INVESTIGATION INTO STRUCTURAL AND FUNCTIONAL RELATIONSHIPS OF SHORT-CHAIN DEHYDROGENASES AND REDUCTASES (SDRs) USING A COMPOUND LIBRARY

CHITRA BHATIA

WOLFSON COLLEGE

UNIVERISTY OF OXFORD

MICHAELMAS TERM 2018

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY SUBMITTED TO THE NUFFIELD DEPARTMENT OF ORTHOPAEDICS, RHEUMATOLOGY AND MUSUCLOSKELETAL SCIENCES (NDORMS)

DECLARATION

I declare that there are no parts of this thesis or its research herein have been reproduced or accepted for another award or degree or diploma at any other university or learning institution. This thesis contains no other person's work except where stated in text.

ACKNOWLEDGEMENTS

This PhD journey has been the most challenging, thought provoking and over whelming experience I have embarked on so far. It has been a privilege to have worked with so many talented individuals at the SGC who have shared their wealth of knowledge and experience with me over the years. There are many people to acknowledge that have made this thesis possible.

I would firstly like to thank my Supervisor Professor Udo Oppermann for his guidance, and his depth of knowledge of SDRs. A special thanks to Dr Wyatt Yue for helping me to ask the right questions, and showing me the finer but important details in the world of research. I am also very grateful to Dr Frank Niesen who first set me off onto this journey and pushed me to think outside the box. I would like to thank Professor Chas Bountra for always checking in on me and providing me with sound advice, even though it may have taken a while to absorb.

The biotechnology group has played an immense role by cloning so many of the important SDR contructs and partially adopting me into their group in the lab. My heartfelt gratitude to Dr. James Bray for his bioinformatics expertise in SDRs, and whom always motivated and steered me in the right direction. I would also like to express my gratitude to all those in the SGC who always took the time in helping, enlightening and showing me lighter side of science, especially a young lady called Caroline Sanvitale.

To the girls who have always given me unconditional support throughout the years (I would be lost without you). A special tribute goes to my husband Ashvin, for encouraging me to take the plunge and who knew I could do it even before I did. You make me a better person in very single way. Finally to my Son Luka, who I set the example of wanting to achieve and who I do everything for.

ABSTRACT

The short-chain dehydrogenases/reductases (SDR) protein family is one of the largest and most evolutionarily conserved protein superfamilies known to date. SDRs show a remarkable ability to accommodate hundreds of reactions/substrates by using the versatile nucleotide binding domain as their central scaffold. SDRs are implicated in many different disease pathways, and represent a 'druggable' enzyme class, which has generated extensive biotechnological and pharmaceutical interest. For 28 human SDR members, the three-dimensional structures have been determined. Despite this, almost 25 % of human SDR members still have unknown functions, and many members lack any chemical descriptors. To address the dearth of ligand knowledge, 28 SDRs were cloned, expressed and purified to further characterise members by the application of differential scanning fluorimetry (DSF). A family annotation through cofactor preference, substrate accommodation and subcellular localisation has shown a few distinctive traits within the SDR family. This in silico approach has led to the creation of a systematic flow chart, which provides a starting point to identify substrate classes for orphan SDRs. Cofactor and inhibitor screening of SDRs by DSF has shown T_m shifts that are in-line with previous experimental data. DSF screening showed the decryption of 6 cofactor preferences for orphan SDRs, five of which have been confirmed in crystal structures. In contrast, screening of substrates in DSF produce little thermal stability (15 % detection rate), rendering it difficult to distinguish specific substrates for orphan SDRs. However, over 70 % of SDRs screened in DSF with known enzyme activity produce thermal shifts with compounds that resemble their substrate structures. This can therefore infer substrate classes for orphan SDRs that compare well with those from computational predictions. DSF screening against a ligand library has led to the discovery of the first small

molecule binders for Nmr-like family domain-conatining protein 1 (NMRAL1) and selective inhibitors for Hydroxyprostaglandin dehydrogenase (HPGD). The use of small molecule ligands identified from DSF has also contributed to the success of crystal formation in the case of Hydroxysteroid dehydrogenase-like protein 2 (HSDL2) and Dehydrogenase/reductase SDR family member 2 (DHRS2). Overall this study has provided a platform for ligand screening to further structurally characterise the SDR family.

CONTENTS

DECLARATION 2

ACKNOWLEDGEMENTS 3

ABSTRACT 4

CONTENTS 6

LIST OF FIGURES 10

LIST OF TABLES 13

ABBREVIATIONS 15

CHAPTER 1: INTRODUCTION 20

1.1 DEHYDROGENASES AND REDUCTASES 20

1.2 THE ROSSMANN-FOLD AS A STRUCTURAL MOTIF IN ENZYMES 23

1.3 ALCOHOL DEHYDROGENASES 25

1.4 SHORT-CHAIN DEHYDROGENASES/REDUCTASES 26

1.4.1 BACKGROUND 26

1.4.2 STRUCTURAL FEATURES 27

1.4.3 CONSERVED MOTIFS AND RESIDUES 29

1.4.4 SDR FOLDS AND TYPES 32

1.4.5 CATALYTIC MECHANISM OF SDRs 39

1.5 EVOLUTION AND FAMILY COMPARISON 42

1.5.1 GENOME COMPARISONS 42

1.5.2 FAMILY COMPARISONS 43

1.6 MEDICAL RELEVANCE 46

1.7 AIMS OF THESIS 50

1.7.1 STRUCTURE OF THESIS 52

CHAPTER 2: MATERIALS AND METHODS 53

2.1 INTRODUCTION 53

2.2 DELETION CONSTRUCTS 54

2.3 ESCHERICHIA COLI EXPRESSION 55

2.3.1 CLONING 55

2.3.2 TRANSFORMATION 56

2.3.3 EXPRESSION....56
2.4 BACULOVIRUS PRODUCTION AND EXPRESSION....57
2.4.1 INTRODUCTION....57
2.5 PURIFICATION....59
2.5.1 PURIFICATION BUFFERS....59
2.5.2 TEST PUFRICATION FROM E.COLI....59
2.5.3 TEST PURIFICATION FROM BACULOVIRUS....61
2.5.4 LARGE SCALE PURIFICATION FROM E.COLI....61
2.5.5 LARGE-SCALE PURIFICATION FROM BACCULOVIRUS....64
2.6 PROTEIN VERIFICATION BY MASS SPECTROMETRY....64
2.7 ANALYTICAL GEL FILTRATION....65
2.8 PROTEIN PRODUCTION SUMMARY....66
2.9 DIFFERENTIAL SCANNING FLUORIMETRY....69
2.9.1 BACKGROUND....69
2.9.2 EXPERIMENTAL SETUP FOR COFACTOR DETERMINATION....70
2.9.3 EXPERIMENTAL SETUP FOR LIGAND SCREENING....71
2.10 X-RAY CRYSATALLOGRAPHY....72
2.10.1 BACKGROUND....72
2.10.2 HIGH-THROUGHPUT CRYSTALLISATION OF PROTEINS....76
2.10.3 CRYSTAL MOUNTING AND DATA COLLECTION....76
2.10.4 MODEL BUILDING AND REFINEMENT....77
2.11 SUBSTRATE SCREENING FOR ORPHAN SDRS....77
CHAPTER 3: SDR FAMILY ANNOTATION....79
3.1 INTRODUCTION....79
3.2 COFACTOR DISTRIBUTION OF HUMAN SDRS....83
3.3 SUBSTRATE DISTRIBUTION....86
3.4. SUBCELLULAR LOCALISATION....87
3.4.1 SUBCELLULAR LOCALISATION DISTRIBUTION....87
3.4.2 SUBCELLULAR LOCALISATION AND SUBSTRATE CLASS ANALYSIS....88
3.4.3 TARGETING SIGNALS....89
3.5 TRANSMEMBRANE ASSOCIATIONS....92
3.6 SUBCELLULAR LOCALISATION PREDICTION....94

3.7 DISCUSSION 96

CHAPTER 4 – COFACTOR DETERMINATION BY DIFFERENTIAL SCANNING FLUORIMETRY (DSF) 110

4.1 INTRODUCTION 110

4.2 COFACTOR DETERMINATION BY TM SHIFT 112

4.3 COFACTOR PREDICTION BY SEQUENCE ANALYSIS 113

4.4 COFACTOR DETERMINATION RESULTS 115

4.5 COFACTOR DETERMINATION BY UNFOLDING GRADIENT (ΔI/Δ°C) 117

4.6 DISCUSSION 119

CHAPTER 5 - LIGAND IDENTIFICATION BY DSF 122

5.1 INTRODUCTION 122

5.2 LIGAND LIBRARY COMPOSITION 123

5.3 SUBSTRATE IDENTIFICATION BY DSF 124

5.3.1 IDENTIFICATION OF SUBSTRATES OF FUNCTIONALLY ANNOTATED SDRS BY DSF 124

5.3.2 SUBSTRATE CLASS IDENTIFICATION BY DSF 133

5.3.3 DETECTING SUBSTRATE SCAFFOLDS 134

5.3.4 SUBSTRATE CLASS ANALYSIS FOR ORPHAN SDRs 138

5.4 INHIBITOR IDENTIFICAION BY DSF 141

5.5 HIGH-AFFINITY INHIBITOR DETECTION FOR HPGD 143

5.5.1 LIGAND LIBRARY DSF SCREENING FOR HPGD 143

5.5.2 INVESTIGATION OF SELECTIVE HPGD INHIBITORS BY DSF 146

5.6 IDENTIFICATION OF LIGANDS FOR REGULATORY SDRs 148

5.7 PRIORTISING SDR TM SHIFTS 149

5.8 DISCUSSION 151

CHAPTER 6 – APPLICATION OF THERMAL SHIFT DATA FOR STRUCTURAL STUDIES 162

6.1 INTRODUCTION 162

6.2 DHRS4 STRUCTURAL ELUCIDATION BY COFACTOR THERMAL STABILITY 163

6.2.1 DHRS4 CRYSTALLISATION AND STRUCTURE DETERMINATION 164

6.2.2. OVERALL STRUCTURE 166

6.2.3 COFACTOR BINDING AND ACTIVE SITE ARCHITECTURE 168

6.3 STRUCTURAL CHARACTERISATION OF SMALL-MOLECULE LIGANDS FOR NMRAL1 . 170

6.3.1 BACKGROUND 170
6.3.2 CO-CRYSTALLISATION AND STRUCTURE DETERMINATION OF NMRAL1 WITH NFL AND ZZ0 171
6.3.3 CRYSTAL STRUCTURE REVEALS A HYDROPHOBIC POCKET FOR SMALL-MOLECULE BINDERS 172
6.3.4 OLIGOMERIC STATE OF NMRAL1 175
6.4 CRYSTAL FORMATION OF HSDL2 AND DHRS2 USING AN IDENTIFIED COMPOUNDS FROM DSF 177
6.4.1 CRYSTALLISATION TRIALS OF DHRS2B AND HSDL2 178
6.5 SUBSTRATE SCREENING FOR ORPHAN SDRS 183
6.5.1 ACTIVITY SCREENING RESULTS 183
6.5.2 HSDL2 AND DHRS11 ACTIVE SITE ANALYSIS 183
6.6 DISCUSSION 185
CHAPTER 7 – CONCLUSIONS AND FUTURE WORK 190
7.1 CONCLUSIONS 190
7.1.1 PROTEIN PRODUCTION 190
7.1.2 SDR FAMILY ANNONATION 191
7.1.3 SDR COFACTOR AND LIGAND SCREENING 193
7.1.4 STRUCTURAL ELUCIDATION OF SDRS USING THERMAL SHIFT DATA 196
7.2 FUTURE WORK 197
7.3 SUMMARY 199
REFERENCES 200
APPENDIX 221

LIST OF FIGURES

Figure 1.1 Structure of Nicotinamide Adenine Diphosphates..... 22

Figure 1.2 Classic dehydrogenase Rossmann Fold..... 24

Figure 1.4.1 Diversity in SDR substrate examples..... 27

Figure 1.4.2.1 SDR reaction types..... 30

Figure 1.4.4.1 SDR schematic diagrams to highlight the ligand binding regions that are significantly different between the two folds..... 33

Figure 1.4.4.2 Schematic diagram of the sequence clusters in the human SDR superfamily..... 35

Figure 1.4.4.3 Comparison of C-terminus alignment of SDR slusters A) classical type I SDR B) Classical type II and III – membrane bound C) Extended Fold – Epimerase family. 37

Figure 1.4.5.1 Postulated reductive reaction mechanism of 3β/17β-HSD involving NADH and steroid substrate (5α-androstane, 3-one, 17ol)..... 40

Figure 1.4.5.2 Active site features of the 'divergent' SDR dienoyl-CoA reductase (1w6u). 41

Figure 1.5.2.1 Ribbon diagram comparison of classical SDR, MDR, LDR, AKR and halohydrin halogenase enzymes..... 44

Figure 1.5.2.2 Superimposed Tyr and Lys active site of classical SDR and AKR enzymes..... 45

Figure 2.1.1 SGC core workflow processes..... 54

Figure 2.4.1 Transposition process of the donor plasmid into bacmid..... 58

Figure 2.5.2.1 50 mL test expression SDS-PAGE of NMRAL1 constructs from PNIC28Bsa4 60

Figure 2.5.4.1 Flow chart of the general purification process for SDR proteins..... 61

Figure 2.5.4.2: Schematic of the gel filtration chromatography method (GE Healthcare). 63

Figure 2.5.4.3 SDS-PAGE of DHRS4- M1-V250..... 64

Figure 2.6.1 Mass spectrometry of DHRS4 before (A) and after TEV cleavage (B)..... 65

Figure 2.9.1.1 DSF unfolding states and midpoint calculation..... 70

Figure 2.10.1.1 Crystal formation..... 74

Figure 2.11.1 SDR Substrate screen setup..... 78

Figure 3.1.1 Venn diagram of SDR substrate specificity 80

Figure 3.1.2 Human SDRs analysed by cofactor, substrate class and subcellular localisation..... 82

Figure 3.2.1 Human SDR Cofactor preference distribution..... 83

Figure 3.3.1 SDR substrate distribution in relation to SDR clusters..... 86

Figure 3.4.1.1 Subcellular localisation in SDR clusters..... 87

Figure 3.4.2.1 SDR substrate distribution in relation to subcellular localisation..... 88

Figure 3.4.3.1 Analyses of targeting signals in human SDRs..... 91

Figure 3.5.1 TM helix distribution..... 94

Figure 3.7.1 HSDL1 active site evolution..... 98

Figure 3.7.2 Sequence alignment of N-terminal of DECR1 proteins..... 100

Figure 3.7.3 Flow chart used in this work to identify potential substrate classes for human orphan SDR..... 104

Figure 3.7.4 Human SDR family subcellular localisation and substrate accommodation diagram..... 109

Figure 4.1 Comparison of the folding topologies between the crystal structures of apo and holo-forms of 3-hydroxysteroid dehydrogenase from *Pseudomonas sp B-0831*..... 111

Figure 4.3.1 NAD vs. NADP residue binding..... 113

Figure 4.5.1 TSTA3 cofactor determination by T_m 117

Figure 4.5.2 Human SDR cofactor determination by dsf slope of transition data..... 118

Figure 4.6.1 TSTA3 cofactor accomodation..... 120

Figure 5.2.1 SDR ligand library for DSF..... 123

Figure 5.3.2.1 SDR Substrate classes for DSF screening..... 133

Figure 5.3.2.2 Substrate class identification by DSF..... 134

Figure 5.5.1.1 Biosynthesis of Eicasanoids..... 144

Figure 5.5.2.1 Correlation between thermal stabilisation of HPGD in the presence of A) NAD B) NADH…… 147

Figure 5.7.1 T_m shift variable analysis…… 150

Figure 5.8.1 Human and Bacterial GALE residues that interact with GDP-glucose…….... 153

Figure 5.8.2 GMDS active site in complex with NADP and GDP…… 155

Figure 5.8.3 Active site accommodation of Human and Mouse HSD11β1…… 158

Figure 5.8.4 Crystal structure of Human MAT2B…….…….... 160

Figure 6.2.1.1 DHRS4 crystal with NADPH…….…….... 165

Figure 6.2.2.1 DHRS4 structure…… 167

Figure 6.2.3.1 DHRS4 cofactor binding and active site coordination…… 169

Figure 6.3.2.1 - NMRAL1 crystals with A) $NMRAL1_{NFL}$- 1.8 Å diffraction B) $NMRAL1_{ZZO}$… 171

Figure 6.3.3.1 NMRAL1 structural form comparison…… 174

Figure 6.3.4.1 NMRAL1 dimer formation in the crystal…… 175

Figure 6.3.4.2 NMRAL1 analytical gel filtration (Superdex 75 10/30) …….. 177

Figure 6.4.1.1 Crystal of A) DHRS2B and B) HSDL2…… 179

Figure 6.4.2.1 HSDL2 in complex with NADP…… 182

Figure 6.5.2.1 Active site coordination of HSDL2 and DHRS11…… 185

Figure 7.1.2.1 Metabolism of NAD(P) in eukaryotic cells…… 192

LIST OF TABLES

Table 1.1 Oxidoreductase EC subclasses..... 23

Table 1.4.3.1 Sequence motifs found in SDR enzymes..... 32

Table 1.4.4.1 Conserved sequence motifs in the classical and the extended SDR families 34

Table 1.4.4.2 SDR Types and their designations..... 38

Table 1.4.5.1 Active site motif alternatives to YxxxK 41

Table 1.6.1 Human SDRs and their associated diseases in the OMIM database..... 49

Table 2.8.1 Protein production summary of human SDRs..... 68

Table 2.9.2.1 Master mix components for coenzyme DSF screen..... 71

Table 2.9.3.1 Master mix components for ligand DSF screen..... 72

Table 2.11.1 Activity screening setup for HSDL2 and DHRS11..... 78

Table 3.2.1 Comparison of NAD/H and NADP/H SDR genomes..... 84

Table 3.4.3.1 Summary of targeting signals in eukaryotic proteins..... 90

Table 3.6.1 Evaluation of subcellular prediction methods on human SDR family..... 95

Table 3.7.1 Human SDR cofactor preferences summary by subcellular localisation and substrate class..... 98

Table 3.7.2 SDR Cluster summary of predominant forms of cofactor preference, substrate class and subcellular localisation..... 99

Table 3.7.3 Predominant substrate class from subcellular localisation analysis..... 101

Table 3.7.4 Substrate prediction for human orphan SDRS..... 107

Table 4.4.1 Cofactor determination summary and comparison T_m Shift data to other data sources..... 115

Table 5.3.1.1 DHRS2B T_m shift summary..... 124

Table 5.3.1.2 GALE T_m shift summary..... 126

Table 5.3.1.3 USX1 nucleotide sugars T_m shifts..... 127

Table 5.3.1.4 PECR CoA substrate T_m shift summary..... 129

Table 5.3.1.5 HPGD prostaglandin T_m shifts..... 130

Table 5.3.1.6 HSD17β10 steroid T_m shifts..... 132

Table 5.3.3.1 GMDS nucleotide T_m shifts..... 135

Table 5.3.3.2 HSD17 β8 hydroxyacl-CoA substrate T_m shifts..... 136

Table 5.3.3.3 HSD17β8 steroid compounds shifts with NADH..... 138

Table 5.3.4.1 HSDL2 shift summary..... 139

Table 5.3.4.2 DHRS1 shift summary..... 140

Table 5.3.4.3 DHRS11 shift summary..... 141

Table 5.3.4.4 DSF substrate group identification for orphan SDRs..... 141

Table 5.4.1. HSD11β1 steroid shift summary..... 142

Table 5.5.1.1 HPGD inhibitor T_m shift summary..... 146

Table 5.6.1 NMARL1 T_m shifts..... 149

Table 5.8.1 Summary of T_m substrate shifts (≥ 1.4°C) for DSF screened SDRs..... 152

Table 5.8.2 DSF substrate class identification for orphan SDRs..... 156

Table 6.2.1.1 DHSR4 co-crystallisation trials..... 165

Table 6.3.2.1 Co-crystallisation details of NMRAL1 with NFL and ZZO..... 171

Table 6.3.4.1 PDBePISA NMRAL1 dimer formation in solution..... 176

Table 6.4.1.1 DHRS2B and HSDL2 crystallisation trials with cofactor and DSF identified compounds summary..... 179

Table 6.5.1.1 Substrate class screening results..... 183

Table 6.6.1 Human SDR crystal formation with cofactor redox states..... 186

ABBREVIATIONS

5α-DHT	Dihydrotestosterone
AA	Amino acid
ADH	Alcohol dehydrogenase
ADP	Adenosine diphophosphate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
AKR	Aldo-keto reductases
ASS1	Arginosuccinate synthase
BDH2	3-hydroxybutyrate dehydrogenase, type 2
CBR1	Carbonyl reductase 1
CBR3	Carbonyl reductase 3
CBX	Carbenoxolone
cDNA	Complimentary deoxyribonucleic acid
CoA	Coenzyme A
Cryo-EM	Cryo electron microscopy
DECR1	2,4-dienoyl-CoA reductase, mitochondrial
DECR2	2,4-dienoyl-CoA reductase, peroxisomal
DCXR	Dicarbonyl/L-xylulose reductase
DHRS1	Dehydrogenase/reductase (SDR family) member 1
DHRS11	Dehydrogenase/reductase (SDR family) member 11
DHRS2B	Dehydrogenase/reductase (SDR family) member 2B
DHRS4	Dehydrogenase/reductase (SDR family) member 4
DHRS7B	Dehydrogenase/reductase (SDR family) member 7B

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSF	Differential Scanning Fluorimetry
EC	Enzyme Commission
EDTA	Ethylene Diamine Tetraacetic Acid
ER	Endoplasmic reticulum
ESI-TOF	Electrospray ionisation time-of-flight
FAD	Flavin adenine dinucleotide
FAS	Fatty acid synthase
FMN	Flavin mononucleotide
GALE	UDP-glucose-4-epimerase
GDP	Guanosine diphosphate
GF	Gel filtration
GMDS	GDP-mannose 4,6-dehydratase
GMP	Guanosine monophosphate
GPDH	Glyceraldehyde-3-phosphate dehydrogenase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGNC	HUGO Gene Nomenclature Committee
HCl	Hydrochloric acid
HPGD	15-hydroxyprostaglandin dehydrogenase
HSD	Hydroxysteroid dehydrogenase
HSD11β1	11β-hydroxysteroid dehydrogenase 1
HSD11β2	11β-hydroxysteroid dehydrogenase 2
HSD17β10	17β-hydroxysteroid dehydrogenase 10

HSD17β12	17β-hydroxysteroid dehydrogenase 12
HSD17β4	17β-hydroxysteroid dehydrogenase 4
HSD17β8	17β-hydroxysteroid dehydrogenase 8
HSDL2	Hydroxysteriod dehydrogenase like 2
HTATIP2	HIV-1 Tat interactive protein 2
HUGO	Human genome organisation
IMAC	Immobilised Metal Ion Affinity Chromatography
IPTG	Isopropyl-Beta-D-Thiogalactoside
KDa	Kilodaltons
L	Litre
LADH	Liver alcohol dehydrogenase/hydrogenase
LDH	Lactate dehydrogenase
LDR	Long-chain dehydrogenase/reductase
LIC	Ligation-independent cloning
m	Milli
M	Molar
MAT2B	Methionine adenosyltransferase II, β
MDH	Malate dehydrogenase
MDR	Medium-chain dehydrogenase/reductases
MOI	Molarity of infection
MPD	2-methyl-2,4-pentadiol
MTS	Mitochondrial targeting signal
MWCO	Molecular Weight Cut-Off
NAD/H	Nicotinamide adenine dinucleotide (oxidised/reduced)

NADP/H	Nicotinamide adenine dinucleotide phosphate (oxidised/reduced)
NFL	Niflumic acid
Ni-NTA	Nickel-nitriloacetic acid
NLS	Nucleus localisation signal
NMR	Nuclear magnetic resonance
NMRAL1	NmrA-like family domain-containing protein 1
NO	Nitric oxide
NSAIDS	Non-steroidal anti-inflammatory drugs
OMIM	Online Mendelian Inheritance in Man
PECR	Peroxisomal trans-2-enoyl-CoA reductase
PEI	Polyethylenimine
PGE2	Prostaglandin PGE2
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PEG	Polyethylene glycol
pI	Isoelectric point
PPAR	Peroxisome proliferate-activated receptor
PTS	Peroxisomal targeting signal
QDPR	Quinoid dihydropteridine reductase
qRT-PCR	quantitative real time PCR
Rcf	Relative centrifugal force
RDH	Retinol dehydrogenases
RNA	Ribonucleic acid
SAM	S-adenosyl-L-methionine-dependent methyltransferase

SDR	Short-chain dehydrogenase/reductases
SDR39U1	Short chain dehydrogenase/reductase family 39U, member 1
SDS-PAGE	SDS-polyacrylamide gel
SEC	Size Exclusion Chromatography
SGC	Structural Genomics Consortium
SPR	Surface plasmon resonance
SPR	Sepiapterin reductase
TB	Terrific Broth
TCEP	Tris(2-carboxyethyl)phosphine
TDP	Thiamine diphosphate
TEV	Tobacco Etch Virus
TIM	Triophosphate isomerase
T_m	midpoint of transition of thermal unfolding
TM	Transmembrane
TSTA3	Tissue specific transplantation antigen P35B
µ	Micro
UDP	Uridine diphosphate
UTP	Uridine triphosphate
UXS1	UDP-glucuronate decarboxylase 1
v/v	volume by volume
WWOX	WW domain containing oxidoreductase
w/v	Weight per volume
ZZO	2-(4-chloro)-anilino]-nicotinic acid

CHAPTER 1: INTRODUCTION

1.1 DEHYDROGENASES AND REDUCTASES

Most cellular reactions occur through enzymes that serve as biological catalysts and are crucial to all life forms. The majority of enzymes are proteins which broadly subdivide into six groups based on their chemical reaction types namely oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases (Young 1977, Yamanishi et al. 2009). An oxidoreductase is an enzyme that specifically catalyses the transfer of electrons from a reductant (electron donor) to an oxidant (electron acceptor). The reaction generally obeys the following scheme:

$$A^- + B \rightarrow A + B^-$$

where A is the reductant and B is the oxidant. Oxidoreductases are among the most widespread enzymes in living organisms, playing a vital role in many metabolic and biosynthetic pathways (Moller et al. 1993, Upcroft and Upcroft 1999, Kerscher et al. 2000).

Oxidoreductases typically employ a catalytic component known as a cofactor, which form part of the active site of an enzyme upon binding to a protein for biological activity. An apoenzyme is the inactive enzyme without cofactor, while the complete enzyme with cofactor is the holoenzyme. There are two types of cofactors: organic cofactors (also known as coenzymes), such as flavin, or NAD(P)(H), and inorganic cofactors, such as the metal ions Mg^{2+}, Cu^{2+}, Zn^{2+}, or iron-sulphur clusters. Coenzymes are small molecules compared to the protein part of the enzyme. Vitamins often function as precursors of coenzymes (e.g., vitamin B1, B2, B6, B12, riboflavin and folic acid) or as coenzymes themselves (e.g., vitamin

C is a cofactor used by enzymes involved in mammalian collagen synthesis) (Burmeister et al. 2000, Linster and Van Schaftingen 2007, Mandl et al. 2009). However, vitamins can have other functions in the body, for example binding and activating steroid receptors, or being involved in carboxylation reactions (Bolander 2006). Many organic cofactors also include a nucleotide, such as the electron carriers Nicotinamide Adenine Dinucleotide (NAD) (Le-Quoc and Le-Quoc 1989), Flavin Adenine Dinucleotide (FAD) (Manson and Modi 1957), or coenzyme A (CoA) (Watmough and Frerman 2010), which carries acyl groups involved in transfer reactions. Oxidoreduactses can be oxidases or reductases. Oxidases utilise dioxygen as an electron acceptor of hydride ions, whereas dehydrogenases are enzymes that transfer hydrogen to an acceptor e.g. NAD^+, $NADP^+$ or a FAD.

NAD (Figure 1.1A) consists of an adenine (magenta) and nicotinamide (green) connected through their phosphate groups (blue) and ribose. Nicotinamide is derived from niacin, otherwise known as Vitamin B3. NAD functions as an electron donor in many enzymatic reactions, for example, glycolysis (Masters et al. 1987). NAD is also required for non-redox reactions, such as involvement in transfer reactions of adenosine diphosphate (ADP) ribose (Ziegler 2000). Nicotinamide adenine dinucleotide phosphate (NADP) (Figure 1.1.B) is chemically similar to NAD, with only the addition of a phosphate group on the 2' position of the ribose ring in NADP. However, $NADP^+$ functions in reductive biosyntheses, such as fatty acid, lipid and cholesterol synthesis in mammals, which requires NADPH as a reducing agent (Murai et al. 1971, Bloch 1975). Some mitochondrial enzymes can synthesise NAD into NADP (Magni et al. 2006). The principal source is the pentose phosphate pathway

(Revilla et al, 1987) followed by NADP-dependent such as malic enzyme, isocitrate dehydrogenase and glutamate dehydrogenase (Hanukoglu and Rapoport 1995).

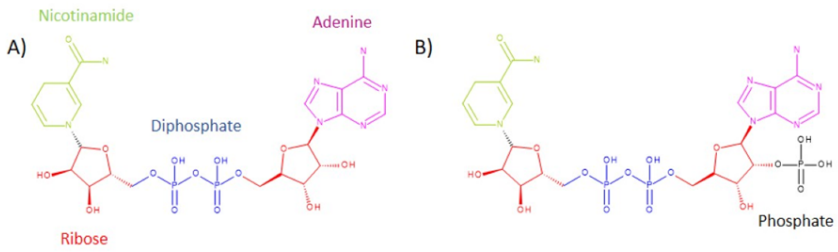


Figure 1.1 Structure of Nicotinamide Adenine Diphosphates; A) NAD B) NADP –nicotinamide (green), ribose (red), diphosphate (blue), adenine (magenta)

$$NADPH + B + H^{+} \leftrightarrow BH_2 + NADP^{+}$$

There are currently 22 divisions of oxidoreductase reactions within the Enzyme Commission (EC) members (Korzybski 1962, Thompson 1962, Matkovics et al. 1972), which is a scheme for enzymes based on the different groups of donors they act on, such as alcohols, aldehydes, sulphur, heme and phenols. For example, oxidoreductases that act on CH-OH groups are classified as EC 1.1 and comprise several subclasses (Table 1.1).

The continuous requirement for NAD results from the coenzyme consumption in regulatory pathways and reactions, such as glycolysis in the cytoplasm and oxidative phosphorylation in the mitochondria (Belenky et al. 2007, Canto et al. 2015). Redox reactions between oxidised and reduced forms of NAD do not change the overall levels of the coenzyme (Belenky et al. 2007). Most cofactors are universal to all forms of life, except for methanogens that have evolved unique cofactors such as, coenzyme B, methanofuran and tetrahydromethanopterin (Rouviere and Wolfe 1988).

EC Number	Acts on
EC 1.1	CH-OH group of donors
EC 1.2	Aldehyde or oxo group of donors
EC 1.3	CH-CH group of donors
EC 1.4	CH-NH_2 group of donors
EC 1.5	CH-NH group of donors
EC 1.6	NAD(H) or NADP(H)
EC 1.7	Other nitrogenous compounds as donors
EC 1.8	Sulphur group of donors
EC 1.9	Heme group of donors
EC 1.10	Diphenols and related substances as donors
EC 1.11	Peroxide as an acceptor
EC 1.12	Hydrogen as donors
EC 1.13	Single donors with incorporation of molecular oxygen
EC 1.14	Paired donors with incorporation of molecular oxygen
EC 1.15	Superoxide radicals as acceptors
EC 1.16	Oxidise metal ions
EC 1.17	CH or CH_2 groups
EC 1.18	Iron-sulphur proteins as donors
EC 1.19	Reduced flavodoxin as a donor
EC 1.20	Phosphorus or arsenic in donors
EC 1.21	X-H and Y-H to form an X-Y bond
EC 1.97	Other oxidoreductases

Table 1.1 Oxidoreductase EC 1 subclasses – Nomenclature committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) (Source:https://www.qmul.ac.uk/sbcs/iubmb/enzyme/EC1/).

1.2 THE ROSSMANN-FOLD AS A STRUCTURAL MOTIF IN ENZYMES

In a metabolic context, oxidoreductases are commonly referred to as dehydrogenases or reductases. Dehydrogenases and reductases utilising nicotinamide adenine dinucleotides [NAD and NADP] are confined to the EC 1.1 subclass (Table 1.1) being the most prevalent. Many nucleotide binding proteins display a common structural fold, known as the Rossmann-fold (Lesk 1995). Named after Michael Rossmann, who observed the frequently occurring fold in nucleotide binding proteins from the alignment of four dehydrogenases, lactate dehydrogenase (LDH), malate dehydrogenase (MDH), liver alcohol dehydrogenase

(LADH) and glyceraldehyde-3-phosphate dehydrogenase (GPDH) (Rossmann et al. 1974), the fold occurs across several enzyme families.

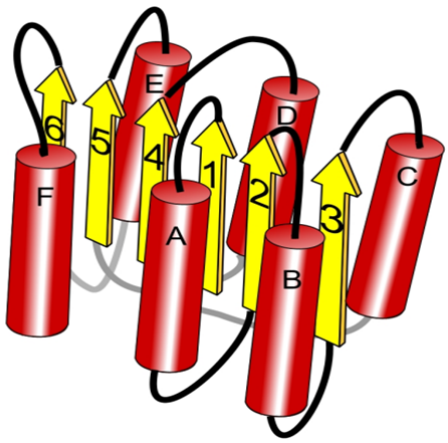


Figure 1.2 Classic dehydrogenase Rossmann Fold: The order is 1-A-2-B-3-C-4-D-5-E-6-F where numbers are β-strands (yellow) and letters are α-helices (red). (http://www.ebi.ac.uk/pdbe/quips?story=Phaser)

The Rossmann fold structure in Figure 1.2 shows a core comprised of six parallel beta strands flanked by two pairs of antiparallel alpha helices in the order β-α-β-α-β. Therefore, the helices create a layer packing against the sheet (Rao and Rossmann 1973). Rossmann folds can exist as pairs and bind one dinucleotide molecule, such as NAD(P), per Rossmann fold. Single Rossmann folds can bind mononucleotides such as the cofactor flavin mononucleotide (FMN) (Rao and Rossmann 1973) and SAM (S-adenosyl-L-methionine-dependent methyltransferase) binding in methyl transferases (Yue et al. 2007).

Thought to have evolved early in proteobacteria (Duax et al. 2009) the NAD Rossmann fold is one of the oldest and most common protein folds, and is observed in a large number of enzyme families. Examples include, flavin binding oxidoreductases (Macheroux et al. 2011), lactate dehydrogenases (Read et al. 2001) and Zn containing-alcohol dehydrogenases

(Jornvall et al. 2003, Nordling et al. 2002a) (ADHs; members of the family of medium-chain dehydrogenases/reductases, MDRs), demonstrating the functional diversity of the Rossmann fold (Danielsson et al. 1994, Lesk 1995, Kallberg and Persson 2006). In the PDB alone, well over 10,000 structures with Rossmann fold have been deposited, of which 1895 are NAD(P)-binding Rossmann domains. Protein families that harbour the NAD(P)-binding Rossman domain include the Long, medium and short-chain dehydrogenase/reductase oxidoreductases, that are found across genomes from all kingdoms of life (Yang et al. 2005a). However, some NAD-dependent proteins use different topologies, for example, a class of bacterial enzymes involved in amino acid (AA) metabolism bind NAD, but lack the Rossmann-fold (Goto et al. 2005).

1.3 ALCOHOL DEHYDROGENASES

Alcohol dehydrogenases (ADH) (EC 1.1.1.1) encompases a large group of enzymes that catalyse the reversible reaction converting alcohols to ketones or aldehydes using NAD(P)(H).

$$\boldsymbol{Alcohol + NAD(P) \rightleftharpoons Aldehyde + NAD(P)H}$$

ADH reactions can serve as a general detoxification process in animals, and they participate in the biodegradation of various metabolites. Unlike other enzymes that carry out similar processes, such as cytochrome P450 and glutathione S-transferases (Nelson 1999), the dehydrogenase reaction lacks free radical formation that can damage cells and contribute to disease.

The comparison of peptide sequences from insect, yeast and mammalian ADHs illustrated differences in protein type and regions of importance in dehydrogenases (Jornvall et al.

1981, Jornvall et al. 1984a) leading to a system of family distinctions of NAD(P)(H)-dependent dehydrogenases (Jornvall et al. 2010, Kallberg et al. 2002a, Oppermann et al. 2003, Persson et al. 1991a). Evolutionary divergence within the ADHs has given rise to five superfamilies; MDRs, SDRs, long-chain dehydrogenases/reductases (LDR), aldo-keto reductases (AKR) and iron activated alcohol dehydrogenases (Nordling et al. 2002a, Nordling et al. 2002b).

1.4 SHORT-CHAIN DEHYDROGENASES/REDUCTASES

1.4.1 BACKGROUND

Evolutionary events have created protein families with similar activities. Analysis of distinct sequence motifs, protein chain-lengths, mechanistic and structural features, a classification of short-, medium- and long-chain dehydrogenases/reductases have been established (Persson et al. 1991a, Jornvall et al. 1995, Nordling et al. 2002a). With the availability of many new genomes from genome sequencing projects, it is now evident that SDRs represent a large group of gene products within almost every genome (Pearl et al. 2005), and possibly constitute the largest protein class known to date (Bray et al. 2009, Yooseph et al. 2007).

The SDR family spans across multiple kingdoms, including plant, archaea, bacteria and animals (Kallberg et al. 2002a, Kallberg et al. 2002b, Persson et al. 2003, Webb et al. 2004, Gavidia et al. 2007, Belyaeva et al. 2009, Kisiela et al. 2012), and currently 73 SDR members have been identified (Persson et al. 2009) within the human genome (Appendix, Table 1). There are over 148,000 members in sequence databases from prokaryotic-archaea and

eukaryotic genomes, and over 280 crystal structures have been deposited in the Protein Database Bank (PDB).

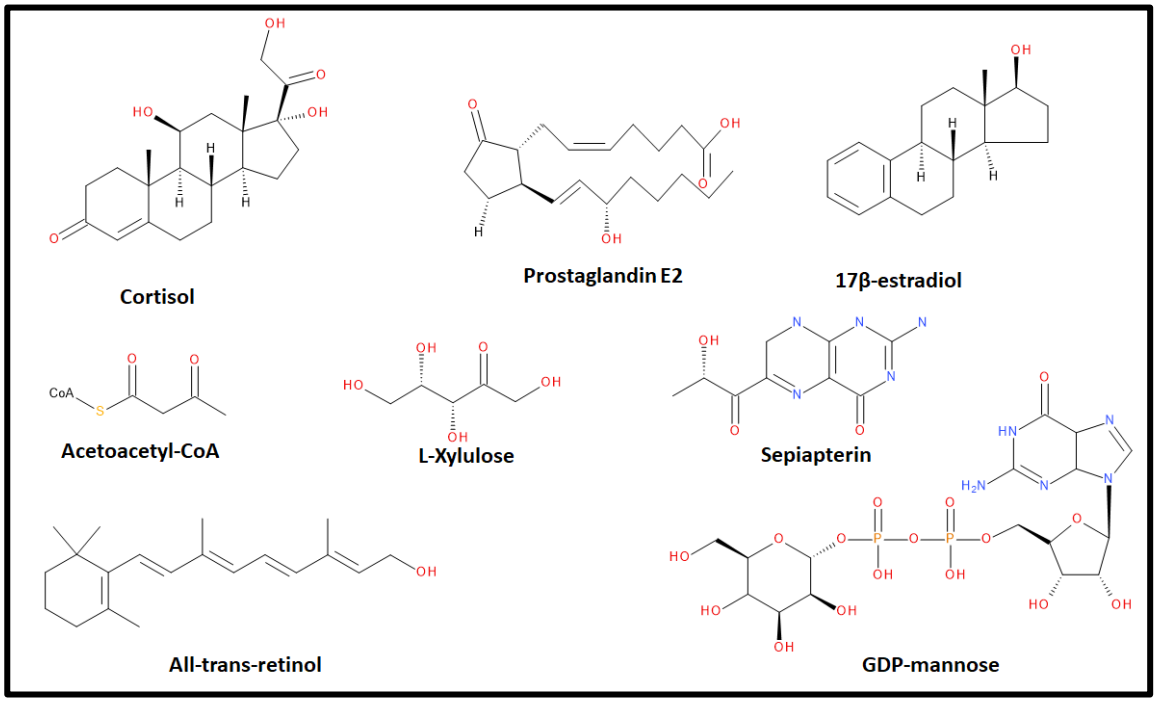


Figure 1.4.1 Diveristy in SDR substrate accomodation - a few examples of types of substrates such as, amino acid derivatives, eicosanoids, fatty acids, retinoids, steroids and sugars.

1.4.2 STRUCTURAL FEATURES

Originally, structures of prokaryotic ribitol dehydrogenase (Dothie et al. 1985) and *Drosophila melanogaster* alcohol dehydrogenase (Schwartz and Jornvall 1976, Thatcher 1980) were elucidated, and were found to differ from the well-known liver and yeast alcohol dehydrogenases (Section 1.3). One of the first SDR member to be discovered was *Drosophila* alcohol dehydrogenase, leading to assignment as 'insect', or 'short-chain' alcohol dehydrogenases (Villarroya et al. 1989, Persson et al. 1991b, Shaw et al. 1992). The discovery of this dehydrogenase was the first step towards the distinction of the SDR superfamily.

The SDR family members are around 250-350 amino acid residues in chain-length and usually contain one domain (Schwartz and Jornvall 1976, Jornvall et al. 1981, Jornvall et al. 1984b, Persson et al. 1991b). Unlike many other protein families, the SDRs exhibit a low sequence identity between members (15-30 % in pairwise alignments), with few but distinct sequence motifs. These motifs are largely conserved, but there is no individual residue strictly conserved.

Despite the low sequence identity, the protein structures reveal a high degree of similarity in the three-dimensional (3D) fold and a central antiparallel 7-stranded sheet surrounded by alpha-helices (a three-layer sandwich of α-β-α; Figure 1.1). This is a variation of the typical six strand NAD(P)(H) Rossmann-fold. Remarkably, the preservation of this 3D fold with conserved cofactor binding properties distinguishes the SDR superfamily from other Rossmann protein families.

SDRs often exhibit N- or C-terminal transmembrane (TM) regions or signal peptides that guides subcellular localisation. For example, two corticosteroid isoenzymes HSD11β1 and HSD11β2 exert dehydrogenase (11β-HSD1 and 11 β -HSD2, cortisol to cortisone) and reductase (11β-HSD1, cortisone to cortisol). Both enzymes localise to the endoplasmic reticulum (ER) membrane due to the N-terminus sequence anchors signals. In the cell, HSD11β1 the N-terminus is cytoplasmic, and the catalytic C terminus protrudes into the ER lumen. This orientation causes HSD11β1 to mainly function as a reductase producing active cortisol. However, in HSD11β2 the N-terminus projects into the ER, and the catalytic domain is facing the cytoplasm. *In vivo*, HSD11β2 principally catalyses dehydrogenation of cortisol. The N-terminal transmembrane segments of 11β -HSD enzymes are required for proper function and orientation in the ER membrane (Odermatt et al. 1999).

The structurally variable C-terminal region allows for different substrate specificities. For example, human SDRs can perform oxidoreductase based reactions including enoyl CoA-reduction (Figure 1.4.2.1A), carbon-alcohol oxidation (Figure 1.4.2.1.B) isomerisation (Figure 1.4.2.1C), epimerisation (Figure 1.4.2.1D), decarboxylation (Figure 1.4.2.1E) and dehydration (Figure 1.4.2.1F) (Mason et al. 1998, Furster 1999, Alphey et al. 2005, Timson 2005, Kavanagh et al. 2008, Pilka et al. 2009, Egger et al. 2011).

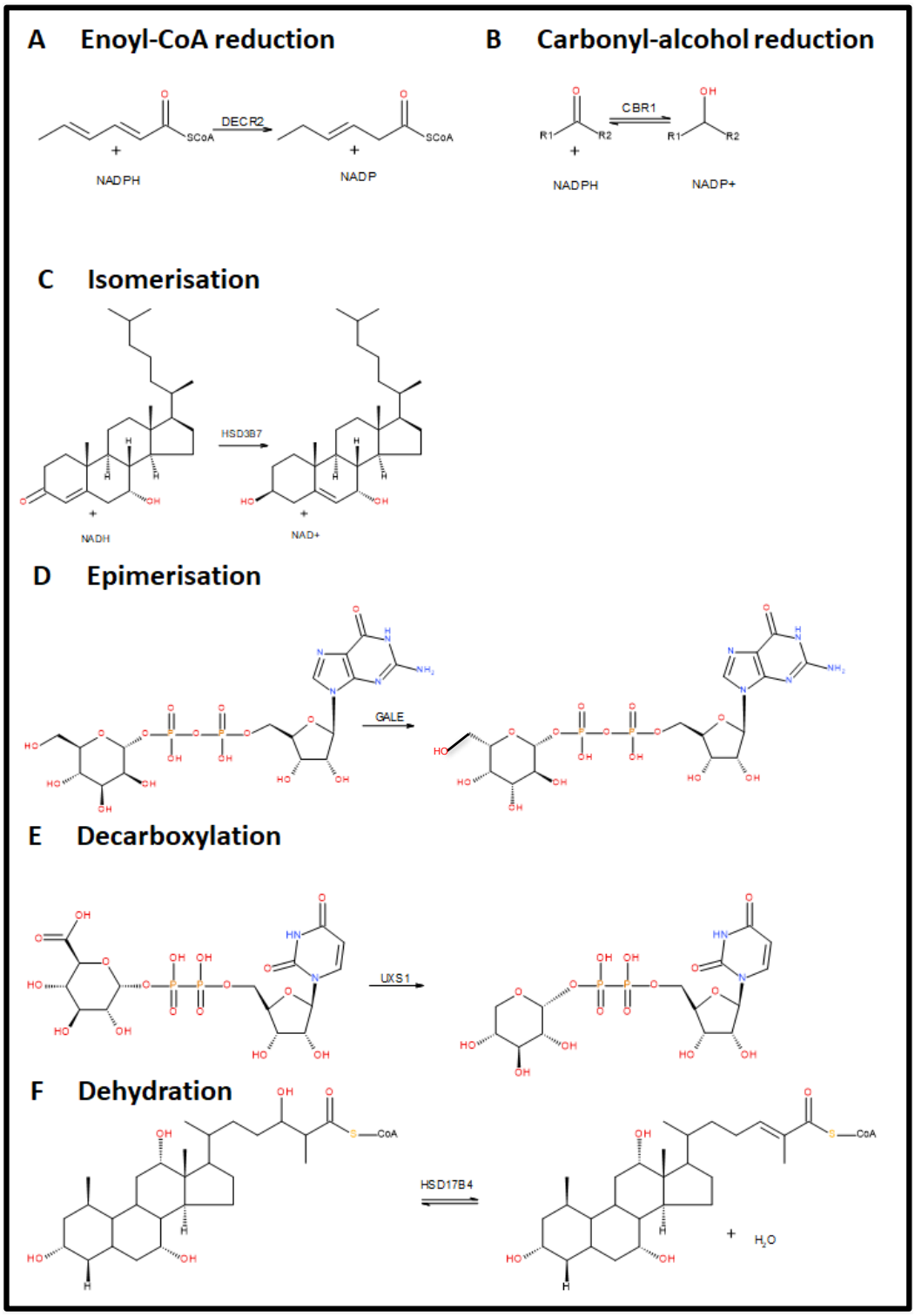


Figure 1.4.2.1 SDR reaction types A) Enoyl-CoA reduction B) Carbonyl-alcohol reduction C) Isomerisation D) Epimerisation E) Decarboxylation F) Dehydration (Adapted from Kavanagh et al. 2008). NB – Isomerisation in C) is two processes i.e isomerisation and reduction.

1.4.3 CONSERVED SEQUENCE MOTIFS AND RESIDIES

The SDR coenzyme NAD(P)(H) binding motif is located in the N-terminal region (Ghosh et al. 1991, Ghosh et al. 1994, Ghosh et al. 1995), where a conserved pattern, Thr-Gly-X-X-X-Gly-X-Gly (located after β1, where X denotes any amino acid) is observed (Rossmann et al. 1974). An acidic residue (Asp or Glu) located approximately 20 residues downstream of the Gly-rich binding motif determines NAD(H) specificity, whereas a basic residue (Arg or Lys) dictates NADP(H) specificity within the Gly-rich sequence and/or in the loop after the second strand (Lesk 1995, Kallberg and Persson 2006). Across many genomes, NAD is the preferred cofactor followed by NADP, and FAD being the least preferred cofactor. In humans, 57 % of members express a preference for NADP/H over NAD/H (44 %) (Bray et al. 2009).

The catalytic motif Tyr-X-X-X-Lys is the most conserved segment typically found between residues 152-156 (numbering adopted from human NAD-linked 15-hydroxyprostaglandin dehydrogenase, HPGD sequence). Multiple sequence alignments and mutagenic studies show Tyr is the most conserved residue and is essential for catalytic activity. Lys156 succeeds Tyr152 as the second most conserved residue throughout the whole family (Jornvall et al. 1981, Krook et al. 1990, Obeid and White 1992, Chen et al. 1993). The active site constitutes a tetrad of Asn111, Ser138, Tyr152 and Lys156 (sequence adopted from HPGD) (Filling et al. 2002).

Another recurring motif Asn-Asn-Ala-Gly is found around positions 90-93 (numbering adopted from human HPGD), and provides additional stabilisation to the central β-Sheet, which are critical to coenzyme binding and the direction of reaction (Filling et al. 2002). In particular, mutations at Thr11 and Asn91 cause the loss of directional activity (Oppermann et al. 1997). Other notable residues for SDR reactions include a single Asp around position 60 that provides stabilisation of the adenine ring pocket, and a Pro-Gly motif around position 180 that is followed by a conserved Thr (at position 188, which provides the reaction direction and hydrogen bonding to carboximide of the nicotinamide ring, respectively (Jornvall et al. 1995, Filling et al. 2001a, Ghosh et al. 2001, Filling et al. 2002). Table 1.4.3.1 shows a SDR motif summary and structural position.

SDR motif	Position	Function
TGxxxGxG	**β1 + α1**	Coenzyme binding region
D	**β3 + α3**	Stabilisation of adenine ring binding of coenzyme
NNAG	**β4**	Stabilisation of central β-sheet
N	**α4**	Connection of substrate binding loop and active site
YSK	**α 5**	Catalytic centre
PG	**β 6**	Reaction direction
T	**β 6**	H-bonding to carboxamide of nicotinamide ring

Table 1.4.3.1 Sequence motifs found in SDR enzymes (Kallberg et al. 2002b) – where 'x' denotes any amino acid.

1.4.4 SDR FOLDS AND TYPES

Based on sequence alignments and structural analysis, the SDR superfamily can be divided into at least two large subfamilies, "classical" and "extended". The "classical" SDRs have a relatively small substrate binding site located between strand 6 and 7 (Figure 1.4.4.1A), whereas the "extended" SDRs have a significantly larger substrate binding site including a two stranded parallel sheet and an α-helix bundle containing 3 helices (Figure 1.4.4.1B). The classical SDR members' exhibit a wide range of oxidoreductase activity towards steroids,

retinoid and carbonyl xenobiotics. Extended members are around 350 long residues, and exhibit activities that are often non-EC1.1 (oxidoreductase) related functions, but clearly include an oxidoreductive component, for example dehydratase, epimerase or isomerase (Kavanagh et al. 2008, Mazurkova et al. 2008). Extended SDRs often have defined roles in carbohydrate metabolism, and display similarities to sugar-metabolising enzymes from bacterial organisms and archaea.

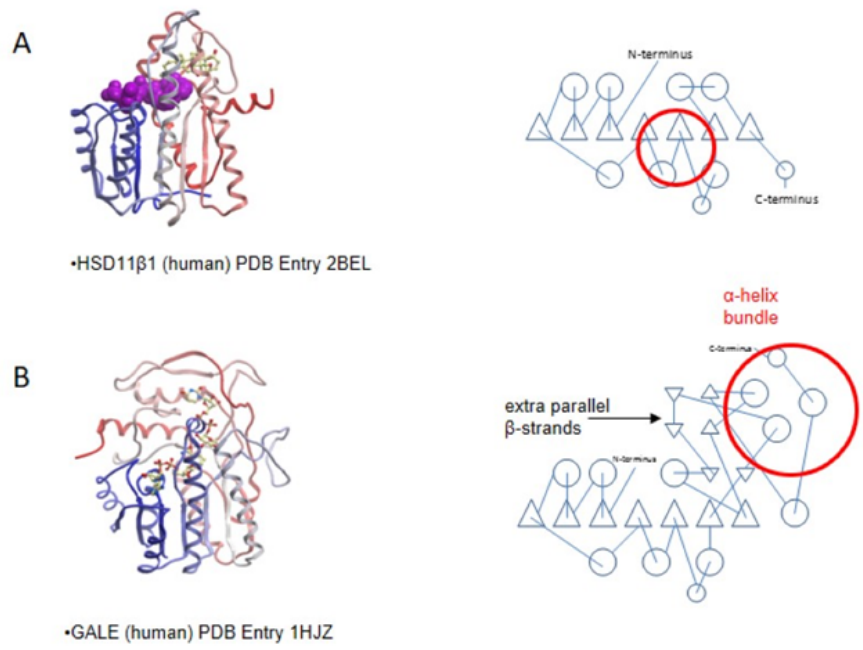


Figure 1.4.4.1 SDR schematic diagrams to highlight the ligand binding regions (red circle) that are significantly different between the two folds. A) 11β-hydroxysteroid-1 (HSD11β1), classical fold bound to NAD/H (ball-and-stick representation); B) UDP-galactose-4-epimerase (GALE), extended fold. The ligand complex contains NAD/H and UDP-acetylglucosamine (Adapted from Bray et al. 2009).

Table 1.4.4.1 highlights differences between two types of SDRs, which shows amino acids of the coenzyme-binding Gly-motif and active site motifs (Kallberg et al. 2002b). In contrast to classical SDRs, the majority of the extended SDRs are NAD/H binding rather than NADP/H binding.

Position	SDR motif	
	Classical	**Extended**
β1 + α1	TGxxxGhG	TGxxGhaG
β3 + α3	Dhx[cp]	DhxD
β4	GxhDhhhNNAGh	[DE]xhhHxAA
α4	hNhxG	hNhhGTxxhhc
α 5	Yx[AS][ST]K	hhhxPYxx[AS]Kxxh[DE]
β 6	h[KR]h[NS]xhxPGxxxT	h[KR]xxNGP

Table 1.4.4.1 Conserved sequence motifs in the classical and the extended SDR families. In the motifs, 'a' denotes an aromatic residue, 'c' a charged residue, 'h' a hydrophobic residue, 'p' a polar residue and 'x' any residue. Alternative amino acids at a motif position are given within brackets (Kallberg et al. 2002b).

In the human genome, the classical SDRs represent the largest proportion of sequences, with 54 members, followed by the extended members with 11 and the atypical with 8 members. Other multicellular organisms such as yeast and *Escherichia coli* also possess the classical type as the predominant form. However, the extended form has a far greater prevalence in smaller unicellular genomes, such as *Mycoplasma genitalium* and *Helicobacter pylori* (Jornvall 1999). The atypical members encompass the smallest group of human SDRs, and display significant sequence diversity, irregular active site motifs or unusual predicted structural features (Figure 1.4.4.2). Some atypical members are part of multifunctional enzyme complexes present in all forms of life, for example, fatty acid synthase (FAS).

