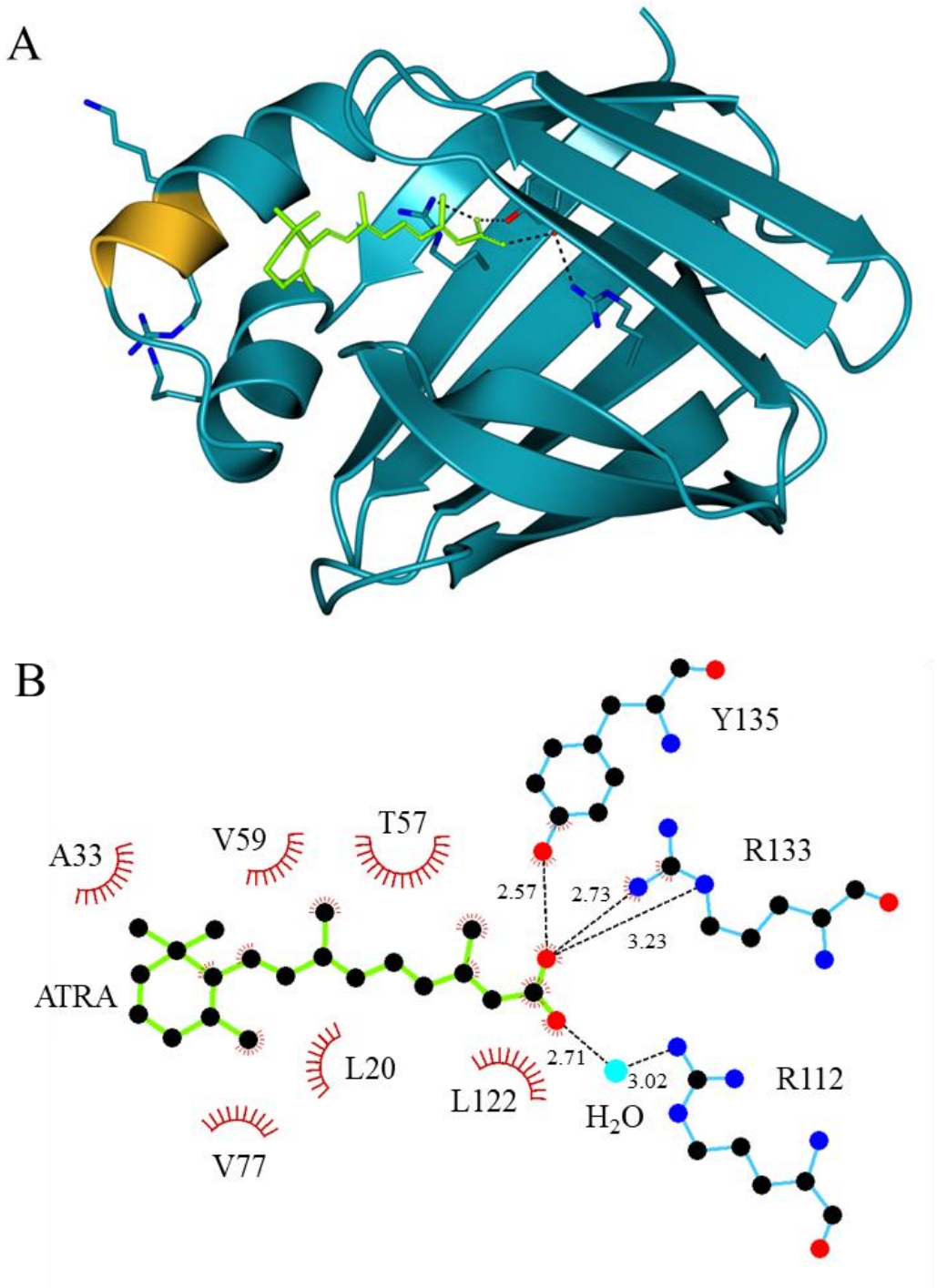


Figure 1.4: A - Structure of ATRA-bound CRABPII (1CBS) in ribbon representation. (Kleywegt *et al.*, 1994) Main chain (dark blue, ribbon), key binding residues (centre, dark blue, cylinder, O-blue, N-red) and NLS (left, dark blue, cylinder, O-blue, N-red). The portion of α-helix stabilised by ligand binding is highlighted (orange, ribbon). ATRA occupies the hydrophobic ligand binding site (green, cylinder) and can be seen interacting via hydrogen bonds (dashed, black) with the binding triad. B - Ligplot+ plane diagram (Laskowski & Swindells, 2011) of the hydrophobic binding site, highlighting key residues that interact with ATRA (green bonds, C-black, O-red). Binding triad residues are labelled (C-black, O-red, N-blue), hydrophobic residues are indicated (red semi-circles, hashed) and conserved water molecule is shown (cyan). Hydrogen bond (black, dashed) distances are labelled (Å).

Whilst the NLS-forming residues of CRABPII are conserved in CRABPI, the latter protein demonstrably does not enter the nucleus, likely due to other features of the CRABPII/nuclear pore interaction. (Budhu & Noy, 2002, Sessler & Noy, 2005) As a result, the protein is found localised entirely within the cytoplasm where it is generally believed to buffer retinoid concentrations, and aids in the degradation of retinoids v*ia* cytochrome P450 mechanisms (see 1.2.4). (Boylan & Gudas, 1992, Ross & Zolfaghari, 2011) There is also strong evidence that CRABPI modulates several putative non-genomic pathways. The ERK1/2 kinase pathway (responsible for cell cycle signalling in response to extracellular growth factors) is perceptibly activated by CRABPI-ATRA *via* sequential interactions with Raf, Mek and Erk kinases, in a dose dependant manner. (Park *et al.*, 2019, Persaud *et al.*, 2013) In mouse models, the presence of CRABPI reduces autophosphorylation of calmodulin-dependent protein kinase (CaMKII - broadly responsible for Ca^{2+} cellular responses and homeostasis) by competing directly with calmodulin, and prevents hypertrophy of cardiac tissues. (Park *et al.*, 2018)

1.2.2. Nuclear receptors RAR and RXR

The retinoic acid receptor proteins (RARs) form the basis of transcriptional control over retinoic acid signalling and were first identified through homology to steroid and thyroid nuclear receptors in cloned cDNA libraries. (Giguere *et al.*, 1987, Petkovich *et al.*, 1987) Upon binding to ATRA by direct transfer from CRABPII (Delva *et al.*, 1999, Budhu *et al.*, 2001), heterodimerisation with RXR proteins, and subsequent association with co-activator proteins, RAR:RXR dimers are capable of binding to DNA at retinoic acid response elements (RAREs) and altering the transcriptional output of a cell. (Bastien & Rochette-Egly, 2004)

The RAR proteins are split into three distinct isoforms (α, β and γ), generated by expression from three separate genes. (Egea *et al.*, 2000) Whilst RARα and RARβ are closely conserved, with ~80% positive sequence identity overall and 95% ligand binding domain similarity, RARγ deviates, both overall and specifically in the ligand binding domain, at 54% identity. (UniProt Consortium, 2017, Altschul *et al.*, 1990) Each isoform also exhibits multiple expression isoforms, generated by control under different promoters and by altered splicing, which creates variance in the N-terminal region and the 5′ untranslated region. (Kastner *et al.*, 1990, Leid *et al.*, 1992, Bastien & Rochette-Egly, 2004)

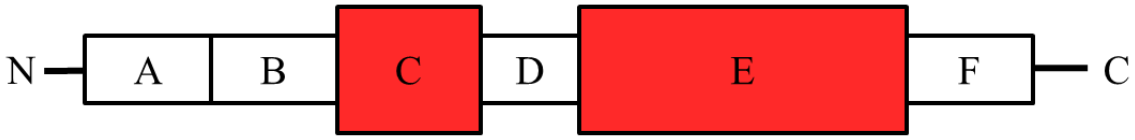


Figure 1.5: Schematic of conserved RAR protein domains: A/B - transcriptional activation, C - DNA binding domain (DBD), D - flexible linker region, E - Ligand binding domain (LBD), F - C terminal extension with uncharacterised function, missing in RXR proteins that share similar structure.

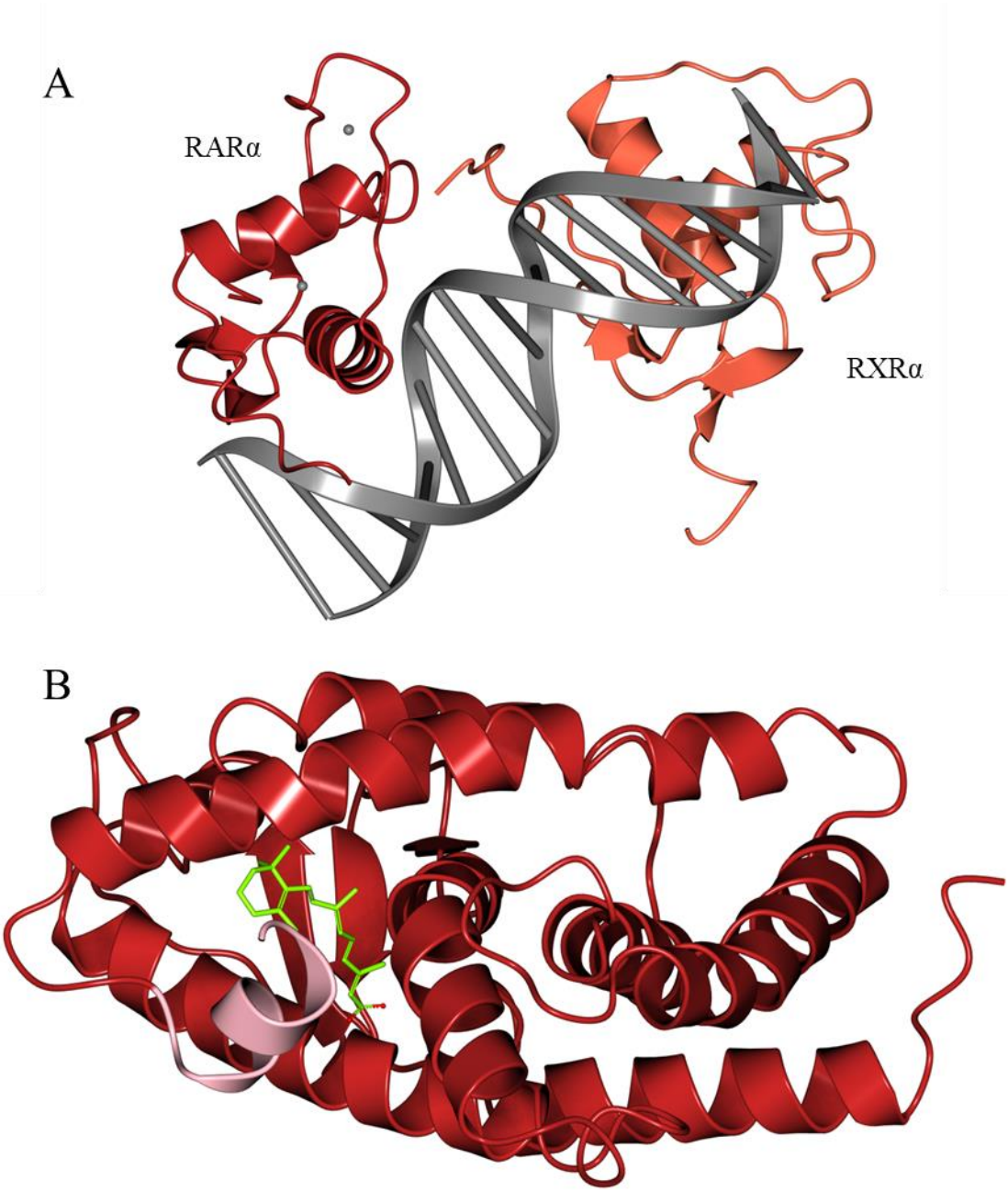


Figure 1.6: A - DNA binding domain zinc fingers of RARα (dark red, ribbon) and RXRα (pale orange, ribbon) bound to DNA retinoic acid response element (RARE - grey, ribbon) and incorporating structural zinc atoms (silver, spheres). 1.70 Å resolution - PDB code 1DSZ. (Rastinejad *et al.*, 2000) B - Ligand binding domain (LBD - residues 153-421) of RARα (dark red, ribbon) incorporating ATRA (green, cylinder) showing nuclear receptor α-helical sandwich architecture. Key helix H12 (pink, ribbon) can be seen folded into the ligand bound conformation. From a heterodimeric RARα/RXRα structure (2.75 Å) - PDB code 3A9E. (Sato *et al.*, 2010)

The RAR proteins contain six distinct domains (labelled A-F, Figure 1.5) and follow the general structure of the nuclear hormone receptor family. The C domain is highly conserved (94-97 %) and consists of a dual zinc finger DNA binding domain (Figure 1.6 - A), specific to the retinoic acid response elements (RAREs). (Chambon, 1996, Rastinejad *et al.*, 2000) This motif follows the canonical nuclear receptor fold, crossed α-helices with a β-linker region, and contains two structural zinc ions coordinated by highly conserved cysteine residues. A high resolution X-ray crystallographic structural determination of the domain, in concert with DNA, corroborates the fold previously identified by NMR studies (Knegtel *et al.*, 1993, Holmbeck *et al.*, 1998) and shows how the domain occupies the major groove of the DNA at the RARE site. (Rastinejad *et al.*, 2000) A flexible linker (D) follows the DNA binding domain, connecting the C domain to the ligand binding domain (E). Consisting of a three layer, globular, α-helical sandwich, the ligand binding domain is responsible for binding to ATRA (Figure 1.6 - B) and subsequent formation of a heterodimer with the retinoid-X receptor (RXR). (Leid *et al.*, 1992, Sato *et al.*, 2010) The binding of ATRA to RARα (Figure 1.6 - B) is centred on a hydrophobic pocket, in a similar manner to CRABPII, covered by the labile C-terminal helix H12 which is understood to control recruitment and expulsion of co-repressor and co-activator proteins, based on ligand occupancy. (Sato *et al.*, 2010, Cordeiro *et al.*, 2019) By movement between multiple ordered states, H12 can engineer co-repressor recruitment (antagonist bound or in the absence of ligand), or co-activator recruitment (presence of ATRA/agonist ligand), which triggers downstream signalling effects. (Nagy & Schwabe, 2004, Cordeiro *et al.*, 2019) With the understanding that CRABPII-RAR interaction is direct and relies on contact between the proteins, (Delva *et al.*, 1999, Budhu *et al.*, 2001, Budhu & Noy, 2002) it is likely that release of the ligand from CRABPII and binding to the RAR are concerted, and this triggers the re-ordering of H12.

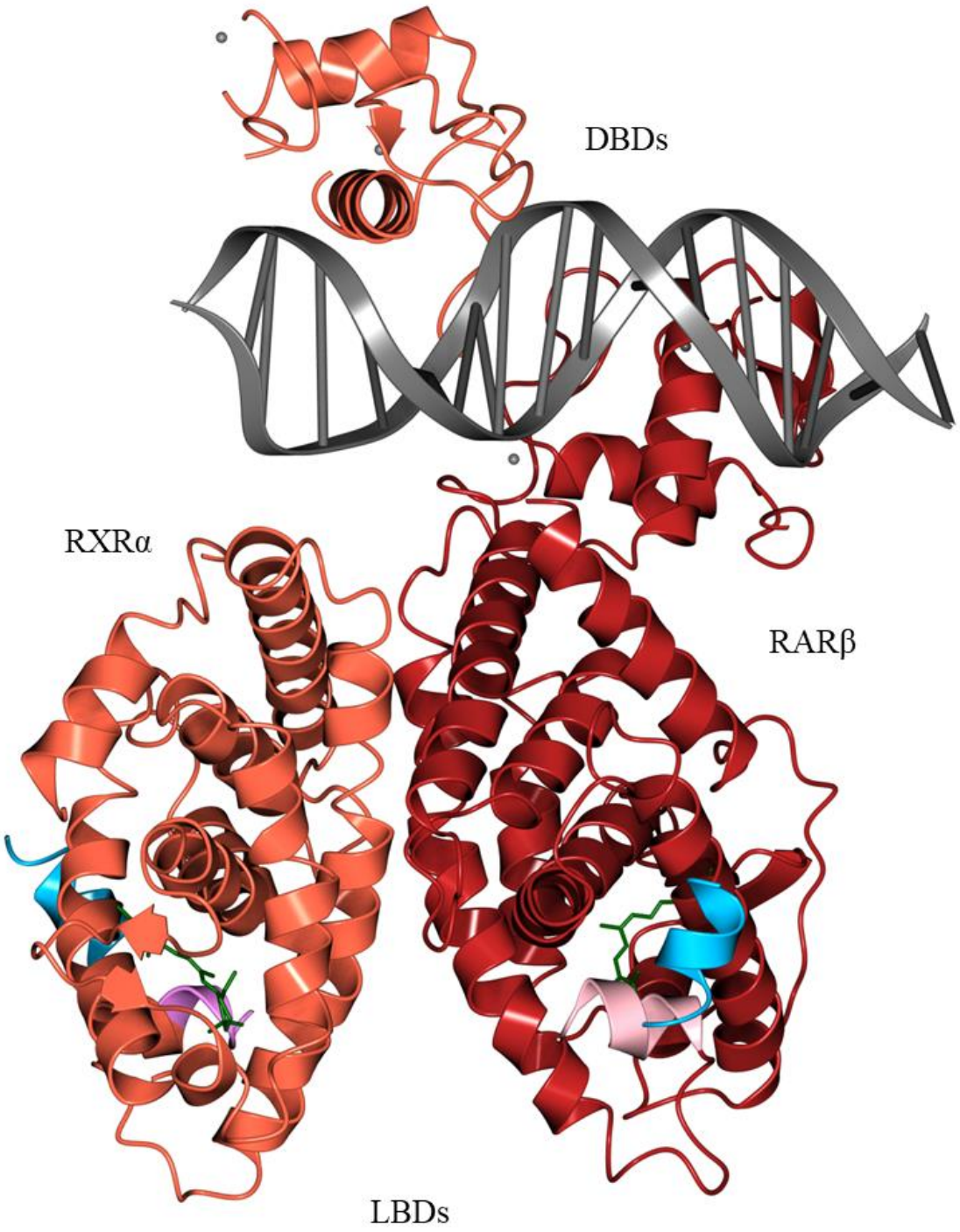


Figure 1.7: Complex of RARβ (dark red, ribbon) and RXRα (pale orange, ribbon) LBD and DBDs bound to DNA (grey, ribbon). Highlighted are H12 helices in active conformation (pinks, ribbon) and coactivator peptides (blues, ribbon) bound into the site they create. Structural zinc finger atoms (silver, spheres) and retinoid ligands (greens, cylinder) are shown. From PDB structure 5UAN, 3.5 Å. (Chandra *et al.*, 2017)

Given their size, and the flexibility between domains, it is unsurprising that there is no successful recorded crystallisation of an entire RAR protein, or a complex between RAR and CRABPII. The X-ray crystal structure of a RARβ/RXRα heterodimer at 3.5 Å has been successfully determined (Figure 1.7), containing electron density for RARβ (LBD with flexible linker to DBD) complexed with RXRα, the hinge region of which is disordered. (Chandra *et al.*, 2017) This linker disorder is consistent with the observation that RXR proteins are also capable of heterodimerisation with a variety of non-RAR nuclear receptors (Mangelsdorf & Evans, 1995, Chandra *et al.*, 2008, Rochel *et al.*, 2011, Lou *et al.*, 2014), the different sizes and modes of action of which require the increased flexibility seen in this structure. (Chandra *et al.*, 2008)

Both proteins contain their native ligands, RARβ ATRA and RXRα 9-*cis*-RA (Figure 1.7 - green), and adopt the active H12 fold (Figure 1.7 - pink). The structure contains a DR1 repeat RARE to which the DNA binding zinc fingers are complexed, as well as LXXLL-containing synthetic co-activator peptides (Figure 1.7 - blue). Crucially, the conformation of the linked ligand binding domains and DNA binding domains is consistent with those previously determined in individual states. This suggests that the derivative work on surface interactions and co-repressor/activator recruitment is valid. (Chandra *et al.*, 2017)

Paired with the RARs are the retinoid X-receptors, RXRs, which are structurally distinct nuclear receptors (Mangelsdorf *et al.*, 1990) with preference for 'rexinoid' ligands. (Heyman *et al.*, 1992) Possessing a comparable structure and domain layout to the RARs (51% sequence identity), though a markedly different ligand binding pocket, the RXRs are also found in α, β and γ isoforms and a host of expression-based isoforms. RXRs do not preferentially bind ATRA as a ligand and have been shown to accept 9-*cis* retinoic acid and fatty acids, including docosahexaenoic-acid (DHA), as ligands in a relatively L-shaped binding pocket. (Heyman *et al.*, 1992, Levin *et al.*, 1992, de Urquiza *et al.*, 2000, Egea *et al.*, 2002) In a similar manner to RAR, the dynamic RXR H12 helix governs recruitment of co-activator and repressor proteins, ultimately controlling signalling capability. (Zhang, Hu, *et al.*, 1999, Nahoum *et al.*, 2007, Cordeiro *et al.*, 2019) As is the case with RAR, it is likely that dynamic effects beyond those visible in crystallographic data are ultimately responsible for the behaviour of the nuclear receptors. (Lu *et al.*, 2006) With a minimum of nine heterodimer configurations, and many more when expression-based isoforms are considered, the RAR:RXR pairing affords a large variety of possible transcription factors, which then control a host of different transcriptional responses via binding to different DNA control elements. (Leid *et al.*, 1992)

1.2.3. Retinoic acid response elements

The retinoic acid response elements, RAREs, are DNA sequences located in the promoter of retinoic acid responsive genes, and are the targets for RAR:RXR heterodimer binding. (Al Tanoury *et al.*, 2013, Bastien & Rochette-Egly, 2004) The core motif of the RARE is a hexameric sequence, (A/G)G(G/T)TCA, which is usually found as a 5-base-pair spaced direct repeat (DR5). There are also examples of heterodimer binding at DR1 (1bp spaced) and DR2 (2bp spaced) sites. (Bastien & Rochette-Egly, 2004) In the case of DR5 and DR2 sites, the first motif (5′) is occupied by the RXR, and the second (3′) by the RAR. (Chambon, 1996) In DR1 sites the reverse is the case and there is evidence to suggest that RXR homodimers and other nuclear receptor dimers can also bind DR1 in a 'promiscuous fashion'. (Bastien & Rochette-Egly, 2004, Chambon, 1996) The RARβ gene is one key example of a DR5 RARE promoter; similarly, the gene responsible for CRABPII expression is controlled by a DR2 RARE. (del Mar Vivanco-Ruiz *et al.*, 1990, Durand *et al.*, 1992) There is additional evidence that 'non-canonical' RARE sequences DR0, DR8, and IR8 (inverted repeat) are also capable of RAR:RXR heterodimer binding, increasing the already diverse portfolio of retinoic acid responsive genes. (Moutier *et al.*, 2012)

The recruitment of retinoid-bound RAR:RXR heterodimers to RARE DNA sites triggers expression of proteins and subsequent phenotype change. Prior to retinoid binding, heterodimers are still capable of binding to RARE sites, though act repressively by recruiting histone deacetylase complexes, preventing the unravelling of chromatin which is required for gene expression. (Al Tanoury *et al.*, 2013) Once ligand interaction and binding at the RAR/RXR occurs, the repressive activity ceases and a coactivator protein (typically SRC-1, with typical consensus sequence LXXLL (Heery *et al.*, 1997, Al Tanoury *et al.*, 2013)) is recruited through the RAR, beginning the process of unravelling chromatin around the target gene. (Glass & Rosenfeld, 2000, Dilworth & Chambon, 2001, Al Tanoury *et al.*, 2013) After this, the normal machinery of mRNA transcription is recruited, and transcription/translation of gene products begins. (Dilworth & Chambon, 2001)

1.2.4. Downstream and non-genomic signalling

One of the most powerful displays of retinoid signalling effects was undertaken through the implantation of retinoid-releasing beads into the developing wings of chicken embryos. The resulting development of extra digits in a mirrored pattern implies a concentration-, and therefore distance-, mediated morphogenic effect that drives normal development of the tissue. By changing the concentration of additional RA, the morphogenic effect could be moderated, from minimal impact to full mirroring of the developing digits. (Tickle *et al.*, 1985) Similar effects can be observed by introducing cells from the zone of polarising activity, or purified sonic hedgehog (SHH) protein, extracellularly, solidifying a mechanism for the dose dependant morphogenesis seen with ATRA. (Tickle, 2006, Riddle *et al.*, 1993) There is additionally evidence that RA is a mediating factor in the expression of the fibroblast growth factor (*Fgf*8) and the *Hox* gene series, which specify tissue identities during embryonic development. (Mercader *et al.*, 2000, Koop *et al.*, 2010, Cunningham & Duester, 2015) Whilst expression of the majority of the *Hox* series are controlled by recruitment of an active RAR:RXR heterodimer to RARE sequences (Cunningham & Duester, 2015), a subset are inhibited by the binding of the heterodimer, contrary to the general mechanism. (Studer *et al.*, 1994, Kumar & Duester, 2014, Cunningham & Duester, 2015) This body of evidence creates a picture of RA as a powerful morphogen and mediator of cellular development; from embryogenesis and into adulthood, largely through the conveyance of positional data to developing cells, driving their differentiation and creating the patterning observed in adult tissues. The entire system is driven by the cellular feedback mechanisms underpinned by CRAB and CYP26 proteins that attenuate the lifetime and concentration of retinoids in the cell. (White *et al.*, 2007, Schilling *et al.*, 2016)

Outside of the traditional RAR-mediated signalling of ATRA there is evidence to suggest that signalling through other pathways, termed ‘non-genomic’, is also occurring in tissues. RAR phosphorylation and subsequent modification of signalling is driven by kinase cascades triggered by the presence of ATRA. (Masiá *et al.*, 2007, Rochette-Egly, 2015) More specifically, the extracellular signal-regulated kinase (ERK1/2) pathway is modified dose-dependently by CRABPI-ATRA in sequential interactions with Raf, Mek and Erk kinases. (Park *et al.*, 2019, Persaud *et al.*, 2013) In mouse models, autophosphorylation of the broad-spectrum calcium-signalling protein calmodulin-dependent protein kinase (CaMKII) is directly modified by the presence of CRABPI. This interaction competes directly with calmodulin binding and prevents hypertrophy of cardiac tissues. (Park *et al.*, 2018) These newly identified and atypical interactions of ATRA add to the already complex web of signalling interactions, nuclear receptor cross talk, and holistic tissue effects that constitute the retinoic acid signalling pathway. The retinoid signalling pathway, and related non-genomic effects, are summarised as a signalling schematic in Figure 1.8.

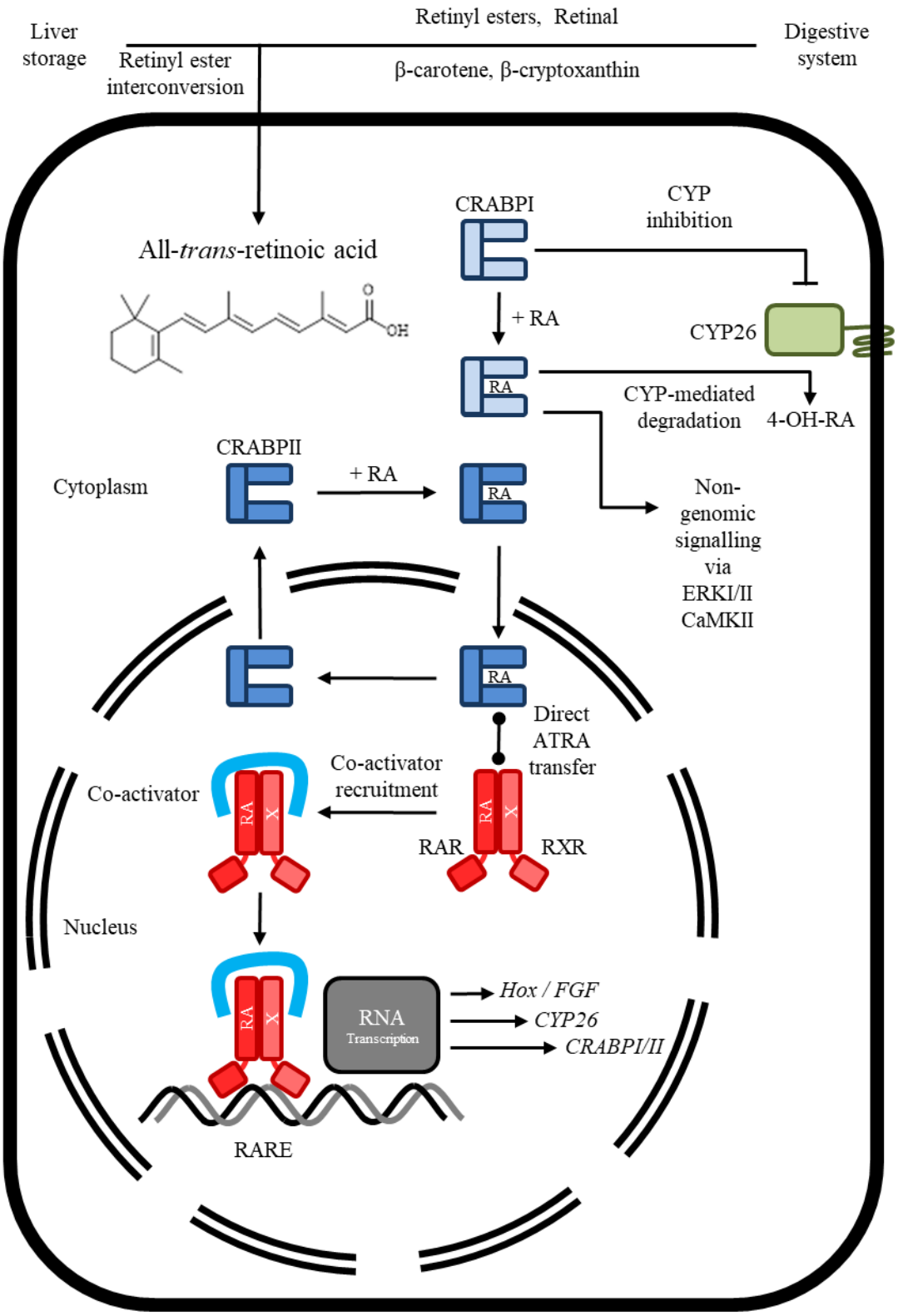


Figure 1.8: Simplified schematic of the retinoid signalling pathway, including CRABPI, CRABPII, RAR and RXR functions, degradation via CYP26 pathway and non-genomic signalling via CRABPI. RA = all-*trans* retinoic acid, X = 'rexinoid' RXR ligand.

The clearance of ATRA is broadly mediated by cytochrome P450 degradation mechanisms, and specifically the ATRA-induced CYP26 family. (Ross & Zolfaghari, 2011) Degradation proceeds by oxidation to 4-OH-RA catalysed by one of three family members, CYP26A1, CYP26B1 or CPY26C1, which are highly conserved between species (70-80% sequence identity) but individually diverse (40-50% sequence identity). (Thatcher & Isoherranen, 2009) These enzymes rely on typical P450 haem cofactors and are likely membrane-anchored in the endoplasmic reticulum of most cells in which they are expressed. (Stevison *et al.*, 2015) In the absence of only one of the three enzymes, mCYP26A1, mouse models showed extreme birth defects akin to the teratogenesis typically seen with an excess of ATRA. This suggests that the enzymes are not interoperable, and that ATRA depletion is as vital as uptake for proper embryonic development. (Abu-Abed *et al.*, 2001) The expression of degradation enzymes from CYP genes are seemingly RARE mediated under the control of typical DR5 promoter sequences. (Loudig *et al.*, 2000, Loudig *et al.*, 2005, Ross & Zolfaghari, 2011) In the case of CYP26A1, there are four RARE sites, identified in HepG2 liver cells and highly conserved in human and rat genomes, that all contribute to gene expression efficiency. (Zhang *et al.*, 2010) The B1 and C1 CYP enzymes are not directly appended to RARE sequences, but are highly proximal to the CYP26A1 promoter which may offer indirect enhancement of expression. (Ross & Zolfaghari, 2011) CRAB proteins modify metabolism of ATRA via the CYP26 family through direct protein-protein interaction, and the ratio of ligand-bound to unbound CRAB protein exerts positive control over the rate of degradation. (Nelson *et al.*, 2016) This points to a system in which low levels of ATRA (and hence unoccupied CRABPs) inhibit the degradation of the signalling molecule, whilst excess ligand (and therefore occupied protein) promotes removal. Given tissue specific expression levels of the CRAB proteins, the exact working of this feedback loop is likely to be highly site-dependant. (Boylan & Gudas, 1992, Nelson *et al.*, 2016, Ghaffari & Petzold, 2018) CYP26 family enzymes can be inhibited by RA metabolism blocking agents (RAMBAs), which enhance the overall retinoid signalling output by preventing ligand degradation. These treatments effectively increase intracellular retinoid concentration and can, as a result, enhance treatments based on ATRA dosing. (Gomaa *et al.*, 2012, Bilip *et al.*, 2020) These factors outline a tightly controlled homeostasis system for ATRA and other retinoids within cells, mediated by the presence of the ligand itself and the amplitude of the signalling system output in general.

1.3. Retinoid drugs

Retinoids and retinoid derivatives enjoy a wide range of drug, and drug-like, applications – specifically in the cancer and neurodegenerative disease fields because of their intrinsic link to cell development and differentiation.

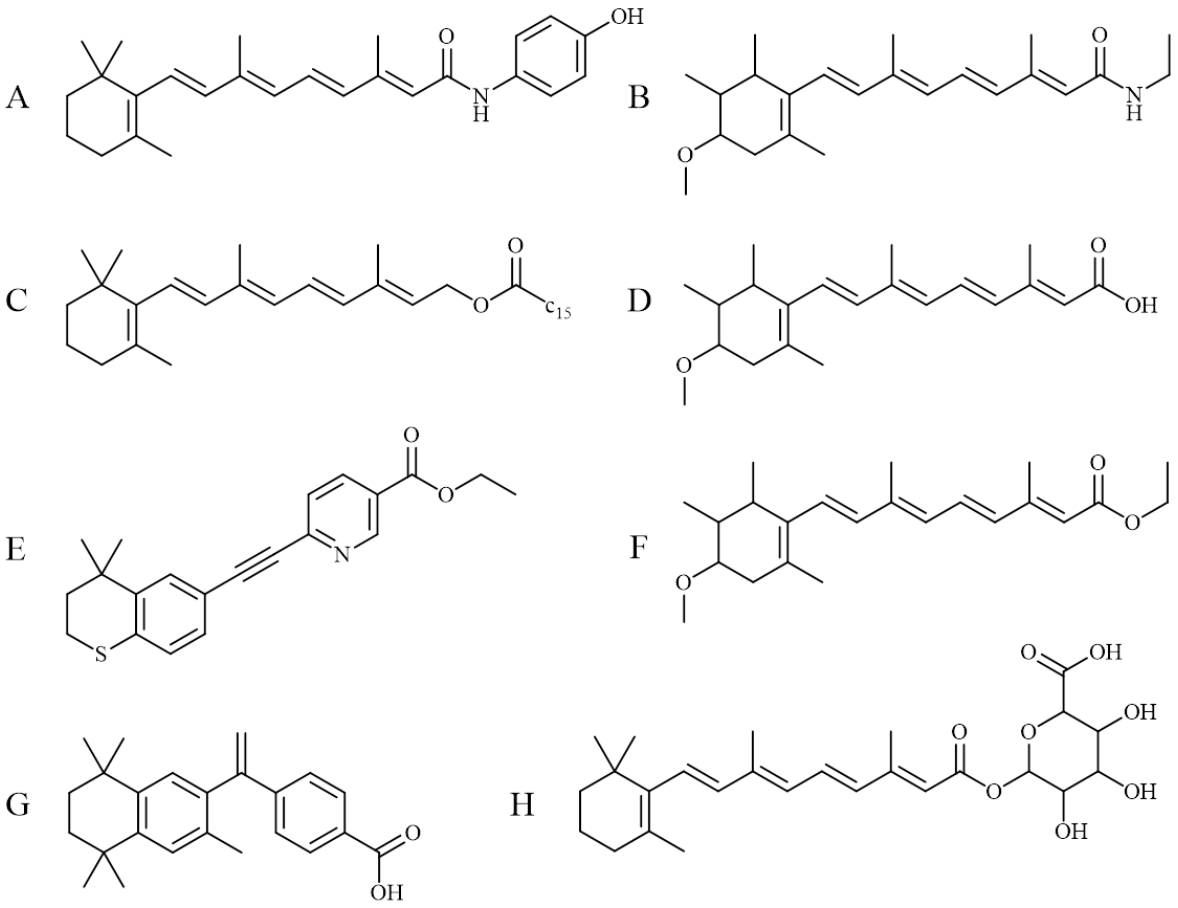


Figure 1.9: The diverse structures of currently approved retinoid and synthetic retinoid drugs. A - fenretinide, B - motretinide, C - retinyl palmitate, D - acitretin, E - tazarotene, F - etretinate, G - bexarotene, H - retinoyl β-glucuronide. Structures prepared in ChemDraw® (PerkinElmer Infomatics).

Oncological research has been a particular focus for retinoid-based therapies, with several synthetic retinoid-based drugs undergoing late phase clinical trials. Whilst retinoids act as morphogens in many tissues simultaneously, as with most chemotherapy agents, the weight of the change is focused on rapidly dividing tumorous growth, which outweighs the potential side effects. One such compound is fenretinide or 4-hydroxyphenylretinamide (Figure 1.9 - A), a hydroxyphenyl amide modification of ATRA, which has been proven to inhibit breast cancer development by causing tumour cell apoptosis in rat studies. (Moon *et al.*, 1979) The compound showed reduced activity compared to retinyl acetate which was used as a control, but far less toxicity in comparison over an extended treatment period, due to specific accumulation in mammary tissues. The specific mode of action of fenretinide is not fully understood, but the resulting apoptosis has been linked to enhanced expression of the caspase zymogen. (Simeone & Tari, 2004, DiPietrantonio *et al.*, 2000) More

recently, fenretinide has been deployed in phase I-II trials against non-Hodgkin lymphoma in concert with rituximab antibodies (MabThera produced by Roche), where it displayed increased remission lengths. (Cowan *et al.*, 2017) This deployment of a simple modification to ATRA shows how effective small changes in structure can be. Given this, there is still great scope for improving on both activity and specificity by use of synthetic retinoid derivatives. One such synthetic molecule licenced for medical use is bexarotene/Targretin (Figure 1.9 - G), which displays activity against T-cell lymphomas. (Lowe & Plosker, 2000) The molecule is based on a dihydronapthalene core with benzoic acid head group, and is RXR selective (making it a 'rexinoid'), likely due to topographic similarity to 9-*cis* retinoic acid.

Neurodegenerative diseases, particularly Parkinson's and Alzheimer's diseases, are heavily underserved by current treatment regimens, which are only able to reduce the observable symptoms and offer no reversal of the neurodegeneration suffered. As a result, compounds that may slow or reverse degeneration are a key focus for on-going research. (Cummings *et al.*, 2014) There is evidence to suggest that ATRA, as well as 9-*cis*- and 13-*cis*-RA, are capable of reducing levels of amyloid-β (A-β) built up in neurological tissues, albeit through an unknown mechanism. (Sahin *et al.*, 2005) This development may offer a route to treatment of Alzheimer's disease, and other disorders thought to be related to A-β, and a pathway that can be leveraged in understanding more about the disease. (Sahin *et al.*, 2005) Additionally, it is reported that individuals with high levels of retinoic acid from dietary vitamin A sources display decreased incidence of Alzheimer's disease. Despite this positive finding, the wide-ranging effects of retinoic acid make it a less than ideal drug candidate, and as such future treatments are likely to be based on targetable synthetic derivatives or co-treatment agents.

The high level of retinoid signalling activity in the skin makes the tissue a common target for retinoid-based treatments. Tretinoin (ATRA), isotretinoin (13-*cis*-RA), tazarotene (Figure 1.9 - E), motretinide (Figure 1.9 - B), retinoyl β-glucuronide (Figure 1.9 - H), and retinal all have an established clinical use as acne treatments. (Krautheim & Gollnick, 2004, Thielitz *et al.*, 2006, Layton, 2009) Acitretin (Figure 1.9 - D) is similarly used for treatment of psoriasis, a chronic proliferation of cells creating scaly skin, which, although effective in treating the ailment, is strongly contraindicated for pregnant women due to its teratogenic effects. (Katz *et al.*, 1999) Current retinoid derivatives in general all offer a potential teratogenic effect, and are commonly metabolised in a similar manner to ATRA, producing retinyl esters (acitretin to etretinate (Figure 1.9 - F)). Whilst the acitretin free acid is relatively short lived, and a good active treatment, the etretinate ester form of the drug is very stable and capable of lasting within the body for up to three years. (DiGiovanna *et al.*, 1989, Katz *et al.*, 1999) The precautions required for administering these compounds highlight that though retinoic acid derivatives can be powerful treatment agents, they can also carry strong secondary effects. Given their efficacy, the development of retinoid-based treatments that are more target specific, and hence display fewer side effects, is a priority.

There is an additional market for 'cosmeceutical' retinoids – compounds integrated into cosmetic products that maintain tentative bioactivity. There are currently several such retinoid-based products available on the market, all of which are primarily aimed at skin improvement. (Sorg *et al.*, 2006) Amongst claims of activity there is evidence that retinyl ester compounds applied topically can reduce UV-B absorption and prevent cellular apoptosis and photo-damage of DNA. (Sorg *et al.*, 2005) A similar study showed that the activity of 2% topically applied retinyl palmitate (Figure 1.9 - C) is analogous to that of commercially available sun protection factor (SPF) 20 UV-blockers. (Antille *et al.*, 2003) Given the natural propensity of retinoid derivatives to absorb UV, often isomerising in the process, it is likely that these topically applied molecules are doing the same, reducing UV exposure of DNA in the process.

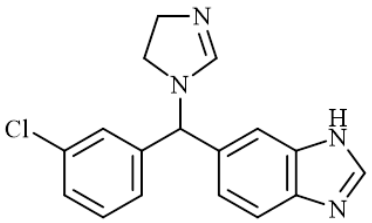


Figure 1.10: Chemical structure of Liarozole, the only retinoic acid blocking agent (RAMBA) currently in medical use. Structure prepared with ChemDraw® (PerkinElmer Infomatics).

One final class of retinoid-based treatments are the RAMBAs, or retinoic acid metabolism blocking agents. These compounds are not retinoids in the traditional sense, but co-treatment agents that prevent the breakdown of ATRA by inhibiting CYP26, increasing RA signalling output in tissues (1.2.4). (Gomaa *et al.*, 2012) Liarozole (Figure 1.10) is the only RAMBA currently in clinical use, deployed against prostate cancer, breast cancers, and in skin disorders including psoriasis and ichthyosis. (Njar *et al.*, 2006) 'Second generation' RAMBAs, currently under development, also show promising effects in prostate and breast cancer cell lines as well as neuroblastoma models. (Patel *et al.*, 2007, Huynh *et al.*, 2006, Bilip *et al.*, 2020) Whilst they are not considered retinoids, and have no signalling activity of their own, the potent inhibition of ATRA metabolism makes RAMBAs a worthy consideration in the development of new treatment regimes, as intrinsic signalling boosters or co-treatment agents to enhance true-retinoid drugs.

The efficacy of retinoid derivatives has been demonstrated in several medical fields, although few deviations from the core ATRA structure have been made. Most therapeutics use natural or isomeric forms of ATRA, as well as some displaying simple structural modifications, or are based on the small number of synthetic derivatives licenced for use. There is significant chemical space to be investigated that may offer enhanced specificity and mode of action, requiring structure-led development programs to best optimise the efficacy of new compounds.

1.4. Quantitative and qualitative assays for retinoid binding proteins

To harness and modify retinoid signalling for research and medicinal purposes, it is often necessary to determine the binding efficiency of compounds to specific proteins. Understanding binding, and being able to modify it based on that information, paves the way for improved research tools, comprehension of the signalling pathway, and pharmacological candidate molecules. Efficiency of screening, and subsequently the development of high-throughput methods, is a vital part of this process and many techniques have been developed to better understand protein-retinoid interactions. (Allenby *et al.*, 1993, Heyman *et al.*, 1992, Stafslien *et al.*, 2007, Tomlinson *et al.*, 2018, Yamada *et al.*, 2019)

The photo-isomerisation that typifies ATRA and related isomers, and their oxidation under ambient conditions, makes direct studies difficult. (Murayama *et al.*, 1997) Indirect measurements are therefore the preferred method by which efficacy is tested for both endogenous and novel retinoid compounds. Cellular reporter assays, radiological probes, FRET techniques, and fluorescence measurements have all been utilised to improve understanding of these interactions.

1.4.1. Cellular expression systems

The earliest methods used to characterise protein-retinoid interactions were based around cellular reporter assays. Suitable cell strains were transfected with an expression system for the protein of interest (initially RAR/RXR isoforms) as well as a RARE/RXRE promoter sequence, coupled to a measurable gene product (often luciferase or chloramphenicol acetyl transferase). The RXR ligand properties of 9-*cis*-RA were determined in insect cell culture of *Drosophillia melanogaster* (Schneider S2 cells) transfected with plasmids for expression of the RXR, and a luciferase gene under the control of an RXRE sequence. The binding of the retinoid compound to the receptor was measured by proxy as luminescence, triggered by binding of an active heterodimer to the RXRE and resulting expression of the luciferase gene product. (Heyman *et al.*, 1992) The high quantum yield of luciferase, processing luciferin and ATP into oxyluciferin, makes it an ideal reporter for *in cellulo* assays. (Gould & Subramani, 1988, Totter, 1981, Seliger & McElroy, 1960) 9-*cis*-RA was purified from the cell lysate of activated cells by HPLC, and was the first evidence that the isomeric version of ATRA was the natural ligand of the RXR nuclear receptor. Other retinoids co-eluted and bound to protein were tested separately, determining that ATRA was undergoing isomerisation in culture. (Heyman *et al.*, 1992)

There are advantages to cell-based assays, mainly the determination of downstream signalling efficacy; however, the possibility of *in vivo* isomerisation, unexpected signalling crosstalk in heterologous cell culture, and presence of endogenous ligands in mammal cell culture is not easily quantified. Preparing, transforming with expression systems, and maintaining insect or mammal cell cultures can additionally be time consuming and costly, especially in comparison to *in vitro*

methods. (Assenberg *et al.*, 2013, van Oers *et al.*, 2015) Such experiments are not trivial to scale and do not always lend themselves to a high-throughput format.

A CV1 (primate cell culture) expression system was used to first demonstrate the ligand binding of ATRA to the RAR proteins. (Giguere *et al.*, 1987, Jensen *et al.*, 1964) Testing of endogenous retinoic acids and derivatives was carried out in a system expressing chloramphenicol acetyl transferase (CAT) under control of a RARE sequence. (Giguere *et al.*, 1987) Mammalian cells lack endogenous CAT activity so reporting is tightly controlled; ^{14}C labelled chloramphenicol is used as a substrate, and purification and identification of the subsequent radio-products allows the signalling to be quantified. (Gorman *et al.*, 1982) CAT activity was also used in the initial characterisation of the RXR receptors, demonstrating that these proteins had greater ligand specificity than RARα when exposed to a series of retinoids. A thyroid hormone receptor (TREp - AGGTCA) palindromic sequence was used in lieu of the as of yet uncharacterised RARE/RXRE sequence. Later, experiments into specific retinoic acid derivatives were carried out in CV1 cells using an analogous approach. (Mangelsdorf *et al.*, 1990)

Further RXR activating factors, including docosahexaenoic acid (DHA), were identified using a luciferase reporter construct in human choriocarcinoma JEG-3 cells. (de Urquiza *et al*., 2000, Gould & Subramani, 1988) Treatment with extract from embryonic and adult mouse brain tissues demonstrated activation of RXR in the presence of DHA. The reporter system, using RXRα as a binding target, was used to better characterise the protein-ligand interaction of DHA and other fatty acids with RXR, demonstrating dose response in the 50-100 μM range. *(*de Urquiza *et al*., 2000*)* Palmitoleic (53 μM), oleic (82 μM) and myristic acids (>150 μM) also show RXR ligand affinity. (Goldstein *et al.*, 2003)

1.4.2. Radiolabelling techniques

Radiolabelling uses tritiated derivatives of ATRA (all-*trans*-[^{3}H]-RA) or 9-*cis*-retinoic acid (9-*cis*-[^{3}H]-RA) to determine the binding activity of proteins both *in vivo* and *in vitro*. Measurement of the ratio of labelled ligand up-taken *vs* that remaining in solution allows for direct measurement of binding, as well as displacement effect caused by the binding of other (non-labelled) ligands.