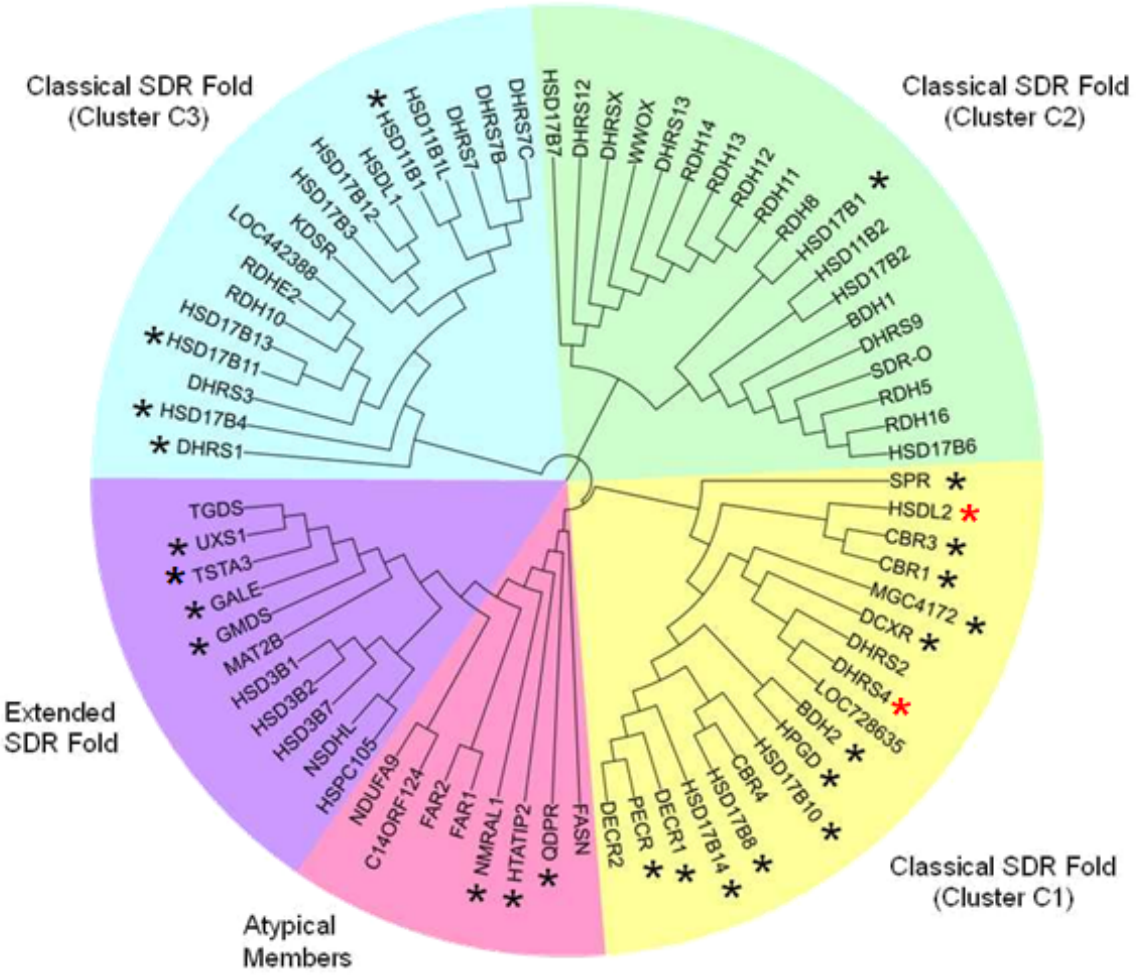


Figure 1.4.4.2 Schematic diagram of the sequence clusters in the human SDR superfamily. There are five main categories annotated as classical clusters C1, C2, C3, extended and atypical. Each superfamily member is annotated using the HUGO gene ID and an asterisk symbol (*) denotes one or more PDB structures is available for the SDR domain. Short branch lengths illustrate that proteins are close homologs, while longer branch lengths indicate a more distant evolutionary relationship (Bray et al. 2009). Red asterisk symbol (*) denotes novel structures discovered in this thesis.

Sequence and predicted secondary structure analyses show human classical SDRs can be subdivided into three main sequence clusters (Figure 1.4.4.2) (Bray et al. 2009). Cluster C1 contains mostly non-membrane bound proteins, and feature a typically 'short' C-terminus after strand 7 (Figure 1.4.4.3A). Cluster C1 accommodates a broader spectrum of substrates compared to cluster C2 and C3 that include CoA, steroids, prostaglandins, and quinone-like molecules.

In general, cluster C2 and C3 are membrane-associated and classically turn over retinoids and steroids. The main feature that makes most of these proteins different from those in C1 is a predicted N-terminal TM helix and a C-terminal helix immediately after β-strand 7. The substrate binding region in cluster C2 and C3 (Figure 1.4.4.3B) is considerably longer than in C1 (Figure 1.4.4.3A). The majority of 3D-SDR structures experimentally solved, lie within the classical SDR cluster C1 because members of the cluster C1 are mostly soluble using heterologous expression systems, in comparison to the other classical SDR clusters, C2 and C3. The N-terminus TM helix in C2 and C3 could attribute to the difficulty in obtaining soluble protein amenable to crystallisation (Figure 1.4.4.4B).

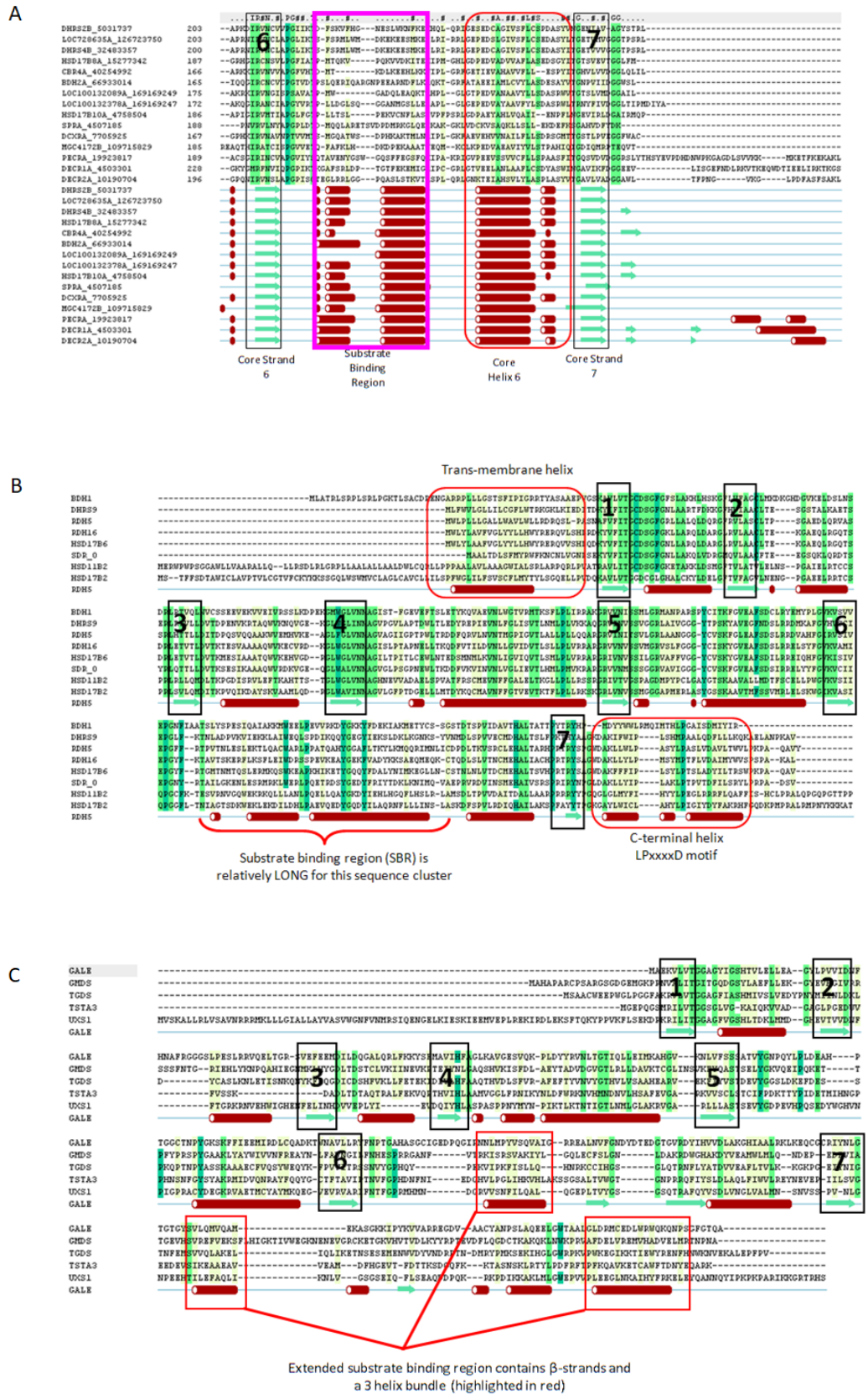


Figure 1.4.4.3 Comparison of C-terminus alignment of SDR slusters - A) classical type I SDR B) Classical type II and III – membrane bound C) Extended Fold – Epimerase family

The vast number of SDR sequences present a problem with similar and multiple names for certain proteins. To overcome this, a nomenclature system based on 48 distinct types of human SDR by Hidden Markov Models has been developed (Persson et al. 2009, Kallberg et al. 2010). A family number is designated to each SDR with the lowest number having the most common human members, and prioritising those with other eukaryotic members, followed by bacterial and archaeal members. The family number follows a letter indicating the SDR type (Table 1.4.4.2). After the SDR type, a number follows to represent an SDR individual member, which prevents confusion between various designations, descriptions, and synonyms (Appendix, Table 1). This system provides a quick way to identify a type of SDR and distinguish individual members, for example, SDR4E1 represents an individual SDR of the extended type, this case Tissue specific transplantation antigen P35B (TSTA3).

The division of the human SDR superfamily into 5 clusters by sequence and structure can be further extended into subsets, since proteins within the same cluster can differ greatly due to their evolutionary distances. For example, there are two types of extended fold SDR proteins, epimerases and 3beta-HSD/delta4-5 isomerase proteins. Epimerases act on nucleotide sugar substrates and are not membrane-associated, whereas 3-HSD family enzymes act on steroids and are membrane-associated (Figure 1.4.4.3C)

SDR Type	Single Letter Nomenclature Designation
Classical	C
Extended	E
Intermediate	I
Divergent	D
Complex	X
Atypical	A
Unknown	U

Table 1.4.4.2 SDR Types and their designations (Bray et al. 2009)

1.4.5 CATALYTIC MECHANISM OF SDRs

Along with the Tyr residue, the active site comprises a catalytic tetrad of Tyr, Lys, Asn, and Ser residues (Oppermann et al. 1997, Filling et al. 2001a, Filling et al. 2002, El-Hawari et al. 2009). The majority of SDRs use Tyr151 as their catalytic acid/base centre, whereas Ser138 stabilises the substrate, and Lys155 forms hydrogen bonds with nicotinamide ribose and lowers the pKa of the Tyr-OH to promote proton transfer (sequence adapted from 3β/17β-HSD) (Benach et al. 1996, Fujimoto et al. 2001). Asn111 located within helix α5 interacts with main-chain atoms of non-conserved residues, and the side-chain carbonyl group ligates a water molecule, which produces a characteristic 'kink' in the helix α4 backbone (Filing et al. 2002). The ligated water molecule is then able to hydrogen bond to the active site Lys155 instead of binding to the main-chain as expected from a helical structure. The function of Asn111 is to support active site geometry through stabilisation to Lys155 via a water molecule, which establishes a proton relay mechanism (Filling et al. 2002). An SDR family proton relay system consists of the deprotonation of Tyr151, and hydride transfer from the protein to the NAD(P) via Ser138 and Lys155 (Figure 1.4.5.1).

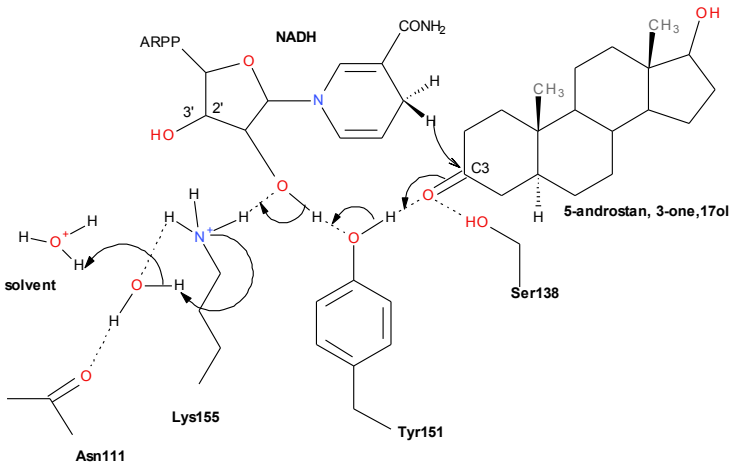


Figure 1.4.5.1 Postulated reductive reaction mechanism of 3β/17β-HSD involving NADH and steroid substrate (5α-androstane, 3-one, 17ol). Catalysis is initiated by proton transfer from Tyr-151 hydroxyl to the substrate carbonyl followed by hydride transfer to C3 of the steroid. A proton relay is formed and involves the 2′ OH of the ribose, the Lys-155 side chain, and a water molecule bound to the backbone carbonyl of Asn-111. *ARPP*, the adenosine ribose pyrophosphate moiety of NADH. Adapted from Filling et al. 2002.

Not all SDRs possess the conserved active site fingerprint Y-x-x-x-K, such as, hydroxysteroid dehydrogenase like 1 (HSDL1), 2,4-dienoyl-CoA reductase, mitochondrial (DECR1), 2,4-dienoyl-CoA reductase, peroxisomal (DECR2), NmrA-like family domain-containing protein 1 (NMRAL1) and C14ORF124 in humans, which highlights the sequence variation in the active site. In DECR1, Ser210 replaces the conserved tyrosine, and an alternative tyrosine that is instead located at position 199 on helix four (Alphey et al. 2005). Figure 1.4.5.2 shows an overlay of DECR1 and 17β-hydroxysteroid dehydrogenase 8 (HSD17β8) structures (an SDR with the conserved Y-x-x-x-K motif) where Tyr199 and Ser210 occupy positions that differ from other SDRs. Ser210 occupies the position of DECR1 corresponding to catalytic Tyr152 of HSD17β8 and Thr197 replaces the catalytic Ser139. It is a serine that acts as the catalytic base in DECR1, and the tyrosine appears to bind the substrate, and to keep the three-dimensional geometry with other residues to stabilise the transition state (Alphey et al. 2005).

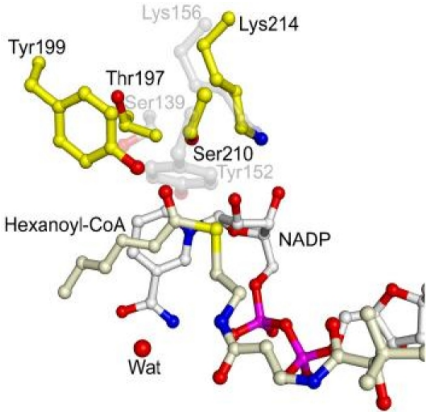


Figure 1.4.5.2 Active site features of the 'divergent' SDR dienoyl-CoA reductase (1w6u). Active site residues are shown with yellow carbons and labelled, while the active site residues of the classical SDR 3α/20β-HSD are superimposed in grey, and shown semi-transparent for comparison. A water molecule that is accessible to the bulk solvent and is proposed to be involved in the reaction mechanism is shown as a red sphere This figure is taken from (Kavanagh et al. 2008).

Similarly in DECR2, Alanine substitutes the catalytic tyrosine, which is located towards the begining of helix five at position 149. Alanine, in this case is not a proton donor and cannot serve as the catalytic base in SDRs. Interestingly, it is an aspartate on helix four that acts as the acid/case catalyst (Hua et al. 2012). A Histidine replaces tyrosine in NMRAL1; however, there is no nearby tyrosine to suggest any catalytic activity and instead it reportedly exhibits a regulatory role with arginosuccinate synthase (ASS1) (Stammers et al. 2001, Zhao et al. 2008, Dai et al. 2009). Table 1.4.5.1 shows that HSDL1 has a phenylalanine in place of the catalytic Tyr151 and at present no activity has been observed. HSDL1 exhibits a weak reductase activity with estrone and androsterone when the catalytic sequence Tyr-x-x-x-Lys is restored (Meier et al. 2009).

SDR HUGO ID	Active site motif	Alternative to YxxxK
DECR1	SxxSK	S110 & T197
DECR2	AxxxK	D137
SDR39U1	YxxxS	Unknown
HSDL1	FxxxK	None
NMRAL1	HxxxK	None

Table 1.4.5.1 SDR active site motif alternatives to YxxxK

1.5 EVOLUTION AND FAMILY COMPARISON

The origin of new functions is central to an understanding of evolutionary diversification. SDRs are thought to arise from gene fusion of a common ancestral coenzyme domain with various substrate-specific domains (Jornvall et al. 1999, Jornvall et al. 2010). Recent studies of *Drosophila* SDR structures and catalytic mechanisms suggest that the evolution of novel functionalities can be achieved by incorporating new amino acids within an active site, while retaining those essential to catalyse a partial reaction; or by changing the amino acids responsible for substrate binding, while retaining those essential for the complete reaction (Benach et al. 2001, Zhang et al. 2004, Zhang et al. 2010).

1.5.1 GENOME COMPARISONS

Genomic studies of human ortholog SDRs show a broad conservation across species in the classical type (Kallberg et al. 2002a). In smaller genomes, such as archaea and algae, extended SDR forms are more prominent, (Persson et al. 1999). Extended SDRs display a higher proportion of conserved residues (24–46 %), in contrast to the classical SDRs with 15-30 % conservation. This implies that extended SDR enzyme activities are similar across different organisms and represents an early form of metabolic processing.

Interestingly, a SDR member that is a tumour suppressor, WWOX (WW domain containing oxidoreductase) (Yang and Zhang 2008) has a high occurrence of homologs in humans, fruit fly, worms and plant, implying this may be an early member of the classical form. Notably there are specific mammalian SDRs that have no gene homologs in smaller genomes.

Mammalian-specific SDRs are mainly steroid or retinol metabolising enzymes (e.g. DHRS9, SDR-O, RDH5, RDH16, HSD17β6 and HSD17β13), which reflect the development of regulation and metabolic conversions in mammals.

1.5.2 FAMILY COMPARISONS

The SDR family has many closely related cousins with similar activities. Through sequence analysis distinct features for each family have been established (Persson et al. 1991b, Jornvall et al. 1995, Nordling et al. 2002a). The SDR and MDR families appear to have a similar architecture comprising a Rossmann fold and cofactor binding site. MDRs possess two domain subunits around 350-375 residues (Figure 1.5.2.1B), and frequently associate with a catalytic zinc atom liganded by two Cys and one His residue (Persson et al. 2008). Long-chain dehydrogenases (Figure 1.5.2.1C) are of similar domain architecture as MDRs with the active-site located in the cleft between the two domains. However, they are usually between 350 and 560 amino acids long and, frequently utilise a Lys-based catalytic centre (Persson et al. 1991b, Kavanagh et al. 2002). FAS (fatty acid synthase) is a multidomain enzyme that catalyses the formation of long-chain fatty acids, in particular palmitate from acetyl- and malonyl-CoA. FAS contains a β-ketoacyl reductase SDR domain and an enoyl reductase MDR domain demonstrating that the two superfamilies can combine for function (Fisher et al. 2000, Airenne et al. 2003).

AKRs are around 320 amino acids in length and use a wide range of substrates like the SDRs, including steroids and monosaccharides. AKRs are monomeric proteins that bind NAD(P)(H) without a Rossmann fold. Structurally the AKR belong to the $(\alpha/\beta)_8$ barrel or triophosphate

isomerase (TIM) barrel protein family (Figure 1.5.2.1E). AKRs display an example of convergent evolution with an active site conformation nearly superimposable to that of SDRs with conserved Tyr and Lys residues (Figure 1.5.2.2) (Jornvall et al. 1995, Oppermann and Maser 2000).

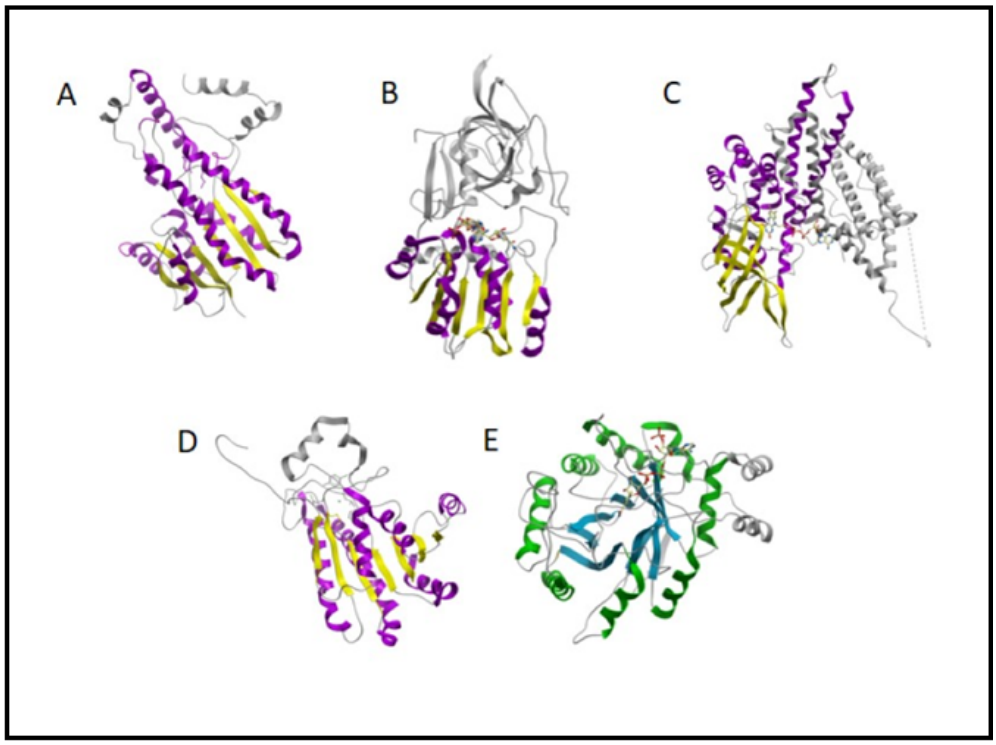


Figure 1.5.2.1 Ribbon diagram comparison of classical SDR, MDR, LDR, AKR and halohydrin halogenase enzymes. The Rossmann-fold motif is depicted with beta strands in yellow and helices in purple; additional domains and secondary structural elements are shown in grey. The TIM-barrel is depicted with beta strands in blue and helices in green. The nucleotide cofactor is drawn in ball-and-stick representation. A) Classical SDR (human HSD17β1; PDB 1BHS), *B*) MDR (human ZADH1; PDB 2VNA), C) LDR (human ACADVL; PDB 2UXW), D) Halohydrin dehalogenase (bacterial; PDB 1PWX), E) AKR (human AKR1C1; PDB 2HDJ). Adapted from Kavanagh et al. 2008).

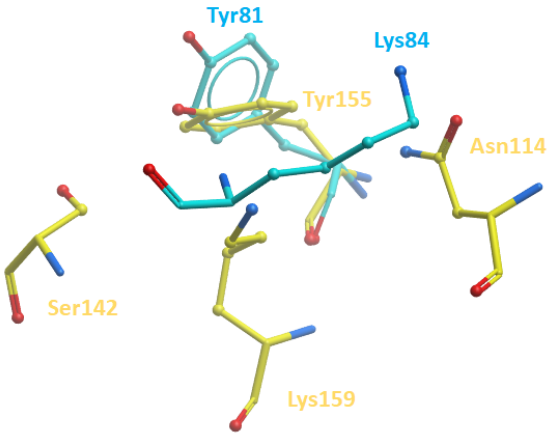


Figure 1.5.2.2 Superimposed Tyr and Lys active site of classical SDR and AKR enzymes. SDR catalytic tetrahad coordinates from human HSD17β1; Asn114, Ser142, Lys159 and Tyr155 (PDB 1BHS - yellow), AKR active site coordinates of Tyr81 and Lys84 of human AKR1C1 (PDB 2HDJ - cyan) Apdated from (Oppermann and Maser 2000).

Halohydrin dehalogenases (Figure1.5.2.1D) are related proteins to SDRs that exhibit low sequence idenitity to each other. A halide-binding site substitutes the SDR cofactor binding motif that facilitates haloalcohol dehalogenation. Site directed mutagenesis studies show that SDRs share a similar catalytic mechanism utilised by halohydrin dehalogenases, where Tyr acts as the catalytic base, and Ser stabilises the substrate. The catalytic Lys is replaced with Arg and also lowers the pK_a of Tyr (Smilda et al. 2001, Spelberg et al. 2001, van Hylckama Vlieg et al. 2001, Kavanagh et al. 2008).

HIV-1 Tat interactive protein 2 (HTATIP2) (also known as, CC3/TIP30), originally thought to be a kinase based on a N-terminal adenosine triphosphate (ATP)-like binding motif (Xiao et al. 1998, Xiao et al. 2000), however was identified to be a SDR with tumour suppressor properties through bioinformatics analysis and structure determination (El Omari et al. 2005).

The tumour suppressor effect of HTATIP2 is implied from the inhibition of nuclear transport by binding to β-importins or regulating transcription through interaction in a complex with co-activator independent of AF-2 function (CIA) and the c-myc gene (King and Shtivelman 2004, Chen and Shtivelman 2010). Mutagenesis work on its nucleotide binding motif suggests that NADPH binding is essential for the biological activity of HTATIP2, including interaction with importins and the CIA/c-myc system (Jiang et al. 2004, King and Shtivelman 2004).

As previously mentioned, NMRAL1 lacks enzyme activity due to the absence of the conserved active-site tyrosine. In the current model, NMRAL1 functions as a NADPH sensor in response to changes in the NADPH/NADP ratio (Zhao et al. 2008, Dai et al. 2009). NMRAL1 interacts with ASS1 to controls its activity in the nitric oxide (NO) synthesis (Zhang et al. 2002). Both HTATIP2 and NMRAL1 provide examples of the adaptation of a metabolic enzyme fold to incorporate a regulatory role. Despite a highly analogous three dimensional architecture, all these proteins display almost unrecognisable sequence homologies (Figure 1.5.2.1).

1.6 MEDICAL RELEVANCE

The functional roles of human SDRs shown by their enzymology and evolution imply them in many disease-associated pathways. Human SDRs represent an enzyme class, which has generated extensive biotechnological and pharmaceutical interest for medicinal targets (Gileadi et al. 2007, Wu et al. 2007, Prehn et al. 2009). Of the 75 human SDR genes listed in

the Swiss-Prot database, 23 SDRs have a phenotypic description associated with a genetic disease in the OMIM database (Online Mendelian Inheritance in Man). A further 37 are listed in the genetic association database (GAD; Becker et al. 2004). Thus, many (or even all) of these enzymes are medically important, see Table 1.6.1 for associated diseases in humans.

An area of pharmaceutical interest is the dysfunction of sex steroid hormones (e.g. androgens and estrogens), which can contribute to the pathogenesis of various severe diseases in humans, for instance, cancer, hypertension, neoplasia, and renal disease. Many steroid SDRs have well-characterised 3D structures that reveal key positions as a basis for selective inhibition. For example, the dehydrogenase activity by hydroxysteroid 11β dehydrogenase 2 (HSD11β2) producing inactive cortisone is necessary to protect mineralocorticoid receptors to regulate sodium homeostasis (Funder et al. 1988). Patients with HSD11β2 abnormalities lose this protection from cortisol, causing mineralocorticoid excess leading to sodium retention, hypokalemia, and hypertension (Atanasov et al. 2007, Bailey et al. 2008). Thus, drug development efforts continue to pursue specific hydroxysteroid 11β dehydrogenase 1(HSD11β1) inhibitors towards its reduction control the production of active cortisol (Barf et al. 2002, Hult et al. 2006, Oppermann 2006, Wan et al. 2011, Blum et al. 2012).

Another group of enzymes of considerable pharmaceutical interest are 17β-HSDs which catalyse the conversion between inactive 17-keto steroids and the 17β-hydroxy steroids.

They are, therefore, potential drug targets in order to lower the local sex steroid concentrations in conditions such as endometriosis, breast cancer and non-small cancer lung carcinomas (Develoux et al. 2014 and He et al. 2016). Several of these enzymes are promising drug targets, and inhibitors are under development for HSD17β1, HSD17β2, HSD17β3, HSD17β7, HSD17β10 and HSD17β12 (Bagi et al. 2008, Chen et al. 2008, Day et al. 2009, Day et al. 2010, Fournier and Poirier 2009, Laplante et al. 2009, Lin et al. 2013, Starčević et al. 2011).

Some microbial SDRs have also been identified as anti-microbial drug targets owing to their role in growth factor synthesis, for example, pteridine reductase or enoyl-CoA reductases (Gourley et al. 2001, Goodman and McFadden 2007). This has allowed the development of novel antibiotics, herbicides and pesticides, demonstrating the versatility in biological applications of this family. The functions of about half of the human SDR enzymes are unknown, and potential candidates for many human diseases may be identified in the future (Bagi et al. 2008, Chen et al. 2008, Day et al. 2008, Day et al. 2009, Fournier and Poirier 2009, Laplante et al. 2009).

Hugo Gene ID	Full Name	Associated disease	Source	OMIM ID
DCXR	Dicarbonyl/L-xylulose reductase	Pentosuria	(Nakagawa et al. 2002)	608347
DECR1	2,4-Dienoyl-CoA reductase	Reductase Deficiency	(Roe et al. 1990)	222745
FAR1	Fatty acyl CoA reductase 1	Peroxisomal fatty acyl-CoA reductase 1 disorder	(Buchert et al. 2014)	616107
GALE	UDP-Galactose-4-epimerase	Galactose epimerase deficiency	(Alano et al. 1998)	230350
HPGD	15-hydroxyprostaglandin dehydrogenase	Cranioosteoarthropathy, digital clubbing, isolated congenital, Hypertrophic osteoarthropathy	(Latos-Bielenska et al. 2007, Uppal et al. 2008, Tariq et al. 2009)	601688
HSD3β2	3β-hydroxysteroid dehydrogenase 2	3-beta-hydroxysteroid dehydrogenase, type II, deficiency	(Rheaume et al. 1992)	613890
HSD3β7	3β-hydroxy-δ5-steroid dehydrogenase	Congenital defect in bile acid synthesis and neonatal cholestasis	(Schwarz et al. 2000, Cheng at al.2003)	607764
HSD11β1	11β-hydroxysteroid dehydrogenase 1	Cortisone reductases deficiency	(Jamieson et al. 1999)	600713
HSD11β2	11β-hydroxysteroid dehydrogenase 2	Mineralocorticoid excess	(Albiston et al. 1994)	614232
HSD17β3	17β-hydroxysteroid dehydrogenase 3	Male Pseudohermaphroditism with gynecomastia	(Geissler et al. 1994)	605573
HSD17β4	17β-hydroxysteroid dehydrogenase 4	D-bifunctional protein deficiency, Perrault syndrome	(Suzuki et al. 1997)	261515
HSD17β10	Hydroxyacyl-CoA dehydrogenase 2	17-beta-hydroxysteroid dehydrogenase X deficiency, X-linked syndromic 10 mental retardation	(Reyniers et al. 1999, Ofman et al. 2003)	300256
KDSR	Follicular lymphoma variant translocation 1	Lymphoma/leukaemia, B-cell, variant	(Rimokh et al. 1993)	136440
NDUFA9	Nadh-ubiquinone oxidoreductase 1 alpha subcomplex, 9	Leigh syndrome due to mitochondrial complex I deficiency	(van den Bosch et al. 2012)	603834
NSDHL	NAD(P)H steroid dehydrogenase-like protein	CHILD syndrome, CK syndrome	(Konig et al. 2000)	607764
QDPR	Quinoid dihydropteridine reductase	BH4-deficient Hyperphenylalaninemia	(Howells et al. 1990)	612276
RDH5	Retinol dehydrogenase 5	Fundus albipunctatus	(Yamamoto et al. 1999)	61213
RDH11	Retinol dehydrogenase 11	Retinal dystrophy, juvenile cataracts, and short stature syndrome	(Xie et al. 2014)	607849
RDH12	Retinol dehydrogenase 12	Leber congenital amaurosis 13	(Haeseleer et al. 2002)	608830
SDR9C7	Short-chain dehydrogenase/reductase family 9C member 7	Ichthyosis, congenital, autosomal recessive 13	(Shigehara et al. 2016)	609769
SPR	Sepiapterin reductase	dopa-responsive dystonia associated with sepiapterin deficiency	(Bonafe et al. 2001)	612716
TGDS	TDP-glucose 4,6-dehydratase	Catel-Manzke syndrome	(Ehmke et al. 2014)	616146
WWOX	WW domain containing oxidoreductase	Epileptic encephalopathy, early infantile, Esophageal squamous cell carcinoma, Spinocrebellar ataxia, autosomal recessive 12	(Abdel-Salam et al. 2014, Kuroki et al. 2002, Mallaret et al. 2015)	605131

Table 1.6.1 Human SDRs and their associated genetic diseases in the OMIM database

1.7 AIMS OF THESIS

Many SDRs are associated with diseases, and represent a druggable enzyme class, which has generated a great deal of pharmaceutical interest. However, structural data are unknown for over half of the human SDR enzymes, and many have unknown substrates or specific inhibitors. Phylogenetic analysis based on characterized homologs can provide indicators for cellular functions of uncharacterised SDRs. However, in reality the closest 'characterised' homologs are distantly related so that they become unsuitable models. As a starting point, this thesis will ascertain an indepth annotation of human SDRs between cofactor preference, substrate accommodation and subcellular localisation of human SDR family to aid functional predictions of uncharacterised SDRs. This study also exploits a family-wide expression technique to purify human SDRs that represent all clusters and encompass a variety of metabolic function for downstream processing (differential scanning fluorimetry (DSF), crystallography and activity screening). Finally, this study is based on the premise that thermal shift data from DSF can be used to widely explore the protein-ligand interactions that occur across the SDR superfamily.

The study will focus on the areas below:

- Develop a large-scale parallel approach to heterologously express human SDRs in sufficient quantities for downstream processing.
- Validation of a DSF data in comparison to literature and bioinformatics predictions e.g. cofactor preference and ligand binding data

- Co-crystallisation experiments with compounds that show an increase in T_m (midpoint of transition) to identify small-molecule binders and elucidate unknown (novel) SDR structures

- Activity assays to identify functional properties of orphan SDRs. This would help in understanding the selectivity and specificity of individual SDR enzymes.

1.7.1 STRUCTURE OF THESIS

The next chapter (Chapter 2) provides the detail, and the reasons why specific methods were chosen for expression and purification strategies required in downstream processing, and a summary of all the human SDRs successfully purified. An annotation of the human SDR superfamily (Chapter 3) to detail what is currently known in terms of substrates and subcellular localisation follows. The main results chapters are divided into cofactor determination (Chapter 4) and ligand screening (Chapter 5). Both provide validation of the DSF method by using established compounds known to bind with SDRs (e.g. cofactors and inhibitors) and identification of new ligands that can be applied in co-crystallisation studies and enzyme kinetics (Chapter 6).

CHAPTER 2: MATERIALS AND METHODS

2.1 INTRODUCTION

The first step in this study of the human SDR family was the production of pure proteins on a large scale, critical for downstream processing by DSF and crystallisation studies (Figure 2.1.1). Traditional purification procedures are developed for individual proteins, however for large-scale studies this can often be time consuming. The exploitation of high-efficiency expression systems and tags has allowed for the development of standard family-based purification protocols. The developed procedures for the protein purification consist of an anion exchange step followed by a size-exclusion step. Affinity purification was chosen as the preferred method due to high selectivity, fast purification and the possibility for high salt concentration. A systematic approach to protein production has been established at the SGC (Oxford), by using small to large-scale expression, purification techniques, DSF and a selection of crystallisation methods, to aid the understanding of protein families (Figure 2.1.1) (Savitksy et al. 2010). A library of some human SDRs existed prior to start of this project at the SGC, and other SDRs were cloned, expressed and purified when needed.

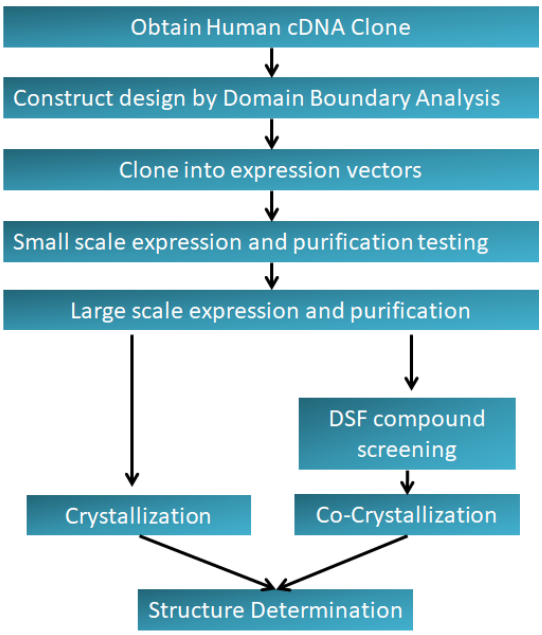


Figure 2.1.1 SGC core workflow processes – (Gileadi et al. 2007)

2.2 DELETION CONSTRUCTS

SDR members were identified using sequence profile searches of human sequences performed with PSI-BLAST (Position-Specific Iterated - basic local alignment search tool) using profiles from the Pfam domain family database (http://pfam.sanger.ac.uk). Gene symbol assigned to each human SDR sequence was from the HUGO Gene Nomenclature Committee (HGNC, www.genenames.org). Amino acid sequences for all the known human SDRs were collated, entered into the PSIPRED (PSI-BLAST based secondary structure PREDiction) server (McGuffin et al. 2000) and the resulting secondary structure predictions analysed. Sequence selection and domain boundary analysis were on bioinformatics predictions of structural conserved domains (PFAM), sequence alignments with proteins in the PDB (Genthreader), prediction of secondary structures and disordered regions (PsiPred) that then guided the design of all the SDR deletion inserts. Domain boundary analysis of

each human SDR allowed the identification of the regions for multiple deletion inserts. The rationale is that truncating protein into a well-ordered state may improve one or more of the deletion inserts to express soluble, stable protein (Structural Genomics et al. 2008). Of additional importance regarding the design of deletion inserts is the removal of flexible or interdomain regions, which may inhibit crystallisation of the protein. Per target, 5-7 different inserts were typically generated (total: 498 constructs for 63 human SDRs).

2.*3 ESCHERICHIA COLI* EXPRESSION

2.3.1 CLONING

A "deletion insert" may be used in several different cloning vectors, which differ in the tags or host expression organism (bacteria or baculovirus) (Strain-Damerell et al. 2014). The full-length cDNAs for 68 human SDR genes were obtained from natural, or synthetic genes (Medigene, Sigma). Corresponding DNA fragments were produced from full-length cDNAs by PCR, and subcloned by ligation into pET derived vectors that incorporate ligase independent cloning (LIC) compatible sites (Appendix, Figures 1-5) (Aslanidis and de Jong 1990). Resulting expression constructs were all sequence-verified. All bacterial vectors contain either an ampicillin or kanamycin resistance gene, a T7-promoter, LacI repressor for regulation of induction, a hexahistidine tag on the C-or N-terminus, followed by TEV protease cleavage site. Construct generation was kindly provided by SGC-Biotechnology (https://www.thesgc.org/biotech), and Metabolic Enzymes and Complexes groups (https://www.thesgc.org/science/mob).

2.3.2 TRANSFORMATION

The resulting expression vectors with inserts were transformed into chemically competent BL21DE3-(pRARE2)-R3 (phage resistant) cells using the heat shock method (Sambrook 1982). All constructs were transformed into the *E. coli* BL21*DE3* strain. The mixture of transformed cells is incubated with shaking at 37 37 ºC for an hour, and then spread onto an agar plate containing 50 µg/mL of the appropriate antibiotic for selection (kanamycin or ampicillin) so that only transformed cells are able to grow. Colonies from the plate were picked, and grown overnight at 37 ºC with shaking at 500-600 rpm. Addition of 60 % glycerol and flash freezing with liquid nitrogen, preserved the cells and stored at -80 °C until further use.

2.3.3 EXPRESSION

Glycerol stocks were expressed in 50 mL to provide a rapid method of screening for soluble proteins for subsequent large-scale production. Start-up cultures were grown in a 1.5 mL deep well block containing 50 µg/mL of the appropriate antibiotic (ampicillin/kanamycin) and 34 µg/mL chloramphenicol in 50 mL terrific broth (TB) overnight at 37 °C with rigorous shaking at 200 rpm. The overnight cultures were inoculated into TB with 50 µg/mL of the corresponding antibiotics. Cells were grown at 37°C and 160 rpm, to an optical density of 1.5 (OD_{600}) before decreasing the temperature to 18^{o}C and adding 0.2 mM IPTG (Isopropyl β-D-1-thiogalatopyranoside) to induce protein expression. Large-scale expression was carried out in 6-12 L for downstream purification.

2.4 BACULOVIRUS PRODUCTION AND EXPRESSION

2.4.1 INTRODUCTION

In some cases, preparation of the recombinant proteins requires a baculoviral expression system due to insoluble expression in *E. coli*. In this system, the gene of interest (or deletion insert) is cloned into a pFastBac donor plasmid (pFB-Lic-Bse, Figure 6 - Appendix) (Savitsky et al. 2010) through LIC cloning. The resulting plasmid has the gene of interest under the influence of a polyhedron promoter flanked by two Tn7 recombination sites. The generation of recombinant baculovirus is based on the transposition of the donor plasmid into a baculovirus shuttle vector (bacmid). Bacmid propagation is achieved by transformation into DH10Bac^{TM} competent cells (Invitrogen), which harbours the viral DNA with a lacZ complementation factor (Figure 2.4.1). Transposition of the insert from the pFAstBac donor plasmid into bacmid interrupts the lacZ operon causing recombinant plasmids to appear white and empty plasmids as blue on selective DH10Bac agar plates containing 50 μg/mL kanamycin, 10 μg/mL tetracycline and 1 μg/mL gentamycin.

Bacmids were isolated using Millipore Montage Plasmid Miniprep HTS 96 Kit (LSKP09624) and used to transfect Sf9 cells using Cellfectin (Invitrogen), thus generating a virus stock (P0 virus). Trypan blue Stain (Invitrogen) and a hemocytometer were used to determine a viable Sf9 cell count. The P0 virus stock was amplified in two successive rounds at a 0.1 molarity of infection (MOI) to produce a higher titre virus suitable for protein expression (P2 virus). For large-scale expression Sf9 cells seeded at density of 2 x 10^6/mL were infected with 1 mL P2 virus stock/500 mL of cell cμture. After 72 hours post-infection the cultures were harvested by centrifugation for 20 min at 2000 rpm (Beckman JLA 8.1).

Cell pellets from each culture flask were washed with 15 mL of ice cold PBS and recentrifuged. Pellets were stored at -80 °C until required. Standard baculovirus protocols were carried out developed for medium-throughout baculovirus expression (Mahajan et al. 2014).

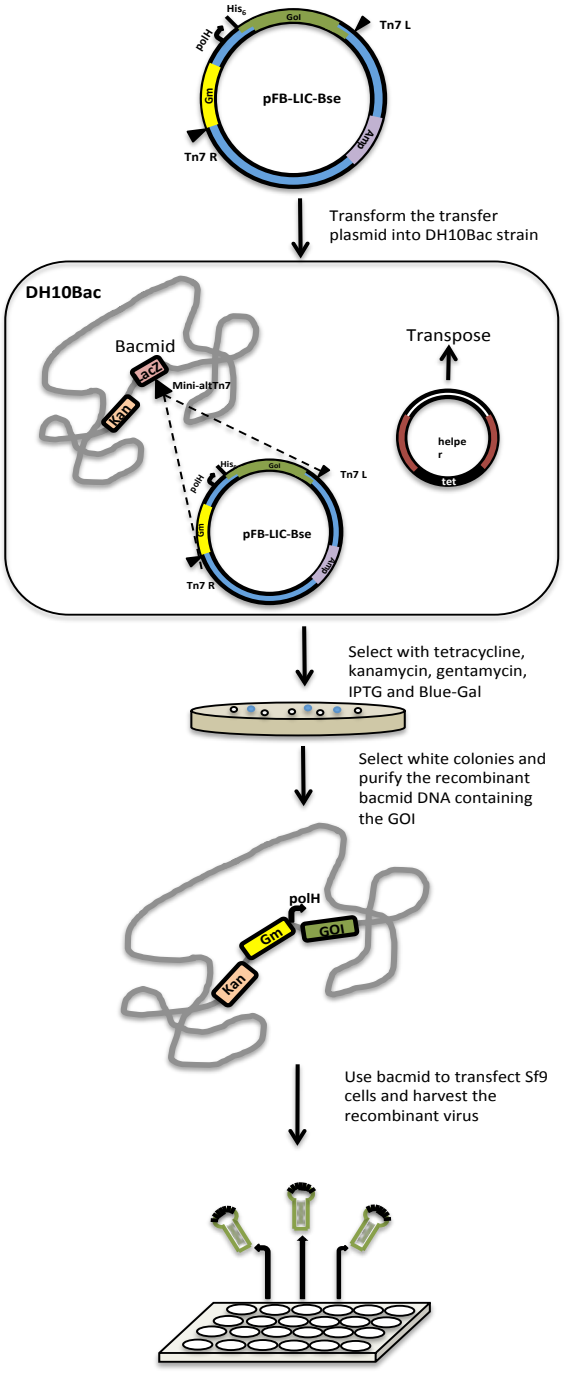


Figure 2.4.1 Transposition process of the donor plasmid into bacmid - The donor plasmid is transformed into the DH10Bac™ *E. coli* strain, which contains both bacmid DNA and a helper plasmid The transposase, expressed from the helper plasmid, will facilitate the transfer of the transposable element including the gene of interest into the bacmid. The recombinant bacmid DNA can then be purified and used directly to transfect Sf9 insect cells (Mahajan et al. 2014).

2.5 PURIFICATION

2.5.1 PURIFICATION BUFFERS

Lysis buffer: 50 mM Hepes, pH 7.5; 500 mM NaCl; 5 % glycerol; 10 mM imidazole**,** 0.1 mM TCEP , protease inhibitors, 5 U/mL Benzonase

Binding buffer: 50 mM Hepes, pH 7.5; 500 mM NaCl; 5 % glycerol; 10 mM imidazole**,** 0.1 mM TCEP

Wash buffer: 50 mM Hepes, pH 7.5; 500 mM NaCl; 5 % glycerol; 20 mM imidazole**,** 0.1 mM TCEP

Elution buffer: 50 mM HEPES, pH 7.5; 500 mM NaCl; 5 % glycerol; 250 mM imidazole, 0.1 mM TCEP

SEC: 10 mM Hepes, pH 7.5; 500 mM NaCl; 0.5 mM TCEP

2.5.2 TEST PUFRICATION FROM *E.COLI*

After 12 hours of induction at 160 rpm, the cultures were harvested by centrifugation (Beckman Avanti, JLA 8.1, 4000 rpm for 10 min at 4 °C). The supernatant was decanted, and the pellets were resuspended in 1.5 mL lysis buffer. Sonication was used to break the cells. A fraction of the lysed cells representing the total expression fraction was kept for SDS-PAGE analysis. The gels were stained using a Coomassie-based stain (InstantBlue, Expedeon). The resulting lysate was centrifuged to pellet the insoluble protein and cell debris at 10,000 rpm,

at 4 ^{o}C for 20 min. The supernatant was applied to a Ni-NTA resin (300 µL of 50 % suspension Ni-NTA equilibrated in 5 mL of binding buffer) in 96-well deep well filter block, then washed with 5 mL binding buffer, followed by 1 mL wash buffer. The protein was then eluted into 1 mL fractions using elution buffer and then 1000 rpm (Beckman Avanti centrifugation for 1 minute (elution fraction).

A number of indicators from the test expression determine whether a construct is suitable for large-scale (>1 litre) expression and purification: the protein's abundance in the total and soluble fractions, the purity of affinity purification. An example of test expression results in Figure 2.5.2.1 shows the soluble protein for several constructs of NMRAL1 in pNIC28Bsa4. Constructs A1, A2, A4, A5 and A7 showed high expression, and high solubility in the elution fraction (Figure 2.5.2.1). The other three constructs (A3, A8 and A9) were low, and either produced no or low protein levels.

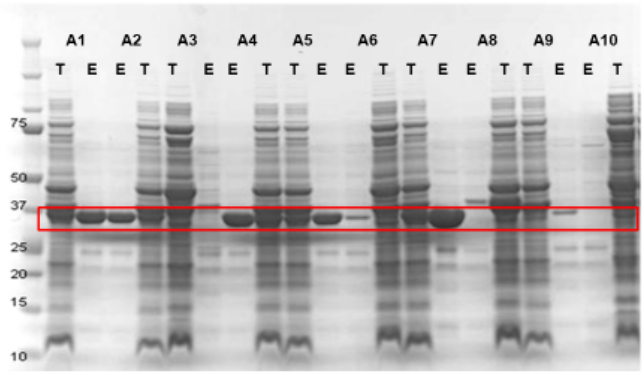


Figure 2.5.2.1 50 mL test expression SDS-PAGE of NMRAL1 constructs from PNIC28Bsa4. Total protein (T), Elution fraction (E) for each construct shown.

2.5.3 TEST PURIFICATION FROM BACULOVIRUS

3 mL expression volumes were grown in Sf9 cells for test purification of SDR constructs. Cells were harvested by centrifugation at 2800 rpm (Beckman Avanti, JLA. 8.1) for 20 min after 48 after induction. The resulting pellets were resuspended in 1 mL of ice cold PBS and re-centrifuged. Pellets were resuspended in 1 mL lysis buffer and processed like *E.coli* cells for test purification (Section 2.5.2).

2.5.4 LARGE SCALE PURIFICATION FROM *E.COLI*

Bacterial expression pellets were resuspended in binding buffer and flash frozen in liquid nitrogen and stored at -80°C until required. A flow chart of the overall large-scale purification method for human SDRs using affinity tags and size exclusion chromatography is summarised in Figure 2.5.4.1.

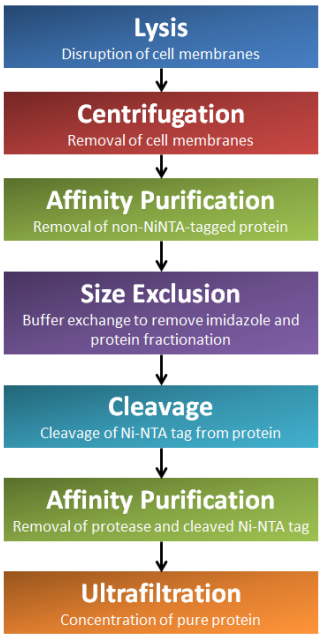


Figure 2.5.4.1 Flow chart of the general purification process for SDR proteins. Each purification step is with the primary goal of the purification step below.

Thawed cells already resuspended in lysis buffer (1 mL/g wet-weight of cells) were placed on ice and supplemented with Benzonase (3µL per 25 m. 30 U/mL). They were subsequently passed through a homogeniser 3-5 times (Emulsiflex C-5, Avestin, Canada) at 15 kpsi at 4°C. The resulting homogenate was supplemented with 1 mL of 0.5 U/mL polyethylenimine (PEI) and left to stand on ice for 10 minutes to precipitate the DNA. The lysate was centrifuged to remove insoluble protein and cell debris (Beckman, Avanti, JA-16.25, 16,000 rpm at 4°C for 60 minutes). The supernatants were decanted off and the pellets discarded.

Large-scale purification employs the same first nickel affinity chromatography step as the test expressions (Section 2.5.2). In Large-scale methods, nickel affinity is followed by size exclusion chromatography (SEC), also known as gel filtration (GF). The eluted fractions from the nickel column were analysed by SDS-PAGE for the selection of fractions to be for SEC. The pooled fractions were then concentrated down to <5 mL using a 10 kDa molecular weight cut-off (MWCO) Amicon Ultracel centrifugal. The eluate was applied to SEC column (HiPrep 16/60 SuperDex 200) pre-equilibrated with SEC buffer. Fractions collected in a 96-deep well block were analysed by SDS-PAGE to identify the location of the target protein. Results from SEC are shown as an elution profile or chromatogram that illustrates the variation in concentration (typically in terms of UV absorbance, for proteins usually at 280 nm) of sample components as they elute from the column from the largest to the smallest molecular weight. Molecules that do not enter the bead matrix are eluted together in the void volume (v) as they run directly through the column at the same speed as the buffer. Figure 2.5.4.2 shows a theoretical chromatogram of a high-resolution separation.

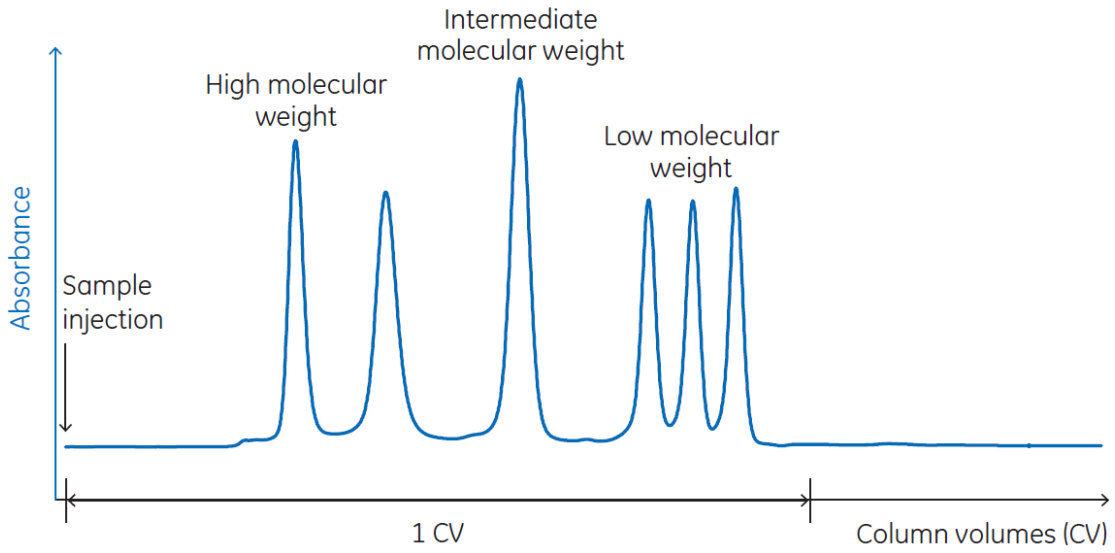


Figure 2.5.4.2: Schematic of the SEC chromatography method (GE Healthcare) - After the sample injection moleculles elute size order (highest molecular weight to lowest). https://www.gelifesciences.com/en/us/solutions/protein-research/knowledge-center/protein-purification-methods/size-exclusion-chromatography

Wells harbouring the target protein identified by SDS-PAGE were pooled together and cleaved using TEV protease. To remove TEV from the sample, the mixture was passed through a Ni-NTA resin pre-equilibrated with SEC buffer. The proteins were concentrated using 10 kDa (MWCO) Amicon Ultracel (Millipore). The concentrated proteins were ready for downstream applications or aliquoted, flash frozen in liquid nitrogen and stored at -80 °C until further use. Figure 2.5.4.3A shows Dehydrogenase/reductase (SDR family) member 4 (DHRS4) is highly expressed and fraction 1-9 were collected and applied to the gel filtration column, and then chromatograms were analysed by SDS-PAGE (Figure 2.5.4.3B) and fractions containing the target protein were pooled for TEV cleavage.

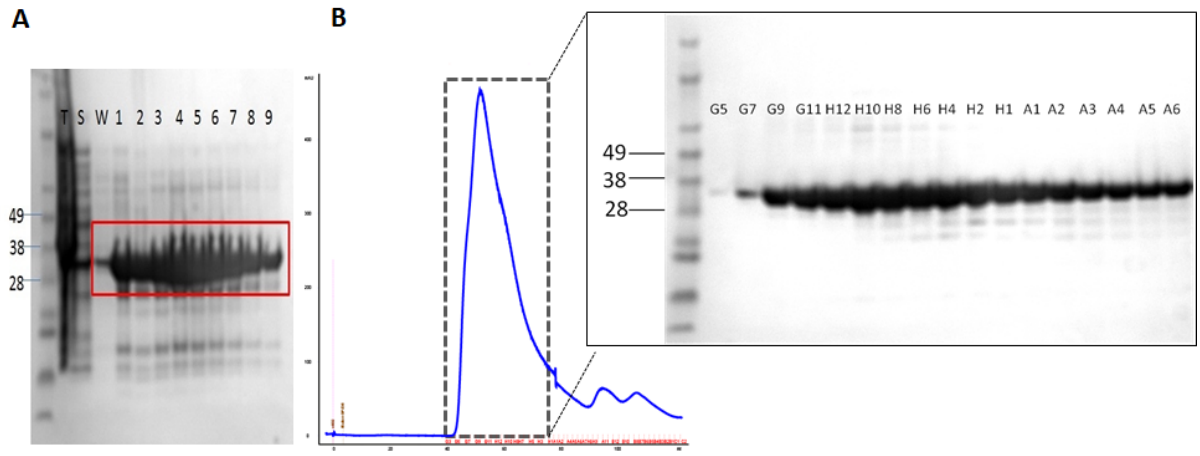


Figure 2.5.4.3 SDS-PAGE of DHRS4- M1-V250 A) after Ni-NTA column. Total protein (T), Soluble fraction (S), Wash (W), 5 mL Elution fractions (E1-9); B) after SEC column fractions (G5-A6)

2.5.5 LARGE-SCALE PURIFICATION FROM BACCULOVIRUS

Large-scale baculovirus expression was typically done in 1-3 litre volumes for a particular SDR construct. Pellets were resuspended in lysis buffer (1 mL/g wet-weight of cells) and homogenised like *E.coli* cells (Section 2.5.3). To pellet the cell debris, the homogenised lysate was centrifuged at 21,000 rpm using a JA-24.50 rotor for 30 min at 4 °C. The lysed cells were applied to a DE-52 column (10 mL diluted in 50 mL of 2.5M NaCl) pre-equilibrated with 100 mL of binding buffer. The flow-through was processed through a nickel affinity and SEC in the same manner as *E.coli* cells for large-scale purification (2.5.3).

2.6 PROTEIN VERIFICATION BY MASS SPECTROMETRY

Purified protein was verified by electrospray ionisation – time of flight (ESI-TOF) mass spectrometry (Agilent Technologies, Stockport, UK) to ensure the correct mass of the expressed protein. The protein was diluted to 0.1 mg/mL in 0.1 % formic acid. The solution was injected into the LC-ESI-TOF. The LC component of the system is a C2 reverse phase

column, which allows the system to desalt the protein effectively prior to application to the ESI-TOF. The system washes the column in 0.1 % formic acid, 5 % acetonitrile, and then elutes the column with a gradient from 5 % to 60 % acetonitrile. The resulting data was interpreted by QS Analyst (Applied Biosystems) and MagTran software (Zhang and Marshall 1998).

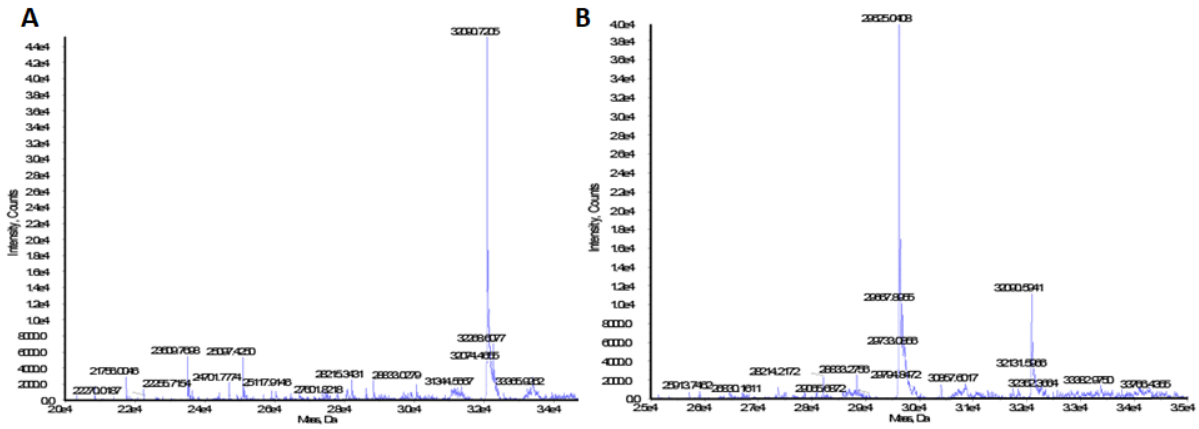


Figure 2.6.1 ESI-TOF mass spectrometry of DHRS4 before (A) and after TEV cleavage (B)

2.7 ANALYTICAL GEL FILTRATION

Analytical GF was used to study the oligomeric state of NMRAL1. Analytical GF was accomplished by loading protein sample onto Superdex 200 HiLoad 10/10 column (GE Healthcare) equilibrated with (50 mM HEPES pH 7.5, 300 mM KCl, 2.5 mM $MgCl_2$, 5 % glycerol). A 500 µg solution of NMRAL in sample buffer was used in each experiment. The column was calibrated using proteins of known molecular weight, including bovine milk α-lactalbumin (14.2 kDa), bovine erythrocytes carbonic anhydrase (29 kDa), chick egg albumin (45 kDa), BSA (66 kDa – monomer, 132 kDa – dimer), and Jack bean urease (272 kDa – trimer, 545 kDa – hexamer) (Sigma) as standards.

2.8 PROTEIN PRODUCTION SUMMARY

HUGO gene ID	SDR Cluster	Cloned	Test Expression	Large Scale Purification	Pure protein	Domain Boundary	Vector	DSF Assay
BDH1	C2	✓	✓	✓	×	×	pNIC-CTHF	×
BDH2	C1	✓	✓	✓	✓	L4-P236	p11	✓
SDR39U1	Atypical	✓	✓	✓	✓	M1-P282	pNIC-CTHF	✓
CBR1	C1	✓	✓	✓	✓	M1-F268	pLIC-SGC1	✓
CBR3	C1	✓	✓	✓	✓	M1-Q267	pNIC28-Bsa4	✓
CBR4	C1	✓	✓	×	×	×	-	×
DCXR	C1	✓	✓	✓	✓	M1-T234	p11	✓
DECR1	C1	✓	✓	×	×	×	-	×
DECR2	C1	✓	✓	×	×	×	-	×
DHRS1	C3	✓	✓	✓	✓	A3-P254	p11	✓
DHRS12	C2	×	×	×	×	×	-	×
DHRS13	C2	✓	✓	×	×	×	-	×
DHRS2B	C1	✓	✓	✓	✓	M23-I271	pNIC-CTHF	✓
DHRS3	C3	✓	✓	×	×	×	-	×
DHRS4	C1	✓	✓	✓	✓	M1-V250	pNIC-CTHF	✓
DHRS4L	C1	×	×	×	×	×	×	×
DHRS7	C3	×	×	×	×	×	×	×
DHRS7B	C3	✓	✓	✓	✓	V44-G281	p11	✓
DHRS7C	C3	×	×	×	×	×	×	×
DHRS9	C2	×	×	×	×	×	×	×
FAR1	Atypical	✓	✓	×	×	×	-	×
FAR2	Atypical	✓	✓	×	×	×	-	×
FASN	Atypical	✓	✓	×	×	×	-	×
GALE	Extended	✓	✓	✓	✓	M1-W336	p11	✓
GMDS	Extended	✓	✓	✓	✓	R23-V363	pLIC-SGC1	✓
HPGD	C1	✓	✓	✓	✓	V3-T247	p15	✓
HSD11β1	C3	✓	✓	✓	✓	E26-I275	p11	✓
HSD11β1L	C3	×	×	×	×	×	-	×
HSD11β2	C3	✓		×	×	×	-	×
HSD17β1	C2	✓		×	×	×	-	×
HSD17β10	C1	✓	✓	✓	✓	M1-R252	p11	✓
HSD17β11	C3	✓	✓	×	×	×	-	×

HSD17β12	C3	✓	✓	✓	✓	N41-L277	pFB-LIC-Bse	✓
HSD17β13	C3	✓	✓	×	×	×	-	×
HSD17β14	C1	✓	✓	×	×	×	-	×
HSD17β2	C2	✓	✓	×	×	×	-	×
HSD17β3	C3	✓	✓	×	×	×	-	×
HSD17β4	C1	✓	✓	✓	✓	M1-I295	p11	✓
HSD17β6	C2	✓	✓	×	×	×	-	×
HSD17β7	C2	✓	✓	×	×	×	-	×
HSD17β8	C1	✓	✓	✓	✓	Q6-S252	pNIC-CTHF	✓
HSD3β1	Extended	✓	✓	×	×	×	-	×
HSD3β2	Extended	✓	✓	×	×	×	-	×
HSD3β7	Extended	✓	✓	×	×	×	-	×
HSDL1	C2	✓	✓	×	×	×	-	×
HSDL2	C1	✓	✓	✓	✓	M1-K284	pET28	✓
HSPC105	Atypical	✓	✓	×	×	×	-	×
HTATIP2	Atypical	✓	✓	✓	✓	M1-K228	pNIC-CTHF	✓
KDSR	C2	✓	✓	×	×	×	-	×
MAT2B	Extended	✓	✓	✓	✓	M27-Q341	pNIC-CTHF	✓
DHRS11	C1	✓	✓	✓	✓	M1-T356	p11	✓
NDUFA9	Atypical	✓	✓	×	×	×	-	×
NMRAL1	Atypical	✓	✓	✓	✓	M1-L289	pNIC28-Bsa4	✓
NSDHL	Extended	✓	✓	×	×	×	-	×
PECR	C1	✓	✓	✓	✓	M1-294	p11	✓
QDPR	Atypical	✓	✓	✓	✓	E9-T233	p11	✓
RDH10	C2	✓	×	×	×	×	-	×
RDH11	C2	✓	×	×	×	×	×	×
RDH12	C2	✓	×	×	×	×	×	×
RDH13	C2	✓	×	×	×	×	×	×
RDH14	C2	✓	×	×	×	×	×	×
RDH16	C2	×	×	×	×	×	×	×
RDH5	C3	✓	✓	×	×	×	×	×
RDH8	C3	✓	✓	×	×	×	×	×
RDHE2	C2	✓	✓	×	×	×	×	×
SDR-O	C3	×	×	×	×	×	×	×

SPR	C1	✓	✓	✓	✓	M1-Q239	pLIC-SGC1	✓
TGDS	Extended	×	×	×	×	×	×	×
TSTA3	Extended	✓	✓	✓	✓	S7-W311	pNIC28-Bsa4	✓
UXS1	Extended	✓	✓	✓	✓	E85-K393	pNIC28-Bsa4	✓
WWOX	C2	✓	✓	✓	✓	M1-L404	PFastBac	✓

Table 2.8.1 Protein production summary of human SDRs - ✓ carried out and completed ✓collaborator provided ×not carried or unsuccessful NB: vectors in green used in large scale purification

The stages of human SDRs purification from cloning to purification are in Table 2.8.1. The table also shows the domain boundary and vector used for the pure protein achieved. In total 63 human SDRs have been cloned successfully, 29 were expressed at large scale and 28 produced pure protein for downstream processing (chapter 4-6).

2.9 DIFFERENTIAL SCANNING FLUORIMETRY (DSF)

2.9.1 BACKGROUND

DSF is a method that exploits the thermal unfolding of a native protein with the presence of a fluorescent hydrophobic dye (e.g. SYPRO orange). In the native state, hydrophobic residues are compacted within the protein interior, and upon unfolding become accessible to the reporter dye which fluoresces at 490 nm (Cummings et al. 2006) (Figure 2.9.1.1). As the dye binds to the exposed hydrophobic residues during protein unfolding, an increase in fluorescence occurs (Niesen et al. 2007) The melting temperature, $T_{m,}$ corresponds to the midpoint between the lowest and highest value of the fluorescence curve, representing a half population of native and unfolded states (Figure 2.9.1.1). T_m was calculated using custom EXCEL-based calculation software accessible at (ftp://ftp.sgc.ox.ac.uk/pub/biophysics), and fitting data to a Boltzmann plot as performed in GraphPad PRISM 5.0. The difference between the T_m control (T_{m1}) and sample-containing compound (T_{m2}) was expressed as the T_m shift (Figure 2.9.1.1). A positive T_m shift represents the stabilisation of the protein by the compound, and is as an indicator of ligand binding. The unfolding gradient ($\Delta I/\Delta T$) often reflects the nature of protein unfolding: steep transitions are indicative for highly cooperative unfolding while shallow transitions indicate high flexibility unfolding (Lavinder et al. 2009).

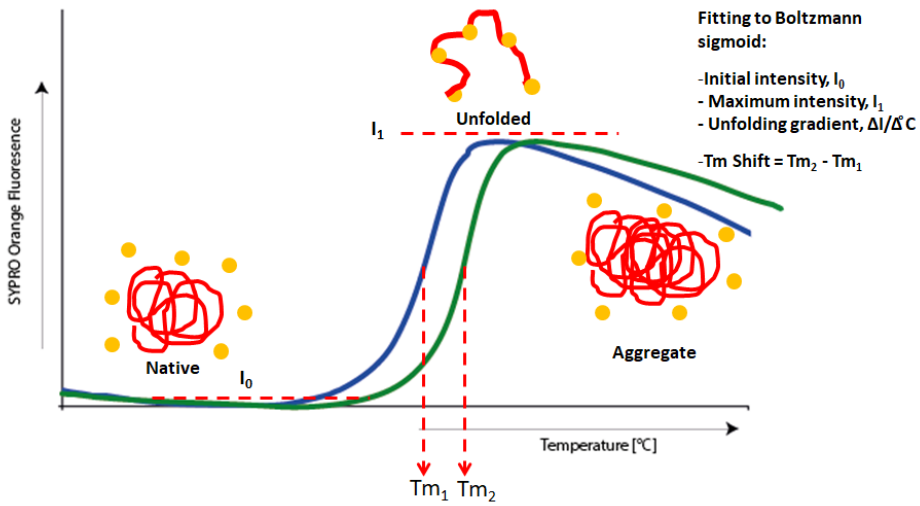


Figure 2.9.1.1 DSF unfolding states and midpoint calculation – calculation of midpoint (T_m), T_m shift and Unfolding gradient ($\Delta I/\Delta^{\circ}C$). Native globular protein at the baseline of the curve becomes unfolded, as temperature increases. SYPRO orange (oragnge circles) binds to hydrophobic patches as the protein nfolds and fluorescence increases. Following the maximum in intensity, a gradual decrease is observed, which is mainly explained by protein being removed from solution owing to precipitation and aggregation. Adapted from (Niesen et al. 2007).

2.9.2 EXPERIMENTAL SETUP FOR COFACTOR DETERMINATION

SYPRO Orange is the preferred dye with its much higher signal to noise ratio, and its superior fluorescence characteristics that have minimal interference at its emission wavelength (Niesen et al. 2007). The resulting plate was sealed with PCR film, centrifuged at 1000 rpm for 1 min and analysed using a real-time PCR machine Mx3005P (Stratagene Inc.), with the temperature being increased at 1°C per minute from 25 to 95°C, and monitoring fluorescence (excitation 492nm, emission 610nm) continuously.

Final Concentration	Reagent	Volume in 4200 µL
	10mM HEPES pH7.5, 150mM NaCl	e.g. 4191 µL
1 µM	Protein	e.g. 4.8 µL
1 X	SYPROrange	4.2 µL

4200 µL total solution (stock mix).

Table 2.9.2.1 Master mix components for coenzyme DSF screen - Cofactor screening compounds were dissolved in DMSO to 2mM and stored in 96-well storage plate. A master mix solution (39 µL) was distributed into each well of 96-well PCR plate. 1 µL of compounds from the storage plate were added to the appropriate wells resulting in a final concentration of 50 µM.

2.9.3 EXPERIMENTAL SETUP FOR LIGAND SCREENING

The ligand library was screened with the determined cofactor (Section 4.4) in two distinct sets, reduced [NAD(P)H)] and the oxidised [NAD(P)]. SDRs showing no distinct T_m shifts for either NAD/H or NADP/H were screened against the ligand library using the cofactor sourced from experimental data in literature or sequence predictions.

The set-up for DSF ligand screening for SDRs is similar to cofactor screening (section 2.9.2). The master mix solution (39 µl) of protein, dye, cofactor and buffer was distributed into each well of 96-well PCR plate (Table 2.9.3.1). The resulting plate was sealed, centrifuged and run as described in section 2.9.2. T_m shift values have been colour-coded (1.4-2.4°C – red; 2.5-3.9°C yellow; ≥4 °C – green) to produce a graphical representation for easy identification of thermal shifts across of the SDR-family. T_m shift values below 1.4 °C are annotated as below threshold (b.t), and they considered non-significant.

Final Concentration	Reagent	Volume in 4200 µl
	10mM HEPES pH7.5, 150mM NaCl	e.g. 4186.8
1 µM	Protein	e.g. 4.8
1 X	SYPROrange	4.2 µl
200 µM	NAD(P)(H)	4.2 µl

Table 2.9.3.1 Master mix components for ligand DSF screen - Ligand screening compounds were dissolved in DMSO to 0.4 mM and stored in 96-well storage plate. The master mix solution (39 µl) of protein, dye, cofactor and buffer was distributed into each well of 96-well PCR plate. 1 µL of compounds from the storage plate were added to the appropriate wells resulting in a final concentration of 10 µM.

2.10 X-RAY CRYSATALLOGRAPHY

2.10.1 BACKGROUND

X-ray crystallography is a technique for protein structure determination, wherein a homogenous protein sample undergoes slow precipitation in an aqueous solution to form a crystal. These crystals can be subjected to X-rays to provide diffraction patterns for the determination of 3D atomic structures, where the position and interactions of each atom in the 3D space can be defined. The 3D structures enhance the understanding of protein function in terms of; how molecules interact, how enzymes catalyse reactions, and how to aid design of novel drugs that target a particular protein.

The starting point of protein crystallisation requires a highly homogeneous soluble protein sample. For crystallisation to occur**,** protein molecules must separate from the solution and self-assemble into an ordered lattice. The crystallisation is achieved by reducing the solubility of the protein in the solution by using cooling, precipitants or evaporation.

Addition of chemicals such as, polyethylene glycol (PEG), 2-methyl-2,4-pentanediol (MPD), or ammonium sulphate are the most common ways of reducing solubility.

A crystallisation process involves three main steps, i.e., the development of supersaturation, followed by nucleation, and then growth. Figure 2.10.1.1 shows solubility represented by the phase diagram, in a typical crystallisation experiment (McPherson 2004). The solubility curve represents the line in which protein solute is in equilibrium with the buffer (saturated). Towards the left of the solubility curve (low protein and low precipitant concentrations), protein solute is in solution (undersaturated), and no crystals will form. Towards the right of the solubility curve, the protein solute begins to precipitate out of solution (metastable), and within the region of supersaturation (increasing protein or precipitant concentration, or both) crystal growth occurs. Supersaturation is a non-equilibrium condition, and formation of stable nuclei will continue to grow into crystals until the system regains equilibrium (McPherson 2004).

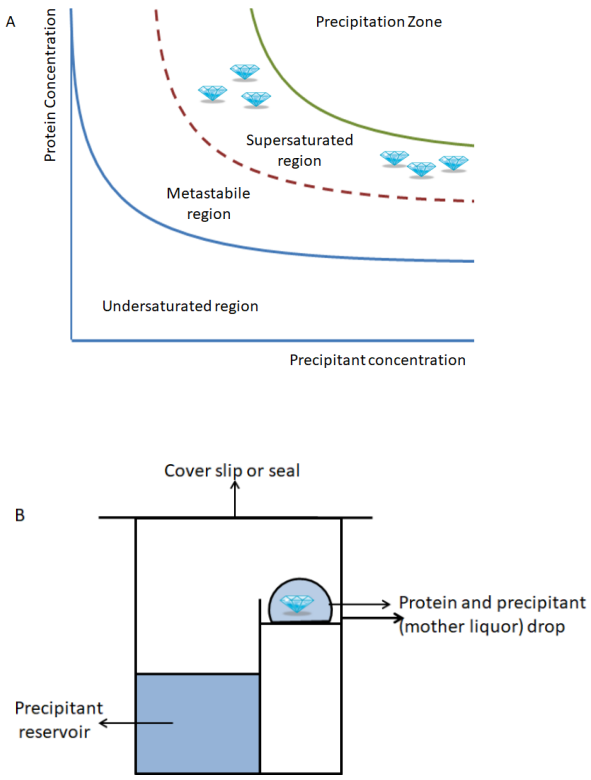


Figure 2.10.1.1 Crystal formation – A) phase diagram for crystallisation – solubility of a protein in a crystallisation solution as a function of the concentration of the precipitant present B) Sitting drop method for crystallisation – unsaturated precipitant containing protein solution is over a reservoir. Through-vapour equilibration of the droplet and reservoir causes the protein solution to reach a supersaturation level where nucleation and initial growth occur (Weber 1991).

In crystallisation, proteins are subjected to a variety of conditions (e.g. salts, pH, temperature and additives) systematically to influence a supersaturated state and hence crystal formation. The molecular diversity between individual proteins means it is difficult to predict precise crystallisation conditions. For example, some proteins form precipitate at low salt concentration and others at high concentrations. The physical process of causing a protein to come out of solution is usually carried out through a process called vapour diffusion. In this technique, water vaporises from the drop containing protein and precipitant, and drives the droplet from a state of unsaturated to supersaturated state,

where crystallisation can occur. In this thesis 'sitting' drop vapour diffusion was employed (Figure 2.10.1.1B).

The aim for an X-ray diffraction experiment is to grow single crystals of suitable size (0.2 - 0.4 mm in at least two of the three dimensions). Crystals are mounted using a nylon loop and cryoprotected before rapidly cooling in liquid nitrogen. The crystal is mounted on a goniometer for measurements so that the X-ray beam is through the crystal and rotated. Crystals act as 3D diffraction gratings, which gives rises to constructive and destructive interference. The diffracted radiation is on a detector as an array of spots, which are as known reflections. The intensity of these reflections is measured to determine the distribution of each electron in the crystal resulting in an electron density map. The wavelength and diffraction angle can be used to calculate the spacing between atomic planes from Braggs Law (Eq 2.10.1.1)

$n\lambda = 2d\sin\theta$ (Equation 2.10.1.1)

Where *n* is an integer, λ is the wavelength of the incident X-ray beam, *d* is the spacing between the atomic layers in a crystal, and θ is the angle of incidence of the X-ray beam.

The processing of diffraction to electron density maps requires mathematical calculations using the amplitude and phases of X-rays in each reflection. Amplitudes can be directly measured from diffracting crystals, however phases are lost during the experiment, which is known as the phase problem. The most rapid method and common method is molecular replacement, whereby a protein known with similar 3D structure is used as a model to generate approximate phases for the protein under investigation. Once a preliminary

model is generated, refinement of phases is needed to improve the agreement of the model (R_{fact}) with the observed data in the electron density map (R_{free}).

2.10.2 HIGH-THROUGHPUT CRYSTALLISATION OF PROTEINS

Purified SDR proteins were concentrated to around 5-10 mg/mL prior to crystallisation, and supplemented with the relevant cofactors in five times of the molar concentration of protein. The SGC adopts a high-throughput approach in crystallisation using automated robots, where protein and mother liquor can be dispensed in nanolitre volumes across 96-well plates, which allows a variety of conditions to be in parallel. Crystallisation solutions are in a high capacity liquid handling platform (Perkin Elmer liquid handler 96 well pipetting system). Sitting drop vapour diffusion crystallisation experiments are set up using an automated liquid handler (Mosquito, Labtech TTP) in 3-well sitting drop plates (Greiner BioOne). Three ratios of protein: crystallisation reagent (2:1, 1:1 and 1:2) were used for 150 nL drops to increase the number of crystallisation conditions and the chance of crystal formation. The plates were incubated for crystal growth in 4 °C and 20 °C incubators (Minstrel, Rigaku Corp) coupled with CrystalMation for plate storage and handling, imaging and analysis.

2.10.3 CRYSTAL MOUNTING AND DATA COLLECTION

SGC staff mounted crystals using a nylon loop and cryoprotected before rapidly cooling in liquid nitrogen. Diffraction data were collected either in the in-house X-ray diffractometer (Rigaku-FRE) coupled to a diffraction detector, or using synchrotron radiation at the

Diamond Light Source by SGC Staff. Diffraction data collected was processed and scaled with MOSFLM and SCALA from the CCP4 suite (Winn et al. 2011). The program Mosflm (Leslie 1992) was used to analyse the diffraction patterns and to calculate the space group of all protein crystals. Diffraction. Symmetry-related reflections were scaled by SCALA (Winn et al. 2011).

2.10.4 MODEL BUILDING AND REFINEMENT

The structures were solved by molecular replacement using the program *PHASER* (McCoy et al. 2005). Automated model building was performed with ARP/wARP (Perrakis et al. 2001), followed by iterative cycles of restrained refinement and model building using COOT (Emsley and Cowtan 2004) and REFMAC5 (Murshudov et al. 1997).

2.11 SUBSTRATE SCREENING FOR ORPHAN SDRS

Substrate screening for the Dehydrogenase/reductase (SDR family) member 1 (DHRS1), Dehydrogenase/reductase (SDR family) member 11 (DHRS11) and hydroxysteroid dehydrogenase like 2 (HSDL2), was carried out with compounds from their identified substrate classes from DSF (Section 5.3.4). Reactions were recorded using a SpectroStar Omega (BMG Labtech), following a change in absorbance at 340nm for NADP/H-dependent oxidoreductase activity. Reactions were performed at room temperature in 40 µL reactions in a 384-well microtiter plates at different pH values (5, 7.5 and 9). Eventhough enzymes are naturally active at 37 °C, room temperature was used, as the proteins were stable for longer

periods. Different pHs were tested to account for preferences in directionality (i.e reductases favour pH < 7.5 and dehydrogenases favour pH > 7.5). The set-up for substrate screening is shown in Table 2.11.1. Figure 2.11.1 shows the setup of a typical screen using a 96-well compound stock plate.

			Volume (µL)		
Component	**Final Concentration**	**Test**	**Control 1**	**Control 2**	**Control 3**
Protein	500nM	4	-	4	-
Cofactor	100 mM	10	10	10	10
Compound	10 mM	1	1		1
Buffer	-	25	29	25	29
DMSO	-	-	-	1	-
		Total 40 µL			

Table 2.11.1 Activity screening setup for HSDL2 and DHRS11

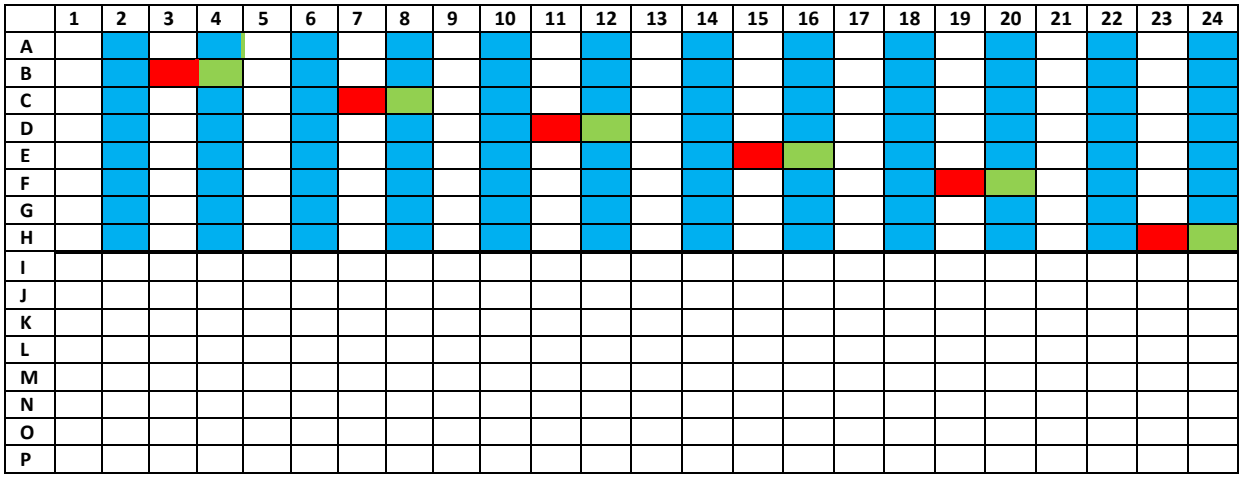


Figure 2.11.1 SDR Substrate screen setup ❑Test condition Control 1 Control 2 Control 3

CHAPTER 3: SDR FAMILY ANNOTATION

3.1 INTRODUCTION

This chapter focuses on the annotation of the human SDR family by analysing computational and experimental data to provide an overview of what is currently known in terms of cofactor preference, substrate class and subcellular localisation for each member. Exploring the relationships among these factors can further aid functional predictions of uncharacterised SDRs. Of 73 human SDR in total, there are currently 16 members that lack any description of their biological roles (Appendix, Table 1), and 23 members that are deficient in the description of subcellular localisation (Appendix, Table 2). Information on protein substrate and subcellular localisation is important to map how human SDRs make up the functions in a living cell, and to better understand any links to function across a protein family. Functional descriptions of proteins are often annotated in large databases such as Uniprot (Magrane and Consortium 2011) and GeneCard (Rappaport et al. 2013), providing information of function, catalysis, ligands and subcellular localisation. However, they often do not present up-to-date information from scientific papers. Moreover, computational predictions are often from untraceable sources, and therefore can be unreliable. Published experimental data on substrate activity/function and subcellular localisation is summarised in Table 1 and Table 2 (Appendix), respectively.

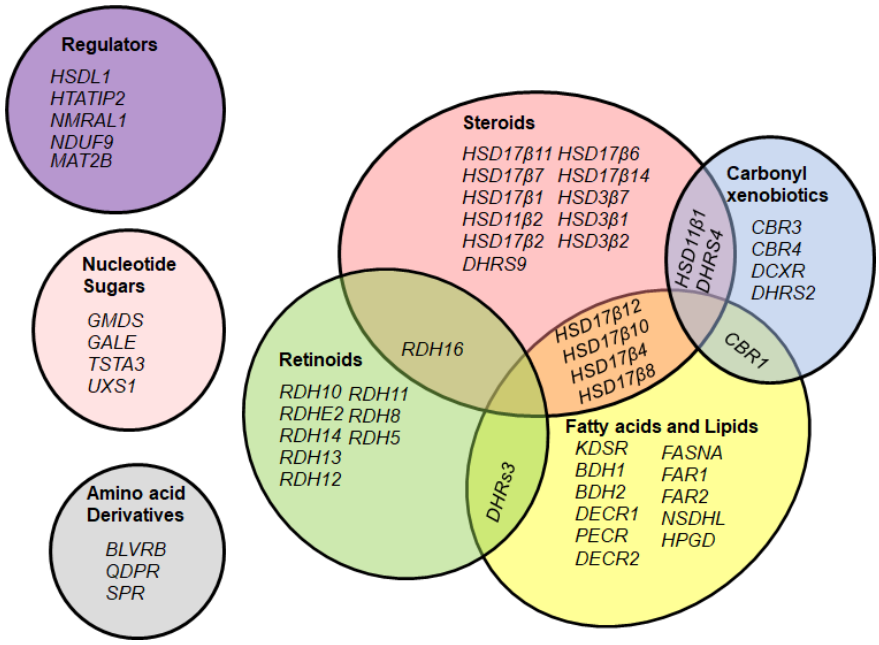


Figure 3.1.1 Venn diagram of SDR substrate specificity

To gain insight into the overall functional makeup of the human SDR family, protein members have been categorised into different substrate classes (carbonyl xenobiotics, fatty acids and lipids, amino acid derivative, retinoids, regulators, steroids, and sugars) based on their catalytic activities, and functions (Table 1, Appendix). SDRs with unknown substrates or functions are classified as orphan: they often possess the necessary residues for catalytic activity, but their substrates or functions are yet to be determined. The distribution of human SDRs with known substrates is depicted as a Venn diagram (Figure 3.1.1), which represents a quick overview of human SDRs that have activity towards a particular substrate class, and demonstrates that some SDRs have overlapping substrates (CBR1, DHRS3, DHRS4, HSD11β1, HSD17β4, HSD17β8, HSD17β12, HSD17β10 and RDH16). Human SDRs that turnover nucleotide sugars, amino acid derivatives, or hold a regulatory function represent discrete substrate classes that do not overlap with other substrates. The largest group of

human SDRs are those that turnover steroid compounds, shortly followed by fatty acids and lipids. SDRs that have activity towards amino acid derivatives represent the smallest substrate class in the SDR family. Human SDRs within a substrate class usually fall into different sequence clusters of the SDR superfamily map (Figure 1.4.4.2). This observation means we are unable to derive functional roles from protein sequence or cluster location alone, as they do not have a uniform correlation with each other (Bhatia et al. 2014).

In an attempt to predict functions for orphan human SDRs, this chapter explores relationships in cofactor preference (Section 3.2), substrate accommodation (Section 3.3) and subcellular localisation (section 3.3). human SDRs from each sequence cluster (C3, C2, C1, Extended and Atypical) are presented and recorded against cofactor preference (NAD/H or NADP/H), substrate class and subcellular localisation (Table 3.1.2).

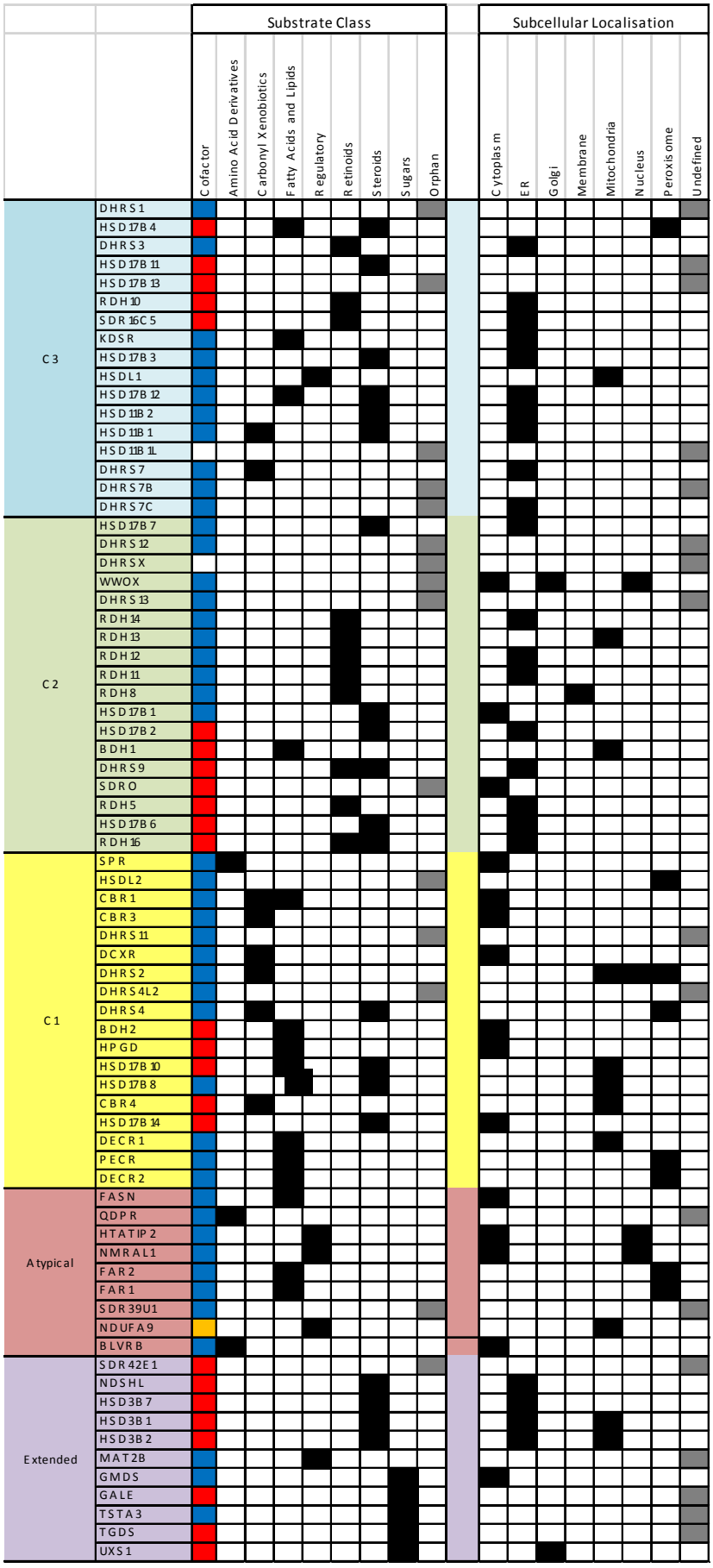


Figure 3.1.2 Human SDRs analysed by cofactor, substrate class and subcellular localisation – Information on cofactor preferences were acquired from the Uniprot database. Substrate class and subcellular localisation data were collated from Table 1-2 of Appendix.

NAD/H

NADP/H

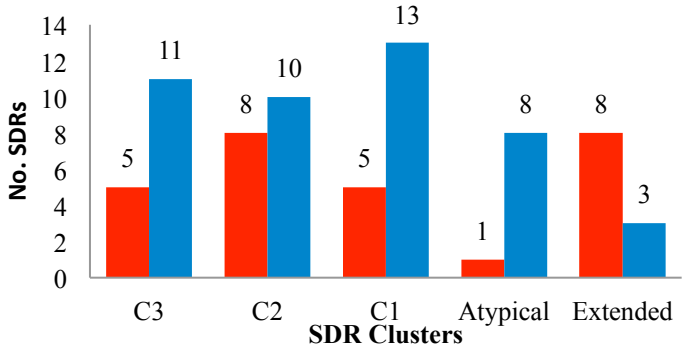
FAD

characterised by SDR activity assay

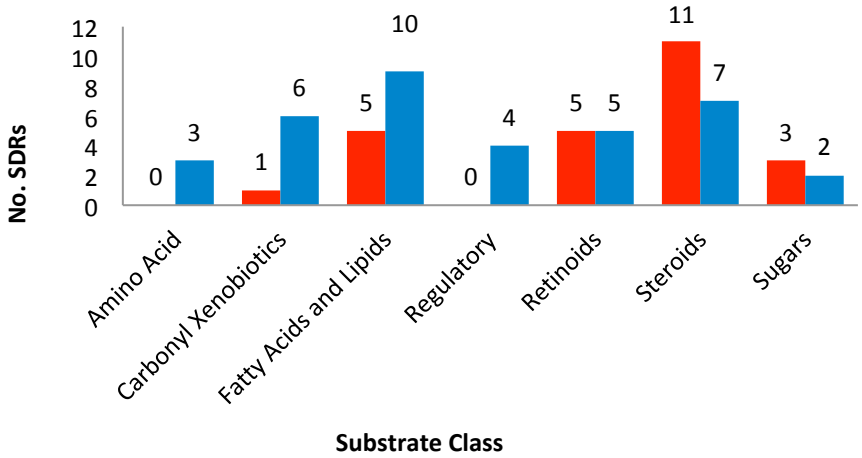
Unknown function or subcellular localisation

3.2 COFACTOR DISTRIBUTION OF HUMAN SDRS

A



B



C

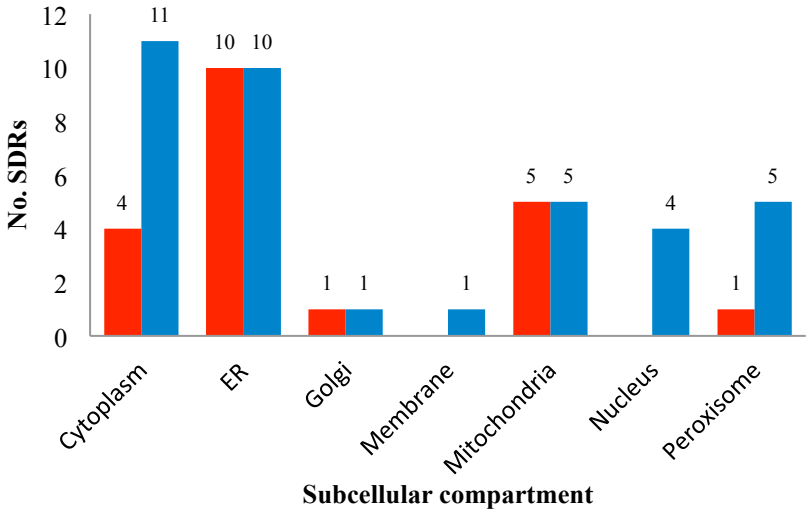


Figure 3.2.1 Human SDR Cofactor preference distribution A) SDR clusters B) Substrate class C) Subcellular localisation NAD/H NADP/H

Domain	Kingdom	Phylum	Organism	Total No. SDRs	Cofactor Preference	
					NAD/H	NADP/H
Eukaryotic	Animalia	Chordata	*Homo sapiens*	73	24	**45**
			Mus musculus	82	35	**46**
			Danio rerio	94	27	**65**
		Anthropoda	*Drosophila melanogaster*	37	6	**31**
		Nematoda	*Caenorhabditis elegans*	33	12	**21**
	Plantae	Angiosperms	*Arabidopsis thaliana*	121	44	**77**
	Fungi	Ascomycota	*Saccharomyces cerevisiae*	33	6	**27**
	Protist	Mycetosoa	*Dictyostelium discoideum*	8	4	3
Prokaryotic	Proteobacteria	Eubacteria	*Escherichia coli (Strain K12)*	14	10	4
			Bacillis subtilis (STRAIN 168)	32	13	22
Archaea	Euryarchaeota	Euryarchaeota	*Haloarcula hispanica*	9	5	4
			Methanosarcina acetivorans	5	2	3

Table 3.2.1 Comparison of NAD/H and NADP/H preference of SDRs across different genomes.

Cofactor preferences for the human SDR members have been analysed in relation to each SDR sequence cluster (Figure 3.2.1A), substrate class (Figure 3.2.1B) and subcellular localisation (Figure 3.2.1C). The human SDR family shows an overall preference for NADP/H (n=45, 62.5 %) as compared to NAD/H (n=24, 33.3 %). NADP/H is the preferred cofactor in each SDR cluster, with the exception of the extended cluster that displays a preference for NAD/H to NADP/H (8:3) (Figure 3.2.1A). This reinforces the idea that NAD/H may have been the initial SDR cofactor as the extended family contains some of the oldest SDR members. Other eukaryotic organisms such as *Drosophila melanogaster, Arabidopsis thaliana, Saccharomyces cerevisiae* also exhibit an overall preference for NADP/H. Smaller SDR genomes (e.g. *Haloarcula hispanica*), however, show no distinct preference between NAD/H and NADP/H (Table 3.2.1).

Figure 3.2.1B displays the cofactor preference samongst each SDR substrate class. SDRs that turnover retinoids, steroids or sugars show no distinct preference. SDRs that act on fatty acids and lipids display a preference for NADP/H over NAD/H (10:5), as do SDRs that turnover carbonyl xenobiotics (6:1). SDRs that have a regulatory function or turnover amino acid derivative substrates are the only class to show complete partiality towards one cofactor, NADP/H.

Subcellular localisation analysis (Figure 3.2.1C) shows an even distribution between NAD/H and NADP/H for SDRs found in the Golgi (1:1), mitochondria (5:5) and ER (10:10). Human SDRs found in the nucleus bind NADP/H as their cofactor. SDRs that localise to the peroxisome show a preference for NADP/H over NAD/H (5:1). Cytoplasmic SDRs also favour NADP/H binding (11:4). RDH8, the only membrane localised SDR that is found in the outer segments of photoreceptor cells, binds NADP/H (Rattner et al. 2000).

3.3 SUBSTRATE DISTRIBUTION

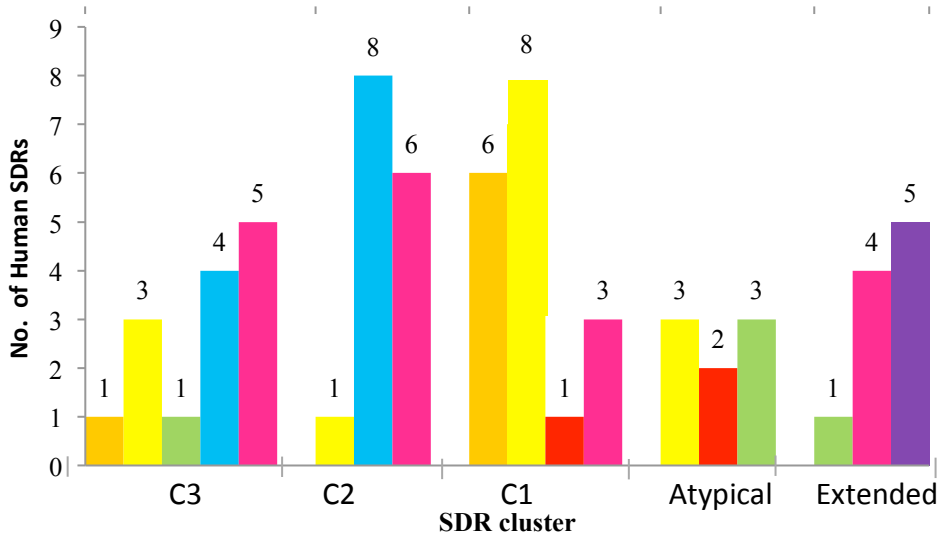


Figure 3.3.1 SDR substrate distribution in relation to SDR clusters - Amino Acid Derivatives Carbonyl Xenobiotics Fatty Acids and Lipids Regulatory Retinoids Steroids Sugars

SDR members with assigned substrate classes have been analysed in relation to each sequence cluster (Figure 3.3.1), using the data from Table 3.1.2. This shows that substrate classes across the human SDR family are distributed unevenly across clusters. SDR enzymes that turnover amino acid derivatives are found/located exclusively to C1 and Atypical clusters. SDRs that turnover carbonyl xenobiotics are concentrated in cluster C1. Similarly, the highest occurrence of SDRs active towards fatty acids is in C1, but also found in C3, C2 and Atypical clusters. Most SDRs with a regulatory function belong to the atypical cluster, with one (MAT2B) in the extended cluster. SDRs that turnover retinoids are exclusively found in clusters C3 and C2. SDRs that catalyse steroids also show the highest frequency in clusters C3 and C2, but are additionally located in C1 and extended clusters. Unlike the other substrate classes, the extended cluster only represents SDRs with catalytic activity towards sugars.

3.4. SUBCELLULAR LOCALISATION

This section analyses aspects of subcellular localisation for each human SDR, namely the distribution amongst SDR clusters (Section 3.4.1), correlations with substrate class (Section 3.4.2), and the study of targeting signals that can guide subcellular localisation (Section 3.4.3).

3.4.1 SUBCELLULAR LOCALISATION DISTRIBUTION

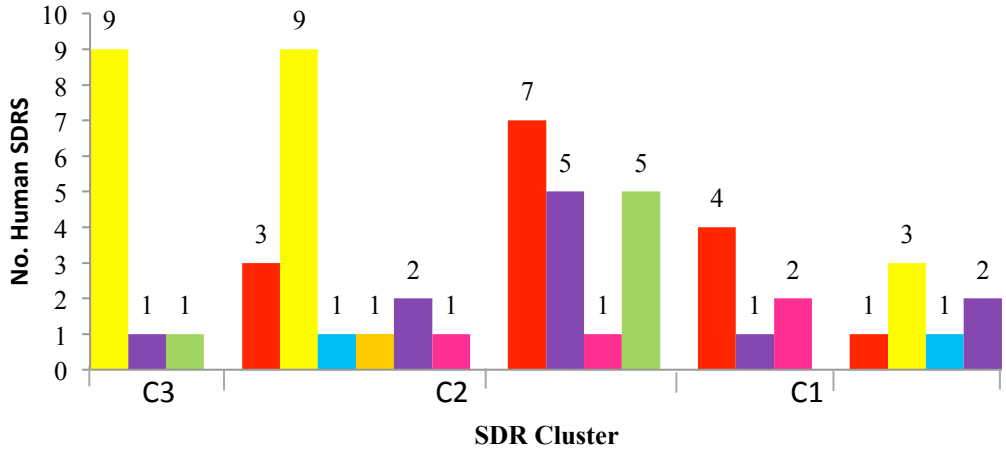


Figure 3.4.1.1 Subcellular localisation in SDR clusters - Cytoplasm ER Golgi Membrane Mitochondrial Nucleus Peroxisome

Figure 3.4.1.1 shows SDRs with subcellular localisation data (Figure 3.1.2) in relation to cluster location. The figure reveals that no particular SDR subcellular localisation is confined to one sequence cluster. The most prominent subcellular localisation for SDRs in cluster C3 is the ER (9/11). In comparison, cluster C2 has a broader range of subcellular compartments than C3, with about half of SDRs localising to the ER (9/17). Cluster C1 has an equal spread of SDRs across the subcellular compartments such as the cytoplasm (7/18), mitochondria

(5/18) and peroxisome (5/18). The atypical SDRs localise to the cytoplasm (4/7) but also to the nucleus (2/8) and mitochondria (1/8). Extended SDR members are distributed amongst the ER (3/7), mitochondria (2/7), cytoplasm (1/7) and Golgi (1/7).

3.4.2 SUBCELLULAR LOCALISATION AND SUBSTRATE CLASS ANALYSIS

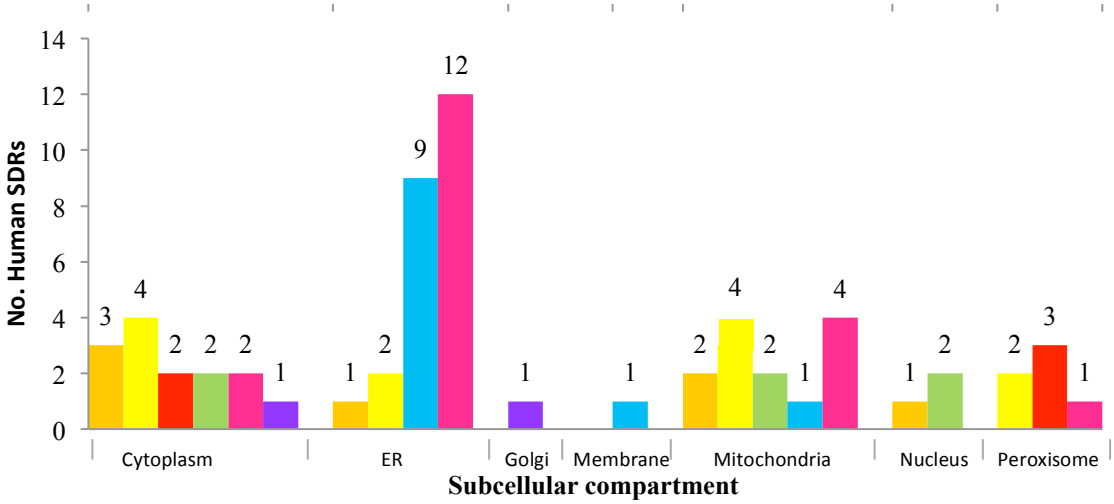


Figure 3.4.2.1 SDR substrate distribution in relation to subcellular localisation - Amino Acid Derivatives Carbonyl Xenobiotics Fatty Acids and Lipids Regulatory Retinoids Steroids Sugars

The distribution of substrate classes in subcellular compartments (Figure 3.4.2.1) illustrates there is no obvious correlation between the two, though a few patterns exist. For example, SDRs that localise in the ER mostly turnover steroids (12/24) or retinoids (9/24). SDRs that act on amino acid derivatives appear only in the cytoplasm. SDRs with catalytic activity towards carbonyl xenobiotics are found in the cytoplasm, ER, mitochondria and peroxisome. Only two SDRs out of five SDRs with sugar substrates have been determined for subcellular localisation (UXS1 - Golgi and GMDS - cytoplasm). It is therefore, difficult to derive any pattern with this substrate class in regard to subcellular localisation. SDRs with regulatory functions tend to localise in the nucleus or cytoplasm. Mitochondria predominantly host SDRs that act on steroids (4/13), or fatty acids and lipids (4/13).

3.4.3 TARGETING SIGNALS

In general eukaryotic proteins are synthesised in the cytosol, and many are sorted via targeting signals into their final subcellular destinations to carry out their functions. Mutations in targeting signals can lead to improper subcellular localisation, which causes protein dysfunction and disease. For example, primary hyperoxaluria type 1 an autosomal recessive disorder that results in kidney failure is caused by mistargeting of alanine glyoxylate aminotransferase (AGT) to the mitochondria instead of the peroxisome (Purdue et al. 1991). Targeting signals are sequences that direct proteins to the relevant organelles inside the cell, or the extracellular space via secretion. Targeting signals are recognised by receptors on the exterior of an organelle, and then the protein is translocated across the membrane into the organelle interior. Table 3.4.3.1 shows a summary of some well-known targeting signals, including signal peptides, signal anchors, mitochondrial targeting signals (MTS), nucleus localisation signals (NLS), peroxisomal targeting signals (PTS), and ER retention signals.

Signal Location	Targeting Sequence Type	Features	Source
N-terminus	Signal peptide	- 20 AA stretch. -Basic N-term, Hydrophobic core , Polar C-term -Cleavable peptide site by signal peptidase.	(Kall et. al 2004)
	Signal Anchor	-Longer hydrophobic stretches than signal peptides -Remain attached to the membrane of an organelle. - No cleavable site	(Nielsen and Krogh 1998)
	Mitochondrial Targeting Signal (MTS)	- 10-70 AA - Alternating pattern of hydrophobic and positive residues. - Often contain cleavable sites by matrix processing peptidase	(Claros 1995)
Central	Nuclear Localisation Signal (NLS)	-One stretch around 5 basic residues OR two stretches of basic residues separated by ~10 AA -Preceded by a helix-breaking residue (e.g. Pro, Gly)	(McLane and Corbett 2009)
C-terminus	Endoplasmic Reticulum (ER) retention	(H/K)DEL KKxx or KxKxx (where x is any AA)	(Stornaiuolo et al. 2004)
	Nuclear Export Signal (NES)	4–5 hydrophobic residues within a region of ~10 amino acids $L_{xxx}L_{xx}L_xL$ (where L is a hydrophobic residue, usually Leu, and x is any AA)	(La Cour et al. 2004)
	Peroxisomal Targeting Signal 1 (PTS1)	C-terminal tripeptide sequence with 5 polar AA upstream (S/A) (K/R) (L/M)	(Neuberger et al. 2003)

Table 3.4.3.1 Summary of targeting signals in eukaryotic proteins.