Combining tritiated-derivative retinoids with cell cultures heterologously expressing RAR or RXR proteins allowed for the determination of binding for common endogenous retinoids including ATRA, and 7-, 9-, 11- and 13-*cis*-RA. (Gluzman, 1981, Allenby *et al.*, 1993) The radio-labelled probe introduced to the heterologous expression system undergoes mass uptake and binding, creating a pool of radio-ligand associated protein. Fixed concentration competitor molecules are added to the prepared cells and size exclusion chromatography of the resulting lysate is carried out. The fraction of radio-probe still bound to protein *vs* free in solution is calculated by radiography, indicating the rate of uptake of the competitor. (Allenby *et al.*, 1993) Displacement ratios at

different competitor-ligand concentrations can then be calculated and the affinity determined. (Levin *et al.*, 1992, Allenby *et al.*, 1993)

Determination that RAR isoforms have different affinities for 9-*cis*-RA and ATRA was also carried out using radio-displacement assay. (Allenby *et al.*, 1994) Expression in COS-1 cells, separation by HPLC, and scintillation counting were all carried out after introduction of the radio-ligands. RARα, β and γ all shared similar ATRA binding affinities (IC_{50} ~ 5 nM), whilst 9-*cis*-RA was displaced six-fold quicker from RARγ than RARβ. (Allenby *et al.*, 1994) All-*trans*-retinol and all-*trans*-retinaldehyde RAR binding studies also used a radio-ligand approach. (Repa *et al.*, 1993) Further experimentation demonstrated that whilst all-*trans*-retinol had binding affinity for RAR proteins, only 4-oxoretinol (the metabolic breakdown product of ATRA) (Leo *et al.*, 1989, Leo & Lieber, 1985) was capable of inducing transcription. (Achkar *et al.*, 1996)

There are well established syntheses for the tritiated retinoid compounds, both directly and using isomerisation and purification for enrichment. (Cahnmann, 1995) Regioselective techniques allow tritiation at the 2,3 or 3,4 positions in the cyclohexene ring with high radiochemical yield. (Bennani & Boehm, 1995) Anhydrous benzene and Wilkinson's catalyst (Meakin *et al.*, 1972) allow for hydrogenation using hydrogen, deuterium or tritium as appropriate. (Bennani & Boehm, 1995) The compounds still undergo photoisomerisation and oxidation under laboratory conditions, making appropriate storage and prompt use paramount. The primary downside of radiolabelling techniques is the specialist facilities, expertise, and overall cost of synthesising, storing, and utilising the tritiated retinoid derivatives required. (Boehm *et al.*, 1994)

1.4.3. Förster Resonance Energy Transfer (FRET) techniques

Förster Resonance Energy Transfer (FRET) is a distance-dependant energy transfer between fluorescent donor and acceptor molecules, which can be harnessed to "measure" intermolecular distances and investigate interactions between molecules. (Förster, 2012, 1948) Typically, the fluorescence of a donor and acceptor pair, absorbing energy from an external source and from the subsequent emission of the donor, respectively, is measured. The transfer of energy from donor to acceptor creates a drop in the emission spectrum of the donor molecule, and a corresponding rise in that of the acceptor. This interaction can be carried out at distances of 10-100 Å and allows for measurement of contacts between biomolecules both *in vitro* and *in vivo*. (Stryer, 1978, Tsien, 1998) Homogenous Time Resolved Fluorescence (HTRF) assays are most widely used, combining FRET measurements with time resolution in a small volume assay to measure interactions for nuclear receptors (including RARs), GPCR signalling proteins, kinases, and other protein-protein, protein-peptide, and protein-DNA interactions. (Degorce *et al.*, 2009)

The FRET/HTRF method has been applied to both the RARs and RXRs in determining the binding of endogenous retinoids and derivatives. RXRβ expressed heterologously (in SF9 insect cell

culture) was investigated using a terbium/fluorescein donor/acceptor pair and both 9-*cis-RA* and ATRA, as well as fatty acids phytanic acid (PA) and docosahexaenoic acid (DHA). (Stafslien *et al.*, 2007) *In vitro* recruitment of coactivator peptides labelled with fluorescein (with LXXLL motif) to the RXR LBD was measured, using a GST-tag and terbium-antibody pairing to provide the donor. The labelled coactivator is recruited to the RXR upon ligand binding and rearrangement of the H12 helix, and generates a FRET signal between the terbium (340 nm) and fluorescein (520 nm) pair. (Stafslien *et al.*, 2007) 9-*cis*-RA, the natural ligand of RXRα (Heyman *et al.*, 1992), caused the greatest measurable FRET response. Phytanic acid (a dietary fatty acid (Verhoeven *et al.*, 1998)) was found to be similarly active, with an affinity greater than that of ATRA; DHA followed up with the poorest affinity of the tested binding compounds. (Stafslien *et al.*, 2007)

The Tb-fluorescein pair was also used in determining the binding of RAs and EC23 derivatives to RARα/β/γ in the form of a commercial experimental kit (Invitrogen Lanthascreen®). (Haffez *et al.*, 2017) Other Lanthascreen test series demonstrate that a retinyl ester synthetically derived from ATRA preferentially binds to RXRα over RARα and is capable of inhibiting the growth of gastric cancer cells. (Hu *et al.*, 2014)

FRET assays can be costly but are advantageous in measuring interactions between multiple proteins, DNA sections, or both. The dose dependencies (EC_{50} values) calculated from FRET are non-equivalent to the actual receptor binding affinities, as the measurement takes into account several equilibrium events. Ligand recruitment, receptor conformational changes, and enlistment of the coactivator peptide must all occur before the FRET signal is detectable, reducing the direct relationship between ligand binding and measurable signal. (Stafslien *et al.*, 2007) FRET/HTRF is therefore a strong qualitative asset but may lack the quantitative element required for greater understanding of binding and ligand-receptor relationships, which can be better achieved using other techniques.

1.5. Development of fluorescent retinoids

Naturally occurring retinoic acid derivatives are inherently fluorescent due to their conjugated structure, though this signal is weak and requires painstaking monitoring to differentiate above the background. (Vivas *et al.*, 2015) Additionally, photo-isomerisation at UV wavelengths (similar to those of excitation) reduces their biological efficacy *in vivo*. (Murayama *et al.*, 1997) Similarly, EC23 and EC19 synthetic retinoids are somewhat fluorescent, at excitation and emission wavelengths 300 nm and 380 nm, respectively; this makes them likely to cause cellular damage when used as *in vivo* probes. The development of improved fluorescent retinoid derivatives, capable of excitation at visible wavelength and with good quantum yield, was carried out in order to present a set of compounds suitable for use as molecular probes. (Chisholm *et al.*, 2016, Chisholm *et al.*, 2019, Chisholm & Whiting, 2020)

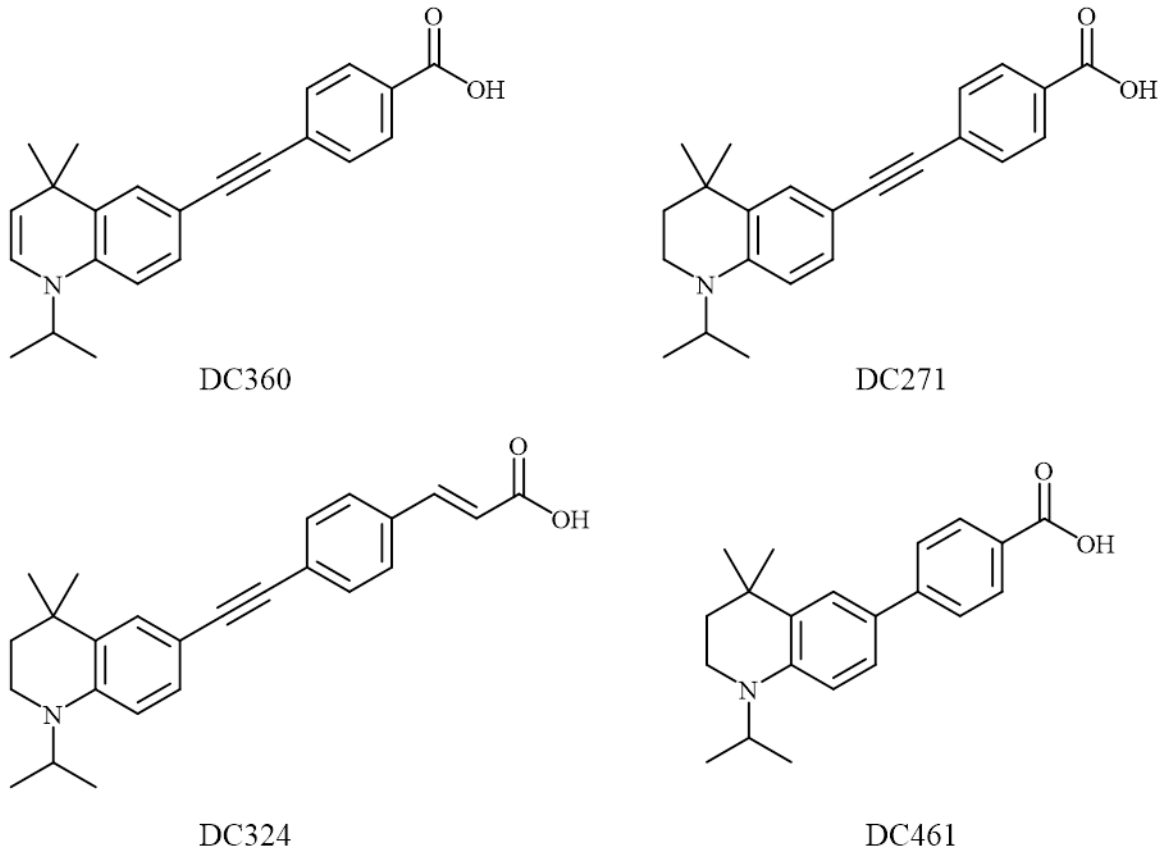


Figure 1.11: Chemical structures of core DC series compounds, based on extended di-tetrahydroquinoline structure. Images prepared in ChemDraw® (PerkinElmer Infomatics)

The result of this work was a library of fluorescent compounds with varying structures (Figure 1.11), based on the di/tetrahydroquinoline/alkyne/benzoic acid framework, that exhibit charge transfer fluorescence at wavelengths suitable for work *in vivo (*typically absorbing ~370 nm). (Chisholm *et al.*, 2016, Chisholm *et al.*, 2019, Chisholm & Whiting, 2020) The increase in the π-electron donating nature of the di/tetrahydroquinoline group (*vs* that of naphthalene-based EC23/19) and the π-accepting nature of the acidic groups creates a drastic bathochromic shift in

excitation wavelength. Some compounds are direct EC23 mimics, whereas others deviate to create different conformations in solution. For example, DC324 (Figure 1.11 - bottom left) contains an extended, conjugated section separating the carboxylic acid moiety from the aromatic ring to increase the overall length of the molecule. DC461 (Figure 1.11 - bottom right) forgoes an alkyne linker in favour of directly bridging the two aromatic rings, creating an overall shorter structure. (Chisholm *et al.*, 2016, Chisholm *et al.*, 2019, Chisholm & Whiting, 2020)

Evidence suggests that DC271 (Figure 1.11 - top right) undergoes nuclear localisation in HaCaT keratinocytes. (Chisholm *et al.*, 2019) Imaging using fluorescence microscopy shows a clear distinction between the bulk of the retinoid in the nucleus, and minimal concentrations in the cytoplasm and mitochondria. RNA sequencing confirms DC271 activates transcription of genes in a similar pattern to that of ATRA or EC23. This evidence, combined with X-ray crystallographic structural determination (4.2.2), suggests that DC271 is capable of correct handling by CRABPII (resulting in nuclear localisation) and therefore potentially a natural mode of action in cells. (Chisholm *et al.*, 2019)

The hydrophobic structure of the DC synthetic retinoids creates a solvatochromic effect. Each compound exhibits strongly increased fluorescence emission under hydrophobic conditions (for example, when solvated in hexane). It is hypothesised that under aqueous conditions the retinoids associate with each other in a hydrophobic manner, creating a fluorescence quenching effect. (Chisholm *et al.*, 2019) Once incorporated into a hydrophobic location, such as the binding site of CRAB or RAR proteins, fluorescence emission is increased by an order of magnitude. This characteristic should make the fluorescent retinoid derivatives an excellent tool for probing retinoid acid signalling.

1.6. Project Aims

The retinoid signalling pathway, due to wide ranging phenotype influence over cell differentiation and fine-tuned control mechanisms, offers an excellent opportunity for the development of future treatments. Underserved conditions, including neurodegenerative diseases and oncological pathologies, can be positively influenced by control of cell differentiation, and so make ideal targets for retinoid-based drugs. High-throughput screening of compounds, for activity against signalling proteins, is a vital technique for the development of new candidate molecules, and novel fluorescent retinoids offer an opportunity to develop a new screening methodology based on solvatochromism.

High-throughput screening will be achieved using fluorescent retinoid derivative DC271 as a competitive probe for CRAB proteins, as well as the nuclear retinoid receptors. The ability of the compound to increase fluorescence upon incorporation into a hydrophobic location will make an ideal probe for the encapsulated binding sites of both the CRAB and RAR proteins.

In tandem, co-crystallography of target proteins with synthetic retinoid ligands will help define and describe the mode of action of each probe, reinforcing the validity of the assay methodology. Better understanding of how binding occurs, and how the protein commensurately reacts, will inform choices about future ligand modification.

The aims of this project are therefore:

- Characterise the solvatochromic fluorescence of novel fluorescent retinoids and their ability to act as ligands for relevant signalling proteins;
- Develop a high-throughput screening assay for ligand investigation, based on the solvatochromic fluorescence of DC271, which will allow binding constants to be determined for competing compounds;
- Co-crystallise relevant binding proteins with synthetic retinoids, demonstrating binding site incorporation and characterising protein-ligand interactions in atomic detail.

The project will engender a better understanding of retinoid binding proteins, their affinities for fluorescent retinoid derivatives, and how they can be used as high-throughput ligand probes for relevant proteins.

2: Expression and Purification of CRAB and RAR Proteins

2.1. Introduction

Heterologous expression of proteins in culture has paved the way for a myriad of modern developments in the biological and biochemical sciences. The use of *Escherichia coli*, *Saccharomyces/Pichia* strains, insect cell systems, and other organisms capable of growth in culture, has allowed for the production of proteins outside of their normal environment, and in large quantities surpassing that which can be extracted from natural sources. (Baneyx, 1999, Cherbas & Cherbas, 2007, van Oers *et al.*, 2015) *E. coli* in particular grow to high cell densities, in media that is inexpensive and commonly available, and are amongst the best genetically characterised organisms, with correspondingly well engineered vector systems. (Baneyx, 1999)

The BL21(DE3) strain is amongst the most commonly used *E. coli* expression hosts, containing an expression system for the T7 RNA polymerase under control of a *lac*UV5 promoter. (Studier & Moffatt, 1986, Jeong *et al.*, 2009) In this it differs from the core BL21 strain, and is capable of expression, controlled with IPTG, of constructs using a T7 promoter sequence. This system gives a high level of control over expression, and can over-produce protein to levels 10-30% of the whole cell quantity. (Hannig & Makrides, 1998)

Affinity chromatography, based on production of protein appended to an affinity tag, is the most common method of purification after recombinant expression. Commonly used tags include glutathione-S-transferase protein (GST), maltose binding protein (MBP) or histidine metal affinity (His_6) tags, all of which provide affinity to an immobilised substrate. (Porath *et al.*, 1975, Smith & Johnson, 1988, di Guana *et al.*, 1988, Porath, 1992, Harper & Speicher, 2011) Capturing protein using affinity to the immobilised substrate, and subsequently removing contaminants by washing, creates a sample of high purity directly from cell lysate which can be eluted using common buffer components. Tags, including GST, MBP, and SUMO protein (Malakhov *et al.*, 2004, Marblestone *et al.*, 2006), can also reduce incorporation of proteins into inclusion bodies by promoting correct folding of the protein of interest. Maltose binding protein forms a large (40 kDa) tag and, along with other large tags, is a common crystal contaminant if purification is not suitably rigorous. (Niedzialkowska *et al.*, 2016) Sequence specific cleavage of tags using a protease recognition site (Waugh, 2011) is usually necessary, unless crystallisation depends on the presence of a tag (Townsend *et al.*, 2014), or the tag sequence is suitably inconspicuous.

Size-exclusion chromatography uses the properties of a porous solid phase to slow the passage of lower molecular weight molecules relative to the larger constituents of a sample. (Porath & Flodin, 1959) Modern size-exclusion matrices are based on dextran covalently linked with highly polymerised agarose. (Hellberg *et al.*, 1996) Separation by size is a useful ‘polishing’ step for protein purification, typically carried out after immobilised metal ion affinity chromatography

(IMAC) or other affinity technique, and allows for the washing out of contaminants over a long solid phase bed to leave a relatively pure protein sample.

The over-expression and purification of CRAB and RAR proteins in milligram quantities allowed suitable scale crystallisation experiments and assaying to be undertaken. Ideally, optimisation of purity improves the chances of successful crystallisation, and reduces the chance of disruptive effects under assay conditions.

2.2. Results

2.2.1. Expression and purification of CRABPII

Heterologous expression of CRABPII in *E. coli* was established based on existing protocols and using a gifted construct. (Redfern & Wilson, 1993) The pGEX2T-hCRABPII plasmid was transformed into *E. coli* strain BL21(DE3) (NEB) for expression, as well as DH5α (NEB) for long-term plasmid storage. The design of the pGEX (Smith & Johnson, 1988) construct appends a *Schistosoma japonicum* glutathione-S-transferase (GST) tag to the protein of interest, linked by a thrombin cleavage site (LVPR|GS) (Waugh, 2011) that allows for efficient proteolysis after initial purification steps, and yields the canonical (Uniprot P29373) CRABPII sequence. (UniProt Consortium, 2017) Sequence 1 shows the primary amino acid sequence of CRABPII after cleavage, and purified for use in assays and crystallisation experiments.

GSGPNFSGNW KIIRSENFEE LLKVLGVNVM LRKIAVAAAS
KPAVEIKQEG DTFYIKTSTT VRTTEINFKV GEEFEEQTVD
GRPCKSLVKW ESENKMVCEQ KLLKGEGPKT SWTRELTNDG
ELILTMTADD VVCTRVYVRE

15764 Da

Sequence 1: Primary sequence of CRABPII from pGEX2t-hCRABPII after tag cleavage, used in expression and purification of the protein. Tag sequence is marked in red, remnants of the cleavage site are underlined. Molecular weight calculated by Expasy MW. (Wilkins *et al.*, 1999, Gasteiger *et al.*, 2003)

Purification was carried out using standard GST-affinity chromatography, on a GSTrap™ HP 1ml column (GE Healthcare), which allows the binding of the GST tag under normal buffer conditions to resin-captured glutathione. Subsequent addition of 15 mM reduced glutathione preferentially binds to the GST tag protein and causes elution. (Harper & Speicher, 2011, Smith & Johnson, 1988) The resulting protein was analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) (Laemmli, 1970) to determine purity and confirm the mass of the protein was correct. The gel bands (Figure 2.1) correspond well to the expected masses, approximately 15 kDa for the cleaved protein, and 25 kDa for the GST tag. Calculations of concentration were made by spectrophotometry at 280 nm, based on extinction coefficients calculated by Expasy ProtParam. (Wilkins *et al.*, 1999, Gasteiger *et al.*, 2003)

After purification of CRABPII by GST affinity and cleavage of the tag on the column, protein fractions were pooled and concentrated using centrifuge spin methods, before storage at 4 °C, or dialysed overnight to exchange buffers. Typical yields from 2 L of *E. coli* culture were ~5 mg.

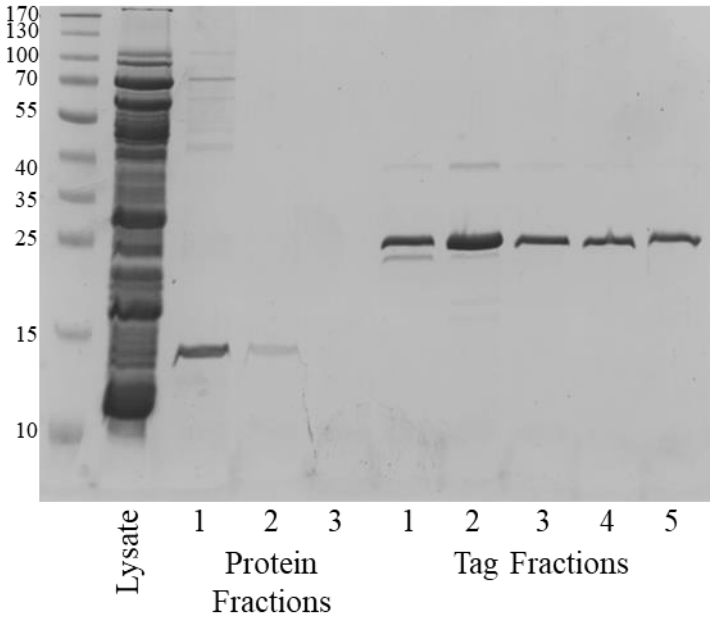


Figure 2.1: 12% SDS PAGE gel from purification of BL21(DE3) pGEX2T-CRABPII. Molecular weight markers labelled in kDa. Lysate lane represents a sample of cell lysate prior to purification. Protein fractions represent those fractions eluted from the gel after proteolysis with thrombin (~15 kDa). Tag fractions represent protein eluted using reduced glutathione, primarily cleaved GST tag (~25 kDa).

2.2.2. Expression and purification of CRABPI-L29C

CRABPI-L29C was expressed using a pET28a vector (Rosenberg *et al.*, 1987), with an insertion coding for hCRABPI (Uniprot 29762) with a single L29C modification. (UniProt Consortium, 2017) Heterologous expression was carried out in BL21(DE3) *E. coli*, and transformations into DH5α were carried out for plasmid storage. The vector appends a His_6 tag to the protein through a linker containing the thrombin cleavage site LVPR|GS, which allowed for removal of the tag sequence after expression. Sequence 2 shows the primary sequence resulting from expression, including the site used for cleavage and the residues left appended to the protein as a result. Expression was tested (Figure 2.2 - A), which demonstrates somewhat 'leaky' expression even under control of the T7 promoter.

MGSS**HHHHHH** SSGLVPRGSH MASMTGGQQM GRGSAMPNFA
GTWKMRSSEN FDELLKALGV NAMCRKVAVA AASKPHVEIR
QDGDQFYIKT STTVRTTEIN FKVGEGFEEE TVDGRKCRSL
ATWENENKIH CTQTLLEGDG PKTYWTRELA NDELILTFGA
DDVVCTRIYV RE

Construct - 19170 Da | Post cleavage - 17288 Da

Sequence 2: Primary amino acid sequence of CRABPI-L29C from pET28a-hCRABPI-L29C, used in expression and purification of the protein. Tag sequence is marked in red, cleavage site is underlined, His_6 tag shown in bold, mutation site is shown in blue. Molecular weight calculated by Expasy MW. (Wilkins *et al.*, 1999, Gasteiger *et al.*, 2003)

Purification was carried out using nickel affinity chromatography (Ni IMAC) (Porath *et al.*, 1975, Porath, 1992) to immobilise the His_6-tagged protein onto the column. Subsequent washing with a gradient of imidazole-containing buffer released the protein, which was collected and analysed by SDS PAGE (Figure 2.2). The pooled protein (~19 kDa) was cleaved by the addition of thrombin protease, the activity of which was calibrated versus the protein (Figure 2.2 - C), and re-purified by nickel affinity. The resulting protein (~17 kDa) was once again analysed by SDS PAGE, and concentration estimated using A_{280} nm. Mass spectrometry was used to confirm the mass of the protein both before and after proteolysis (Appendix 3 and Appendix 4). The two stage method of purification was successful and yielded ~15 mg per 2 L culture.

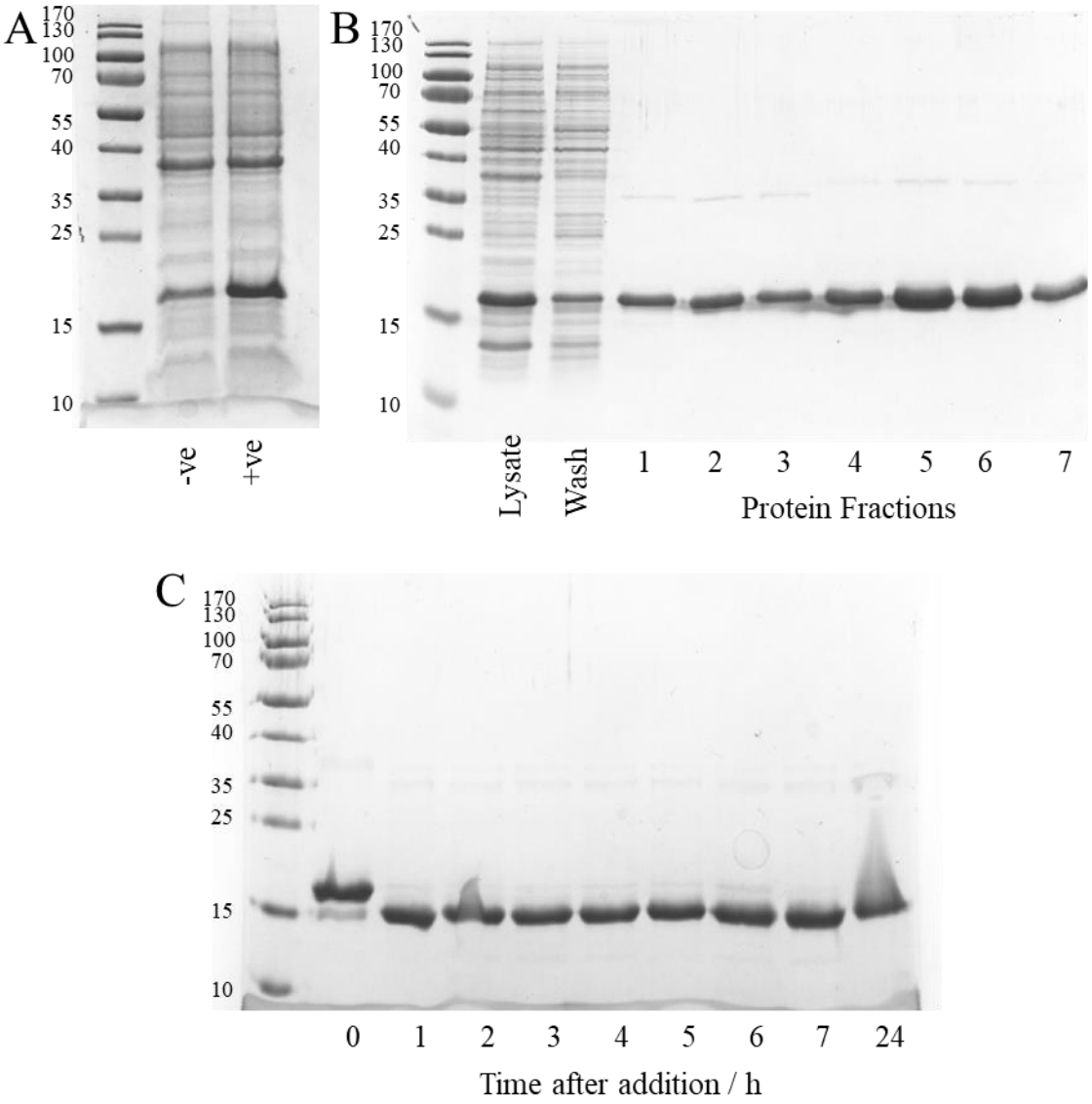


Figure 2.2: SDS PAGE gels from Purification of CRABPI-L29C, molecular weight markers labelled in kDa. A - Expression test of pET28a-CRABPI-L29C in BL21(DE3), induced and non-induced samples both show levels of CRABPI-L29C expression (~19 kDa). B - Elution profile of CRABPI-L29C in imidazole gradient elution from Ni IMAC. Protein can be seen contained in the cell lysate, as well as the protein fractions 1-7. C - Test cleavage of CRABPI-L29C using ~80 units thrombin.

After crystal structure determination demonstrated partial dimerisation through the modified residue (see 4.2.6), analytical size exclusion (SEC) was carried out to confirm the dimerisation status (Figure 2.3). Two peaks were identified at 59.3 ml and 69.5 ml, with molecular weights calculated based on column calibration of 44.1 kDa and 23.1 kDa respectively. Given the non-perfect stokes radius equivalency of both monomer and dimer, it is reasonable to conclude that these represent dimeric (34.4 kDa) and monomeric (17.2 kDa) CRABPI-L29C. Concentration estimates based on absorbance at 280 nm suggest that the approximately one third of the protein was dimerised at the time of separation, making it likely that dimerisation was transient in solution and driven in favour of the dimer species during crystallisation (see 4.2).

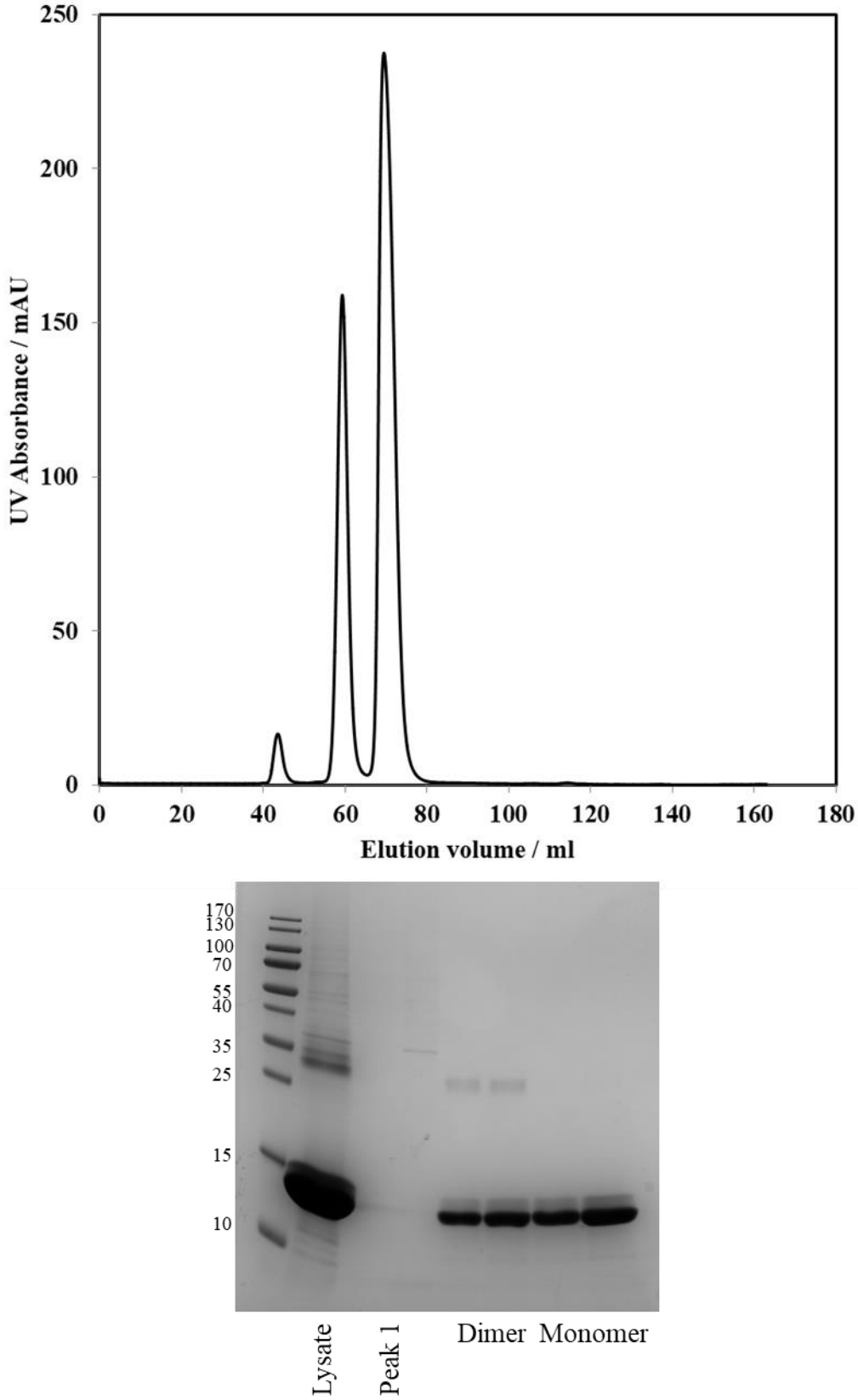


Figure 2.3: Analytical size-exclusion chromatography trace for CRABPI-L29C. Peak at 59.3 ml corresponds to a 44.1 kDa moiety; peak at 69.5 ml corresponds to a 23.1 kDa moiety. Peak at 43.6 ml is within the void volume of the column. SDS PAGE indicates that fractions from the column only contain CRABPI-L29C, and a small dimer band visible (~34 kDa) in the dimer peak fractions.

2.2.3. Expression and purification of RARα and RARγ

The expression of the ligand binding domains of RARα and RARγ proteins was carried out using vectors from the pOPIN vector system (Bird, 2011), and commercially produced inserts, codon optimised for *E. coli* expression containing the proteins of interest (Sequence 3). Plasmids for expression were recovered from frozen stocks prepared by K. Zmarzly. Insert sequences were based on literature examples of crystallised LBDs. (le Maire *et al.*, 2010, Chebaro *et al.*, 2017)

RARα 176-421

MA**HHHHHH**GS DSEVNQEAKP EVKPEVKPET HINLKVSDGS
SEIFFKIKKT TPLRRLMEAF AKRQGKEMDS LRFLYDGIRI
QADQTPEDLD MEDNDIIEAH REQISSGLEV LFQGPESYTL
TPEVGELIEK VRKAHQETFP ALCQLGKYTT NNSSEQRVSL
DIDLWDKFSE LSTKCIIKTV EFAKQLPGFT TLTIADQITL
LKAACLDILI LRICTRYTPE QDTMTFSDGL TLNRTQMHNA
GFGPLTDLVF AFANQLLPLE MDDAETGLLS AICLICGDRQ
DLEQPDRVDM LQEPLLEALK VYVRKRRPSR PHMFPKMLMK
ITDLRSISAK GAERVITLKM EIPGSMPPLI QEMLENSEGL
D

Construct - 40962 Da | Post cleavage - 27920 Da

RARγ 178-423

MA**HHHHHH**GS DSEVNQEAKP EVKPEVKPET HINLKVSDGS
SEIFFKIKKT TPLRRLMEAF AKRQGKEMDS LRFLYDGIRI
QADQTPEDLD MEDNDIIEAH REQISSGLEV LFQGPDSYEL
SPQLEELITK VSKAHQETFP SLCQLGKYTT NSSADHRVQL
DLGLWDKFSE LATKCIIKIV EFAKRLPGFT GLSIADQITL
LKAACLDILM LRICTRYTPE QDTMTFSDGL TLNRTQMHNA
GFGPLTDLVF AFAGQLLPLE MDDTETGLLS AICLICGDRM
DLEEPEKVDK LQEPLLEALR LYARRRRPSQ PYMFPRMLMK
ITDLRGISTK GAERAITLKM EIPGPMPPLI REMLENPEMF
E

Construct - 41055 Da | Post cleavage - 28013 Da

Sequence 3: Primary amino acid sequences of RARα and RARγ from pOPINS3C vector constructs, used in expression and purification of the protein. Tag sequence (including SUMO solubility tag) is marked in red, cleavage site is underlined. His_6 affinity sequence is marked in bold. Molecular weight calculated by Expasy MW. (Wilkins *et al.*, 1999, Gasteiger *et al.*, 2003)

The pOPIN vector series is constructed around a common backbone and cloning site that allows for the use of the same insert and primer set for cloning into multiple vectors. The pOPINS3C vector was primarily used for its additional SUMO solubility tag (Malakhov *et al.*, 2004, Marblestone *et al.*, 2006), and also provided an N-terminal His_6 tag as well as a human rhinovirus 3C (HRV3C) cleavage site for removal of the entire tag section. HRV3C is closely related to the tobacco etch virus (TEV) protease, and cleaves at a LEVLFQ|GP recognition site. (Waugh, 2011) Related vectors pOPINK (GST tagged) and pOPINF (His_6 without solubility tag) were also considered. In-Fusion® cloning (Takara biosciences) was used to perform the recombinational cloning step, inserting the construct into the vector. (Park *et al.*, 2015) The resulting plasmids, pOPINS3C-hRARα and pOPINS3C-hRARγ were transformed into DH5α (NEB) for plasmid storage and BL21(DE3) (NEB) for expression.

Purification of the RAR proteins was carried out using Ni IMAC (Porath *et al.*, 1975, Porath, 1992), adapted from a standard protocol for HisTrap™ HP columns (GE Healthcare). Where necessary, size exclusion chromatography was carried out to enhance final purity (Figure 2.4 - C).

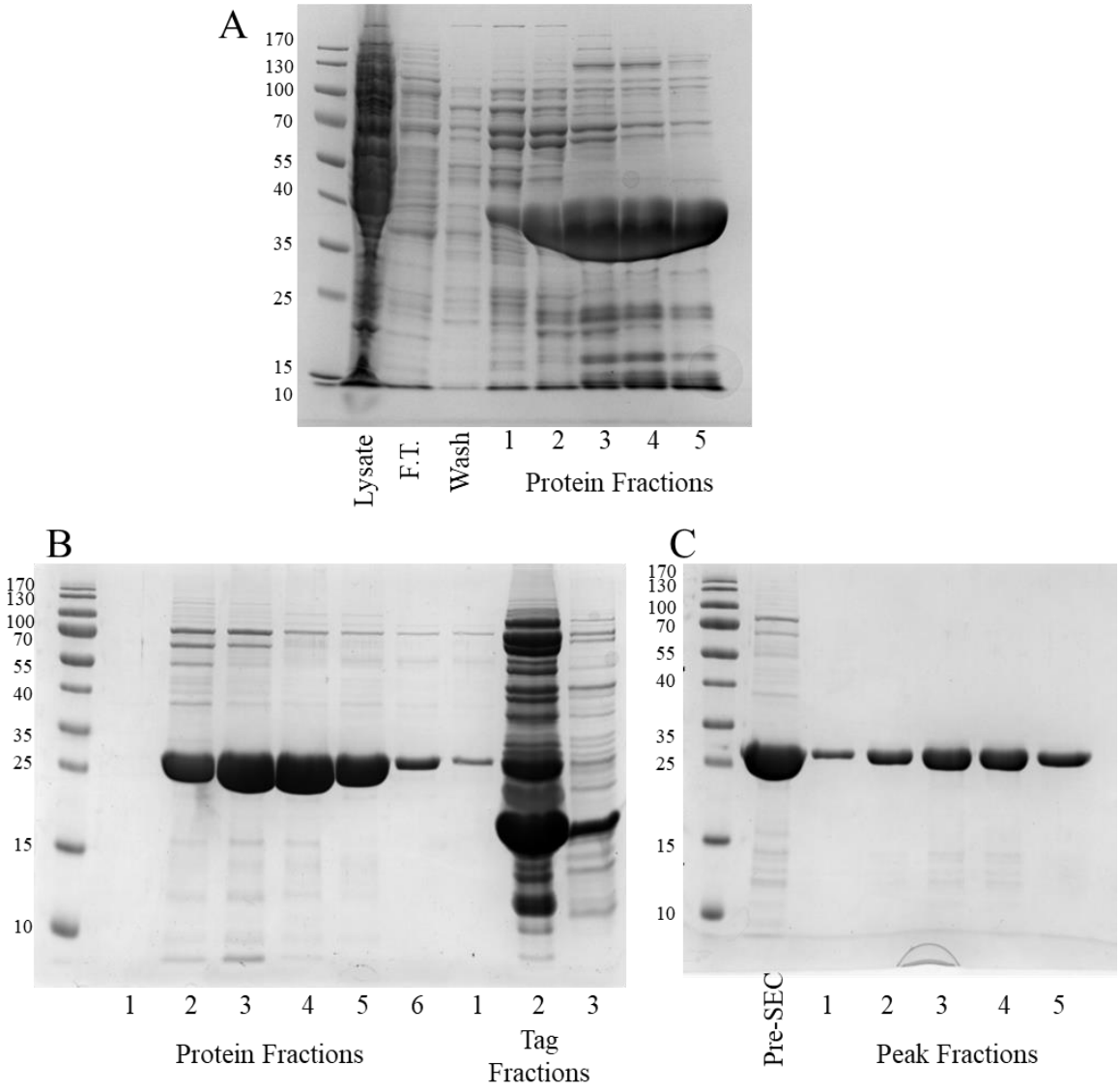


Figure 2.4: SDS PAGE gels from purification of RARα, molecular weight markers labelled in kDa. A - Elution of BL21(DE3) pOPINS3C-RARα using an imidazole gradient program, protein is visible as an overloaded band in protein fractions 1-5 (~40 kDa). B - RARα Ni IMAC elution fractions after cleavage with HRV3C protease, protein visible in fractions 1-6 (~28 kDa). C - SEC for polishing of RARα before assays and crystallisation experiments. The single captured peak contains the pure protein.

RARα was purified routinely by Ni IMAC, with subsequent cleavage of the expression and solubility tag using HRV3C protease. Analysis by SDS PAGE (Figure 2.4 A and B) shows that initial elution by imidazole gradient released a series of impurities that were captured by the second IMAC step in which the protein of interested had been cleaved from its affinity tag. Future purifications improved first stage purity by increasing the concentration of imidazole in the binding buffer, creating a more stringent binding environment. Where necessary for purity, SEC was carried out (HiLoad 16/600 Superdex 75 pg - GE Healthcare) conserving fractions that made up the primary peak. Mass spectrometry (Appendix 5) was used to confirm correct protein mass (27920 Da). Typical RARα purification from 2 L of culture yielded 4 mg of protein after cleavage and all purification steps.

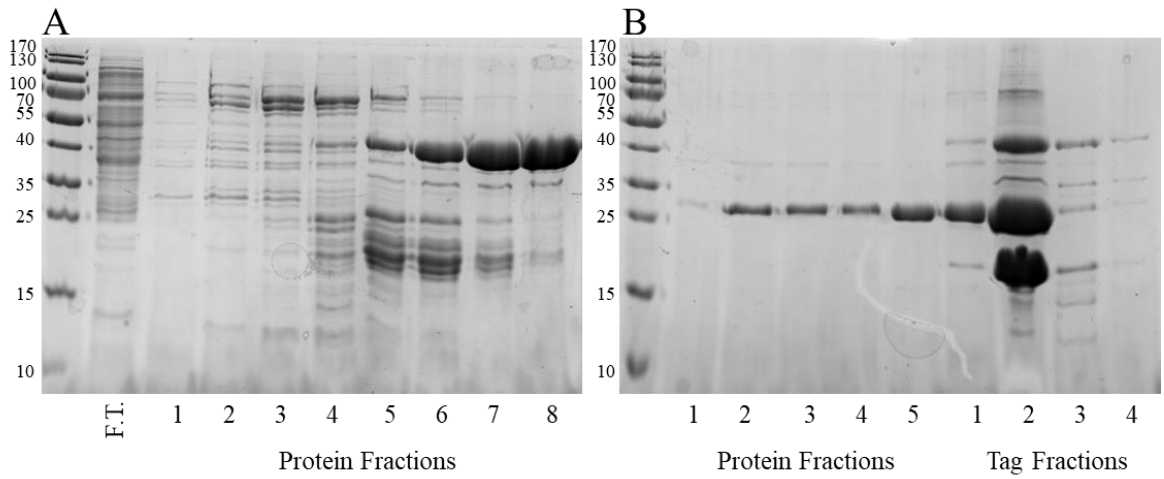


Figure 2.5: SDS PAGE gels from purification of RARγ, molecular weight markers labelled in kDa. A - Elution of RARγ by imidazole gradient. Protein is visible in protein fractions 5-8 as SUMO-S3C_RARγ (~40 kDa). B - RARγ in protein fractions post cleavage with HRV3C protease. Nickel affinity is used to capture the tag fractions containing His_6 tagged SUMO tag as well as contaminant proteins.

Purification of the RARγ isoform was carried out using an identical protocol as RARα, and yielded similar results. Expression and solubilisation of the protein using the pOPINS3C-hRARγ vector was successful. Protein capture by IMAC showed initial impurities (Figure 2.5 - A), and stringency of binding was increased with greater imidazole concentration. Second round purification after HRV3C cleavage generally yielded pure protein for use in assays and crystallisation experiments (Figure 2.5 - B). If necessary, size exclusion chromatography could be carried out to 'polish' prior to storage of the protein. Mass spectrometry (Appendix 6) confirms the identity of the expressed protein by mass (28012 Da). RARγ purification generally yielded 3 mg of protein per 2 L culture for further use.

2.2.4. Expression and purification of RARβ

Expression of RARβ was attempted using three different constructs. Primarily, a pOPIN3SC-hRARβ vector was assembled in the same manner as RARα and RARγ, consisting of a His_6 purification tag, SUMO solubility tag (Malakhov *et al.*, 2004, Marblestone *et al.*, 2006) and HRV3C cleavage site. A second vector produced commercially (pJOE-hRARβ) appends a His_6 tag to the protein directly without a protease recognition site for cleavage. A third vector, received as a gift and used previously in RARβ crystallisation experiments (Osz *et al.*, 2012), was based on the pET15-b framework.

pOPINS3C-RARβ 176-421

MA**HHHHHH**GS DSEVNQEAKP EVKPEVKPET HINLKVSDGS
SEIFFKIKKT TPLRRLMEAF AKRQGKEMDS LRFLYDGIRI
QADQTPEDLD MEDNDIIEAH REQISSGLEV LFQGPESYEM
TAELDDLTEK IRKAHQETFP SLCQLGKYTT NSSADHRVRL
DLGLWDKFSE LATKCIIKIV EFAKRLPGFT GLTIADQITL
LKAACLDILI LRICTRYTPE QDTMTFSDGL TLNRTQMHNA
GFGPLTDLVF TFANQLLPLE MDDTETGLLS AICLICGDRQ
DLEEPTKVDK LQEPLLEALK IYIRKRRPSK PHMFPKILMK
ITDLRSISAK GAERVITLKM EIPGSMPPLI QEMLENSEGH
E

Construct - 41037 Da | Post cleavage - 27866

pJOE-RARβ 182-417

MTMIT**HHHHH** **H**GSAELDDLT EKIRKAHQET FPSLCQLGKY
TTNSSADHRV RLDLGLWDKF SELATKCIIK IVEFAKRLPG
FTGLTIADQI TLLKAACLDI LILRICTRYT PEQDTMTFSD
GLTLNRTQMH NAGFGPLTDL VFTFANQLLP LEMDDTETGL
LSAICLICGD RQDLEEPTKV DKLQEPLLEA LKIYIRKRRP
SKPHMFPKIL MKITDLRSIS AKGAERVITL KMEIPGSMPP
LIQEMLENS

28193 Da

pET15b-RARβ 169-414

GSTSKQECTE SYEMTAELDD LTEKIRKAHQ ETFPSLCQLG
KYTTNSSADH RVRLDLGLWD KFSELATKCI IKIVEFAKRL
PGFTGLTIAD QITLLKAACL DILILRICTR YTPEQDTMTF
SDGLTLNRTQ MHNAGFGPLT DLVFTFANQL LPLEMDDTET
GLLSAICLIC GDRQDLEEPT KVDKLQEPLL EALKIYIRKR
RPSKPHMFPK ILMKITDLRS ISAKGAERVI TLKMEIPGSM
PPLIQEML

27980 Da

Sequence 4: Primary amino acid sequences of RARβ from pOPINS3C, pJOE and pET15b vector constructs, used in expression and attempted purification of the protein. Tag sequence (including SUMO solubility tag) is marked in red, cleavage site is underlined. Molecular weight calculated by Expasy MW. (Wilkins *et al.*, 1999, Gasteiger *et al.*, 2003)

Extraction of soluble protein from cell culture proved largely unsuccessful. In the case of pOPINS3C-hRARβ expressed in BL21(DE3) *E. coli*, the protein was absent from the soluble fractions and no meaningful quantity could be captured by Ni IMAC (Figure 2.6 - A).

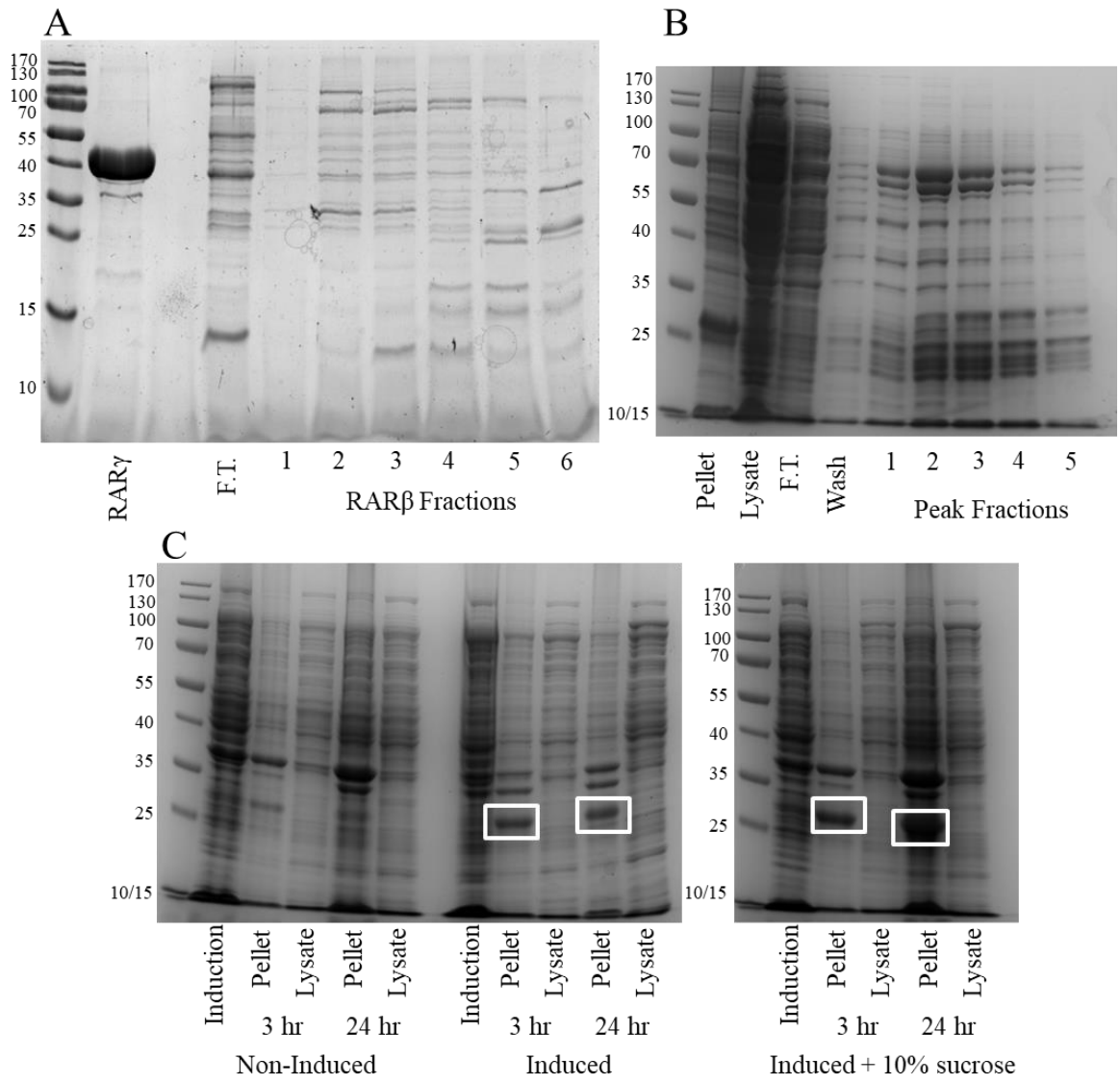


Figure 2.6: SDS PAGE gels from purification of RARβ, molecular weight markers labelled in kDa. A - Purification of BL21(DE3) pOPINS3C-RARβ (~40 kDa) showing no protein present in the flow through (F.T.) or wash fractions, or in the collected peak fraction. B - Purification of BL21(DE3) pJOE-RARβ showing protein concentrated in the pellet fraction (~25 kDa), and absent in the cell lysate, F.T., wash and peak fractions. C - Test expression series of pET15b-RARβ in BL21(DE3), the protein (~25 kDa) is found only in the pellet of induced cells.