The protein sequences for all the human SDR members were analysed for targeting signals to aid subcellular prediction and identify any association with function (Appendix, Table 2). Using SignalP v4.0 (Petersen et al. 2011) and phobius (Kall et al. 2004) servers, all human SDRs with predicted N- and C-terminal signalling sequences were identified (Appendix, Table 2). SDR sequences were also analysed by PTS1 predictor (Neuberger et al. 2003). Mitoprot was used to analyse the SDR family for MTS sequences (Claros 1995). Other targeting signals were obtained from the Uniprot database and published experimental data (Appendix, Table 2).

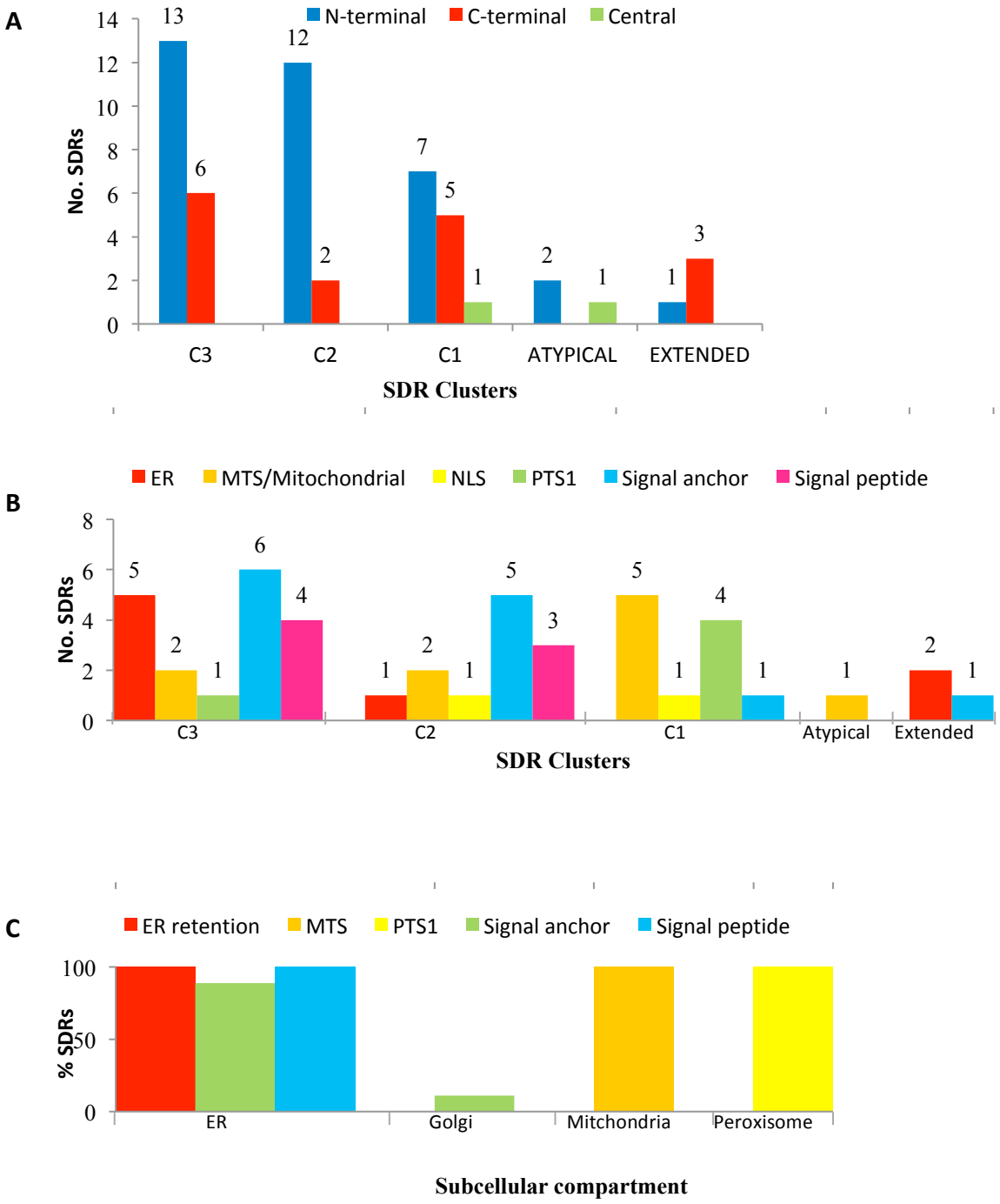


Figure 3.4.3.1 Analyses of targeting signals in human SDRs A) Distribution of targeting signals B) Distribution of targeting signal type C) experimental localisation of SDRs with targeting signals

Figure 3.4.3.1A shows the various targeting signals exist in over 60 % of all human SDRs. Around 66 % of the targeting signals in the SDR family are located in the N-terminus. Only two SDRs contain signalling sequence in the central region of the protein. Clusters C3 and C2 have the highest frequency of N-terminal targeting signals, whereas cluster C1 shows a relatively equal number of N- and C-terminal targeting signals (Figure 3.4.3.1A). There is a very low occurrence of targeting signals detected in the atypical and extended clusters, which suggests many are located in the cytosol (Figure 3.4.3.1A).

Figure 3.4.3.1B illustrates the distribution of targeting signal types across the human SDR family. Signal anchor sequences and signal peptides appear to concentrate in clusters C3 and C2. The presence of signal anchor sequences implies these are membrane-bound/anchored proteins. Cluster C1 shows no signal anchor sequences, which correlates with the lack of organelle membrane associations. Instead, cluster C1 shows the presence of MTS (5/11) and PTS1 (4/11). Extended SDR members display ER retention (2/3) and signal anchors sequences (1/3). NES (5/53) and NLS (3/53) motifs occur in the lowest frequencies across the SDR family. Figure 3.4.3.1C shows the localisation of human SDRs with the most abundant types of targeting signals. SDRs that contain signal peptides or signal anchors have a tendency to localise in the ER.

3.5 TRANSMEMBRANE ASSOCIATIONS

Membranes form a partition for each eukaryotic subcellular compartment, which allows the exchange of small molecules (e.g. H_2O, Ca^{2+}, sugars and amino acids), send and receive signal via receptors (e.g. growth factors, hormones, neurotransmitters) and allow the translocation cargo molecules such as, polypeptides and oligonucleotides. A TM helix is one or more segments within a protein that either anchors or spans a membrane. Analysis of TM helices may link human SDRs to a particular subcellular compartment and help in elucidating its functional role. TM α-helices are usually around 20 AA in length, and hydrophobic in nature to integrate or span the hydrophobic phospholipid bilayer. Table 3 (Appendix) shows predicted TM helices from Uniprot analyses and their specific localisation for human SDRs. It is important to be cautious of some predicted TM helices, as they are incompatible with the catalytic site of SDRs. For example, the active site motif between positions 188-192 in DHRS3 is predicted as a TM helix 170-190 (Haeseleer et al. 1998). TM helices are frequently found in Cluster C3, followed by C2 and extended clusters. Cluster C1 contains no predicted TM helices. SDRs in clusters C2 and C3 often show a non-conserved hydrophobic α-helix around residues 4-24 (numbering adapted from HSD11β1) in the N-terminus. The extended members display C-terminus TM α-helices that localise to the ER, and tend to turnover either steroids or retinoids (Figure 3.5.1). Other less common TM locations occur in the SDR family, for example, some SDRs in Cluster C2 harbour TM helices in the region of 132-151 and/or 159–178 (numbering adapted from HSD17β6).

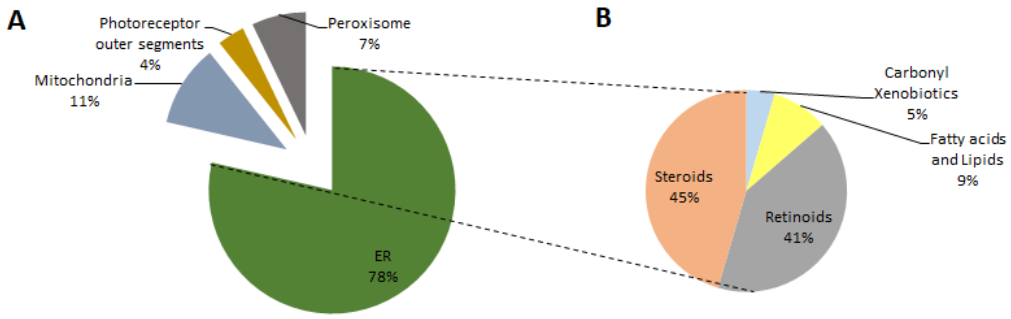


Figure 3.5.1 TM helix distribution A) subcellular localisation B) Substrate turnover for ER localised TM helices

3.6 SUBCELLULAR LOCALISATION PREDICTION

In the absence of experimental data, subcellular localisation can be inferred from an amino acid sequence to help functional decryption of orphan SDRs. Several computational prediction programs exist for this analysis, but present varying degrees of accuracy. Four selected servers, Hum-mPloc (Shen and Chou 2007), MultiLoc (Hoglund et al. 2006), CELLO (Yu et al. 2006) and WoLF PSORT (Horton et al. 2007), were chosen to compare the prediction based on the following criteria: suitability for eukaryotic systems, acceptance of multiple protein sequence batches, free availability, the ability to predict the major subcellular locations (cytoplasm, ER, extracellular region Golgi, lysosome, mitochondria, nucleus, peroxisome and plasma membrane), and distinct variation of algorithms used between the programs. Each method is evaluated for the proportion of correctly predicted subcellular location for human SDR when judged against experimental data (Table 3.6.1). It is useful to survey different programs to gauge their sensitivities, and filter predictions based on their limitations.

	Total SDRs with Experimental Subcellular Localisation	Hum-mPloc	MultiLoc	CELLO	Wolf PSORT
All Subcellular compartments	**58**	**44**	25	20	20
Cytoplasm	**15**	8	7	**12**	9
ER	**21**	**20**	9	0	3
Golgi	**2**	0	**1**	0	0
Mitochondria	**11**	7	7	7	6
Nucleus	**4**	**3**	0	1	2
Peroxisome	**5**	**5**	1	0	0

Table 3.6.1 Evaluation of subcellular prediction methods on human SDR family

Overall, subcellular location predictions by Hum-mPLoc server are the most consistent (43/58) with experimental data determined for SDRs (Table 3.6.1). The other three prediction servers tend to produce a large proportion of mismatches i.e. the predicted subcellular location does not match the experimentally determined subcellular location. SDRs localised to the cytoplasm, or mitochondria have the highest proportion of correct predictions by all four servers. Perhaps this pertains to MTS being well-defined by their peptide sequences, where as the lack of signal sequences for cytoplasmic proteins makes subcellular predictions simpler. The most accurate prediction for cytoplasmic SDRs is by CELLO (12/15). Hum-mPloc identifies 20 out of 21 of the SDRs with experimental ER localisations, while the other servers under-performed (<50 % correctly identified) for this compartment. The Golgi organelle is an under-represented subcellular compartment in the SDR family, and almost all servers poorly predict this location. Golgi proteins do not contain obvious targeting signals, are even challenging to identify experimentally, and often undergo post-translational modifications (e.g. glycosylation), which makes them difficult to identify by prediction tools. Subcellular localization programs often only consider Golgi protein with TM-helices in their algorithms, which can misidentify their localisation. Hum-

mPLoc predicts the nucleus and peroxisome localisation accurately despite the low number of SDR proteins in these compartments. Certain prediction servers commonly produce mismatches for some subcellular compartments. For example, SDRs localised to the cytoplasm are often predicted as mitochondrial despite the lack of targeting signals. CELLO, multi-Loc and WoLF PSORT frequently predict ER-localised SDRs as plasma membrane or peroxisomal. These three programs also predict SDRs located in the peroxisome to be mitochondrial. To filter out these kinds of disparities from prediction, it is important to analyse a sequence for targeting signals and TM-associations to further support accurate subcellular prediction.

3.7 DISCUSSION

The chapter has focused on the annotation of the human SDR family by investigating relationships across distribution of cofactor preferences, substrate class, subcellular localisation, targeting signals and TM helices. Cofactor analysis of the human SDR family clearly shows the overall preference for NADP/H. Genome comparisons show many other organisms also exhibit this trait. Unicellular organisms such as *Dictyostelium discoideum, E. coli* and *Haloarcula hispanica* exhibit a relatively equal preference between NAD/H and NADP/H, although their SDR genomes are considerably smaller. As organisms have evolved to become multicellular, SDR proteins tend to utilise NADP/H over NAD/H. SDRs have most likely evolved cofactor preference in accordance with their cellular compartmentalisation conditions (e.g. availability of oxidised or reduced cofactor, pH, redox environment) that can dictate reaction directionality (Agarwal and Auchus 2005, Go and Jones 2009). In oxidised conditions, such as the lumen of the ER, reductases principally use NADPH to drive

reactions. Similarly, in reduced conditions such as the cytosol, dehydrogenases largely use NAD to drive reactions. This could explain the NAD dependent activity for cytoplasmic 17β-HSD1, and the NADPH-dependent activity for 17β-HSD2 in the ER (Filling et al. 2001b). SDRs found in the cytosol, nucleus and peroxisomes favour NADP/H. The limited cytosolic free $NADP^+$/NADPH suggests a high number of protein with bound NADP/H, as seen by the preference in SDR proteins (Figure 3.2.1C). Subcellular compartments such as the mitochondria display an equal preference for either cofactor, which is in line with the NADP/H:NAD/H ratio in this organelle (Filling et al. 2001b). The preference for NADP/H over NADH also shows in some SDR substrate classes such as AA derivatives, carbonyl xenobiotics, and regulatory proteins.

The majority of SDR proteins with a regulatory function bind NADP/H. There are precedents of NADP/H binding regulatory SDRs (i.e. NMRAL1 and HSDL1) that lack the signature catalytic tetrad and are devoid of enzymatic activity. Methionine adenosyltransferase II, β (MAT2B) is an SDR member that harbours the conserved S-Y-K-N active site residues, but functions to modulate the activity of Methionine adenosyltransferase II, α (MAT2A) (González et al. 2012). It is possible at some point during evolution that some of these regulatory SDRs may have acquired catalytic activity. For example, a BLAST analysis of HSDL1, reveals HSDL1 and HSD17B12 have evolved from a common SDR member with steroid activity (Figure 3.7.1A – red). A point mutation of Phe218 to Tyr in human HSDL1 results in weak activity towards steroid and retinoid substrates (Meier et al. 2009). Interestingly, *Gallus gallus* HSDL1 contains the catalytic active site Tyr residue, but still may possess regulatory functions (Figure 3.7.1B -red). It appears that the preference of NADP/H

for these proteins seems to have been established before the regulatory functions, as demonstrated by HSDL1. Examination of cofactor preferences across the human SDR family does show differences in subcellular localisation and substrate class accommodation, which are summarised in Table 3.7.1.

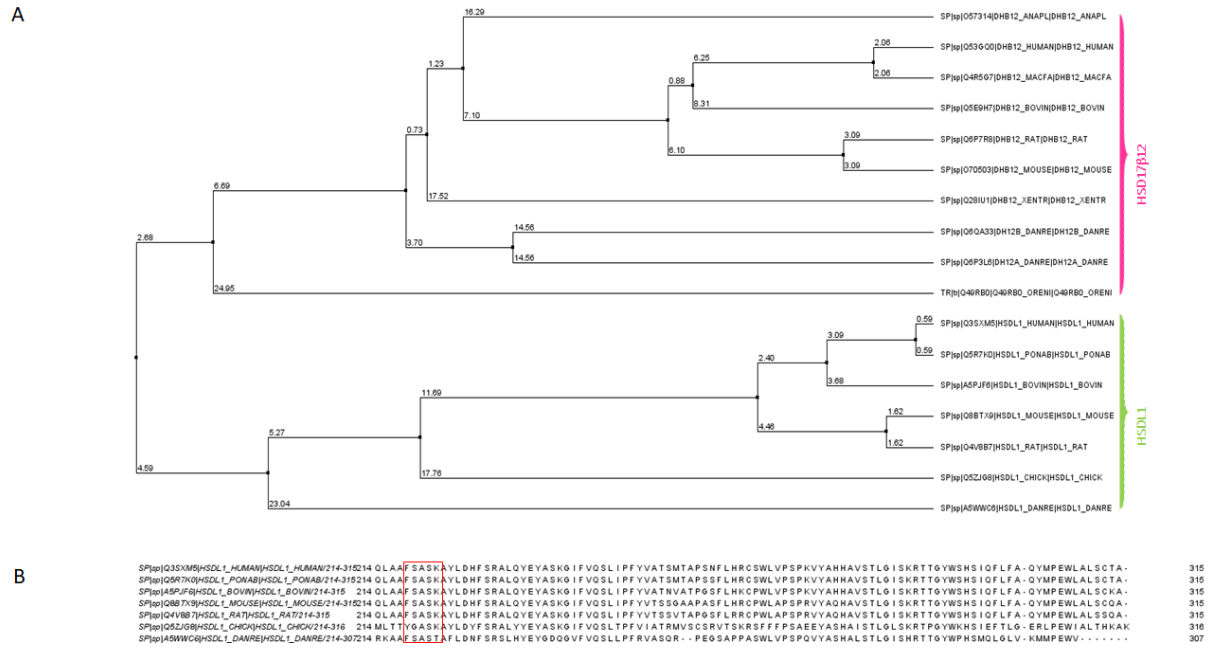


Figure 3.7.1 HSDL1 active site evolution - A) Phylogenetic tree of HSDL1 (green) and HSD17β12 (pink) B) Sequence alignment of active site of vertebrae HSDL1 (red). Phylogenetic tree created by ClustalW2 Phylogeny and sequence aligments created by Clustal Omega.

Cofactor	Subcellular Localisation	Substrate Class
NAD/H	ER, Golgi, Mitochondria	Fatty acids and lipids, Retinoids, Steroids, Sugars
NADP/H	Cytoplasm, ER, Golgi, Mitochondria, Nucleus, Peroxisome	Carbonyl xenobiotics, Fatty acids and lipids, Amino acid derivatives, Regulatory, Retinoids, Steroids, Sugars

Table 3.7.1 Human SDR cofactor preferences summary by subcellular localisation and substrate class.

Sequence cluster and substrate analysis shows there are some distinct traits in each cluster. For example, SDRs with catalytic activity towards sugars are restricted to the extended cluster. Each cluster also has a predominant substrate class preference (Table 3.7.2). For example, cluster C1 has a high concentration of SDRs that utilises carbonyl xenobiotics or fatty acids and lipids, whereas the most prevalent substrate classes in cluster C2 are retinoids and steroids. Clusters C1 and C3 harbour the most diverse range of substrate classes, which may be attributed by the larger evolutionary distances between each member as indicated by the SDR superfamily in Figure 1.4.4.2. It is important to note, as SDR orphans become functionally characterised that the distribution of substrate classes in each sequence may alter. For instance, Cluster C2 contains seven SDR orphans, which may play a significant part in altering the most prevalent form of a substrate class. However, there is only one SDR orphan in atypical and extended clusters, which will not drastically change the ratio of substrate preferences.

SDR cluster	Cofactor preference	Subcellular Localisation	Substrate class
C3	NAD/H or NADP/H	ER	Fatty acids and lipids/Retinoids/ Steroids
C2	NAD/H or NADP/H	ER	Retinoids/Steroids
C1	NAD/H or NADP/H	Cytoplasm/Mitochondria/ Peroxisome	Carbonyl xenobiotics/Fatty acids and lipids
Atypical	NADP/H	Cytoplasm	Fatty acids and lipids/Amino acid derivatives/ Regulatory
Extended	NAD/H	ER/Mitochondria	Steroids/Sugars

Table 3.7.2 SDR Cluster summary of predominant forms of cofactor preference, substrate class and subcellular localisation.

Similar to substrate analysis, subcellular localisation shows an uneven distribution across the family. No single subcellular compartment is confined to a particular sequence cluster. It

is thought that over time enzymes of a particular pathway acquire targeting signals for a new compartment independently, and selected biochemical functions would exist in new compartments (Martin 2010). The enoyl reductases in the SDR family, DECR1, DECR2 and peroxisomal trans-2-enoyl-CoA reductase (PECR), all arise from a common ancestor. DECR2 and PECR contain PTS1 signals at the C-terminus, whereas DECR1 harbours an N-terminal MTS (Alphey et al. 2005, Kim and Battaile 2001). Interestingly, the mitochondrial signal in DECR1 appears to have evolved only in vertebrates, whereas DECR2 and PECR have acquired their PTS1 signals earlier on in evolution. Although the three human enzymes perform similar functions in beta-oxidation of unsaturated fatty acids, they have evolved into separate subcellular compartments.

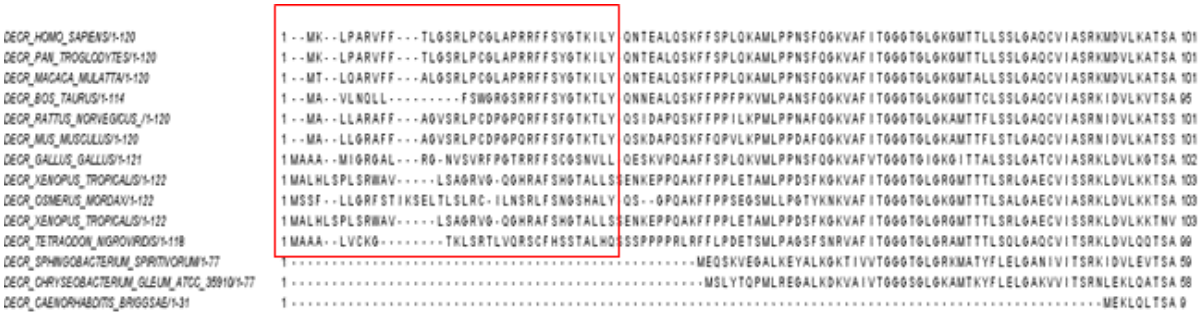


Figure 3.7.2 Sequence alignment of N-terminal of DECR1 proteins - □ denotes sequences with mitochondrial targeting signal. Sequences aligned with ClustalW2 Omega.

ER localisation dominates the SDRs found in clusters C3 and C2 is the ER (Figure 3.4.1.1). In contrast, SDRs in the C1 and atypical cluster lack signal anchors or TM helices, which means many localise in the cytoplasm. A summary of cofactor preference, subcellular localisation and substrate class accommodation in relation to each sequence cluster is given in Table 3.7.2.

Proteins perform specific functions in the environment provided by the organelle, and hence are likely to work towards similar physiological processes e.g. glycolysis in the cytoplasm, citric acid cycle in mitochondria and beta-oxidation of fatty acids in mitochondria and peroxisomes. This chapter illustrates an observation that SDRs localised to the same subcellular compartment often turnover similar types of substrates. For example, SDRs in the ER tend to turnover either retinoids or steroids. SDRs that act on amino acid derivatives exclusively locate to the cytoplasm. Subcellular compartments such as the cytoplasm and mitochondria display the most diverse range of substrate class used by SDRs. Identifying potential substrate classes using the patterns uncovered in this chapter is a step towards understanding potential functions for orphan SDRs. A summary for substrate classes based on subcellular localisation is shown in Table 3.7.3.

Subcellular Localisation	Substrate Classes Prediction
Cytoplasm	Amino acid derivatives/Carbonyl xenobiotics/Fatty acid and Lipids/Regulatory/Steroids
ER	Retinoids/ Steroids
Mitochondria	Carbonyl xenobiotics/Fatty acids and Lipids/Regulatory/Steroids
Nucleus	Regulatory
Peroxisome	Carbonyl xenobiotics/Fatty acids and Lipids

Table 3.7.3 Predominant substrate class from subcellular localisation analysis

As mentioned in Section 3.4.2, targeting signals are important sequences that direct proteins to the proper subcellular compartment. Mutations in targeting signals can often lead to a number of human genetic diseases, such as cystic fibrosis. Mutations in the TM

segment sequence of cystic fibrosis transmembrane conductance regulator (CFTR) does not allow correct orientation to the cytoplasm and the protein becomes misfolded in the ER, where it is then degraded (Skach 2000). Speech abnormalities are observed in patients with deletions in the NLS within the forkhead box (FOX)P2 gene, where the protein results in an aggregate in the cytoplasm (Mizutani et al. 2007). *Fundus albipunctatus* (stationary night blindness) is linked to mislocalisation of RDH5 from the ER to the perinuclear region as result in mutations to the N-terminal signal peptide and C-terminal TM helices (Liden et al. 2001).

Most targeting signals in the human SDR family are found in the N-terminus (Figure 3.4.3.1A). MTS1 and signal peptides are the most common type of targeting signals, whereas NES, NLS and ER retention signals are infrequent signal types across the entire human SDR family. Analysis of targeting signals for proteins without subcellular localisation is a good method to indicate potential subcellular compartments, which in turn can be used to characterise orphan SDRs (Table 3.7.3).

Targeting signals clearly provide a good indication of subcellular compartmentalisation. Computational programs were assessed as an additional method to predict subcellular localisation. From the selected prediction servers, Hum-mPLoc predictions generally corresponded to experimental subcellular localisations of human SDRs. Each server exhibited strengths and weakness on different subcellular compartments. Cytoplasmic prediction for human SDRs is only 50 % accurate for all servers with the exception of CELLO,

which predicted 80 % of the cytoplasmic localisation for human SDRs. Unlike many of the other subcellular compartments; there are no universal Golgi signal sequences to indicate translocation to the Golgi, which makes localisation prediction difficult by sequence analysis. Golgi prediction servers do exist; however, prediction is based on transmembrane domains (Yuan and Teasdale 2002), which many Golgi localised proteins do not contain. Dipeptide composition analysis is an alternative method for the prediction of Golgi proteins, although this has shown high success rates, it is lengthy to calculate, and only a large scale server for plant sequences currently exists (Chou et al. 2010). Overall, subcellular predictions for SDRs are difficult to detect by prediction programs alone. However, they provide a good indication of where a human SDR may localise in the cell when used in conjunction with other factors such as cluster locations, targeting signals and TM helices.

Application of the data in Table 3.7.1-3 can be used to narrow down substrate classes for SDRs with no functional information. A systematic flow chart (Figure 3.7.3) shows the important factors when identifying potential Substrate classes for orphan human SDRs. The first step is to identify whether an active site exists (box 1) to establish whether an SDR protein is capable of catalytic activity or may possess a regulatory function. Looking at cofactor preference and sequence cluster (boxes 2 and 3) can then narrow down potential substrate classes using Tables 3.7.1-2. The latter stages (boxes 4-6) of substrate type identification focus on pinpointing subcellular localisation, and then Table 3.7.3 can be applied to refine potential substrate types to test human SDRs for activity.

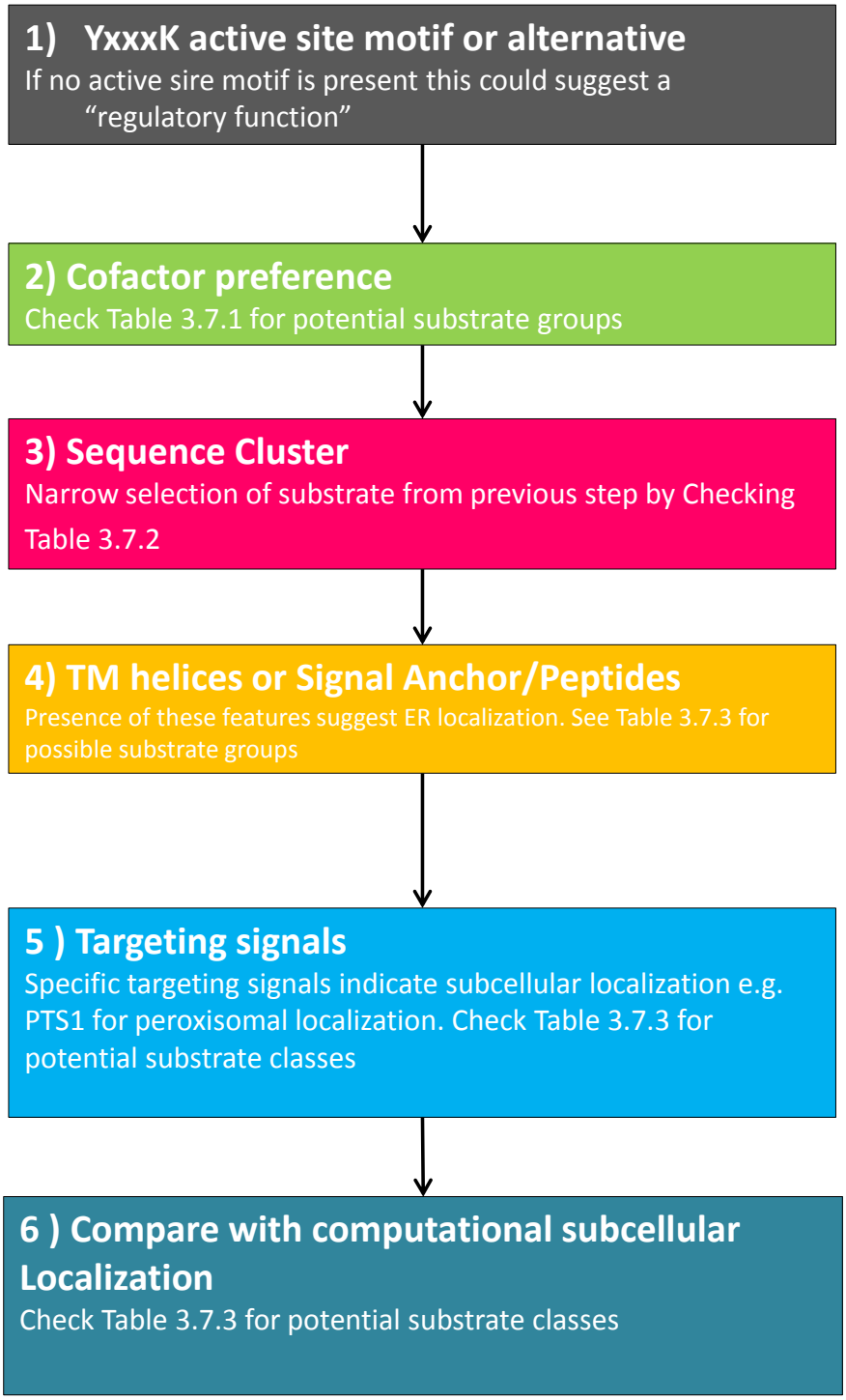


Figure 3.7.3 Flow chart used in this work to identify potential Substrate classes for human orphan SDR

This flow chart approach has been applied to establish potential substrate classes in the following examples:

(i) HSD17β13 retains an active site motif suggesting a catalytic activity. Its presence in the C3 sequence cluster, and possession of a signal peptide suggest ER as potential subcellular localisation, which is in agreement with Hum-mPLoc prediction. The ER subcellular prediction leads to the conclusion that HSD17β13 is likely to turnover steroids and/or retinoids (Table 3.7.3). Recently, HSD17β13 has been discovered to localise to lipid droplets in the liver (Abud-Husn et al. 2018, Adam et al. 2018) and turnover steroids and lipids.

(ii) DHRS1, another C3 cluster member with a classic active site fingerprint, binds NADP/H. It lacks any N- or C-terminal signal sequences, which indicates cytoplasmic localisation, and is likely to turnover fatty acid and lipids and/or steroids.

(iii) Unlike HSD17β13 and DHRS1, HSDL1 is deficient of the catalytic Tyr (or a suitable alternative). Studies have shown there is still little or no activity, even when a Tyr is mutated into the active site of HSDL1, and is therefore thought to possess a regulatory function.

(iv) Dehydrogenase/reductase (SDR family) member 7B (DHRS7B) contains an intact active site motif, and is a C3 member with a MTS signal suggesting a mitochondrial localisation. DHRS7B is therefore likely to turnover a steroid substrate.

(v) WWOX a known tumour suppressor (Del Mare and Agelian 2015, Pospeich et al. 2018, Yang and Zhang 2008), which is an unusual NADP/H dependent SDR in the

C2 cluster, owing to its multiple sites in the cell (cytoplasm, nucleus and Golgi) and lacks TM-helices. The C2 allocation suggestions a strong possibility of steroid turnover, retinoid is unlikely as these have membrane associations (Table 3.7.2). The NADP/H preference along with the cytoplasmic and nucleus localisations highlights a possible regulatory function (Table 3.7.3). Although WWOX retains an active site motif, it may be similar to MAT2B in having a regulatory role. There is data from crude protein extracts that suggest weak steroid turnover by WWOX (Saluda-Gorgul et al. 2011). Predictions of substrate class preferences for all orphan SDRs are summarised in Table 3.7.4.

SDR	YxxxK site	Cofactor preference	Cluster	Signal Sequence	TM helices	Experimental Subcellular localisation	Hum-mPloc Subcellular prediction	Manual Subcellular prediction	Substrate class Prediction
DHRS1	YES	NADP/H	C3	NONE	NONE	*n/a*	ER	Cytoplasm	Fatty acids and lipids/Steroid
HSD17β13	YES	NAD/H	C3	Signal Peptide	NONE	*n/a*	ER/cytoplasm	ER	Retinoid/Steroid
HSDL1	NO	NADP/H	C3	MTS	NONE	Mitochondria	ER	Mitochondria	Regulatory
HSD11β1L	YES	NADP/H	C3	Signal Anchor	NONE	n/a	ER	ER	Steroid
DHRS7B	YES	NADP/H	C3	MTS	YES	n/a	Mitochondria	Mitochondria	Steroid
DHRS7C	YES	NADP/H	C3	Signal Peptide ER retention	NONE	ER	ER	ER	Retinoid/Steroid
DHRS12	YES	NADP/H	C2	NONE	NONE	n/a	Cytoplasm	Cytoplasm	Steroid
WWOX	YES	NADP/H	C2	NLS	NONE	Cytoplasm/ Golgi/ Nucleus	Nucleus	Cytoplasm/ Golgi/ Nucleus	Regulatory/Steroid
DHRS13	YES	NADP/H	C2	Signal Anchor	NONE	n/a	Mitochondria	ER	Steroid/Retinoid
HSDL2	YES	NADP/H	C1	PTS1	NONE	Peroxisome	Mitochondria/ Peroxisome	Peroxisome	Carbonyl xenobiotics/Fatty acids and Lipids
DHRS11	YES	NADP/H	C1	Signal Anchor	NONE	n/a	Mitochondria	Mitochondria	Carbonyl xenobiotics/Fatty acids and lipids/Steroid
DHRS4L2	YES	NADP/H	C1	MTS NLS	NONE	n/a	Mitochondria /Peroxisome	Peroxisome/	Carbonyl xenobiotics/Fatty acids and lipids
SDR39U1	NO	NADP/H	Atypical	NES	NONE	n/a	Mitochondria	Cytoplasm	Regulatory
SDR42E1	YES	NAD	Extended	NONE	YES	n/a	ER	ER	Steroid

Table 3.7.4 Substrate prediction for human orphan SDRS

As substrates are discovered and subcellular localisation is determined, it is likely to reinforce or slightly alter the predictions made in Table 3.7.4. For example, predictions for carbonyl xenobiotics and regulatory functions are likely to alter due to the small number of characterised SDRs in these substrate classes compared to steroids and retinoids (Figure 3.3.1). At present, there is a low number of SDRs experimentally determined to localise in the peroxisome and nucleus. These two compartments are predicted accurately by Hum-mPLoc (Table 3.6.1). Subcellular prediction programs that consider targeting signals will enhance subcellular prediction, and thus substrate prediction. In general, it is harder to predict substrate classes for SDR sequences with multiple sequence signals or more than one subcellular compartments e.g. WWOX. Improved cytoplasmic prediction by computational programs such as Cell-ploc would aid substrate prediction for orphan SDRs, as Hum-mPloc only results in approximately 50 % accuracy for human SDRs.

For the first time, a cell map for all human SDRs has been created using the experimental and predicted data (Figure 3.7.4). This map is a quick visual representation of how human SDRs are distributed within a cell. The data accumulated in this chapter will be used as a basis to compare DSF generate data in chapters 4 and 5. The next chapter explores cofactor determination of human SDRs by DSF, which is compared to experimental data and prediction from sequence analysis. In addition, the substrate class analysis and orphan SDR predictions will help to evaluate DSF data accumulated in Chapter 5 for human SDR substrates.

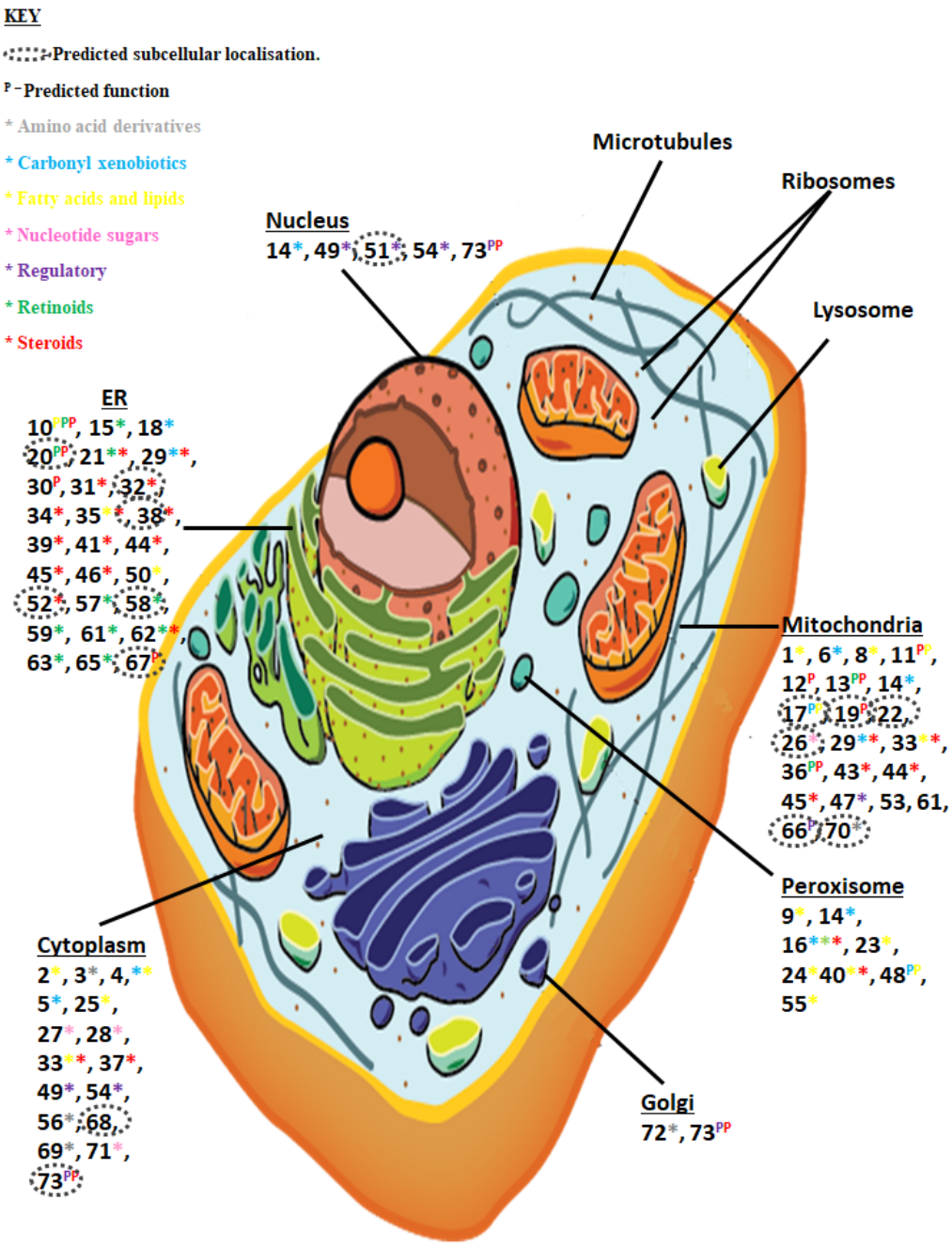


Figure 3.7.4 Human SDR family subcellular localisation and substrate accommodation diagram – 1) BDH1, 2) BDH2, 3)BLVRB, 4)CBR1, 5) CBR3, 6) CBR4, 7) DCXR, 8) DECR1, 9) DECR2, 10) DHRS1, 11) DHRS11, 12) DHRS12, 13) DHRS13, 14) DHRS2, 15) DHRS3, 16) DHRS4, 17) DHRS4L2, 18) DHRS7, 19) DHRS7B, 20) DHRS7C, 21) DHRS9, 22) DHRSX, 23) FAR1, 24) FAR2, 25) FASN, 26) GALE, 27) GMDS, 28) HPGD, 29) HSD11β1, 30) HSD11β1L, 31) HSD11β2, 32) HSD17β1, 33) HSD17β10, 34) HSD17β11, 35) HSD17β12, 36) HSD17β12, 37) HSD17β13, 38) HSD17β2, 39) HSD17β3, 40) HSD17β4, 41) HSD17β6, 42) HSD17β7, 43) HSD17β8, 44) HSD3β1, 45) HSD3β2, 46) HSD3β7, 47) HSDL1, 48) HSDL2, 49) HTATIP2, 50) KDSR, 51) MAT2B, 52) NDUFA9, 53) NMRAL1, 54) NSDHL, 55) PECR, 56) QDPR, 57) RDH10, 58) RDH11, 59) RDH12, 60) RHD13, 61) RDH14, 62) RDH16, 63) RDH5, 64) RDH8, 65) SDR16C5, 66) SDR39U1, 67) SDR42E1, 68) SDR9C7, 69) SPR, 70) TGDS, 71)TSTA3 72)UXS1 73)WWOX - Each human SDR is represented by a number and colour coded with an asterisk to indicate substrate class turnover. Predicted substrate classes are annotated with a colour coded 'P' and predicted subcellular localisations are enclosed by a dashed circle. (Bhatia et al. 2014)

CHAPTER 4 – COFACTOR DETERMINATION BY DIFFERENTIAL SCANNING FLUORIMETRY (DSF)

4.1 INTRODUCTION

The previous chapter highlights the distribution and relationships between cofactor, substrate and subcellular localisation data across the human SDR family. The following chapter investigates the determination of cofactor preference for human SDRs. Cofactor binding is critical for SDR active site formation and correct substrate orientation for catalytic activity. Structural re-arrangement upon NAD(P)(H) binding is a common phenomenon in SDRs and usually occurs between strand βF and helix αG in the Rossmann fold (Figure 4.1). For example, structural comparison of apo (Figure 4.1A and C) and holo-protomers (Figure 4.1B and D) from bacterial 3α-hydroxysteroid dehydrogenase (PDB ID 3DKN) reveals that NAD/H binding causes the formation α-helices in the otherwise disordered region 185-207, and establishes the substrate binding pocket (Nakamura et al. 2006). Structural studies of alcohol dehydrogenase from *Drosophila Lebanonesis* (DADH) also show that a small helix forms between strand βF and helix αG upon cofactor binding (Benach et al. 1999), which creates the active site cavity for substrate binding (PDB ID 1B14). Determination of cofactor preference to the active site of SDR proteins is, therefore, an integral component of understanding their functional activities and ligand interactions.

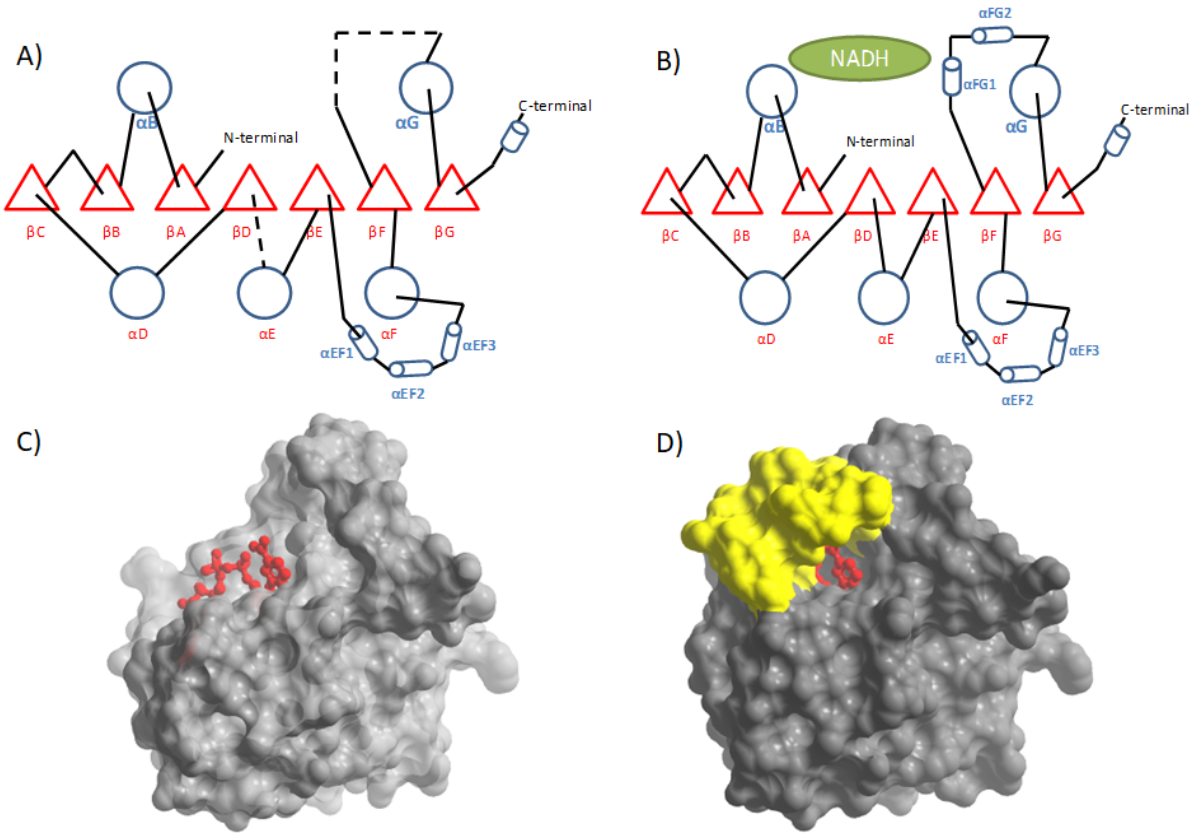


Figure 4.1 Comparison of the folding topologies between the crystal structures of apo and holo-forms of 3-hydroxysteroid dehydrogenase from *Pseudomonas sp B-0831.* Topology of the A) apo-form and *B*) holo-form - Helices are represented as *circles*, helix loop regions with *cylinders*, β-strands as *triangles*, and disordered loops as *dotted lines.* Surface representation of the C) apo form and D) holo form – the bound NADH molecule is represented in red balls and sticks. Loop 185-207, which is disordered in the apo structure, is ordered in the holo structure and highlighted in yellow. Adapted from Nakamura et al. 2006.

Traditionally, kinetic assays that measure spectrophotometric changes in substrate turnover with the addition of a cofactor is used to detect cofactor preferences for SDRs. However, enzymatic assays require an active protein with a known substrate, and achieving optimal conditions can expend more reagents and be a time-consuming process. Differential scanning fluorimetry (DSF) is an alternative for determining cofactor preferences, which does not depend on the use of a substrate or enzymatic reaction. DSF is a rapid and inexpensive approach that detects protein-ligand thermal stabilisation, independent of its enzymatic function (Niesen et al. 2007; Fedorov et al. 2012). DSF utilises a relatively small

amount of protein, and requires no addition of fluorescent tags (Section 2.9). DSF streamlines the investigation for large-scale studies of protein families by using the same concentration of protein and ligand in each experiment. This chapter describes the use of DSF to distinguish cofactor preferences of 28 human SDRs.

4.2 COFACTOR DETERMINATION BY T_M SHIFT

For a family-wide DSF screen of human SDRs, 28 members were selected based on expression, representation of each sequence cluster, biological significance or orphan status. The 28 target SDRs were expressed as recombinant proteins and purified (Chapter 2), and subsequently screened against a focused compound library comprising two parts: known cofactors (n = 48, described in this chapter) and other ligands that comprise a variety of substrates, inhibitors and their analogues (n=500, described in Chapter 5). All SDRs were screened against the cofactor library first to distinguish between NAD/H and NADP/H preference. The ligand library was subsequently screened with the cognate cofactor in two distinct sets, reduced [NAD(P)H)] and oxidised [NAD(P)] forms (Chapter 5). SDRs that show no distinct thermal shifts for either NAD/H or NADP/H are determined as 'ambiguous', and these proteins were subsequently screened against the ligand library using the cofactor information sourced from experimental data in literature or bioinformatics predictions. Table 4.4.1 shows the data for 28 human SDRs screened for cofactor preference. Among them, 22 SDRs have had cofactor preferences determined experimentally by other methods than DSF prior to this work.

4.3 COFACTOR PREDICTION BY SEQUENCE ANALYSIS

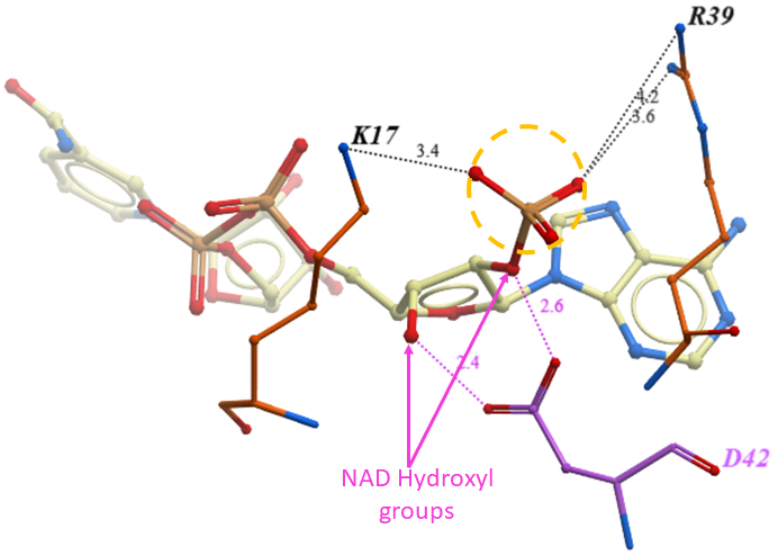


Figure 4.3.1 NAD vs. NADP residue binding – NAD hydroxyl groups (purple arrows) interaction with HSD17β8 Asp42 (purple carbon atoms), NADP phosphate (orange circle) interaction with DCXR Lys17 and Arg39 (orange carbon atoms). NADP and amino acids in ball and stick representation: nitrogen (blue); oxygen (red); phosphate (orange) carbon atoms in the cofactor (cream).

Computational determination of SDR cofactor preference can be predicted from analysing key residues in/around the active site. Previous studies (Kallberg et al. 2002b) show that NAD/H preferring SDRs contain an acidic residue (Asp or Glu) on strand β2 (about 20 residues downstream of the cofactor binding motif, Gly-X-X-X-Gly-X-Gly) that interacts with hydroxyl groups of the cofactor adenine ribose (Figure 4.3.1 – pink). NADP/H preferring enzymes contain a basic residue (Arg or Lys) just after strand β2 that interacts with the additional phosphate group of NADP/H (Figure 4.4.1 – orange circle). For example, the key residue for NADP/H preference in dicarbonyl/L-xylulose reductase (DCXR) is Arg39 located after strand β2 (PDB ID 1WNT; Lima et al. 2006). Some NADP/H preferring enzymes also contain a basic residue in the cofactor-binding motif before the second Gly (Kallberg et al. 2002a). Secondary structure prediction of each target SDR was generated by Jpred server

(Cole et al. 2008), and then analysed for key residues to predict cofactor preference (summarised in Table 4.4.1).

4.4 COFACTOR DETERMINATION RESULTS

SDR	Literature	Cofactor preference sequence prediction (this study)	Key Residues	Cofactor preference DSF determination (this study)	T_m Shift °C	
					NAD/H	NADP/H
BDH2	NAD/H	NAD/H	D37	Ambiguous	b.t/b.t	b.t/b.t
CBR1	NADP/H	NADP/H	R38	NADP/H	b.t/1.6	6.6/5.8
CBR3	NADP/H	NADP/H	R38	NADP/H	b.t/b.t	1.8/2.9
DCXR	NADP/H	NADP/H	K11 and R39	NADP/H	b.t/b.t	5.4/9.6
DHRS1	No Data	NADP/H	R39	NADP/H	b.t/b.t	2.4/2.0
DHRS11	No Data	NADP/H	R43	NADP/H	b.t/b.t	5.7/12.4
DHRS2B	NADP/H	NADP/H	R68 and K69	NADP/H	b.t/b.t	2.4/3.9
DHRS4	NADP/H	NADP/H	R64	NADP/H	b.t/b.t	b.t/b.t
DHRS7B	No Data	NADP/H	R84	Ambiguous	b.t/b.t	b.t/b.t
GALE	NAD/H	NAD/H	D33	Ambiguous	b.t/b.t	b.t/b.t
GMDS	NADP/H	NADP/H	R55 and R56	NADP/H	b.t/b.t	b.t/14.7
HPGD	NAD/H	NAD/H	D36	NAD/H	3.5/13.7	b.t/b.t
HSD11B1	NADP/H	NADP/H	R66	NADP/H	b.t/b.t	3.1/b.t
HSD17B10	NAD/H	NAD/H	D41	NAD/H	1.9/13.1	b.t/b.t
HSD17B12	NADP/H	NADP/H	R82	Ambiguous	b.t/b.t	b.t/b.t
HSD17B4	NAD/H	NAD/H	D40	Ambiguous	b.t/b.t	b.t/b.t
HSD17B8	NAD/H	NAD/H	D42	NAD/H	b.t/1.6	b.t/b.t
HSDL2	No Data	NADP/H	K42	NADP/H	b.t/b.t	b.t/b.t
HTATIP2	NADP/H	NADP/H	R86,R87 and K88	NADP/H	b.t/b.t	1.3/4.1
MAT2B	NADP/H	NADP/H	R61 and R62	NADP/H	b.t/b.t	4.5/9.7
NMRAL1	NADP/H	NADP/H	R37	NADP/H	b.t/b.t	7.6/8.6
PECR	NADP/H	NADP/H	R50 and K51	NADP/H	b.t/b.t	4.3/8.6
QDPR	NAD/H	NAD/H	D41	NAD/H	1.3/7.5	b.t/b.t
SDR39U1	No Data	NADP/H	R32 and K33	NADP/H	1.6/b.t	-/5.3
SPR	NADP/H	NADP/H	R42	NADP/H	b.t/b.t	2.4/3.3
TSTA3	NADP/H	NAD/H	D37	NADP/H	1.5/b.t	6.4/9.8
UXS1	NAD/H	NAD/H	D119	NAD/H	4.6/6.1	b.t/b.t
WWOX	No Data	NADP/H	R196	NADP/H	b.t/b.t	1.8/b.t.

Table 4.4.1 Cofactor determination summary and comparison T_m Shift data to other data sources. NAD/H (green), NADP/H (purple), 'No data' = no experimental data. Ambiguous = further screening required for determination of cofactor. b.t = below threshold, red = cofactor determined by unfolding gradient (section 4.5)

The thermal shift data collected for the 28 SDRs shows that 23 of them show a distinct preference for either NAD/H or NADP/H, and 5 of them show no clear preference for either ('ambiguous'). Of these 23 SDRs, 18 have their cofactor preference reported in literature, and all correspond correctly with cofactor determined by DSF. In comparison, there are 21 cases in where the sequence prediction agrees with data from literature, and one instance (TSTA3) where there is disagreement.

From the study set of 28 SDRs, 6 previously uncharacterised proteins (DHRS1, HSDL2, MAT2B, DHRS11, SDR39U1 and WWOX) have had their cofactor preference determined for the first time (Table 4.4.1). Structures for DHRS1 (PDB ID: 2QQ5), DHRS11 (PDB ID: 1XG5), MAT2B (PDB ID 2YDX), SDR39U1 (PDB ID 4B4O), HSDL2 (PDB ID: 3KVO, section 6.4) and DHRS4 (PDB ID 3O4R, section 6.2) have all been subsequently determined at the SGC in the presence of their respective cofactors determined here. Among these, the preferred cofactor for DHRS11, HSDL2, MAT2B, SDR39U1 and DHRS4 can be observed in the electron density, which supports results obtained from cofactor determination by DSF.

4.5 COFACTOR DETERMINATION BY UNFOLDING GRADIENT (ΔI/Δ°C)

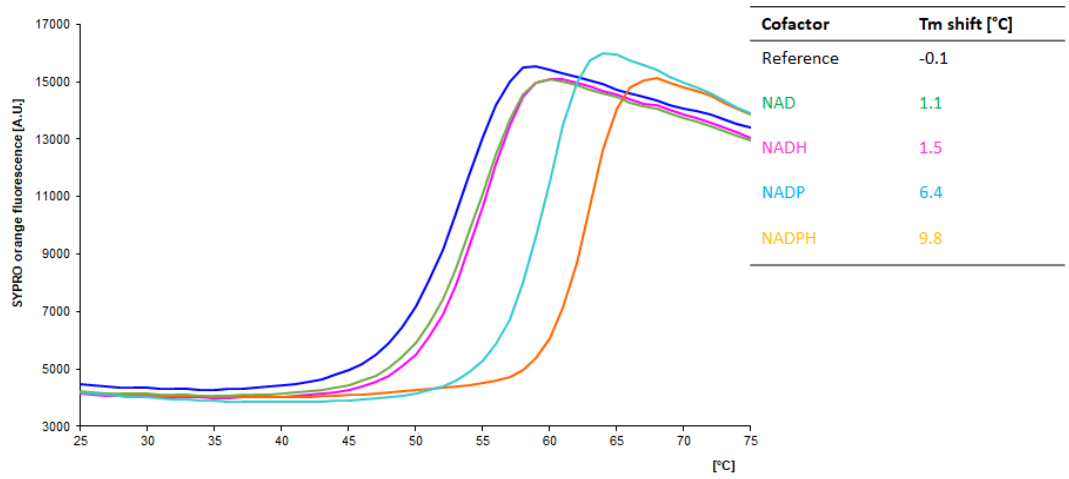


Figure 4.5.1 TSTA3 DSF curves and cofactor determination by T_m - reference line (blue), NAD (green), NADH (magenta), NADP (cyan) and NADPH (orange).

For the majority of SDRs screened, good comparable transition curves were used for midpoint calculations, as in the example for TSTA3 (Figure 4.5.1). For the less stable SDR proteins isolated from recombinant systems, initial screening can show dissimilar curve profiles for cofactors and T_m shift values do not represent the best indicators of a preferred cofactor. Using the unfolding gradient (first derivative) as an additional method to deterimine T_m Values from DSF data, which has been often used in DSC.

For example, SDR36U1 produces irregular plots, in that each DSF curve has a different height and slope, including the reference (Figure 4.5.2A). Notably, the addition of NADP creates a more characteristic unfolding curve of a stable protein with a lower starting fluorescence. The unfolding gradient of SDR39U1 shows NADP (16.9 ΔI/Δ°C) to have the highest value indicating it is also the steepest curve (Figure 4.5.2A). Similarly, DHRS4 does

not exhibit a significant increase in T_m with the addition of the cofactors (Figure 4.5.2B). The addition of NADPH causes the formation of a considerably steeper unfolding transition (8.1 ΔI/Δ°C) compared to shallow transition of the other cofactors (Figure 4.5.2B). HSDL2 (Figure 4.5.2C) shows a 1.2 °C shift with NAD and NADH, however the initial fluorescence is significantly lower and unfolding gradients are higher with NADP (24.1 ΔI/Δ°C) and NADPH (19.7 ΔI/Δ°C). The steeper slope in the data indicates a better folded native protein and, therefore NADP is concluded to be the preferred cofactor. DSF cofactor determination in all three cases agrees with sequence predictions (Table 4.4.1). HSDL2 (Section 6.4) and DHRS4 (Section 6.2) have been crystallised in the presence of their DSFdetermined cofactor and are bound in the active site (See Chapter 6).

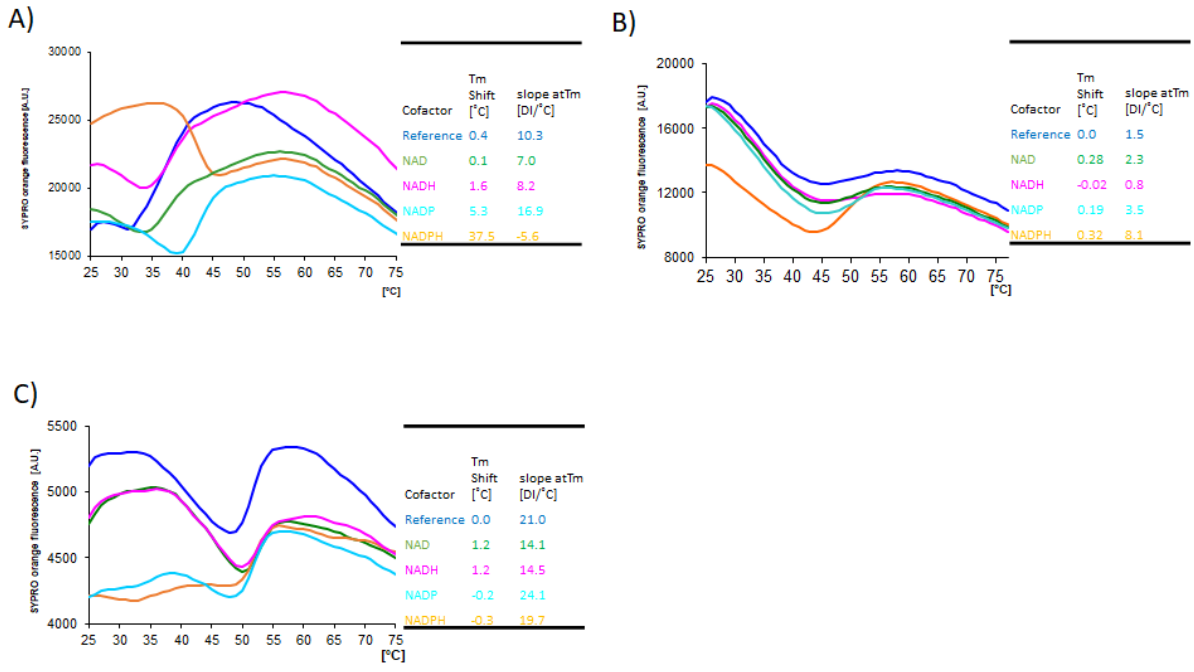


Figure 4.5.2 Human SDR cofactor determination by DSF slope of transition data A) SDR39U1 B) DHRS4 C) HSDL2

4.6 DISCUSSION

In SDRs, cofactor binding is essential for conformational rearrangement for substrate binding (Küssau at al. 2008, Nakamura et al. 2006). Cofactor determination by spectrophotometric methods usually relies on the knowledge of working substrates for a given SDR. In this chapter, the successful application of DSF to determine cofactor preference for any given SDR show the feasibility of this method for this application. Section 4.4 demonstrates that most SDRs shows a distinct preference for either NAD/H or NADP/H, which coincides with data reported in literature and crystal structures. This implies that DSF is a reliable alternative method for determining SDR cofactor preference. Regulatory (e.g. NMRAL1, MAT2B) and orphan SDRs (e.g. HSDL2, WWOX, SDR39U1) that lack activity or functional knowledge, respectively, escape conventional kinetic methods to determine cofactor preference. For the first time, the decryption of cofactor preference for 6 orphan SDRs was achieved and was verified by crystal structures produced within the group at SGC. It is clear that for the majority of cases, cofactor binding creates a more thermally stable SDR protein, which correlates with the formation of more ordered helices in the Rossmann fold.

Although DSF correctly identifies TSTA3 as NADP/H preferring, it is the only SDR that evades prediction by sequence analysis (Table 4.4.1). Kinetic data (Tonetti et al. 1996), and 3D structural data (PDB 4E5Y) show that NADP/H is the necessary cofactor for activity. As mentioned in Section 1.4, TSTA3 is a member of the extended sequence cluster, which presents a slight variation from the classical Gly-motif, Gly-X-X-Gly-X-X-Gly (Kallberg et al. 2002). Jpred secondary structure prediction generates a map of where the potential α-

helices and β-sheets exist (Figure 4.6.1A). The key residue from this analysis would be Glu36, as it is positioned on strand β2, and no basic residue directly after or within the Gly-rich motif, suggesting that TSTA3 is a NAD/H preferring enzyme. The 3D-structure of TSTA3 shows that β2 resides between residue 37-40 and the basic residue Lys44 determines NADP/H preference (Figure 4.6.1B-C). Although bioinformatics analyses provide a reliable way to predict cofactor preference for many SDRs, a few SDRs escape correct prediction. This is due to the lack of precision in secondary structure prediction, and ultimately requires experimental validation.

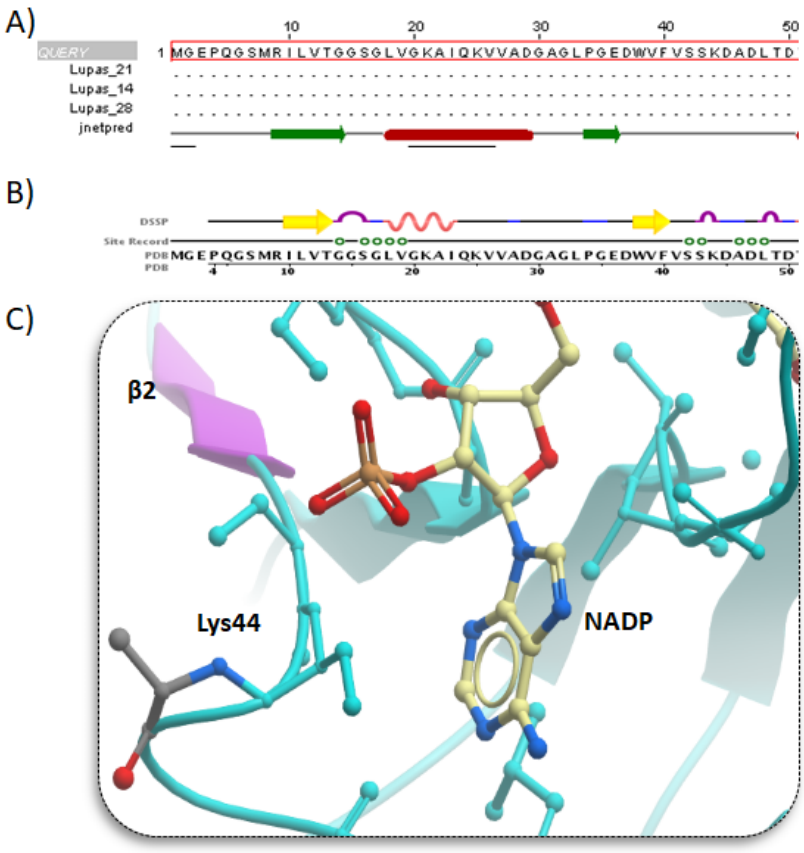


Figure 4.6.1 TSTA3 cofactor accomodation comparison of Jpred and crystal structure – A) Jpred secondary structure prediction – α-helices (red barrels), β-sheet (green arrows) B) PDB ID 4B8W - α-helices (pink), β-sheet (yellow arrows) C) PDB ID 4B8W – crystal structure shows NADP accommodation by Lys44 after B2 strand (purple).

In this chapter the application of the unfolding gradient as a secondary DSF calculation to determine cofactor preference was examined. Proteins that are less stable in solution tend to produce irregular, or variable curves are difficult to determine by T_m shift alone. The cofactor preference for such SDRs is assigned by the highest unfolding gradient (ΔI/Δ°C). DHRS4 (section 6.2), HSDL2 (section 6.4) and SDR39U1 have all been crystallised in the presence of their cofactor, as determined by the unfolding gradient. This is the first reported application of unfolding gradient values (ΔI/Δ°C) used to determine cofactor preference when thermal shift values are ambiguous or inconclusive.

Five of the 28 SDRs screened were determined as ambiguous, and could not be distinguished by unfolding gradient values as they presented curves that all resembled the reference. In these cases the cofactor preference was ascertained by using either the literature or sequence prediction. Perhaps these particular SDRs have been saturated with cofactor during expression, and therefore, present little or no thermal stabilisation when screened with additional cofactor. For example, the human NMRAL1 structure shows NADP bound when no additional cofactor was added to the crude prep prior to crystallisation (Dai et al. 2006).

There is considerable interest in exploring protein stabilisation, particularly for use in downstream processing such as crystallisation (Deller et al. 2006). The use of DSF to assess cofactor preference of SDRs allows the potential of improving crystallisation behavior, and enabling substrate screening for orphan SDRs by kinetic means (Chapter 6). This study of cofactor determination by DSF is the first step to conduct studies on further cofactor dependencies and ligand binding (Chapter 5 and 6).

CHAPTER 5 - LIGAND IDENTIFICATION BY DSF

5.1 INTRODUCTION

The preceding chapter showed the successful application of DSF to determine cofactor specificity for human SDRs. Using the identified cofactors for each target SDR in Chapter 4, this chapter explores ligand profiling across the human SDR family by using DSF. Ligand profiling of a diverse functional family as the SDRs against a focused compound library can help identify ligand interactions and patterns among different chemical classes (Niesen et al, 2010). A focus of ligand profiling for human SDRs by DSF is to assess the ability to assign substrates and inhibitors for the proteins. Large-scale inhibitor studies on other protein families, such as protein kinases, have already shown the successful use of DSF in this application (Fedorov et al. 2012). Human SDRs have been screened against a focused SDR library of about 500 compounds comprised of cofactors, substrates, inhibitors and analogues thereof.

5.2 LIGAND LIBRARY COMPOSITION

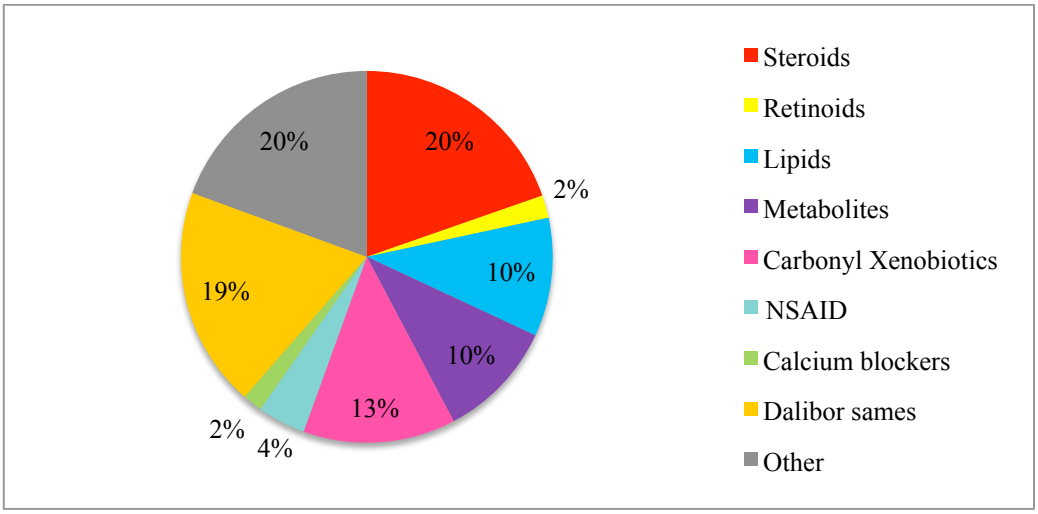


Figure 5.2.1 SDR ligand library composition for DSF - Calcium channel blockers Carbonyl xenobiotics Dalibor Sames Lipids Metabolites NSAIDs Retinoids Steroids Other

The SDR ligand library (Figure 5.2.1) comprises various types of compounds, including steroids, retinoids, non-steroidal anti-inflammatory drugs (NSAIDs) and pharmacological compounds such as calcium blockers. Many compounds were sourced from literature and analogues were searched in MolCart (https://www.molsoft.com/molcart-compounds.html). The range of compounds encompasses a diverse range of known substrates and inhibitors for a variety of SDRs. A collection of potential steroid analogues was also provided by our collaborators from the Dalibor Sames group, Columbia University (http://www.columbia.edu/cu/chemistry/groups/sames/).

5.3 SUBSTRATE IDENTIFICATION BY DSF

5.3.1 SUBSTRATE IDENTIFICATION OF SDRS WITH KNOWN SUBSTRATE BY DSF

Typically large-scale, systematic DSF studies have focussed on inhibitor profiling for one or more closely related proteins (Fedorov et al. 2012, Wahlberg et al. 2012). Substrate detection studies by DSF have received less attention, and are generally carried out on non-oxidoreductase enzymes, such as transporters (Boucher and Noll 2011). In this study, detection of specific substrates by DSF for human SDRs has been evaluated by looking at T_m shifts ≥ 1.4 °C, as more false positives occur below this threshold. Compounds in the DSF ligand library that are known substrates for human SDRs are summarised with their corresponding T_m shift in Table 4, Appendix. In this study, DSF was only able to identify 14.1 % (10/71) of all the possible SDR substrate reactions. The low detection rate is attributed to enzymatic turnover of the substrate in the reaction mixture, which does not allow stable complexes to form. For example, Dehydrogenase/reductase (SDR family) member 2B (DHRS2B) displays no T_m shifts with its dicarbonyl compounds, but shows 1.4 ^{o}C T_m shift with NADP and Palmitoleyl-CoA (Table 5.3.1.1).

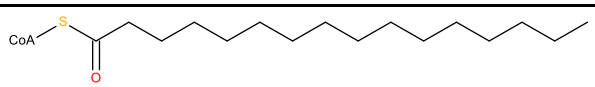
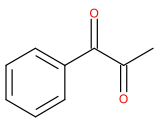
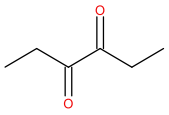
Compound	T_m °C NADP/H	Structure
Palmitoleyl-CoA	1.4/b.t	
1-Phenyl-1,2-propanedione	b.t/b.t	
3,4-Hexanedione	b.t/b.t	

Table 5.3.1.1 DHRS2B T_m shift summary b.t= below threshold

Only 6 SDRs (Appendix, Table 4) achieve T_m shifts over 1.4 °C with one or more of their specific substrates. These examples will be described in the following paragraphs. Uridine disphosphate (UDP)-glucose 4-epimerase (GALE) performs the last step in the Leloir pathway in galactose metabolism, and catalyses the NAD/H dependent reversible conversion of UDP-galactose to UDP-glucose (Wohlers et al. 1999, Timson 2005). GALE does not undergo standard EC 1.1 oxidoreductase reactions; instead it catalyses the stereochemical inversion of an asymmetric carbon in its substrates. Both UDP-galactose and UDP-glucose achieve T_m shifts > 1.5 °C with NAD. Human GALE isoforms also catalyse the formation of UDP-*N*-acetylgalactosamine (UDP-GalNAc) from UDP-*N*-acetylglucosamine (UDP-GlcNAc) (Schulz et al. 2004). Only UDP-GalNAc attains an increase in thermal stability with NADH (2.7 °C). Similarly, UDP-glucuronate and TDP (thiamine diphosphate) that are not substrates for GALE generates a 1.4 °C shift with NADH, implying the nucleoside moiety common to these ligands contributes to the thermal stability (Table 5.3.1.2). Unfortunately UDP was not tested, as it was not included in the initial design of the screening library.

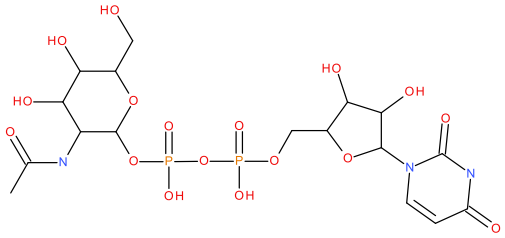
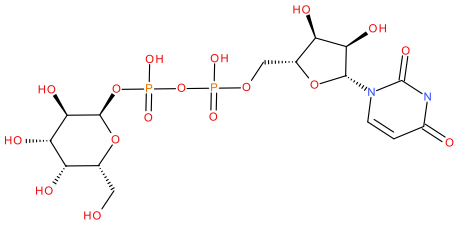
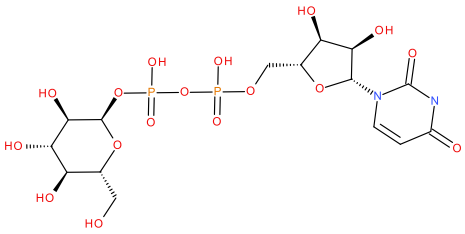
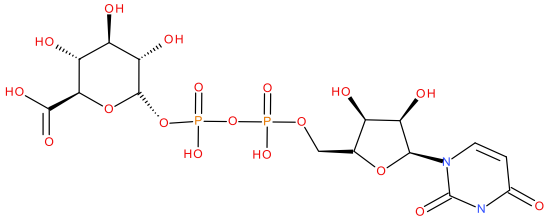
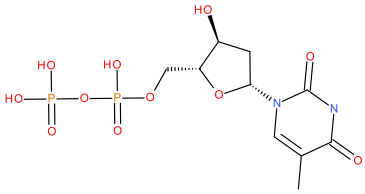
Compound	T_m °C NAD/H	Structure
UDP-α-D-GalNac	b.t/2.7	
UDP-glucose	b.t/1.7	
UDP-galactose	b.t/1.6	
UDP-glucuronate	b.t/1.4	
TDP	b.t/1.4	

Table 5.3.1.2 GALE T_m shift summary with substrates and analogues b.t= below threshold *compound names in red denote substrates.

Another extended SDR, UDP-glucuronate decarboxylase 1 (UXS1), is a decarboxylase that catalyses the formation of UDP-xylose from UDP-glucuronate using NAD/H (Bakker et al. 2009). UXS1 produces a shift of 7.9 °C with its substrate UDP-glucuronate in the presence of NADH. Most T_m shifts for UXS1 are NADH dependent, and resemble the substrate e.g UDP-glucose (2.0 °C) and UDP-galactose (1.9 °C). Interestingly, TDP produces the highest T_m shift

(12.3 °C) with NADH, and the only compound to shift with both cofactors (NAD 1.8 °C)(Table 5.3.1.3). No increase in thermal stability is observed in the presence of TDP-glucose, confirming that UXS1 is more selective towards uridine.

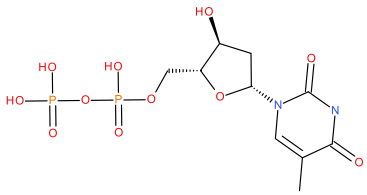
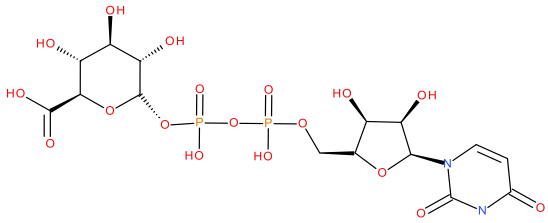
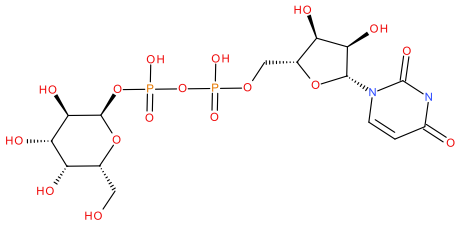
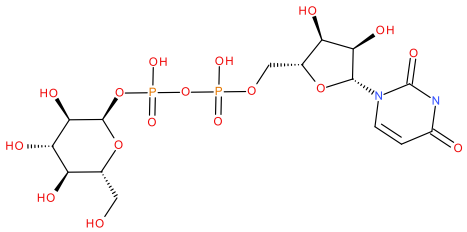
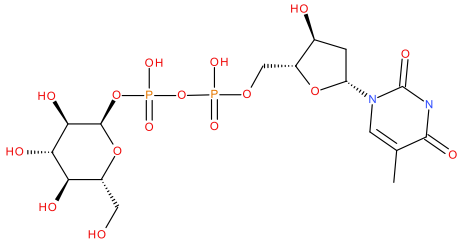
Compound	T_m shift °C NAD/H	Structure
TDP	1.8/12.3	
UDP-glucuronate	b.t/7.9	
UDP-Glucose	b.t/2.0	
UDP-Galactose	b.t/1.9	
TDP-Glucose	b.t/b.t	

Table 5.3.1.3 USX1 nucleotide sugars T_m shifts b.t= below threshold

PECR (peroxisomal enoyl-CoA reductase) a non-EC 1.1 SDR, shows T_m shifts with its substrate. PECR specifically acts on unsaturated carbon bonds in CoAs converting them into their constituent trans-enoyl-CoA in the presence of NADP (Das et al. 2000). PECR generates T_m shifts with most acyl-CoA compounds; in particular the medium-chain CoAs (decanoyl-CoA, lauroyl-CoA and hexanoyl-CoA) produce the highest NADP-dependent T_m shifts over 3.5 °C. The presence of NADPH results in lower T_m shifts, except for palmitoleyl-CoA (3.0 °C)(Table 5.3.1.4).

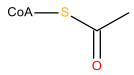
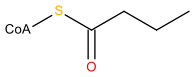
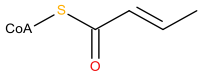
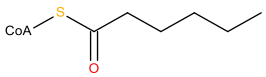
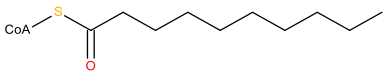
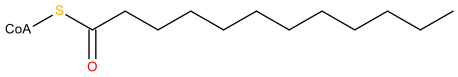
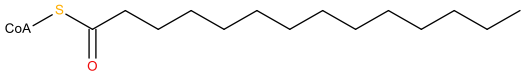
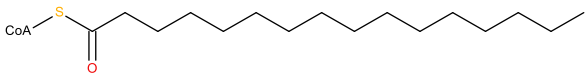
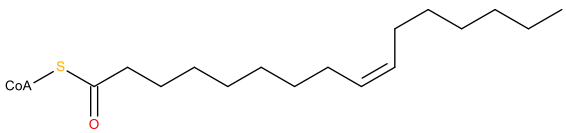
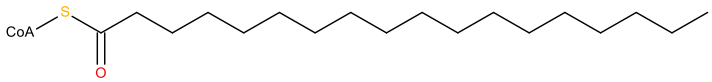
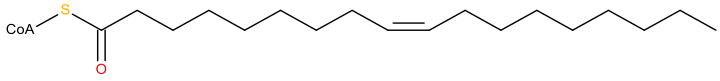
Substrate	Structure	Carbon chain length	C=C Bond	T_m Shift °C NADP/H
Acetyl-CoA		2	No	b.t/b.t
Butyryl-CoA		4	No	b.t/b.t
Crotonyl-CoA		4	Yes	b.t/b.t
Hexanoyl-CoA		6	No	3.5/b.t
Decanoyl-CoA		10	No	5.7/2.5
Lauroyl-CoA		12	No	5.1/1.1
Myristoyl-CoA		14	No	b.t/b.t
Palmitoyl-CoA		16	No	b.t/b.t
Palmitoleyl-CoA		16	Yes	1.1/3.0
Stearoyl-CoA		18	No	b.t/b.t
Oleoyl-CoA		18	Yes	b.t/b.t

Table 5.3.1.4 PECR CoA substrate T_m shift summary b.t= below threshold

A few SDRs that perform the standard EC 1.1 oxidoreductase reaction also attain T_m shifts over 1.4 °C with substrates. For instance, HPGD is the key NAD-dependent prostaglandin, leukotriene and lipoxin degrading enzyme, and it catalyses the oxidation of prostaglandin E2 (PGE2) to 11α-OH-9,15-dioxoprost-13-enoate (15-keto-PGE2) (Tai et al. 2002). Table 5.3.1.4 shows that HPGD produces NAD-dependent T_m shifts with PGE2 (5.1 °C) and prostaglandin A2 (PGA2) (7.5 °C). The carbonyl group at position '9' and double bond at position '5' are key factors in HPGD thermal stabilisation, as the removal of these groups causes a significant decrease in T_m shift (Table 5.3.1.5).

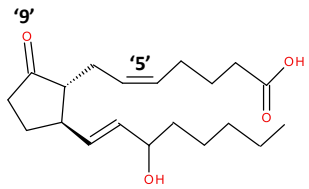
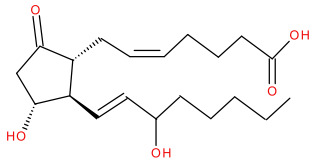
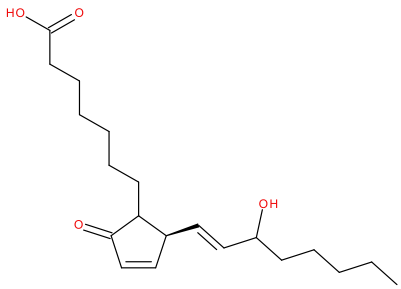
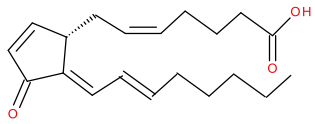
Compound	T_m shift °C NAD/H	Structure
PGA2	7.5/b.t	
PGE2	5.1/b.t	
PGA1	b.t/b.t	
15-Deoxy-delta(12,14)-prostaglandin J2	b.t/b.t	

Table 5.3.1.5 HPGD prostaglandin T_m shifts b.t= below threshold

HSD17β10 (17β-hydroxysteroid dehydrogenase 10 or Hydroxyacyl-CoA dehydrogenase 2) is a multifunctional SDR that performs dehydrogenase activity towards an assortment of substrates, such as 3-hydroxyacyl-CoA, acetoacetyl-CoA, steroids and bile acids (Ofman et al. 2003; Shafqat et al. 2003). HSD17β10 also operates as part of a complex wth MRPP1 and/or MRPP3 for posttranslational modifications (methylation and 5' end processing, respectively) of mitochondrial precursor tRNA (Oreum et al. 2017, Oreum et al. 2018).

The HSD17β10 active site accommodates a variety of molecules by allowing different orientations in relation to the active-site residues Ser155 and Tyr168 and NAD/H (Shafqat et al. 2003). The HSD17β10 substrates, dihydroandrosterone (5.4 °C), estradiol (3.0 °C), 5α-Androstane,3β,17β-diol (1.9 °C), and testosterone (1.5 °C) produce NAD dependent T_m shifts (Table 5.3.1.6). Interestingly, the ketone equivalents (androsterone, estrone, and androstenedione) do not exhibit any thermal stabilisation with NAD or NADH. Other HSD17β10 substrates which do not show increased thermal stability, such as cortisone and chenodeoxycholic acid, contain other substitutions for the hydroxyl group at position '17'. Therefore, it is the '17'-OH group on the steroid substrate that appears to play a key part in the thermal stabilisation of HSD17β10. A reduction in thermal stability occurs when the hydroxyl group '3α' in dihydroandrosterone changes to the '3β' position in 5α-androstane, 3β,17β-diol (Table 5.3.1.6).

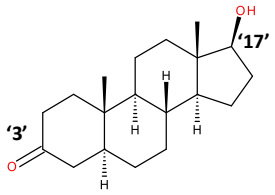
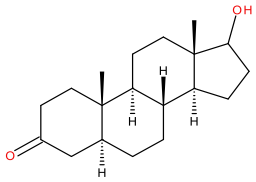
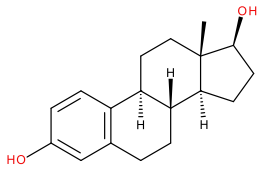
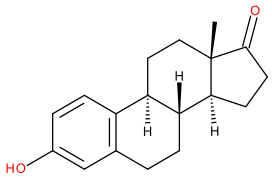
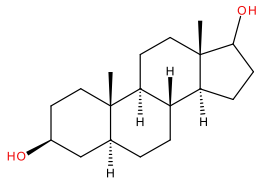
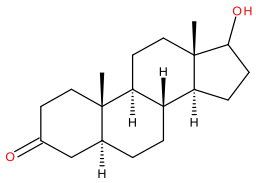
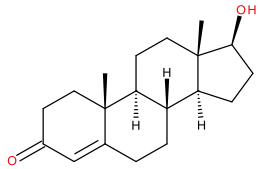
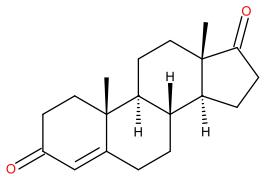
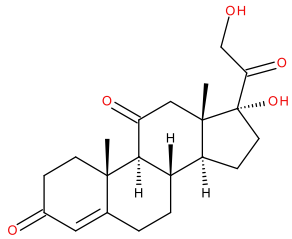
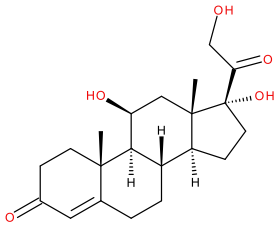
Steroid (-OH)	T_m shift °C NAD/H	Steroids (=O)	T_m shift °C NAD/H
 **Dihydroandrosterone** **(5α-Androstane, 3α,17β-diol)**	5.4/b.t	 5α-DHT	b.t/b.t
 **Estradiol**	3.0/b.t	 Estrone	b.t/b.t
 **5α-Androstane, 3β,17β-diol**	1.9/b.t	 5α-DHT	-
 **Testosterone**	1.5/b.t	 Androstenedione	b.t/b.t
 **Cortisone**	b.t/b.t	 Cortisol	b.t/b.t

Table 5.3.1.6 HSD17β10 steroid T_m shifts – b.t = below threshold

5.3.2 SUBSTRATE CLASS IDENTIFICATION BY DSF

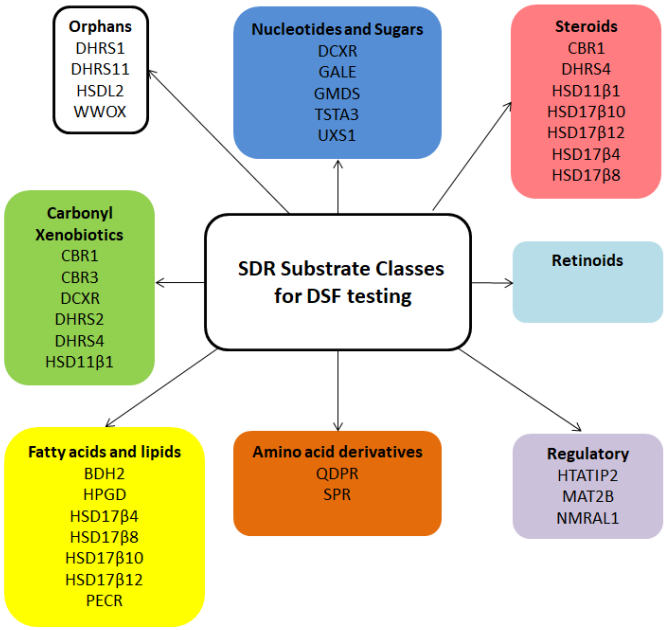


Figure 5.3.2.1 SDR Substrate classes for DSF screening

Section 5.3.1 has shown that most SDRs do not display increased thermal stability in the presence of their substrate. However, it is of note that T_m shifts with substrate-like compounds can be detected for SDRs. To this end, all screened SDRs with known enzymatic activity have been analysed (Table 5, Appendix) to see if they also exhibit T_m shifts with non-substrate compounds in their substrate class (Figure 5.3.2.1). Approximately 70 % of the human SDRs with known enzymatic activity tested in DSF showed at least one or more T_m shifts (≥ 1.4 °C) with a compound within the same substrate class (Figure 5.3.2.2). Figure 5.3.2.2 also shows a high detection rate (>65 %) of T_m shifts for SDRs utilising a particular substrate class. It is difficult to deduce individual substrates for human SDRs by DSF; however, identifying substrate classes is readily accomplished, and could form the basis for further enzymatic studies.

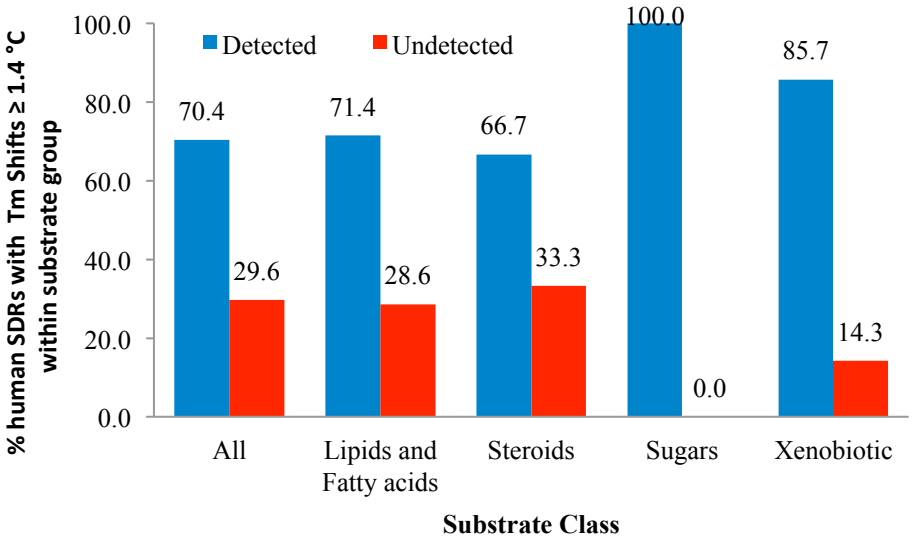


Figure 5.3.2.2 Substrate class identification by DSF

5.3.3 DETECTING SUBSTRATE SCAFFOLDS

Some SDRs produce T_m shifts within their substrate class, and display scaffold patterns that determine features for thermal stability. For example, the extended SDR member GMDS (GDP-mannose 4,6-dehydratase) catalyses the NADP-dependent conversion of GDP-mannose to GDP-4-keto-6-deoxymannose. This is the first step in the synthesis of GDP-fucose from GDP-mannose (Sullivan et al. 1998). Guanosine nucleotides present shifts with both NADP and NADPH, with the latter always higher. GMP (1.5/1.8 °C, NADP/H) is the nucleotide with the lowest shift, whereas GMP-PNP (5′ -Guanylyl-imidodiphosphate) (7.1/9.9 °C, NADP/H) is the highest. GMP-PNP is a GTP analogue, where an oxygen on the phosphate moiety is replaced with an amine, thus rendering it non-hydrolyzable and would explain the high thermal shift. GDP (3.0/8.1 °C, NADP/H) also produces a higher shift than GMP. ATPγS is also a purine based nucleotide that possesses an amine group at position c6 instead of a carbonyl group, and is the sole non-guanosine nucleotide to shift with only NADPH (2.9 °C).

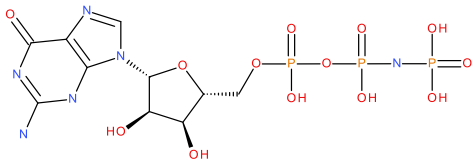
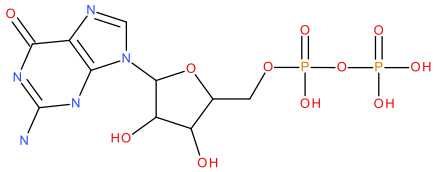
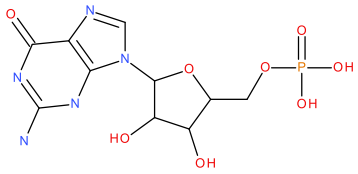
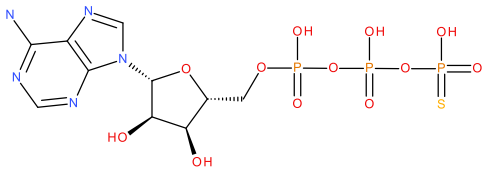
Compound	T_m Shift °C NADP/H	Structure
GMP-PNP	7.1/9.9	
GDP	3.0/8.1	
GMP	1.5/1.8	
ATPγS	b.t/2.9	

Table 5.3.3.1 GMDS nucleotide T_m shifts b.t= below threshold

Hydroxysteroid 17β dehydrogenase 8 (HSD17β8) is a NAD-dependent SDR that forms a heterotetramer with carbonyl reductase 4 (CBR4) has NADH-dependent 3-ketoacyl-acyl carrier protein reductase activity, and thereby plays a role in mitochondrial fatty acid biosynthesis (Venkatsean et al. 2014). Like many other SDRs HSD17β8 does not show any stability with its fatty acid substrates (Table 5.3.3.2). HSD17β8 is also known to inactivate sex steroids, such as estradiol, testosterone, and dihydrotestosterone (5α-DHT). No HSD17β8 substrates have any significant increase in T_m stability, however scaffold patterns within the steroid substrate class are apparent. Cholesterols produce the highest shifts (2.4-3.2 °C) with NADH (Table 5.3.3.3). Cholesterols that possess a structural modification at position '17', where a hydrocarbon chain is present may cause the increase in thermal

stability. Many shifts appear with non-substrate steroid compounds with both NAD and NADH, such as 5α-androstane-3β-17-diol (0.5/1.4 °C), 5-androstendiol (1.3/1.6 °C), progesterone (1.3/2.3 °C), and allopregnanolone (2.0/2.0 °C) (Table 5.3.3.3). Thermal stability begins to decrease when as the hydrocarbon chain at position '17' is substituted for carbonyl and hydroxyl groups. The androgens in Table 5.3.3.3 exhibit the lowest shifts and bear the closest resemblance to the substrates. 5α-androstendiol and 5α-androstane-3β-17-diol possess hydroxyl groups at position '3' instead of ketone groups, which prevents enzymatic activity, and may account for thermal stability. Interestingly, shifts with NAD do not follow any obvious pattern.

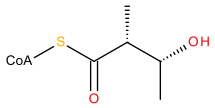
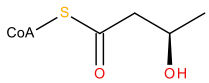
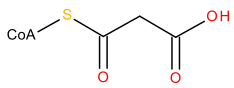
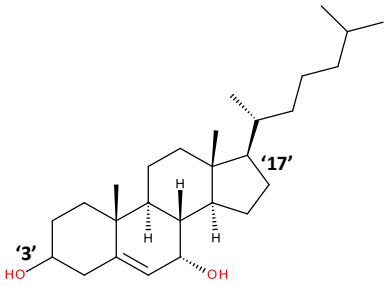
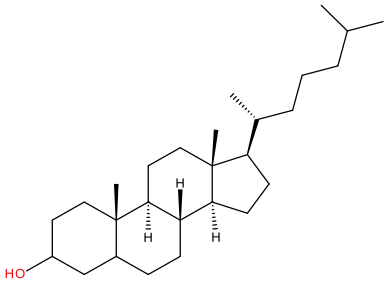
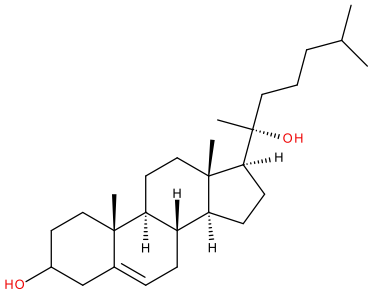
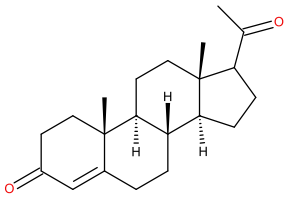
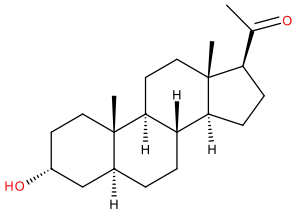
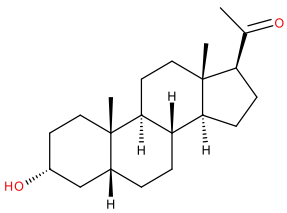
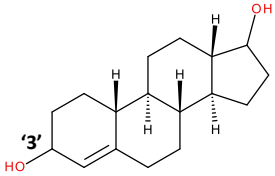
Compound	T_m °C NAD/H	Structure
2-methyl-3-hydroxy-butyryl-CoA	b.t/b.t	
DL-hydroxybutyryl-CoA	b.t/b.t	
Malonyl-CoA	b.t/b.t	

Table 5.3.3.2 HSD17β8 hydroxyacl-CoA substrate T_m shifts b.t =below threshold

Compound	T_m °C NAD/H	Structure	Type
7b-hydroxycholesterol	1.1/3.1		Cholesterols
5α-cholestan-3β-ol	2.2/3.0		
20α-hydroxycholesterol	0.9/2.4		
Progesterone	1.3/2.3		Progestens
Allopregnanolone	2.0/2.0		
Pregnanolone	b.t/2.0		
5α-androstendiol	1.3/1.6		Androgens

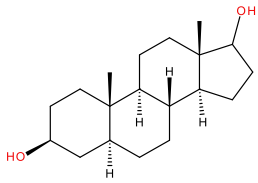
5α-androstane-3β-17-diol	0.5/1.4	

Table 5.3.3.3 HSD17β8 steroid compounds shifts with NADH b.t= below threshold

5.3.4 SUBSTRATE CLASS ANALYSIS FOR ORPHAN SDRs

Substrate validation has shown a high detection rate for SDRs with known substrates, and could be applied to narrow down possible groups to test for enzymatic activity for orphan SDRs and compare computational predictions (Table 3.7.4). For example, hydroxysteroid dehydrogenase like 2 (HSDL2) is an orphan SDR that contains a Sterol-carrier protein-2 (SCP-2) domain. The SCP-2 domain is only found in 3 other human proteins, SCP2, stomatin-like protein 1 (STOML1) and 17β-hydroxysteroid dehydrogenase 4 (HSD17β4). The human SCP-2 domain is known to participate in the intracellular transport of cholesterol and various other lipids (Szyperski et al. 1993). Suggestions of involvement in fatty acid metabolism could be possible due its peroxisomal localisation (Kowalik et al. 2009). HSDL2 generates shifts with steroid based compounds, such as 22-hydroxycholesterol (3.0 °C), while estradiol (2.5 °C) only produces a shift with NADP. For long-chain CoAs such as lauroyl- and decanoyl-CoA, T_m shifts occur in the presence of both NADP and NADPH (Table 5.3.4.1). The thermal shifts with long chain CoAs seem to coincide with the fatty acid from Chapter 3, and would correspond to the nature of an SCP2 domain at the C-terminus of HSDL2.

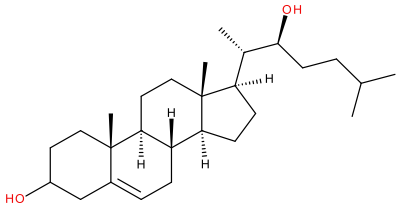
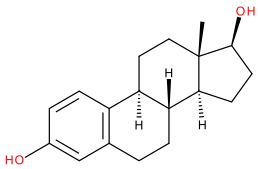
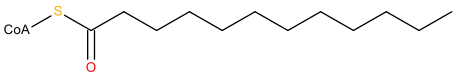
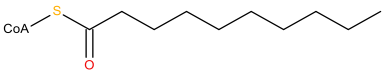
Compound	T_m °C NADP/H	Structure
22-hydroxycholesterol	3.0/b.t.	
Estradiol	2.5/b.t	
Lauroyl-CoA	3.0/1.2	
Decanoyl-CoA	1.5/1.0	

Table 5.3.4.1 HSDL2 shift summary b.t= below threshold

DHRS1 and DHRS11 are both orphan SDRs with structures elucidated by X-ray crystallography. However, neither of these proteins had had their enzymatic function elucidated. DSF screening shows retinoid and steroids compounds thermal stability with DHRS1, which presents them as potential substrate classes to test for activity (Table 5.3.4.2). Similarly, DHRS11 produces T_m shifts with quinones, which would advocate screening with carbonyl xenobiotics for activity (Table 5.3.4.3).

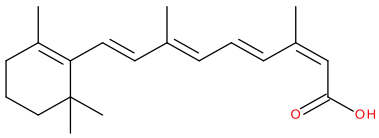
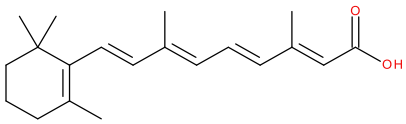
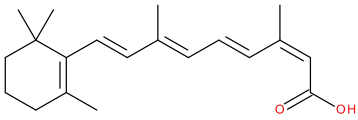
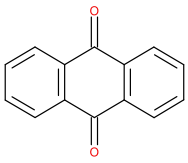
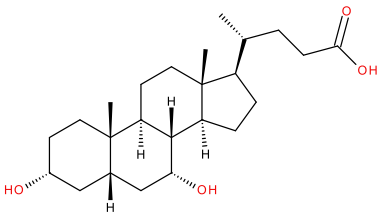
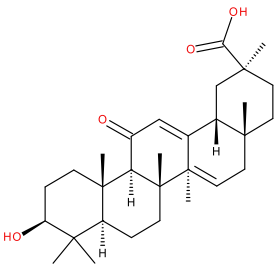
Compound	T_m °C NADP/H	Structure
9-cis retinoic acid	7.1/b.t.	
All trans retinoic	6.3/b.t.	
13-cis retinoic acid	4.3/b.t	
Anthraquinone	4.1/b.t.	
Chenodeoxycholic acid	1.5/b.t.	
Enoxolone	1.4/b.t	

Table 5.3.4.2 DHRS1 shift summary b.t= below threshold

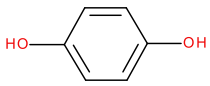
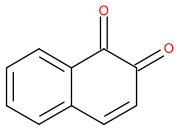
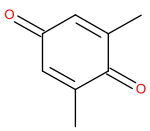
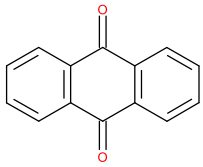
Compound	T_m °C NADP/H	Structure
Hydroquinone	b.t./2.7	
1,2-Naphthoquinone	b.t./1.4	
2,6-Dimethyl-p-benzoquinone	b.t./1.4	
Anthraquinone	b.t./2.6	

Table 5.3.4.3 DHRS11 shift summary b.t= below threshold

SDR	Substrate class with T_m Shifts ≥1.4 °C
DHRS1	Carbonyl xenobiotic/ Retinoid/ Steroid
DHRS11	Carbonyl xenobiotics
HSDL2	Carbonyl xenobiotics/ Fatty acids and Lipids/ Steroid

Table 5.3.4.4 DSF substrate class identification for orphan SDRs

5.4 INHIBITOR IDENTIFICATION BY DSF

In this section, T_m shifts for known inhibitors of human SDRs are investigated to assess DSF as a suitable screening method for ligand identification. Typically, inhibitors produce higher shifts than substrates, as the protein can be fixed into one stable conformational state. The rigid form of ligand bound protein increases stability, and can often prove ameable to crystallisation. DSF was able to detect 90 % of human SDR inhibitors that were in the compound library summarised in Table 6, Appendix. For example, HSD11β1 shifts 7.0/6.6 °C (NADP/H) with carboxenolone (CBX), whereas its substrates exhibits a shift below 0.5°C (Table 5.4.1). Interestingly, DSF was unable to detect N-acetyl-serotonin thermal stability with Sepiapterin reductase (SPR), which could be attributed by weak binding (Ki > 100

µM)(Smith et al. 1992). The high detection rate of SDR inhibitors sets a good precedent to explore new ligands for SDRs by means of DSF.

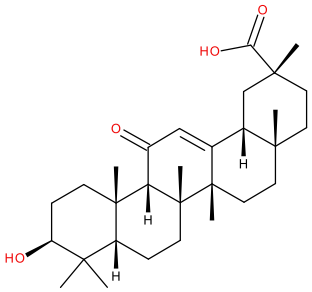
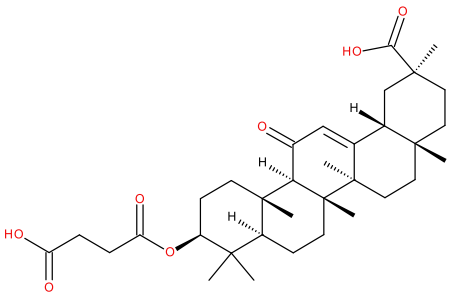
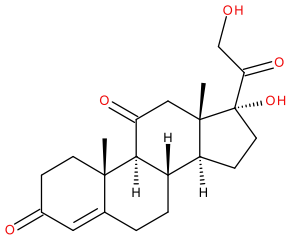
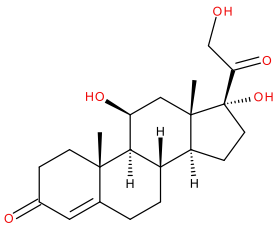
Compound	Structure	T_m °C NADP/H
Glycyrrhetinic acid		4.7/8.1 °C
CBX		7.0/6.6 °C
Cortisone		b.t./b.t
Cortisol		b.t/b.t

Table 5.4.1. HSD11β1 steroid shift summary b.t = below threshold

5.5 HIGH-AFFINITY INHIBITOR DETECTION FOR HPGD

5.5.1 LIGAND LIBRARY DSF SCREENING FOR HPGD

DSF provides a highthroughput method to detect inhibitors for SDR members. Here we explore the potential of DSF to identify new inhibitor scaffolds for the SDR member, 15-hydroxyprostaglandin dehydrogenase (HPGD). Prostaglandins (PGs) are bioactive signalling molecules derived from arachidonic acid by the activity of COX and PG synthase enzymes (Figure 5.5.1.1). PGs have a range of physiological functions e.g. mediating signalling events in inflammation (Ricciotti and FitzGerald 2011), relaxation of smooth muscle (Altura and Altura 1976), and cell differentiation (Portanova et al. 1996). PGE2 is the major eicosanoid implicated in various disease pathologies, such as arthritis (Fattahi and Mirshafiey 2012), inflammatory bowel disease (Sheibanie et al. 2007), and a variety of cancers including prostate, colon, and breast (Rigas et al. 1993, Park et al. 2011, Peng et al. 2012). HPGD is essential in regulating PGE2 levels, and represents a key target to understand the downstream pathways of prostaglandin metabolism. Current inhibitors such as pioglitazone or rosiglitazone lack potency and specificity (Hazra et al. 2007; Subbaramaiah et al. 2012), which makes it difficult to understand the precise effects on prostaglandin metabolism. The need for high-affinity inhibitors with nanomolar applications for HPGD represents a key tool to explore the biological function and intracellular interactions of HPGD.