Similarly, expression via the pJOE-hRARβ vector (produced commercially by GenScript) demonstrated ample protein production, but incorporation into the insoluble fraction found in the cell pellet (Figure 2.6 - B). The third, gifted, pET15b-hRARβ vector was used in concert with an established protocol, and addition of 10 % sucrose to the expression media was tested. (Osz *et al.*, 2012) A ~28 kDa protein band can be seen in the insoluble pellet of both the 'induced' and 'induced + 10 % sucrose' test expression stages, with a noticeable advantage to the addition of

sucrose and overnight growth at 25 °C. Nevertheless, the protein is demonstrably insoluble and would not be recoverable using standard Ni IMAC purification. It is likely that future expression of soluble RARβ will be possible with further optimisation of the expression protocol, incorporating literature techniques as well as known methods for solubilising or refolding of proteins. (Nishihara *et al.*, 2000, Osz *et al.*, 2012, Lobstein *et al.*, 2016, Selas Castiñeiras *et al.*, 2018)

2.3. Conclusions

Heterologous expression systems for CRABPII, CRABPI-L29C, RARα, and RARγ were successfully established in *E. coli*, and purification was carried out using standard GST affinity or IMAC techniques. By design, two-stage GST and IMAC purification techniques comprised both capture and purification steps of a typical protein purification, with polishing carried out with SEC (when necessary). Analytical SEC was carried out to determine the dimerisation status of CRABPI-L29C, which demonstrated a dimeric form that had been observed in crystallisation experiments. Tuning of imidazole elution gradients in IMAC to improve purity were particularly successful and typically rendered protein pure enough for assay and crystallisation experiments.

The expression of CRABPII was carried out successfully based on previous work, and repeatedly produced pure protein for use in binding assays (3.2) and crystallisation experiments (4.2). CRABPI-L29C was similarly well expressed and purified based on standard Ni IMAC techniques. The resulting protein was crystallised with natural and synthetic ligands, and its binding to select retinoid compounds was determined (5.2).

RARα and RARγ were expressed using the pOPIN vector system to append a SUMO solubility tag. This process worked well and the protein produced was used for crystallisation experiments, as well as binding of a library of retinoids (5.2.3).

Attempts were made to express and purify RARβ, but the protein was consistently found incorporated into the insoluble fraction (Figure 2.6). It is likely that this can be remedied in future, through the use of refolding, or chaperone strains, as well as further refinement of the expression system. (Nishihara *et al.*, 2000, Selas Castiñeiras *et al.*, 2018)

2.4. Methods

2.4.1. Bacterial transformation

Transformation of plasmids into BL21(DE3) and DH5α competent cells were carried out using protocols supplied by New England Biolabs (NEB). Competent cells were allowed to defrost on ice for 10 min. 50-100 ng DNA were added, and allowed to combine on ice for 30 min. Heat shock (42 °C) was carried out for 10 s (BL21(DE3)) or 30 s (DH5α) before incubation on ice for 5 min. Volume was made up to 1 ml with SOC media, and shaken (150 rpm) at 37 °C for 60 min. Selection plates with suitable antibiotics were prepared and spread with culture (100 µl and 10:90 µl dilution in SOC media). Plates were incubated overnight (37 °C) and colonies selected for growth.

2.4.2. Plasmid recovery

Plasmids were extracted from 5 ml cell culture using a standard mini-prep (GeneJet - Thermo Scientific) alkaline lysis procedure. Plasmid was frozen in recovery buffer at −80 °C.

2.4.3. CRABPII Expression

CRAPBII protein (GST-tagged) was expressed in BL21(DE3) competent *E. coli* cells (NEB) transformed with a pGEX 4T1 vector. Starter cultures (20 g/L LB broth, 100 µg/ml ampicillin, 25 ml) were grown overnight and inoculated into 1 L LB media grown at 37 °C. Expression was induced after 4 h shaking (37 °C, 150 rpm) with IPTG (1 mM in culture), before shaking overnight (20 h). The resulting cultures were pelleted using an Avanti Hi-Speed centrifuge (JLA 8.1000, 4000 rpm, 25 min, 4 °C), before removal of supernatant and freezing at −80 °C.

2.4.4. CRABPII Purification

Defrosted cell pellets were resuspended in 15 ml of lysis buffer (10 mM DTT, 100 mM $MgCl_2$, PBS pH 7.5, 4 °C) and lysed with the addition of 0.5 mg/ml lysozyme for 30 min. The resulting suspension was sonicated (2 min, 40% power, on ice) and centrifuged (Avanti Hi-Speed JLA 25.50, 20,000 rpm, 50 min, 4 °C). The supernatant was loaded onto a GSTrap™ HP affinity column (1 ml, GE healthcare) and washed using 30 ml PBS pH 7.5. The column was then loaded with 1 ml thrombin in PBS (GE Healthcare, ~80 units) and reserved at room temperature for 16 h to cleave the GST tag. The cleaved protein was eluted using PBS pH 7.5 (3 x 1 ml), and bound GST tag was eluted with reduced glutathione (10 mM in PBS, 10 ml). The resulting protein fractions were analysed by SDS PAGE and concentration estimated by absorbance at 280 nm using a Thermo Scientific Nanodrop A_{280}.

2.4.5. CRABPI-L29C expression

CRABPI-L29C was expressed in BL21(DE3) competent *E. coli* (NEB) transformed with a pET-28a vector incorporating an N-terminal His_6-tag. Starter cultures (20 g/L LB broth, 50 μg/ml kanamycin, 25 ml) were grown overnight (150 rpm, 37 °C) then transferred into 1 L expression flasks (LB / kanamycin) and grown with shaking at 37 °C, 150 rpm. Induction with IPTG (1 mM in culture) was carried out at an OD_{600} of 0.6, before shaking overnight (20 h) at 25 °C. Cultures were pelleted using an Avanti Hi-Speed centrifuge (JLA 8.1000, 4000 rpm, 25 min, 4 °C), before removal of the supernatant and freezing at −80 °C.

2.4.6. CRABPI-L29C purification

Defrosted cell pellets were resuspended with 15 ml loading buffer (10 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 8). The resulting suspension was sonicated on ice (40% power, 2 min) and centrifuged (Avanti Hi-Speed JA25.50, 20,000 rpm, 50 min, 4 °C). Supernatant was loaded onto a HisTrap™ HP column (1 ml, GE healthcare) and washed with 30 ml of loading buffer. Elution was carried out in a smooth gradient against elution buffer (500 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 8) on an ÄKTA pure FPLC. The resulting protein-containing fractions were analysed by SDS PAGE and concentration was estimated using Thermo Scientific Nanodrop A_{280}. Isolated protein was incubated overnight with thrombin (~80 units) to cleave the His-tag, and dialysed into buffer without imidazole (500 mM NaCl, 20 mM Tris-HCl, pH 8). The digest mixture was loaded onto a HisTrap™ HP column (1 ml, GE healthcare) and cleaved protein was eluted without binding using 3 ml loading buffer into 3x1 ml fractions. Bound tag was then eluted with elution buffer and the column washed to remove any other contaminants. The resulting protein and tag fractions were analysed by SDS PAGE and Nanodrop A_{280}, before use in assays and crystallisation experiments.

2.4.7. In-Fusion® Cloning

In-Fusion® cloning (Park *et al.*, 2015) was carried out according to instructions provided by Takara Biosciences. A PCR reaction to amplify a commercial gene insert (GenScript) was carried out using 12.5 μl HiFi PCR mix (Takara), 1 μl pUC57 plasmid containing the gene insert, 2.5 μl each forward and reverse primers (5 μM stock), 0.25 μl DMSO and 6.25 μl ddH_2O. Denaturation took place at 98 °C (10 s), annealing at 62 °C (15 s) and elongation at 72 °C (5 min), over a period of thirty cycles. The resulting product was separated on a 1.5% agarose gel, the primary band of which was recovered and inserts extracted using a commercial kit (NucleoSpin).

The pOPIN vector, recovered from miniprep of DH5α storage cells, was linearised using HindIII and KpnI overnight, at 37 °C. The resulting product was similarly recovered using a NucleoSpin gel extraction kit.

In-Fusion annealing was carried out using 2 µl In-Fusion premix (Takara) and a range of vector/insert ratios, in 10 µl solution made up with ddH_2O. Each reaction was incubated at 50 °C for 15 min, cooled on ice and analysed by agarose gel electrophoresis for insertion of the construct. The resulting vector product was used in transformations of Stellar *E. coli* cells (Takara) following standard procedures. Successful transformants were selected using ampicillin/LB/agar plates, and colonies picked for PCR insert confirmation using the same program as above. The colonies that showed amplification of the insert were grown overnight and the plasmid recovered before service Illumina sequencing and transformation into DH5α and BL21(DE3) *E. coli* strains.

2.4.8. RARα and RARγ expression

RARα/γ protein was heterologously expressed in BL21(DE3) cells (NEB) transformed with a pOPINS3C vectors carrying the gene construct and containing a His-3C-SUMO tag used for solubility. (Bird, 2011) Transformations were carried out under standard conditions, using a 10 s heat shock. Successful transformants were grown overnight in LB and frozen for storage at −80 °C in 50% glycerol. Expression cultures were inoculated from glycerol stocks into 25 ml of LB media (Melford) with 100 µg/ml ampicillin (Sigma Aldrich) and grown overnight with shaking (37 °C, 150 rpm). The overnight cultures were transferred into 1 L expression flasks (LB/ampicillin 100 µg/ml) and grown with shaking at 37 °C, 150 rpm. Induction was carried out at an OD_{600} of 0.6, with 1 ml of 1 M IPTG (1 mM in culture), before shaking overnight (20 h) at 25 °C. The resulting cultures were spun down into pellets using an Avanti Hi-Speed centrifuge (JLA 8.1000, 4000 rpm, 25 min, 4 °C), before the supernatant was removed and the bacterial pellet frozen at –80 °C.

2.4.9. RARα and RARγ purification

RARα/γ (popinS3C, BL21(DE3) (NEB)) was resuspended from pellet with 15 ml buffer A (10 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 8) and the resulting suspension was sonicated on ice (40% power, 2 min) before centrifuging (Avanti Hi-Speed JA25.50, 20,000 rpm, 50 min, 4 °C). Supernatant was loaded onto HisTrap™ HP column (1 ml, GE healthcare) and washed using 30 ml buffer A. The column was washed with buffer A until UV traces returned to baseline, before elution in gradient with buffer B (500 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 8). The resulting protein containing fractions were analysed by SDS PAGE and concentration was estimated using Thermo Scientific Nanodrop A_{280}. RAR proteins were cleaved overnight from 3C-SUMO-His tag using 1 unit/100 µg protein of HRV3C protease (Merck), based on Nanodrop A_{280} calculated protein concentration, and dialysed into fresh buffer without imidazole (500 mM NaCl, 20 mM Tris-HCl, pH 8). Solution after digest was loaded onto a HisTrap™ column (1 ml, GE healthcare) and cleaved protein was eluted immediately using 3 ml buffer A into 3x1 ml fractions. Bound tag was then eluted with buffer B and the column washed to remove any other contaminants. The resulting protein and tag fractions were analysed by SDS PAGE (120 V, 70 min), before use in assays and crystallisation experiments.

2.4.10. RARβ expression

RARβ protein was heterologously expressed in BL21(DE3) cells (NEB) transformed with one of three vectors carrying the gene construct - pOPINS3C-hRARβ (amp/IPTG), pJOE-hRARβ (amp/rhamnose) and pET15b-hRARβ (amp/IPTG).

Transformations were carried out under standard conditions, using a 10 s heat shock, and plated on selective media (LB/amp). Successful transformants were picked and grown overnight in LB before freezing for storage at −80 °C in 50% glycerol. Expression cultures were inoculated from glycerol stocks into 25 ml of LB media (Melford) with appropriate antibiotic and grown overnight with shaking (37 °C, 150 rpm). pET15b-hRARα cultures were adulterated with 10% sucrose. The overnight cultures were transferred into 1 L expression flasks (LB/antibiotic) and grown shaking at 37 °C, 150 rpm. Induction was carried out at an OD_{600} of 0.6, with 1 ml of 1 M IPTG (1 mM in culture - pOPINS3C and pET15b) or rhamnose (0.2 % in culture, pJOE), before shaking overnight (20 h) at 25 °C. The resulting cultures were spun down into pellets using an Avanti Hi-Speed centrifuge (JLA 8.1000, 4000 rpm, 25 min, 4 °C), before the supernatant was removed and the bacterial pellet frozen at −80 °C.

2.4.11. RARβ purification

RARβ purification was attempted using the same method as RARα and RARγ. SDS PAGE analysis indicates that any recombinant protein was incorporated into the insoluble fraction.

2.4.12. Mass Spectrometry

Time-of-Flight Electrospray Ionization Mass Spectrometry (ToF-ESI-MS) was carried out by departmental services, using a Quad Time-of-Flight (QToF) Premier spectrometer (Waters). Samples were buffer exchanged into ddH_2O by spin exchange (VivaSpin 5000 MWCO spin column - GE Healthcare), and 100 µl aliquots at 1 mg/ml were submitted for analysis.

3: Solvatochromism of DC retinoids and development of a competitive binding assay

3.1. Introduction

Retinoid signalling, and its potentially far reaching consequences for cellular development, has focused research efforts into influencing the binding of retinoic acid and related compounds to the signalling proteins that ultimately influence phenotype. The design and synthesis of new retinoid-like molecules has been a strong theme of the drug development process in several areas and has generated demand for the probing of protein-ligand interactions at a molecular level. (Chisholm *et al.*, 2019, Chisholm *et al.*, 2016, Haffez *et al.*, 2017, Yamada *et al.*, 2019) Understanding of these interactions, and the concurrent development of new technologies, has aided in comprehending the full extent of the signalling pathway, nucleating the development of new and potent effectors of gene regulation.

In particular, the development of small, synthetic retinoid mimics has created an opportunity for the creation of a comprehensive toolkit for studying the behaviour of retinoid signalling proteins *in vivo* and *in vitro*. Key to this are the promising solvatochromic effects displayed by the DC series of retinoids, specifically their differing fluorescence response in polar *vs* apolar solvents. This behaviour suggests that they will act as an indicator of ligand binding site occupancy when incorporated into a suitable hydrophobic environment, such as in proteins capable of retinoid binding (CRAB as well as RAR/RXR proteins). (Chisholm *et al.*, 2019) Demonstrating that a suitable compound can indicate the occupancy of the ligand binding site through differentiation between fluorescence in the binding site *vs* the bulk solvent, it will be possible to probe the activity of a second compound directly without relying on second-hand measurement of genetic output or FRET. The implementation of this effect in a large scale format, and advantages offered over other methods (1.4.1 to 1.4.3), will result in a new high-throughput technique suitable for determining the binding of candidate molecules to retinoid binding protein in a rapid and effective manner.

3.2. Results

3.2.1. Determination of binding affinities for DC series retinoids

In order to assess the binding ability of the DC series retinoids for CRABPII, and make a determination of which would be most suitable going forward for competitive binding studies, the binding affinity of each compound for the protein was determined using cuvette fluorimetry. A titration of recombinantly expressed CRABPII protein was carried out against a 50 nM solution of each ligand, expected to be a suitable concentration for binding affinities based on those of the endogenous ligands, including ATRA (K_d = 14.2 nM by intrinsic fluorescence). (Redfern & Wilson, 1993) Fluorescence emission data were collected and subject to background subtraction in order to determine the level of fluorescence related to incorporation of the ligands into the binding site of CRABPII.

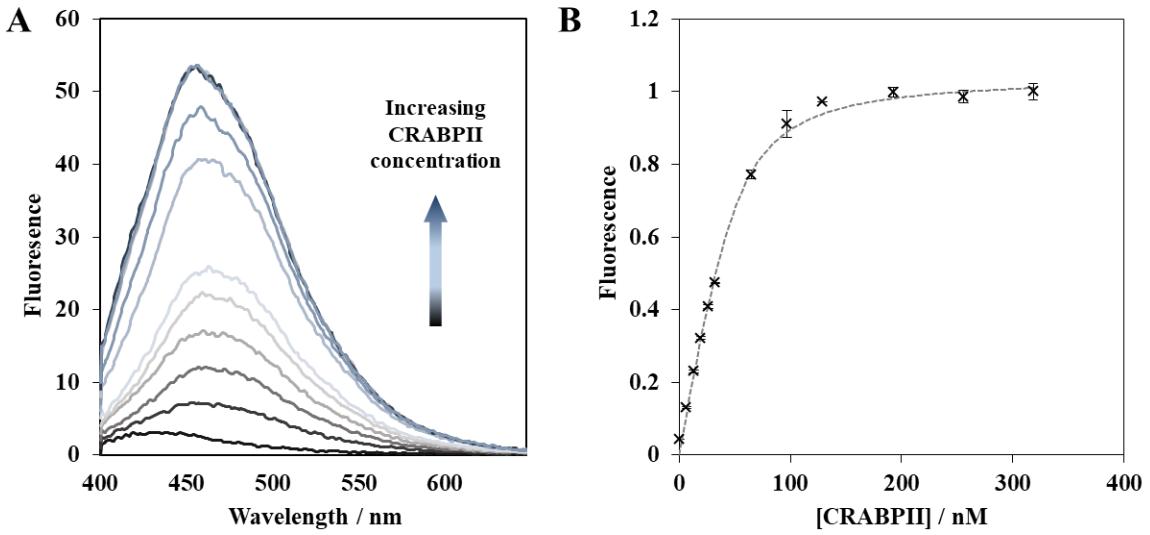


Figure 3.1 - A - Fluorescence emission curve (excitation 340 nm) for cuvette titration of CRABPII (0 - 300 nM) into 50 nM DC360. B - Isotherm at 455 nm for CRABPII/DC360 titration (n = 2) with curve fit using Dynafit.

The solvatochromic effect, demonstrated clearly in Figure 3.1 - A, creates an increasing fluorescence signal as the ligand is incorporated into the binding site of CRABPII. Whilst the concentration of ligand remains the same, the availability of CRABPII binding sites increases and ligand moves from the bulk solvent to the hydrophobic pocket. By taking an isotherm from the binding data at the fluorescence emission maxima of 455 nm, a binding curve which reaches saturation can be plotted (Figure 3.1 - B). Fitting of this curve gives an estimation of the binding constant, and was carried out using a least-squares fitting through the Dynafit software (Kuzmič, 1996). Dynafit uses a simply expressed reaction mechanism to create a set of rate equations that can be minimised to calculate the K_d. Included in Figure 3.2 are K_d values calculated using Dynafit for binding of fluorescent DC retinoids to CRABPII.

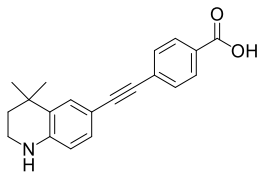
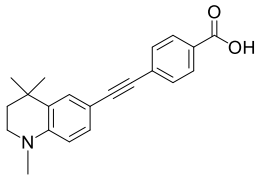
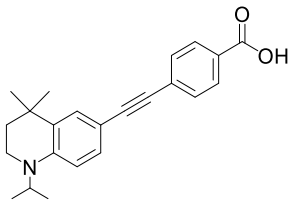
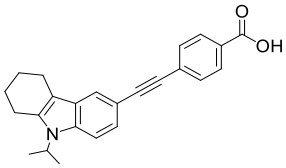
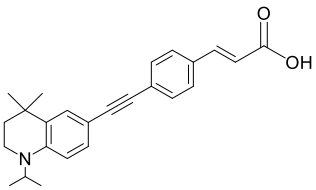
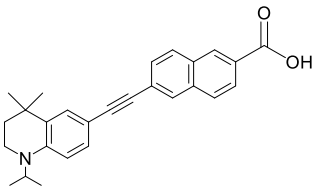
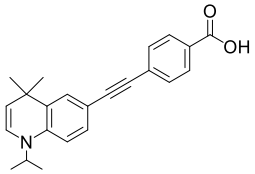
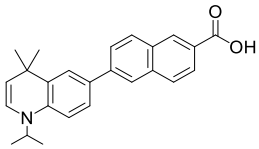
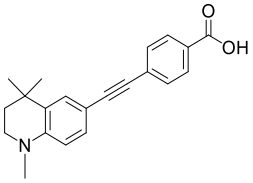
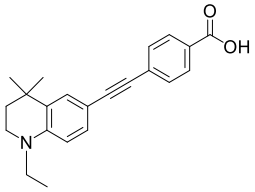
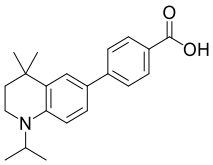
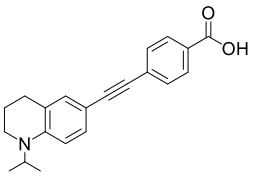
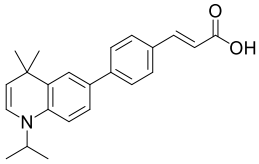
 **DC122** **139 ± 19 nM**	 **DC128** **80 ± 21 nM**	 **DC271** **32.3 ± 8.6 nM**
 **DC318** **235 ± 35 nM**	 **DC324** **no binding detected**	 **DC329** **110.6 ± 9.2 nM**
 **DC360** **9.6 ± 1.4 nM**	 **DC375** **51 ± 12 nM**	 **DC440** **216 ± 43 nM**
 **DC444** **166 ± 47 nM**	 **DC461** **556 ± 64 nM**	 **DC476** **106 ± 21 nM**
 **DC479** **0.031 ± 0.44 nM**		

Figure 3.2: DC series retinoid ligands; with chemical structures and CRABPII binding constants (nM) from least squares regression (Dynafit).

To corroborate DC271 and DC360 binding to CRABPII, co-crystallisation was carried out to determine how the fluorescent ligands are incorporated into the binding site; the molecular basis of binding is determined and discussed in 4.2. Incorporation of the ligand into the binding site in the same manner as the natural ligand, demonstrated by crystallography, as well as the high solvatochromic fluorescence seen in cuvette testing, makes DC271 an excellent candidate for competitive binding assays.

3.2.2. Determination of CRABPI-L29C-DC271 binding affinity

CRABPI-L29C, discussed in greater detail in 4.2.6, 4.2.7, and 5.2.2, was the primary CRABPI construct used for crystallography as well as competitive binding studies with selected ligands. To improve the speed and precision with which the binding affinity of the protein could be assessed, efforts were made to make the determination using a 96-well plate assay. This approach reduced the amount of retinoid compound, protein, and equipment time required, and gave a result in greater multiplicity than could reasonably be achieved using the cuvette technique.

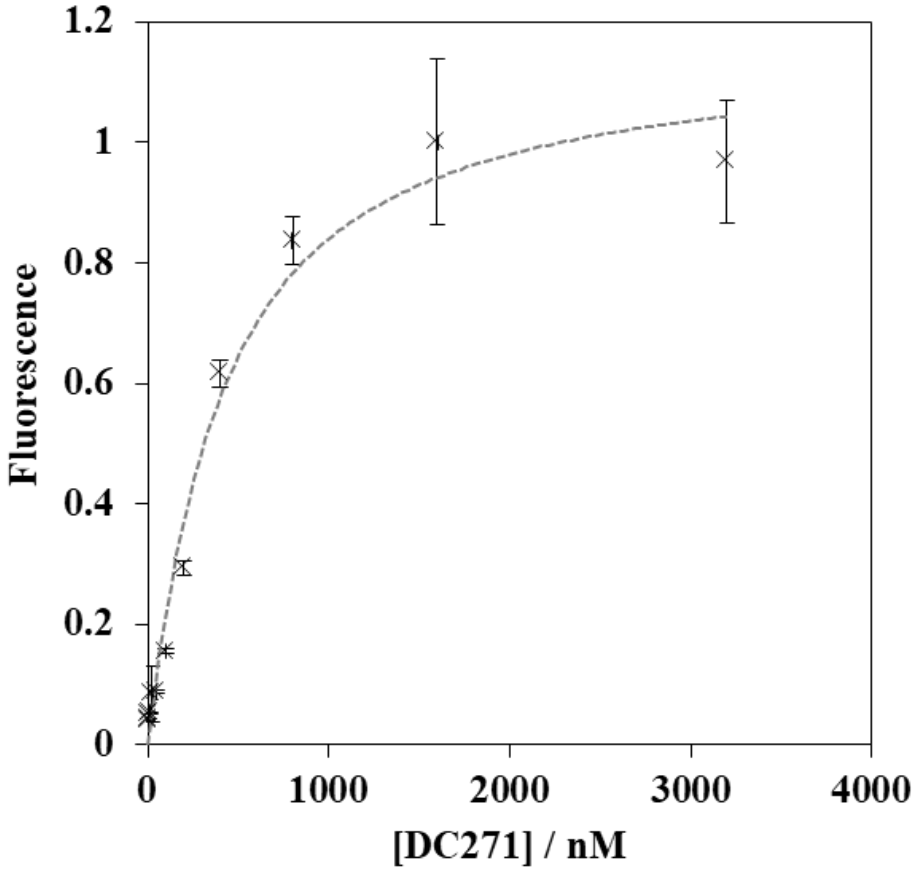


Figure 3.3 - The binding of CRABPI-L29C to the fluorescent retinoid DC271 was determined by 96-well plate fluorescence assay, K_d = 356±70 nM, normalized fluorescence response (n = 3, ± standard deviation).

Shown in Figure 3.3 is the resulting binding curve for the interaction between CRABPI-L29C and DC271 at a range of concentrations. Using the 96-well plate method, described in section 3.4.2, and subsequent curve fitting, a K_d value of 356±70 nM was determined for the pair, approximately tenfold weaker than that detected for DC271 binding to CRABPII.

3.2.3. Development of a competitive binding assay

The unique solvatochromic fluorescence displayed by DC271 (and associated compounds) creates an opportunity to measure the binding of a second compound by means of displacement, and hence was used as the indicative probe in the development of a solvatochromic fluorescence displacement assay. Only previously achieved using gene-product assays, radiolabelling techniques, or close contact FRET based methods, the direct measurement of binding efficacy will allow for rapid, high-throughput assaying with minimal material requirements. Only a source of the test protein, the desired test compounds, fluorescent retinoid used as an indicator of binding site occupancy, and a fluorescence 96-well plate reader are required. This method has been developed for use on CRABPII, and can readily be applied to close cousin CRABPI, though offers far greater potential in differentiating between binding of compounds to the nuclear receptors (RARs - see 5.2.3). Any protein with a retinoid binding site that is suitably hydrophobic could be assayed using the technique, providing a fluorescent derivative retinoid can found to suit its particular binding site. In this case, DC271 was chosen for its relatively tight binding affinity for CRABPII, and because it displayed the most intense fluorescence signal under the titration conditions.

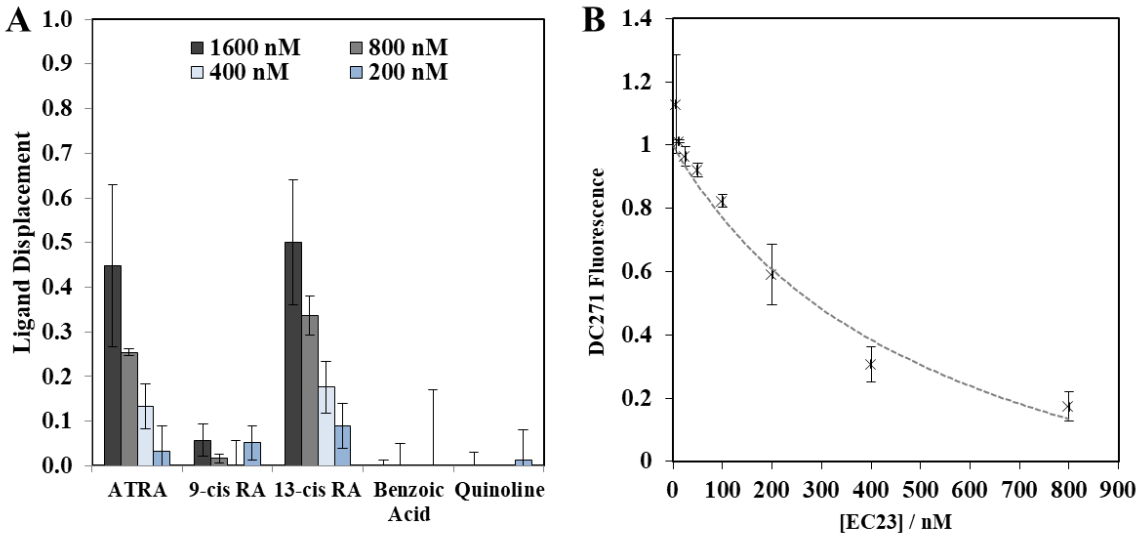


Figure 3.4: A - Displacement of DC271 from the binding site of CRABPII by select retinoids and related structural components (n=3, α=0.05), designed to indicate qualitatively whether each compound is capable of binding to CRABPII. B - Quantitative measurement of DC271 fluorescence to indicate ligand displacement by EC23 synthetic retinoid (n=3, α=0.05). Curve fit to a competitive binding equation using DynaFit19. (Kuzmič, 1996)

The method was initially designed based on the effective concentration ranges determined in the fluorescence cuvette assays, working in the low nanomolar range for concentrations of both protein and solvatochromic retinoid. Similarly, the wavelengths selected were based upon the absorption/emission maxima of DC271. To demonstrate the effectiveness of the assay, a selection of endogenous retinoid ligands and synthesis fragments from synthetic retinoids were tested for binding (Figure 3.4 - A). Carrying out the study at multiple concentrations gave a qualitative

indication of binding efficacy, but does not produce a full binding curve that can be fit to give a K_d value. This first methodology was intended to simulate a rapid, high-throughput screening method, as might be carried out with a large commercial screening library, to be followed up by detailed study of any ligands that indicated binding. A second experiment using EC23 was designed to simulate this process, capturing the full binding curve, and was subsequently fit to give a K_d value of 160 nM, a mid-range binding efficiency compared to other known ligands. (Redfern & Wilson, 1993) Fitting was carried out using a competitive binding simulation in the Dynafit software, and minimizing least squares to produce a K_d value.

Across both methodologies a Z' value of greater than 0.6, indicating the statistical "width" between minimum and maximum values, was routinely achieved. This key indicator suggests that the assay is valid, and any results are likely to stand prominent above the background noise. (Zhang, Chung, *et al.*, 1999) The data collected as part of the testing process indicates that ATRA and 13-*cis* retinoic acid bind to CRABPII as measured by displacement of the fluorescent ligand. Conversely, 9-*cis* retinoic acid, with its bent shape a poor fit for the ligand binding site, demonstrates comparatively poor binding and displacement of DC271. The synthesis fragments quinolone and benzoic acid also had no measurable effect, which was expected given the depth of the ligand binding site and multiple interactions normally required to create a good binding.

The final protocol decided upon after the test experiments, and used in future binding determination for CRABPII, CRABPI-L29C and RAR proteins, is described fully in Appendix 8.

3.3. Discussion

All of the DC series compounds, with the exception of DC324, tested for solvatochromic binding in CRABPII proved competent, justifying their design as compounds closely related to retinoids ATRA and EC23. DC479, and closely related compound pair DC360 and DC271, were the compounds with the tightest experimentally determined affinity. From the data presented in Figure 3.1 it can be inferred that compound length, oxidation status of the di/tetrahydroquinoline group and substituents made to that same group all cause changes in binding efficacy. DC324 (elongated with an additional carbon-carbon double bond) and DC461 (shortened by removal of the central alkyne bond) demonstrate that neither deviation from the core EC23 structure is favoured. DC324 does not bind CRABPII at all, or shows no solvatochromism if it does, whilst DC461 has a binding affinity two orders of magnitude worse than the tightest binders. With a smaller distinction, it is apparent that the dihydroquinoline (DHQ) group of DC360 creates a slight, yet significant, improvement in binding capability *vs* that of the more flexible tetrahydroquinoline (THQ) in DC271. Furthermore, substituent groups to the DHQ/THQ section also seem to impact the binding efficiency. DC271, with a gem-dimethyl substituent at the para position to the DHQ-nitrogen atom, has a binding constant nearly twice improved over that of DC476, where the group is absent. All of these factors affecting binding efficiency are logical, and likely relate closely to the behaviour of the compounds in the deep, hydrophobic binding pocket of the CRAB proteins. Similarly, the preference for CRABPII over CRABPI-L29C displayed by DC271 is likely to be due to a specific binding interaction not replicated between the two otherwise closely related proteins. DC479 stands out as the tightest binding compound, with a K_d value ostensibly lower than that of ATRA (9.95 nM *vs* 14.2 nM). (Redfern & Wilson, 1993) The overall length of DC479 is similar to that of ATRA, due to the incorporation of an additional carbon-carbon double bond in the head group and removal of the alkyne bond seen in EC23 (Figure 3.2). This change likely makes DC479 overall more flexible, in line with flexibility exhibited by ATRA, and possibly assists in creating a low energy conformation in the ligand binding site. The decision to use DC271 going forward was based on its strong binding constant for CRABPII (32.3 nM) as well as CRABPI-L29C (302 nM), intense absolute fluorescence under titration conditions, and the relative abundance in which it had been supplied.

The wholesale switch to 96-well plate-based assays improved the overall reliability of the experiment, and reduced the number of abstractions that had to be made for dilution effects. In addition, preparation of a 96-well experiment, including making dilution series and aliquoting of materials, was faster and more accurate than titration and mixing within a quartz cuvette. The method is also natively suited to expansion into larger plate sizes and theoretically the automation of pipetting tasks, to achieve a high-throughput of tested compounds from libraries or fragment series. Calculation of a binding constant for the interaction between CRABPI-L29C and DC271 was undertaken as part of an experiment to determine the binding of the protein to other retinoid

compounds through competitive binding (DC645, see 5.2) and to probe why the interactions of CRABPI-L29C differ from CRABPII given their close sequence similarity. The fact that there is a tenfold difference in the binding affinity for DC271 between the two proteins, in favour of CRABPII, suggests a difference in the binding site or other structural effect that could be identified with structural investigation.

Performing large plate-based assays requires suitable controls and measures of effectiveness, and the Z' value is the standard method of quantifying this success. (Zhang, Chung, *et al.*, 1999) By measuring the width between the minimum and maximum fluorescence achieved in the assay, and considering the error associated with each, the likelihood that a test value will be distinguishable above the baseline can be expressed as a decimal value. A Z' value of greater than 0.5 is considered an 'excellent assay'. (Zhang, Chung, *et al.*, 1999) The high solvatochromic fluorescence of DC271 made it an excellent candidate for competitive binding assays in combination with CRABPII or CRABPI-L29C; after accounting for background, CRABPII-DC271 fluorescence routinely gave a Z' value of greater than 0.7. Special controls were taken into account when a test molecule had some measure of fluorescent activity (see 5.2) which kept the separation between minima and maxima, and the corresponding Z' value, high.

The data presented in 3.2.3 shows how the assay technique can be applied in both qualitative and quantitative modes, suitable for both initial screening of large numbers of compounds and more focused collection of binding data for compounds with known (or suspected) retinoid activity. Comparison to known CRABPII binders (ATRA and 13-*cis*-RA) demonstrates that the effect being measured is a genuine displacement of the fluorescent retinoid from the binding site by the competitor ligand, and that it cannot be falsely replicated (false positive) by non-binders such as 9-*cis*-RA. Additionally, fragments selected from the syntheses of the DC retinoid series (quinolone and benzoic acid) are incapable of displacing DC271 at relevant concentrations, demonstrating that the binding of the compound is genuine and based on holistic interactions between the entire molecule and the ligand binding pocket. Unlike other methods, measurement of displacement as a function of solvatochromic fluorescence is a direct indicator for ligand binding site occupancy and does not require separation of unbound probes as with radiolabelling techniques. (Allenby *et al.*, 1993, Levin *et al.*, 1992) Solvatochromic fluorescence displacement has further been used to determine binding constants for non-fluorescent, synthetic retinoid derivatives to CRABPII as well as both RARα and RARγ isoforms (see 5.2.3).

3.4. Methods

3.4.1. Fluorescence cuvette titrations

Three cuvettes (quartz, 3 ml, 10 mm path) were prepared with DC retinoid (50 nM in PBS, from 100% EtOH stock solution) and control buffer (PBS only) solutions. Blank fluorescence emission spectra (400-650 nm, excitation 340 nm, ex/em slit 5/10 nm, scan speed = 200 nm/min) were recorded before titration of CRABPII solution (12.5 μM) into each cuvette. Following addition of CRABPII, each solution was stirred for 5 min before recording spectra. Titration was ceased when a plateau was reached.

Data processing was carried out to subtract background fluorescence of CRABPII and DC271, and to correct for the changing concentration of CRABPII during the titration; spectra and an isotherm at the fluorescence emission maximum were taken to investigate the binding efficiency of each compound.

Data analysis was carried out by least squares fitting of a simple protein-ligand interaction (P + L ⇋ P.L) event using the Dynafit software, determining a binding affinity (K_d) for each compound. (Kuzmič, 1996)

3.4.2. CRABPI-L29C DC271 Affinity determination (96 well plate)

Solutions of DC271 (9.6 μM - 4.68 nM, <1% EtOH) and CRABPI-L29C (150 nM in 20 mM HEPES, 300 mM NaCl, pH 7.5) were prepared and combined (50 μl + 100 μl) in a Corning, black non-binding surface (NBS) fluorescence plate (96-well), using HEPES buffer in place of CRABPI-L29C when necessary for controls. The plate was spun for 2 min (1500 rpm) to ensure incorporation, and fluorescence read at excitation/emission 355/440 nm using a Synergy H4 plate reader. The total volume per well was 150 μl, the final concentration of protein was 100 nM.

Data analysis was carried out as per 3.4.1.

3.4.3. CRABPII assay validation

Solutions of DC271 (300 nM, <1% EtOH) and CRABPII (300 nM, 20 mM K_2HPO_4, 100 mM KCl, pH 7.4 + 0.05% Pluronic F127) were prepared and mixed before use. A Corning black, NBS fluorescence plate (96-well) was cleaned, and Max and Mid wells were loaded with 60 μl CRABPII/DC271 solution. Min wells were loaded with 60 μl control solution of DC271 mixed with assay buffer. Mid wells were then loaded with 30 μl EC23 (600 nM in 1% EtOH), and Max and Min wells with 1% EtOH solution (30 μl). Plates were spun for 2 min (1500 rpm) and read using a Synergy H4 plate reader, at ex/em 355/460 nm. The total well volume was 90 μl, the CRABPII and DC271 concentration was 100 nM.

3.4.4. CRABPII competitive binding assay

Solutions of DC271 (300 nM, <1% EtOH) and CRABPII (300 nM, 20 mM K_2HPO_4, 100 mM KCl, pH 7.4 + 0.05% Pluronic F127) were prepared and combined before use. A Corning black, NBS fluorescence plate (96-well) was cleaned using compressed air, and maximum signal (Max), mid-inhibition (Mid), and all test compound wells were loaded with 60 µl CRABPII/DC271 solution. Minimum signal wells (Min) were loaded with 60 µl control solution of DC271 mixed with assay buffer. Mid wells were then loaded with 30 µl EC23 (600 nM, < 1% EtOH), and Max and Min wells with 1% EtOH solution (30 µl). Dilution series of test compounds were prepared and aliquoted to the plate (30 µl). Plates were spun for 2 min (1500 rpm) to ensure mixing and read using a Synergy H4 plate reader (ex/em 355/460 nm). The total well volume was 90 µl, the on plate concentration of CRABPII and DC271 was 100 nM. Later experiments increased well volume to 150 µl, combining 50 µl each CRABPII, DC271 and test compound.

Data were processed to subtract background fluorescence of CRABPII and DC271. Data analysis was carried out using the Dynafit software (Kuzmič, 1996), minimising least squares to fit a competitive binding event in which two ligands compete for the same binding site (P + L1 ⇋ P.L1 *vs* P + L2 ⇋ P.L2). This fitting calculates a K_d value for the competing interaction, based on a known K_d value for the CRABPII-DC271 interaction.

4: Structural biology - binding of retinoids to CRABPI and CRABPII

4.1. Introduction

The ability to determine the atomic structure of proteins and their ligands, and to recognise interactions between the two, has been a core component in understanding the mechanics of life and the subsequent development of pharmaceuticals for the past seventy five years. Advancements in automation (O'Hea *et al.*, 2018), robotics, the development of high frequency pixel detectors (Broennimann *et al.*, 2006, Henrich *et al.*, 2009, Casanas *et al.*, 2016), and provision of automatic data processing (Winter, 2010) has made the crystallographer's job ever faster and more accessible. Whilst cryo-electron microscopy matures and develops at a prodigious rate, and complementary techniques including NMR structural interpretation are available, X-ray crystallography remains pre-eminent in the determination of atomic resolution protein structures and fine detail ligand interaction.

Understanding how retinoids interact with their carrier and receptor proteins is at the centre of determining how these processes can be manipulated for therapeutic benefits. The first structure of ATRA-bound CRABPII was solved using molecular replacement with structures of semi-related fatty acid binding proteins, retinol binding protein and myelin P2. (Kleywegt *et al.*, 1994) In 1998, CRABPII was crystallised for the first time with an unoccupied binding site using an R111M mutant of the wild type (Chen *et al.*, 1998), followed by determination of the wild type structure at 1.35 Å in 2006. (Vaezeslami *et al.*, 2006) These structural determinations demonstrate a consistently globular protein based around an orthogonal β-sandwich, capped with a pair of α-helices that form part of the ligand binding site which does not display any large, ligand-induced conformational changes (see Figure 4.1 to Figure 4.9). This wealth of structural data has allowed for the determination of a pseudo nuclear localisation sequence (Sessler & Noy, 2005) which permits the protein to enter the nucleus once retinoid bound, as well as the interaction surface necessary for CRABPII and RARα interaction in the nucleus. (Delva *et al.*, 1999, Budhu *et al.*, 2001)

Determination of crystal structures of CRABPII bound to synthetic retinoids will demonstrate correct incorporation of the molecules into the binding site, indicating a retinoid-like mode of action *in vivo*, and verify that the solvatochromic effect seen and utilised for fluorescence binding assays is based on true ligand binding and not an off-site or other hydrophobic effect. Further structural data may also aid in future understanding of how the retinoid signalling pathway works, with particular focus on the differential between genomic and non-genomic signalling (Persaud *et al.*, 2013, Park *et al.*, 2019), and the predicted but not structurally characterised interaction with RAR/RXR proteins. (Budhu *et al.*, 2001, Delva *et al.*, 1999, Dong *et al.*, 1999)

4.2. Results

The crystallography and structural determination of CRABPII and CRABPI-L29C is discussed below; both proteins adopted the same, well-known globular structure previously discussed, with minor differences based on ligand binding and incorporation. Further details of data collection and refinement are contained in the methods section and crystallographic tables (Appendix 9 and Appendix 10). Omit (ligand absent) and polder (ligand absent and solvent excluded) maps verifying ligand density for all CRABPII and CRABPI-L29C structures can be found in Appendix 14 and Appendix 15, respectively. Equations used to calculate R-factors are included in Appendix 17.

4.2.1. Crystal structure of CRABPII-DC360

CRABPII, expressed and purified recombinantly in *E. coli*, crystallised reliably in sodium sulphate-based conditions and incorporated ligands proficiently by combining the two in equimolar amounts prior to drop setting. The data for solution of CRABPII-DC360 was provided based on a previous project (Chisholm, 2017), and was processed and refined to the level presented here.

Crystallisation of CRABPII with synthetic retinoid DC360 was carried out in part to demonstrate that the ligand occupies the hydrophobic binding site in the canonical manner. The structure was determined at 1.8 Å, refined to a final R/R_{free} of 0.17/0.24. This forms key evidence that the solvatochromic effect used for the determination of binding (3.2.1) is genuine and based on incorporation of the ligand into the binding site, rather than an external hydrophobic region or as part of a larger complex formation.

The ligand, designed to mimic ATRA and based on the EC23 framework, incorporates a carboxylic acid head group in the "all-*trans*" orientation, and a hydrophobic aromatic body centred on an alkyne bond. The head group is buried in the hydrophobic tunnel that makes up the ligand binding site of CRABPII, and interacts with the triad of residues (R112, R133, Y135) and conserved water molecule typically found hydrogen bonded to retinoid ligands (Figure 4.1 - A). (Kleywegt *et al.*, 1994) The incorporation of the ligand creates a notable inversion of R29 (Figure 4.1 - white), generating the NLS required for passage to the nucleus. (Sessler & Noy, 2005) This key structural change suggests that DC360 would be trafficked to the nucleus *in vivo.*

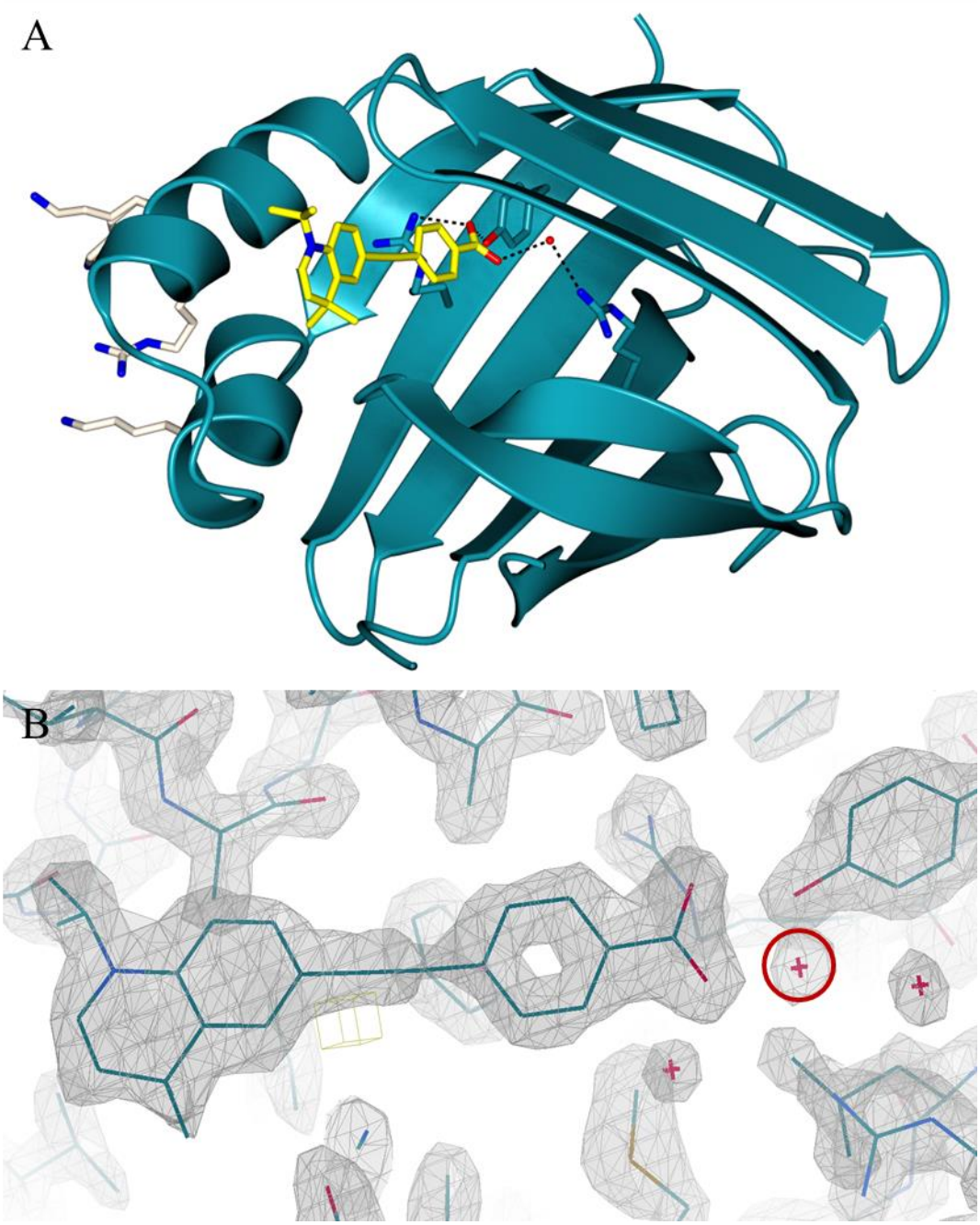


Figure 4.1: A - Crystal structure of CRABPII (dark blue, ribbon diagram) with synthetic retinoid DC360 (yellow, cylinder form); binding triad (dark blue, O-red, N-blue, cylinder form), conserved water molecule (red, sphere) and nuclear localisation sequence (white, cylinder form) are highlighted. B - 2Fo-Fc map of DC360 in the ligand binding site, from refinement including the ligand in the model (contour at σ = 1.6). Conserved water molecule is highlighted (red circle).

4.2.2. Crystal structure of CRABPII-DC271

DC271 was selected for co-crystallisation with CRABPII to demonstrate correct incorporation into the binding site, after the molecule was chosen as the key compound in fluorescent displacement assays.

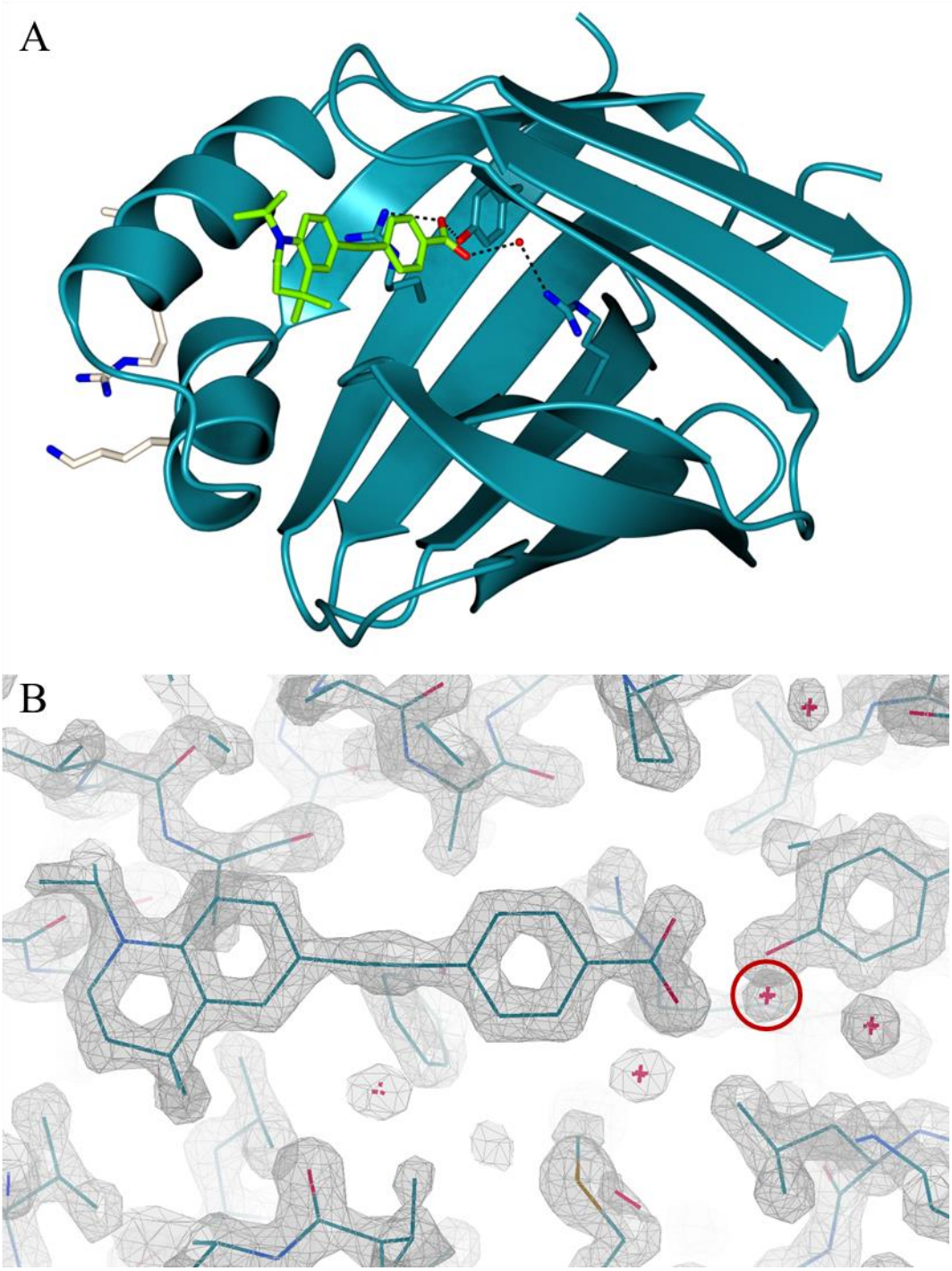


Figure 4.2: A - Crystal structure of CRABPII (dark blue, ribbon diagram) with synthetic retinoid DC271 (green, cylinder form), binding triad (dark blue, O-red, N-blue, cylinder form) and nuclear localisation sequence (white, cylinder form) are highlighted. B - 2Fo-Fc map density of DC271 in the ligand binding site, from refinement calculations including the ligand (contour at $\sigma = 2.0$). Conserved water molecule is highlighted (red circle).

Based closely on the structure of DC360, DC271 incorporates a tetrahydroquinoline tail group rather than the dihydroquinoline used in the former. This difference imparts slightly more flexibility to the DC271 tail group, allowing minor puckering of the ring not seen in DC360, at a cost to the overall binding affinity (3.2.1). The high resolution (1.5 Å, final R/Rf$_{ree}$ 0.14/0.17) of the structure shows extremely well resolved electron density, particularly around the ligand (Figure 4.2 - B), which suggests a very rigid conformation for the molecule once incorporated into the binding site. Interactions with the ligand binding triad follow the canonical pattern, incorporating the conserved water molecule that forms part of the hydrogen bonding network. In the same manner as DC360, DC271 creates a viable NLS (Figure 4.2 - white) external to the helical cap.

4.2.3. Crystal structure of CRABPII-EC19

EC19 was designed to mimic 13-*cis* retinoic acid, bent at the head group to form a distorted 'L' shape. Whilst 13-*cis* retinoic acid is known to bind to CRABPII, and was demonstrated earlier to do so by fluorescence displacement, it is prone to isomerisation under laboratory conditions, and so can prove difficult to work with *in vitro* and *in vivo.* EC19 was designed to improve upon these weaknesses, and was shown to cause differentiation of stem cell cultures into epithelial tissues, rather than the neuronal tissues developed with ATRA-mimic EC23. (Christie *et al.*, 2008) The fact that EC19 and EC23 cannot easily interconvert in solution likely enhances this effect, above what would be expected of a 13-*cis* retinoic acid solution prone to isomerization. To investigate how EC19 interacts with CRABPII, the pair were co-crystallised, yielding a structure at 2.8 Å (Figure 4.3) in a condition of 30% PEG 3350, 0.1 M Tris-HCl and 0.2 M sodium acetate. The final R/R$_{free}$ was 0.23/0.31.

Table 4.1: Key data processing statistics from Aimless/Pointless (Evans & Murshudov, 2013) for two data sets collected on CRABPII-EC19. Outer shell values shown in parentheses.

Dataset	Unit cell volume (Å^3)	No. of molecules	R_{meas}	R_{pim}	I/σ(I)	CC½
$P2_13$	483736.6	1	0.086 (0.333)	0.016 (0.061)	33.6 (11.0)	0.999 (0.987)
H3	1644533.0	4	0.684 (5.126)	0.215 (1.666)	4.8 (1.0)	0.977 (0.265)

The structure of CRABPII-EC19 presents a cubic unit cell (a=b=c=78.5 Å, α=β=γ=90.0°) in space group $P2_13$, though a second data set made the case for a H-centred trigonal (H3) setting (a=b=110.33 Å, c=135.10 Å, α=β=90°, γ=120°). Both solutions had fair initial R/R$_{free}$ of around 0.3, but $P2_13$ was selected for its higher symmetry and improved R_{meas}/R_{pim}, I/σ(I), and CC½ values (Table 4.1). When considering non-crystallographic symmetry, the H3 setting displayed the same overall symmetry as $P2_13$, but with four molecules in the asymmetric unit rather than one. It is likely that the observed difference in space group was a result of radiation damage to the crystal

during data acquisition. Comparison of unit cells and crystal packing for both solutions are found in Appendix 11 ($P2_13$), Appendix 12 (H3) and Appendix 13 (comparison).

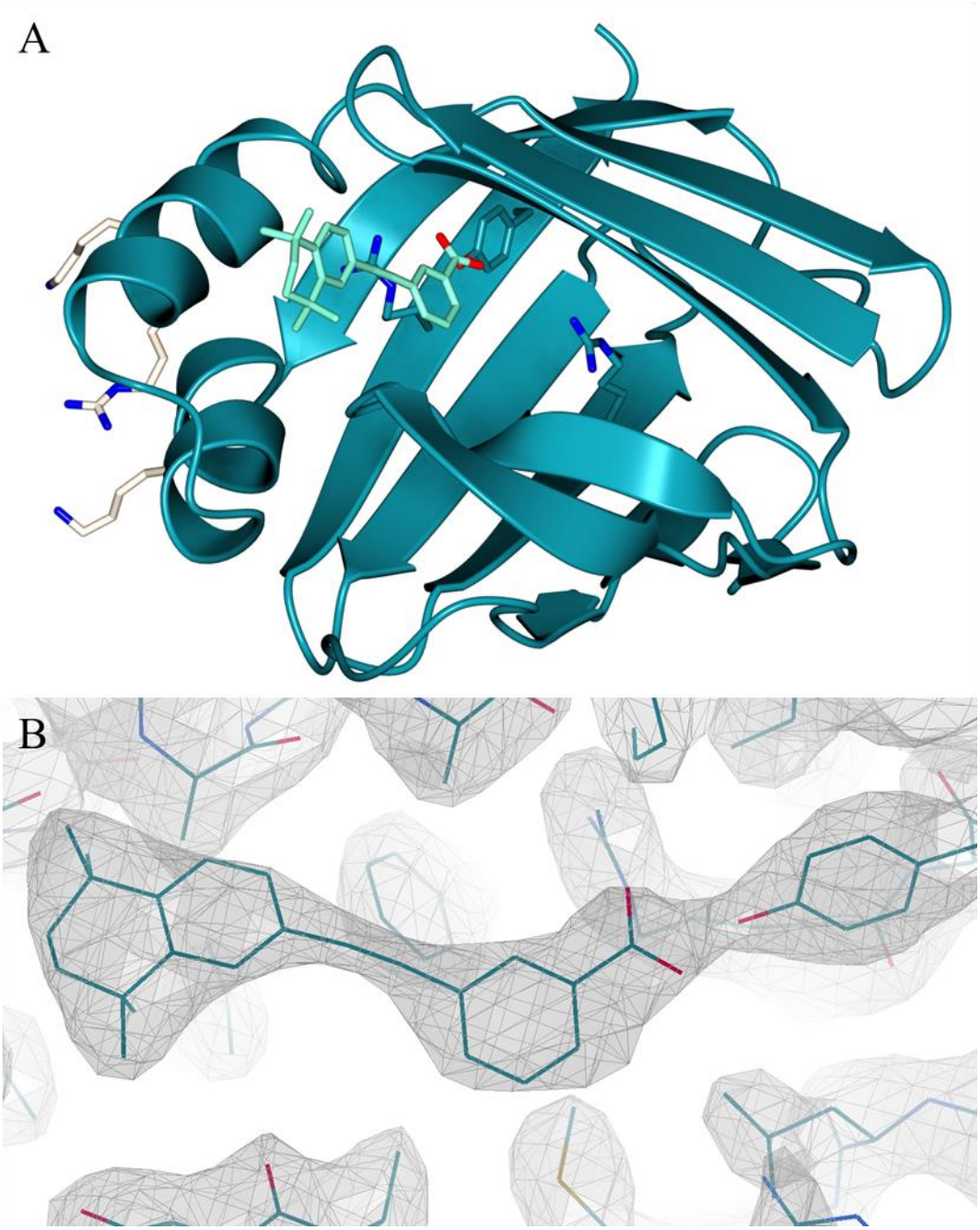


Figure 4.3: A - Crystal structure of CRABPII (dark blue, ribbon diagram) with EC19 (aqua, cylinder form); binding triad (dark blue, O-red, N-blue, cylinder form) and nuclear localisation sequence (white, cylinder form) are highlighted. B - 2Fo-Fc map density of EC19 in the ligand binding site, from refinement calculations including the ligand (contour at $\sigma = 1.0$).

The final structure (Figure 4.3) comprises a single molecule with EC19 visibly incorporated into the binding site, albeit without recognisable hydrogen bonding interactions with binding triad

residues (occupying conserved positions). The *para*- positioning of the carboxylic acid group prevents it from interacting with R112, and whilst there is space for a conserved water molecule it is not resolvable at the resolution of the data. The overall geometry of the site prevents adoption of an alternate conformation with better access for the carboxylic acid group. In particular, hydrophobic residues F16, A37, and V77 create a space only 8 Å wide around the benzoic acid head group. With an approximate 3 Å clearance between the ligand and each residue, there is limited space for reorientation to better accommodate the *para*-benzoic acid in the binding site.

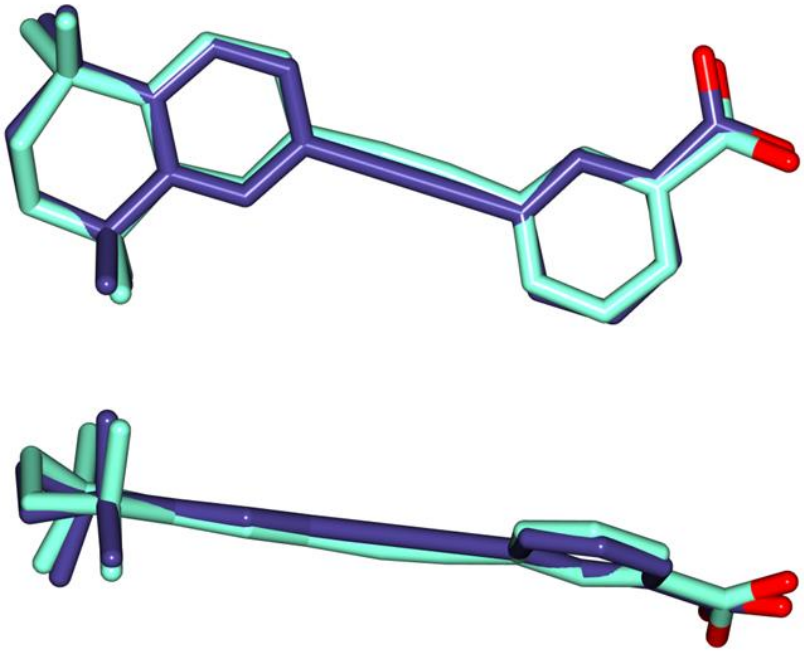


Figure 4.4: Front and side view of EC19 from CRABPII-EC19 (aqua, cylinder form) aligned with energy minimised EC19 structure (purple, cylinder form). Oxygen atoms are coloured red. Minimisation carried out in Avogadro, using the general Amber force field. (Wang *et al.*, 2004, Hanwell *et al.*, 2012)

As a result of the close confines, EC19 adopts a visibly strained position in the binding site, and comparison to an energy minimised ligand structure confirms this (Figure 4.4). The deviation from 180° bond angles visible in the alkyne linker of EC19 highlights that the binding position is not 'comfortable'; either the data represents averaging of the molecule across two straightened positions or the ligand is forced into that unusual conformation by the geometry of the hydrophobic pocket and binding triad. It is possible that the limited resolution of this structure relative to other CRABPII structures is due to disorder generated by EC19, and the binding irregularities it generates. Nevertheless, the positioning of the ligand does generate a NLS (Figure 4.3 - white) and therefore EC19 would likely be ported to the nucleus.

4.2.4. Crystal structure of CRABPII-DC479

DC479 was selected from the DC series of retinoids because of the extended carbon chain that improves the flexibility of the head group (and carboxylic acid group) of the molecule. (Chisholm *et al.*, 2019) The omission of the alkyne midsection compensates and overall the molecule is of a similar length to ATRA, with more flexibility than the fixed conformers seen in DC360 and DC271. DC479 also has a high affinity for CRABPII (see 5.2) and this could be due to conformations accessible only because of the flexible head group. DC479 has also been shown to preferentially activate non-genomic pathways over genomic signalling, demonstrated by lack of neurite outgrowth *in vitro.* (Khatib *et al.*, 2020)

The structure of CRABPII-DC479 (1.8 Å, R/R_{free} 0.19/0.25, Figure 4.5) follows previously determined CRABPII structures, including 5OGB (RMSD, Cα = 0.61 Å) which was used for molecular replacement. The binding pocket incorporates DC479, with full interaction between the carboxylic head group, the binding triad, and a conserved water molecule. Incorporation of the ligand and externalisation of the nuclear localisation sequence residues suggest that it is not CRABPII interaction that prevents genomic signalling in the case of DC479, which is discussed further below.

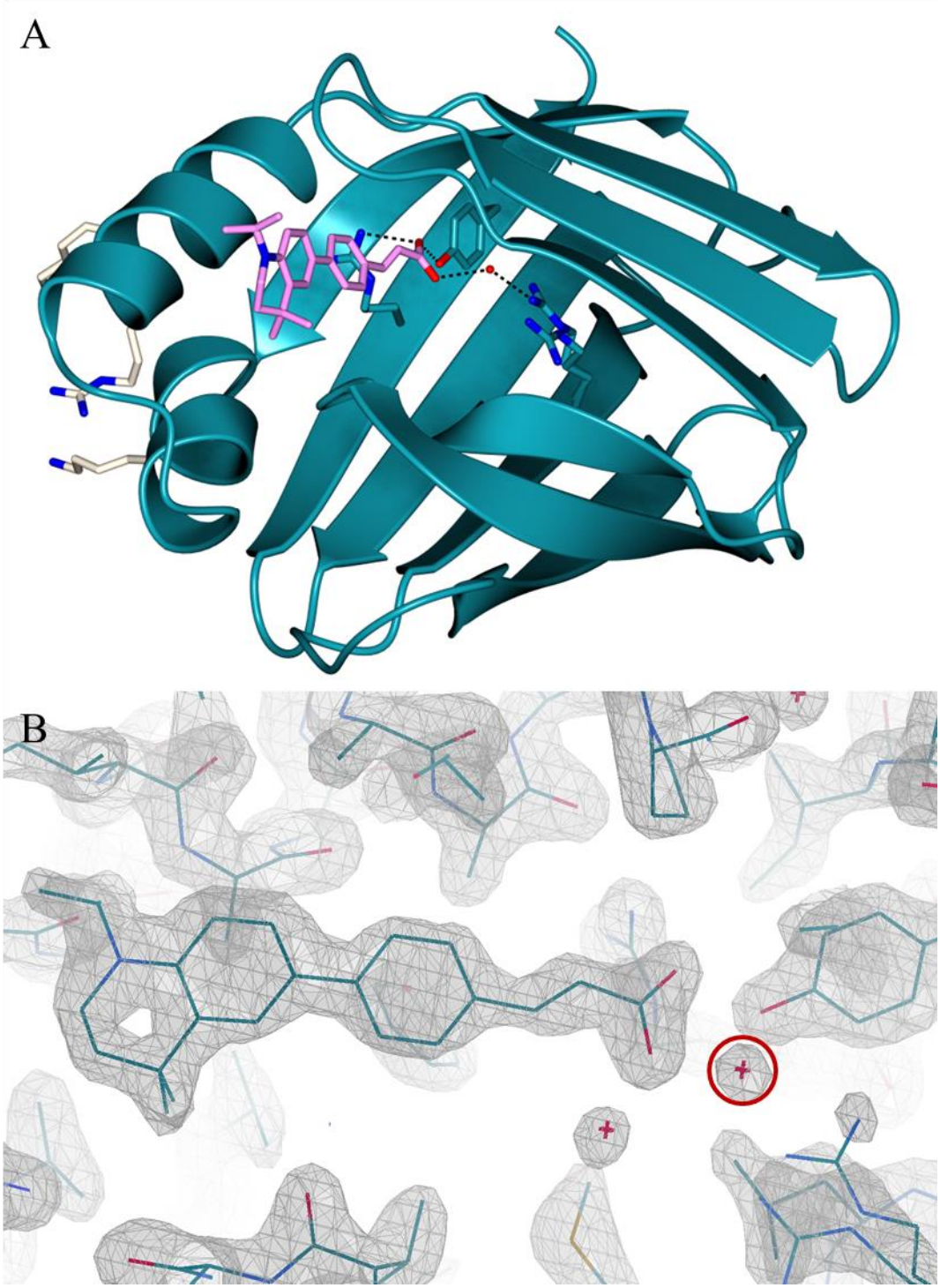


Figure 4.5: A - Crystal structure of CRABPII (dark blue, ribbon diagram) with synthetic retinoid DC479 (pink, cylinder form diagram); binding triad (dark blue, O-red, N-blue, cylinder form) and nuclear localisation sequence (white, cylinder form) are highlighted. B - 2Fo-Fc map density of DC479 in the ligand binding site, from refinement calculations including the ligand (contour at σ = 1.5). Conserved water molecule is highlighted (red circle).

4.2.5. Crystal structure of CRABPII-DC645

Non-fluorescent EC23 analogue DC645 was selected for ligand binding studies because of its beneficial outcome in studies of neurite outgrowth. (Khatib *et al.*, 2020) Incorporating a tetrahydroquinoxiline in place of the hydroquinoline or hydronapthalene structures used in solvatochromic DC compounds, DC645 only exhibits low intrinsic fluorescence emission at 440 nm, likely the tail of a sub 400 nm fluorescence peak. This structure creates increased opportunity for hydrogen bonding in the tail section, and may influence the overall interaction with CRABPII, for which it has a lower affinity than other DC retinoids (see 5.2.1).

Determined at 1.7 Å (R/R_{free} 0.18/0.20) and solved by molecular replacement (5OGB, CRABPII-DC360, RMSD Cα = 0.73 Å), the structure of CRABPII with DC645 shows complete incorporation of the retinoid into the binding site, including the expected interactions with the binding triad and conserved water molecule. Increased hydrogen bonding, to solvent and R60, can be seen around the binding pocket entrance where the tetrahydroquinoxiline tail group sits. The nuclear localisation sequence (Figure 4.6 - A, white) can be seen adopting an external position, indicating that nuclear transport would occur *in vivo*.

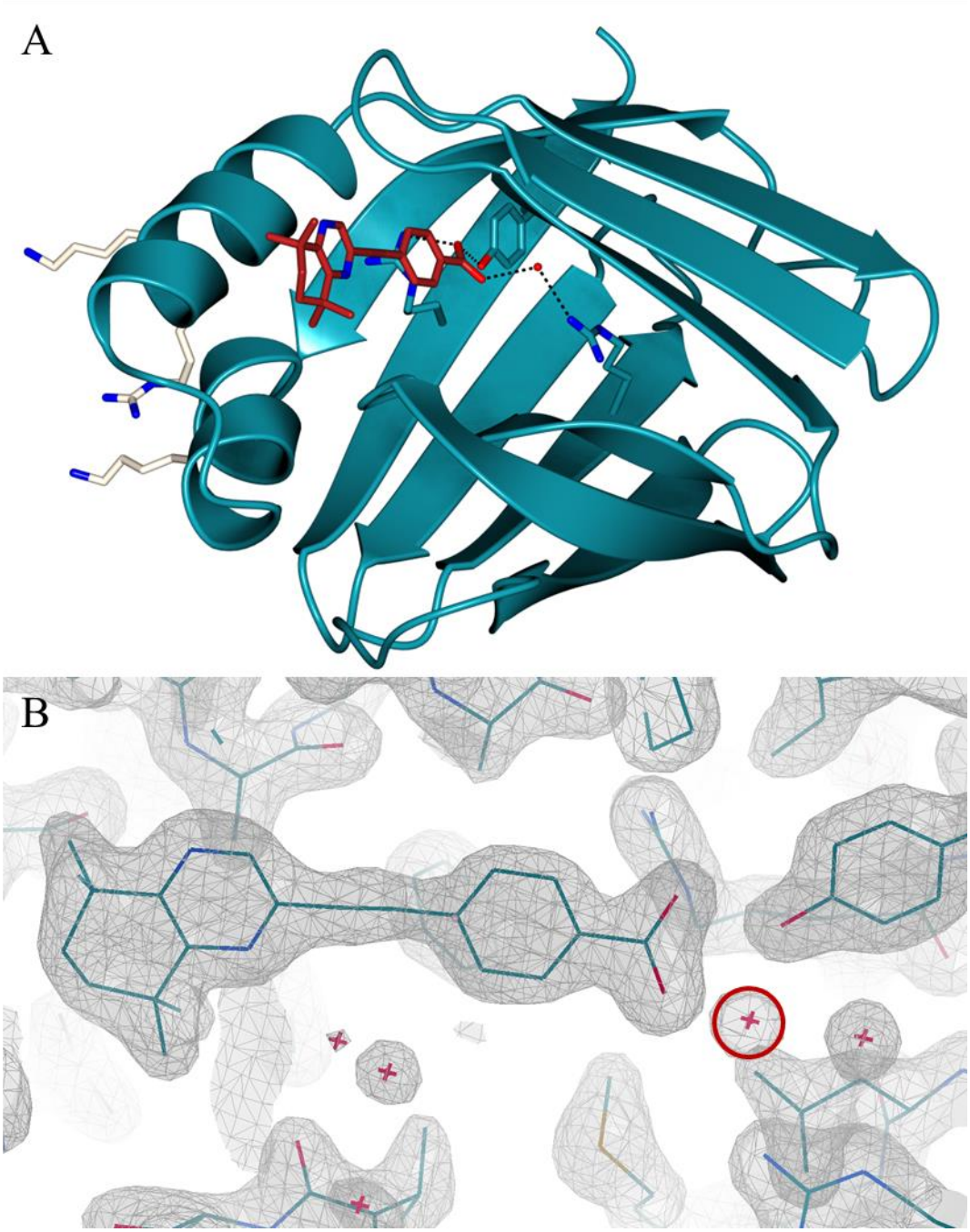


Figure 4.6: A - Crystal structure of CRABPII (dark blue, ribbon diagram) with synthetic retinoid DC645 (brick red, cylinder form); binding triad (dark blue, O-red, N-blue, cylinder form) and nuclear localisation sequence (white, cylinder form) are highlighted. B - 2Fo-Fc map density of DC645 in the ligand binding site, from refinement including the ligand in calculation (contour at σ = 1.5). Conserved water molecule is highlighted (red circle).

4.2.6. Crystal structure of CRABPI-L29C in complex with fatty acids

The first structure of CRABPI was reported in 1994 using protein from *Mus musculus* and *Bos taurus*, which share the same primary sequence, highly conserved (99% from NCBI BLAST) with that of hCRABPI. (Altschul *et al.*, 1990, Kleywegt *et al.*, 1994) Depositions of CRABPI structures currently extend only to a resolution of 2.7 Å (1CBI). (Thompson *et al.*, 1995) The hCRABPI-L29C mutant was identified as part of crystal screening as allowing more efficient crystallisation

than the wild type, and diffraction to atomic resolution; the resulting structures suggest that this is due to dimerisation at the mutated surface residue cysteine 29. The modified residue is found on the exterior of the helical cap and as such does not modify the environment of the ligand binding pocket.

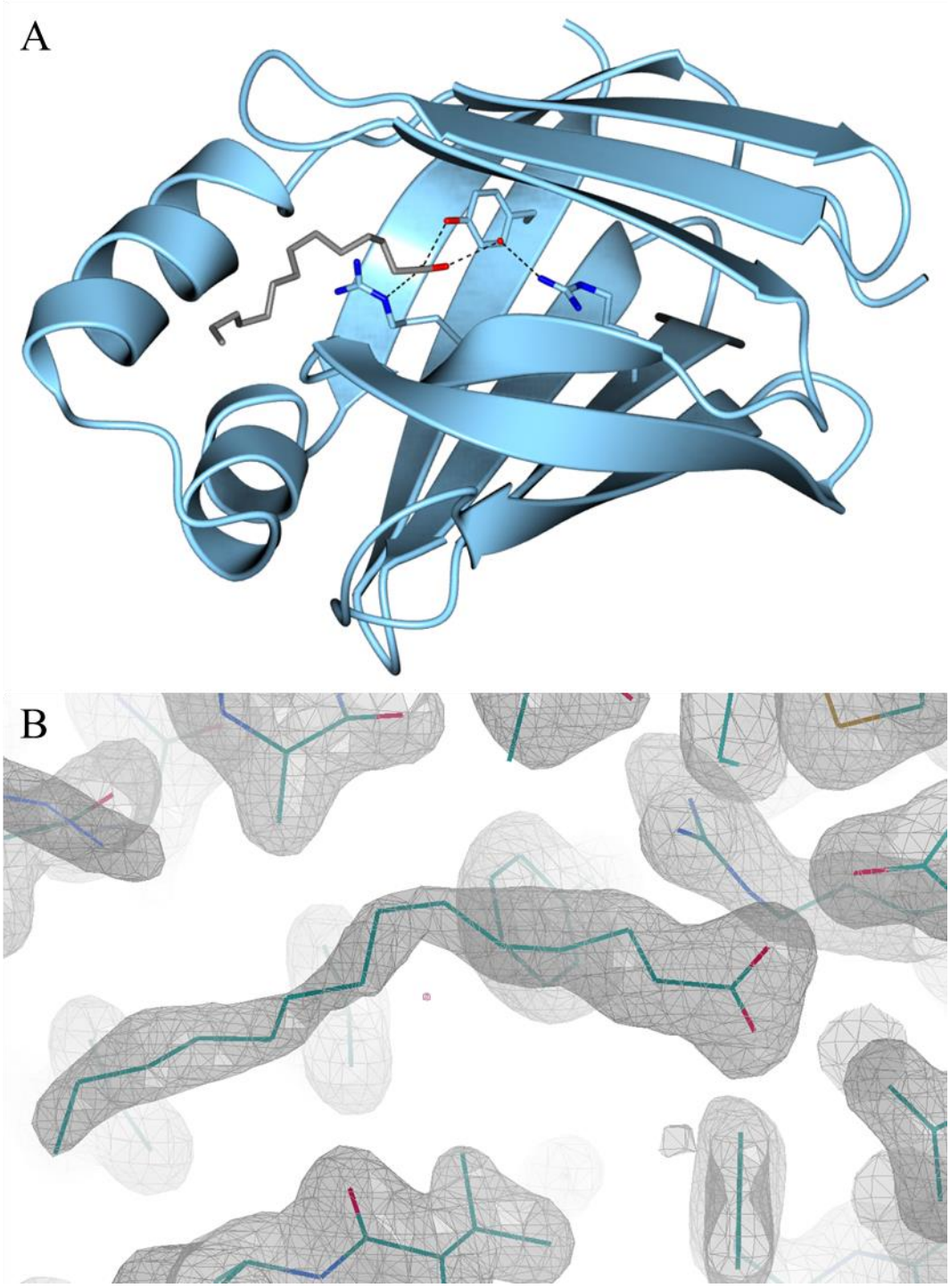


Figure 4.7: A - Crystal structure of CRABPI-L29C (pale blue, ribbon diagram) with proposed fatty acid TDA (grey, cylinder form) and binding triad (pale blue, O-red, N-blue, cylinder form). The protein forms a dimer with non-crystallographic symmetry which, for clarity, has been omitted. The A chain is displayed, with RMSD at Cα to chain B of 0.53 Å. B - 2Fo-Fc map density of tridecanoic acid (TDA) in the ligand binding site, from refinement calculations including the ligand (contour at $\sigma = 1.5$).

Initial efforts to co-crystallise CRABPI-L29C with DC retinoids yielded a structure (1.6 Å, R/R_{free} 0.19/0.22) that did not incorporate the expected ligand; instead, a fourteen-carbon fatty acid myristic acid (MYR) and thirteen-carbon tridecanoic acid (TDA) were discovered occupying the binding sites of each monomer, respectively. It is likely that a range of fatty acids resulting from expression in *E. coli* are to be found in the binding site, and the modelling as TDA and MYR was based entirely on evident electron density (Figure 4.7 - B). Both fatty acids can be seen interacting with the binding triad of CRABPI-L29C (R112, R132, Y134) through hydrogen bonding to their carboxylic acid groups, as well as to solvent molecules. Given the similarity of CRABPI to other members of the fatty acid binding protein family, and the known roles for fatty acids in the retinoid signalling pathway (Goldstein *et al.*, 2003, de Urquiza *et al.*, 2000), the incorporation of such ligands as a by-product of expression is a logical conclusion to draw based on the electron density.

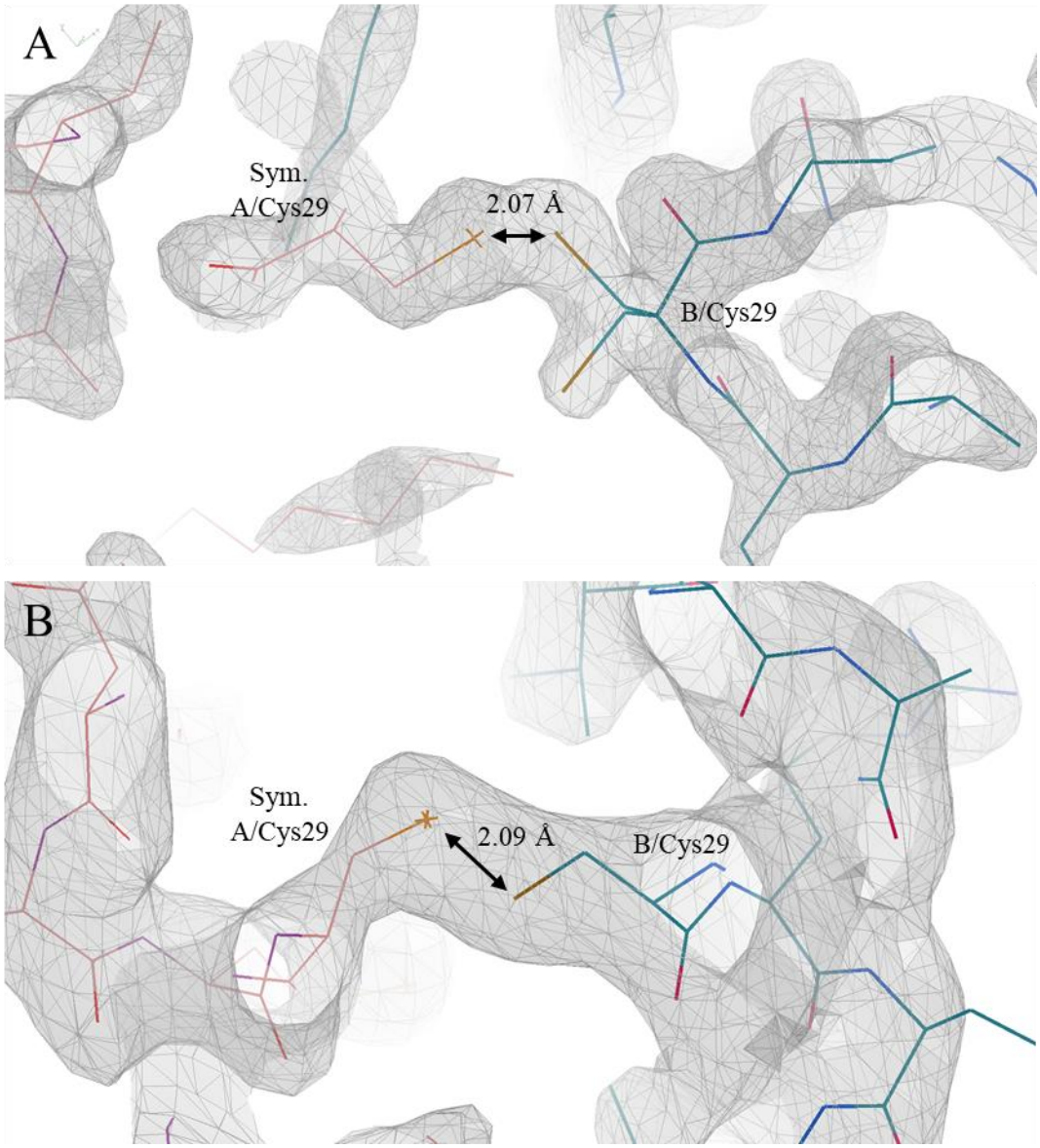


Figure 4.8: Comparison of disulphide bond formation between symmetry mates in CRABPI-L29C-MYR and CRABPI-L29C-DC645. Main chain (dark blue), symmetry mate (pale pink), atoms coloured O-blue, N-red, S-yellow. A - Dual conformation of CRABPI-L29C-MYR B/C29 creates partially occupied disulphide bond. B - Single conformer B/C29 creates fully occupied disulphide bond in CRABPI-L29C-DC645.

The structure, modelled containing TDA and MYR, resolution 1.6 Å, contains two molecules in the asymmetric unit related by non-crystallographic symmetry. Dimerisation can be seen to occur across the symmetry boundary, through C29, which in the B chain occupies two conformations, only one of which fulfils the parameters of a disulphide bond of 2.1 Å (Figure 4.8). This dimerisation is corroborated by size exclusion chromatography (HiLoad 16/600 Superdex 75 pg), which gives an approximate ratio for the protein of one third dimerised to two thirds monomeric (in the absence of ligand). It is likely that both states exist in equilibrium. For comparison, CRABPI-L29C-DC645 (Figure 4.8 - B) contains a single conformer for C29 which fulfils the disulphide bond parameters.

4.2.7. Dimeric crystal structure of CRABPI-L29C with incorporated DC645

Proper inclusion of a synthetic ligand was pursued with DC645 using conditions from crystallisation screening and an equimolar combination of protein and compound, which resulted in the structure shown in Figure 4.9.

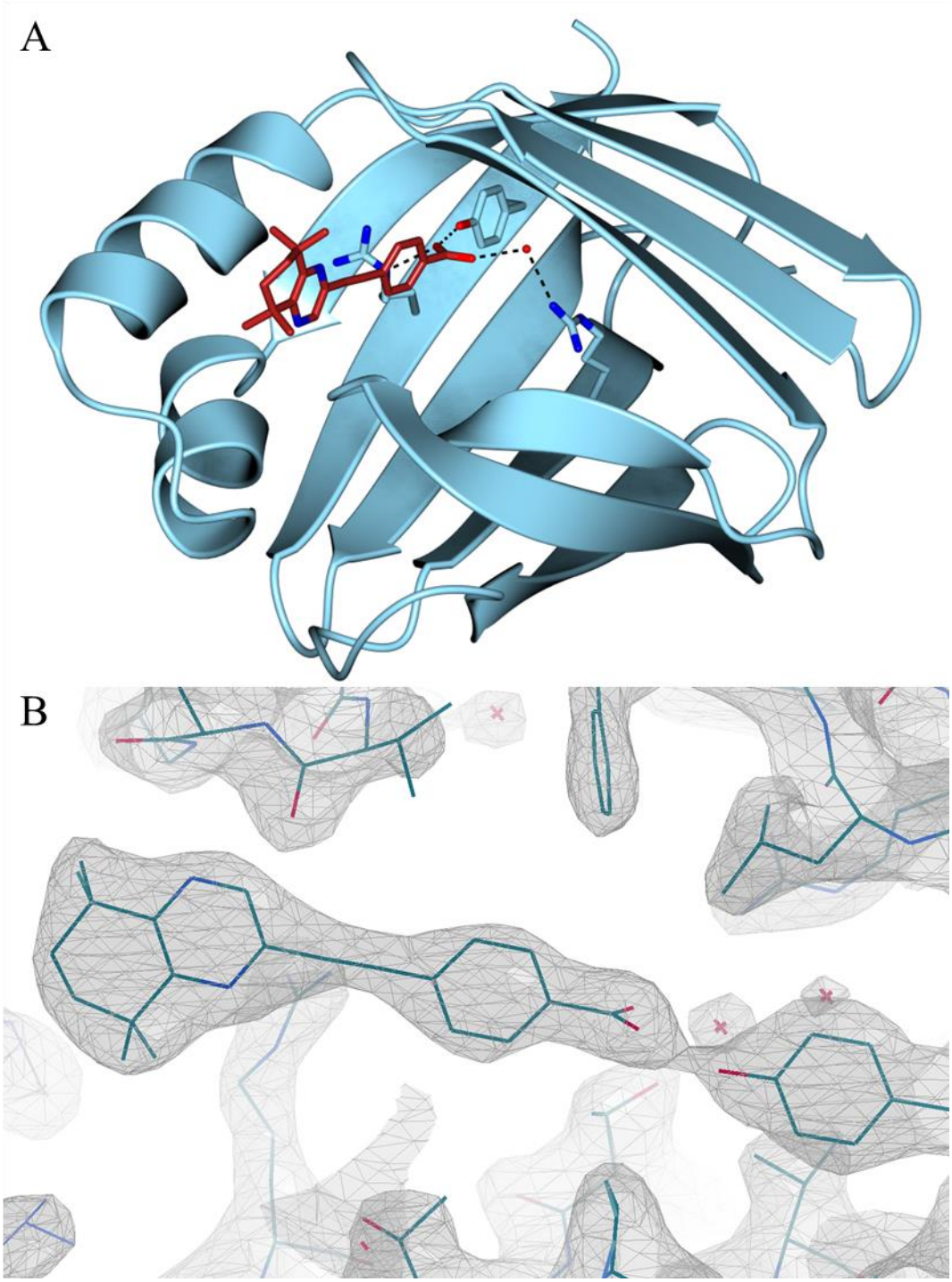


Figure 4.9: A - Crystal structure of CRABPI-L29C (pale blue, ribbon diagram) with synthetic retinoid DC645 (brick red, cylinder form) and binding triad (pale blue, O-red, N-blue, cylinder form). B - 2Fo-Fc map density of DC645 in the ligand binding site, from refinement calculations including the ligand (contour at $\sigma = 1.5$).

Resolved at 2.4 Å (R/R_{free} 0.20/0.25), the structure also contains a dimer; with one binding site occupied by DC645 and one empty of any electron density. When aligned, the two monomers are seen to be closely related (RMSD Cα = 0.66 Å) with key differences discussed below (4.3.2). The asymmetric unit, similarly to the TDA/MYR structure, contains two molecules related by non-crystallographic symmetry with dimerisation seen across the symmetry boundary. In this case C29 on both A and B chains is structured to form the disulphide bridge correctly, with no evidence of any other conformation (Figure 4.8). Whilst one ligand binding site is occupied by DC645, it is evident that the other is truly empty with no electron density that would suggest incorporation of an alternate ligand. DC645 can be seen to occupy the binding site as would typically be expected, interacting with the CRABPI binding triad as well as the conserved water molecule that bridges hydrogen bonding between the carboxylic acid and R112. The carboxylic acid moiety of the ligand is rotated relative to the position of the same atoms in the CRABPII-DC645 structure, due to the difference in binding site geometry. The two carboxylic acid groups actually interact with binding triad residues in the same orientation, meaning the 'body' of the ligand is rotated by ~45° (Figure 4.10).

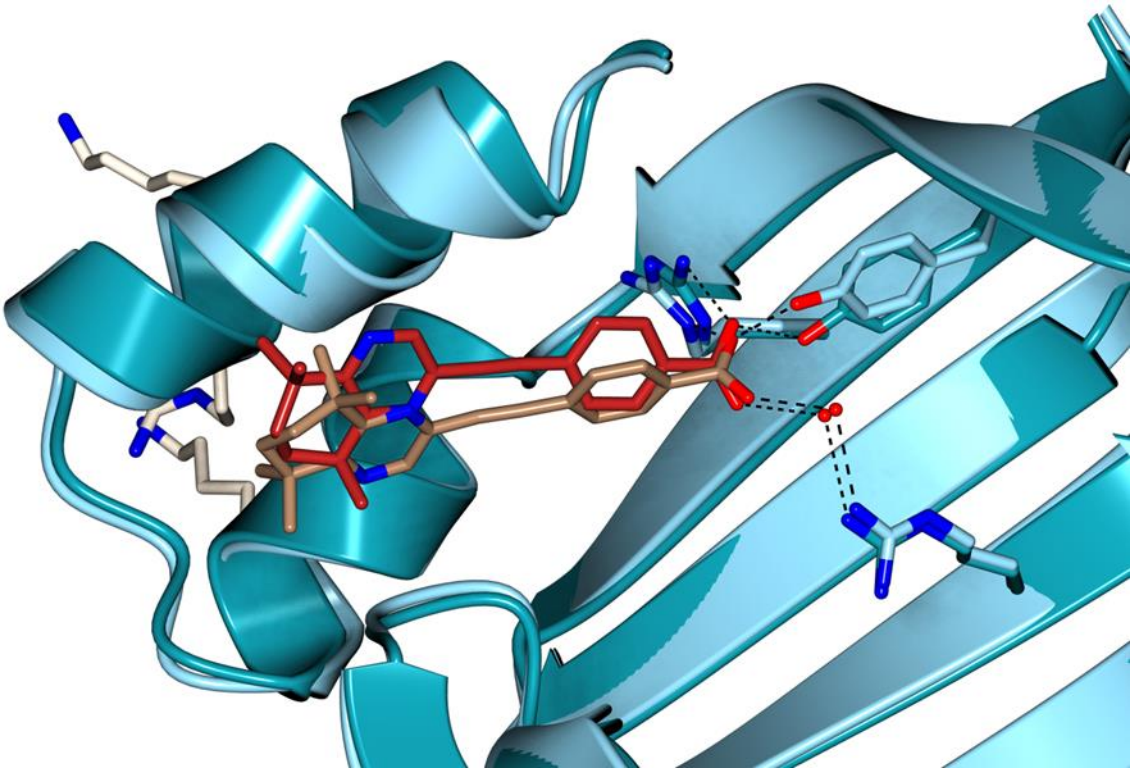


Figure 4.10: Superposition of CRABPII-DC645 (ribbon representation, dark blue, O-red, N-blue) and CRABPI-L29C-DC645 (ribbon representation, pale blue, O-red, N-blue) cut away to show similarity of ligand pose (cylinder representation, CRABPII-DC645-brick red, CRABPI-L29C-DC645-brown). CRABPII nuclear localisation sequence is shown (cylinder representation, white, N-blue). RMSD at Cα is 0.81 Å.

4.3. Discussion

Structural analysis of CRABPII in concert with DC retinoid derivatives was used to verify that each ligand that would be used as part of fluorescence displacement assays properly inhabited the hydrophobic binding site. It can be seen in the case of all CRABPII co-crystal structures presented above that this is the case, and in particular DC271, which was chosen as the key compound for assays, interacts with the ligand binding site in the expected manner. This canonical interaction consists of the hydrogen bonding patterns discussed and outlined above, as well as interactions with the conserved water molecule, and general hydrophobic interactions in the binding pocket. Understanding how each ligand interacts with the binding site increases the foundations on which new molecules can be designed. Confirmation that each co-crystallised compound is capable of interaction with the conserved water molecule in the same manner as the endogenous ligands, and do not depose it in favour of direct ligation to R112, suggests that this feature should be retained in future ligand design.

4.3.1. Analysis of CRABPII binding and DC645 interactions

One key factor in ligand interaction for CRABPII, other than simple incorporation into the binding site, is the sidechain conformation of residues K20, R29 and K30 that make up the nuclear localization sequence. (Sessler & Noy, 2005)

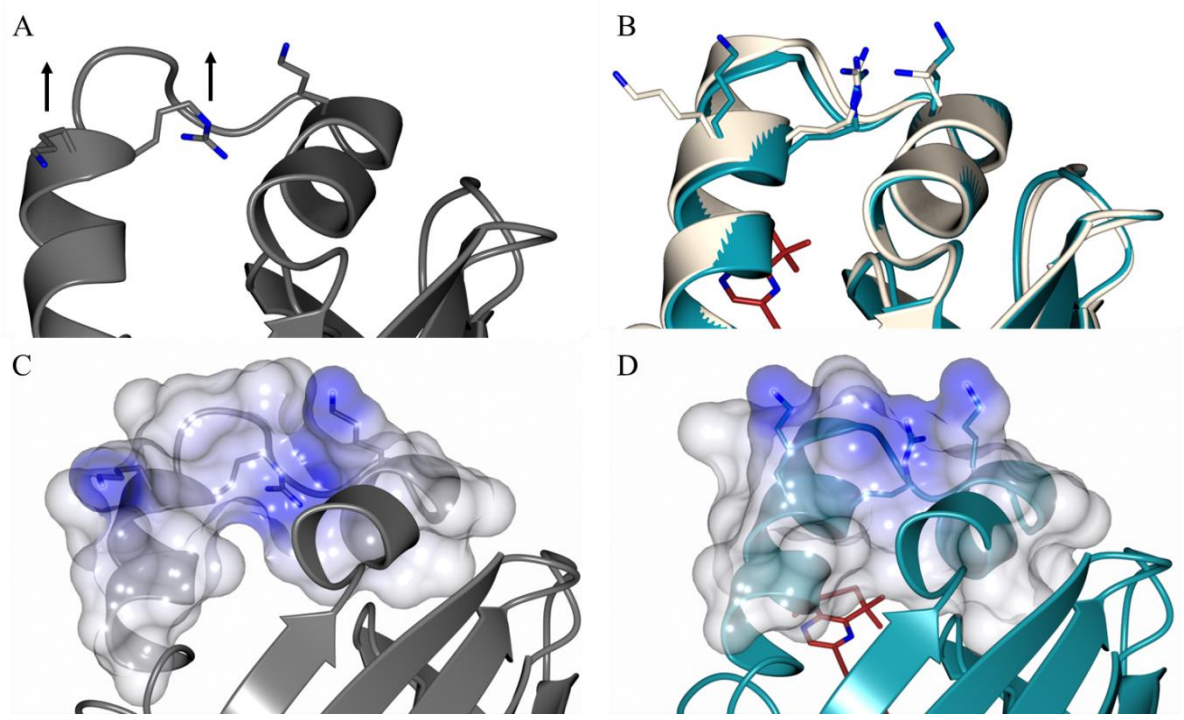


Figure 4.11: Unoccupied CRABPII (A, 1XCA, grey) in comparison with DC645 ligand bound complex (B, dark blue) aligned with ATRA ligand bound complex (B - 1CBS - white). Arrows (A) denote the direction of movement as residues adopt the ligand-bound conformation and form the nuclear localisation sequence. C and D illustrate the change to the resulting electrostatic surface that allows nuclear import (computed surface encompassing residues 20-35, white/neutral, blue/positive, generated in CCP4mg). (Sessler & Noy, 2005, McNicholas *et al.*, 2011)

Upon DC645 binding there is a key shift in the positions of residues R29 and K30 (Figure 4.11 - A/B); both sidechains move to an externalised position as residues 26-30 are rigidified into a helical conformation, creating a patch of positive charge in the electrostatic surface of the protein (Figure 4.11 - C/D). This has previously been identified as a pseudo nuclear localisation sequence by its structural and spatial (though not sequence) homology to known sequences in other proteins. (Sessler & Noy, 2005) Figure 4.11 demonstrates this shift in the case of CRABPII-DC645, when compared to ligand-free structure 1XCA and ligand bound structure 1CBS used in the literature study that first determined the NLS. All other CRABPII co-crystal structures presented above also adopt this conformation, and rear elevation views of the proteins including electrostatic surface are included in Appendix 16. The adoption of this conformation suggests that a compound would be imported into the nucleus *in vivo* and hence would be made available to the downstream RAR/RXR signalling system. In the case of DC645, nuclear import likely leads to upregulation of the downstream signalling pathway, evidenced by its neurite outgrowth-causing potential (Khatib *et al.*, 2020). Contrarily, DC479, shown to only act through the non-genomic signalling pathway, likely fails to interact favourably with the system after nuclear import has been achieved.