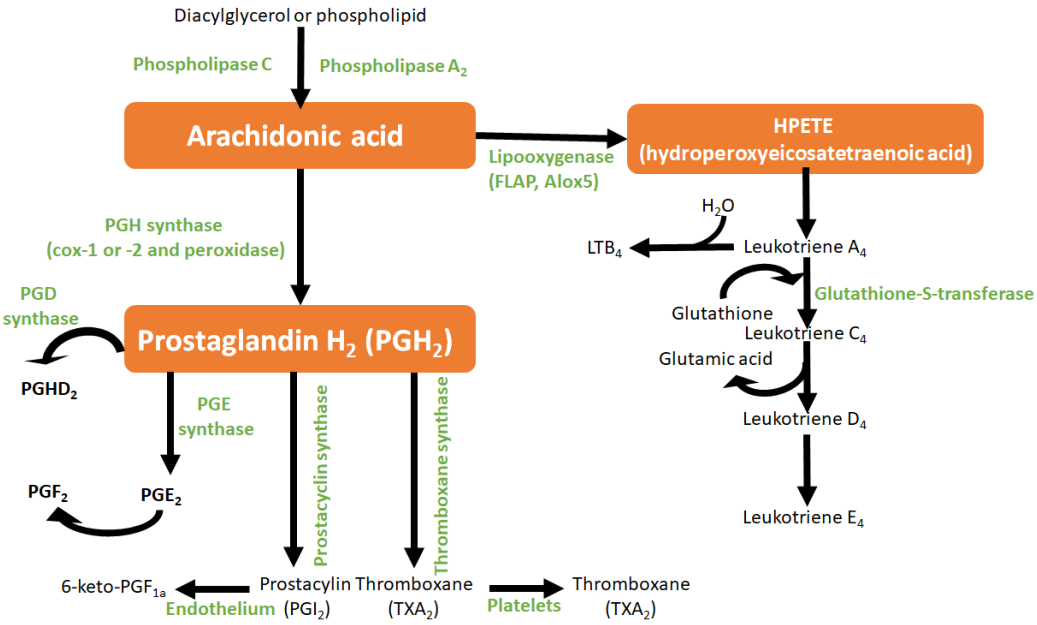
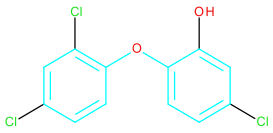
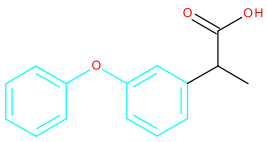
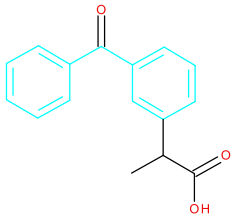
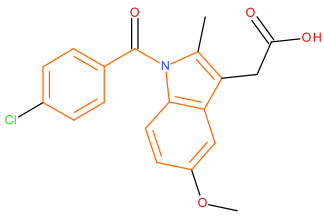
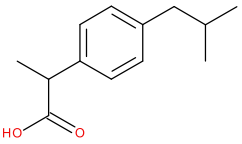
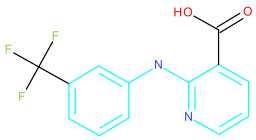
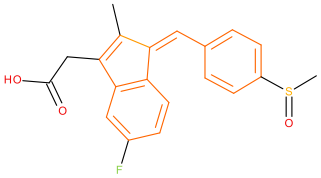
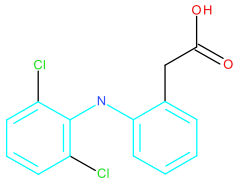
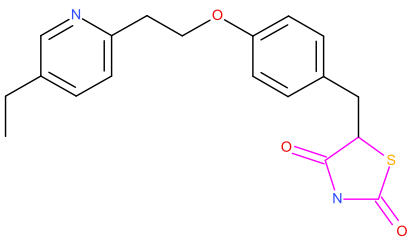


Figure 5.5.1.1 Biosynthesis of Eicasanoids (Garcia-de-Lorenzo et al. 2015)

HPGD generates T_m shifts (Table 5.5.1.1) with well-known prostaglandin inhibitors with NAD, such as fenoprofen (3.8 °C), ketoprofen (3.2 °C), ibuprofen (1.9 °C), niflumic acid (1.9 °C), and diclofenac (1.4 °C). COX inhibitors that cause thermal stability contain a propanoic acid group and possess a benzylbenzene scaffold (Table 5.5.1.1 turquoise). Indomethacin has a benzyl-indene (Table 5.5.1.1 orange) instead of benzylbenzene, which generates a shift of 2.1 °C in the presence of NAD. Peroxisome proliferator-activated receptor (PPAR) inhibitors also produce thermal stability, and are well-established inhibitors of HPGD (Hazra et al. 2007, Subbaramaiah et al. 2012). The thiazolidinedione based HPGD inhibitors (Table 5.5.1.1 purple) pioglitazone and rosiglitazone achieve good T_m shifts (≥ 2.5 °C) with NAD. Interestingly, triclosan is an antibacterial agent that contains the benzylbenzene-bridged scaffold that lacks a propanoic acid, yet produces a high shift with both NAD and NADH of 16.3 and 2.4 °C, respectively. Triclosan, like other coloured compounds, has a tendency to be less specific, and produces T_m shifts with many SDRs (Section 5.7).

Compound	T_m °C NAD/H	Structure	Ligand type
Triclosan	16.3/2.4		Antibacterial
Fenoprofen	3.8/b.t		COX 1&2
Ketoprofen	3.2/b.t		
Indomethacin	2.1/b.t		
Ibuprofen	1.9/b.t		COX 2
Niflumic acid	1.9/b.t		
Sulindac	1.5/b.t		
Diclofenac	b.t/1.4		
Pioglitazone	5.2/1.5		PPAR receptors

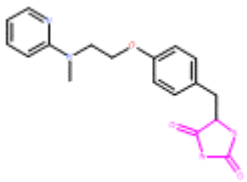
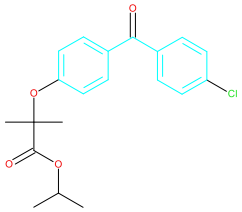
Rosiglitazone	2.7/b.t		PPAR receptors
Fenofibrate	1.5/b.t		PPAR receptors

Table 5.5.1.1 HPGD inhibitor T_m shift summary – scaffold types; benzylbenzene (turquoise), benzyl-indene (orange), **thiazolidinedione** (purple)

5.5.2 INVESTIGATION OF SELECTIVE HPGD INHIBITORS BY DSF

To identify possible inhibitors, The National Institute of Health (NIH) performed a high-throughput screen (qHTS) by measuring HPGD activity from the fluorescence of NADH production (Niesen et al. 2010). Eighty-seven compounds were selected for re-testing in triplicate for IC50 data. The compounds were ranked based on criteria including concentration-response curves, ≥80 % maximum inhibition and R^2>0.9 and were selected for further evaluation by thermal stabilisation properties towards HPGD, in the absence, and presence of NAD or NADH (Table 7, Appendix). DSF was performed as described in Section 4.2 and 5.2, with slight alterations. HPGD was diluted in 100mM Tris pH 8.0, containing 0.01 % Tween-20 (to prevent aggregation) to a final concentration of 1 µM. The final assay volume was 20 µL, and compounds were added to a final concentration of 50 or 250 µM. Lastly, the DSF curves were recorded by increasing the temperature gradient 2 °C/min.

The T_m of ligand-free HPGD was increased by more than 4 °C and 10 °C in the presence of NAD and NADH, respectively. The substrate PGE2 or inhibitors failed to stabilise without cofactor, implying that the cofactor is necessary for the enzyme activity as discussed in

Section 4.3. In the presence of cofactor, the tested compounds exhibited a range in thermal stabilisation up to 12.2 °C in the presence of NAD and up to 13.5 °C in the presence of NADH.

Plotting HPGD IC50 against T_m shift data in the presence of cofactor reveals chemotype dependent correlations (Figure 5.5.2.1). Several chemotype clusters exhibit a weak correlation, while other individual chemotype clusters display a distinct correlation with T_m. In general, compounds in cluster 1 (Figure 5.5.2.1 dark green) show greater T_m shifts with NADH than NAD, as exemplified by the most potent compound 13 (1-(3-methylphenyl)-1H-benzimidazol-5-yl)(piperidin-1-yl)methanone), (IC50 of 56nM, ΔT_m = 7.9/13.5 °C NAD/H). Compound 72 in cluster 9 (Table 7, Appendix) shows a similar profile with a higher T_m shift in the presence of NADH. Members of cluster 5 exhibit the reverse profile, compound 61 (ML147) generates an IC50 of 141nM and T_m shift of 12.2 and 2.9 °C (NADH).

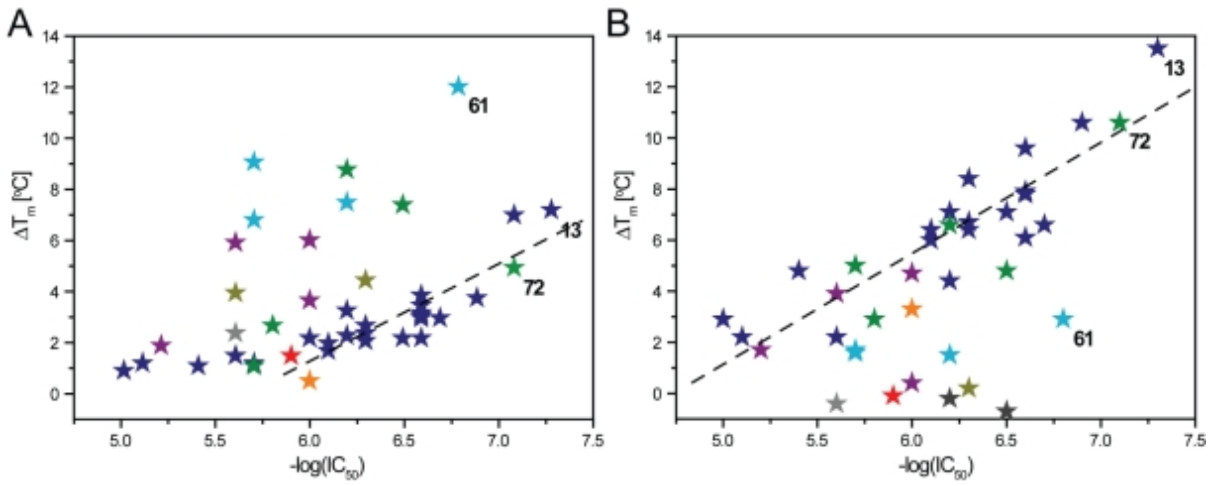


Figure 5.5.2.1 Correlation between thermal stabilisation of HPGD in the presence of A) NAD B) NADH – ΔT_m and pI_{50} (-log (IC50), values are means of at least three independent measurements. The inhibitors are represented by colour-coded symbols according to their chemical clusters, respectively: blue, cluster 1; dark greeb, cluster 2; orange, cluster 3; purple, cluster 4; light blue, cluster 5; light green, cluster 8; grey, cluster 9; dark yellow, cluster 11; red, singletons (a T_m cold not be determined for the tested compounds in clusters 6, 7 & 10). The dotted line in each graph A & B denotes correlations for the compounds in cluster 1, respectively. Numbers within the graphs assign the positions for the inhibitors of particular interest (Niesen et al, 2010).

5.6 IDENTIFICATION OF LIGANDS FOR REGULATORY SDRs

As mentioned before DSF is an assay independent of substrate knowledge, and represents a direct method of monitoring thermodynamic effects of small molecules on regulatory proteins. Other ligand interaction assays, such as Differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC) can require relatively large amounts of protein, and are difficult to perform in a high-throughput manner. In this section, T_m shifts with small molecules have been identified to explore ligand binding sites for regulatory SDRs, such as NMRAL1 and MAT2B.

NMRAL1 belongs to a subset of SDRs including NmrA (Stammers et al. 2001), CC3 (El Omari et al. 2005) and CtBP (Zhang et al. 2002) which adopt the canonical Rossmann fold harbouring the NADP(H) binding motif, but lack the classical YxxxK motif in which Tyr serves as the catalytic base (Filling et al. 2002; Bray et al. 2009). NMRAL1 does not have small molecule binders identified, and exhibits shifts with benzylbenzene scaffolds in the presence NADP/H (Table 5.6.1). Two notable compounds, niflumic acid and triclosan produce a high (>3.0 °C) NADP and a low (<1.5 °C) shift with NADPH. Niflumic acid (NFL) is an interesting candidate for co-crystallisation studies (Section 6.3), compared to molecules such as triclosan that have a tendency to produce T_m shifts with several other SDRs (Section 5.7). This prompted the screening of other nicotinic acid derivatives, such as 2-(4-chloro)-anilino]-nicotinic acid (ZZ0), which produced a moderate temperature shift of 1.8°C (NADP) and 1.3°C (NADPH).

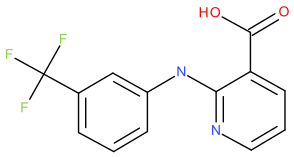
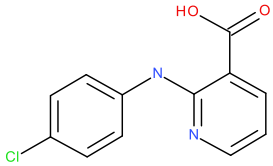
Compound	Structure	T_m Shift NADP/H
Niflumic acid		4.3/1.1
2-(4-chloro)-anilino]-nicotinic acid		1.8/1.3

Table 5.6.1 NMARL1 T_m shifts

5.7 PRIORITISING SDR T_M SHIFTS

When screening against a family of proteins with a large compound library, sometimes individual proteins can produce a large number of hits that can be difficult to prioritise for downstream processing e.g. crystallography. Analysis of variables such as cofactor, number of substrate classes, and oligomeric states may help filter out promiscuous compounds (i.e. molecules that often present hits with several different proteins). The average number of T_m shift ≥ 1.4 °C from all the SDRs screened against the library was calculated for each variable (Figure 5.7.1 A – Cofactor, B – number of substrate classes, and C - oligomeric state). The percentage of SDRs with T_m shift ≥ 1.4 °C against each compound was calculated for each compound class (Figure 5.7.1D).

Cofactor analysis, shows that NAD/H (25 shifts) produces a slightly higher number of average shifts ≥ 1.4 °C for a SDR compared to NADP/H (17 shifts) (Figure 5.7.1A). Similarly, there is no obvious link beween number of substrate classes utilised by an SDR and the number of T_m shifts generated over 1.4 °C (Figure 5.7.1B). SDRs with higher multimeric states produce a greater number T_m shifts than those in a monomeric form (Figure 5.7.1C). However, it is important to note that out of 23 SDRs with known oligomeric state that were

screened only 3 were monomeric (CBR1, CBR3 and HTATIP2). Examination of each compound class (Figure 5.7.1D) shows that quinone-like compounds provided by our collaborator Dalibor Sames and carbonyl xenobiotics compounds generate a large proportion (80 %) of T_m shifts with screened SDRs. The promiscuous character of these compounds may stem from their 'flourphores' scaffolds that have similar excitation and emission ranges to the SYPRO orange dye. Evaluation of compound classes may help narrow, and identify genuine hits from false positives.

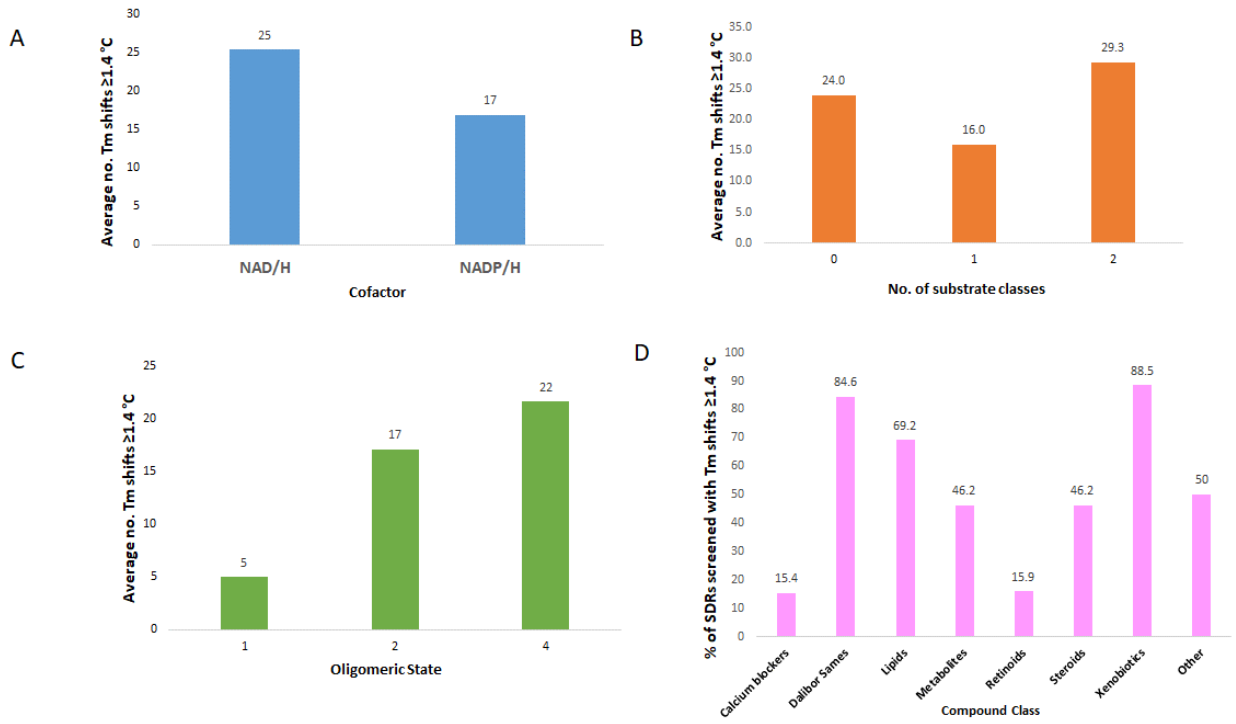


Figure 5.7.1 T_m shift variable analysis – Average number of hits across the compound library vs: A) Cofactor comparison B) Number of different substrate classes turned over by SDRs - *SDRs with a regulatory role are defined as having '0' number of substrate classes C) oligomeric states D) Compound class

5.8 DISCUSSION

This chapter evaluates whether substrates and inhibitors of the human SDR family can be discerned through thermal shift changes in DSF. Thermal shift profiles have been derived for each of the selected SDRs against each compound in the ligand library. It is rare to observe any changes in thermal stability with a particular SDR substrate by DSF, which is most likely due to substrate turnover in the reaction mixture. In addition, many of the SDRs will perform reversible reactions, converting the product back into the original form. This changeable form of the enzyme does not provide increased thermal stability, and therefore, is difficult to detect by DSF.

Table 5.8.1 shows a summary of all the SDRs screened by DSF with substrate T_m shifts over 1.4 °C. The few SDRs that exhibit thermal stability is caused by no obvious single factor. However, a number of factors may contribute to the manifestation of T_m shifts with substrate: for example, most of the SDRs that attain thermal stability with a substrate are NAD/H binding. Although the majority of SDRs that appear in Table 5.8.1 are EC 1.1 oxidoreductases that act on the CH-OH donor group, only 3 out of 16 (HPGD, HSD11β1, HSD17β10) display substrate-induced thermal stability, whereas 3 out of 4 non-EC 1.1 SDRs (GALE, UXS1, and PECR) display T_m shifts with their substrates. From this and previous studies, it appears that non-oxidoreductase enzymes including transferases (Lea and Simeonov 2012), transporters (Boucher and Noll 2011) and synthases (Niesen et al. 2007) manifest substrate T_m shifts. Another observation is that an increase in thermal stability cannot be attributed to substrate class e.g., steroid, xenobiotic or nucleotide. For instance,

HSD11β1 only produces a T_m shift with the xenobiotic substrate menadione, a common substrate often used in kinetic assays to test for xenobiotic activity (Miura et al. 2008). Increased thermal stability is not displayed with other SDRs (CBR1 and DHRS4, Table 4 – Appendix) that turnover menadione. Equally, steroids substrates that produce T_m shifts with HSD17β10 do not with other SDRs that turnover steroids. Additional factors, such as reaction type, low substrate turnover, and the position of reaction equillibrium may also contribute to T_m Shift for substrates.

Enzyme	EC Class	Cofactor	Enzyme Class	Donor Group
GALE	**5.1.3.2**	**NAD/H**	Isomerase	Carbohydrate and derivatives
HPGD	**1.1.1.41**	**NAD/H**	Oxidoreductase	CH-OH
HSD11β1	**1.1.1.146**	**NADP/H**	Oxidoreductase	CH-OH
HSD17β10	**1.1.1.35**	**NAD/H**	Oxidoreductase	CH-OH
	1.1.1.51		Oxidoreductase	CH-OH
PECR	**1.3.1.38**	**NADP/H**	Oxidoreductase	CH-CH
UXS1	**4.1.1.35**	**NAD/H**	Lyase	C=C

Table 5.8.1 Summary of SDRs that show T_m substrate shifts (≥ 1.4°C) from DSF.

The non EC 1.1 extended SDRs do display specific shifts with nucleotide based substrate compounds. The crystal structure of human GALE complexed with UDP-glucose (PDB ID 1EK6, Figure 5.8.1) shows that the majority of hydrogen bonding occurs between the nucleoside moiety of UDP-glucose and main chain residues Asn187, Asn207, Leu208 Asn224, Phe226, Arg239 and Arg300. Work on bacterial GALE (Thoden and Holden 1998) shows the various residues that anchor different parts of UDP-glucose to the enzyme (Figure 5.8.1, orange). The superimposition of human (PDB ID 1EK6) and bacterial (PDB ID 1UDC) GALE structures show the specific accommodation of uracil, and the high amino acid conservation of the active site residues. The accommodation of a specific nucleoside is reflected in the

DSF data as mainly uracil-based molecules attain T_m shifts (Table 5.3.1.2). TDP, another purine base is the only other nucleotide to generate a T_m shift (Table 5.3.1.2).

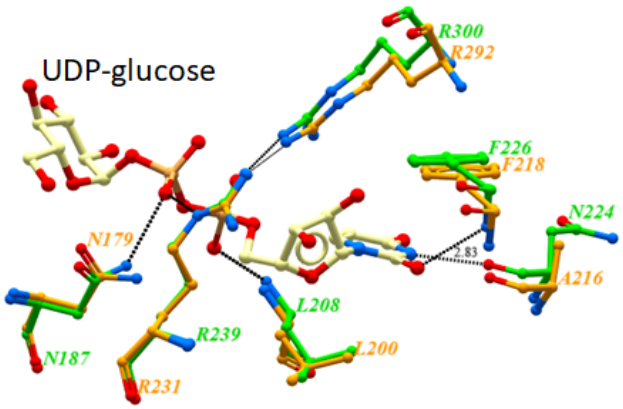


Figure 5.8.1 Human and Bacterial GALE residues that interact with GDP-glucose - Human GALE (green carbon atoms), Bacterial GALE (orange carbon atoms) UDP glucose (cream carbon atoms), Oxygen atoms (red), Nitrogen atoms (blue), Hydrogen bonds (dotted lines) Thoden at el 2000.

The lyase UXS1 (EC 4.1.1.35) also shows increased thermal stability with its substrate UDP-glucuronate. Similar to GALE, UXS1 also displays thermal stability with UDP-based sugars (Table 5.3.1.3) and the highest T_m Shift occurs with TDP. Compared to UDP, TDP lacks an oxygen at position C2 and contains a methyl group at position C5. Interestingly, unlike UDP-glucose, TDP-glucose does not show increased thermal stability, which suggests that stability is based on the nucleotide and not the sugar moiety. Unlike the EC 1.1 oxidoreductases, PECR acts on the CH-CH double bond of substrates (EC 1.3, Table 5.8.1) and attains T_m shifts with its acyl-CoA substrates. The observation that higher degrees of thermal stabilisation are achieved with medium-chain CoAs (6-12 carbon chain length) is consistent with kinetic studies in *Cavia porcellus* PECR (Das et al. 2000).

Despite the low detection rate of individual substrates for human SDRs, most (> 70 %) SDRs show T_m shifts with compounds within their substrate class (Figure 5.3.2.2). Many display T_m shifts with compounds similar to their substrates and reveal which part of the scaffold causes the increased thermal stability (Section 5.3.3). For example, analysis of shifts with GMDS show that T_m shifts are dependent on the presence of a guanosine moiety of a nucleotide. (Table 5.3.3.1). Examination of the GMDS active site in complex with NADP and GDP (Figure 5.8.2 - Green) shows the interaction between the nucleotide and amino acid residues. The diphosphate moiety forms hydrogen bonds to 3 residues (Asn208, Arg247 and Arg325) and the nucleoside moiety forms 5 hydrogen bonds with 4 residues (Asn217, Lys222, Gly241 and Glu328) (Figure 5.8.2). Table 5.3.3.1 shows that the replacement of the diphosphate to monophosphate causes a reduction in T_m shift, which is most likely due to the hindrance of possible hydrogen bond formation to residues Asn208, Arg247, and Arg325. DSF like other studies on protein families show that hydrogen bonding may be linked to thermal stability (Vogt and Argos 1997).

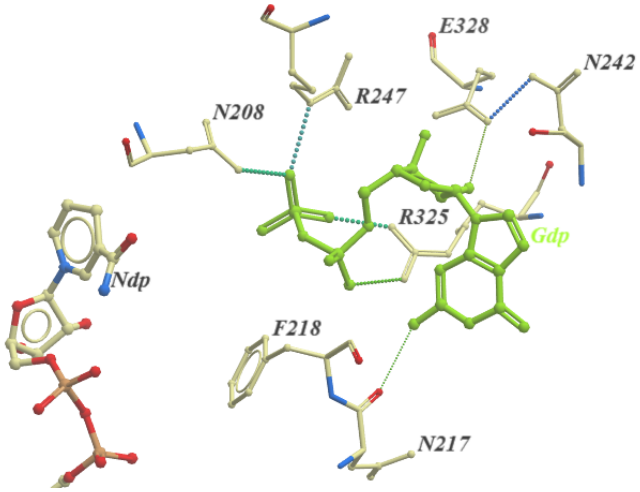


Figure 5.8.2 GMDS active site in complex with NADP and GDP - GDP (green carbons), NADP (grey carbons), GMDS amino acids (cream carbons), Oxygen atoms (red), Nitrogen atoms (blue), Hydrogen bonds (black dotted lines)

HSD17β8 is another SDR that displays scaffold patterns among its substrate class, the steroids. HSD17β8 typically acts on the hydroxyl group of testosterone at position '17' and position '3' of bile acids or pregnanes (Fomitcheva et al. 1998). The T_m shifts with cholesterols, progestens and androgens all show modifications to a carbonyl or hydroxyl group at either position '17' or '3' (Table 5.3.3.3). T_m shifts begin to decrease as the substitution groups begin to resemble HSD17β8 substrates, which reiterates that increased thermal stability is difficult to achieve with a substrate.

The success of substrate class identification for known SDRs prompted the analysis of orphan SDRs. Table 5.8.2 shows a comparison of substrate classes identified (Table 5.3.4.4) for orphan SDRs and those identified by theoretical predictions in Chapter 3 (Table 3.7.4). Table 5.8.2 not only aids the assessment of substrate classes identified by DSF for orphan SDRs, but also highlights matching substrate classes between theoretical and DSF

predictions (Table 5.8.2 red) that could be used to test for enzymatic activity (See Chapter 6.5).

SDR	YxxxK site	Substrate class Computation Prediction (Chapter 3)	Substrate class T_m Shifts ≥1.4 °C
DHRS1	YES	Fatty acids and lipids/ **Steroid**	Carbonyl xenobiotic/ Amino acid derivatives/ Retinoid/ **Steroid**
HSDL2	YES	**Fatty acids and Lipids**	Carbonyl xenobiotics/ **Fatty acids and Lipids/** Steroid
DHRS11	YES	**Carbonyl xenobiotics**/ Fatty acids and lipids/ Steroids	Amino acid derivatives/ **Carbonyl xenobiotics**

Table 5.8.2 DSF substrate class identification for orphan SDRs (matching computational and DSF substrate classes - red)

Similar to substrate class detection, recognition of known SDR ligands also proved successful. In general, inhibitors achieve higher T_m shifts than substrates, as they are able to form rigid complexes with the protein. The crystal structures of human HSD11β1 illustrate that an inhibitor such as CBX forms additional hydrogen bonding with residues Leu217 and Met233 compared to the substrate, thus restricting movement in the active site (Figure 5.8.3). In rodents, HSD1β11 carries out the interconversion of corticosterone and 11-dehydrocorticosterone (Huang et al. 2010). The ternary complex structure of HSD11 β 1 from **Mus musculus complexed shows the orientation of corticosterone and NADPH in the active site (Figure 5.8.3 – PDB ID 1YR5). The overlap of the human HSD11 β 1 shows the same orientation of catalytic residues Tyr183 and Lys187. It also demonstrates how the inhibitor CBX binds in a similar fashion to the substrate in the active site but forms hydrogen bonds to Leu217 and Met233 (Figure 5.8.3, green). These additional hydrogen bonds allow the CBX molecule to form a thermally stable complex.**

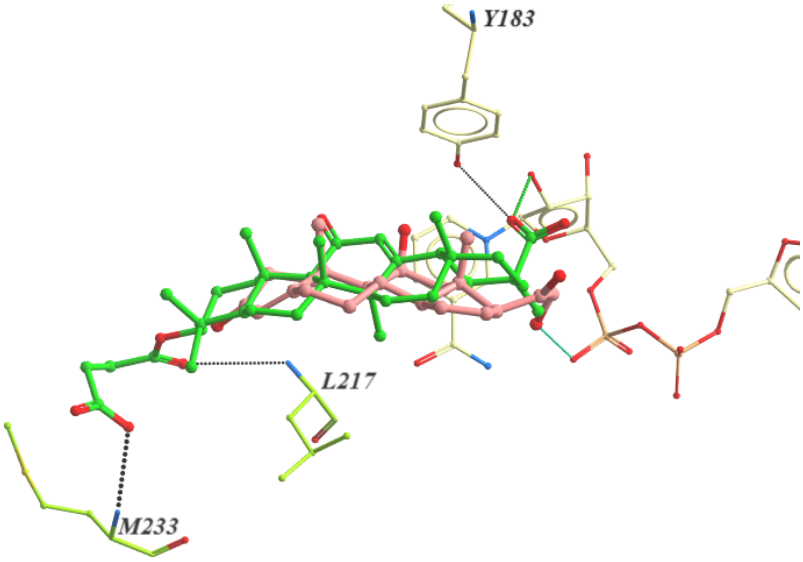


Figure 5.8.3 Active site accommodation of Human and Mouse HSD11β1 – CBX in human HSD11β1 (PDB ID 2BEL - green carbons), Corticosterone in Mouse HSD11β1 (PDB ID 1Y5R - pink carbons), NADP (cream carbons), Oxygen atoms (red), Nitrogen atoms (blue), Hydrogen bonds (dotted lines), Human HSD11β1 residues (lime green) (Zhang et al. 2005).

Comparing SDR ligands to compounds with similar chemical backbones pinpoints which part of the scaffold are key to thermal stability. For example, well-known HPGD inhibitors from different functional classes such as COX and PPAR cause thermal stability (Table 5.5.1.1). The DSF data shows that compounds that cause a T_m shift with HPGD are based on either a benzylbenzene, benzyl-indene, thiazodidinedione scaffold. (Table 5.5.1.1). These common features that are repeated across drug classes emphasises that thermal stabilisation is not based on functional similarity but a compound's basic chemical scaffold.

There are many well-known HPGD inhibitors, but most lack specificity and potency. A search for potential selective HPGD inhibitors by DSF showed a positive correlation between Log IC_{50} and T_m shift that revealed several cofactor dependent chemotype clusters (Figure

5.5.2.1). Kinetic analysis demonstrated that each chemotype represented a different mode of HPGD inhibition (Jadhav et al. 2010, Niesen et al. 2010).

Mode of action studies on compound 13, 61 and 72 shows different mechanisms of inhibition for each chemotype (Niesen et al. 2010) (Table 7, Appendix). Compound 13 displays a 40-50 % decrease in Vmax and Km with respect to PGE2, suggesting an uncompetitive mode of inhibition. In contrast, NAD titration causes a lower Vmax, but not Km that implies a non-competitive mode of inhibition. Compound 61 displays a competitive mechanism of inhibition, as there is no significant change in Vmax but an increase in the Km with respect to PGE2. With respect to NAD, compound 61 exhibits uncompetitive inhibition, as it shows 50 % reduction in Vmax and Km uncompetitive inhibition. Similar to inhibitor 13, compound 72 also incurred a reduction of approximately 50 % in Vmax with respect to PGE2. Unlike the other two compounds, compound 72 displayed no effect on the Km and a decrease in catalytic efficiency (Kcat).

Compounds 13, 61 and 72 also display low reactivity profiles against proteins with similar roles in PG signalling (Niesen et al. 2010), which provides support for the potential application of the selective inhibitor probes for HPGD. The availability of multiple inhibitors with diverse chemical structures and mechanisms can enhance cellular research by observing the effects on specific locations of the target protein. The new inhibitor chemotypes with nanomolar potencies and low-cross reactivity verified by DSF and kinetic

studies provides a set of tools, for functional studies on the role of HPGD in prostaglandin signalling pathways.

Often SDRs screened by DSF achieve several T_m shifts with various types of compounds. When selecting compounds for further downstream analysis (i.e. kinetic studies or crystallisation), it is beneficial to consider factors such as oligomeric state, and compound class to remove false positives. For instance, when dealing with SDRs with oligomeric states higher than 2, the T_m shift significance level may have to be increased to ≥ 1.4 °C. In addition, compound classes such as the carbonyl xenobiotics generate significant T_m shifts with 80 % of the screened SDRs, suggesting possible non-specific interaction with the protein (section 5.7.1).

DSF is also a direct method of monitoring proteins independent of substrate and, therefore, represents a method to investigate potential non-catalytic ligands for regulatory proteins. For example, the subunit MAT2B (MATIIβ) is an extended SDR member that functions by regulatory association with MAT2A (MATIIα) by lowering the Km for its substrate L-methionine. DSF screening identified resveratrol, as a thermal stabiliser of MAT2B by 1.8 °C in the presence of NADP/H. Co-crystallisation structures MAT2B shows two resveratrol binding sites per monomer (Figure 5.8.4A PDB ID 2YDX). One resveratrol is located next to the cofactor binding site, where it forms interactions with NADP and the Tyr159 from the conserved YxxxK motif (Figure 5.8.4B). The second resveratrol molecule is situated between the dimer interface of the two monomers, which suggests its binding provides a stabilised

structure for crystal packing interactions (Figure 8.7.4C) (Shafqat et al. 2013). As in the case of MAT2B, DSF illustrates that compounds have an increase in T_m can be identified as potential ligands.

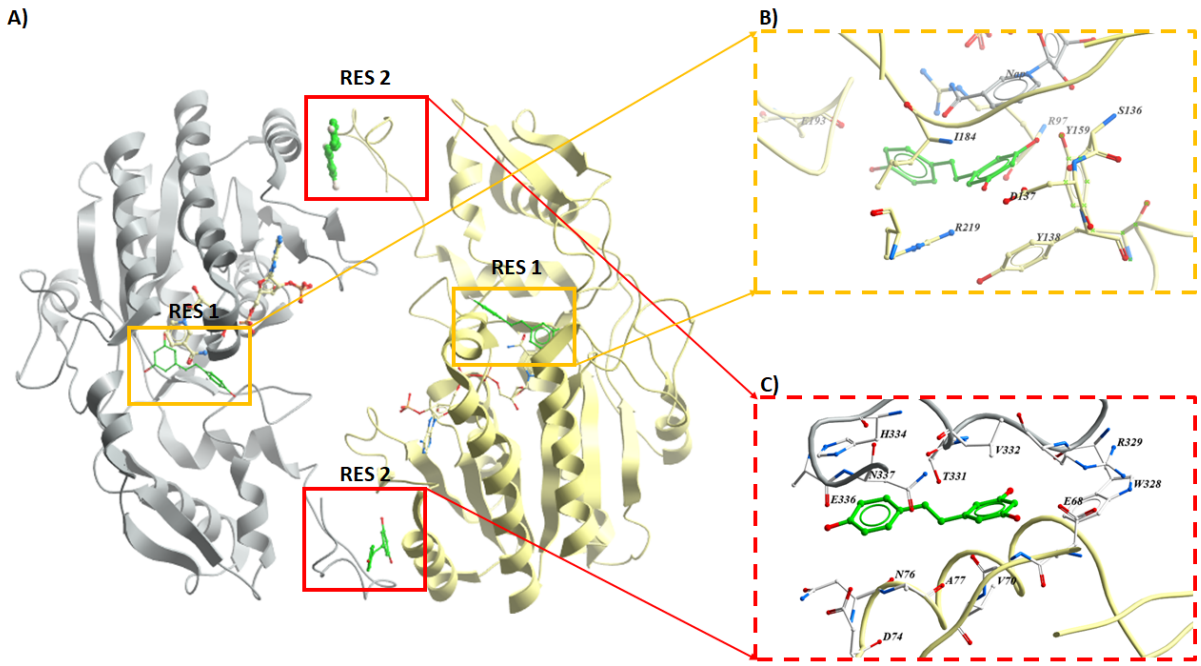


Figure 5.8.4 Crystal structure of Human MAT2B – A) MAT2B dimer with resveratrol, B) Binding of Resveratrol to SDR substrate pocket, C) Binding of resveratrol to C-terminal region. MAT2B Chain A (grey), MAT2B chain B (yellow), NADP (cream), Oxygen atoms (red), Nitrogen atoms (blue), and Hydrogen bonds (dotted lines) (Shafqat et al. 2013).

In conclusion, DSF is able to recognise substrate classes for an SDR enzyme for further testing in activity assays. It also provides a way to identify thermally stable scaffolds that can help in the design of novel inhibitors or probes. DSF also represents an important tool for the discovery of small molecule binders for regulatory proteins such as NMRAL1, an aspect

that is further explored in Section 6.3. The subsequent chapter explores the application of DSF in crystallisation and kinetic studies to aid characterisation of human SDRS.

CHAPTER 6 – APPLICATION OF THERMAL SHIFT DATA FOR STRUCTURAL STUDIES

6.1 INTRODUCTION

In Chapters 4 and 5, family-wide ligand profiling by DSF for 28 human SDRs has been described. Much of the T_m shift data has been compared to existing information in the literature to evaluate the reliability of DSF as a screening tool for ligands that bind to proteins. DSF is a valuable tool for determining cofactor preference independent of any substrate knowledge and displayed thermal stability with documented inhibitors that bind to human SDRs. It is thought that the increase in the midpoint of thermal unfolding (i.e. T_m) is associated with increased protein stability. There are many instances that illustrate an increase in T_m is linked to protein stability. For example, in DSC experiments it is shown that the presence of calcium ions raises the free energy of unfolding by 13.9 kcal/mol and the T_m by 20 °C for βγ-crystallin (Kretschmar and Jaenicke 1999; Kretschmar et al. 1999). Also, thermal denaturation of cytochrome c showed stabilisation by hydrophobic amino acids (Taneja and Ahmad 1994). Stable proteins are a prerequisite for structure determination projects and optimisation of conditions such as concentration, pH, buffer and additives (e.g. small molecules) (Taneja and Ahmad 1994) may improve the success rate of crystallisation. The first part of this chapter explores the application of T_m shift data from Chapters 4 and 5 for the crystallisation of structurally uncharacterised human SDRs (6.2-6.5).

The SDR family of proteins performs a diverse range of functions and accommodates a variety of substrates (El-Kabbani et al. 2004; Guo et al. 2006; Luu-The et al. 2006; Bateman

et al. 2008). However, many members lack substrate or functional knowledge, and therefore a need to explore these orphan SDRs to bridge the gap between sequence information and biochemical function exists. Current methods for determining a substrate include computational docking, mass spectrometry and enzymatic/kinetic assays, which may involve screening very large compound libraries before possible ligands are found. Substrate exploration often begins by identifying ligand recognition sites using homology modelling to proteins of known structure and function. Although this method can provide insight of specific substrate accommodation, it is dependent on the availability of structural and functional information of homologues and paralogues. The second part of this chapter uses the identified substrate class from Chapter 5 (Table 5.3.4.2) as starting point for activity screening to decrypt substrates for orphan SDRs.

6.2 DHRS4 STRUCTURAL ELUCIDATION BY COFACTOR THERMAL STABILITY

The crystal structure of porcine DHRS4 reveals that the C-terminal PTS1 signal residues (Ser-Arg-Leu) forms an interface between two monomers in an overall tetrameric structure (2ZAT, Tanaka et al. 2008). Previously at the SGC, human DHRS4 was expressed in *E. coli* using p11 vector (Appendix, Figure 1), which adds Gly-Ser residues directly after the C-terminal SRL tripeptide, as part of the vector-encoded sequences. The Gly-Ser addition may have caused an obstruction in the formation of tetrameric DHRS4. To produce a stable tetrameric protein in solution, constructs encompassing the SRL tripeptide were cloned into a pNIC-CTHF vector (Appendix, Figure 5). The vector contains an N-terminal His_6-tag and TEV protease cleavage site, and does not incorporate additional residues on the C-terminus, which would allow the DHRS4 tetramer to form.

In Section 4.5 (Figure 4.5.2B), DSF determined DHRS4 to have a NADP/H cofactor preference. The deviations in the unfolding curves for DHRS4 meant T_m shift values were not good indicators of thermal stability and thus, the slope of transition had been used to distinguish cofactor preference. Interestingly, only the addition of NADPH caused a characteristic unfolding curve to form a steep transition (Figure 4.5.2B - 8.1 ΔI/Δ°C). DSF screening against the ligand library in the presence of cofactors displayed a similar effect, as adding NADP would generate non-characteristic transition curves and the majority of T_m shifts occurred in the presence of NADPH. In an attempt to solve the crystal structure of DHRS4, crystallisation experiments with all cofactors (NAD/H and NADP/H) were conducted. The parallel nature of the experiments would show if crystal formation had any correlation to the cofactor patterns in the DSF data.

6.2.1 DHRS4 CRYSTALLISATION AND STRUCTURE DETERMINATION

DHRS4 (18.5 mg/mL) was supplemented with 6.7 mM of one cofactor (NAD, NADH, NADP or NADPH) prior to crystallisation to test the effect on crystal formation (Table 6.2.1.1). Initial crystallisation trials only showed granular precipitation in the presence of NADPH. The DHRS4 protein was then transferred into a lower salt (250 mM NaCl) buffer for crystallisation, with the rationale that the protein will have a higher tendency to come out of solution, a pre-requisite of crystal formation. A lower-salt buffer indeed caused crystals to appear, only in the presence of NADPH, and not NADP/NAD/NADH (Figure 6.2.1.1). All conditions were tested at 20 °C and Table 6.2.1.1 summarises the resulting outcomes.

Protein Conditions	GF Buffer NaCl (mM)	Precipitation	Crystal growth Condition	Protein:Reservoir Solution Ratio
DHRS4 + NAD	500	None	N/A	N/A
DHRS4 + NADH	500	None	N/A	N/A
DHRS4 + NADP	500	None	N/A	N/A
DHRS4 + NADPH	500	Granular precipitation	N/A	N/A
DHRS4 + NAD	250	None	N/A	N/A
DHRS4 + NADH	250	None	N/A	N/A
DHRS4 + NADPH	250	None	N/A	N/A
DHRS4 + NADPH	250	None	0.1 M potassium thiocyanate and 30 % w/v PEG ME 2000, pH 7.0	1:2

Table 6.2.1.1 DHSR4 co-crystallisation trials with cofactors and NaCl reduction

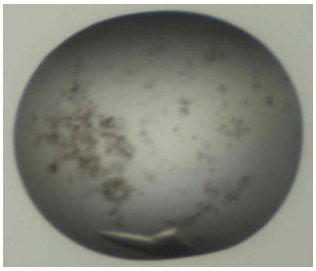


Figure 6.2.1.1 DHRS4 crystal with NADPH - DHRS4 crystal in 0.1 M potassium thiocyanate pH 7.5 and 30 % w/v PEG ME 2000 gave rise to 1.7 Å diffraction at Diamond Light Source beamline I03.

Diffraction data was collected at the Diamond Beamline I03 by Dr. Emilie Ugochukwu (part of the crystallography group at SGC Oxford (https://www.thesgc.org/groupprofile). The structure of human DHRS4 was solved by molecular replacement, using the porcine reductase coordinates (PDB ID 2ZAT, Tanaka et al. 2008) as search model. The model was refined to a 1.7 Å resolution with final R_{FACT} 0.212 and R_{FREE} 0.25. Refinement statistics are summarised in Table 8, Appendix.

6.2.2. OVERALL STRUCTURE

The final model of DHRS4 is a tetramer with a NADPH molecule bound in each monomer (Figure 6.2.2.1A). Residues 34-270 form the typical SDR Rossmann fold with a seven stranded β-sheet (β1-7), which is flanked by three parallel α-helices on each side (α1-3 and α4-6) (Figure 6.2.2.1A). Residues Cys248-Cys256 forms an additional helix (α7) between α6 and β7, which constitutes the substrate-binding region of the C-terminus. The two longest helices (α4 and α5) form the interfacial four-helix bundle with residues Glu131, Asp135, Lys143, Asn183, Thr195, characteristic of dimeric and tetrameric SDRs (Figure 6.2.2.1B). Analogous to the porcine orthologue, human DHRS4 also forms inter-subunit interactions via the C-terminal PTS1 signal buried in the interior of the tetrameric structure (Figure 6.2.2.1C).

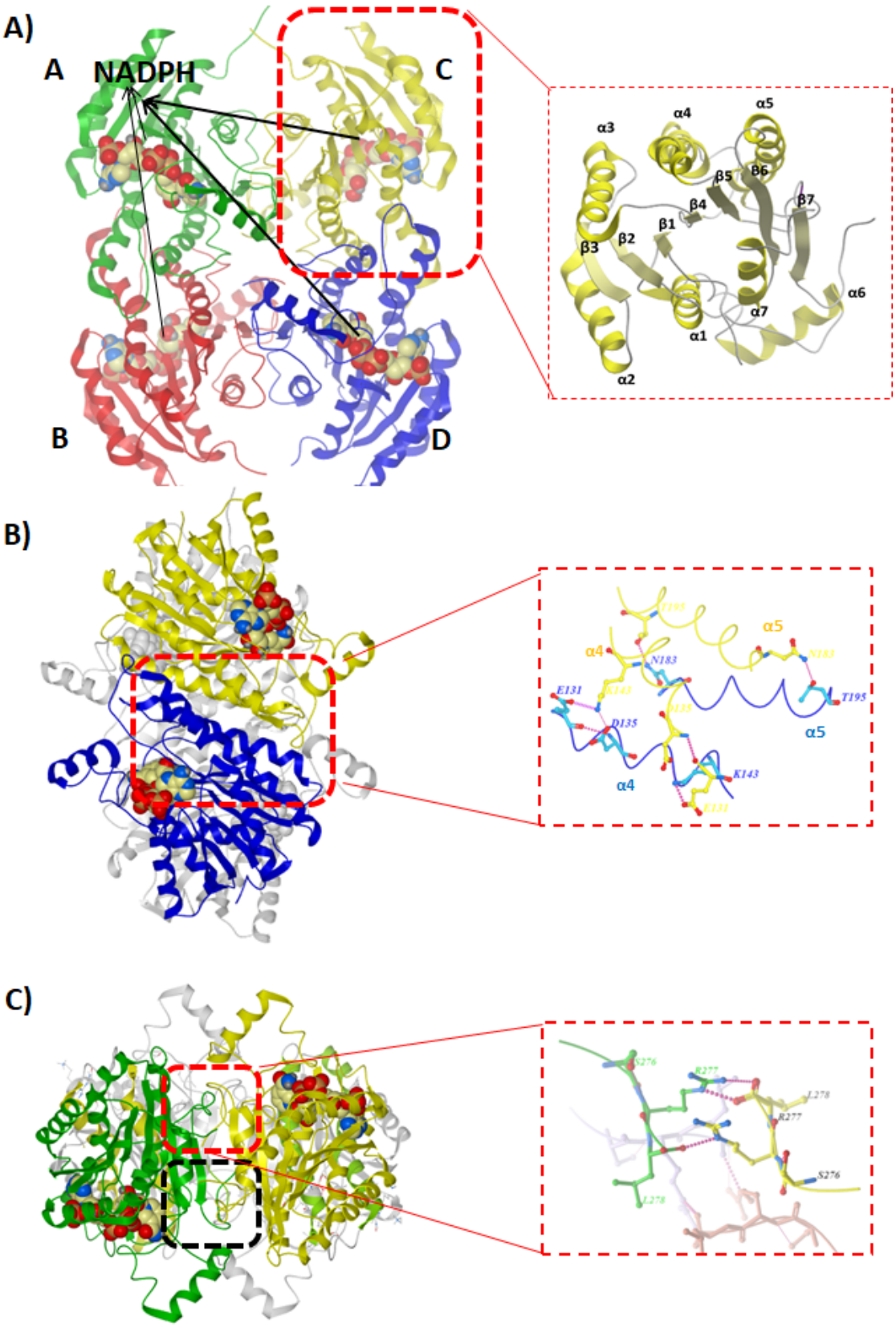


Figure 6.2.2.1 DHRS4 structure –A) tetrameric formation and overall Rossmann fold - β-sheet (light yellow), α-helices (yellow) B) Subunit C and D interface; D) Subunit A and C interface of PTS1 signal; sphere model is of a bound NADPH molecule.

6.2.3 COFACTOR BINDING AND ACTIVE SITE ARCHITECTURE

Each NADPH molecule of the DHRS4 tetramer adopts an extended conformation within a hydrophobic pocket (13.4 Å between C2 of nicotinamide and C6 of the adenine ring). DHRS4 shows a variation in the cofactor-binding motif sequence from classical SDRs (G-XXX-G-X-G), in that the first Gly residue is replaced with Ala. Ala39 with its additional methyl group replaces the Gly residue in a similar orientation and has no significant structural consequences, when superimposed with SDRs containing the preserved Gly residue (Figure 6.2.3.1A). The negatively charged phosphate substituent of NADPH at the O2′ adenine atom is accommodated by the presence of two basic residues Arg64 and Lys65 (Figure 6.2.3.1B).

As observed with many SDRs, DHRS4 retains the conserved catalytic tetrad Asn141, Ser168, Tyr182 and Lys186 (Figure 6.2.3.1A). Inspection of catalytic residues by H++ 3.0 (Gordon et al. 2005) estimates the pKa values of Tyr182 and Lys186 to be >12.0 and 7.6, respectively. In the crystallised conditions of DHRS4 at pH 7.0, Tyr182 and Lys186 would have been protonated, allowing the formation of a proton relay network that accommodates a hydride ion at position C4 of the nicotinamide ring in NADPH (Jornvall et al. 1995, Oppermann et al. 1997, Ghosh et al. 2001, Filling et al. 2002). A water molecule hydrogen-bonds to protonated Tyr182 and is further coordinated by Ser169, thus imitating the position of carbonyl a substrate (Figure 6.2.3.1B). The torsion angle C2N-C3N-C7N-O7N represents the planarity of the cofactor carboxamide in relation to its attached pyridine ring and is indicative of the cofactor redox state (Figure 6.2.3.1C). In each DHRS4 subunit, this angle is significantly less than 180° (15.4°), and suggests NADPH, the reduced form of the cofactor, to be present in the crystal (Tanaka et al. 1996).

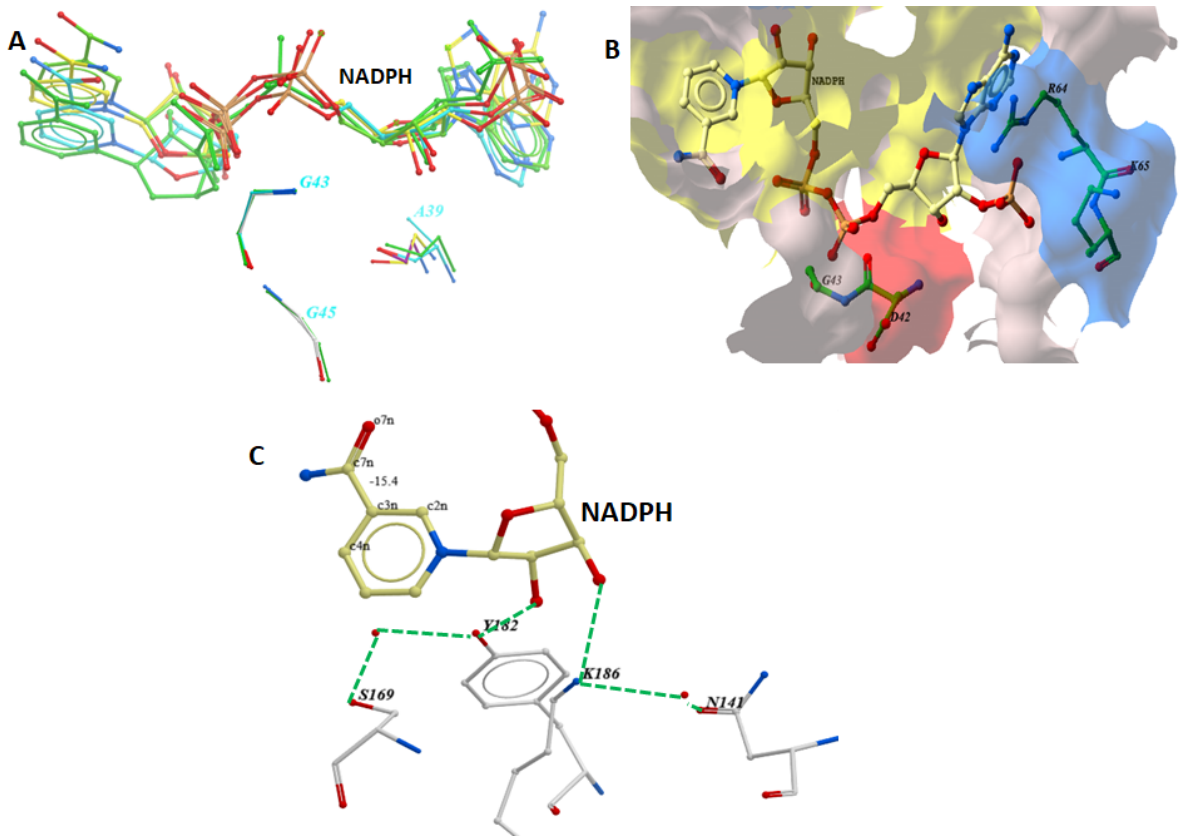


Figure 6.2.3.1 DHRS4 cofactor binding and active site coordination – A) superimposed NADPH of human DHRS4 with CBR3 (orange), DCXR (green) and SPR (yellow) B) cofactor binding pocket hydrophobicity - DHRS4 residues (green carbon atoms Ala39), hydrophobic residues (yellow), Negative charge (red), positive charge (blue) and uncharged (grey) C) Active site hydrogen bonding interactions and nicotinamide torsion angle; NADP represented as ball and stick, Oxygen atoms (red), Nitrogen atoms (blue), Hydrogen bonds (green dotted lines).

6.3 STRUCTURAL CHARACTERISATION OF SMALL-MOLECULE LIGANDS FOR NMRAL1

6.3.1 BACKGROUND

As shown with DHRS4 (section 6.2) it is common to co-crystallise SDRs with nicotinamide cofactors, as it permits active site formation, which makes a more stable protein. Here we test ligand hits identified from DSF in crystallisation, and determine the co-crystal structures to visualise binding of small molecules for NMRAL1, a redox sensor protein.

NMRAL1 was first proposed as an NADP/H sensor on the basis of its reported crystal structure, showing an asymmetric dimer composed of one NADP-bound ($NMRAL1_{holo}$) subunit and one cofactor-free ($NMRAL1_{apo}$) subunit (Zheng et al. 2007). Structural superposition of the $NMRAL_{apo}$ and $NMRAL_{holo}$ subunits reveals substantial loop-to-helix rearrangements that occur upon NADP/H binding, in a region close to the dimer interface. Interestingly the dimerisation region of NMRAL1 also overlaps with the binding region of ASS1, a protein that regulates the nitric oxide (NO) cycle (Zhao et al. 2008). Therefore, the structural switch and dimerisation in NMRAL1 may represent a potential mechanism for the modulation of ASS1 function in response to intracellular NADP/H level. Modulation of ASS1 activity by NMRAL1 could represent a potential regulatory point to maintain a balance between the intracellular redox state and NO level. Small molecule binders could provide a basis for therapeutic and cellular applications to study the interactions between ASS1 and NMRAL1. As as first step, Niflumic acid (NFL) derivatives were identified from DSF ligand screening in Chapter 5 (Section 5.6). Here, we describe the structures of NMRAL1 complexed with NFL and 2-(4-chloro-phenylamino)-nicotinic acid (ZZ0 – an analogue of NFL).