4.3.2. Analysis of CRABPI-L29C-DC645 Apo and Holo conformations

The understanding of CRABPI structure has previously been limited to structures of non-human analogues, determined at medium resolution. The use of the L29C mutant improved crystallisation, and the fatty acid-bound CRABPI-L29C structure shown above has been resolved to 1.6 Å, approaching atomic resolution. The use of human CRABPI also allows for insights to be drawn into the interaction behaviour of the protein, as the key difference (A86P) that defines the *Bos taurus* and *Mus musculus* analogues is located in the basal β-sandwich section reported to be the location of non-genomic signaling interactions. (Park *et al.*, 2019)

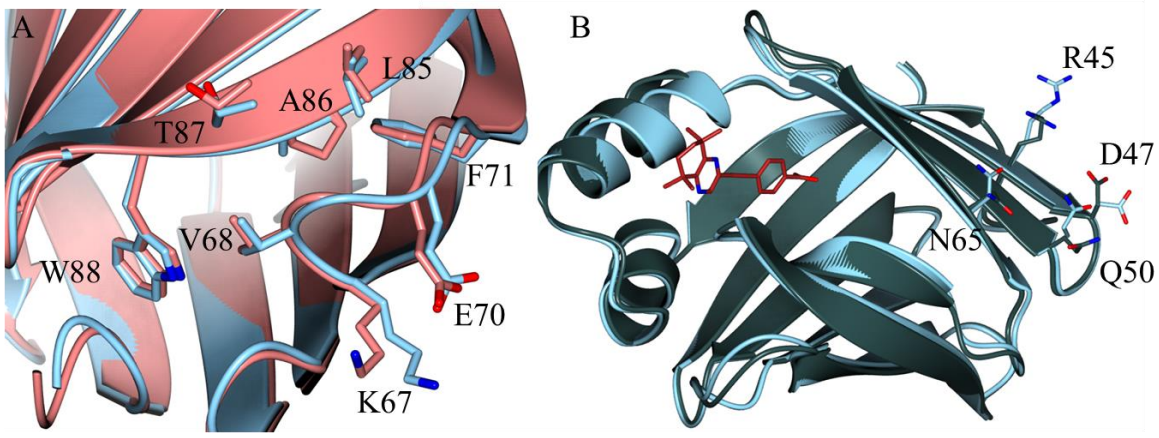


Figure 4.12: A - Least-squares superposition of CRABPI-L29C-DC645 (A chain, pale blue, *Homo sapiens*) with CRABPI (1CBR, pale red, *Mus musculus*) focused on the β-strand region 70-90, wherein the A86P modification that differentiates the two is found. RMSD at Cα = 0.69. B - Alignment of the unoccupied (B chain, slate grey) and ligand-bound (A chain, pale blue) monomers of CRABPI-L29C-DC645 which shows differences between the two forms (RMSD at Cα = 0.66).

Figure 4.12 (1CBR alignment, A) demonstrates that the A86P amino acid shift seen in non-human CRABPI does not affect the overall configuration of the protein, or the specific orientation of any side chains in the locality of the modification. This suggests that previous work done using analogues, both structurally and with binding or cellular data, is credible and applicable in the case of analogous human pathways.

Previous studies based on NMR data have suggested that CRABPI interacts with the Raf kinase, involved in non-genomic pathway signalling, at the base of the β-sandwich opposite to the ligand binding pocket. (Park *et al.*, 2019) Whilst there is no evaluation of specific residues, this suggests a method of protein-protein interaction, centred on the β-sandwich area, which is not yet fully understood. In CRABPI-L29C-DC645 (Figure 4.12 - B) there are several residues that adopt altered side chain conformations after ligand binding (aligning ligand bound A chain to unoccupied B chain). It is notable that these residues fall inside the dimerisation region of the pair of molecules in the asymmetric unit, which could be the cause of their alternate conformations or be a direct result of it, allowing for the unoccupied/ligand-bound cross-dimerisation to occur. It is also of

interest that the four highlighted residues (R45, D47, Q50 and R83) are amongst the few which make up the sequence variation between the otherwise closely related CRABPI and II. As previously, there is no observable structural rearrangement outside of the typical constriction of the binding pocket seen when a ligand is introduced to the protein. It is difficult, without further corroborating evidence, to draw any specific conclusion as to the role of these side chains, and whether or not they are part of a protein-protein interaction site; however, this high resolution determination of the structure of CRABPI-L29C creates a good starting point for further understanding of this complex and previously undefined system.

The structures presented above may also help define why the affinity of CRABPII for DC645 is greater than that of CRABPI-L29C (see 5.2.2). Residue R60 in the ligand binding site is positioned to create a hydrogen bonding network between the ligand and several water molecules in CRABPII. Whilst an analogous residue is also present in CRABPI it does not have the same orientation relative to the ligand and so could not act to increase binding efficiency. It could be through this small interaction that CRABPII gains its binding advantage over its cytoplasm bound cousin.

The determination of the structure of human CRABPI-L29C to a high resolution demonstrates that the protein compares closely to murine and bovine orthologs, and highlights interesting surface effects generated by ligand binding. DC645 demonstrably binds to CRABPI-L29C in the typical fashion: fulfilling hydrogen bonding patterns between the binding triad and the carboxylic acid head group, as well as creating a notable alteration to side chain orientations in the β-sandwich area previously characterised as a plausible protein-protein interaction site.

4.4. Methods

4.4.1. Crystallisation of CRABPI-L29C and CRABPII

Crystals for diffraction were grown by screening of conditions from commercially available crystallisation kits (Molecular Dynamics) including JCSG-plus™ HT-96 ECO screen and PACT premier™ HT-96 ECO screen. (Newman *et al.*, 2005) Purified proteins were combined with their co-crystallised retinoid (dissolved in EtOH) in a 1:1 equimolar fashion, using suitable concentrations to keep [EtOH] < 1%. Where necessary, screening was carried out using a mosquito Xtal3 robot (SPT Labtech) and sitting drop experiments (100+100 nl and 200+100 nl protein+reservoir droplet sizes), and optimised in sitting drop experiments (400 µl reservoir solutions) with varying drop sizes (1+1 µl and 2+1 µl protein+reservoir). The resulting crystals were inspected with a Leica MZ16 stereomicroscope and were mounted for diffraction using standard uni-puck style pins before flash freezing in $N_{2(l)}$ (Teng, 1990)

CRABPI-L29C (5 mg/ml, ~320 µM) was crystallised in 0.1 M sodium acetate pH 5.0, with PEG 6000 in a variety of concentrations.

CRABPII (5 mg/ml, ~320 µM) was generally crystallised in 0.1 M Tris-HCl pH 8.0, 0.2 M sodium sulphate with PEG 4000 in varying concentrations. CRABPII-EC19 was crystallised in 0.1 M Tris-HCl pH 8.0, 0.2 M sodium acetate and 30% PEG 3350.

4.4.2. Diffraction experiments

Diffraction experiments were carried out at Diamond Light Source as part of the BAG program. Data was collected on Pilatus detectors (Henrich *et al.*, 2009, Broennimann *et al.*, 2006, Casanas *et al.*, 2016), from cryo-protected, flash-frozen, crystals (Teng, 1990) using the rotation method (Rupp, 2009) and monochromatic X-rays. I02 was used to collect diffraction data for CRABPII-DC360. I24 was used to collect data for CRABPII-DC645 (remote collection); I03 was used for data collected for CRABPI-MYR (remote collection), CRABPII-DC271 and CRABPII-DC479 (remote collection). I04 was used for CRABPI-L29C-DC645 (remote collection). Further details are summarised in the crystallographic tables (Appendix 9 and Appendix 10).

4.4.3. Data processing

Data processing was carried out using the CCP4 suite of programs, as well as XDS (Kabsch, 2010), iMosflm (Battye *et al.*, 2011), and DIALS (Winter *et al.*, 2018) as part of the autoprocessing (Winter, 2010) done at Diamond Light Source and for reprocessing of data. Scaling was carried out with aimless/pointless (Evans & Murshudov, 2013). Phasing was performed via molecular replacement with Phaser (McCoy *et al.*, 2007), and manual modifications made with *Coot*. (Emsley *et al.*, 2010) REFMAC5 was used for refinement. (Murshudov *et al.*, 1999, Nicholls *et al.*, 2012, Winn *et al.*, 2003, Murshudov *et al.*, 1997, Vagin *et al.*, 2004, Murshudov *et al.*, 2011)

The disulphide bonds of CRABPI-L29C were added using SSBOND comments in the PDB file prior to refinement, to improve definition of the interaction of A/C29 with B/C29 of the symmetry mate using REFMAC5. (Murshudov *et al.*, 1999, Nicholls *et al.*, 2012, Winn *et al.*, 2003, Murshudov *et al.*, 1997, Vagin *et al.*, 2004, Murshudov *et al.*, 2011)

Local NCS restraints generated by REFMAC5 were used where appropriate to improve the refinement of structures containing a dimer, or two monomer units. (Uson *et al.*, 1999, Murshudov *et al.*, 2011)

Alignments and figure making was carried out with CCP4mg. (McNicholas *et al*., 2011) *Coot* was used for representations of electron density and omit maps. (Emsley *et al.*, 2010) Polder maps were generated using phenix.polder. (Liebschner *et al.*, 2017)

5: Application of Fluorescent Binding assays to CRABPII and RAR proteins

5.1. Introduction

Target-led drug design requires exemplary knowledge of a signalling pathway, and relies on the collection of structural data, the design of assays, and the screening of large numbers of candidate molecules. The fields of genomics and proteomics have led the way in uncovering targets for novel drugs, allowing for in-depth analysis of specific promoter sequences and gene expression patterns. (Kiriiri *et al.*, 2020) Phenotype-based (classically empirical) screening comparatively requires less foreknowledge of a process or pathway, and rather a strong cellular model of the system as a whole upon which outcomes can be tested. (Comley, 2015)

The retinoic acid signalling pathway is a potent controller of neurite development and cell differentiation. (Corcoran & Maden, 1999, Ross *et al.*, 2000, Bastien & Rochette-Egly, 2004) The potential for modification and control of the pathway using synthetic retinoid derivatives creates an opportunity to improve the treatability of Alzheimer's disease, and other conditions currently considered underserved by the available pharmacopoeia. (Cummings *et al.*, 2014, Sahin *et al.*, 2005) Data on the potency of several retinoic acid derivatives from the DC series of compounds has been collected based on neurite outgrowth assays in cell culture. (Khatib *et al.*, 2019, Khatib *et al.*, 2020) Whilst this data is indicative of which retinoids have the most potent outcomes on neurite outgrowth, and by proxy the most potential as an anti-neurodegenerative, it offers no indication of how each compound is acting in the cell, by what signalling method or pathway. It is known that both genomic and non-genomic signalling have the potential to influence overall cell growth, and specific knowledge of what mechanism of action is being undertaken will allow for better evaluation of compounds and compound design going forward.

Using fluorescence displacement assays, the binding constants of compounds as ligands for CRABPI-L29C or CRABPII (linked to non-genomic and genomic signalling respectively) will be determined and similarly, affinities for different RAR isoforms will be calculated. Determination of K_d values for these interactions will create a grading of compounds by their effectiveness, which can be taken into account in the design of new molecules. Given the large number of possible RAR:RXR heterodimers that can be formed, targeting a specific protein and isoform will be vital in controlling genomic output effectively.

5.2. Results

5.2.1. CRABPII-DC645 solvatochromic fluorescence binding assays

Having been chosen for further investigation due to its interesting *in cellulo* properties (Khatib *et al.*, 2020), DC645 was crystallised in the binding site of both CRABPII and CRABPI-L29C (4.2.5 and 4.2.7). Successful co-crystallisation of the ligand with CRABPII demonstrated binding site incorporation, full interaction with the ligand binding triad, and generation of the nuclear localisation sequence, suggesting validity in investigating DC645 with fluorescence displacement binding. DC645 differs from the purposefully fluorescent analogues of EC23, being based on tetrahydroquinoxiline and having more hydrogen bonding opportunities as a result. Whilst this does not appear to affect incorporation into the CRABPII binding site, it may have an impact on the binding affinities and therefore preference for particular retinoid binding proteins. Determining these preferences is vital for understanding the cellular effects of the retinoid and for the development of future compounds designed with specific binding in mind.

The ability to test large libraries of compounds for activity against a target is a core principle of structure-based drug design. Processing large numbers of compounds, in a high-throughput manner, and interpreting the results in concert with other key data is vital in finding novel ways to control signalling pathways. Binding activity determination, from high-throughput assays, is often cross-compared with phenotype testing data from cellular assays, or directly analysed by 3D quantitative structure activity relationship (QSAR) to derive a model for the development of future compounds. High-throughput screening can also guide other experimental work, by narrowing down a number of candidates to test further from libraries that routinely contain thousands of compounds. Proving that an assay is scalable, and produces interpretable data, is a vital first step towards this process.

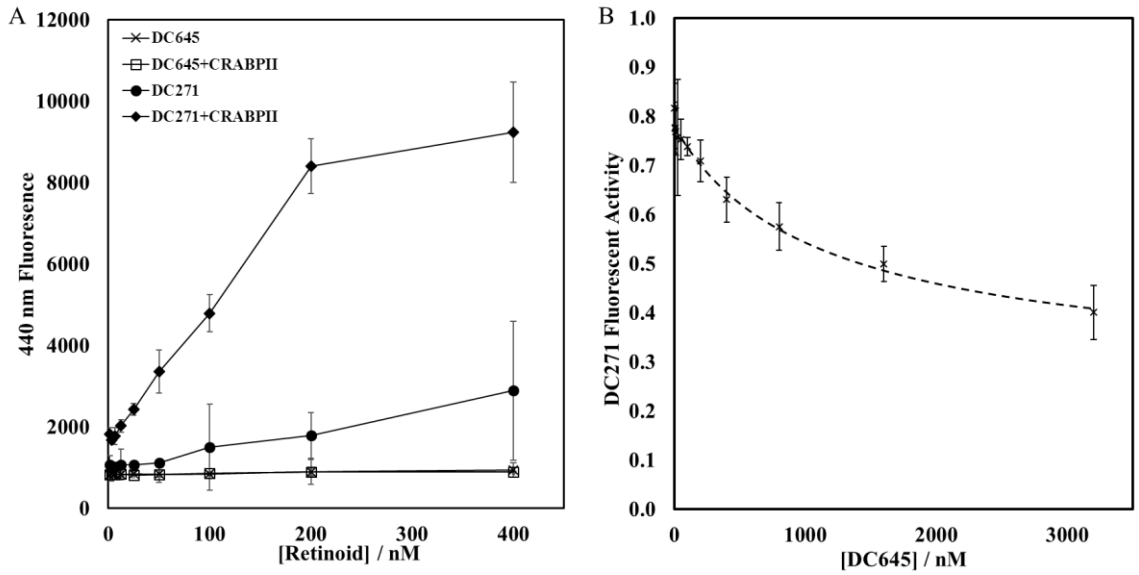


Figure 5.1: A - Retinoid fluorescence analysis of DC271 and DC645 in the presence and absence of CRABPII (n = 3) at ex/em 350/440 nm. DC271 presents a strong solvatochromic effect in the presence of CRABPII due to inclusion into the binding site. (Chisholm *et al.*, 2019) DC645 instead displays low intrinsic fluorescence and no solvatochromic effect. B - Competitive binding analysis of DC645 to CRABPII by displacement of DC271, curve fitting performed using Dynafit (n = 6, α = 0.05). (Kuzmič, 1996)

To characterise the protein-ligand interaction of DC645 with CRABPII, binding was assessed using the solvatochromic retinoid displacement method (3.2.3). By interpreting the displacement of the solvatochromic, fluorescent retinoid DC271 from the hydrophobic binding pocket, the relative binding affinity of DC645, which would otherwise be hard to detect, was determined. DC645 displays some small intrinsic fluorescence at the typical emission wavelength 440 nm, which forms the tail of a larger fluorescence in the sub 400 nm region. As a result, the compound was included into the background fluorescence control of the experiment. Prior to experimentation it was also demonstrated that DC645 was not solvatochromic with respect to CRABPII (Figure 5.1 - A), *i.e.* absolute fluorescence of the molecule in solution was not dependent on CRABPII concentration and therefore not dependent on incorporation into the binding site. This allowed the assay to be undertaken as normal, with background controls accounting for the low level DC645 fluorescence. Using Dynafit to fit the data (Figure 5.1 - B) a K_d value of 0.25 ± 0.06 μM was determined for DC645 binding to CRABPII. (Kuzmič, 1996)

5.2.2. CRABPI-L29C-DC645 solvatochromic fluorescence binding assay

Given the impact of DC645 on both genomic (due to RAR signalling and facilitated by CRABPII shuttling (Durand *et al.*, 1992, Bastien & Rochette-Egly, 2004)) and non-genomic (purported to be due to CRABPI interaction with other cytoplasm proteins (Persaud *et al.*, 2013, Park *et al.*, 2018, Park *et al.*, 2019)) signalling pathways, investigation and comparison of its binding to CRABPI (CRABPI-L29C) was carried out using fluorescence displacement. Previous crystallographic evidence shows that DC645 is capable of inclusion into the binding site of CRABPI-L29C under co-crystallisation conditions (~ 320 μM equimolar mixture), but occupancy of only half of the monomer units suggests its affinity may vary from that of other tightly binding retinoids.

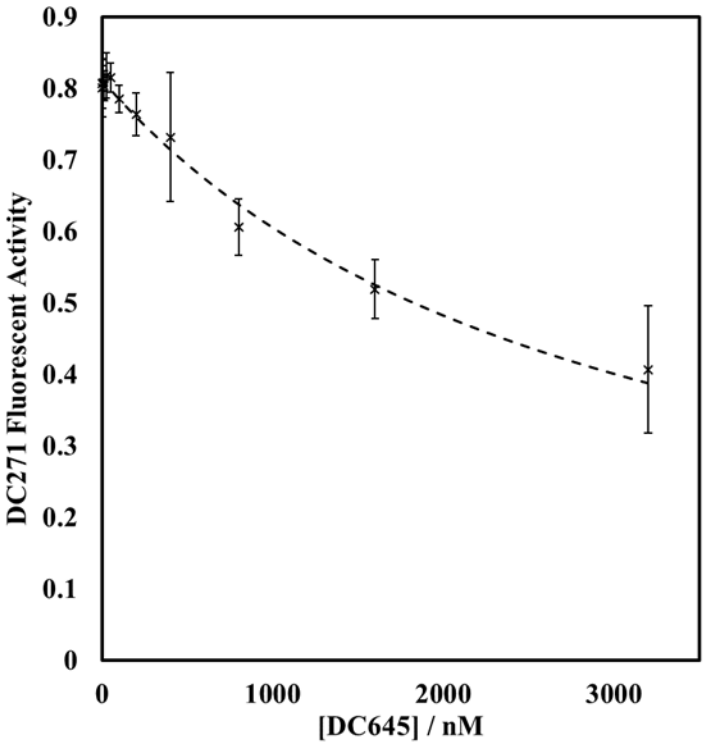


Figure 5.2: Competitive displacement of DC271 from the hydrophobic binding pocket of CRABPI-L29C by DC645 was carried out using previously discussed methodology; curve fitting was carried out by least-squares regression using Dynafit. (n = 6, $\alpha = 0.05$) (Kuzmič, 1996)

Competitive binding analysis using DC271 (Figure 5.2) shows that CRABPI-L29C interacts less favourably with DC645 than CRABPII - a much larger concentration of the ligand is required to displace DC271 to the same degree. Least-squares regression of the curve gives a K_d value of 1.94 ± 0.11 μM for the pair. This may explain why only one of two ligand binding sites seen in the crystal structure (4.2.7) are ligand-occupied, the interaction being considerably weaker than other retinoid/binding protein pairs, and suggests that greater than equimolar combination may be required to capture a fully occupied structure.

5.2.3. RARα and RARγ solvatochromic fluorescence binding assays

To prove the efficacy, and demonstrate the medium-/high-throughput nature, of the fluorescent retinoid displacement assay, a series of DC compounds were selected and tested for binding affinity to RARα and RARγ. Each was from a series of compounds designed to explore the chemical space around a common core, with targeted modifications to chemical functionalities. (Chisholm *et al.*, 2016, Chisholm & Whiting, 2020) Most are based around the EC23 core ligand design, with a tetra/dihydronapthalene 'tail' appended to a benzoic acid 'head' via an alkyne linker. Added to this framework are halide (fluorine and chlorine) and ether groups at both head and tail locations, as well as modifications to methyl group substituents. Fluorine substituents in particular are widely harnessed in the design of compounds for medicinal chemistry due to their lipophilicity and stability modifying effects. (Yerien *et al.*, 2016, Huchet *et al.*, 2015) Introduction of fluorine atoms alters polarity within a molecule, whilst minimally impacting overall volume, and can act to slow degradation by cytochrome p450 pathways *in vivo* (Huchet *et al.*, 2015). Additionally, whilst fluorine is capable of hydrogen bonding, PDB and CCDC data show it rarely does so, generally only acting to increase polarity and bonding via Van der Waals interactions. (Howard *et al.*, 1996, Dunitz & Taylor, 1997)

DC645 and DC712 substitute tetra/dihydronapthalene for tetrahydroquinoxaline, changing the pattern of their hydrogen bond donors. DC716, DC718, DC719, and DC721 substitute the alkyne linker for a thiazole ring, creating a bent profile predicted to better bind to RXR proteins in the manner of 9-*cis* retinoic acid. (Heyman *et al.*, 1992, Levin *et al.*, 1992) Analysis of the output data (Figure 5.3 and Figure 5.4) will allow for correlation between molecular structure and binding efficiency, as well as cross correlation with genomic output assays (Khatib *et al.*, 2019, Khatib *et al.*, 2020), in the design of future lead compounds for affecting the retinoid signalling pathway, and targeting specific signalling proteins in a precise manner.

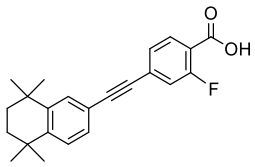
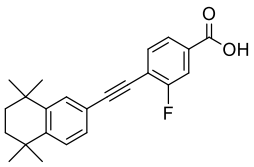
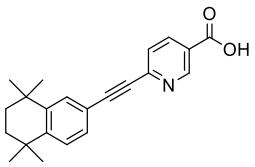
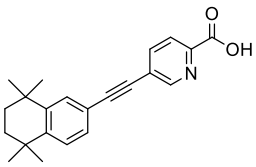
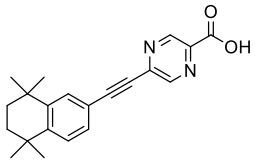
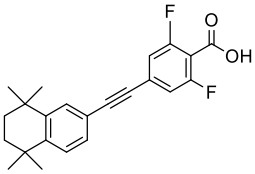
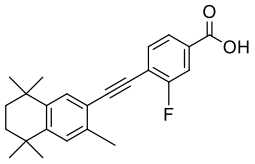
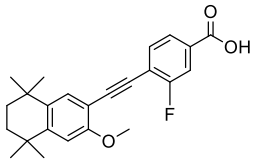
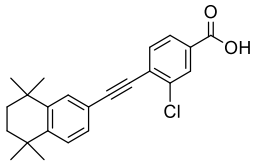
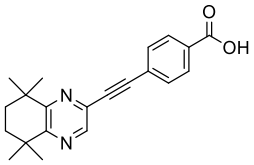
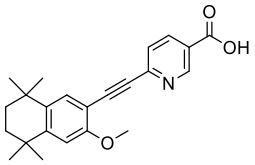
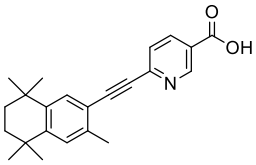
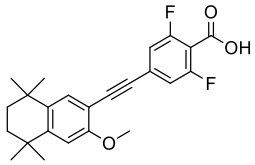
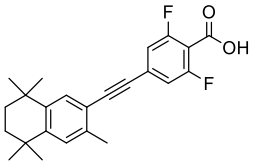
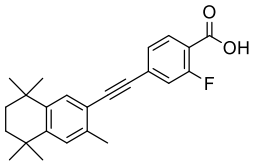
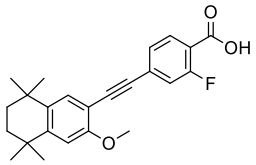
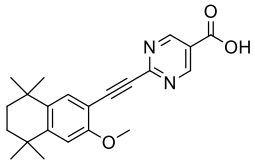
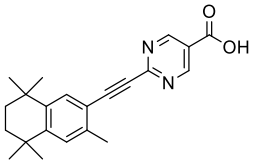
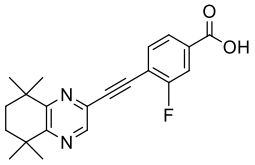
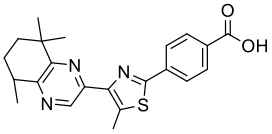
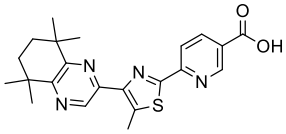
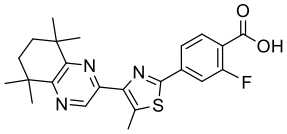
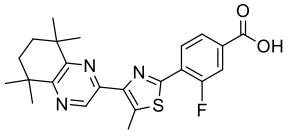
 **DC525** **0.81±0.68 / -**	 **DC526** **22.0±4.30 / -**	 **DC527** **29.0±13.2 / 97.6±34.3**	 DC528 30.8±12.7 / 106.6±36.2
 DC529 - / -	 **DC540** **1.4±0.8 / 14±2.5**	 DC559 53.6±1.4 / -	 DC564 563.3±127.5 / -
 **DC567** **13.0±8.1 / 10.9±1.6**	 DC645 93.0±12.9 / 92.0±8.6	 DC646 - / -	 DC656 52.3±31.5 / -
 DC657 78.7±19.3 / 313.3±100.4	 **DC661** **- / 22.3±7.2**	 **DC667** **2.8±7.1 / 52.7±21.7**	 DC673 - / -
 DC706 - / -	 DC707 416.7±62.5 / 893.3±37.9	 DC712 148.7±10.0 / 101.5±120.7	 DC716 369.3±68.5 / 336.7±87.2
 DC718 - / -	 DC719 133.0±68.9 / 240.0±65.7	 DC721 - / -	

Figure 5.3: DC series compounds from the RARα/γ test set. K_d values were calculated with Dynafit (Kuzmič, 1996) as previously described (n = 3 α = 0.05). Each compound is arrayed with its structure, name, and K_d for RARα / RARγ expressed in nM. [-] indicates no binding was detected. Compounds with sub 30 nM affinity to at least one protein are captioned in bold text, compounds where binding is preferential to one protein are captioned in red.

It is evident from the wide range of binding coefficients (0.8 nM to 0.8 μM) that variation in molecular structure, even around the common EC23 core, creates a wide variation in affinities to RARα and RARγ proteins (Figure 5.3). Introduction of ether groups appended to the di/tetrahydronapthalene ring appears to reduce affinity relative to similar compounds, or prevent detectable binding in totality. Comparing DC540 (α/γ 1.4/14 nM) and DC657 (α/γ 78.7/313.3 nM) it is evident that introduction of the ether group in DC657 has shifted binding from strong to relatively weak, and that dual fluorination in the *ortho* position favours binding to RARα by approximately tenfold. Similarly, comparing DC667 (α/γ 2.8/52.7 nM) and DC673 (no detectable binding), the change from methyl group to ether, respectively, disrupts binding to the extent that the ether-containing DC673 does not perceptibly bind to either RAR isoform.

Introduction of pyrazidine-carboxylic acid as a head group in place of benzoic acid (DC706 and DC707) also seems to impede binding to both RAR proteins, with the DC706 ether substituted molecule having no detectable binding compared to the DC707 methyl substituted version.

DC645, based on the dihydroquinoxiline structure and a key promoter of neurite outgrowth (Khatib *et al.*, 2020), can be compared to DC712. Fluorination at the *meta* position in DC712 appears to decrease the affinity to RARα by 50 nM relative to DC645, whilst binding to RARγ remains stable.

All molecules based on the bent, thiazole linker structure display relatively large binding constants, and so weak binding, when compared to the strongest binding affinities detected, and those of natural ligands. (Giguere *et al.*, 1987, Allenby *et al.*, 1993, Allenby *et al.*, 1994) DC718 and DC721 display no binding affinity within the range of the assay, potentially due to solubility issues or modifications at their acid head group, using pyridine and *meta*-fluorinating, respectively. DC716 shows no preference for either RARα or RARγ (K_d 369.3 nM *vs* 336.7 nM respectively), whilst DC719, which is *ortho* modified with fluorine, shows an overall stronger affinity for both proteins, and reasonably strong preference for RARα (α/γ K_d 133.0 nM *vs* 240.0 nM).

5.2.4. RAR isoform selectivity

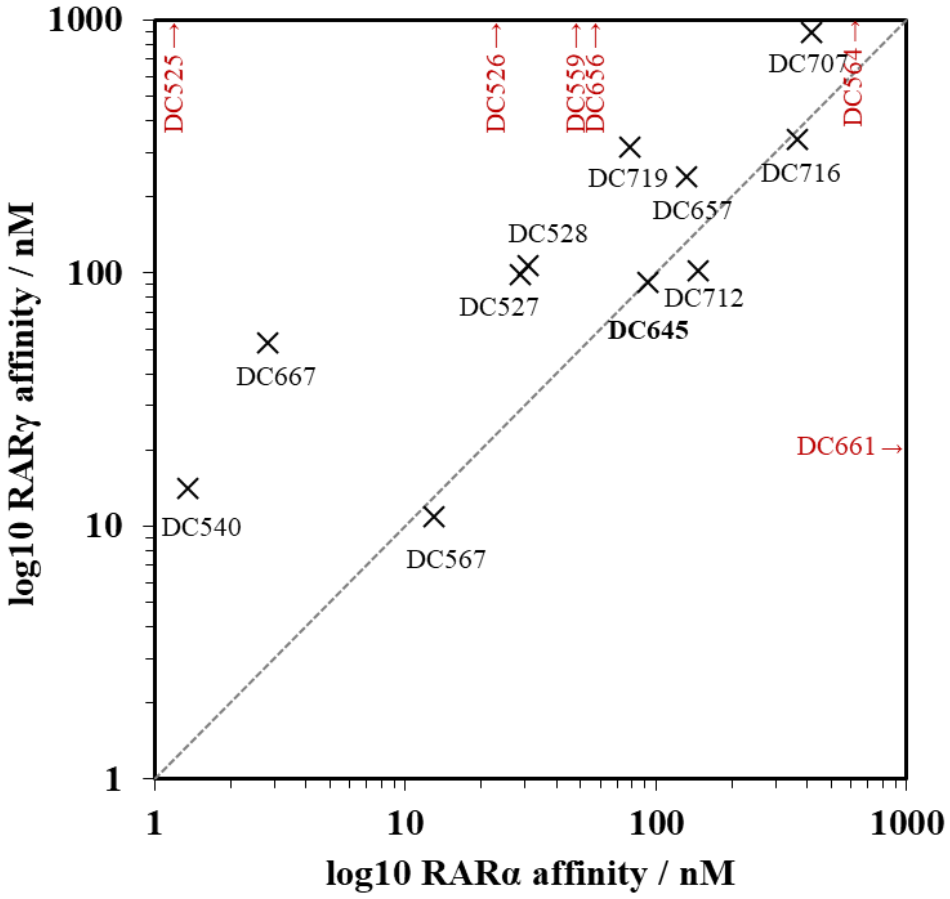


Figure 5.4: Crosswise analysis of RARα and RARγ binding data (n = 3, α = 0.05) to show preference of each DC compound: compounds marked in red have affinity for only one protein, lead compound DC645 is marked in bold. Compounds above the median favour RARα, those below favour binding to RARγ. Logarithmic axes assist in the display data in the low nanomolar K_d range.

Further analysis of the binding affinities for compounds to both RARα and RARγ highlights those which show a definite preference for one receptor protein, and those who bind ambivalently. Of particular notice is DC645 (α/γ 93.0/92.0 nM), which sits right on the meridian in Figure 5.4 with equal binding affinity to both proteins. As lead compound with known strong activity in neurite outgrowth assays (Khatib *et al.*, 2020), the binding preferences of DC645 could be instrumental in understanding its phenotypic outcome. Given matched binding to both RARα and RARγ, it could be an overall strong up-regulator of retinoid signalling (in the style of ATRA), or be favouring action via RARβ, still to be uncovered. Other compounds with similar matched binding include DC567 (tighter binding overall than DC645), DC712 (which displays high variance) and DC716 (lower overall affinity *vs* DC645).

Compounds that show a particular preference for one receptor protein are key to understanding how preferential binding can be recreated in other molecules. DC661, preferentially binding to RARγ due to a lack of detected binding to RARα, displays di-fluorination in the benzoic acid head group and a methylated tetrahydronapthalene. When compared to DC667, which is mono-fluorinated and binds with good affinity to RARα but reduced affinity to RARγ, it can be deduced that di-fluorination in that position has limited the RARα binding affinity. However, all other compounds with instances of *ortho*-fluorination are seen as RARα selective, with reduced RARγ affinities, suggesting that it is the combination of di-*ortho*-fluorination and methylation of the tetrahydronapthalene group that weaken RARα binding in DC667. Similarly DC525, which shows huge affinity for RARα but no binding to RARγ, displays the same fluorination pattern as DC667 but lacks the methyl group appended to the tetrahydronapthalene. This methyl group can be seen again in DC527 and DC656, similarly reducing the affinity to RARα. An overall pattern begins to emerge, that *ortho*-fluorination (mono or di) favours binding to RARα and that methylation at the di/tetrahydronapthalene opposes it. Apart from DC661, which has been discussed, no compound favours RARγ in a meaningful fashion.

5.3. Discussion

Fluorescence displacement assays were carried out to determine the binding of a range of compounds to CRABPII, CRABPI-L29C and RAR isoforms α and γ. Data shows that CRABPII and CRABPI-L29C differ in their binding to DC645, and that binding affinities for a series of compounds with RAR isoforms differed greatly based on modifications to chemical functional groups.

5.3.1. DC645 binding - CRABPII *vs* CRABPI-L29C

The affinities of DC645 for both CRABPII (0.25 ± 0.06 μM) and CRABPI-L29C (1.94 ± 0.11 μM) are sufficient to indicate strong binding interactions, but are orders of magnitude weaker than those of natural ligands or other synthetic derivatives such as ATRA and EC23. It is known that DC645 modulates the retinoid signalling pathway via both genomic and non-genomic means (Khatib *et al.*, 2020), which tallies with binding data; CRABPII binding is vital for passage into the nucleus to effect traditional signalling via the RAR proteins (Sessler & Noy, 2005), whilst CRABPI is implicated in signalling through non-genomic means including the Raf kinase pathway. (Persaud *et al.*, 2013, Park *et al.*, 2018, Park *et al.*, 2019)

Key comparisons are derivative cousins DC271 (CRABPII binding affinity 32.3 nM) and DC360 (CRABPII binding affinity 9.6 nM) based on di/tetrahydroquinoline rather than tetrahydroquinoxaline. Whilst the binding of DC645 to CRABPII is credible, it is an order of magnitude weaker than that of other compounds, and CRABPI-L29C binding is weaker again. It is plausible that the increased hydrogen bonding and reduced hydrophobicity of this tail moiety, observed in the crystallisation of both proteins with DC645, is the cause of the reduced affinity. Increased hydrogen bonding allows greater interaction with, and retention of, the network of water molecules that occupy and stabilise the binding site in the absence of a ligand. (Vaezeslami *et al.*, 2008, Vasileiou *et al.*, 2009) The presence of these water molecules once the ligand is bound, and lower overall hydrophobic contribution to binding, may be the cause of the comparatively lower binding constants.

5.3.2. Implications of RARα and RARγ affinity determination

The determination of binding constants for a series of DC compounds with both RARα and RARγ yielded a wide range of values (0.8 nM to 0.8 μM), with the majority of molecules either of neutral preference or binding selectively to RARα. DC661 was the only compound to display RARγ selectivity, with lack of any detectable activity against the α-isoform.

Overall interpretation suggests that some modifications decrease the binding affinity for both proteins, namely the addition of an ether group to the di/tetrahydroquin/quinoxaline tail. Fluorination also seems to result in complex effects, generally increasing preference for RARα,

with methylation of the tail group reducing this effect. Use of pyrazidine-carboxylic acid as a head group, and to some extent using a quinoxaline-based tail group, decreases affinity for both proteins with no observable preference for either isoform. It is likely that this is due to disruption of existing hydrogen bonding networks caused by the introduction of more hydrogen bond donating nitrogen atoms. Similarly, thiazole-based compounds are overall less successful at RAR binding, with only two of four compounds showing any detectable binding.

Interpretation of the effects of specific molecular modifications is valid, based on the data at hand, but should be considered holistically with results from cellular and other assays going forward. All compounds that exhibit a sub-micromolar affinity to either protein could prove useful for further development, but their specific impact on cellular growth, as well as interactions with RXR and other nuclear receptor proteins, needs be considered in tandem. Additional analysis of binding of all compounds to RARβ would also be advantageous, as signalling can be exacted through any of the protein isoforms. RXR proteins are an additional source of signalling complexity, with a similar isoform pattern to the RAR proteins, and should be considered for future screening.

Engineered specificity to one RAR isoform is a key goal in developing new active retinoid compounds, because specific activation of one signalling outcome will allow for finer control than the broad-spectrum activation seen with current derivatives. The discovery of chemical features that promote such behaviour from existing libraries will have great impact going forward with new candidate design.

5.4. Methods

5.4.1. RARα and RARγ competitive binding assay

Solutions of DC271 (300 nM, <1% EtOH) and RAR protein (300 nM,) were prepared and combined before use. A Corning black, NBS fluorescence plate (96-well) was cleaned using compressed air, and Max, Mid and all test compound wells were loaded with 50 µl RAR/DC271 solution. Min wells were loaded with 50 µl control solution of DC271 mixed with assay buffer. Mid wells were then loaded with 50 µl EC23 (600 nM, < 1% EtOH), and Max and Min wells with 1% EtOH solution (50 µl). Dilution series of test compounds were prepared and aliquoted to the plate (50 µl). Plates were spun for 2 min (1500 rpm) to ensure mixing and read using a Synergy H4 plate reader (ex/em 355/460 nm). The total well volume was 150 µl, the on plate concentration of RAR protein and DC271 was 100 nM.

Data processing was carried out as per 3.4.4.

6: Behaviour of CRABPII mutants

6.1. Introduction

CRABPII is known to bind all-*trans*-retinoic acid (and derivatives) with nM affinities (Dong *et al.*, 1999, Chisholm *et al.*, 2019), which induces travel through the nuclear pore (Sessler & Noy, 2005) and allows access to the retinoic acid receptor (RAR) and retinoid-X receptor (RXR) proteins. (Delva *et al.*, 1999, Budhu *et al.*, 2001) The recruitment of ligand bound RAR:RXR heterodimeric complexes to retinoic acid response elements (RAREs) subsequently induces gene expression and changes cell phenotype. (Durand *et al.*, 1992, Dilworth & Chambon, 2001)

In addition to RAR/RXR interactions, yeast two-hybrid screening of cDNA from haematopoetic Hl-60 cells revealed an interaction between cyclin D3 and CRABPII. (Despouy *et al.*, 2003) The cyclin family of proteins are involved in regulation of the cell cycle and the D-type cyclins specifically are expressed in response to mitogenic growth factor in the G1 phase. (Evans *et al.*, 1983, Tatum & Endicott, 2020) The family is large, but typified by interactions with cyclin-dependant kinases (CDKs) and the presence of a cyclin box domain (CBD) used for these interactions. (Adams *et al.*, 1996, Lowe *et al.*, 2002, Tatum & Endicott, 2020) Using hCRABPII as bait, these interactions were shown to be specific to cyclin D3 (not replicated by cyclins D1 or D2) and retinoic acid independent. (Despouy *et al.*, 2003) Subsequent interactions between CRABPII/cyclin D3 and RARα as a ternary complex were also observed independent of ligands, as well as a direct cyclin D3/RARα (absence of CRABPII) interaction mediated by the presence of retinoic acid. Finally, cyclin D3 was shown to increase expression of retinoic acid target genes, likely by facilitating interactions between CRABPII and the RAR proteins. (Despouy *et al.*, 2003) The RXL motif, generally considered to govern binding to cyclin D3 through the cyclin box fold (Adams *et al.*, 1996, Lowe *et al.*, 2002), is absent in CRABPII, suggesting the existence of an alternate cyclin binding site and recognition sequence. Collaborators at the Northern Institute for Cancer Research (NICR, Newcastle, UK) sought to characterise this binding interaction, through the controlled mutation of CRABPII, and investigation of its subsequent binding activity with synthetic retinoids and the cyclin D proteins. The single, double and triple-site mutations included residues notably changed between CRABPI and CRABPII, the ligand binding triad located in the centre of the fold (Kleywegt *et al.*, 1994, Vaezeslami *et al.*, 2006), as well as the pseudo nuclear localisation sequence residues (Sessler & Noy, 2005) found in the helical cap.

Binding of solvatochromic retinoid derivatives in an assay format offered the opportunity to gauge the binding activity of these mutant proteins in a quick and efficient manner. These data, coupled with structural insight from crystallography, will be used to determine how single point mutations change the behaviour of the CRABPII protein with respect to cyclin D3 binding.

6.2. Results

6.2.1. Binding of DC271 to CRABPII mutants

The retinoid binding behaviour of CRABPII mutant variants (Figure 6.1) will offer insight into how their in-solution activity is modified by the changes to the primary sequence.

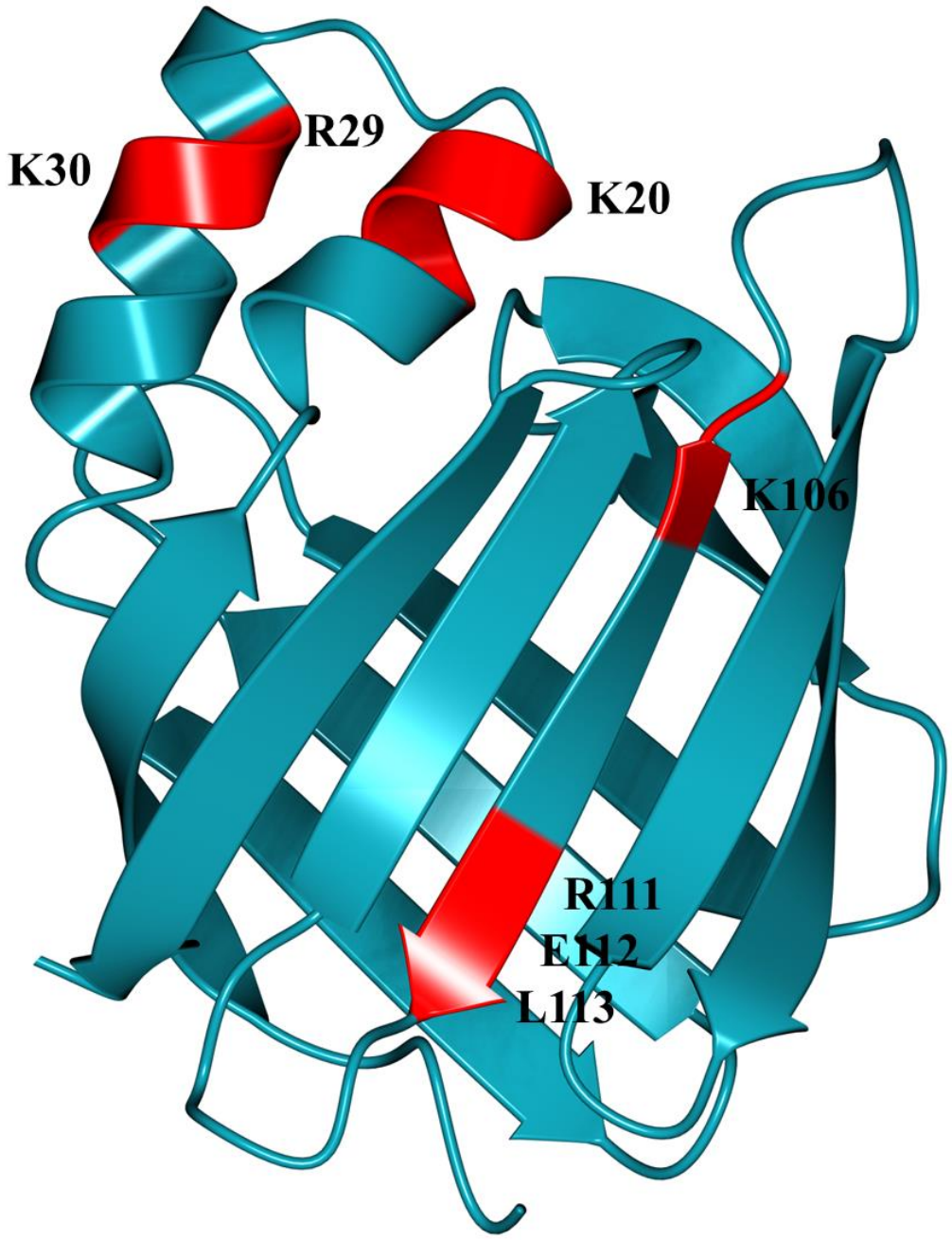


Figure 6.1: WT-hCRABPII (ribbon, dark blue) labelled with the sites of mutants (ribbon, red) used for binding studies. K20, R29 and K30 are externally facing and make up the nuclear localisation sequence that allows recognition by nuclear import proteins. (Sessler & Noy, 2005) R111 is part of the binding triad, and is internal to the ligand binding pocket - cf. Figure 1.4. (Kleywegt *et al.*, 1994) K106 is an external facing lysine residue.

Using the solvatochromic fluorescence method discussed previously, determination of whether each mutation damages, or enhances, retinoic acid binding ability (relative to the wild type) was carried out. Eight key mutants produced by the Endicott group were received as cryogenically frozen aliquots, and defrosted prior to assaying in each case. A sample of CRABPII wild type (treated in the same manner as the mutants) was included amongst these samples to act as a control. Carrying out each assay in a high-throughput 96-well format allowed for quick and efficient determination of binding for each mutant.

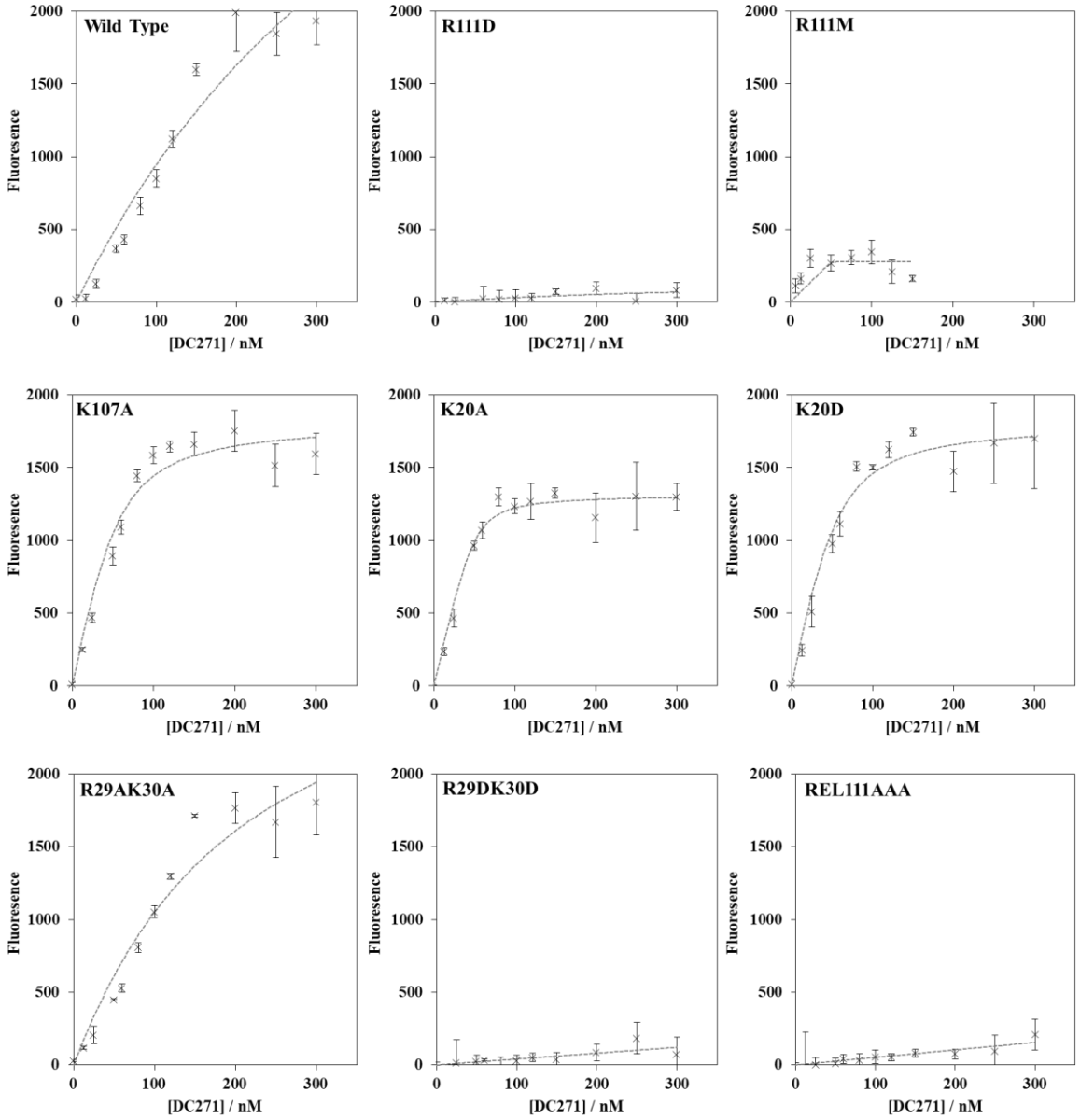


Figure 6.2: Plots of CRABPII wild type and mutant proteins binding to DC271 synthetic retinoid. High solvatochromic fluorescence from DC271 indicates the occupation of the binding site by the probe. R111D and R111M, R29DK30D and REL111AAA mutants display no binding activity. (n = 3 ± standard deviation). Curve fitting carried out using least squares fitting (Dynafit) (see 3.2).

Figure 6.2 shows the relative fluorescence activity of DC271 in the presence of each mutant, including wild type CRABPII supplied as a control and treated (frozen and thawed) in the same manner. It is immediately evident that R111D and R111M samples were not capable of binding to DC271, which by proxy suggests a general lack of retinoid binding activity. Modification of a key ligand binding triad residue was expected to eliminate retinoid binding, and had previously been used to capture an X-ray crystal structure of CRABPII in the absence of a ligand. (Chen *et al.*, 1998) Both aspartic acid (-ve) and methionine (neutral) change the charge as well as the steric bulk of the side chain relative to arginine (+ve); this is likely to have a severe impact on the hydrogen bonding network, location of water molecules that can stabilise the fold, and on the interactions with the carboxylic acid head group of ligands. Triple mutant REL111AAA also shows a failure to bind the ligand, suggesting that the modification of the ligand binding residue R111 to the much smaller (and neutral) alanine has a similar impact to the change in charge discussed above.