6.3.2 CO-CRYSTALLISATION AND STRUCTURE DETERMINATION OF NMRAL1 WITH NFL AND ZZ0

NMRAL1 (9.6 mg/mL) was crystallised by pre-incubation with 1.5 mM NADP and 1.1 mM NFL or ZZ0 (Figure 6.3.2.1). Crystals were grown at 20 °C in 150 nL sitting drops as detailed in Table 6.3.2.1.

	Ligand Conc. (mM)	NMRAL1 conc. mg/mL	NADP conc. (**mM**)	Crystal solvent condition	Protein:Solvent ratio
NiFL	1.0	9.6 (0.28 mM)	1.5	2.1 M sodium malate pH 7.0	1:1
ZZO	1.5	12.7 (0.38 mM)	1.9	2M ammonium sulphate and 0.1 % BIS-TRIS pH 5.5	1:2

Table 6.3.2.1 Co-crystallisation details of NMRAL1 with Niflumic acid (NFL) and 2-(4-chloro-phenylamino)-nicotinic acid (ZZO)

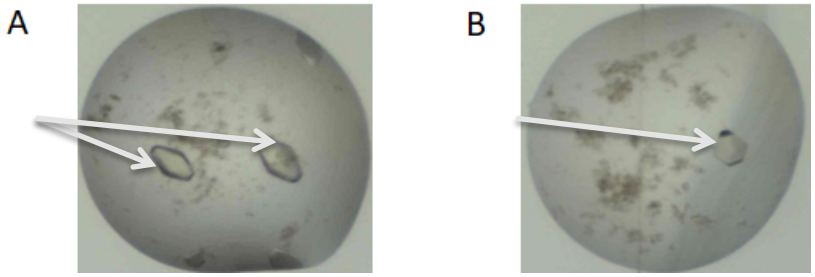


Figure 6.3.2.1 – NMRAL1 crytsals A) with NFL B) ZZO

The structures of NMRAL1 in complex with NADP and NFL ($NMRAL1_{NFL}$), and in complex with NADP and ZZ0 ($NMRAL1_{ZZ0}$), were solved by molecular replacement using the HSCARG structure (PDB ID 2EXX) as search model. The $NMRAL1_{NFL}$ and $NMRAL1_{ZZ0}$ structures were refined to 1.85 Å (R_{FACT} 0.167, R_{FREE} 0.197) and 2.25 Å (R_{FACT} 0.204, R_{FREE} 0.248) resolution, respectively. Refinement statistics are summarised in Table 8, Appendix.

6.3.3 CRYSTAL STRUCTURE REVEALS A HYDROPHOBIC POCKET FOR SMALL-MOLECULE BINDERS

The structures of $NMRAL1_{NFL}$ and $NMRAL1_{ZZ0}$ (Figure 6.3.3.1A) superimpose well with each other (rmsd 0.45 A for all Cα atoms) and reveal the same overall architecture as observed previously in the $NMRAL1_{holo}$ (NMRAL1 and NADP) conformer of the reported asymmetric dimer ($NMRAL1_{apo}$:$NMRAL1_{holo}$) (Zheng et al. 2007).

Both NFL and ZZ0 compounds bind to a surface-exposed pocket in the cofactor-binding site, close to the NADP nicotinamide ring (Figure 6.3.3.1B). The ligand pocket is mainly hydrophobic in nature and accommodates NFL and ZZ0 in similar orientations (Figure 6.3.3.1C-E). The two compounds share a common phenylpyridine scaffold, but differ in the substituent group/position at the phenyl ring. The pyridine ring of each compound stacks against the NADP nicotinamide such that its carboxylate is hydrogen-bonded to the $NADP^+$ ribose O2' and indole nitrogen of Trp82. The phenyl rings of NFL and ZZ0 slot between the side-chains of His162 and Leu257, while their substituent groups engage in different interactions (Figure 6.3.3.1E).

Close inspection of the NMRAL1 complexes shows that there are local differences between the $NMRAL1_{NFL/ZZO}$ complexes and $NMRAL1_{HOLO}$ conformer, which reflect conformational flexibility in two loop regions (Fig. 6.3.3.1B). Region 1 represents an extended loop spanning Leu162-Leu176 that is moved only in the $NMRAL1_{zzo}$ structure (Figure 6.3.3.1E). However, NFL and ZZO affect the orientation of His162 but with no loop movement in $NMRAL_{NFL}$. Region 2 represents the connecting loop (Tyr246-Arg255) between helices α10 and α11, which forms part of the ligand-

binding pocket by contributing Tyr246 to interact with the ligands (Figure 6.3.3.1E). This region is also disordered in the *apo* protomer of the heterodimer (Figure 6.3.3.1B), but forms an ordered loop in the *holo* protomer (Zheng et al. 2007) as well as in the $NMRAL1_{NFL}$ and $NMRAL1_{ZZ0}$ structures, suggesting this is driven by NADP/H binding. In the both ligand bound NMRAL1 structures, the Tyr246 is displaced from its original position by the relative subtituent group (NFL: 5-methyl-trifluoro, ZZO: 4-chloro), pushing the region 2 loop further out, in order to disrupt the asymmetric dimer formation originally described (Zheng et al. 2007)

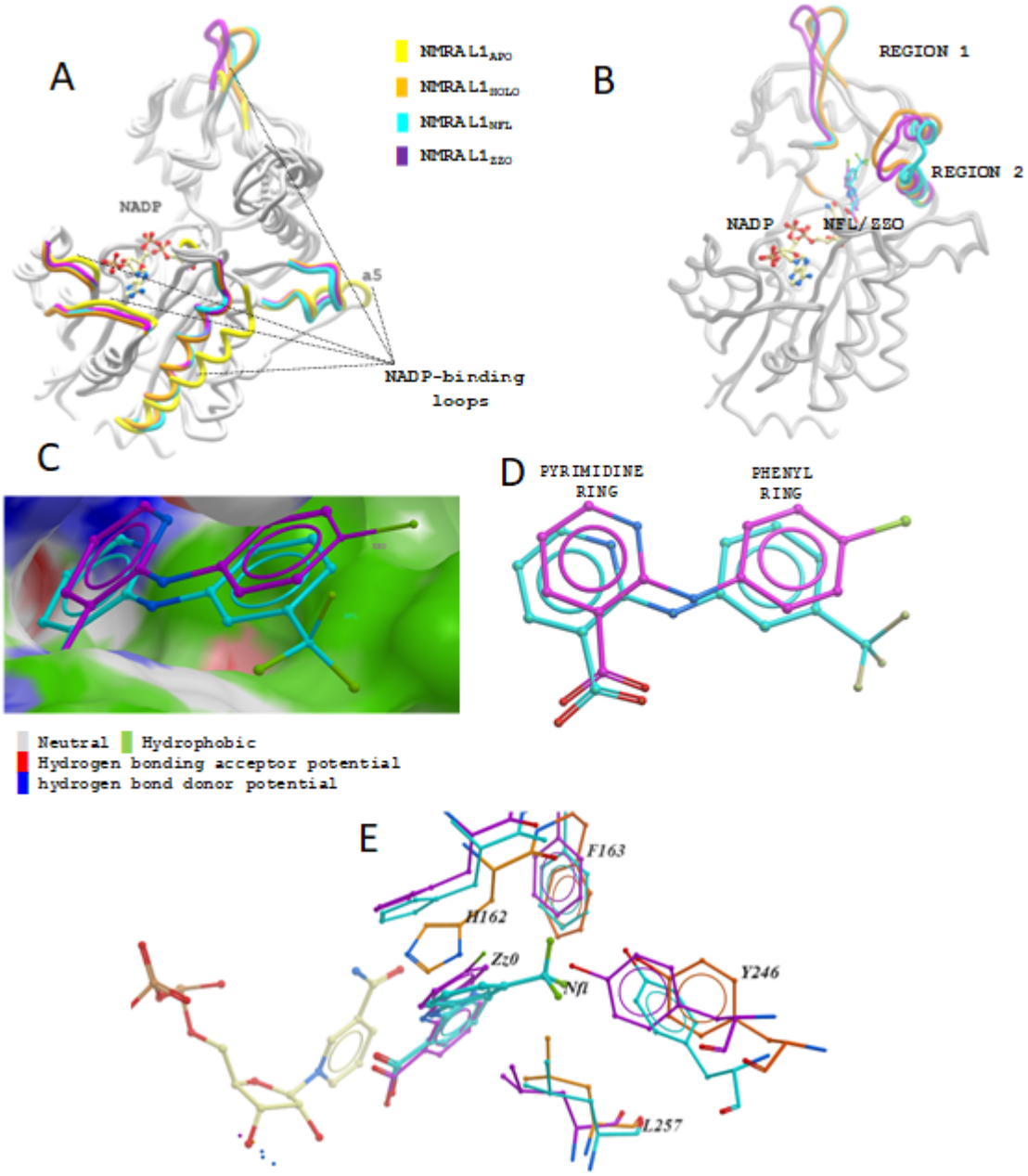


Figure 6.3.3.1 NMRAL1 structural form comparison – A) $NMRAL1_{NFL}$ and $NMRAL1_{ZZO}$ vs. $NMRAL1_{HOLO}$ B) $NMRAL1_{NFL}$ and $NMRAL1_{ZZO}$ vs. $NMRAL1_{APO}$ C) ligand pocket surface analysis - Neutral (grey) Hydrophobic (green), Hydrogen bonding acceptor potential (red), hydrogen bond donor potential (blue) D) NFL and ZZO active site orientation overlap E) NMRAL1 ligand binding pocket with NFL and ZZO ; NMRAL1 residues, NADP, NFL and ZZO represented in a ball-and-stick format. Holo (orange), apo (yellow), NFL bound (cyan), ZZO bound (purple).

6.3.4 OLIGOMERIC STATE OF NMRAL1

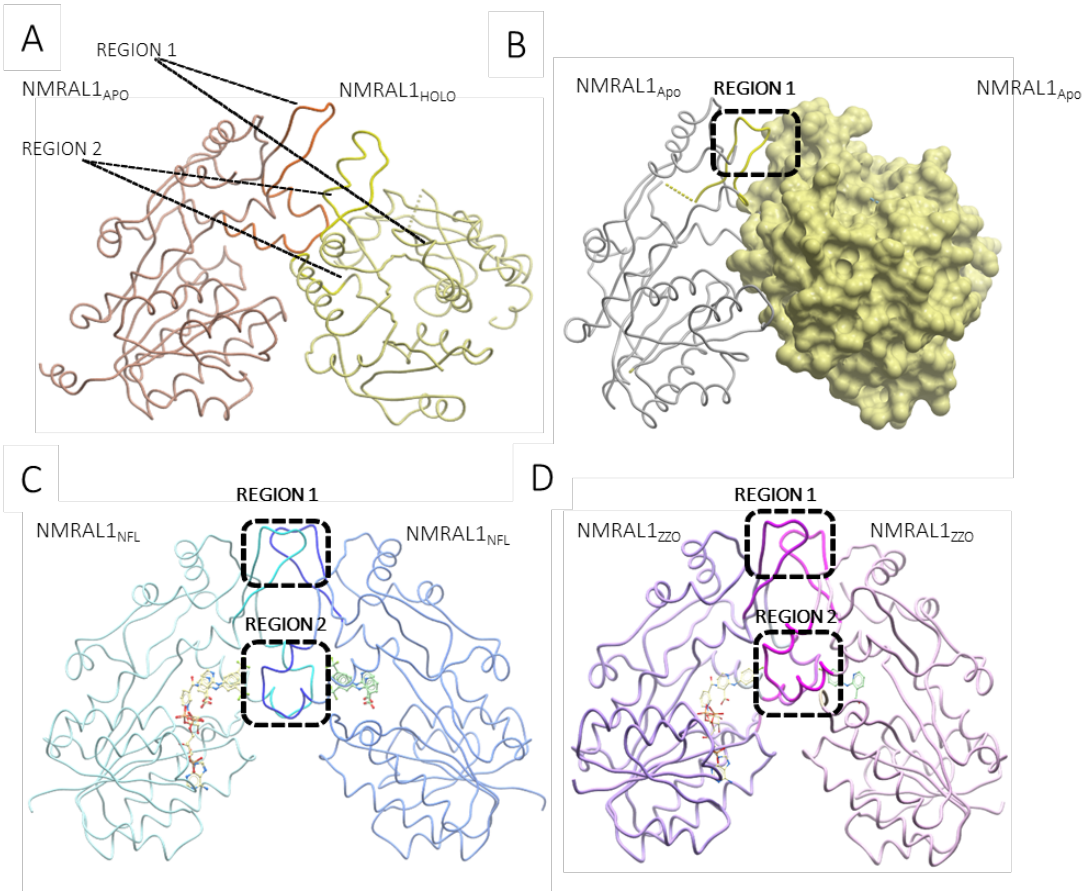


Figure 6.3.4.1 NMRAL1 dimer formation in the crystal - (A) $NMRAL1_{APO}$ (yellow) and $NMRAL1_{HOLO}$ (orange), (B) $NMRAL1_{APO}$ dimer, (C) $NMRAL1_{NFL}$ dimer (D) $NMRAL1_{ZZO}$ dimer. NMRAL1 represented in protein worm diagrams, NADP, NFL and ZZO represented in a ball-and-stick.

Subunit interaction in the previously-observed asymmetric dimer of NMRAL1 involves a major interface contributed specifically from regions 1 and 2 of both the *holo* protomer (Figure 6.3.4.1A, orange) and the *apo* protomer (Figure 6.3.4.1A, yellow). Docking two $NMRAL1_{APO}$ protomer onto each other to form a homodimeric assembly mode results in steric clashes in the interface region of each protomer (Figure. 6.3.4.1 B). The asymmetric dimer assembly is not observed in either of the ligand bound complexes; instead crystal packing shows potential

homodimers by 2-fold symmetry related subunits of $NMRAL_{NFL}$ and $NMRAL1_{ZZO}$ (Figure 6.3.4.1 C-D), where the mode of dimer assembly involves regions 1 and 2 from each protomer, and differs from the previously observed asymmetric dimer (Figure 6.3.4.1 A).

Analysis of the two interacting protomers within the $NMRAL1_{NFL}$ and $NMRAL1_{ZZO}$ structures by the PDBePISA server (Krissinel and Henrick 2007) suggests that a $NMRAL1_{NFL}$ protomer is unlikely to dimerize and may not be stable in solution. $NMRAL1_{NFL}$ dimer in the crystal has a modest buried surface area, a negative ΔG^{diss}, but a very low complexation significance (CSS) score that suggest a monomeric state in solution (Table 6.3.4.1).

NMRAL1	Stable in Solution	Surface Area sq. Å	Buried Area sq. Å	ΔG^{int}	ΔG^{diss}	CSS (%)
Apo: Holo	YES	25850	4250	-8.4	3.5	0.613
NFL:NFL	NO	23450	7820	-32.6	-0.6	0.0027
ZZO:ZZO	YES	23930	5730	-26.1	5.8	0.351

Table 6.3.4.1 – PDBePISA NMRAL1 dimer formation in solution

To further investigate the oligomeric state of $NMRAL1_{NFL,}$ 3 preparations of NMRAL1 namely as-purified (no addition of NADP during purification), NADP saturated, Niflumic acid and NADP saturated were tested in analytical gel filtration (Superdex 75 10/30 column) studies. All preparations showed a monomeric peak around 29KDa with $NRMAL1_{NFL}$ and a lower unidentified peak around 12.4 KDa. Interestingly, analytical gel filtration suggests a monomeric form for not only for $NMRAL1_{NFL}$ in solution, but all forms tested and shows no peak for a dimer in solution Figure 6.3.4.2). It might be that the aysmetric dimer is a weak dimer in solution and perhaps it may only form when only <50 % of NMRAL1 is saturated with NADP.

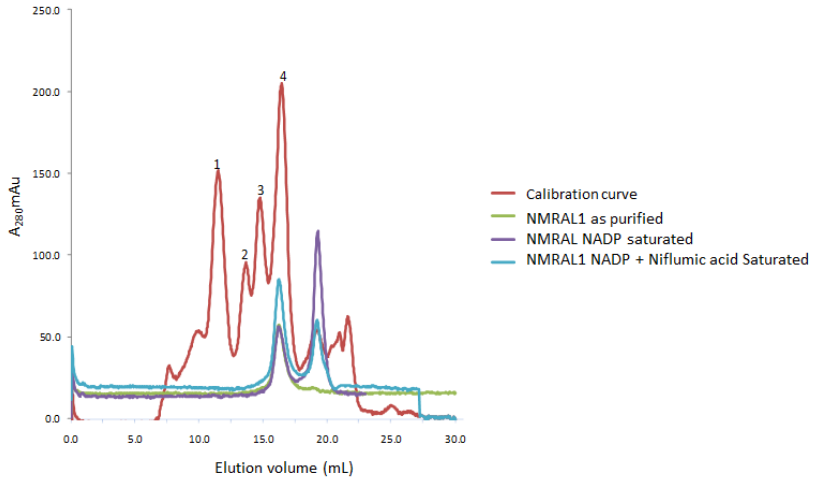


Figure 6.3.4.2 NMRAL1 analytical gel filtration (Superdex 75 10/30) – Calibration curve: 1) Alcohol dehydrogenase (150 kDa) 2) BSA (66 kDa – monomer) 3) chick egg albumin (45 kDa) 4) bovine erythrocytes carbonic anhydrase (29 kDa).

6.4 CRYSTAL FORMATION OF HSDL2 AND DHRS2 USING AN IDENTIFIED COMPOUNDS FROM DSF

From the previous section it is clear that specific compounds that produce a T_m shift have the potential to bind specifically to a protein, and crystallography can reveal its binding mode. In some cases, identification of suitable conditions and ligands via DSF has been used to increase the success of crystallisation of novel proteins, for example, in the case of pyridoxal phosphate (PLP)-dependent enzymes and protein kinases (Ericsson et al. 2006, Vedadi et al. 2006, Phillips and de la Pena 2011, Geders et al. 2012).

HSDL2 and DHRS2 are SDRs located within cluster C1 of the superfamily diagram and both lack structural data. HSDL2 has been suggested to be involved in fatty acid metabolism due to its peroxisomal localisation and contain a sterol carrier protein 2 (SCP2) domain (Kowalik

et al. 2009) (Figure 6.4.1.), whereas, DHRS2 is a known dicarbonyl reductase involved in the inactivation of reactive carbonyls (Shafqat et al. 2006). Initial efforts to crystallise either HSDL2 or DHRS2 with and without the presence of their cofactor NADP/H proved unfruitful. Consequently, potential ligands identified by DSF identified were tested in crystallisation trials in an endeavour to enhance crystal formation.

6.4.1 CRYSTALLISATION TRIALS OF DHRS2B AND HSDL2

DHRS2B generates a T_m shift with palmitoleyl-CoA in the presence of NADP (Table 5.3.1.1), which does not resemble other DHRS2 substrates and thus presented an interesting candidate for crystallisation trials. Similarly, DSF identified estradiol and 22-hydroxycholesterol (Table 5.3.4.1) as potential ligands to test in crystallisation trials with HSDL2. Prior to crystallisation, each protein sample was supplemented with 1.5 mM NADP/H and 0.75 mM of the corresponding compound (Table 6.4.1.1).

DHRS2B crystals were formed in coarse screen conditions after 5 days and were cryo-protected with 25 % glycerol before flash freezing in liquid nitrogen (Figure 6.4.1.1A). The crystals were mounted and generated a diffraction pattern to an estimated 3.3 Å resolution at the Diamond Light source beamline I24. However, diffraction data was not collected, as the quality criteria for collection set by the SGC is 2.8 Å. Further attempts to improve diffraction quality were carried out by setting up grid optimisation and seeding experiments. Grid optimisation plates were setup by varying the Bis-TRIS pH and PEG3350 concentration. Seeding experiments comprised of mixing equal amounts of protein solution and the reservoir solution where the crystal had formed. After one-day incubation, the drops were

micro-seeded from the original crystal hits. All crystals formed from optimisation and seeding experiments, were mounted and cryoprotected in the same manner as the original crystal. However, the all the crystals tested did not improve the diffraction quality.

HSDL2 crystals only formed in the presence of estradiol and NADP at 20 °C (Figure 6.4.1.1B). The addition of 22-hydroxycholesterol caused the HSDL2 protein solution to precipitate, which may be attributed by the high concentration of ~10 % DMSO present in original stock solution required to dissolve the compound.

Protein Conditions	Drop Status	Crystalisation Condition	Drop Ratio
DHRS2B + NADPH	No precipitation	N/A	N/A
DHRS2B + NADP	No Precipitation	N/A	N/A
DHRS2B + NADP + palmitoleyl-CoA	No Precipitation	0.1BIS M BIS-TRIS pH 5.5, 25 % PEG 3350	2:1
HSDL2 + NADPH	No precipitation	N/A	N/A
HSDL2 + NADP	No Precipitation	N/A	N/A
HSDL2 + NADP + estradiol	No Precipitation	0.1BIS M BIS-TRIS pH 5.5, 25 % PEG 3350	2:1
HSDL2 + NADP + 22-hydroxycholesterol	Preciptated	N/A	N/A

Table 6.4.1.1 DHRS2B and HSDL2 crystallisation trials with cofactor and DSF identified compounds summary

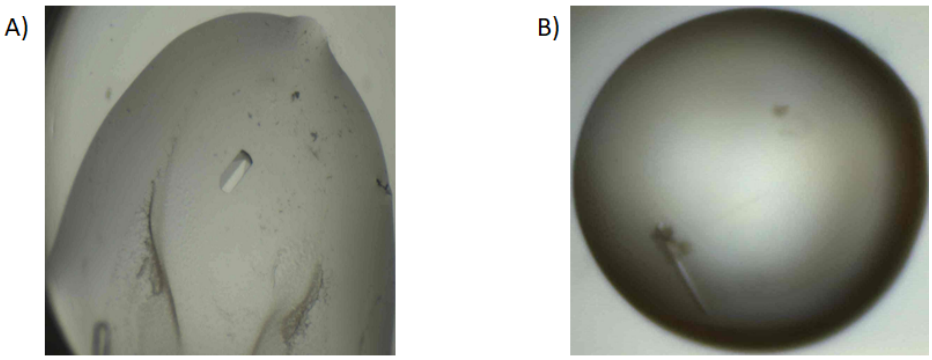


Figure 6.4.1.1 Crystals of A) DHRS2B and B) HSDL2

The structure of human HSDL2 was solved by molecular replacement using the putative dehydrogenase structure from Xanthomonas campestris (PDB ID 3E03) as search model. The model was refined to a resolution of 2.25 Å (R_{FACT} 0.202, R_{FREE} 0.247); Refinement statistics are summarised in Table 10, Appendix.

6.4.2 OVERALL HSDL2 STRUCTURE

The dimeric HSDL2 structure shows the highly conserved α/β Rossmann core (residues 12–196) with a parallel β-sheet (strands β1 to β6) interconnected by α-helices (α2 to α6). An additional eighth helix (α8) and seventh strand (β7) are contributed by the C-terminal region to form a relatively long C-terminus that would adjoin the SCP2 domain of HSDL2, (Figure 6.4.2.1B). The dimer interface is coordinated by the Asp119 (Chain A) side chain forming two hydrogen bonds to the amines of Arg127 (Chain B) and vice versa (Figure 6.4.2.1C).

The HSDL2 structure contains one molecule of NADP in each subunit that resides in a pocket at the core Rossmann fold formed by the conserved TGxxGxG motif (Figure 6.4.2.1A). The positively charged residue Arg20 around the adenine ribose in the NADP complements binding because of its negatively charged phosphate substituent at the O2’ adenine atom (Figure 6.4.2.1D). Residues Arg20, Ala41, Lys42 and Thr43 form hydrogen bonds to the ribose phosphate group, while Asp74 and Val75 hydrogen bond to the amino group of adenine moiety. Tyr168 and Lys172 typically interact with the 2’ and 3’-OH group of the nicotinamide ribose. Ile201 and Ile22 form a hydrogen bond to the carbonyl 7’=O of the nicotinamide and 1’=O of the phosphate group, respectively.

The co-crystallisation of HSDL2 with NADP and estradiol shows clear density for NADP but not estradiol (Figure 6.4.2.1A). To understand if estradiol could be accommodated in an

SDR, we analysed the structure of HSD17β1, which was co-crystallised with NADP and estradiol (1FDT, Breton et al., 1996). Superimposition of the HSD17β1 active site with that of HSDL2 reveals that Ser152 (numbering from HSDL2) is not in the correct position to interact with an estradiol molecule. In estradiol free structures of HSD17β1, the Ser152 is located in the same position, implying that the current HSDL2 model does not harbour an estradiol molecule. Active site analysis for HSDL2 is further discussed in Section 6.5.

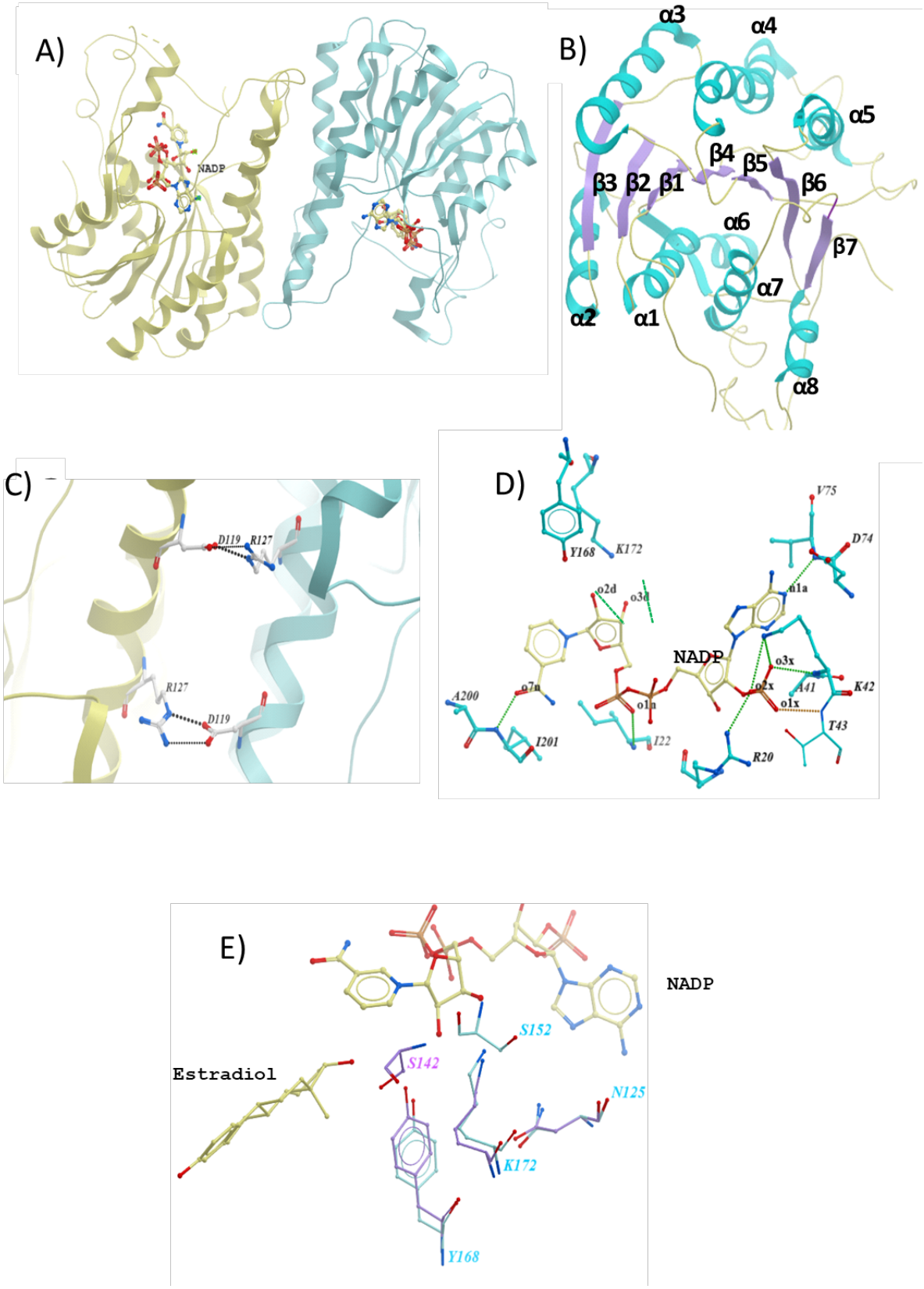


Figure 6.4.2.1 HSDL2 structure in complex with NADP (PDB ID: 3KVO) - A) HSDL2 dimer overview: chain A (yellow), chain B (cyan) B) Overall Rossmann fold - β-sheet (purple), α-helices (blue) C) Hydrogen bonding in dimer interface D)NADP coordination at the active site E) active site overlay of HSDL2 (cyan lines) and HSD17β1 (purple lines) bound with estradiol (PDB ID: 1FDT).

6.5 SUBSTRATE SCREENING FOR ORPHAN SDRS

Following the structural determination of HSDL2, which confirmed the binding of NADP/H as determined by DSF (Figure 4.5.2C), substrate screening for two SDRs (HSDL2 and DHRS11) were selected for substrate screening by DSF using the ligand library (Chapter 5). Both HSDL2 and DHRS11 are orphan SDRs that retain the catalytic tetrad and have had their crystal structures determined by the SGC. At present, there is no extensive screening data to show substrates of HSDL2 or DHRS11. Here, we investigate the application of substrate classes identified by DSF to aid functional characterisation of orphan SDRs.

6.5.1 ACTIVITY SCREENING RESULTS

Human HSDL1 and DHRS11 produced no change in NADP/H florescence at (λ_{ex}/ λ_{em} = 340/450 nm) (as described in Section 2.11) to imply enzymatic turnover towards any of the tested compounds within DSF identified substrate classes.

SDR	Substrate classes tested	Number of compounds within group	Activity Observed
HSDL2	Steroids	53	NO
	Fatty acids and Lipids	35	NO
DHRS11	Carbonyl Xenobiotics	20	NO

Table 6.5.1.1 Substrate class screening results of HSDL2 and DHRS11

6.5.2 HSDL2 AND DHRS11 ACTIVE SITE ANALYSIS

The unsuccessful decryption of specific substrates prompted the analysis of active site orientation in HSDL2 and DHRS11 from their crystals structures, to deduce if enzymatic turnover was plausible. Most of the catalytic tetrad in DHRS11 and HSDL2 superimpose

well with characterised SDRs (Figure 6.5.2.1A). The nicotinamide moiety of NADP in both HSDL2 and DHRS11 is positioned close to the catalytic tetrad, potentially competent for catalytic activity (Oppermann et al. 1997, Filling et al. 2002, Kallberg et al. 2002b) (Figure 6.5.2.1A). The upstream catalytic residues in HSDL2 and DHRS11, Ser152 and Asn125, are in the correct position relative to Tyr168 on helix α5 (numbering from HSDL2). In addition, Lys172 forms hydrogen bonds with the nicotinamide ribose and Asn125 interacts with the main chain amide of Ser152 on α5, thus producing the characteristic kink in α4 (Figure 6.5.2.1B).

Many SDRs bind a water molecule between Asn128 and Lys721 as part of a proton relay for catalytic activity, e.g. in 3β/17β hydroxysteroid dehydrogenase from *Comamonas testosterone* (Filling et al. 2002). A bound water molecule is observed in DHRS11 (Figure 6.5.1.1B – orange) but not in HSDL2. The nearby water molecule in HSDL2 can only coordinate one oxygen on Asn125 and is too far (>3.5 Å) to form hydrogen bonds to Lys172 to provide a proton relay mechanism. Closer inspection of the active site reveals Ser152 in HSDL2 has a different orientation compared to the other SDRs. As mentioned for DHRS4 (section 6.2.4), a water molecule mimicking the substrate is not found to coordinate the residues Tyr168 and Ser152 in DHRS11 and HSDL2 (Figure 6.5.2.1C).

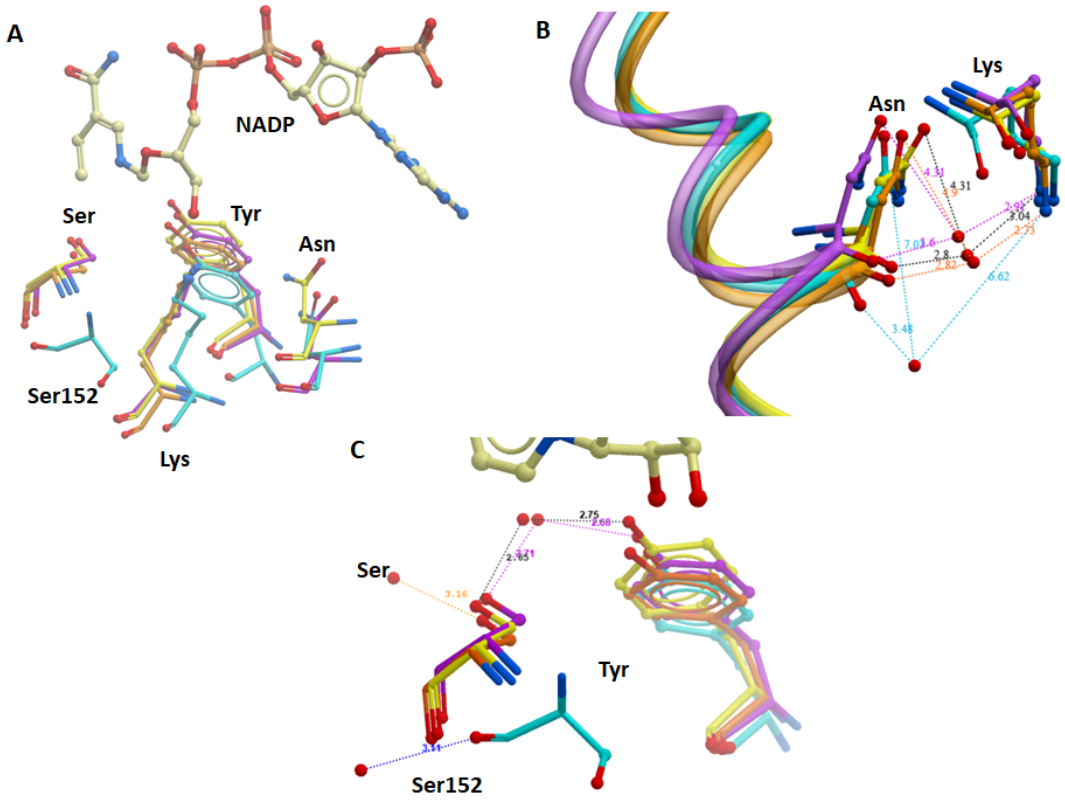


Figure 6.5.2.1 Active site coordination of HSDL2 and DHRS11 – A) superimposition of catalytic tetrad of HSDL2 (cyan),DHRS11 (purple), DHRS4 (purple) and HPGD (yellow) B) water molecule coordination by Lys and Asn in where C) water molecule coordination by Ser and Tyr; HSDL2 (cyan), DHRS11 (orange), DHRS4 (purple) and HPGD (yellow), bond distances (dotted lines)

6.6 DISCUSSION

The primary applications of DSF studies with recombinant proteins are to identify ligands (e.g. activators and inhibitors) for a particular protein (Amaning et al. 2013, Jeankumar et al. 2014), and to compare the stability of mutant variants for proteins (Lyu et al. 2014, Megarity et al. 2014). This chapter highlights different ways the application of DSF data can be adopted downstream of purification to structurally characterise and elucidate the human SDR family. The majority of SDR crystal structures deposited in the PDB are co-crystallised with their appropriate cofactor. As seen in Chapter 4, cofactor binding creates a more

thermally stable SDR protein by forming number of α-helices (Nakamura et al. 2006). The cofactor preference for DHRS4 was identified by DSF, distinguishing between the phosphorylated (NADP/H) and unphosphorylated (NAD/H) forms. DHRS4 crystal trials showed granular precipitation and crystal formation only with the addition of the reduced form of the cofactor (NADPH), which is observed bound to the active site in the crystal structure (section 6.2.4). The ability of DSF to differentiate thermal stability between cofactor redox states (e.g. DHRS4) can be used to pinpoint optimal conditions for use in crystallisation experiments or to prioritise cofactor addition when protein quanity is limiting. Analysis of other human SDR crystal structures (determined from SGC and non-SGC) shows that 12 out of 17 SDRs use the same cofactor and redox state as determined by DSF for crystal formation (Table 6.6.1). Only two SDRs, namely CBR3 and HSD17β8, have been crystallised in the presence of the oxidised form of cofactor, as opposed to the DSF determined reduced form.

SDR ID	Cofactor Preference	Cofactor redox state		PDB ID	Reference
		DSF determined	Crystal formation		
BDH2	NAD/H	Ambiguous	NADH	2AG5	(Guo et al. 2006)
CBR1	NADP/H	NADP	NADP	3BHI	(Bateman et al. 2008)
CBR3	NADP/H	NADPH	NADP	2HRB	(Pilka et al. 2009)
DCXR	NADP/H	NADPH	NADPH	1WNT	(El-Kabbani et al. 2004)
DHRS4	NADP/H	NADPH	NADPH	3OQR	
DHRS11	NADP/H	NADPH	NADPH	1XG5	http://www.thesgc.org/structures/1XG5#mandm
GALE	NAD/H	Ambiguous	NAD	1EK5	(Thoden et al. 2000)
HPGD	NAD/H	NADH	NADH	2GDZ	(Niesen et al. 2010)
HSD17β10	NAD/H	NADH	NADH	2O23	http://www.thesgc.org/structures/2O23#mandm
HSD17β4	NAD/H	Ambiguous	NAD	1ZBQ	http://www.thesgc.org/structures/1ZBQ#mandm
HSD17β8	NAD/H	NADH	NAD	2PD6	http://www.thesgc.org/structures/2pd6#mandm
HTATIP2	NADP/H	NADPH	NADPH	2BKA	(El Omari et al. 2005)
NMRAL1	NADP/H	NADP	NADP	2EXX	(Dai et al. 2006)
QDPR	NAD/H	NADH	NADH	1HDR	(Su et al. 1993)
SPR	NADP/H	NADPH	NADPH	1Z6Z	http://www.thesgc.org/structures/1Z6Z#mandm
SDR39U1	NADP/H	NADPH	NADPH	4B4O	http://www.thesgc.org/structures/4B4O#mandm
TSTA3	NADP/H	NADPH	NADPH	4E5Y	(Zhou et al. 2013)

Table 6.6.1 Human SDR crystal formation with cofactor redox states - red denotes where cofactor determined in DSF and used for crystallography agree.

DSF has also been employed to identify niflumic acid (NFL) as the first small molecule binder for NMRAL1, a regulatory SDR that has been implicated in the nitric oxide cycle (Section 6.3). This has proved that compounds with an increased T_m shift can be identified that bind to the proteins. The crystal structures of NMRAL1 in complex with NFL and ZZO, show a monomeric form of the protein that differs from the previously observed heterodimer (Figure 6.3.3.1A-B). The monomeric $NMRAL1_{NFL}$ structure does not appear to be a crystallographic artefact (Figure 6.3.4.1C, Table 6.3.4.1), as it is unlikely to form a dimer in solution based on docking analyses and analytical GF studies (Figure 6.3.4.1-2). The data shows that the pyridine-phenyl ring in NFL and ZZO represent structural scaffolds as starting point for structure-activity relationships (SAR) to develop selective ligands. Both ligands stack with the nicotinamide ring of NADP that blocks the dimer interface and a possible association with ASS1 due to movement of His162 (Figure 6.3.3.1E), which is the proposed interaction zone (AA 153-189) between ASS1and NMRAL1 (Zhao et al. 2008). This suggests the two compounds could have the potential to inhibit the interaction between ASS1 and NMRAL1. If NFL and ZZO affect the functional associations, the scaffolds could be used as a novel way to observe the effects of ASS1 regulation in the nitric oxide cycle.

HSDL2 cofactor preference was also determined by the slope of transition (Section 4.5) in DSF, however, unlike DHRS4, cofactor addition did not yield any crystals. Addition of estradiol and NADP showed crystal formation and allowed diffraction data to be collected, although no electron density was observed for estradiol. Structural superimposition of HSDL2 and HSD17β1 complexed with estradiol and NADP also suggested that no estradiol was present as the active site serine in HSDL2 was not in the correct orientation, as in

HSD17β1, to bind estradiol as potential substrate. The parallel setup of the crystallisation experiments may suggest that estradiol merely helped to create a supersaturated state by enhancing the thermal stability of the protein in solution. It is possible that HSDL2 crystal formation was a stochastic event independent of the presence of estradiol, but there are many other protein structures where crystal formation occurs in the presence of ligand but no electron density is observed, for example, α1-microglobulin (Meining and Skerra 2012), cholesterol 7α-hydroxylase (Tempel et al. 2014) and malonyl-coenzyme A decarboxylase (Froese et al. 2013). A similar phenomenon occurs with DHRS2B, where it resists the crystallisation process by addition of cofactor alone. Only the addition of palmitoleyl-CoA in the presence of NADP causes DHRS2B crystal formation. The nature of DHRS2B crystal formation is not stochastic as the coarse screen and follow-up plates were repeated several times with similar crystals forming exactly after 5 days. This gives plausibility to the idea that DSF can be used to identify ligands that aid crystal formation.

The second part of this chapter attempted to investigate substrate screening for two orphan SDRs, HSDL2 and DHRS11 from idenitfied substrate classes in Chapter 5. No compounds within the identified substrate classes from DSF showed any sign of reduction or oxidation activity by measuring NADPH fluorescence. Several factors may attribute to not elucidating specific substrates for either HSDL2 or DHRS11: (1) An untested steroid, fatty acid, lipid or carbonyl xenobiotic may yet result in enzymatic activity. (2) Only a partially soluble form of HSDL2 has previously been screened and has shown no activity towards either steroids or retinoids (Kowalik et al. 2009). (3) Enzyme activity of both SDRs may be sensitive to DMSO concentrations and compounds could be better dissolved in an alternative solvent, such as

methanol to rule out potential DMSO effect. (4) Although the active site tetrad in HSDL2 superimposes well with other catalytic SDRs, the unusually orientated active site Ser152 does not allow coordination to the Tyr168 via a nearby water molecule. This could be because Ser152 is located 16 residues upstream of the active site Tyr in HSDL2, whereas the catalytic Tyr in most SDRs is situated 13-14 resdiues upstream. (5) It is also plausible that both SDRs may adopt a regulatory role, as in the case of MAT2B, despite having a catalytic tetrad poised for enzymatic turnover (Shafqat et al. 2013). In the future, coimmunoprecipitation experiments on cells that contain endogenous HSDL2/DHRS11 could detect any target proteins that they may interact with to conclude if they do possess a regulatory function.

In summary, it is clear that a number of human SDRs are more prone to crystallise with ligands that produce a significant increase in protein unfolding temperature (T_m). Routine use of DSF can help to identify cofactors and ligands that bind to the protein and aid the characterisation of proteins by crystallisation.

CHAPTER 7 – CONCLUSIONS AND FUTURE WORK

7.1 CONCLUSIONS

Since the identification of SDRs in the 1980s, the protein family has received increasing interest from the scientific community. The majority of SDR studies either focus on the mechanistic, or therapeutic themes of individual members or types (e.g.17β- and 11β-HSDs), or are larger studies that involve evolutionary analysis of the SDR family (Belyaeva et al. 2014; Kallberg et al. 2002a). The single and multi-membered SDR studies have left a substantial gap in regards to the systematic analyses of substrate and cofactor specificities, subcellular localisation, and inhibitor identification. This thesis has endeavoured to combine bioinformatics and experimental methods (protein purification, DSF, crystallisation) to further characterise the SDR protein family. This study also provides a methodical breakdown into the applications and limitations of DSF in characterising the SDR protein family.

7.1.1 PROTEIN PRODUCTION

The first step in this study was to conduct parallel testing of multiple constructs for each SDR target for expression analysis for large-scale purification. This has allowed the identification of SDRs that can be expressed as stable, soluble proteins in relatively high yields for characterisation by DSF. Of the 63 human SDRs cloned successfully, 28 produced soluble proteins in high yields for downstream processing (Table 2.8.1). Most of the

recombinant purified SDRs came from *E. coli* expression, representing all clusters, diverse metabolic functions, and different degrees of functional or structural characterisation.

7.1.2 SDR FAMILY ANNONATION

Following protein production a comprehensive annotation of the human SDR family by analysis of relationships between cofactor preferences, substrate class, and subcellular localisation provides an overview of the functional distributions across the family, evolutionary traits, putative functional assignment of orphan SDRs and how SDRs make up a living cell (Figure 3.7.4). Cofactor analyses show that human SDRs have adopted an overall increased specificity towards NADP/H (Table 3.2.1). Evolution of subcellular compartments has subjected proteins to their specialised environments, where the availability of cofactor(s) such as NAD(P) is regulated by set of enzymes that interconvert between the forms of NAD(P) (Figure 7.1.2.1, Agledal et al. 2010). Cofactor analyses shows that certain subcellular compartments, e.g. the cytoplasm, favours NADP/H for human SDRs, which is likely to be linked to the low cytoplasmic redox potential of NADP compared to NAD that lends itself to allow for more reduction of molecules such as carbonyl xenobiotics (Krebs and Veech 1969). Most NADP-dependent SDRs found in the cytoplasm (e.g. CBR1, DCXR, BLVRB) are reductases, whereas NAD-dependent SDRs in the cytoplasm are dehydrogenases.

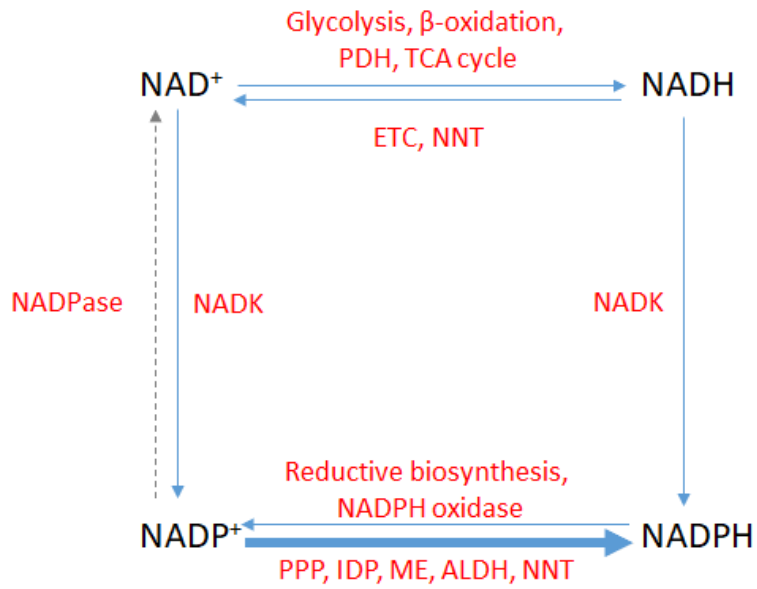


Figure 7.1.2.1 Metabolism of NAD(P) in eukaryotic cells. NADPase (NAD Phosphatase); PDH (pyruvate dehydrogenase); TCA cycle (tricarboxylic acid); ETC (electron transport chain); NNT (nicotinamide nucleotide transhydrogenase); NADK (NAD kinase); PPP (pentose phosphate pathway); IDP (isocitrate dehydrogenase); ME (malic enzyme); ALDH (aldehyde dehydrogenase). This figure was adopted from (**Agledal et al. 2010**).

Another feature that can be observed is the preference of cofactors in some SDR substrate classes. In particular, SDRs with non-enzymatic functions all show NADP-dependence. Both NAD(H) and NADP(H) are involved in a variety of metabolic processes including energy production, but have emerged as signalling molecules that regulate transcription, homeostasis, and apoptosis (Fjeld et al. 2003; Pollak et al. 2007). Interestingly, regulatory or non-enzymatic functions have evolved after the establishment of the cofactor binding sites; as for example, HSDL1 has a lineage associated with steroid activity (Figure 3.7.1).

Sequence cluster analysis shows there are dominant substrate classes and subcellular compartments within each cluster (Table 3.7.2-3). For instance, SDRs with catalytic activity

towards sugars are restricted to the extended cluster and ER localisation is predominant in clusters C2 and C3. Subcellular localisation can be observed at the sequence level through signalling motifs and TM helices in over 60 % of all human SDR members, with the majority located in the N-terminus (Figure 3.4.3.1A). Of particular note are human SDRs with TM helices, which tend to localise to the ER membrane. Signal motifs and TM helices represent good indicators of compartmentalisation and the ability to predict subcellular localisation based on these factors plays an important role in implicating protein function (Section 3.4-3.5). From applying the noticeable features throughout the human SDR family, a systematic flowchart for pinpointing potential substrate classes for orphan SDRs is presented (Figure 3.7.3). Each step shows the key features to consider (active site motif, cofactor preference, sequence cluster, TM helices, targeting signals and subcellular localisation), thus providing a reference point for assigning substrate class candidates.

7.1.3 SDR COFACTOR AND LIGAND SCREENING

Following family annotation, the evaluation of cofactor determination via DSF correctly determined 82 % cofactor preference (Table 4.4.1). Those that do not show any distinction between NAD(H) and NAD(P) through DSF may have been pre-saturated with cofactor through expression in *E.coli*. Only thermal shift data for TSTA3 conflicted with sequence analysis prediction, likely due to the ambiguity in the key residue that determines NAD(H) or NADP(H) binding.

When DSF curves diverge from the typical two-state transition, extrapolation of T_m shift values from these curves may not represent the correct cofactor or ligand. The gradient of thermal unfolding curves has often been associated with stability (Niesen et al. 2007), and is a good secondary measure to identify hits. This has proven successful to determine cofactor

preferences for SDRS for the first time (Section 4.5). Proteins that show small unfolding gradient (ΔI/Δ°C) values point towards less thermally stable proteins, which suggests the protein may be partially unstructured. Hydrophobic residues in the protein could be more accessible to the dye, which results in high intential intensities and a negative pre-transition slope. Cofactor binding in SDRs induces the formation of α-helices, which would explain why this phenomenon can be seen in the case of DHRS11, DHR4 and HSDL2. Thermal unfolding gradients are infrequently used as a validation tool and can be often overlooked. For example, in many instances irregular transitions may be discarded as 'bad data' and using pre-defined analysis software often results in the loss of raw DSF curve plots in final data outputs (e.g bar charts of T_m).

DSF can distinguish redox states of cofactors, which may be crucial in the use of the appropriate forms for inclusion in crystallisation, as shown for the structural elucidation of DHRS4 (Section 6.2) that requires the presence of NADPH, but not NADP, for crystal formation. Therefore knowledge of the most thermostable redox state can help prioritise crystallisation studies, especially when purified protein is in limited supply for crystallisation trials. Together with T_m shift and slope of transition values, the decryption of cofactor preference for six orphan SDRs was accomplished, five of which have all been subsequently verified in crystal structures (DHRS1, HSDL2, DHRS7B, DHRS11, SDR39U1).

DSF assays primarily identify compounds that can promote protein purification, crystallization, and/or discover small molecule inhibitor scaffolds. Here, the efficacy of DSF

as a screening method was evaluated by the identification of known SDR substrate and inhibitors. Less than 15 % of SDR substrates tested showed an increase in thermal stability (Section 5.3.1). Naturally enzymes will turnover substrates in the reaction mixture and, unlike inhibitors, usually do not form stable enzyme-bound complexes. Interestingly, SDRs that perform non-EC 1.1 reactions all exhibit thermal shifts with their respective substrates (Table 5.8.1).

Many SDRs display thermal shifts with compounds that resemble their substrate and can therefore point toward the correct substrate class that the enzyme belongs to, as illustrated by PECR (Table 5.3.1.4) and HSD17β10 (Table 5.3.1.6). For the screened human SDRs with known enzymatic activity, more than 70 % showed a change in T_m (+1.4 C) when screened against a compound in the respective substrate class (Figure 5.3.2.2). Thermal shifts can therefore infer potential substrate class for orphan SDRs and compare well with those from computational prediction (Table 5.8.2). The two orphan SDRs, HSDL2 and DHRS11 that retain the residues for enzymatic turnover, revealed no catalytic activity when tested against compounds from their predicted substrate class (Section 6.5). However, HSDL2 and DHRS11 active site analysis from their crystal structures shows the inability of both these SDRs to form a proton relay network that is crucial to catalytically active SDRs and may imply a regulatory role (Figure 6.5.2.1).

Unlike substrates, inhibitors demonstrate increased thermal stability in DSF experiments and generally attain higher T_m shifts than substrates. For example, HSD11β1 does not

exhibit any thermal shifts with any of its glucocorticoids or xenobiotic substrates. However, a significant shift is observed with HSD11β1 inhibitors in the presence of CBX or glycerrhitinic acid. This ligand binding is further supported by the crystal structure of human HSD11β1 in the presence of CBX (Figure 5.8.3).

The use of small molecule ligands identified from DSF to assist in co-crystallization is well established (Niesen et al. 2007), and has contributed to the success in elucidating ligands to aid crystal formation of human SDRs in this study. Compounds identified for SDRs that have been recalcitrant to crystallization, have aided the formation of crystals, in the case of HSDL2 and DHRS2 (Section 6.4). However, the real power of DSF lies in its ability to readily detect inhibitor scaffolds. The discovery of potent inhibitor scaffolds for HPGD was found through identifying chemotypes by T_m shift and kinetic data (Section 5.5). The diversity in HPGD inhibitor scaffolds and selectivity provides a valuable tool for downstream functional cellular studies.

7.1.4 STRUCTURAL ELUCIDATION OF SDRS USING THERMAL SHIFT DATA

DSF data has been applied structurally to characterize not only enzymatic proteins, but also regulatory and orphan SDRs, as the assay is independent of substrate activity. For example, niflumic acid, a COX-2 inhibitor used in the treatment of muscle pain identified by DSF, binds NMRAL1 in a NADP(H)-dependent manner (Table 5.6.1). Structural elucidation of niflumic acid and its derivative 2-(chloro-phenyllamino)-nicotinic acid by crystallography has shown these to be small molecule binders for NMRAL1 (Section 6.3). This has also reiterated that compounds with an increased thermal shift can be identified and bind in the active sites of

proteins, as in the case of MAT2B and TSTA3. Small molecule binders identified in this study can be developed as probes for use in assays or used as a starting framework to aid structure based drug development, as many SDRs are implicated in various diseases.

7.2 FUTURE WORK

The research conducted in this thesis has created a good basis for further exploration of ligand screening to structurally characterize the SDR family. The crucial limiting factor for further characterization studies is protein production, particularly those SDRs with membrane associations. Membrane associated SDRs typically found in Cluster C2 and C3 are currently structurally underrepresented, many of which are associated in disease (Table 1.6.1) and would benefit from ligand screening programs. Large scale expression testing of membrane associated SDRs in insect cell expression and mammalian systems (e.g transient expression in HEK293) in combination with detergent screening and refolding techniques could potentially allow many non-characterized member to be purified. Not only has insect cell expression proved a successful method for a few of the SDR members in this study (e.g. HSD17β12 and WWOX), it has been applied to other published studies including DHRS3 (Lundová et al. 2015), HSD17β6 and SDR16C5 (Lee et al. 2009; Chetykrin et al. 2001). Mammalian cells that have been used to express SDRs are generally used in cellular studies and generally produce lower yields, which is costly for large-scale ligand screening and crystallization applications. The recent advent of baculovirus-infected mammalian HEK expression (Elegheert et al. 2018) could significantly change this.