DC271 fluorescence in K20A, K20D, and K106A rose more sharply than in the case of the wild type, which may indicate a beneficial effect on the structure of the ligand binding site despite the externalised locations of the residues. Double mutants R29AK30A and R29DK30D show different activity, with R29AK30A consistent with the wild type, and R29DK30D showing a lack of binding activity. The behaviour of R29AK30A is consistent with previous retinoid affinity testing for R29/K30 mutants which used the intrinsic fluorescence of ATRA to detect the occupancy of the binding site. (Sessler & Noy, 2005)

6.2.2. X-ray crystal structure of R29AK30A

In order to understand the binding activity of selected CRABPII mutants, crystallisation trials were undertaken with samples of K20A, K20D, R29AK30A and R29DK30D (expression systems provided by the Endicott group, Newcastle NICR, UK). Of these, K20A and K20D provided crystalline material in robotic screening plates but did not diffract, and double mutant R29AK30A crystallised, was optimised, and diffracted under remote beamline conditions to 1.35 Å. The structure was solved using molecular replacement with CRABPII-DC360 (5OGB) to give the structure below (Figure 6.3).

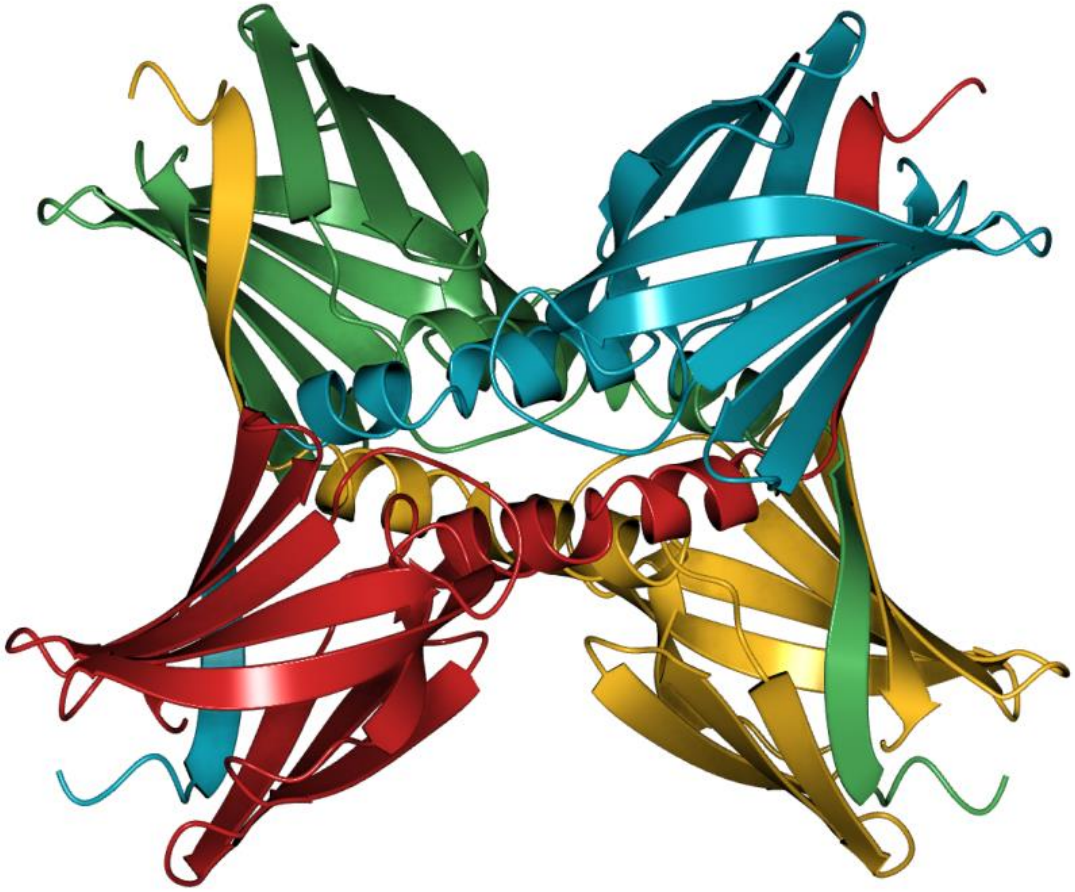


Figure 6.3: Crystal structure of CRABPII-R29AK30A, crystallised with four monomers in the asymmetric unit. All monomers presented in ribbon representation, A - blue, B - red, C - green, D - yellow. RMSD at Cα of each chain, relative to the A chain: B - 0.47 Å, C - 0.54 Å, D - 0.69 Å, calculated by superposition with CCP4mg.

The initial structural solution from Phaser (McCoy *et al.*, 2007) contained three molecules in the asymmetric unit, each consistent with wild type CRABPII. Estimates from the Matthews-coefficient based on the calculated unit cell placed four monomers in the asymmetric unit. A large region of empty space was present in the model calculated from phases of wild type CRABPII, and density fit problems and poorly connected features were identified around the 25-35 turn region of the helical cap in each monomer, pointing to significant structural deviation from the molecular

replacement model. Given the near-atomic resolution of the data, this was highly unexpected and pointed to a previously uncharacterised structure for CRABPII. A fourth copy of the protein was identified in the unallocated space and included. The helix-loop-helix section was rebuilt by removing the relevant sections and refining to reduce model bias from the canonical structure. It was identified that the N-terminus and first twenty five residues of each monomer had undergone a domain swap into the body of the monomer opposite it, creating dimerisation. This was remodelled residue by residue in the A chain using *Coot*, and the structure presented in Figure 6.3 generated by molecular replacement in Phaser. This configuration is consistent throughout the asymmetric unit, creating a dimer-of-dimers, with AB and CD pairs as labelled in Figure 6.3. The final R/R_{free} is 0.21/0.23, compared to 0.47/0.48 straight from Phaser, and 0.31/0.32 after insertion of a fourth monomer. The log likelihood gain (LLG), used by Phaser to gauge solved structures, was 111 in the initial structural solution, and 13508 using the domain-swapped N-terminus model. These factors, as well as the observable fit to the density, are strongly indicative that the proposed dimer-of-dimers structure is accurate.

Analysis of the A chain (Figure 6.4) shows that the dimerised structure largely mimics that of the wild type protein, with the exception that the N-termini have undergone a domain swap between the two monomers. In comparison to the helix-turn-helix motif observed in the wild type, the R29AK30A mutant displays a straightened section comprised almost entirely of α-helix. Residues G24-V25-N26 make up the 'turn' section of the wild type helix-turn-helix motif, whereas the second α-helix (α2) of the mutant proteins shifts three residues back to start at V25. The shift of the helix from residues 27-37 to 25-35 also increases the size of the non-structured chain after the motif, which must make a sharp turn into the first β-pleat. A shorter turn section means the helix-turn-helix becomes linear, and a longer trailing section makes the sharp turn into β-sheet implausible. The combined effect is displayed in the dimerised mutant: the N-terminus is extended because of the factors mentioned above, and naturally finds a suitable environment in the gap left in its opposite monomer. Whilst it cannot be empirically determined, the shift from hydrophillic lysine and arginine to hydrophobic alanine may be inducing a hydrophobic effect that makes the shifted helix conformation favourable over that of the wild type.

Figure 6.4: View of CRABPII-R29AK30A A chain (ribbon, dark blue) with residues 1-26 of chain B (ribbon, dark red) making a ‘complete’ CRABPII unit relative to the wild type. Mutation site for R29AK30A is highlighted in white (ribbon and cylinder), N and C termini are labelled in main chain colours.

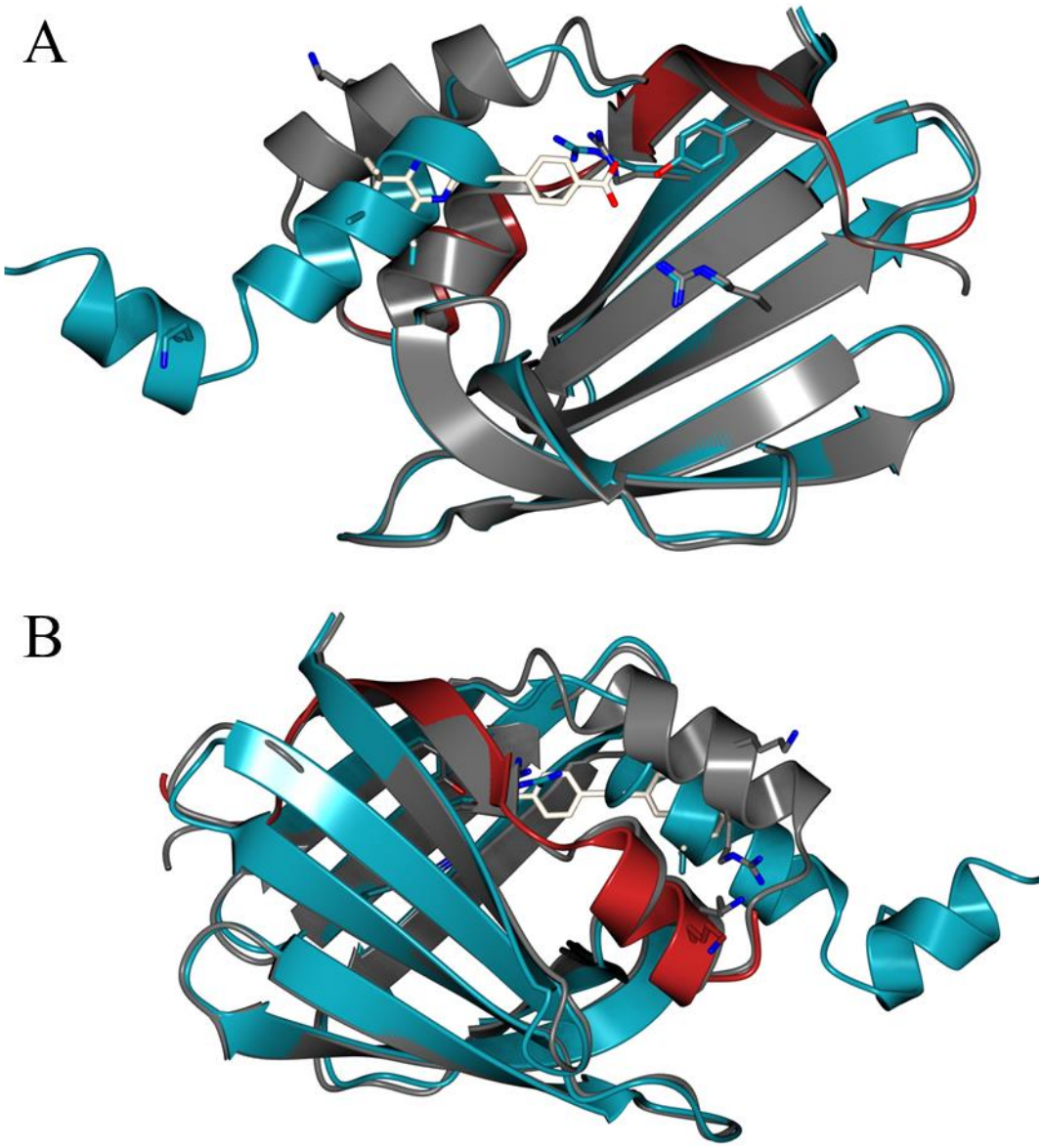


Figure 6.5: A - Superposition of CRABPII-R29AK30A (ribbon, A - dark blue, B - dark red) with CRABPII-DC645 (ribbon, grey), RMSD at Cα is 1.31 Å covering 106 residues. Cut-away of residues 40-67 in both models opens the binding site to view; binding triad (cylinder representations, main chain colours) show close alignment. DC645 (cylinder representation, white) highlights the location of the wild type binding pocket which is heavily occluded by residues 25-35 in the CRABPII-R29AK30A structure. B - Rear alignment shows that NLS sequence residues (wild type K20, R29, K30, cylinder representations, main chain colouring) are relocated by the mutation.

Mutation to the NLS-forming residues effectively collapses the binding site (Figure 6.5 - A), with the α2 helix of the helical cap folded into the pocket entrance relative to the wild type. The location of the nuclear localisation sequence residues (Figure 6.5 - B) is extremely perturbed, with the two alanine mutant residues formed into a straight helical section that would otherwise form the turn section of the helical cap. The ligand binding triad (Figure 6.5 - A) remains remarkably intact, with near identical locations for each residue. This, combined with evidence of the protein still being able to incorporate and bind DC271 (Figure 6.2 - panel 7), suggests that the dimer structure is not fixed in solution and that an 'induced fit' mechanism may still allow entry of the ligand to the

binding site. Under this paradigm, the seemingly labile N-terminus and residues up to N25 could be stabilised into forming the normal ligand binding site, once in the presence of a ligand, but would otherwise be mobile, allowing for formation of the structure present in the crystals. It is also plausible that the dimer conformation is present in solution, and the ligand binding pocket is expanded in the presence of a ligand, with it otherwise collapsed into the crystallised conformation.

6.3. Conclusions

Solvatochromic fluorescence demonstrates that half of the CRABPII mutants from the candidate pool are capable of retinoid binding, with R111M, R111D, and REL112A, which severely disrupt the architecture of the wild type ligand binding triad, being notable exceptions. Loop modified mutant R29DK30D also shows a lack of binding, in contrast to R29AK30A. It is likely that none of these mutants bind retinoids in solution and subsequently that the lack of a NLS would prevent traffic to the nucleus.

Crystallographic evidence highlights interesting behaviour from the R29AK30A mutant, in which key residues of the NLS are mutated, that demonstrates a domain-swapped dimer structure with a collapsed ligand binding pocket. Similar domain swapping events have been observed in mutations of the structurally related cellular retinol binding protein (CRBPII), although based on a larger section: the first sixty residues of the protein.(Assar *et al.*, 2016, Ghanbarpour *et al.*, 2019, Ghanbarpour *et al.*, 2020) The mutations used to generate the CRBPII dimers (and trimers) are dispersed further across the primary sequence than in CRABPII-R29AK30A, so it is possible with further mutation studies that the different dimerisation effects observed in one of the two proteins could be reproduced in the other. (Assar *et al.*, 2016, Ghanbarpour *et al.*, 2019, Ghanbarpour *et al.*, 2020) When considered alongside binding data, it is likely the collapsed-pocket dimer of CRABPII-R29AK30A is not the only structure present in solution, and rather the α2 helical section of the protein is somewhat flexible, able to adopt the collapsed conformation when crystallised, as well as provide a suitable binding site for retinoic acid. This could occur through one of two mechanisms: either an ‘induced fit’ effect which causes opening of the binding site when a ligand is present in solution, or a similar mechanism where dimerisation occurs once the ligand is already present in the interaction space, causing recruitment of the second monomer in a more suitable conformation. Crystallographic data collected by collaborators (Pastok *et al.*, 2020) shows a similar effect in mutant R29DK30D, but, rather than a dimeric structure, CRABPII-R29DK30D adopts a traditional fold, albeit with the α2 helix at residues 25-35 occluding the binding site in a deeper fashion than shown in Figure 6.5 - A. This adds credence to an ‘induced fit’ model, as the non-dimerised, deeply collapsed R29DK30D showed no affinity for DC271, whilst the dimerised and potentially more flexible R29AK30A showed binding comparable to the wild type. Neither mutant affecting the R29/K30 pair would be capable of travel to the nucleus due to disrupted or lacking nuclear localisation sequences (Sessler & Noy, 2005), so *in vivo* would likely result in reduced expression of retinoid related genes.

The CRABPII-cyclin interaction is not retinoic acid mediated (Despouy *et al.*, 2003), and so it is likely that if the NLS is involved in the interface between these two proteins, there is no distinction made between its externalised, retinoid bound form and the invert unbound form. However, HTRF interaction data for CRABPII-cyclin D3 suggests that both mutations altering the NLS (R29AK30A and R29DK30D) by removal of positively charged side chains totally disrupt cyclin

D3 binding. (Pastok *et al.*, 2020) It is likely that conformational changes related to these mutations form part of this loss of function; the collateral damage done to the folding of the helical cap by mutations to the NLS impacts surface interactions with cyclin D3. The correlation of this binding and structural data with information about CRABPII-cyclin D3 interactions will allow for increased understanding of the how the interface between the two proteins occurs.

6.4. Methods

6.4.1. Expression of CRABPII mutants R29AK30A, R29DK30D, K20A and K20D

CRAPBII protein (GST-tagged) was expressed in BL21(DE3) competent *E. coli* cells (NEB) transformed with a pGEX-GST-HRV3c-hCRABPIIR29AK30A vector. Starter cultures (20 g/L LB broth, 100 μg/ml ampicillin, 25 ml) were grown overnight and inoculated into 1 L LB media grown at 37 °C. Expression was induced at OD_{600} 0.6 with IPTG (1 mM in culture), before shaking overnight (20 h). The resulting cultures were pelleted using an Avanti Hi-Speed centrifuge (JLA 8.1000, 4000 rpm, 25 min, 4 °C), before removal of supernatant and freezing at −80 °C.

6.4.2. Purification of CRABPII mutants R29AK30A, R29DK30D, K20A and K20D

Defrosted cell pellets were resuspended in 30 ml of lysis buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 10 mM imidazole). The resulting suspension was sonicated (2 min, 40% power, on ice) and centrifuged (Avanti Hi-Speed JLA 25.50, 20,000 rpm, 50 min, 4 °C). The supernatant was loaded onto a GST affinity column (1 ml, GE healthcare) and washed using 10 ml lysis buffer. The column was then loaded with 1 ml HRV3C protease (Sigma, 20 units) in buffer and refrigerated for 16 h (4 °C) to cleave the GST tag. The cleaved protein was eluted using standard buffer (3 x 1 ml), and bound GST tag was eluted with reduced glutathione (15 mM in buffer, 10 ml). The resulting protein fractions were analysed by SDS PAGE (Appendix 18) and concentration estimated by absorbance at 280 nm using a Thermo Scientific Nanodrop A_{280}.

6.4.3. DC271-CRABPII mutant binding

Solutions of DC271 (600 nM - 25 nM, <1% EtOH) and CRABPII mutants (100 nM in 20 mM HEPES, 300 mM NaCl, pH 7.5, from cryo-frozen samples) were prepared and combined (50 μl + 50 μl) in a Corning non-binding surface black fluorescence 96-well plate, using HEPES buffer in place of CRABPII where necessary for controls. The plate was spun for 2 min (1500 rpm) to ensure incorporation, and fluorescence read at ex/em 355/440 nm using a Synergy H4 plate reader. The total volume per well was 100 μl, the final concentration of protein was 50 nM.

6.4.4. Crystallisation of CRABPII R29DK30D

Conditions for crystal growth were determined by screening using commercially available crystallisation kits and purified CRABPII-R29AK30A. Screening was carried out using a mosquito Xtal3 robot (SPT Labtech) and sitting drop experiments (100+100 nl and 200+100 nl protein+reservoir droplet sizes), and optimised in sitting drop experiments (400 μl reservoir solutions) with varying drop sizes (1+1 μl and 2+1 μl protein+reservoir). The resulting crystals were inspected and mounted for diffraction under a Leica MZ16 stereomicroscope using standard uni-puck style pins before flash freezing in $N_{2(l)}$. (Teng, 1990)

CRABPII-R29AK30A was crystallised in 0.15 M magnesium chloride hexahydrate, 0.1 M Bis-Tris pH 5.5, 25% PEG 3350 based on optimisation of Molecular Dimensions JCSG-plus™ HT-96 ECO screen and SG1™ HT-96 ECO screen hits. (Page *et al.*, 2003, Newman *et al.*, 2005, Fazio *et al.*, 2014)

6.4.5. Data collection

Diamond Light Source I24 beamline was used for data collection as part of the BAG program. Diffraction was recorded on a Pilatus3 6M detector. (Henrich *et al.*, 2009) Further details are summarised in the crystallographic tables (Appendix 19 and Appendix 20).

6.4.6. Data Processing

Data processing was carried out using the CCP4 suite of programs, and DIALS (Winter *et al.*, 2018) as part of the autoprocessing (Winter, 2010) done at Diamond Light Source. Scaling was carried out with aimless/pointless (Evans & Murshudov, 2013), phasing via molecular replacement with Phaser (McCoy *et al.*, 2007), and manual modifications made with *Coot.* (Emsley *et al.*, 2010) REFMAC5 was used for refinement. (Murshudov *et al.*, 1997, Murshudov *et al.*, 1999, Winn *et al.*, 2003, Vagin *et al.*, 2004, Murshudov *et al.*, 2011, Nicholls *et al.*, 2012)

Local NCS restraints generated by REFMAC5 were used to improve the refinement of the dimer of dimers. (Uson *et al.*, 1999, Murshudov *et al.*, 2011)

Alignments and figure making were carried out with CCP4mg. (McNicholas *et al.*, 2011)

7: Conclusions and Further Work

7.1. Solvatochromism of synthetic retinoids

The first series of DC synthetic retinoids were designed to enhance and shift fluorescence emission into the visible range for use as probes *in vivo*. By substituting a di/tetrahydroquinoline tail group in place of the dihydronapthalene used in EC23, a bathochromic shift and increase in quantum yield were achieved. (Chisholm *et al.*, 2019) These retinoid derivatives proved solvatochromic, both with respect to solvent (Chisholm *et al.*, 2019), and by titration with the heterologously expressed protein CRABPII. The increase in their fluorescence signal caused by a hydrophobic environment makes them ideal indicators of the occupancy of the buried, hydrophobic locations such as the CRABPII binding site. The binding affinities calculated for each retinoid molecule highlight that structural diversity has a large impact on binding to CRABPII. Deviation from the core EC23 design by lengthening or shortening the molecule had a negative impact on binding affinity; eliminating and reducing binding by two orders of magnitude respectively. DC479, which maintained the same overall geometry as EC23 but with a flexible alkene carboxylic acid, was observed as binding CRABPII with greatest affinity of all the tested compounds. Conversely, increasing compound flexibility by use of tetrahydroquinoline *vs* dihydroquinoline (DC271 *vs* DC360) reduced the calculated K_d value. Finally, key substituents to the tail of the molecule were observed to generate a strong binding interaction; the tertiary amine with *para* gem-dimethyl group in DC271 exemplifies this.

As a result of a tight binding affinity for CRABPII (32.3 nM), high absolute fluorescence, and abundant availability for testing, DC271 was selected for incorporation into a 96-well assay as an indicator of binding site occupancy. The addition of a second compound was demonstrated to reduce the fluorescence signal of the mixture in a concentration-linked manner, as DC271 was displaced from the binding site into the aqueous environment where its fluorescence is quenched. This key characteristic is the core of the solvatochromic fluorescence binding assay used to investigate the binding properties of other, non-fluorescent retinoids. Testing with ATRA, 9-*cis* and 13-*cis* retinoic acids, and synthesis fragments from synthetic derivatives, demonstrated that the assay performed as expected, only responding to the known CRABPII ligands ATRA and 13-*cis*-RA. Additional testing demonstrated that DC271 was also solvatochromic with respect to CRABPI-L29C, a known effector of non-genomic retinoid signalling (Masiá *et al.*, 2007, Rochette-Egly, 2015, Park *et al.*, 2018, Park *et al.*, 2019, Persaud *et al.*, 2013) later used in crystallographic studies of binding interactions. The fluorescence displacement effect can be used qualitatively to simply indicate a positive binding event, or quantitatively by determination of binding affinity via a concentration *vs* effect curve.

The development of a solvatochromic fluorescence displacement assay was successful, and provided a key tool for understanding the interactions of retinoid binding proteins and synthetic retinoid molecules. The method is simple, and relies on small volumes of reagents and basic lab equipment to carry out. Unlike other methods, there is no separation of bound/unbound ligand required, and the measurement recorded is a direct result of a binding event, not a separate equilibrium interaction. Using standard 96-well plates, the assay can be employed in medium and high-throughput testing of compounds, scaled based on the automation available to the user.

7.2. Co-crystallisation of synthetic retinoids with CRABPII and CRABPI-L29C

In order to fully characterise the interactions of DC retinoids with CRABPII and CRABPI-L29C, co-crystallisation experiments were performed using a simple equimolar combination of protein and ligand. X-ray crystal structures of CRABPII were determined with five synthetic retinoids: DC360 (1.8 Å resolution), DC271 (1.5 Å), EC19 (2.8 Å), DC479 (1.8 Å) and DC645 (1.7 Å), in a variety of crystal forms. The co-crystal structure of CRABPI-L29C was also determined with DC645 (2.4 Å), and with fatty acids MYR and TDA (1.6 Å) resulting from heterologous expression in *E. coli*. Many of these structures represent the protein-ligand interaction at near atomic resolution, and allow for interpretation of binding interactions in detail.

The ligand binding site of CRABPII consists of an enclosed hydrophobic pocket created as a result of an orthogonal β-sheet fold well conserved amongst fatty acid binding proteins. (Kleywegt *et al.*, 1994) A pair of α-helices cap the fold and are reordered as a result of ligand binding. In each case, the co-crystal structures of DC ligands with CRABPII displayed typical binding behaviour, increasing the order observed in the α2 helix section (Chen *et al.*, 1998) and forming appropriate nuclear localisation sequences. (Sessler & Noy, 2005) This formation of an externally facing, negatively charged NLS surface is the key indicator of retinoid-like activity observable in crystallised CRABPII. Each structure (bar CRABPII-EC19) additionally demonstrated canonical interactions with the ligand binding triad residues R112, Y133, and R135. The hydrogen bonding interactions between the carboxylic acid moiety of the ligands and these residues, as well as a conserved water molecule, are key indicators of proper ligand binding site incorporation. In the case of CRABPII-EC19, the modified 13-*cis*-like structure was prevented from hydrogen bonding due to its inflexible nature, and the close contacts of the ligand binding site. Given the formation of a NLS, it is likely that EC19 would still be trafficked to the nucleus.

hCRABPI shares high sequence and structural homology with hCRABPII but is not capable of nuclear transport. (Budhu & Noy, 2002, Sessler & Noy, 2005) The hCRABPI-L29C mutant was observed to improve crystallisation *vs* the wild type, allowing data collection to near-atomic resolution; ostensibly due to disulphide bridge formation and subsequent dimerisation through the mutated residue. Co-crystal structures with DC645, a key compound implicated in neurite outgrowth of pluripotent stem cells, and with fatty acids myristic acid and tridecanoic acid resulting

from expression, were refined at 2.4 Å and 1.6 Å, respectively - both improvements over the current best deposited structure of a mouse homolog at 2.7 Å. (Thompson *et al.*, 1995) Both structures presented as dimers; in the case of the fatty acid-bearing structure, each monomer contained electron densities that were modelled as MYR and TDA, respectively. The CRABPI-L29C-DC645 structure is made of a single ligand bound monomer, dimerised to a second monomer with an empty binding site. Both structures present a typical binding interaction, each molecule interacting with the binding triad in the expected manner. As a result of the ligand-bound/unoccupied dimerisation observed in CRABPI-L29C-DC645, it was possible to identify four β-sheet residues (R45, D47, Q50 and N65) that differ in their surface orientation between the two monomers (ligand bound *vs* unbound). This approximate location has been implicated by NMR studies in binding of CRABPI to non-genomic signalling partners (Park *et al.*, 2019), and so offers a potential avenue for further study.

As a result of co-crystallisation experiments and structural determination, it has been demonstrated that DC retinoids interact with CRAB proteins in the canonical manner. This is a key piece of evidence in validating the mode of action of the fluorescence displacement assays, and offers important insights into protein-ligand interactions of both CRABPI and CRABPII.

7.3. Applications of fluorescence displacement assays

Solvatochromic fluorescence displacement assays were used to characterise the binding of DC645 to both CRABPI-L29C and CRABPII. DC645 has shown great potential as an agonist of retinoid signalling, both via genomic and non-genomic pathways and, as a result, is a compound of particular interest. (Khatib *et al.*, 2020) The binding curves resulting from displacement of fluorescent DC271 show that DC645 is a mid-strength binder of CRABPII (0.25 μM), albeit tenfold weaker than other synthetic derivatives, and approximately tenfold weaker again at binding to CRABPI-L29C (1.94 μM). These results tally with the use of dihydroquinoxaline as a tail group in the compound; introducing additional hydrogen bonding potential into an area of molecule usually responsible for hydrophobic interactions could be expected to reduce overall affinity.

The binding interactions of a series of non-fluorescent DC compounds with both RARα and RARγ offer insights into how structural modifications to a common core can affect ligand affinity. Of the twenty three compounds tested, sixteen were found to have affinity for the RAR proteins, in the nanomolar range (0.8 - 800 nM). Six of sixteen were found to only bind one RAR isoform (five to RARα and one to RARγ). The remaining compounds showed varying levels of dual affinity, with a general trend towards RARα selectivity. Interpretation of these results suggests that specific modifications have large impacts on affinity to the two proteins. Ether groups in the tail location were found to generally reduce binding affinity, likely due to increased hydrogen bonding opportunities not present in the parent molecule EC23. Similarly, introduction of hydrogen bond donors through the use of pyrazine-carboxylic acids (head group) or quinoxaline (tail) evidently

reduces overall binding affinities for similar reasons. Fluorination at the head group was seen to generally increase affinity for RARα, based on the compounds in the series, though this effect was reduced by the presence of tail methylation. The thiazole-based compounds tested, with their overall bent topology, were generally less successful binders than the linear EC23 derivatives. These interpretations, based on the available data, are valid, but the design of future compounds should be based on holistic consideration of additional factors. In particular, the evidence from cell differentiation assays is vital for determining which compounds will ultimately be most effective as moderators of retinoid signalling, as any compound with sub-micromolar affinity could be considered 'sufficient' to trigger downstream signalling effects. Additional analysis of compound binding to other proteins, and RARβ in particular, will be welcome as part of any decision making process.

The characterisation of compounds binding to signalling proteins is an important factor in structure-led design of molecules, and has been achieved here using solvatochromic fluorescence assays. One key compound, DC645, has been characterised binding to both CRABPI-L29C and CRABPII. The additional discovery of chemical features which promote specific binding to one RAR isoform over another, based on data from existing compounds, will aid in development of new candidates going forwards.

7.4. Characterising CRABPII mutants

Seeking to structurally characterise the interaction of CRABPII with cell cycle control protein cyclin D3, the Endicott group at the NICR - Newcastle designed a series of mutant proteins with specific focus on surface exposed and binding triad residues. (Despouy *et al.*, 2003, Pastok *et al.*, 2020) Using the solvatochromic fluorescence of DC271, it was possible to qualitatively determine the retinoid binding ability of each of these mutants in comparison to the wild type. Mutations to the binding triad residues were found to largely disrupt compound binding, likely due to different charge interactions or steric positioning of the modified side chains. Surface mutations made less impact on ligand interactions in general, though R29DK30D mutations to the NLS disrupted binding.

Based on this, and collaborator's HTRF interaction data with the cyclin D3 protein, CRABPII R29AK30A was selected for crystallisation studies. A structure at atomic resolution (1.35 Å) was determined for the mutant, and revealed a novel geometry not previously characterised in CRABPII. The structure comprised a dimer-of-dimers, with each pair of monomers having undergone a domain swap *via* the first thirty five residues starting at the N terminus. This is proposed to be a result of the mutation to alanine of two key residues in the helical cap. Reordering of the α2 helix is observed, reducing the loop length of the helix-turn-helix motif and preventing it from forming. Instead, a linear section is incorporated into the body of the dimer mate creating the overall domain swap effect. Whilst the overall structure is broadly consistent with wild type

CRABPII morphology, the α2 helix is collapsed into the hydrophobic pocket, likely to satisfy the hydrophobic interactions of the two mutant residues. Based on the fact that DC271 binding was still observed, it is suggested that this effect is reversible, or that an induced fit model causes rearrangement of the structure in the presence of a suitable ligand allowing it to enter the hydrophobic binding site. Additional evidence from collaborators shows that a second CRABPII double mutant, R29DK30D, also adopts a collapsed helix conformation, without a domain swapping effect (Pastok *et al.*, 2020), although since no binding to DC271 was detected it is likely this represents a more rigid conformation.

The data collected for CRABPII mutants binding to DC271 have helped to better characterise a novel interaction between the protein and cyclin D3, a cell cycle regulator. HTRF data suggests that an intact CRABPII NLS sequence is necessary for cyclin D3 interaction, although as the interaction is not RA mediated (Despouy *et al.*, 2003) it is likely that full formation of the sequence by externalisation of R29 is not necessary. Based on crystallographic structural evidence of CRABPII-R29AK30A it is likely that a disrupted structure, and formation of domain swapped dimer, contributes to this loss of binding function. The correlation of both retinoid binding and structural data with CRABPII/cyclin D3 interaction data has created increased understanding of the interaction between the two proteins.

7.5. Future work

The development of a powerful tool for interpreting ligand binding to retinoic acid binding proteins is a key success of the project, and will aid in development of new lead compounds and potential treatments for underserved neurological and oncological diseases. The fluorescence displacement assay is backed up with rigorous structural evidence demonstrating ligand binding to CRABPII and CRABPI-L29C in the canonical manner. Deployment of the assay in determining binding of synthetic retinoids to both CRAB and RAR proteins has demonstrated efficacy, and produced a data set that reveals key structural elements responsible for changes to binding affinities. Determination of the binding of DC271 to CRABPII mutants, as a proxy for retinoids in general, has aided in the validation and structural characterisation of a novel protein-protein interaction between CRABPII and cyclin D3. A new domain swapped dimer structure of CRABPII has additionally been determined as part of that study.

The general success of the project creates opportunities to further increase the understanding of the structural and behavioural characteristics of the retinoic acid signalling pathway and retinoid binding proteins specifically. One key avenue for further research is the interaction surface used by CRABPI as part of the non-genomic signalling pathway. A general area for the interaction between CRABPI and the Raf kinases has previously been highlighted as part of NMR based studies. (Park *et al.*, 2019) The determination of a mixed ligand-bound and unoccupied dimer of CRABPI-L29C in this project has highlighted several residues within this general location that change orientation

upon ligand binding. This offers a jumping off point for further investigation of interactions between the two proteins, and a starting point in terms of residues to focus on for mutational or crystallographic studies.

Additional characterisation of the interactions of DC compounds with RARβ would be welcome, as the protein forms a key additional element in the nuclear portion of the retinoic acid signalling pathway. Determination of binding constants for the library of compounds tested against RARα and RARγ would allow for in depth interpretation of chemical space modifications, as well as correlation with phenotype assays to further enhance the understanding of each isotype role in the signalling system. Furthermore, investigation of the retinoid-X receptors (RXRs) should easily be within the capabilities of the assay system, and proof of this would open up another large field of investigation. The RXRs are somewhat more promiscuous that the RARs, interacting with other nuclear receptors, and offer another avenue for potential control over retinoid signalling and, more generally, the growth and development of cells.

Finally, co-crystallisation studies of the RAR proteins in concert with DC retinoids would further validate the solvatochromic florescence displacement assay, and offer greater insight into how the compounds are oriented the ligand binding site. Understanding ligand morphology, and specific interactions with the binding site, will allow for better interpretation of binding data. Matching the modifications made in a compound to specific interactions in the ligand binding site will make interpretation of K_d values more powerful and allow for development of new compounds with potentially greater specificity.

As a whole, there are still many avenues to explore that will create better understanding of the retinoid signalling pathway, allowing for greater overall control, modification of signalling, and design of compounds to enact these modifications. The understanding brought about by the work presented here will be key in the development of new and potent affecters of retinoid signalling, towards the treatment of critically underserved pathologies.

References

Abu-Abed, S., Dollé, P., Metzger, D., Beckett, B., Chambon, P. & Petkovich, M. (2001), The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev.* **15**, 226-240.

Achkar, C. C., Derguini, F., Blumberg, B., Langston, A., Levin, A. A., Speck, J., Evans, R. M., Bolado, J., Nakanishi, K. & Buck, J. (1996), 4-Oxoretinol, a new natural ligand and transactivator of the retinoic acid receptors. *Proc. Natl. Acad. Sci. USA* **93**, 4879-4884.

Adams, P. D., Sellers, W. R., Sharma, S. K., Wu, A. D., Nalin, C. M. & Kaelin, W. G., Jr. (1996), Identification of a cyclin-cdk2 recognition motif present in substrates and p21-like cyclin-dependent kinase inhibitors. *Mol. Cell. Biol.* **16**, 6623-6633.

Al Tanoury, Z., Piskunov, A. & Rochette-Egly, C. (2013), Vitamin A and retinoid signaling: genomic and nongenomic effects Thematic Review Series: Fat-Soluble Vitamins: Vitamin A. *J. Lipid Res.* **54**, 1761-1775.

Allenby, G., Bocquel, M.-T., Saunders, M., Kazmer, S., Speck, J., Rosenberger, M., Lovey, A., Kastner, P., Grippo, J. F. & Chambon, P. (1993), Retinoic acid receptors and retinoid X receptors: interactions with endogenous retinoic acids. *Proc. Natl. Acad. Sci. USA* **90**, 30-34.

Allenby, G., Janocha, R., Kazmer, S., Speck, J., Grippo, J. F. & Levin, A. A. (1994), Binding of 9-cis-retinoic acid and all-trans-retinoic acid to retinoic acid receptors alpha, beta, and gamma. Retinoic acid receptor gamma binds all-trans-retinoic acid preferentially over 9-cis-retinoic acid. *J. Biol. Chem.* **269**, 16689-16695.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990), Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Antille, C., Tran, C., Sorg, O., Carraux, P., Didierjean, L. & Saurat, J.-H. (2003), Vitamin A exerts a photoprotective action in skin by absorbing ultraviolet B radiation. *J. Investig. Dermatol.* **121**, 1163-1167.

Assar, Z., Nossoni, Z., Wang, W., Santos, E. M., Kramer, K., McCornack, C., Vasileiou, C., Borhan, B. & Geiger, J. H. (2016), Domain-Swapped Dimers of Intracellular Lipid-Binding Proteins: Evidence for Ordered Folding Intermediates. *Structure* **24**, 1590-1598.

Assenberg, R., Wan, P. T., Geisse, S. & Mayr, L. M. (2013), Advances in recombinant protein expression for use in pharmaceutical research. *Curr. Opin. Struct. Biol.* **23**, 393-402.

Balmer, J. E. & Blomhoff, R. (2002), Gene expression regulation by retinoic acid. *J. Lipid Res.* **43**, 1773-1808.

Baneyx, F. (1999), Recombinant protein expression in Escherichia coli. *Curr. Opin. Biotechnol.* **10**, 411-421.

Barnard, J. H., Collings, J. C., Whiting, A., Przyborski, S. A. & Marder, T. B. (2009), Synthetic retinoids: structure–activity relationships. *Chem. Eur. J.* **15**, 11430-11442.

Bastien, J. & Rochette-Egly, C. (2004), Nuclear retinoid receptors and the transcription of retinoid-target genes. *Gene* **328**, 1-16.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. (2011), iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **67**, 271-281.

Bennani, Y. L. & Boehm, M. F. (1995), Syntheses of High Specific Activity 2, 3-and 3, 4-[3H] 2-9-cis-Retinoic Acid. *J. Org. Chem.* **60**, 1195-1200.

Bilip, M., Shah, S., Mathiyalakan, M., Tagalakis, A. D., Hart, S. L., Maeshima, R., Eaton, S., Orford, M., Irving, E., Di Florio, A., Simons, C. & Stoker, A. W. (2020), Liposomal delivery of hydrophobic RAMBAs provides good bioavailability and significant enhancement of retinoic acid signalling in neuroblastoma tumour cells. *J. Drug Targeting* **28**, 643-654.

Bird, L. E. (2011), High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. *Methods* **55**, 29-37.

Blaner, W. S., O'Byrne, S. M., Wongsiriroj, N., Kluwe, J., D'Ambrosio, D. M., Jiang, H., Schwabe, R. F., Hillman, E. M. C., Piantedosi, R. & Libien, J. (2009), Hepatic stellate cell lipid droplets: A specialized lipid droplet for retinoid storage. *Biochim. Biophys. Acta* **1791**, 467-473.

Blomhoff, R. (1994). *Vitamin A in health and disease*. CRC Press.

Boehm, M. F., McClurg, M. R., Pathirana, C., Mangelsdorf, D., White, S. K., Hebert, J., Winn, D., Goldman, M. E. & Heyman, R. A. (1994), Synthesis of high specific activity tritium-labeled [3H]-9-cis-retinoic acid and its application for identifying retinoids with unusual binding properties. *J. Med. Chem.* **37**, 408-414.

Boylan, J. F. & Gudas, L. J. (1992), The level of CRABP-I expression influences the amounts and types of all-trans-retinoic acid metabolites in F9 teratocarcinoma stem cells. *J. Biol. Chem.* **267**, 21486-21491.

Broennimann, C., Eikenberry, E., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M. & Tomizaki, T. (2006), The PILATUS 1M detector. *J. Synchrotron Radiat.* **13**, 120-130.

Brünger, A. T. (1997). *Methods Enzymol.*, pp. 366-396: Academic Press.

Budhu, A., Gillilan, R. & Noy, N. (2001), Localization of the RAR interaction domain of cellular retinoic acid binding protein-II. *J. Mol. Biol.* **305**, 939-949.

Budhu, A. S. & Noy, N. (2002), Direct channeling of retinoic acid between cellular retinoic acid-binding protein II and retinoic acid receptor sensitizes mammary carcinoma cells to retinoic acid-induced growth arrest. *Mol. Cell. Biol.* **22**, 2632-2641.

Bull, J. N., West, C. W., Anstöter, C. S., da Silva, G., Bieske, E. J. & Verlet, J. R. R. (2019), Ultrafast photoisomerisation of an isolated retinoid. *Phys. Chem. Chem. Phys.* **21**, 10567-10579.

Burri, B. J., La Frano, M. R. & Zhu, C. (2016), Absorption, metabolism, and functions of β-cryptoxanthin. *Nutr. Rev.* **74**, 69-82.

Cahnmann, H. (1995), A fast photoisomerization method for the preparation of tritium-labeled 9-cis-retinoic acid of high specific activity. *Anal. Biochem.* **227**, 49-53.

Casanas, A., Warshamanage, R., Finke, A. D., Panepucci, E., Olieric, V., Nöll, A., Tampé, R., Brandstetter, S., Förster, A., Mueller, M., Schulze-Briese, C., Bunk, O. & Wang, M. (2016), EIGER detector: application in macromolecular crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **72**, 1036-1048.

Chambon, P. (1996), A decade of molecular biology of retinoic acid receptors. *FASEB J.* **10**, 940-954.

Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T. P. & Rastinejad, F. (2008), Structure of the intact PPAR-γ–RXR-α nuclear receptor complex on DNA. *Nature* **456**, 350-356.

Chandra, V., Wu, D., Li, S., Potluri, N., Kim, Y. & Rastinejad, F. (2017), The quaternary architecture of RARβ–RXRα heterodimer facilitates domain–domain signal transmission. *Nat. Commun.* **8**, 868.

Chebaro, Y., Sirigu, S., Amal, I., Lutzing, R., Stote, R. H., Rochette-Egly, C., Rochel, N. & Dejaegere, A. (2017). Allosteric Regulation in the Ligand Binding Domain of Retinoic Acid Receptorγ, Vol. 12, PLOS ONE, p. e0171043.

Chen, X., Tordova, M., Gilliland, G. L., Wang, L., Li, Y., Yan, H. & Ji, X. (1998), Crystal structure of apo-cellular retinoic acid-binding protein type II (R111M) suggests a mechanism of ligand entry. *J. Mol. Biol.* **278**, 641-653.

Cherbas, L. & Cherbas, P. (2007), Drosophila Cell Culture and Transformation. *Cold Spring Harb. Protoc.* **2007**, pdb.top6.

Chisholm, D. (2017). Thesis, Durham University.

Chisholm, D. R., Tomlinson, C. W., Zhou, G.-L., Holden, C., Affleck, V., Lamb, R., Newling, K., Ashton, P., Valentine, R., Redfern, C., Erostyakgh, J., Makkaight, G., Ambler, C. A., Whiting, A. & Pohl, E. (2019), Fluorescent retinoic acid analogues as probes for biochemical and intracellular characterization of retinoid signaling pathways. *ACS Chem. Biol.* **14**, 369-377.

Chisholm, D. R. & Whiting, A. (2020). *Methods Enzymol.*, edited by E. Pohl, pp. 453-491: Academic Press.

Chisholm, D. R., Zhou, G.-L., Pohl, E., Valentine, R. & Whiting, A. (2016), Practical synthetic strategies towards lipophilic 6-iodotetrahydroquinolines and-dihydroquinolines. *Beilstein J. Org. Chem.* **12**, 1851.

Christie, V. B., Barnard, J. H., Batsanov, A. S., Bridgens, C. E., Cartmell, E. B., Collings, J. C., Maltman, D. J., Redfern, C. P., Marder, T. B., Przyborski, S. & Whiting, A. (2008), Synthesis and evaluation of synthetic retinoid derivatives as inducers of stem cell differentiation. *Org. Biomol. Chem.* **6**, 3497-3507.

Comley, J. (2015), PHENOTYPIC DRUG DISCOVERY. *Drug Discovery* **16**, 25.

Corcoran, J. & Maden, M. (1999), Nerve growth factor acts via retinoic acid synthesis to stimulate neurite outgrowth. *Nat. Neurosci.* **2**, 307-308.

Cordeiro, T. N., Sibille, N., Germain, P., Barthe, P., Boulahtouf, A., Allemand, F., Bailly, R., Vivat, V., Ebel, C., Barducci, A., Bourguet, W., le Maire, A. & Bernadó, P. (2019), Interplay of Protein Disorder in Retinoic Acid Receptor Heterodimer and Its Corepressor Regulates Gene Expression. *Structure* **27**, 1270-1285.e1276.

Cowan, A. J., Stevenson, P. A., Gooley, T. A., Frayo, S. L., Oliveira, G. R., Smith, S. D., Green, D. J., Roden, J. E., Pagel, J. M. & Wood, B. L. (2017), Results of a phase I-II study of fenretinide and rituximab for patients with indolent B-cell lymphoma and mantle cell lymphoma. *Br. J. Haematol.*

Cummings, J. L., Morstorf, T. & Zhong, K. (2014), Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimer's Res. Ther.* **6**, 37.

Cunningham, T. J. & Duester, G. (2015), Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat. Rev. Mol. Cell Biol.* **16**, 110-123.

D'Ambrosio, D. N., Clugston, R. D. & Blaner, W. S. (2011), Vitamin A Metabolism: An Update. *Nutrients* **3**, 63-103.

Dawson, M. I., Chan, R. L. S., Derdzinski, K., Hobbs, P. D., Chao, W. R. & Schiff, L. J. (1983), Synthesis and pharmacological activity of 6-[(E)-2-(2,6,6-trimethyl-1-cyclohexen-1-yl)ethen-1-yl)]- and 6-(1,2,3,4-tetrahydro-1,1,4,4-tetramethyl-6-naphthyl)-2-naphthalenecarboxylic acids. *J. Med. Chem.* **26**, 1653-1656.

de Urquiza, A. M., Liu, S., Sjöberg, M., Zetterström, R. H., Griffiths, W., Sjövall, J. & Perlmann, T. (2000), Docosahexaenoic acid, a ligand for the retinoid X receptor in mouse brain. *Science* **290**, 2140-2144.

Degorce, F., Card, A., Soh, S., Trinquet, E., Knapik, G. P. & Xie, B. (2009), HTRF: a technology tailored for drug discovery–a review of theoretical aspects and recent applications. *Curr. Chem. Genom. Translat. Med.* **3**, 22.

del Mar Vivanco-Ruiz, M., Tiollais, P., Stunnenberg, H. & Dejean, A. (1990), Identification of a Retinoic Acid Responsive Element in the Retinoic Acid Receptor (Beta) Gene. *Nature* **343**, 177.

Delva, L., Bastie, J. N., Rochette-Egly, C., Kraïba, R., Balitrand, N., Despouy, G., Chambon, P. & Chomienne, C. (1999), Physical and functional interactions between cellular retinoic acid binding protein II and the retinoic acid-dependent nuclear complex. *Mol. Cell. Biol.* **19**, 7158-7167.

Despouy, G., Bastie, J.-N., Deshaies, S., Balitrand, N., Mazharian, A., Rochette-Egly, C., Chomienne, C. & Delva, L. (2003), Cyclin D3 Is a Cofactor of Retinoic Acid Receptors, Modulating Their Activity in the Presence of Cellular Retinoic Acid-binding Protein II. *J. Biol. Chem.* **278**, 6355-6362.

Devery, J. & Milborrow, B. V. (1994), beta-Carotene-15,15'-dioxygenase (EC 1.13.11.21) isolation reaction mechanism and an improved assay procedure. *Br J Nutr* **72**, 397-414.

di Guana, C., Lib, P., Riggsa, P. D. & Inouyeb, H. (1988), Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* **67**, 21-30.

Diederichs, K. & Karplus, P. A. (1997), Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269-275.

DiGiovanna, J. J., Zech, L. A., Ruddel, M. E., Gantt, G. & Peck, G. L. (1989), Etretinate: persistent serum levels after long-term therapy. *Arch. Dermatol.* **125**, 246-251.

Dilworth, F. J. & Chambon, P. (2001), Nuclear receptors coordinate the activities of chromatin remodeling complexes and coactivators to facilitate initiation of transcription. *Oncogene* **20**, 3047.

DiPietrantonio, A. M., Hsieh, T. C., Juan, G., Traganos, F., Darzynkiewicz, Z. & Wu, J. M. (2000), Fenretinide-induced caspase 3 activity involves increased protein stability in a mechanism distinct from reactive oxygen species elevation. *Cancer Res.* **60**, 4331-4335.

Dong, D., Ruuska, S. E., Levinthal, D. J. & Noy, N. (1999), Distinct roles for cellular retinoic acid-binding proteins I and II in regulating signaling by retinoic acid. *J. Biol. Chem.* **274**, 23695-23698.

Dunitz, J. D. & Taylor, R. (1997), Organic Fluorine Hardly Ever Accepts Hydrogen Bonds. *Chem. Eur. J.* **3**, 89-98.

Durand, B., Saunders, M., Leroy, P., Leid, M. & Chambon, P. (1992), All-trans and 9-cis retinoic acid induction of CRABPII transcription is mediated by RAR-RXR heterodimers bound to DR1 and DR2 repeated motifs. *Cell* **71**, 73-85.

Egea, P. F., Klaholz, B. P. & Moras, D. (2000), Ligand–protein interactions in nuclear receptors of hormones. *FEBS Lett.* **476**, 62-67.

Egea, P. F., Mitschler, A. & Moras, D. (2002), Molecular Recognition of Agonist Ligands by RXRs. *Mol. Endocrinol.* **16**, 987-997.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010), Features and development of Coot. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 486-501.

Evans, P. R. & Murshudov, G. N. (2013), How good are my data and what is the resolution? *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **69**, 1204-1214.

Evans, T., Rosenthal, E. T., Youngblom, J., Distel, D. & Hunt, T. (1983), Cyclin: A protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell* **33**, 389-396.

Fazio, V. J., Peat, T. S. & Newman, J. (2014), A drunken search in crystallization space. *Acta Crystallogr F Struct Biol Commun* **70**, 1303-1311.

Förster, T. (1948), Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der physik* **437**, 55-75.

Förster, T. (2012), Energy migration and fluorescence. *Journal of biomedical optics* **17**, 011002.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. (2003), ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784-3788.

Ghaffari, H. & Petzold, L. R. (2018), Identification of influential proteins in the classical retinoic acid signaling pathway. *Theor. Biol. Med. Model.* **15**, 16.

Ghanbarpour, A., Pinger, C., Esmatpour Salmani, R., Assar, Z., Santos, E. M., Nosrati, M., Pawlowski, K., Spence, D., Vasileiou, C., Jin, X., Borhan, B. & Geiger, J. H. (2019), Engineering the hCRBPII Domain-Swapped Dimer into a New Class of Protein Switches. *J. Am. Chem. Soc.* **141**, 17125-17132.

Ghanbarpour, A., Santos, E. M., Pinger, C., Assar, Z., Hossaini Nasr, S., Vasileiou, C., Spence, D., Borhan, B. & Geiger, J. H. (2020), Human Cellular Retinol Binding Protein II Forms a Domain-Swapped Trimer Representing a Novel Fold and a New Template for Protein Engineering. *ChemBioChem* **21**, 3192-3196.

Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. (1987), Identification of a receptor for the morphogen retinoic acid. *Nature* **330**, 624.

Glass, C. K. & Rosenfeld, M. G. (2000), The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev.* **14**, 121-141.

Gluzman, Y. (1981), SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **23**, 175-182.

Goldstein, J. T., Dobrzyn, A., Clagett-Dame, M., Pike, J. W. & DeLuca, H. F. (2003), Isolation and characterization of unsaturated fatty acids as natural ligands for the retinoid-X receptor. *Arch. Biochem. Biophys.* **420**, 185-193.

Gomaa, M. S., Bridgens, C. E., Illingworth, N. A., Veal, G. J., Redfern, C. P. F., Brancale, A., Armstrong, J. L. & Simons, C. (2012), Novel retinoic acid 4-hydroxylase (CYP26) inhibitors based on a 3-(1H-imidazol- and triazol-1-yl)-2,2-dimethyl-3-(4-(phenylamino)phenyl)propyl scaffold. *Biorg. Med. Chem.* **20**, 4201-4207.

Gorman, C. M., Moffat, L. F. & Howard, B. H. (1982), Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**, 1044-1051.

Gould, S. J. & Subramani, S. (1988), Firefly luciferase as a tool in molecular and cell biology. *Anal. Biochem.* **175**, 5-13.

Haffez, H., Chisholm, D. R., Valentine, R., Pohl, E., Redfern, C. & Whiting, A. (2017), The molecular basis of the interactions between synthetic retinoic acid analogues and the retinoic acid receptors. *MedChemComm* **8**, 578-592.

Hannig, G. & Makrides, S. C. (1998), Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends Biotechnol.* **16**, 54-60.

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E. & Hutchison, G. R. (2012), Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics* **4**, 17.

Harper, S. & Speicher, D. W. (2011), Purification of proteins fused to glutathione S-transferase. *Methods Mol. Biol.* **681**, 259-280.

Heery, D. M., Kalkhoven, E., Hoare, S. & Parker, M. G. (1997), A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**, 733-736.

Hellberg, U., Ivarsson, J.-P. & Johansson, B.-L. (1996), Characteristics of Superdex® prep grade media for gel filtration chromatography of proteins and peptides. *Process Biochem.* **31**, 163-172.

Henrich, B., Bergamaschi, A., Broennimann, C., Dinapoli, R., Eikenberry, E. F., Johnson, I., Kobas, M., Kraft, P., Mozzanica, A. & Schmitt, B. (2009), PILATUS: A single photon counting pixel detector for X-ray applications. *Nucl. Instrum. Methods Phys. Res., Sect. A* **607**, 247-249.

Heyman, R. A., Mangelsdorf, D. J., Dyck, J. A., Stein, R. B., Eichele, G., Evans, R. M. & Thaller, C. (1992), 9-cis retinoic acid is a high affinity ligand for the retinoid X receptor. *Cell* **68**, 397-406.

Higgins, D. G. & Sharp, P. M. (1988), CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237-244.

Holmbeck, S. M. A., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, H. J. & Wright, P. E. (1998), High-resolution solution structure of the retinoid X receptor DNA-binding domain. *J. Mol. Biol.* **281**, 271-284.

Howard, J. A. K., Hoy, V. J., O'Hagan, D. & Smith, G. T. (1996), How good is fluorine as a hydrogen bond acceptor? *Tetrahedron* **52**, 12613-12622.

Hu, K.-W., Pan, X.-H., Chen, F.-H., Qin, R., Wu, L.-M., Zhu, H.-G., Wu, F.-R., Ge, J.-F., Han, W.-X. & Yin, C.-L. (2014), A novel retinoic acid analog, 4-amino-2-trifluoromethyl-phenyl retinate, inhibits gastric cancer cell growth. *Int. J. Mol. Med.* **33**, 415-422.

Huchet, Q. A., Kuhn, B., Wagner, B., Kratochwil, N. A., Fischer, H., Kansy, M., Zimmerli, D., Carreira, E. M. & Müller, K. (2015), Fluorination Patterning: A Study of Structural Motifs That Impact Physicochemical Properties of Relevance to Drug Discovery. *J. Med. Chem.* **58**, 9041-9060.

Huxley, R. R., Lloyd, B. B., Goldacre, M. & Neil, H. A. W. (2000), Nutritional research in World War 2: The Oxford Nutrition Survey and its research potential 50 years later. *Br. J. Nutr.* **84**, 247-251.

Huynh, C. K., Brodie, A. M. H. & Njar, V. C. O. (2006), Inhibitory effects of retinoic acid metabolism blocking agents (RAMBAs) on the growth of human prostate cancer cells and LNCaP prostate tumour xenografts in SCID mice. *Br. J. Cancer* **94**, 513-523.

Jensen, F. C., Girardi, A. J., Gilden, R. V. & Koprowski, H. (1964), Infection of human and simian tissue cultures with Rous sarcoma virus. *Proc. Natl. Acad. Sci. USA* **52**, 53.

Jeong, H., Barbe, V., Lee, C. H., Vallenet, D., Yu, D. S., Choi, S.-H., Couloux, A., Lee, S.-W., Yoon, S. H., Cattolico, L., Hur, C.-G., Park, H.-S., Ségurens, B., Kim, S. C., Oh, T. K., Lenski, R. E., Studier, F. W., Daegelen, P. & Kim, J. F. (2009), Genome Sequences of Escherichia coli B strains REL606 and BL21(DE3). *J. Mol. Biol.* **394**, 644-652.

Kabsch, W. (2010), XDS. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 125-132.

Kastner, P., Krust, A., Mendelsohn, C., Garnier, J., Zelent, A., Leroy, P., Staub, A. & Chambon, P. (1990), Murine isoforms of retinoic acid receptor gamma with specific patterns of expression. *Proc. Natl. Acad. Sci. USA* **87**, 2700-2704.

Katz, H. I., Waalen, J. & Leach, E. E. (1999), Acitretin in psoriasis: An overview of adverse effects. *J. Am. Acad. Dermatol.* **41**, S7-S12.

Khatib, T., Chisholm, D. R., Whiting, A., Platt, B. & McCaffery, P. (2020), Decay in Retinoic Acid Signaling in Varied Models of Alzheimer's Disease and In-Vitro Test of Novel Retinoic Acid Receptor Ligands (RAR-Ms) to Regulate Protective Genes. *J. Alzheimer's Dis.* **73**, 935-954.

Khatib, T., Marini, P., Nunna, S., Chisholm, D. R., Whiting, A., Redfern, C., Greig, I. R. & McCaffery, P. (2019), Genomic and non-genomic pathways are both crucial for peak induction of neurite outgrowth by retinoids. *Cell. Commun. Signal* **17**, 40.

Kiriiri, G. K., Njogu, P. M. & Mwangi, A. N. (2020), Exploring different approaches to improve the success of drug discovery and development projects: a review. *Future J. Pharm. Sci.* **6**, 27.

Kleywegt, G. J., Bergfors, T., Senn, H., Motte, P. L., Gsell, B., Shud, K. & Jones, T. A. (1994), Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-*trans*-retinoic acid and a synthetic retinoid. *Structure* **2**, 1241-1258.

Knegtel, R. M. A., Katahira, M., Schilthuis, J. G., Bonvin, A. M. J. J., Boelens, R., Eib, D., van der Saag, P. T. & Kaptein, R. (1993), The solution structure of the human retinoic acid receptor-β DNA-binding domain. *J. Biomol. NMR* **3**, 1-17.

Koop, D., Holland, N. D., Sémon, M., Alvarez, S., de Lera, A. R., Laudet, V., Holland, L. Z. & Schubert, M. (2010), Retinoic acid signaling targets Hox genes during the amphioxus gastrula stage: Insights into early anterior–posterior patterning of the chordate body plan. *Dev. Biol.* **338**, 98-106.

Krautheim, A. & Gollnick, H. P. M. (2004), Acne: Topical treatment. *Clin. Dermatol.* **22**, 398-407.

Kumar, S. & Duester, G. (2014), Retinoic acid controls body axis extension by directly repressing Fgf8 transcription. *Development* **141**, 2972-2977.

Kuzmič, P. (1996), Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* **237**, 260-273.

Laemmli, U. K. (1970), Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680-685.

Lakshman, M. R. (2004), Alpha and Omega of Carotenoid Cleavage. *J. Nutr.* **134**, 241S-245S.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007), Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Laskowski, R. A. & Swindells, M. B. (2011), LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **51**, 2778-2786.

Layton, A. (2009), The use of isotretinoin in acne. *Dermatoendocrinol.* **1**, 162-169.

le Maire, A., Teyssier, C., Erb, C., Grimaldi, M., Alvarez, S., de Lera, A. R., Balaguer, P., Gronemeyer, H., Royer, C. A., Germain, P. & Bourguet, W. (2010), A unique secondary-structure switch controls constitutive gene repression by retinoic acid receptor. *Nat. Struct. Mol. Biol.* **17**, 801-807.

Leid, M., Kastner, P. & Chambon, P. (1992), Multiplicity generates diversity in the retinoic acid signalling pathways. *Trends Biochem. Sci.* **17**, 427-433.

Leo, M. A., Lasker, J. M., Raucy, J. L., Kim, C.-I., Black, M. & Lieber, C. S. (1989), Metabolism of retinol and retinoic acid by human liver cytochrome P450IIC8. *Arch. Biochem. Biophys.* **269**, 305-312.

Leo, M. A. & Lieber, C. S. (1985), New pathway for retinol metabolism in liver microsomes. *J. Biol. Chem.* **260**, 5228-5231.

Levin, A. A., Sturzenbecker, L. J., Kazmer, S., Bosakowski, T., Huselton, C., Allenby, G., Speck, J., Rosenberger, M., Lovey, A. & Grippo, J. F. (1992), 9-cis retinoic acid stereoisomer binds and activates the nuclear receptor RXRα. *Nature* **355**, 359.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019), Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **75**, 861-877.

Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C. & Adams, P. D. (2017), Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **73**, 148-157.

Lindqvist, A. & Andersson, S. (2002), Biochemical Properties of Purified Recombinant Human β-Carotene 15,15′-Monooxygenase. *J. Biol. Chem.* **277**, 23942-23948.

Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P. & Berkmen, M. (2016), Erratum to: SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial cell factories* **15**, 124-124.

Lou, X., Toresson, G., Benod, C., Suh, J. H., Philips, K. J., Webb, P. & Gustafsson, J.-A. (2014), Structure of the retinoid X receptor α–liver X receptor β (RXRα–LXRβ) heterodimer on DNA. *Nat. Struct. Mol. Biol.* **21**, 277-281.

Loudig, O., Babichuk, C., White, J., Abu-Abed, S., Mueller, C. & Petkovich, M. (2000), Cytochrome P450RAI(CYP26) promoter: a distinct composite retinoic acid response element underlies the complex regulation of retinoic acid metabolism. *Mol. Endocrinol.* **14**, 1483-1497.

Loudig, O., Maclean, G. A., Dore, N. L., Luu, L. & Petkovich, M. (2005), Transcriptional co-operativity between distant retinoic acid response elements in regulation of Cyp26A1 inducibility. *Biochem. J* **392**, 241-248.

Lowe, E. D., Tews, I., Cheng, K. Y., Brown, N. R., Gul, S., Noble, M. E. M., Gamblin, S. J. & Johnson, L. N. (2002), Specificity Determinants of Recruitment Peptides Bound to Phospho-CDK2/Cyclin A. *Biochemistry* **41**, 15625-15634.

Lowe, M. N. & Plosker, G. L. (2000), Bexarotene. *Am. J. Clin. Dermatol.* **1**, 245-250.

Lu, J., Cistola, D. P. & Li, E. (2006), Analysis of Ligand Binding and Protein Dynamics of Human Retinoid X Receptor Alpha Ligand-Binding Domain by Nuclear Magnetic Resonance. *Biochemistry* **45**, 1629-1639.

Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D. & Butt, T. R. (2004), SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics.* **5**, 75-86.

Mangelsdorf, D. J. & Evans, R. M. (1995), The RXR heterodimers and orphan receptors. *Cell* **83**, 841-850.

Mangelsdorf, D. J., Ong, E. S., Dyck, J. A. & Evans, R. M. (1990), Nuclear receptor that identifies a novel retinoic acid response pathway. *Nature* **345**, 224.

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X. & Butt, T. R. (2006), Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci.* **15**, 182-189.

Masiá, S., Alvarez, S., de Lera, A. R. & Barettino, D. (2007), Rapid, Nongenomic Actions of Retinoic Acid on Phosphatidylinositol-3-Kinase Signaling Pathway Mediated by the Retinoic Acid Receptor. *Mol. Endocrinol.* **21**, 2391-2402.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007), Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674.

McNicholas, S., Potterton, E., Wilson, K. & Noble, M. (2011), Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **67**, 386-394.

Meakin, P., Jesson, J. & Tolman, C. (1972), Nature of chlorotris (triphenylphosphine) rhodium in solution and its reaction with hydrogen. *J. Am. Chem. Soc.* **94**, 3240-3242.

Mercader, N., Leonardo, E., Piedra, M. E., Martínez, A. C., Ros, M. A. & Torres, M. (2000), Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development* **127**, 3961-3970.

Moon, R. C., Thompson, H. J., Becci, P. J., Grubbs, C. J., Gander, R. J., Newton, D. L., Smith, J. M., Phillips, S. L., Henderson, W. R. & Mullen, L. T. (1979), N-(4-Hydroxyphenyl) retinamide, a new retinoid for prevention of breast cancer in the rat. *Cancer Res.* **39**, 1339-1346.

Moutier, E., Ye, T., Choukrallah, M.-A., Urban, S., Osz, J., Chatagnon, A., Delacroix, L., Langer, D., Rochel, N., Moras, D., Benoit, G. & Davidson, I. (2012), Retinoic Acid Receptors Recognize

the Mouse Genome through Binding Elements with Diverse Spacing and Topology. *J. Biol. Chem.* **287**, 26328-26341.

Murayama, A., Suzuki, T. & Matsui, M. (1997), Photoisomerization of retinoic acids in ethanol under room light: a warning for cell biological study of geometrical isomers of retinoids. *J. Nutr. Sci. Vitaminol.* **43**, 167-176.

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011), REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **67**, 355-367.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997), Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **53**, 240-255.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999), Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **55**, 247-255.

Nagy, L. & Schwabe, J. W. R. (2004), Mechanism of the nuclear receptor molecular switch. *Trends Biochem. Sci.* **29**, 317-324.

Nahoum, V., Pérez, E., Germain, P., Rodríguez-Barrios, F., Manzo, F., Kammerer, S., Lemaire, G., Hirsch, O., Royer, C. A., Gronemeyer, H., de Lera, A. R. & Bourguet, W. (2007), Modulators of the structural dynamics of the retinoid X receptor to reveal receptor function. *Proc. Natl. Acad. Sci. USA* **104**, 17323-17328.

Nelson, C. H., Peng, C.-C., Lutz, J. D., Yeung, C. K., Zelter, A. & Isoherranen, N. (2016), Direct protein–protein interactions and substrate channeling between cellular retinoic acid binding proteins and CYP26B1. *FEBS Lett.* **590**, 2527-2535.

Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005), Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **61**, 1426-1431.

Nicholls, R. A., Long, F. & Murshudov, G. N. (2012), Low-resolution refinement tools in REFMAC5. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **68**, 404-417.

Niedzialkowska, E., Gasiorowska, O., Handing, K. B., Majorek, K. A., Porebski, P. J., Shabalin, I. G., Zasadzinska, E., Cymborowski, M. & Minor, W. (2016), Protein purification and crystallization artifacts: The tale usually not told. *Protein Sci.* **25**, 720-733.

Nishihara, K., Kanemori, M., Yanagi, H. & Yura, T. (2000), Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Appl Environ Microbiol* **66**, 884-889.

Njar, V. C. O., Gediya, L., Purushottamachar, P., Chopra, P., Vasaitis, T. S., Khandelwal, A., Mehta, J., Huynh, C., Belosay, A. & Patel, J. (2006), Retinoic acid metabolism blocking agents (RAMBAs) for treatment of cancer and dermatological diseases. *Biorg. Med. Chem.* **14**, 4323-4340.

O'Hea, J. D., Burt, M. H., Fisher, S., Jones, K. M. J., McAuley, K. E., Preece, G. & Williams, M. A. (2018), BART: Development of a Sample Exchange System for MX Beamlines. *Contribution to Conference Proceedings*.

Ong, D. E. (1993), Retinoid metabolism during intestinal absorption. *J. Nutr.* **123**, 351-355.

Osz, J., Brélivet, Y., Peluso-Iltis, C., Cura, V., Eiler, S., Ruff, M., Bourguet, W., Rochel, N. & Moras, D. (2012), Structural basis for a molecular allosteric control mechanism of cofactor binding to nuclear receptors. *Proc. Natl. Acad. Sci. USA* **109**, E588-594.

Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C. & Lesley, S. A. (2003), Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the Thermotoga maritima proteome. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **59**, 1028-1037.

Park, J., Throop, A. L. & LaBaer, J. (2015), Site-specific recombinational cloning using gateway and in-fusion cloning schemes. *Current protocols in molecular biology* **110**, 3.20.21-23.20.23.

Park, S. W., Nhieu, J., Persaud, S. D., Miller, M. C., Xia, Y., Lin, Y.-W., Lin, Y.-L., Kagechika, H., Mayo, K. H. & Wei, L.-N. (2019), A new regulatory mechanism for Raf kinase activation, retinoic acid-bound Crabp1. *Sci. Rep.* **9**, 10929.

Park, S. W., Persaud, S. D., Ogokeh, S., Meyers, T. A., Townsend, D. & Wei, L.-N. (2018), CRABP1 protects the heart from isoproterenol-induced acute and chronic remodeling. *J. Endocrinol.* **236**, 151-165.

Pastok, M. W., Tomlinson, C. W. E., Tatum, N. J., Baslé, A., Noble, M. E. M., Pohl, E. & Endicott, J. A. (2020), Paper in Preparation.

Patel, J. B., Mehta, J., Belosay, A., Sabnis, G., Khandelwal, A., Brodie, A. M. H., Soprano, D. R. & Njar, V. C. O. (2007), Novel retinoic acid metabolism blocking agents have potent inhibitory activities on human breast cancer cells and tumour growth. *Br. J. Cancer* **96**, 1204-1215.

Persaud, S. D., Lin, Y.-W., Wu, C.-Y., Kagechika, H. & Wei, L.-N. (2013), Cellular retinoic acid binding protein I mediates rapid non-canonical activation of ERK1/2 by all-trans retinoic acid. *Cell. Signal.* **25**, 19-25.

Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. (1987), A human retinoic acid receptor which belongs to the family of nuclear receptors. *Nature* **330**, 444-450.

Porath, J. (1992), Immobilized metal ion affinity chromatography. *Protein Expression Purif.* **3**, 263-281.

Porath, J., Carlsson, J. A. N., Olsson, I. & Belfrage, G. (1975), Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598-599.

Porath, J. & Flodin, P. E. R. (1959), Gel Filtration: A Method for Desalting and Group Separation. *Nature* **183**, 1657-1659.

Rastinejad, F., Wagner, T., Zhao, Q. & Khorasanizadeh, S. (2000), Structure of the RXR–RAR DNA-binding complex on the retinoic acid response element DR1. *EMBO J.* **19**, 1045-1054.

Redfern, C. P. F. & Wilson, K. E. (1993), Ligand binding properties of human cellular retinoic acid binding protein II expressed in E. coli as a glutathione-S-transferase fusion protein. *FEBS Lett.* **321**, 163-168.

Repa, J. J., Hanson, K. K. & Clagett-Dame, M. (1993), All-trans-retinol is a ligand for the retinoic acid receptors. *Proc. Natl. Acad. Sci. USA* **90**, 7293-7297.

Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. (1993), Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401-1416.

Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mély, Y., Svergun, D. I. & Moras, D. (2011), Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat. Struct. Mol. Biol.* **18**, 564-570.

Rochette-Egly, C. (2015), Retinoic acid signaling and mouse embryonic stem cell differentiation: Cross talk between genomic and non-genomic effects of RA. *Biochim. Biophys. Acta* **1851**, 66-75.

Rosenberg, A. H., Lade, B. N., Dao-shan, C., Lin, S.-W., Dunn, J. J. & Studier, F. W. (1987), Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**, 125-135.

Ross, A. C. & Zolfaghari, R. (2011), Cytochrome P450s in the Regulation of Cellular Retinoic Acid Metabolism. *Annu. Rev. Nutr.* **31**, 65-87.

Ross, S. A., McCaffery, P. J., Drager, U. C. & Luca, L. M. D. (2000), Retinoids in Embryonal Development. *Physiol. Rev.* **80**, 1021-1054.

Rupp, B. (2009). *Biomolecular crystallography: principles, practice, and application to structural biology*. Garland Science.

Sahin, M., Karaüzüm, S. B., Perry, G., Smith, M. A. & Alicigüzel, Y. (2005), Retinoic acid isomers protect hippocampal neurons from amyloid-β induced neurodegeneration. *Neurotox. Res.* **7**, 243-250.

Sato, Y., Ramalanjaona, N., Huet, T., Potier, N., Osz, J., Antony, P., Peluso-Iltis, C., Poussin-Courmontagne, P., Ennifar, E., Mély, Y., Dejaegere, A., Moras, D. & Rochel, N. (2010). The

"Phantom Effect" of the Rexinoid LG100754: structural and functional insights, Vol. 5, PLOS ONE, p. e15119.

Schilling, T. F., Sosnik, J. & Nie, Q. (2016), Visualizing retinoic acid morphogen gradients. *Methods in cell biology* **133**, 139-163.

Selas Castiñeiras, T., Williams, S. G., Hitchcock, A. G. & Smith, D. C. (2018), E. coli strain engineering for the production of advanced biopharmaceutical products. *FEMS Microbiol. Lett.* **365**.

Seliger, H. H. & McElroy, W. D. (1960), Spectral emission and quantum yield of firefly bioluminescence. *Arch. Biochem. Biophys.* **88**, 136-141.

Sessler, R. J. & Noy, N. (2005), A Ligand-Activated Nuclear Localization Signal in Cellular Retinoic Acid Binding Protein-II. *Mol. Cell* **18**, 343-353.

Simeone, A. M. & Tari, A. M. (2004), How retinoids regulate breast cancer cell proliferation and apoptosis. *Cell. Mol. Life Sci.* **61**, 1475-1484.

Smith, D. B. & Johnson, K. S. (1988), Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* **67**, 31-40.

Sorg, O., Antille, C., Kaya, G. & Saurat, J. H. (2006), Retinoids in cosmeceuticals. *Dermatol. Ther.* **19**, 289-296.

Sorg, O., Carraux, P., Grand, D., Hügin, A., Didierjean, L. & Saurat, J. H. (2005), Spectral properties of topical retinoids prevent DNA damage and apoptosis after acute UV-B exposure in hairless mice. *Photochem. Photobiol.* **81**, 830-836.

Stafslien, D. K., Vedvik, K. L., De Rosier, T. & Ozers, M. S. (2007), Analysis of ligand-dependent recruitment of coactivator peptides to RXRβ in a time-resolved fluorescence resonance energy transfer assay. *Mol. Cell. Endocrinol.* **264**, 82-89.

Stevison, F., Jing, J., Tripathy, S. & Isoherranen, N. (2015), Role of Retinoic Acid-Metabolizing Cytochrome P450s, CYP26, in Inflammation and Cancer. *Adv. Pharmacol.* **74**, 373-412.

Strickland, S., Breitman, T. R., Frickel, F., Nürrenbach, A., Hädicke, E. & Sporn, M. B. (1983), Structure-Activity Relationships of a New Series of Retinoidal Benzoic Acid Derivatives as Measured by Induction of Differentiation of Murine F9 Teratocarcinoma Cells and Human HL-60 Promyelocytic Leukemia Cells. *Cancer Res.* **43**, 5268-5272.

Stryer, L. (1978), Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem* **47**, 819-846.

Studer, M., Pöpperl, H., Marshall, H., Kuroiwa, A. & Krumlauf, R. (1994), Role of a conserved retinoic acid response element in rhombomere restriction of Hoxb-1. *Science* **265**, 1728-1732.

Studier, F. W. & Moffatt, B. A. (1986), Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.

Tatum, N. J. & Endicott, J. A. (2020), Chatterboxes: the structural and functional diversity of cyclins. *Semin. Cell Dev. Biol.* **107**, 4-20.

Teng, T.-Y. (1990), Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Crystallogr.* **23**, 387-391.

Thatcher, J. E. & Isoherranen, N. (2009), The role of CYP26 enzymes in retinoic acid clearance. *Expert Opin. Drug Metab. Toxicol.* **5**, 875-886.

Thielitz, A., Krautheim, A. & Gollnick, H. (2006), Update in retinoid therapy of acne. *Dermatol Ther* **19**, 272-279.

Thompson, J. R., Bratt, J. M. & Banaszak, L. J. (1995), Crystal structure of cellular retinoic acid binding protein I shows increased access to the binding cavity due to formation of an intermolecular beta-sheet. *J. Mol. Biol.* **252**, 433-446.

Tickle, C. (2006), Making digit patterns in the vertebrate limb. *Nat. Rev. Mol. Cell Biol.* **7**, 45-53.

Tickle, C., Lee, J. & Eichele, G. (1985), A quantitative analysis of the effect of all-trans-retinoic acid on the pattern of chick wing development. *Dev. Biol.* **109**, 82-95.

Tomlinson, C. W. E., Chisholm, D. R., Valentine, R., Whiting, A. & Pohl, E. (2018), Novel Fluorescence Competition Assay for Retinoic Acid Binding Proteins. *ACS Med. Chem. Lett.* **9**, 1297-1300.

Totter, J. R. (1981). Bioluminescence and chemiluminescence: Basic chemistry and analytical applications: Edited by Marlene A. DeLuca and William D. McElroy, Academic Press, New York/London, 1981, Pg 179-180: Academic Press.

Townsend, P. D., Jungwirth, B., Pojer, F., Bußmann, M., Money, V. A., Cole, S. T., Pühler, A., Tauch, A., Bott, M., Cann, M. J. & Pohl, E. (2014), The crystal structures of apo and cAMP-bound GlxR from Corynebacterium glutamicum reveal structural and dynamic changes upon cAMP binding in CRP/FNR family transcription factors. *PLOS ONE* **9**, e113265-e113265.

Tsien, R. Y. (1998), The Green Fluorescent Protein. *Annu. Rev. Biochem*, 509-544.

UniProt Consortium (2017), UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158-D169.

Uson, I., Pohl, E., Schneider, T. R., Dauter, Z., Schmidt, A., Fritz, H.-J. & Sheldrick, G. M. (1999), 1.7 A structure of the stabilized REIv mutant T39K. Application of local NCS restraints. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **55**, 1158-1167.

Vaezeslami, S., Jia, X., Vasileiou, C., Borhan, B. & Geiger, J. H. (2008), Structural analysis of site-directed mutants of cellular retinoic acid-binding protein II addresses the relationship between structural integrity and ligand binding. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **64**, 1228-1239.

Vaezeslami, S., Mathes, E., Vasileiou, C., Borhan, B. & Geiger, J. H. (2006), The structure of Apo-wild-type cellular retinoic acid binding protein II at 1.4 A and its relationship to ligand binding and nuclear translocation. *J. Mol. Biol.* **363**, 687-701.

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004), REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **60**, 2184-2195.

van Oers, M. M., Pijlman, G. P. & Vlak, J. M. (2015), Thirty years of baculovirus–insect cell protein expression: from dark horse to mainstream technology. *J. Gen. Virol.* **96**, 6-23.

Vasileiou, C., Lee, K. S. S., Crist, R. M., Vaezeslami, S., Goins, S. M., Geiger, J. H. & Borhan, B. (2009), Dissection of the Critical Binding Determinants of Cellular Retinoic Acid Binding Protein II by Mutagenesis and Fluorescence Binding Assay. *Proteins* **76**, 281-290.

Verhoeven, N., Wanders, R., Saudubray, J.-M. & Jakobs, C. (1998), The metabolism of phytanic acid and pristanic acid in man: a review. *J. Inherited Metab. Dis.* **21**, 697-728.

Vivas, M., Siqueira, J., Silva, D., de Boni, L. & Mendonca, C. (2015), Investigation of the nonlinear absorption spectrum of all-trans retinoic acid by using the steady and transient two-photon absorption spectroscopy. *RSC Adv.* **5**, 74531-74538.

von Lintig, J. (2012), Metabolism of Carotenoids and Retinoids Related to Vision. *J. Biol. Chem.* **287**, 1627-1634.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. (2004), Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157-1174.

Wang, L. & Yan, H. (1999), NMR study of the binding of all-trans-retinoic acid to type II human cellular retinoic acid binding protein. *Biochim. Biophys. Acta* **1433**, 240-252.

Waugh, D. S. (2011), An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression Purif.* **80**, 283-293.

Weiss, M. S. (2001), Global indicators of X-ray data quality. *J. Appl. Crystallogr.* **34**, 130-135.

Weiss, M. S. & Hilgenfeld, R. (1997), On the use of the merging R factor as a quality indicator for X-ray data. *J. Appl. Crystallogr.* **30**, 203-205.

White, R. J., Nie, Q., Lander, A. D. & Schilling, T. F. (2007), Complex regulation of cyp26a1 creates a robust retinoic acid gradient in the zebrafish embryo. *PLoS Biol.* **5**, e304-e304.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. & Hochstrasser, D. F. (1999), Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531-552.

Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003), Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300-321.

Winter, G. (2010), xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186-190.

Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018), DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **74**, 85-97.

Wolf, G. (2002), The Experimental Induction of Vitamin A Deficiency in Humans. *J. Nutr.* **132**, 1805-1811.

Yamada, S., Kawasaki, M., Fujihara, M., Watanabe, M., Takamura, Y., Takioku, M., Nishioka, H., Takeuchi, Y., Makishima, M. & Motoyama, T. (2019), Competitive Binding Assay with an Umbelliferone-based Fluorescent Rexinoid for Retinoid X Receptor Ligand Screening. *J. Med. Chem.*

Yerien, D. E., Bonesi, S. & Postigo, A. (2016), Fluorination methods in drug discovery. *Org. Biomol. Chem.* **14**, 8398-8427.

Zhang, J.-H., Chung, T. D. & Oldenburg, K. R. (1999), A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* **4**, 67-73.

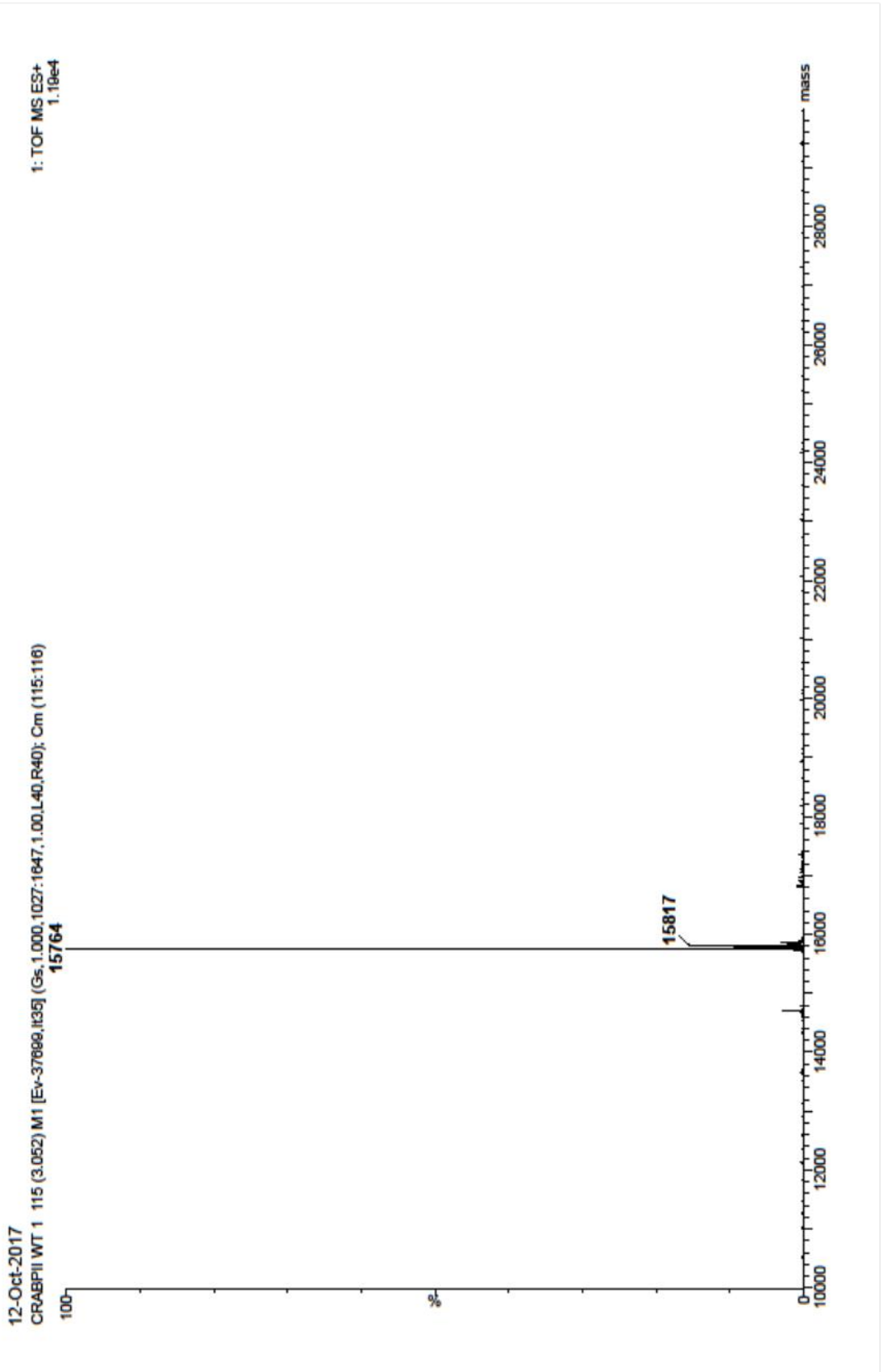
Zhang, J., Hu, X. & Lazar, M. A. (1999), A novel role for helix 12 of retinoid X receptor in regulating repression. *Mol. Cell. Biol.* **19**, 6448-6457.

Zhang, Y., Zolfaghari, R. & Ross, A. C. (2010), Multiple retinoic acid response elements cooperate to enhance the inducibility of CYP26A1 gene expression in liver. *Gene* **464**, 32-43.

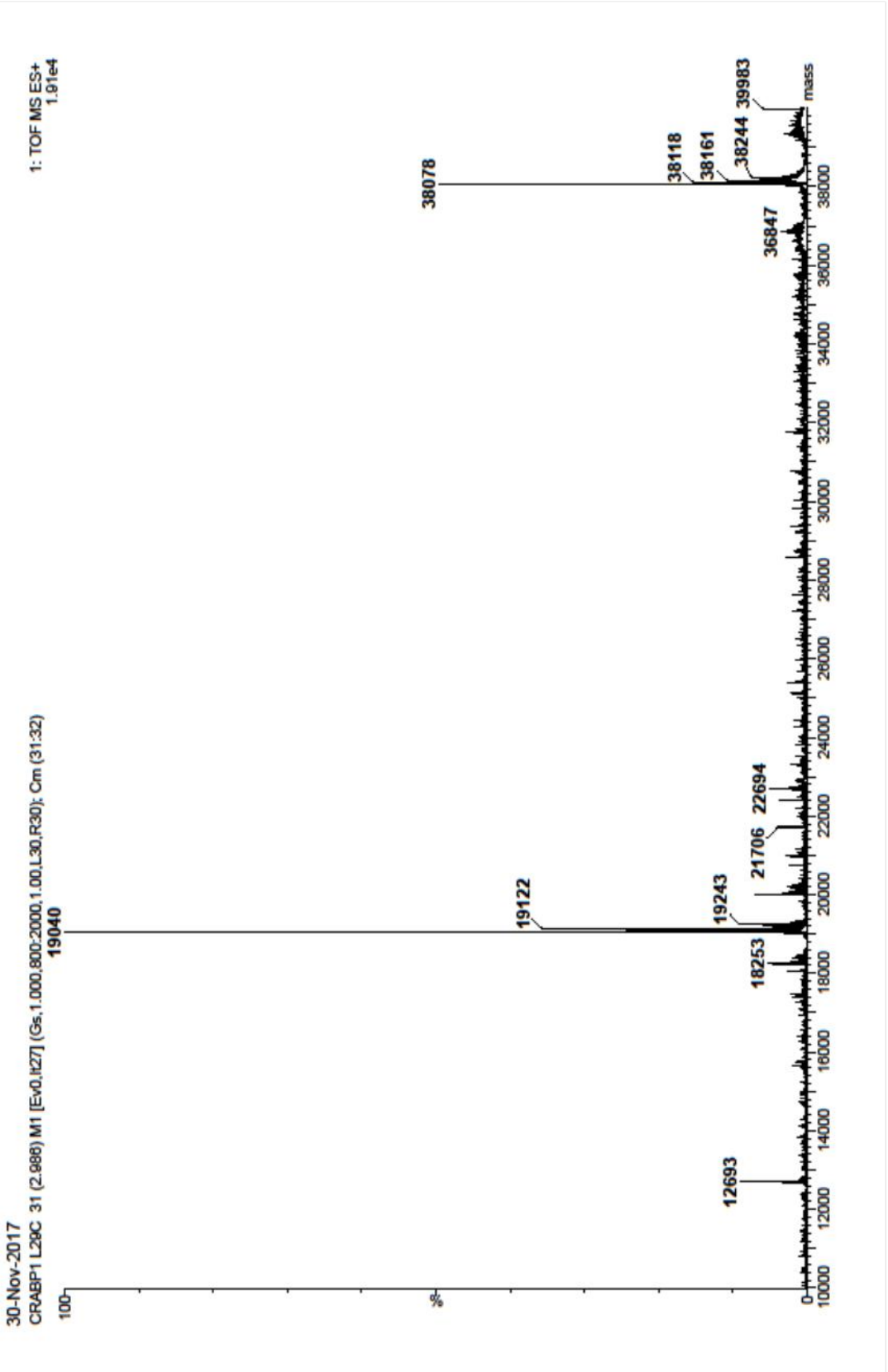
Appendix

```
hCRABPII PNFSGNWKII RSENFEDLLK VLGVNVMLRK IAVAAASKPA VEIKQEGDTF YIKTSTTVRT 60
hCRABPI  PNFAGTWKMR SSENFDELLK ALGVNAMLRK VAVAAASKPH VEIRQDGDQF YIKTSTTVRT 60
         ***:*.**:   ****::*** .****.**** :********  ***:*:** * **********

hCRABPII TEINFKVGE EFEEQTVDGRP CKSLVKWESE NKMVCEQRLL KGEGPKTSWT RELTNDGELI 120
hCRABPI  TEINFKVGE GFEEETVDGRK CRSLATWENE NKIHCTQTLL EGDGPKTYWT RELAND-ELI 119
         *********  ***:*****  *:**..**.* **: * * ** :*:**** ** ***:** ***

hCRABPII LTMTADDIVC TRVYVRE 137
hCRABPI  LTFGADDVVC TRIYVRE 136
         **: ***:** **:****

                          75 % Sequence Identity
```

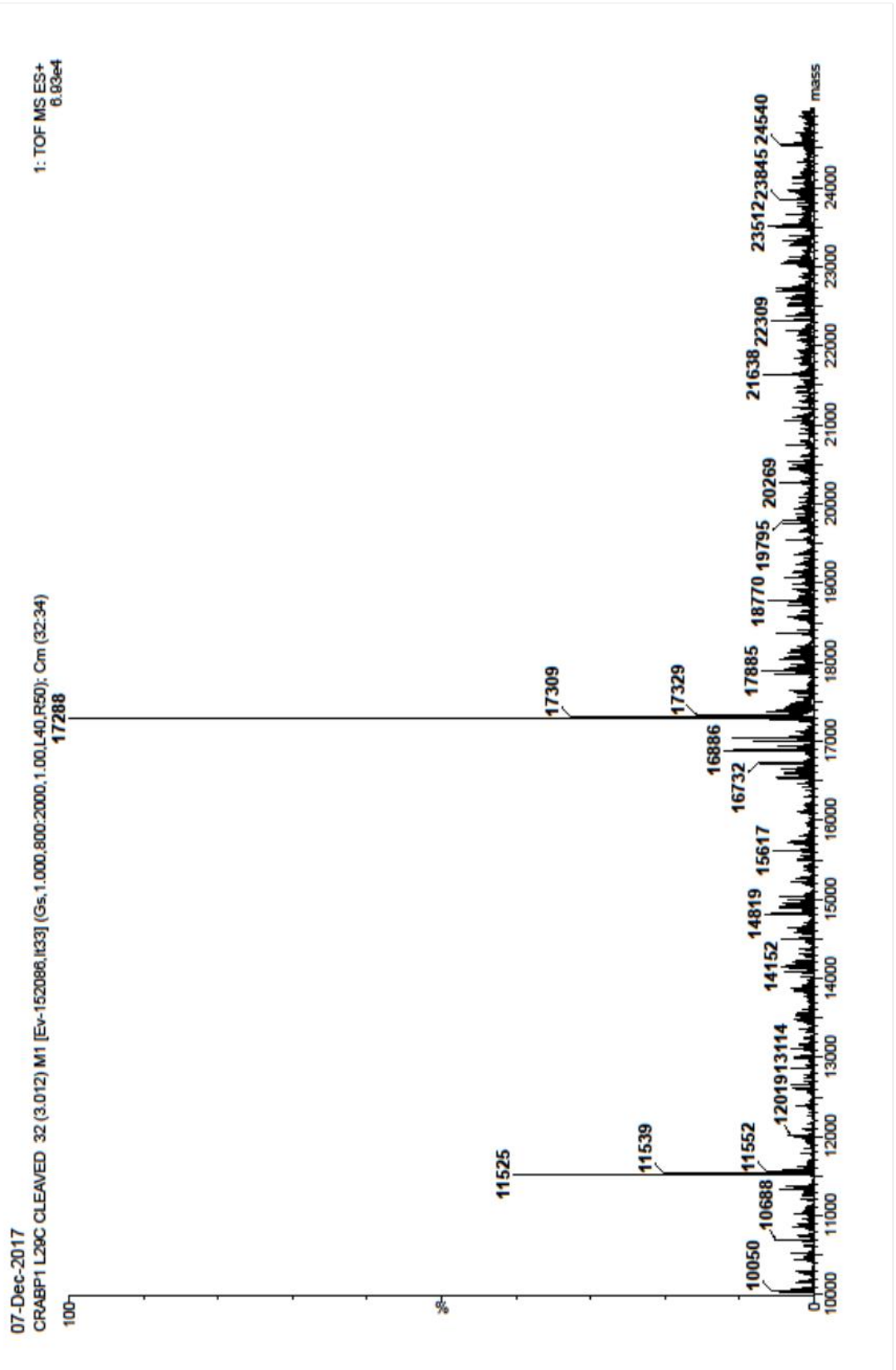
Appendix 1: Alignment of hCRABPII (P29373) and hCRABPI (P29762) primary sequences showing identical (*), high similarity (:), and low similarity (.) residues. Bold residues denote NLS (CRABPII, blue), RAR interaction residues (CRABPII, green) and ligand binding residues (red). Sequence alignment carried out using Clustalω. (Higgins & Sharp, 1988, Larkin *et al.*, 2007)



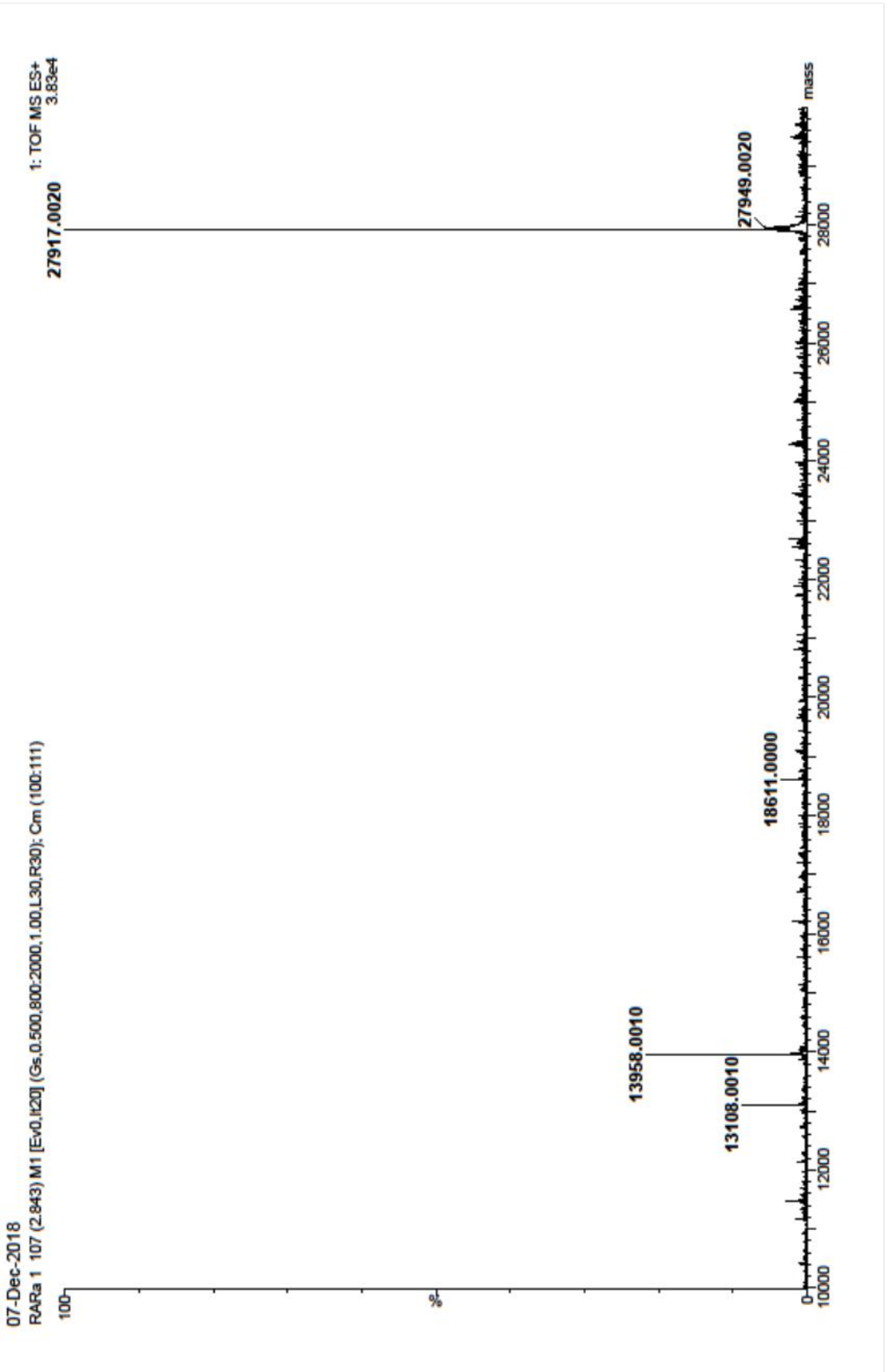
Appendix 2: Service mass spectrometry (TOF) of CRABPII post cleavage with thrombin showing molecular ion peak 15764 Da. Calculated construct mass (Expasy MW) 15764 Da.



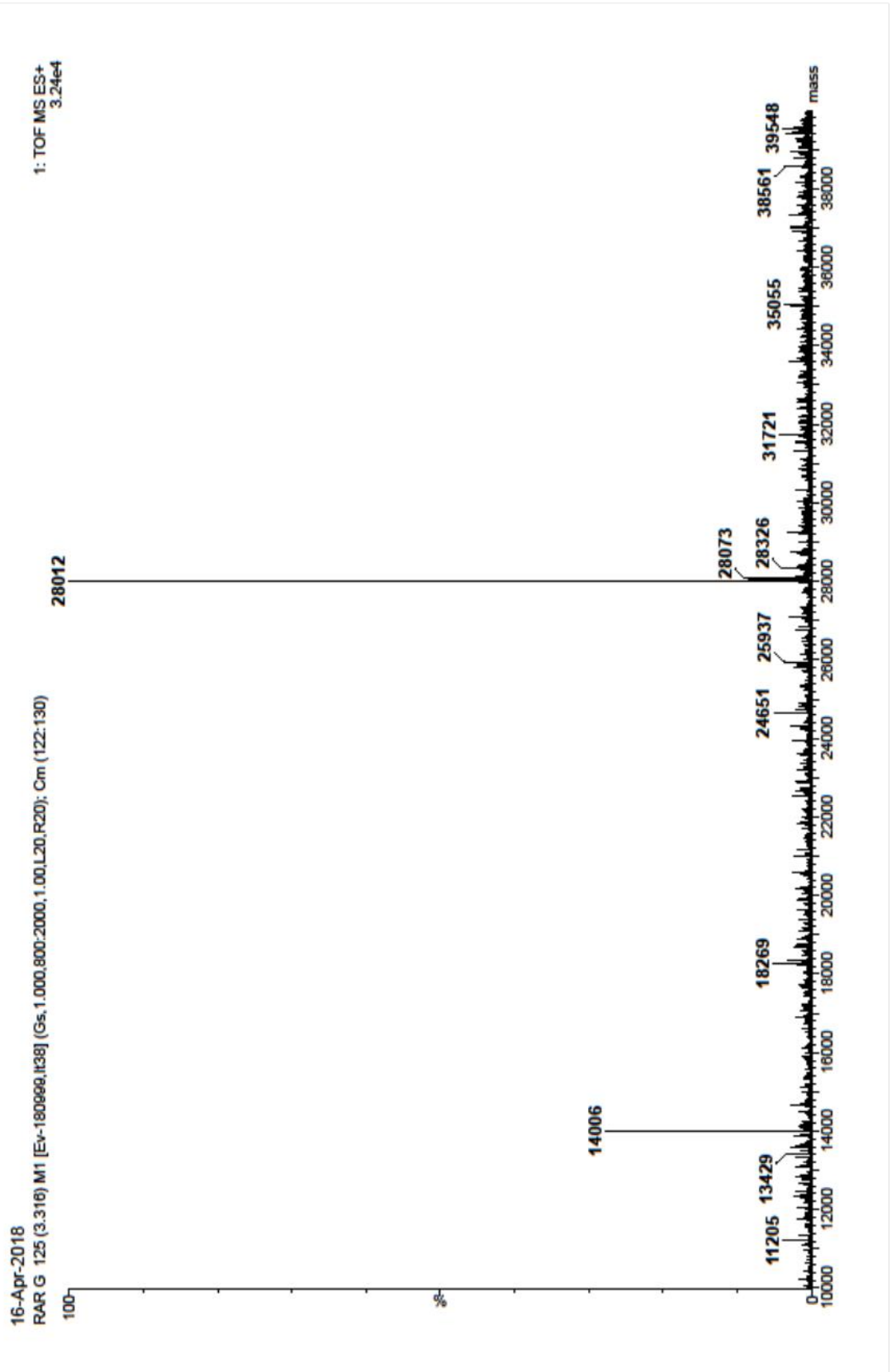
Appendix 3: Service mass spectrometry (TOF) of CRABPI-L29C prior to cleavage showing molecular ion peak 19040 Da. Calculated construct mass (Expasy MW) 19170 Da.



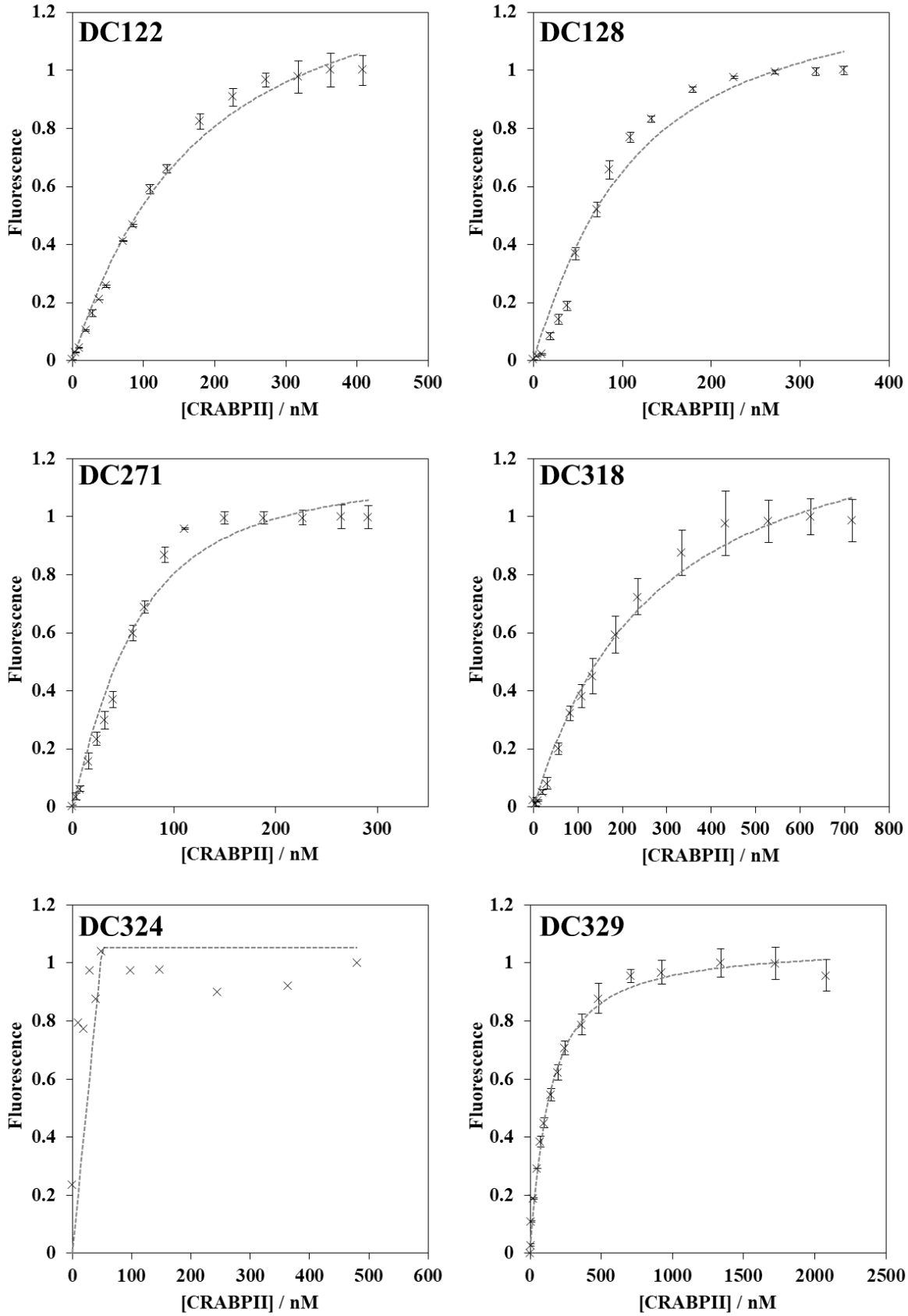
Appendix 4: Service mass spectrometry (TOF) of CRABPI-L29C post cleavage with thrombin, showing molecular ion peak 17288 Da and degradation product 11525 Da. Calculated construct mass (Expasy MW) 17228 Da.



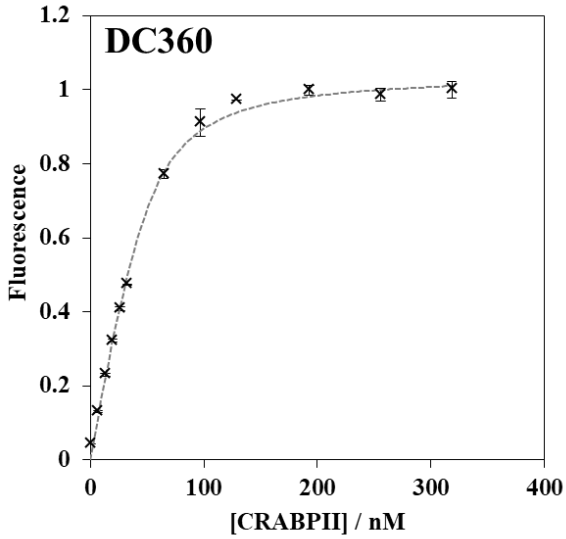
Appendix 5: Service mass spectrometry (TOF) of RARα post cleavage with HRV3C. Molecular ion peak 27917 Da. Calculated construct mass (Expasy MW) 27920 Da.

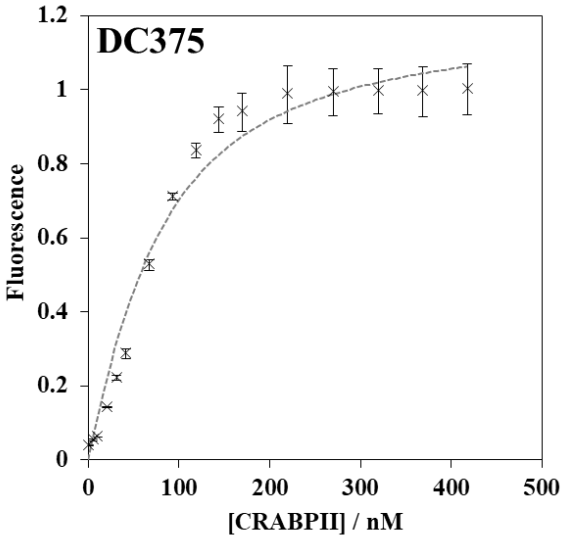


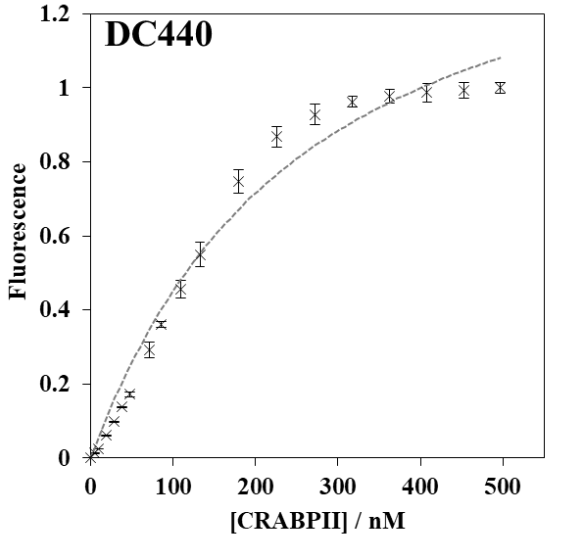
Appendix 6: Service mass spectrometry (TOF) of RARγ post cleavage with HRV3C protease, showing molecular ion peak 28012 Da. Calculated construct mass (Expasy MW) 28013 Da.

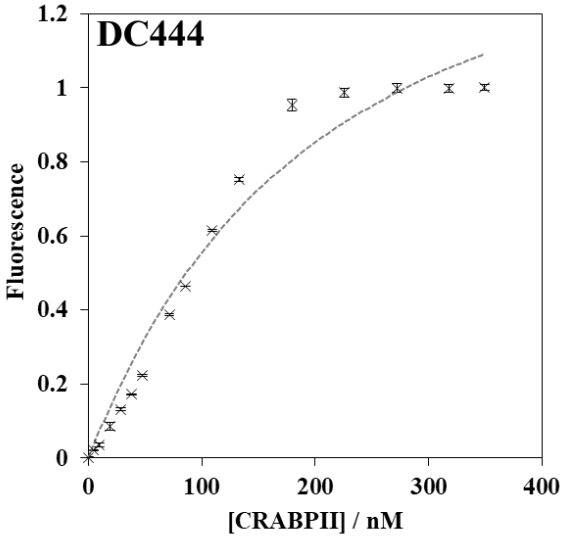


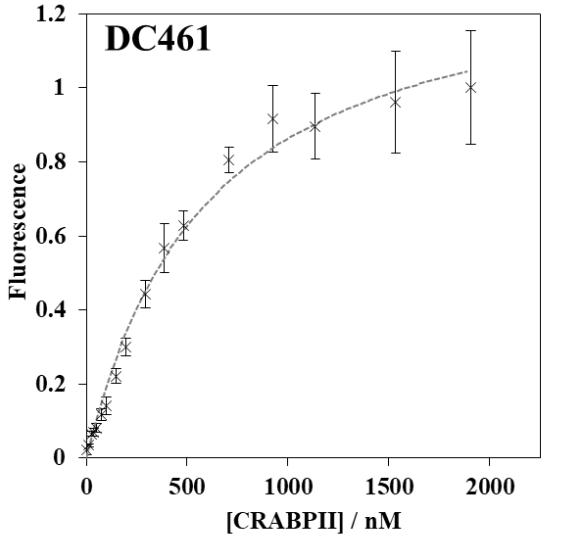
Appendix 7: Cuvette titration data (± standard deviation, n=2) for all DC compounds with CRABPII, showing isotherms at fluorescence maxima from excitation 340 nm. Also shown are curves fit using a least squares calculation (Dynafit). Continued overleaf on pages 133 and 134.

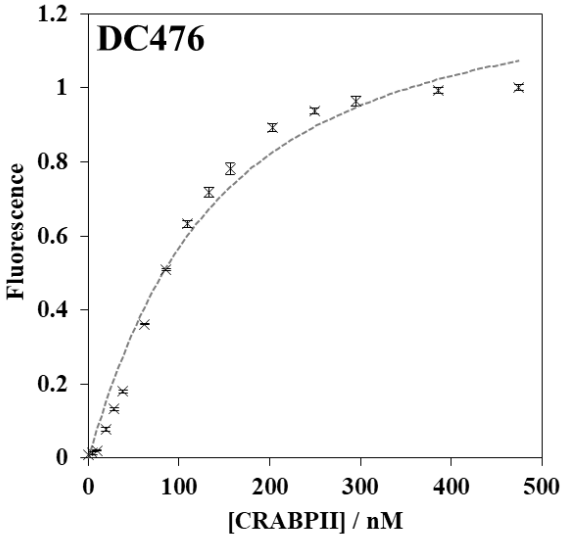

DC360
Fluorescence
[CRABPII] / nM
0
0.2
0.4
0.6
0.8
1
1.2
0
100
200
300
400

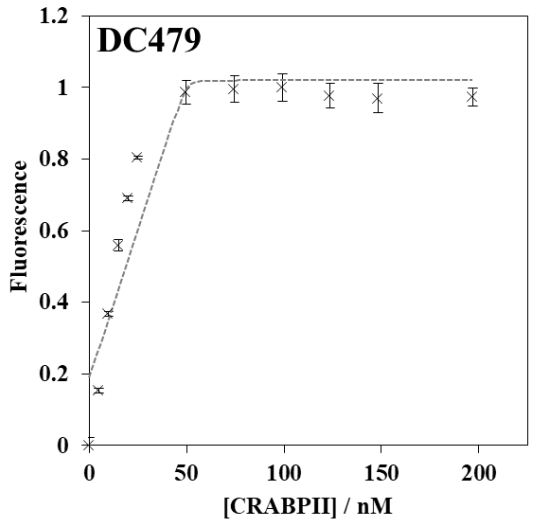

DC375
Fluorescence
[CRABPII] / nM
0
0.2
0.4
0.6
0.8
1
1.2
0
100
200
300
400
500


DC440
Fluorescence
[CRABPII] / nM
0
0.2
0.4
0.6
0.8
1
1.2
0
100
200
300
400
500


DC444
Fluorescence
[CRABPII] / nM
0
0.2
0.4
0.6
0.8
1
1.2
0
100
200
300
400


DC461
Fluorescence
[CRABPII] / nM
0
0.2
0.4
0.6
0.8
1
1.2
0
500
1000
1500
2000


DC476
Fluorescence
[CRABPII] / nM
0
0.2
0.4
0.6
0.8
1
1.2
0
100
200
300
400
500


DC479
Fluorescence
[CRABPII] / nM
0
0.2
0.4
0.6
0.8
1
1.2
0
50
100
150
200

Materials and Equipment

- **CRABPII protein from heterologous expression**
- **DC271 probe retinoid**
- **EC23, ATRA, or other known binder for positive/mid controls**
- **Dilution series of test compounds (12x or 8x based on requirements)**
- **96-well black fluorescence plates with non-binding treatment**
- **Plate reader (ex/em 355/460 nm)**

Protocol

1. **DC271 (150 nM) solvated in <1% EtOH after dilution and CRABPII (150 nM in 10 mM HEPES, 150 mM NaCl, pH 7.5, 0.5 mM EDTA + 0.5mM TCEP) solutions were prepared and mixed before use to create an equimolar solution. *N.B* - Use a suitable buffer for the expression system and protein in use.**
2. **A Corning, Black Non-binding Surface (NBS), 96-well fluorescence plate was cleaned with compressed air and Max response, Mid response (half displacement by known ligand), and all test compound wells were loaded with 100 µl CRABPII/DC271 mixed solution.**
3. **Min response wells were loaded with 100 µl control solution of DC271 mixed with HEPES buffer to control for minimum background fluorescence of the ligand in the absence of the protein.**
4. **Mid response wells were loaded with 50 µl EC23 (600 nM in 1% EtOH), calculated to displace half of the solvatochromic probe. *N.B* - any known binding ligand could be used at a suitable concentration to displace half of the fluorophore.**
5. **Max and Min response wells were loaded with 50 µl of a <1% EtOH solution, to control for the volume effect of the competitor ligand solution.**
6. **Test compounds were prepared as dilution series (typically 1.6 µM to 12.5 nM on plate concentration) and applied as 50 µl aliquots.**
7. **Plates were spun at 1700 rpm, 2 min to ensure mixing.**
8. **Plates were read at ex/em 355/460 nm on a Synergy H4 plate reader. The total well volume was 150 µl, the on-plate concentration of CRABPII and DC271 was 50 nM each.**
9. **Curve fitting was carried out using least-squares regression through Dynafit (Kuzmič, 1996) and the Z' factor for each plate was calculated (Zhang, Chung, *et al.*, 1999).**

Appendix 8: Protocol for CRABPII competitive binding assay using DC271, easily adapted to RAR proteins.

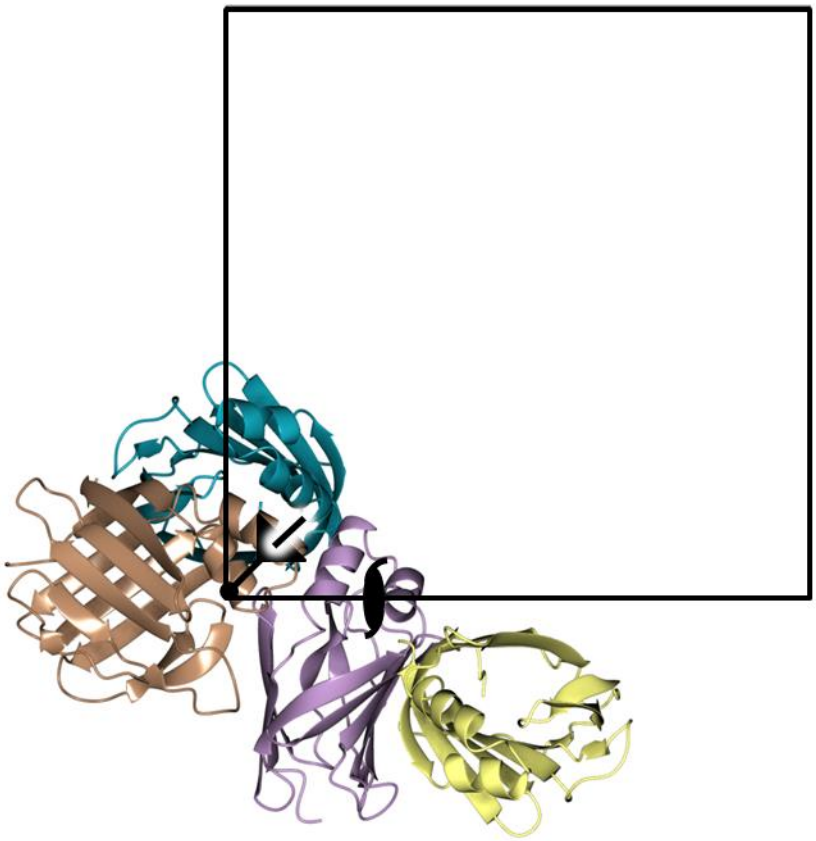
Appendix 9: Table 1: Diffraction and data processing statistics for CRABPI-L29C and CRABPII structures. Outer shell values appear in parentheses

PDB code	5OGB	6HKR	-	7AA0	7AA1	7A9Y	7A9Z
Diffraction source	Diamond Light Source - I02	Diamond Light Source - I03	Diamond Light Source - I03	Diamond Light Source - I03	Diamond Light Source - I24	Diamond Light Source - I03	Diamond Light Source - I04
Wavelength (Å)	1.0388	0.9763	0.9763	0.9763	0.9763	0.9763	0.9795
Temperature (K)	100	100	100	100	100	100	100
Detector	Pilatus 6M	Pilatus 6M	Pilatus 6M	Pilatus 6M	Pilatus 6M	Pilatus 6M	Eiger2 XE 16M
Rotation range per image (°)	0.1	0.1	0.2	0.1	0.1	0.1	0.1
Total rotation range (°)	200	180	280	200	180	200	360
Exposure time per image (s)	0.05	0.025	0.025	0.05	0.10	0.05	0.05
Detector distance / mm	368.30	289.05	289.30	321.83	186.39	213.39	385.58
Space group	$P12_11$	$P2_12_12_1$	$P2_13$	$P2_1$	$P2_13$	I222	$P3_121$
***a, b, c* (Å)**	33.05, 45.58, 36.67	44.49, 47.48, 77.73	77.463, 77.463, 77.463	39.56, 49.92, 67.37	78.56, 78.56, 78.56	36.89, 109.98, 170.34	127.54, 127.54, 52.16
α, β, γ (°)	90.0, 102.9, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.00, 96.04, 90.00	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.00, 90.00, 120.00,
Mosaicity (°)	1.87	0.10	0.00	0.70	0.00	0.00	0.00
Resolution range (Å)	9.78 - 1.80	40.52 - 1.50	44.72 - 2.80	49.92 - 1.82	45.36 - 1.71	36.05 - 1.64	110.46 - 2.41
Total No. of reflections	23289	172350	117813	80669	326013	81769	366266
No. of unique reflections	9030	27224	4018	23023	17775	43112	19102
Completeness (%)	91.1 (67.4)	100.0 (100.0)	99.9 (99.3)	97.6 (96.8)	100.0 (100.0)	99.9 (100)	99.9 (99.3)
Multiplicity	2.6 (1.8)	6.3 (5.1)	29.3 (28.2)	3.5 (3.7)	18.3 (18.3)	1.9 (1.9)	19.2 (16.8)

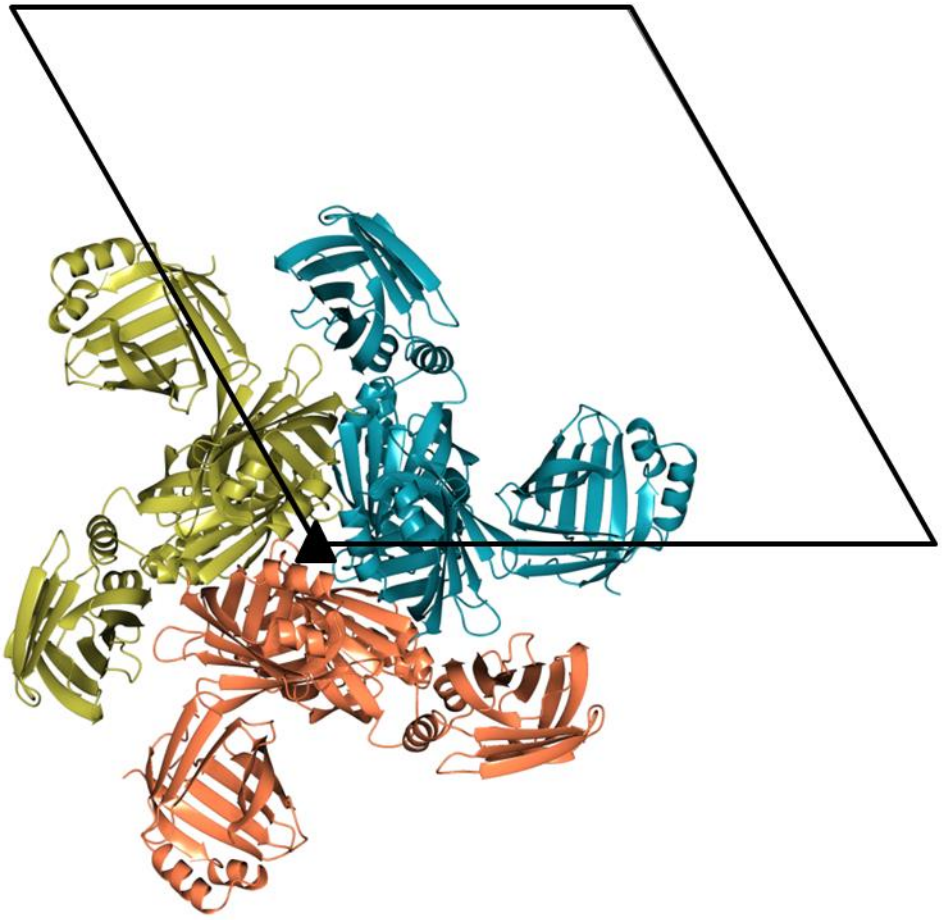
CC(1/2)	0.991 (0.762)	0.991 (0.778)	0.999 (0.987)	0.996 (0.486)	0.999 (0.548)	0.998 (0.480)	0.994 (0.398)
⟨ *I*/σ(*I*)⟩	6.2 (1.9)	8.9 (2.3)	33.6 (11.0)	10.4 (1.6)	18.8 (1.2)	9.8 (1.1)	7.1 (0.5)
Rmeas	0.112 (0.569)	0.134 (0.792)	0.086 (0.333)	0.113 (0.990)	0.082 (2.333)	0.041 (0.791)	0.389 (6.693)
***R*pim**	0.063 (0.358)	0.053 (0.348)	0.016 (0.061)	0.059 (0.504)	0.019 (0.554)	0.029 (0.559)	0.088 (1.624)
Overall *B* factor from Wilson plot (Å^2)	12.39	9.22	53.89	26.78	31.81	26.42	47.34

Appendix 10: Table 2: Statistics from refinement of CRABPI-L29C and CRABPII structures, outer shell values appear in parentheses.

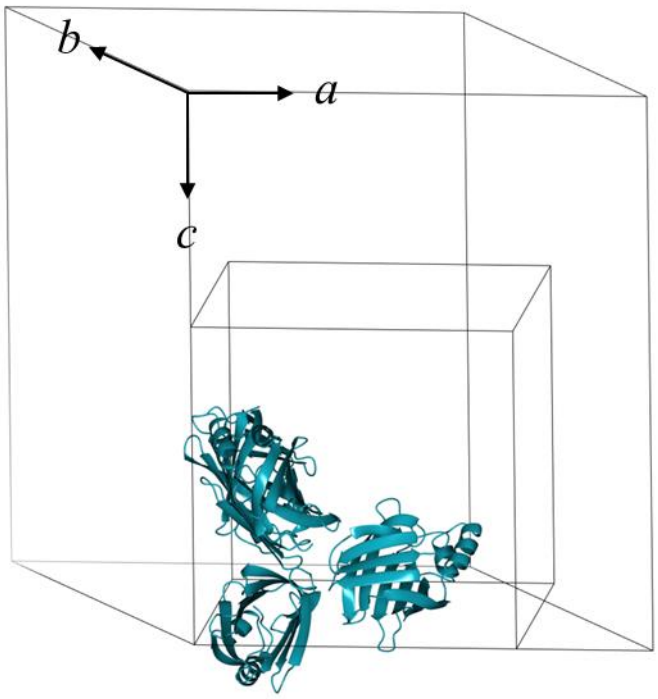
Resolution range (Å)	9.78 - 1.80	40.52 - 1.50	44.72 - 2.80	66.99 - 1.80	45.36 - 1.71	36.05 - 1.64	110.46 - 2.41
Completeness (%)	91.1 (67.4)	100.0 (100.0)	99.9 (99.3)	99.6 (99.1)	100.0 (100.0)	99.9 (100)	99.9 (99.3)
No. of reflections, working set	9002	27170	4005	23016	17753	43108	19078
No. of reflections, test set	468	1374	191	1125	889	2049	848
Final R_{cryst}	0.17	0.14	0.23	0.19	0.18	0.19	0.20
Final R_{free}	0.24	0.17	0.31	0.25	0.20	0.22	0.25
No. of non-H atoms							
Protein	1084	1101	1091	2184	1091	2212	2158
Ligand	26	49	25	52	25	37	25
Water	62	131	9	158	89	154	39
Total	1172	1281	1125	2394	1205	2403	2222
R.m.s. deviations							
Bonds (Å)	0.0138	0.0147	0.0075	0.0083	0.0120	0.0121	0.0088
Angles (°)	1.66	1.903	1.709	1.547	1.745	1.699	1.559
Average *B* factors (Å^2)							
Protein	23.06	14.51	59.08	30.19	39.21	37.11	57.51
Ligand	19.42	20.59	52.85	27.41	32.24	51.40	49.98
Water	34.4	25.95	41.06	37.43	44.93	45.79	44.85
Ramachandran plot							
Favoured (%)	96.18	97.79	95.56	95.94	97.78	97.03	95.88
Allowed (%)	3.05	2.21	3.70	3.32	1.48	2.23	3.00
High energy (%)	0.76	0	0.74	0.74	0.74	0.74	1.12



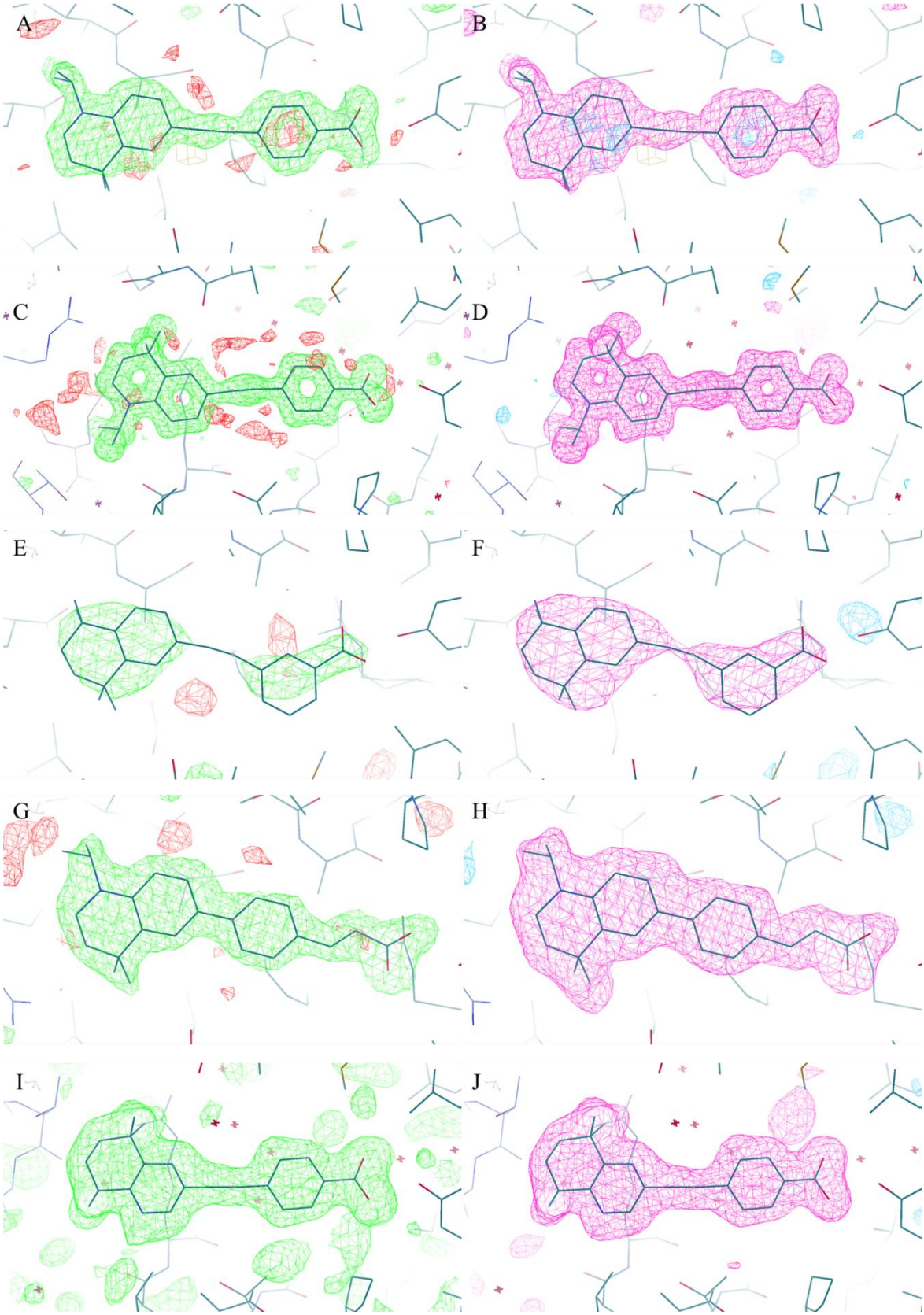
Appendix 11: Packing diagram (ab face) for CRABPII-EC19 in space group $P2_13$. Symmetry related molecules XYZ (blue), ZXY(brown), YZX (purple) and $\bar{X}$+½ ,$\bar{Y}$,Z+½ (Yellow) are shown in ribbon format. Inclined 3-fold axis shown at the origin relates XYZ, ZXY and YZX. Two-fold screw axis shown relates XYZ and $\bar{X}$+1/2, $\bar{Y}$, Z+1/2.



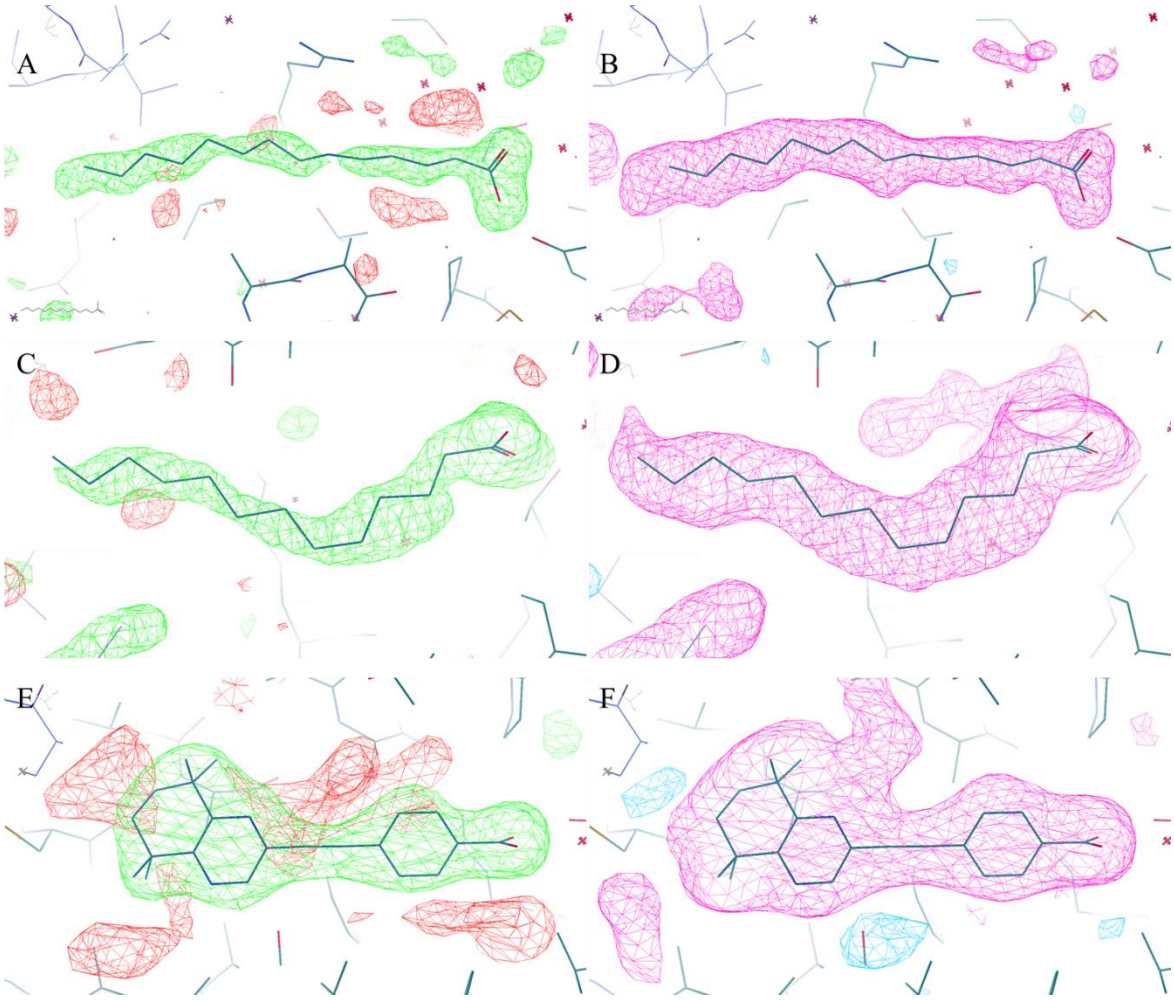
Appendix 12: Packing diagram (ab face) for CRABPII-EC19 in space group H3. Symmetry related molecules XYZ (blue), $\bar{Y}$,X-Y,Z (yellow) and $\bar{X}$+Y,$\bar{X}$,Z (orange) are shown in ribbon format. A three-fold axis at the origin relates all three molecules.



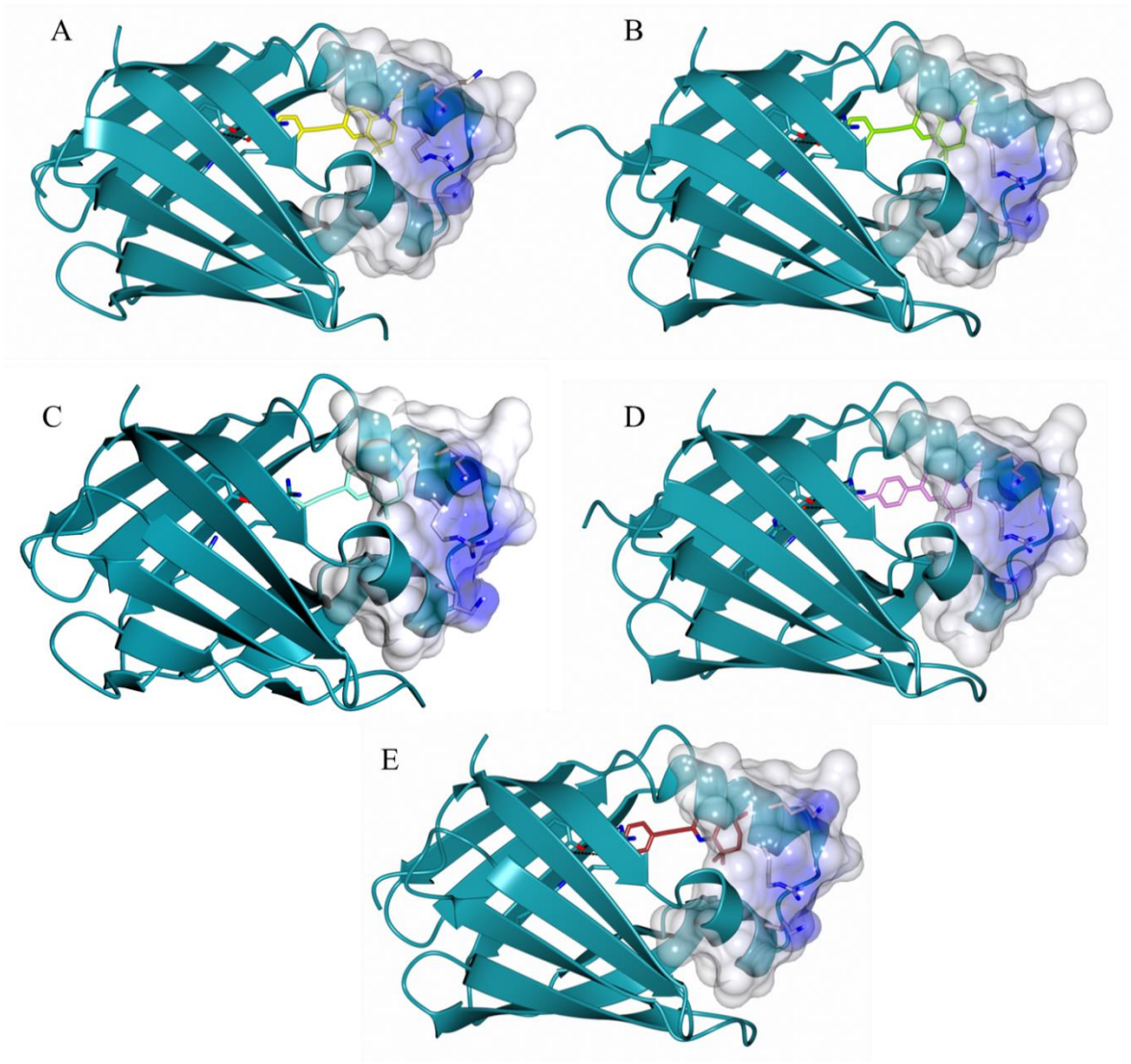
Appendix 13: Comparison of unit cells for CRABPII $P2_13$ and H3 settings. Contents of one H3 asymmetric unit displayed (ribbon, blue). Changing between the settings comes with a 3.4x change in volume (H3: 1645129 $Å^3$ *vs* $P2_13$: 483736 $Å^3$), and 4x change in asymmetric unit contents.



Appendix 14: Omit (left - green) and Polder (right - pink) maps for CRABPII with ligands DC360 (A+B), DC271 (C+D), EC19 (E+F), DC479 (G+H) and DC645 (I+J). Maps generated using phenix.polder and viewed in *Coot***.** (Liebschner *et al.*, 2019, Liebschner *et al.*, 2017, Emsley *et al.*, 2010)



Appendix 15: Omit (left - green) and Polder (right - pink) maps for CRABPI-L29C with ligands Myristic acid (A+B), Tridecanoic acid (C+D), and DC645 (E+F). Maps generated using phenix.polder. (Liebschner *et al.*, 2019, Liebschner *et al.*, 2017, Emsley *et al.*, 2010)



Appendix 16: Rear alignment of CRABPII (dark blue, ribbon diagram) with DC360 (A, yellow), DC271 (B, green), EC19, (C, aqua), DC479 (D, pink) and DC645 (E, brick red). Shown in white are NLS forming residues, overlain with electrostatic surface (white/neutral, blue/positive, calculated in CCP4mg), CRABPII-DC360 (A) contains a dual conformation of K31 due to lability in the crystal structure, CRABPII-DC271 has K31 truncated due to lability and lack of electron density. Both structures likely still form a working NLS based on the location of the residues.

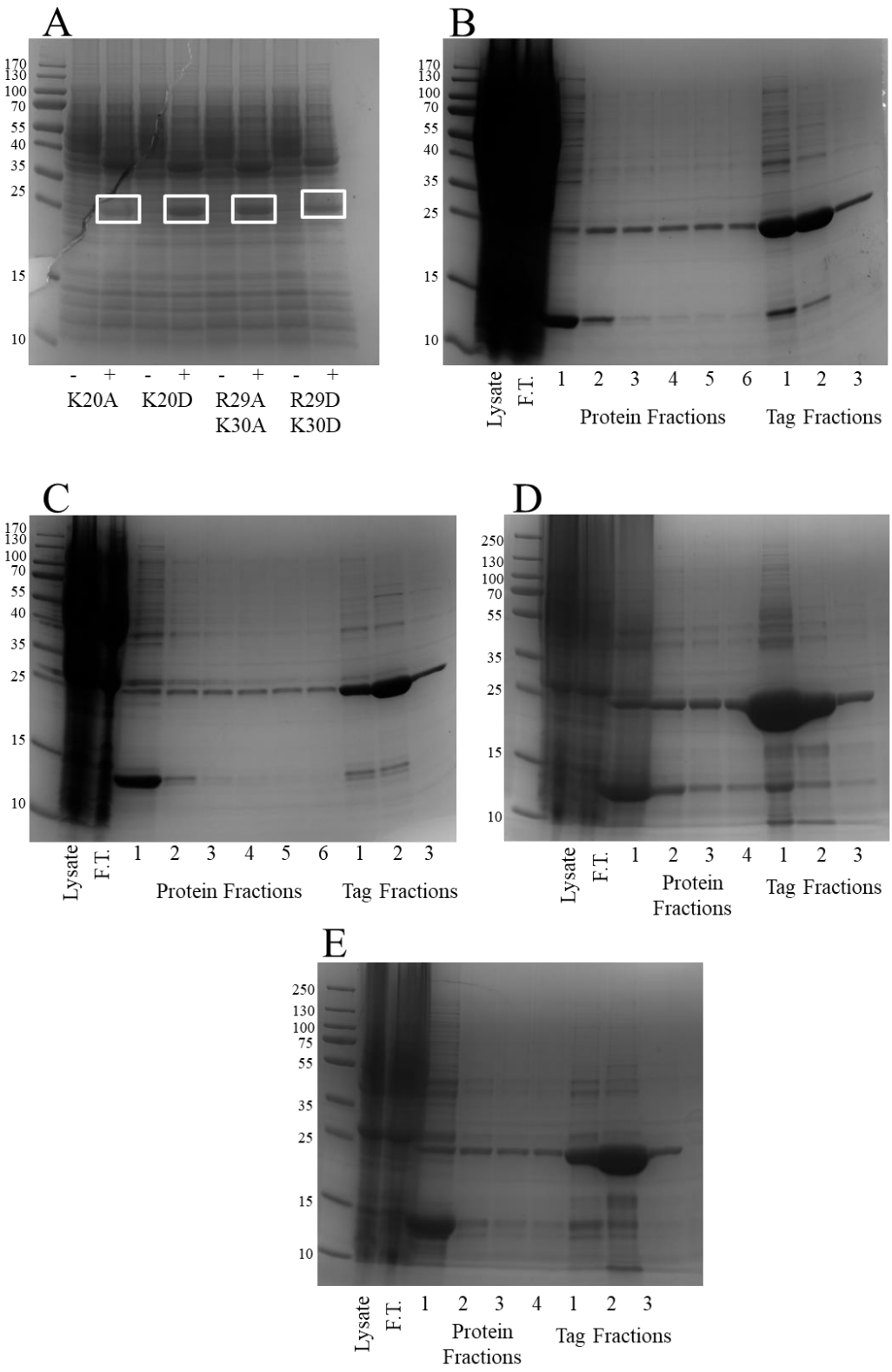
$$(1) \qquad R_{merge} = \frac{\sum_{hkl} \sum_{i} \left| I_{i(hkl)} - \bar{I}_{(hkl)} \right|}{\sum_{hkl} \sum_{i} I_{\mathbf{h}(i)}}$$

$$(2) \qquad R_{meas} = \frac{\sum_{hkl} \sqrt{\left(\frac{N}{N-1} \right)} \sum_{i} \left| I_{i(hkl)} - \bar{I}_{(hkl)} \right|}{\sum_{hkl} \sum_{i} I_{i(hkl)}}$$

$$(3) \qquad R_{p.i.m} = \frac{\sum_{hkl} \sqrt{\left(\frac{1}{N-1} \right)} \sum_{i} \left| I_{i(hkl)} - \bar{I}_{(hkl)} \right|}{\sum_{hkl} \sum_{i} I_{i(hkl)}}$$

$$(4) \qquad R / R_{free} = \frac{\sum_{hkl} w_{hkl} \left| \left| F_{obs}(hkl) \right| - k \left| F_{calc}(hkl) \right| \right|}{\sum_{hkl} \left| F_{obs}(hkl) \right|}$$

Appendix 17: Equations used by data processing for calculation of R-factors. (1) R_{merge} is the linear merging R-value, used as a data quality indicator but superseded by (2) R_{meas}, the redundancy-independent R-factor which accounts for redundancy/multiplicity of the data. (3) $R_{p.i.m}$ is the precision-indicating R-factor, indicating the variance of the distribution of intensities. (4) is the R/R_{free} calculation used to indicated agreement between the refined model and the data, where *w* are weights and *k* is the scale factor. The R calculation uses the entire data set, minus a proportion (usually 5%) set aside to calculate the R_{free} value. (Rupp, 2009, Diederichs & Karplus, 1997, Weiss & Hilgenfeld, 1997, Weiss, 2001, Brünger, 1997)



Appendix 18: Expression test SDS PAGE gel (A) and purification gels of CRABPII: K20A (B), K20D (C), R29AK30A (D), and R29DK30D (E). Molecular weight markers are in kDa. Expression bands at 25 kDa (tagged) and ~15 kDa (cleaved).

Appendix 19: Table 1: Diffraction and data processing statistics for CRABPII-R29AK30A structure. Outer shell values appear in parentheses.

PDB code	-
Diffraction source	Diamond Light Source - I24
Wavelength (Å)	0.9686
Temperature (K)	100
Detector	Pilatus 3 6M
Rotation range per image (°)	0.1
Total rotation range (°)	180
Exposure time per image (s)	0.02
Detector distance / mm	196.0
Space group	$P12_11$
***a, b, c* (Å)**	51.3, 106.0, 55.7,
α, β, γ (°)	90.0, 101.6, 90.0
Mosaicity (°)	0.0
Resolution range (Å)	54.60-1.35
Total No. of reflections	391460
No. of unique reflections	124074
Completeness (%)	96.8 (87.4)
Multiplicity	3.2 (3.2)
CC(1/2)	0.99 (0.39)
⟨ *I*/σ(*I*)⟩	5.6 (0.9)
R_{meas}	0.121 (1.239)
R_{pim}	0.068 (0.682)
Overall *B* factor from Wilson plot (Å^2)	12.65

Appendix 20: Table 2: Statistics from refinement of CRABPII-R29AK30A structure, outer shell values appear in parentheses.

Resolution range (Å)	54.60-1.35
Completeness (%)	96.8 (87.4)
No. of reflections, working set	124026
No. of reflections, test set	6289
Final R_{cryst}	0.21
Final R_{free}	0.24
	No. of non-H atoms
Protein	4426
Ligand	N/A
Water	471
Total	4897
	R.m.s. deviations
Bonds (Å)	0.0127
Angles (°)	1.736
	Average *B* factors (Å^2)
Protein	22.3
Ligand	N/A
Water	32.77
	Ramachandran plot
Favoured (%)	97.97
Allowed (%)	1.66
High energy (%)	0.37