Although DSF has proved successful in ligand and inhibitor profiling, it has proved difficult to pinpoint specific substrates, particularly for the decryption of orphan SDRs. Orthogonal biophysical assays such as, SPR, DSC and ITC could be used to clarify binding parameters as a means to determine specific substrate for SDRs. Site-directed mutagenesis of active site residues could represent another approach to identify specific SDR substrates through DSF. Ideally, mutation/s in the active site tetrad would prevent enzymatic turnover, but still allow the SDR to coordinate a substrate molecule. In other words, the mutations would allow the substrate to behave similarly to an inhibitor where it produces a thermal shift in DSF. For those SDRs suspected to have regulatory roles, coimmunoprecipitation assay (Sambrook and Russell 2006), proximity ligation assay (Bagchi et al. 2015) in cell lines that endogenously express the target SDR experiments would confirm any possible protein partners *in vivo*. Another system to detect unknown protein partners is a technique called next generation sequencing yeast two hybrid (NGS-Y2H) (Suter et al. 2015). This system exploits yeast genetic markers to detect direct binding partners *in vivo* and confirms partners via next generation sequencing.

For proteins where diffraction quality needs to improve before structural data can be collected (i.e. DHRS2), purification buffers could be altered to induce more favourable crystallisation conditions and improve crystal quality. In addition, there are many optimization techniques that could be employed, such as additive screening, limited proteolysis, seeding, and crystal dehydration. There are many other human SDRs without 3D-structures that resist crystallisation, a closer analysis of the construct design, purification and/or addition of compounds may aid the formation of crystals. Alternative methods for

structure determination, include cryo electron microscopy (cryo-EM) and isotope nuclear magnetic resonance (NMR) by isotope labelling could be employed. The monomeric SDRs (<30 kDa) are more suitable for NMR experiments and oligomeric (>2mer) SDRs are ameanable for cryo-EM, where >60 KDa is the lower limit. Cryo-EM, at present it is difficult to detect small molecules, as resolution often the limiting factor.

7.3 SUMMARY

To summarise, the work undertaken in this thesis has provided a framework for high-throughput production of many human SDR proteins, an in-depth annotation describing the distribution of cofactor preferences, subcellular localization and TM helices across the family, the utilization of DSF as a tool for cofactor screening, and the elucidation of potential ligands of SDRs. There are still many uncharacterised SDRs of great interest, especially those with disease associations. With new advances in protein production and structural methods, this workflow can be used to tackle other member's human SDRs family.

REFERENCES

**denotes citation in Appendix only.

Abdel-Salam, G., et al. (2014). "The supposed tumor suppressor gene WWOX is mutated in an early lethal microcephaly syndrome with epilepsy, growth retardation and retinal degeneration." Orphanet J Rare Dis 9: 12.

Abul-Husn N. S., X. Cheng, A. H. Li, Y. Xin, C. Schurmann, P. Stevis, Y. Liu, J. Kozlitina, S. Sten. der, G. C. Wood, A. N. Stepanchick, M. D. Still, S. McCarthy, C. O'Dushlaine, J. S. Packer, S. Balasubramanian, N. Gosalia, D. Esopi, S. Y. Kim, S. Mukherjee, A. E. Lopez, E. D. Fuller, J. Penn, X. Chu, J. Z. Luo, U. L. Mirshahi, D. J. Carey, C. D. Still, M. D. Feldman, A. Small, S. M. Damrauer, D. J. Rader, B. Zambrowicz, W. Olson, A. J. Murphy, I. B. Borecki, A. R. Shuldiner, J. G. Reid, J. D. Overton, G. D. Yancopoulos, H. H. Hobbs, J. C. Cohen, O. Gottesman, T. M. Teslovich, A. Baras, T. Mirshahi, J. Gromada and F. E. Dewey (2018). "A Protein-truncating HSD17B13 variant and protection from chronic liver disease." N Engl J Med 378(12):1096-1106.

Adam M., H. Heikelä, C. Sobolewski, D. Portius, J. Mäki-Jouppila, A. Mehmood, P. Adhikari, I. Esposito, L. Elo, F. P. Zhang, S. T. Ruohonen, L. Strauss, M. Foti and M. Poutanen (2018). "Hydroxysteroid (17 β) dehydrogenase 13 deficiency triggers hepatic steatosis and inflammation in mice." FASEB J 32(6):3434-3447.

Agarwal A. K. and R. J. Auchus (2005). "Minireview: Cellular redox state regulates hydroxysteroid dehydrogenase ativity and intracellular hormone potency". Endocrin 146(6):2531-2538.

Airenne, T. T., J. M. Torkko, S. Van den plas, R. T. Sormunen, A. J. Kastaniotis, R. K. Wierenga and J. K. Hiltunen (2003). "Structure-function analysis of enoyl thioester reductase involved in mitochondrial maintenance." J Mol Biol 327(1): 47-59.

Alano, A., S. Almashanu, J. M. Chinsky, P. Costeas, M. G. Blitzer, E. A. Wulfsberg and T. M. Cowan (1998). "Molecular characterization of a unique patient with epimerase-deficiency galactosaemia." J Inherit Metab Dis 21(4): 341-350.

Albiston, A. L., V. R. Obeyesekere, R. E. Smith and Z. S. Krozowski (1994). "Cloning and tissue distribution of the human 11 beta-hydroxysteroid dehydrogenase type 2 enzyme." Mol Cell Endocrinol 105(2): R11-17.

Alphey, M. S., W. Yu, E. Byres, D. Li and W. N. Hunter (2005). "Structure and reactivity of human mitochondrial 2,4-dienoyl-CoA reductase: enzyme-ligand interactions in a distinctive short-chain reductase active site." J Biol Chem 280(4): 3068-3077.

Altura, B. M. and B. T. Altura (1976). "Vascular smooth muscle and prostaglandins." Fed Proc 35(12): 2360-2366.

Agledal, L., M Niere., M Zeigler (2010). "The phosphate makes a difference: cellular functions of NADP." Redox Rep 15(1): 2-10.

Amaning, K., M. Lowinski, F. Vallee, V. Steier, C. Marcireau, A. Ugolini, C. Delorme, F. Foucalt, G. McCort, N. Derimay, C. Andouche, S. Vougier, S. Llopart, N. Halland and A. Rak (2013). "The use of virtual screening and differential scanning fluorimetry for the rapid identification of fragments active against MEK1." Bioorg Med Chem Lett 23(12): 3620-3626.

Aslanidis, C. and P. J. de Jong (1990). "Ligation-independent cloning of PCR products (LIC-PCR)." Nucl. Acids Res. 18(20): 6069-6074.

Atanasov, A. G., I. D. Ignatova, L. G. Nashev, B. Dick, P. Ferrari, F. J. Frey and A. Odermatt (2007). "Impaired protein stability of 11beta-hydroxysteroid dehydrogenase type 2: a novel mechanism of apparent mineralocorticoid excess." J Am Soc Nephrol 18(4): 1262-1270.

Bagchi S., Fredriksson R., Wallén-Mackenzie Å. (2015) In Situ Proximity Ligation Assay (PLA). In: Hnasko R. (eds) ELISA. Methods in Molecular Biology, vol 1318. Humana Press, New York, NY

Bagi, C. M., J. Wood, D. Wilkie and B. Dixon (2008). "Effect of 17beta-hydroxysteroid dehydrogenase type 2 inhibitor on bone strength in ovariectomized cynomolgus monkeys." J Musculoskelet Neuronal Interact 8(3): 267-280.

Bailey, M. A., J. M. Paterson, P. W. Hadoke, N. Wrobel, C. O. Bellamy, D. G. Brownstein, J. R. Seckl and J. J. Mullins (2008). "A switch in the mechanism of hypertension in the syndrome of apparent mineralocorticoid excess." J Am Soc Nephrol 19(1): 47-58.

Bakker, H., et al. (2009). "Functional UDP-xylose transport across the endoplasmic reticulum/Golgi membrane in a Chinese hamster ovary cell mutant defective in UDP-xylose Synthase." J Biol Chem 284(4): 2576-2583.

Barf, T., J. Vallgarda, R. Emond, C. Haggstrom, G. Kurz, A. Nygren, V. Larwood, E. Mosialou, K. Axelsson, R. Olsson, L. Engblom, N. Edling, Y. Ronquist-Nii, B. Ohman, P. Alberts and L. Abrahmsen (2002). "Arylsulfonamidothiazoles as a new class of potential antidiabetic drugs. Discovery of potent and selective inhibitors of the 11beta-hydroxysteroid dehydrogenase type 1." J Med Chem 45(18): 3813-3815.

Bateman, R. L., D. Rauh, B. Tavshanjian and K. M. Shokat (2008). "Human carbonyl reductase 1 is an S-nitrosoglutathione reductase." J Biol Chem 283(51): 35756-35762.

Becker, K. G., K. C. Barnes, T. J. Bright and S. A. Wang (2004). “The genetic association database.” Nat Genet 36(5):431-432.

Belenky, P., K. L. Bogan and C. Brenner (2007). "NAD^+ metabolism in health and disease." Trends Biochem Sci 32(1): 12-19.

Belayeva, O. V., S. A. Lee, O. V. Kolupaev and N. Y. Kedishvili (2009). "Identification and characterization of retinoid-active short-chain dehydrogenases/reductases in Drosophila melanogaster." Biochim Biophys Acta 1790(10): 1266-1273.

Benach, J., S. Atrian, R. Gonzalez-Duarte and R. Ladenstein (1999). "The catalytic reaction and inhibition mechanism of Drosophila alcohol dehydrogenase: observation of an enzyme-bound NAD-ketone adduct at 1.4 A resolution by X-ray crystallography." J Mol Biol 289(2): 335-355.

Benach, J., S. Atrian, R. Ladenstein and R. Gonzalez-Duarte (2001). "Genesis of Drosophila ADH: the shaping of the enzymatic activity from a SDR ancestor." Chem Biol Interact 130-132(1-3): 405-415.

Benach, J., S. Knapp, U. C. Oppermann, O. Hägglund, H. Jörnvall and R. Ladenstein (1996). "Crystallization and crystal packing of recombinant 3 (or 17) beta-hydroxysteroid dehydrogenase from Comamonas testosteroni ATTC 11996." Eur J Biochem 236(1): 144-148.

Bhatia, C., S. Oreum, J. Bray, K. L. Kavanagh, N. Shafqat, W. Yue and U. Oppermann (2014). “Towards a systematic analysis of human short-chain dehydrogenases/reductases (SDR): Ligand identification and structure-activity relationships.” Chem Biol Interact 234:114-125.

Bloch, K. (1975). "Some aspects of the control of lipid biosynthesis." Adv Exp Med Biol 60: 1-12.

Blum, A., C. Loerz, H. J. Martin, C. A. Staab-Weijnitz and E. Maser (2012). "Momordica charantia extract, an herbal remedy for type 2 diabetes, contains a specific 11beta-hydroxysteroid dehydrogenase type 1 inhibitor." J Steroid Biochem Mol Biol 128(1-2): 51-55.

Bolander, F. F. (2006). "Vitamins: not just for enzymes." Curr Opin Investig Drugs 7(10): 912-915.

Bonafe, L., B. Thony, W. Leimbacher, L. Kierat and N. Blau (2001). "Diagnosis of dopa-responsive dystonia and other tetrahydrobiopterin disorders by the study of biopterin metabolism in fibroblasts." Clin Chem 47(3): 477-485.

Buchert, R., et al. (2014). "A peroxisomal disorder of severe intellectual disability, epilepsy, and cataracts due to fatty acyl-CoA reductase 1 deficiency." Am J Hum Genet 95(5): 602-610.

Boucher, N. and K. M. Noll (2011). "Ligands of thermophilic ABC transporters encoded in a newly sequenced genomic region of Thermotoga maritima MSB8 screened by differential scanning fluorimetry." Appl Environ Microbiol 77(18): 6395-6399.

Bray, J. E., B. D. Marsden and U. Oppermann (2009). "The human short-chain dehydrogenase/reductase (SDR) superfamily: a bioinformatics summary." Chem Biol Interact 178(1-3): 99-109.

Breton, R., et al. (1996). "The structure of a complex of human 17beta-hydroxysteroid dehydrogenase with estradiol and $NADP^+$ identifies two principal targets for the design of inhibitors." Structure 4(8): 905-915.

Burmeister, W. P., S. Cottaz, P. Rollin, A. Vasella and B. Henrissat (2000). "High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base." J Biol Chem 275(50): 39385-39393.

Canto, C., K. J Menzies and J. Auwerx (2015) “NAD(+) metabolism and the control of enery homeostasis: A balancing act between mitochondria and nucleus.” Cell Metab 25(1): 31-53

**Chai, Z., P. Brereton, T. Suzuki, H. Sasano, V. Obeyesekere, G. Escher, R. Saffery, P. Fuller, C. Enriquez and Z. Krozowski (2003). "17 beta-hydroxysteroid dehydrogenase type XI localizes to human steroidogenic cells." Endocrinology 144(5): 2084-2091.

Chen, V. and E. Shtivelman (2010). "CC3/TIP30 regulates metabolic adaptation of tumor cells to glucose limitation." Cell Cycle 9(24): 4941-4953.

Chen, Y., C. L. Sawyers and H. I. Scher (2008). "Targeting the androgen receptor pathway in prostate cancer." Curr Opin Pharmacol 8(4): 440-448.

Chen, Z., J. C. Jiang, Z. G. Lin, W. R. Lee, M. E. Baker and S. H. Chang (1993). "Site-specific mutagenesis of Drosophila alcohol dehydrogenase: evidence for involvement of tyrosine-152 and lysine-156 in catalysis." Biochemistry 32(13): 3342-3346.

**Chen, Z., A. J. Kastaniotis, I. J. Miinalainen, V. Rajaram, R. K. Wierenga and J. K. Hiltunen (2009). "17beta-hydroxysteroid dehydrogenase type 8 and carbonyl reductase type 4 assemble as a ketoacyl reductase of human mitochondrial FAS." FASEB J 23(11): 3682-3691.

Cheng, J. B., E. Jacquemin, M. Gerhardt, H. Nazer, D. Cresteil, J. E. Heubi, K. D. Setchell and D. W. Russell (2003). "Molecular genetics of 3beta-hydroxy-Delta5-C27-steroid oxidoreductase deficiency in 16 patients with loss of bile acid synthesis and liver disease." J Clin Endocrinol Metab 88(4): 1833-1841.

**Cheng, J. B. and D. W. Russell (2004). "Mammalian wax biosynthesis. I. Identification of two fatty acyl-Coenzyme A reductases with different substrate specificities and tissue distributions." J Biol Chem 279(36): 37789-37797.

Chetyrkin, S. V., O. V. Belyaeva, W. H. Gough and N. Y. Kedishvili (2001). “Characterization of a novel type of human microsomal 3alpha -hydroxysteroid dehydrogenase: unique tissue distribution and catalytic properties.” J Biol Chem 276(25):22278-86.

**Chi, A., J. C. Valencia, Z. Z. Hu, H. Watabe, H. Yamaguchi, N. J. Mangini, H. Huang, V. A. Canfield, K. C. Cheng, F. Yang, R. Abe, S. Yamagishi, J. Shabanowitz, V. J. Hearing, C. Wu, E. Appella and D. F. Hunt (2006). "Proteomic and bioinformatic characterization of the biogenesis and function of melanosomes." J Proteome Res 5(11): 3135-3144.

**Cho, H., A. Hamza, C. G. Zhan and H. H. Tai (2005). "Key NAD^+-binding residues in human 15-hydroxyprostaglandin dehydrogenase." Arch Biochem Biophys 433(2): 447-453.

Chou, W. C., Y. Yin and Y. Xu (2010). "GolgiP: prediction of Golgi-resident proteins in plants." Bioinformatics 26(19): 2464-2465.

Claros, M. G. (1995). "MitoProt, a Macintosh application for studying mitochondrial proteins." Comput Appl Biosci 11(4): 441-447.

**Clish, C. B., B. D. Levy, N. Chiang, H. H. Tai and C. N. Serhan (2000). "Oxidoreductases in lipoxin A4 metabolic inactivation: a novel role for 15-onoprostaglandin 13-reductase/leukotriene B4 12-hydroxydehydrogenase in inflammation." J Biol Chem 275(33): 25372-25380.

Cole, C., J. D. Barber and G. J. Barton (2008). "The Jpred 3 secondary structure prediction server." Nucleic Acids Res 36(Web Server issue): W197-201.

Cummings, M. D., M. A. Farnum and M. I. Nelen (2006). "Universal screening methods and applications of ThermoFluor." J Biomol Screen 11(7): 854-863.

Dai, X., X. Gu, M. Luo and X. Zheng (2006). "Protein expression, crystallization and preliminary X-ray crystallographic studies on HSCARG from Homo sapiens." Protein Pept Lett 13(9): 955-957.

Dai, X., Y. Li, G. Meng, S. Yao, Y. Zhao, Q. Yu, J. Zhang, M. Luo and X. Zheng (2009). "NADPH is an allosteric regulator of HSCARG." J Mol Biol 387(5): 1277-1285.

Danielsson, O., S. Atrian, T. Luque, L. Hjelmqvist, R. Gonzalez-Duarte and H. Jornvall (1994). "Fundamental molecular differences between alcohol dehydrogenase classes." Proc Natl Acad Sci U S A 91(11): 4980-4984.

Das, A. K., et al. (2000). "Molecular cloning and expression of mammalian peroxisomal trans-2-enoyl-coenzyme A reductase cDNAs." J Biol Chem 275(32): 24333-24340.

Day, J. M., P. A. Foster, H. J. Tutill, M. F. Parsons, S. P. Newman, S. K. Chander, G. M. Allan, H. R. Lawrence, N. Vicker, B. V. Potter, M. J. Reed and A. Purohit (2008). "17beta-hydroxysteroid dehydrogenase Type 1, and not Type 12, is a target for endocrine therapy of hormone-dependent breast cancer." Int J Cancer 122(9): 1931-1940.

Day, J. M., H. J. Tutill, P. A. Foster, H. V. Bailey, W. B. Heaton, C. M. Sharland, N. Vicker, B. V. Potter, A. Purohit and M. J. Reed (2009). "Development of hormone-dependent prostate cancer models for the evaluation of inhibitors of 17beta-hydroxysteroid dehydrogenase type 3." Mol Cell Endocrinol 301(1-2): 251-258.

Day, J. M., H. J. Tutill and A. Purohit (2010). “17ß-hydroxysteroid dehydrogenase inhibitors.” Minerva Endocrinol. 35(2):87-108.

**De Nys, K., E. Meyhi, G. P. Mannaerts, M. Fransen and P. P. Van Veldhoven (2001). "Characterisation of human peroxisomal 2,4-dienoyl-CoA reductase." Biochim Biophys Acta 1533(1): 66-72.

. **Deisenroth, C., A. R. Thorner, T. Enomoto, C. M. Perou and Y. Zhang (2010). "Mitochondrial Hep27 is a c-Myb target gene that inhibits Mdm2 and stabilizes p53." Mol Cell Biol 30(16): 3981-3993

Del Mare, S. and R. Ageilan (2015). “Tumor Suppressor WWOX inhibits osteosarcoma metastasis by modulating RUNX2 function.” Sci Rep. 5:12959.

Deller, M. C, L. Kong and B. Rupp (2016). “Protein stability: A crystallographer’s perspective”. Acta Crystallogr F Struc Biol Commun. 72(2): 72-95.

Delvoux, B., T. D'Hooghe, C. Kyama, P. Koskimies, R. J. Hermans, G. A. Dunselman and A. Romano (2014). “Inhibition of type 1 17β-hydroxysteroid dehydrogenase impairs the synthesis of 17β-estradiol inendometriosis lesions.” J Clin Endocrinol Metab. 99(1):276-84.

Dothie, J. M., J. R. Giglio, C. B. Moore, S. S. Taylor and B. S. Hartley (1985). "Ribitol dehydrogenase of Klebsiella aerogenes. Sequence and properties of wild-type and mutant strains." Biochem J 230(3): 569-578.

Duax, W. L., R. Huether, V. Pletnev, T. C. Umland and M. W. Charles (2009). "Divergent evolution of a Rossmann fold and identification of its oldest surviving ancestor." Int J Bioinform Res Appl 5(3): 280-294.

Egger, S., A. Chaikuad, K. L. Kavanagh, U. Oppermann and B. Nidetzky (2011). "Structure and mechanism of human UDP-glucose 6-dehydrogenase." J Biol Chem 286(27): 23877-23887.

Ehmke, N., et al. (2014). "Homozygous and compound-heterozygous mutations in TGDS cause Catel-Manzke syndrome." Am J Hum Genet 95(6): 763-770.

Elegheert, J., E. Behiels, B. Bishop, S. Scott, R. E. Woolley, S. C. Griffiths. E. F. X. Byrne, V. T. Chang, D. I .Stuart, E. Y. Jones, C. Siebold and A. R. Aricescu. "Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins." Nat Protoc 13:2991-3017.

El-Hawari, Y., A. D. Favia, E. S. Pilka, M. Kisiela, U. Oppermann, H. J. Martin and E. Maser (2009). "Analysis of the substrate-binding site of human carbonyl reductases CBR1 and CBR3 by site-directed mutagenesis." Chem Biol Interact 178(1-3): 234-241.

El-Kabbani, O., S. Ishikura, C. Darmanin, V. Carbone, R. P. Chung, N. Usami and A. Hara (2004). "Crystal structure of human L-xylulose reductase holoenzyme: probing the role of Asn107 with site-directed mutagenesis." Proteins 55(3): 724-732.

El-Omari, K., L. E. Bird, C. E. Nichols, J. Ren and D. K. Stammers (2005). "Crystal structure of CC3 (TIP30): implications for its role as a tumor suppressor." J Biol Chem 280(18): 18229-18236.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1): 2126-2132.

**Endo, S., T. Matsunaga, Y. Kitade, S. Ohno, K. Tajima, O. El-Kabbani and A. Hara (2008). "Human carbonyl reductase 4 is a mitochondrial NADPH-dependent quinone reductase." Biochem Biophys Res Commun 377(4): 1326-1330.

Ericsson, U. B., B. M. Hallberg, G. T. Detitta, N. Dekker and P. Nordlund (2006). "Thermofluor-based high-throughput stability optimization of proteins for structural studies." Anal Biochem 357(2): 289-298.

Fattahi, M. J. and A. Mirshafiey (2012). "Prostaglandins and rheumatoid arthritis." Arthritis 2012: 239310.

Fedorov, O., F. H. Niesen and S. Knapp (2012). "Kinase inhibitor selectivity profiling using differential scanning fluorimetry." Methods Mol Biol 795: 109-118.

Filling, C., K. D. Berndt, J. Benach, S. Knapp, T. Prozorovski, E. Nordling, R. Ladenstein, H. Jornvall and U. Oppermann (2002). "Critical residues for structure and catalysis in short-chain dehydrogenases/reductases." J Biol Chem 277(28): 25677-25684.

Filling, C., E. Nordling, J. Benach, K. D. Berndt, R. Ladenstein, H. Jornvall and U. Oppermann (2001a). "Structural role of conserved Asn179 in the short-chain dehydrogenase/reductase scaffold." Biochem Biophys Res Commun 289(3): 712-717.

Filling, C., X. Wu, N. Shafqat, M. Hult, E. Martensson, J. Shafqat and U. C. Oppermann (2001b). "Subcellular targeting analysis of SDR-type hydroxysteroid dehydrogenases." Mol Cell Endocrinol 171(1-2): 99-101.

Fisher, M., J. T. Kroon, W. Martindale, A. R. Stuitje, A. R. Slabas and J. B. Rafferty (2000). "The X-ray structure of Brassica napus beta-keto acyl carrier protein reductase and its implications for substrate binding and catalysis." Structure 8(4): 339-347.

Fjeld, C. C., et al. (2003). "Differential binding of NAD^+ and NADH allows the transcriptional corepressor carboxyl-terminal binding protein to serve as a metabolic sensor." Proc Natl Acad Sci U S A 100(16): 9202-9207.

Fomitcheva, J., M. E. Baker, E. Anderson, G. Y. Lee and N. Aziz (1998). "Characterization of Ke 6, a new 17beta-hydroxysteroid dehydrogenase, and its expression in gonadal tissues." J Biol Chem 273(35): 22664-22671.

Fournier, M. A. and D. Poirier (2009). "Estrogen formation in endometrial and cervix cancer cell lines: involvement of aromatase, steroid sulfatase and 17beta-hydroxysteroid dehydrogenases (types 1, 5, 7 and 12)." Mol Cell Endocrinol 301(1-2): 142-145.

Froese, D. S., et al. (2013). "Crystal structures of malonyl-coenzyme A decarboxylase provide insights into its catalytic mechanism and disease-causing mutations." Structure 21(7): 1182-1192.

Fujimoto, K., M. Hara, H. Yamada, M. Sakurai, A. Inaba, A. Tomomura and S. Katoh (2001). "Role of the conserved Ser-Tyr-Lys triad of the SDR family in sepiapterin reductase." Chem Biol Interact 130-132(1-3): 825-832.

Funder, J. W., P. T. Pearce, R. Smith and A. I. Smith (1988). "Mineralocorticoid action: target tissue specificity is enzyme, not receptor, mediated." Science 242(4878): 583-585.

Furster, C. (1999). "Hepatic and extrahepatic dehydrogenation/isomerization of 5-cholestene-3 beta,7 alpha-diol: localization of 3 beta-hydroxy-delta 5-C27-steroid dehydrogenase in pig tissues and subcellular fractions." Biochim Biophys Acta 1436(3): 343-353.

**Gabrielli, F., G. Donadel, G. Bensi, A. Heguy and M. Melli (1995). "A nuclear protein, synthesized in growth-arrested human hepatoblastoma cells, is a novel member of the short-chain alcohol dehydrogenase family." Eur J Biochem 232(2): 473-477.

**Gangloff, A., R. Shi, V. Nahoum and S. X. Lin (2003). "Pseudo-symmetry of C19 steroids, alternative binding orientations, and multispecificity in human estrogenic 17beta-hydroxysteroid dehydrogenase." FASEB J 17(2): 274-276.

Garcoa-de-Lorenzo y Mateos, A., J. C. Montejo González, Quintana Diaz M. (2015) Eicosanoid Synthesis and Respiratory Distress Syndrome in Intensive Medicine. In: Rajendram R., Preedy V.R., Patel V.B. (eds) Diet and Nutrition in Critical Care. Springer, New York, NY

Gavidia, I., R. Tarrio, F. Rodriguez-Trelles, P. Perez-Bermudez and H. U. Seitz (2007). "Plant progesterone 5beta-reductase is not homologous to the animal enzyme. Molecular evolutionary characterization of P5betaR from Digitalis purpurea." Phytochemistry 68(6): 853-864.

Geders, T. W., K. Gustafson and B. C. Finzel (2012). "Use of differential scanning fluorimetry to optimize the purification and crystallization of PLP-dependent enzymes." Acta Crystallogr Sect F Struct Biol Cryst Commun 68(Pt 5): 596-600.

Geissler, W. M., D. L. Davis, L. Wu, K. D. Bradshaw, S. Patel, B. B. Mendonca, K. O. Elliston, J. D. Wilson, D. W. Russell and S. Andersson (1994). "Male pseudohermaphroditism caused by mutations of testicular 17 beta-hydroxysteroid dehydrogenase 3." Nat Genet 7(1): 34-39.

Ghosh, D., V. Z. Pletnev, D. W. Zhu, Z. Wawrzak, W. L. Duax, W. Pangborn, F. Labrie and S. X. Lin (1995). "Structure of human estrogenic 17 beta-hydroxysteroid dehydrogenase at 2.20 A resolution." Structure 3(5): 503-513.

Ghosh, D., M. Sawicki, V. Pletnev, M. Erman, S. Ohno, S. Nakajin and W. L. Duax (2001). "Porcine carbonyl reductase. structural basis for a functional monomer in short chain dehydrogenases/reductases." J Biol Chem 276(21): 18457-18463.

Ghosh, D., Z. Wawrzak, C. M. Weeks, W. L. Duax and M. Erman (1994). "The refined three-dimensional structure of 3 alpha,20 beta-hydroxysteroid dehydrogenase and possible roles of the residues conserved in short-chain dehydrogenases." Structure 2(7): 629-640.

Ghosh, D., C. M. Weeks, P. Grochulski, W. L. Duax, M. Erman, R. L. Rimsay and J. C. Orr (1991). "Three-dimensional structure of holo 3 alpha,20 beta-hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family." Proc Natl Acad Sci U S A 88(22): 10064-10068.

Gileadi, O., S. Knapp, W. H. Lee, B. D. Marsden, S. Muller, F. H. Niesen, K. L. Kavanagh, L. J. Ball, F. von Delft, D. A. Doyle, U. C. Oppermann and M. Sundstrom (2007). "The scientific impact of the Structural Genomics Consortium: a protein family and ligand-centered approach to medically-relevant human proteins." J Struct Funct Genomics 8(2-3): 107-119.

Go, Y. and D. Jones (2009). “Redox compartilization in eukaryotic cells.” Biochim Biophys Acta 1780(11): 1273-1290.

González B., F. Garrido, R. Ortega, M. Martínez-Júlvez, A. Revilla-Guarinos, Y. Pérez-Pertejo, A. Velázquez-Campoy, J. Sanz-Aparicio and M. A. Pajares (2012). $NADP^+$ binding to the regulatory subunit of methionine adenosyltransferase II increases intersubunit binding affinity in the hetero-trimer.

Goodman, C. D. and G. I. McFadden (2007). "Fatty acid biosynthesis as a drug target in apicomplexan parasites." Curr Drug Targets 8(1): 15-30.

Gordon, J. C., J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev (2005). "H++: a server for estimating pKas and adding missing hydrogens to macromolecules." Nucleic Acids Res 33(Web Server issue): W368-371.

Goto, M., H. Muramatsu, H. Mihara, T. Kurihara, N. Esaki, R. Omi, I. Miyahara and K. Hirotsu (2005). "Crystal structures of Delta1-piperideine-2-carboxylate/Delta1-pyrroline-2-carboxylate reductase belonging to a new family of NAD(P)H-dependent oxidoreductases: conformational change, substrate recognition, and stereochemistry of the reaction." J Biol Chem 280(49): 40875-40884.

Gourley, D. G., A. W. Schuttelkopf, G. A. Leonard, J. Luba, L. W. Hardy, S. M. Beverley and W. N. Hunter (2001). "Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites." Nat Struct Biol 8(6): 521-525.

**Green, D., A. R. Marks, S. Fleischer and J. O. McIntyre (1996). "Wild type and mutant human heart (R)-3-hydroxybutyrate dehydrogenase expressed in insect cells." Biochemistry 35(25): 8158-8165.

Guo, K., P. Lukacik, E. Papagrigoriou, M. Meier, W. H. Lee, J. Adamski and U. Oppermann (2006). "Characterization of human DHRS6, an orphan short chain dehydrogenase/reductase enzyme: a novel, cytosolic type 2 R-beta-hydroxybutyrate dehydrogenase." J Biol Chem 281(15): 10291-10297.

Haeseleer, F., G. F. Jang, Y. Imanishi, C. A. Dresden, M. Matsumura, P. S. Nelson and K. Barczewski (2002). "Dual-substrate specificity short chain retinol dehydrogenases from the vertebrate retina." J Boil CChem 277(47): 45537-45546.

**Haeseleer, F., J. Huang, L. Lebioda, J. C. Saari and K. Palczewski (1998). "Molecular characterization of a novel short-chain dehydrogenase/reductase that reduces all-trans-retinal." J Biol Chem 273(34): 21790-21799.

Hanukoglu, I. and R. Rapoport (1995). Routes and regulation of NADPH production in steroidogenic mitochondria. Endocr Res 21(1-2):231-41.

**Hawkins, J., D. Mahony, S. Maetschke, M. Wakabayashi, R. D. Teasdale and M. Boden (2007). "Identifying novel peroxisomal proteins." Proteins 69(3): 606-616.

Hazra, S., R. K. Bart, H. H. Tai, S. Sharma, X. Cui and S. M. Dubiety (2007). "Pioglitazone and rosiglitazone decrease prostaglandin E2 in non-small-cell lung cancer cells by up-regulating 15-hydroxyprostaglandin dehydrogenase." Moll Pharmacology 71(6): 1715-1720.

He, W., M. Gauri, T. Li, R. Wang and S. X Lin (2016). “Current knowledge of the multifunctional 17β-hydroxysteroid dehydrogenase type 1 (HSD17B1).” Gene 588(1): 54-61.

**He, X. Y., Y. Z. Yang, D. M. Peehl, A. Lauderdale, H. Schulz and S. Y. Yang (2003). "Oxidative 3alpha-hydroxysteroid dehydrogenase activity of human type 10 17beta-hydroxysteroid dehydrogenase." J Steroid Biochem Mol Biol 87(2-3): 191-198.

Hoagland, A., P. Donnas, T. Blum, H. W. Adolph and O. Kohlbacher (2006). "MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition." Bioinformatics 22(10): 1158-1165.

Horton, P., K. J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier and K. Nakai (2007). "WoLF PSORT: protein localization predictor." Nucleic Acids Res 35(Web Server issue): W585-587.

Houtkooper, R. H., C. Canto, R. J. Wanders and J. Auwerx (2010). "The secret life of NAD^+: an old metabolite controlling new metabolic signaling pathways." Endocr Rev 31(2): 194-223.

Howells, D. W., S. M. Forrest, H. H. Dahl and R. G. Cotton (1990). "Insertion of an extra codon for threonine is a cause of dihydropteridine reductase deficiency." Am J Hum Genet 47(2): 279-285.

Hua, T., D. Wu, W. Ding, J. Wang, N. Shaw and Z. J. Liu (2012). "Studies on human 2, 4 dienoyl CoA reductase (DCR) sheds new light on peroxisomal beta-oxidation of unsaturated fatty acids." J Biol Chem.

Huang, Y., X. Li, H. Lin, Y. Chu, B. Chen, Q. Lian and R. S. Ge (2010). "Regulation of 11beta-hydroxysteroid dehydrogenase 1 and 2 by IGF-1 in mice." Biochem Biophys Res Commun 391(4): 1752-1756.

**Huang, X. F. and V. Luu-The (2000). "Molecular characterization of a first human 3(alpha-->beta)-hydroxysteroid epimerase." J Biol Chem 275(38): 29452-29457.

**Huang, X. F. and V. Luu-The (2001). "Gene structure, chromosomal localization and analysis of 3-ketosteroid reductase activity of the human 3(alpha-->beta)-hydroxysteroid epimerase." Biochim Biophys Acta 1520(2): 124-130.

Hult, M., N. Shafqat, B. Elleby, D. Mitschke, S. Svensson, M. Forsgren, T. Barf, J. Vallgarda, L. Abrahmsen and U. Oppermann (2006). "Active site variability of type 1 11beta-hydroxysteroid dehydrogenase revealed by selective inhibitors and cross-species comparisons." Mol Cell Endocrinol 248(1-2): 26-33.

Jadhav, A., F. H. Niesen, L. Schultz, U. Oppermann, D. J. Maloney and A. Simeonov (2010). Potent and selective inhibitors of NAD^+-dependent 15-hydroxyprostaglandin dehydrogenase (HPGD). Probe Reports from the NIH Molecular Libraries Program. Bethesda (MD).

Jamieson, A., A. M. Wallace, R. Andrew, B. S. Nunez, B. R. Walker, R. Fraser, P. C. White and J. M. Connell (1999). "Apparent cortisone reductase deficiency: a functional defect in 11beta-hydroxysteroid dehydrogenase type 1." J Clin Endocrinol Metab 84(10): 3570-3574.

**Jayakumar, A., M. H. Tai, W. Y. Huang, W. al-Feel, M. Hsu, L. Abu-Elheiga, S. S. Chirala and S. J. Wakil (1995). "Human fatty acid synthase: properties and molecular cloning." Proc Natl Acad Sci U S A 92(19): 8695-8699.

Jeankumar, V. U., J. Renuka, S. Kotagiri, S. Saxena, S. S. Kakan, J. P. Sridevi, S. Yellanki, P. Kulkarni, P. Yogeeswari and D. Sriram (2014). "Gyrase ATPase Domain as an Antitubercular Drug Discovery Platform: Structure-Based Design and Lead Optimization of Nitrothiazolyl Carboxamide Analogues." ChemMedChem.

Jiang, C., M. Ito, V. Piening, K. Bruck, R. G. Roeder and H. Xiao (2004). "TIP30 interacts with an estrogen receptor alpha-interacting coactivator CIA and regulates c-myc transcription." J Biol Chem 279(26): 27781-27789.

**Jiang, L. L., S. Miyazawa, M. Souri and T. Hashimoto (1997). "Structure of D-3-hydroxyacyl-CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase bifunctional protein." J Biochem 121(2): 364-369.

Jornvall, H., J. Hedlund, T. Bergman, U. Oppermann and B. Persson (2010). "Superfamilies SDR and MDR: from early ancestry to present forms. Emergence of three lines, a Zn-metalloenzyme, and distinct variabilities." Biochem Biophys Res Commun 396(1): 125-130.

Jornvall, H., J. O. Hoog and B. Persson (1999). "SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature." FEBS Lett 445(2-3): 261-264.

Jornvall, H., E. Nordling and B. Persson (2003). “Multiplicity of eukaryotic ADH and other MDR forms.” Chem Biol Interact 143-144:255-261.

Jornvall, H., B. Persson, M. Krook, S. Atrian, R. Gonzalez-Duarte, J. Jeffery and D. Ghosh (1995). "Short-chain dehydrogenases/reductases (SDR)." Biochemistry 34(18): 6003-6013.

Jornvall, H., M. Persson and J. Jeffery (1981). "Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type." Proc Natl Acad Sci U S A 78(7): 4226-4230.

Jornvall, H., H. von Bahr-Lindstrom, K. D. Jany, W. Ulmer and M. Froschle (1984a). "Extended superfamily of short alcohol-polyol-sugar dehydrogenases: structural similarities between glucose and ribitol dehydrogenases." FEBS Lett 165(2): 190-196.

Jornvall, H., H. von Bahr-Lindstrom and J. Jeffery (1984b). "Extensive variations and basic features in the alcohol dehydrogenase-sorbitol dehydrogenase family." Eur J Biochem 140(1): 17-23.

Kall, L., A. Krogh and E. L. Sonnhammer (2004). "A combined transmembrane topology and signal peptide prediction method." J Mol Biol 338(5): 1027-1036.

Kallberg, Y., U. Oppermann, H. Jornvall and B. Persson (2002a). "Short-chain dehydrogenase/reductase (SDR) relationships: a large family with eight clusters common to human, animal, and plant genomes." Protein Sci 11(3): 636-641.

Kallberg, Y., U. Oppermann, H. Jornvall and B. Persson (2002b). "Short-chain dehydrogenases/reductases (SDRs)." Eur J Biochem 269(18): 4409-4417.

Kallberg, Y., U. Oppermann and B. Persson (2010). "Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models." FEBS J 277(10): 2375-2386.

Kallberg, Y. and B. Persson (2006). "Prediction of coenzyme specificity in dehydrogenases/reductases. A hidden Markov model-based method and its application on complete genomes." FEBS J 273(6): 1177-1184.

Kavanagh, K. L., H. Jornvall, B. Persson and U. Oppermann (2008). "Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes." Cell Mol Life Sci 65(24): 3895-3906.

Kavanagh, K. L., M. Klimacek, B. Nidetzky and D. K. Wilson (2002). "Crystal structure of Pseudomonas fluorescens mannitol 2-dehydrogenase binary and ternary complexes. Specificity and catalytic mechanism." J Biol Chem 277(45): 43433-43442.

**Kedishvili, N. Y., O. V. Chumakova, S. V. Chetyrkin, O. V. Belyaeva, E. A. Lapshina, D. W. Lin, M. Matsumura and P. S. Nelson (2002). "Evidence that the human gene for prostate short-chain dehydrogenase/reductase (PSDR1) encodes a novel retinal reductase (RalR1)." J Biol Chem 277(32): 28909-28915.

Kerscher, S. J. (2000). "Diversity and origin of alternative NADH:ubiquinone oxidoreductases." Biochim Biophys Acta 1459(2-3): 274-283.

**Kihara, A. and Y. Igarashi (2004). "FVT-1 is a mammalian 3-ketodihydrosphingosine reductase with an active site that faces the cytosolic side of the endoplasmic reticulum membrane." J Biol Chem 279(47): 49243-49250.

Kim, J. P. and K. P. Battaile (2002). @Burning fat: the structural basis of fatty acid β-oxidation.” Curr Opin Struc Biol. 12(6): 721-728.

King, F. W. and E. Shtivelman (2004). "Inhibition of nuclear import by the proapoptotic protein CC3." Mol Cell Biol 24(16): 7091-7101.

Kisiela, M., A. Skarka, B. Ebert and E. Maser (2012). "Hydroxysteroid dehydrogenases (HSDs) in bacteria: a bioinformatic perspective." J Steroid Biochem Mol Biol 129(1-2): 31-46.

Konig, A., R. Happle, D. Bornholdt, H. Engel and K. H. Grzeschik (2000). "Mutations in the NSDHL gene, encoding a 3beta-hydroxysteroid dehydrogenase, cause CHILD syndrome." Am J Med Genet 90(4): 339-346.

Korzybski, T. (1962). "[The basis for classification and nomenclature of enzymes according to the Report of the Commission on Enzymes of the International Union of Biochemistry]." Postepy Biochem 8: 261-293.

Kowalik, D., F. Haller, J. Adamski and G. Moeller (2009). "In search for function of two human orphan SDR enzymes: hydroxysteroid dehydrogenase like 2 (HSDL2) and short-chain dehydrogenase/reductase-orphan (SDR-O)." J Steroid Biochem Mol Biol 117(4-5): 117-124.

Krebs, H. A. and R. L. Veech (1969). "Equilibrium relations between pyridine nucleotides and adenine nucleotides and their roles in the regulation of metabolic processes." Adv Enzyme Regul 7: 397-413.

Kretschmar, M. and R. Jaenicke (1999). "Stability of a homo-dimeric Ca(2+)-binding member of the beta gamma-crystallin superfamily: DSC measurements on spherulin 3a from Physarum polycephalum." J Mol Biol 291(5): 1147-1153.

Kretschmar, M., E. M. Mayr and R. Jaenicke. (1999). "Kinetic and thermodynamic stabilization of the betagamma-crystallin homolog spherulin 3a from Physarum polycephalum by calcium binding." J Mol Biol 289(4): 701-705.

Krissinel, E. and Henrick, K. (2007). 'Inference of macromolecular assemblies from crystalline state.'. J. Mol. Biol. **372**, 774--797

Krook, M., L. Marekov and H. Jornvall (1990). "Purification and structural characterization of placental NAD(+)-linked 15-hydroxyprostaglandin dehydrogenase. The primary structure reveals the enzyme to belong to the short-chain alcohol dehydrogenase family." Biochemistry 29(3): 738-743.

Küssau, T., M. Flipo, N. Van Wyk, A. Vijoen, V. Olieeric, L. Kremer and M. Blaise (2018). " Structural rearrangements occurring upon cofactor binding in the *Mycobacterium smegmatis B-ketoacyl* carrier protein reductase MabA." Acta Crystallogr D Struct Biol. 74(5): 383-393.

Kuroki, T., F. Trapasso, T. Shiraishi, H. Alder, K. Mimori, M. Mori and C. M. Croce (2002). "Genetic alterations of the tumor suppressor gene WWOX in esophageal squamous cell carcinoma." Cancer Res 62(8): 2258-2260.

La Cour, T., L. Kiemer, A. Mølgaard, R. Gupta, K. Skriver and S. Brunak (2004). "Analysis and prediction of leucine-rich nuclear export signals." Protein Eng Des Sel 17(6):527-536.

Laplante, Y., C. Rancourt and D. Poirier (2009). "Relative involvement of three 17beta-hydroxysteroid dehydrogenases (types 1, 7 and 12) in the formation of estradiol in various breast cancer cell lines using selective inhibitors." Mol Cell Endocrinol 301(1-2): 146-153.

Latos-Bielenska, A., I. Marik, M. Kuklik, A. Materna-Kiryluk, C. Povysil and K. Kozlowski (2007). "Pachydermoperiostosis-critical analysis with report of five unusual cases." Eur J Pediatr 166(12): 1237-1243.

Lavinder, J. J., et al. (2009). "High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering." J Am Chem Soc 131(11): 3794-3795.

Le-Quoc, D. and K. Le-Quoc (1989). "Relationships between the NAD(P) redox state, fatty acid oxidation, and inner membrane permeability in rat liver mitochondria." Arch Biochem Biophys 273(2): 466-478.

Lea, W. A. and A. Simeonov (2012). "Differential scanning fluorometry signatures as indicators of enzyme inhibitor mode of action: case study of glutathione S-transferase." PLoS One 7(4): e36219.

Lee, S. A., O. V. Belyaeva OV and N. Y. Kedishvili (2009). “Biochemical characterization of human epidermal retinol dehydrogenase 2.”Chem Biol Interact 178(1-3):182-7.

**LeGros, H. L., Jr., A. B. Halim, A. M. Geller and M. Kotb (2000). "Cloning, expression, and functional characterization of the beta regulatory subunit of human methionine adenosyltransferase (MAT II)." J Biol Chem 275(4): 2359-2366.

Lesk, A. M. (1995). "NAD-binding domains of dehydrogenases." Curr Opin Struct Biol 5(6): 775-783.

Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography No. 26.

Liden, M., A. Romert, K. Tryggvason, B. Persson and U. Eriksson (2001). "Biochemical defects in 11-cis-retinol dehydrogenase mutants associated with fundus albipunctatus." J Biol Chem 276(52): 49251-49257.

Lima, L. H., C. G. Pinheiro, L. M. De Moraes, S. M. De Freitas And F. A. Torres (2006). “Xylitol dehydrogenase From Candida Tropicalis: Molecular Cloning Of The Gene And Structural Analysis Of The Protein.” Appl Microbio Biotchnol 73(3): 631-639.

Lin, S. X., D. Pourier And J. Adamski (2013). “A challenge For medicinal Chemistry By The 17β-Hydroxysteroid Dehydrogenase Superfamily: An Integrated Biological Function And Inhibition Study.” Curr Top Med Chem. 3(10):1164-1171.

**Lindqvist, A., I. A. Hughes and S. Andersson (2001). "Substitution mutation C268Y causes 17 beta-hydroxysteroid dehydrogenase 3 deficiency." J Clin Endocrinol Metab 86(2): 921-923.

**Lockyer, J., R. G. Cook, S. Milstien, S. Kaufman, S. L. Woo and F. D. Ledley (1987). "Structure and expression of human dihydropteridine reductase." Proc Natl Acad Sci U S A 84(10): 3329-3333.

**Lorence, M. C., C. J. Corbin, N. Kamimura, M. S. Mahendroo and J. I. Mason (1990). "Structural analysis of the gene encoding human 3 beta-hydroxysteroid dehydrogenase/delta 5----4-isomerase." Mol Endocrinol 4(12): 1850-1855.

**Lukacik, P., B. Keller, G. Bunkoczi, K. L. Kavanagh, W. H. Lee, J. Adamski and U. Oppermann (2007). "Structural and biochemical characterization of human orphan DHRS10 reveals a novel cytosolic enzyme with steroid dehydrogenase activity." Biochem J 402(3): 419-427.Lundová T., L. Zemanová, B. Malčeková, A. Skarka, H. Štambergová, J. Havránková, M. Šafr and V. Wsól (2015). “Molecular and biochemical characterisation of human short-chain dehydrogenase/reductase member 3 (DHRS3).” Chem Biol Interact 234:178-187.

Luu-The, V., P. Tremblay and F. Labrie (2006). "Characterization of type 12 17beta-hydroxysteroid dehydrogenase, an isoform of type 3 17beta-hydroxysteroid dehydrogenase responsible for estradiol formation in women." Mol Endocrinol 20(2): 437-443.

Lyu, Q., S. Wang, W. Xu, B. Han, W. Liu and D. N. Jones (2014). "Structural insights into the substrate-binding mechanism for a novel chitosanase." Biochem J 461(2): 335-345.

Macheroux P., B. Kappes B, S. E. Ealick (2011) “Flavogenomics--a genomic and structural view of flavin-dependent proteins.” FEBS J 278(15):2625-2634

Magni, G., G. Orsomando and N. Raffaelli (2006). "Structural and functional properties of NAD kinase, a key enzyme in NADP biosynthesis." Mini Rev Med Chem 6(7): 739-746.

Magrane, M. and U. Consortium (2011). "UniProt Knowledgebase: a hub of integrated protein data." Database (Oxford) 2011: bar009.

Mahajan, P., C. Strain-Damerell, O. Gileadi and N. A. Burgess-Brown (2014). "Medium-throughput production of recombinant human proteins: protein production in insect cells." Methods Mol Biol 1091: 95-121.

Malleret, B., et al. (2015). "Plasmodium vivax: restricted tropism and rapid remodeling of CD71-positive reticulocytes." Blood 125(8): 1314-1324.

Mandl, J., A. Szarka and G. Banhegyi (2009). "Vitamin C: update on physiology and pharmacology." Br J Pharmacol 157(7): 1097-1110.

Manson, W. and V. V. Modi (1957). "Flavin adenine dinucleotide (FAD) metabolism and lactation." Biochim Biophys Acta 24(2): 423-425.

**Marijanovic, Z., D. Laubner, G. Moller, C. Gege, B. Husen, J. Adamski and R. Breitling (2003). "Closing the gap: identification of human 3-ketosteroid reductase, the last unknown enzyme of mammalian cholesterol biosynthesis." Mol Endocrinol 17(9): 1715-1725.

Martin, W. (2010). "Evolutionary origins of metabolic compartmentalization in eukaryotes." Philos Trans R Soc Lond B Biol Sci 365(1541): 847-855.

Mason, J. I., D. Naville, B. W. Evans and J. L. Thomas (1998). "Functional activity of 3beta-hydroxysteroid dehydrogenase/isomerase." Endocr Res 24(3-4): 549-557.

Masters, C. J., S. Reid and M. Don (1987). "Glycolysis--new concepts in an old pathway." Mol Cell Biochem 76(1): 3-14.

Matkovics, B., J. Bariska, M. Simon and S. Zador (1972). "Proposal for the nomenclature classification of the enzymes of all the class-1 oxidoreductases." Enzyme 13(4): 238-239.

**Matsunaga, T., S. Endo, S. Maeda, S. Ishikura, K. Tajima, N. Tanaka, K. T. Nakamura, Y. Imamura and A. Hara (2008). "Characterization of human DHRS4: an inducible short-chain dehydrogenase/reductase enzyme with 3beta-hydroxysteroid dehydrogenase activity." Arch Biochem Biophys 477(2): 339-347.

**Matsunaga, T., S. Shintani and A. Hara (2006). "Multiplicity of mammalian reductases for xenobiotic carbonyl compounds." Drug Metab Pharmacokinet 21(1): 1-18.

Mazurkova, N. A., G. P. Troshkova, T. P. Sumkina, T. D. Kolokol'tsova, M. O. Skarnovich, A. S. Kabanov, L. N. Shishkina and A. N. Sergeev (2008). "Comparative analysis of reproduction of influenza virus strains in cell lines perspective for the creation of cultural vaccines grown on nutrient medium on the basis of rise flour protein hydrolysate." Bull Exp Biol Med 146(4): 547-550.

McCoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni and R. J. Read (2005). "Likelihood-enhanced fast translation functions." Acta Crystallogr D Biol Crystallogr 61(Pt 4): 458-464.

McGuffin, L. J., K. Bryson and D. T. Jones (2000). "The PSIPRED protein structure prediction server." Bioinformatics 16(4): 404-405.

McLane, L. M. and A. H. Corbett (2009). "Nuclear localization signals and human disease." IUBMB Life 61(7): 697-706.

McPherson, A. (2004). "Introduction to protein crystallization." Methods 34(3): 254-265.

Megarity, C. F., H. K. Looi and D. J. Timson (2014). "The Saccharomyces cerevisiae quinone oxidoreductase Lot6p: stability, inhibition and cooperativity." FEMS Yeast Res.

Meier, M., J. Tokarz, F. Haller, R. Mindnich and J. Adamski (2009). "Human and zebrafish hydroxysteroid dehydrogenase like 1 (HSDL1) proteins are inactive enzymes but conserved among species." Chem Biol Interact 178(1-3): 197-205.

Meining, W. and A. Skerra (2012). "The crystal structure of human alpha(1)-microglobulin reveals a potential haem-binding site." Biochem J 445(2): 175-182.

**Mertz, J. R., E. Shang, R. Piantedosi, S. Wei, D. J. Wolgemuth and W. S. Blaner (1997). "Identification and characterization of a stereospecific human enzyme that catalyzes 9-cis-retinol oxidation. A possible role in 9-cis-retinoic acid formation." J Biol Chem 272(18): 11744-11749.

Miura, T., T. Nishinaka and T. Terada (2008). "Different functions between human monomeric carbonyl reductase 3 and carbonyl reductase 1." Mol Cell Biochem 315(1-2): 113-121.

Mizutani, A., A. Matsuzaki, M. Y. Momoi, E. Fujita, Y. Tanabe and T. Momoi (2007). "Intracellular distribution of a speech/language disorder associated FOXP2 mutant." Biochem Biophys Res Commun 353(4): 869-874.

Moller, I. M., A. G. Rasmusson and K. M. Fredlund (1993). "NAD(P)H-ubiquinone oxidoreductases in plant mitochondria." J Bioenerg Biomembr 25(4): 377-384.

**Moon, Y. A. and J. D. Horton (2003). "Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade." J Biol Chem 278(9): 7335-7343.

Murai, T., M. Tokushige, J. Nagai and H. Katsuki (1971). "Physiological functions of NAD- and NADP-linked malic enzymes in *Escherichia coli*." Biochem Biophys Res Commun 43(4): 875-881.

"The subunit composition of the human NADH dehydrogenase obtained by rapid one-step immunopurification."

Murray J., B. Zhang, S. W. Taylor, D. Oglesbee, E. Fahy, M. F. Marusich, S. S. Ghosh and R. A. Capaldi (2003). “Refinement of macromolecular structures by the maximum-likelihood method." J. Biol. Chem 278:13619-13622.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). Acta Crystallogr D Biol Crystallogr 53(Pt 3): 240-255.

Nakagawa, J., S. Ishikura, J. Asami, T. Isaji, N. Usami, A. Hara, T. Sakurai, K. Tsuritani, K. Oda, M. Takahashi, M. Yoshimoto, N. Otsuka and K. Kitamura (2002). "Molecular characterization of mammalian dicarbonyl/L-xylulose reductase and its localization in kidney." J Biol Chem 277(20): 17883-17891.

Nakamura, S., M. Oda, S. Kataoka, S. Ueda, S. Uchiyama, T. Yoshida, Y. Kobayashi and T. Ohkubo (2006). "Apo- and holo-structures of 3alpha-hydroxysteroid dehydrogenase from Pseudomonas sp. B-0831. Loop-helix transition induced by coenzyme binding." J Biol Chem 281(42): 31876-31884.

Nelson, D. R. (1999). "Cytochrome P450 and the individuality of species." Arch Biochem Biophys 369(1): 1-10.

Neuberger, G., S. Maurer-Stroh, B. Eisenhaber, A. Hartig and F. Eisenhaber (2003). "Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence." J Mol Biol 328(3): 581-592.

Nielsen, H. and A. Krogh (1998). “Prediction of signal peptides and signal anchors by hidden Markov model” Proc Int Conf Intell Syst Mol Biol 6:122-130

Niesen, F. H., H. Berglund and M. Vedadi (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." Nat Protoc 2(9): 2212-2221.

Niesen, F. H., L. Schultz, A. Jadhav, C. Bhatia, K. Guo, D. J. Maloney, E. S. Pilka, M. Wang, U. Oppermann, T. D. Heightman and A. Simeonov (2010). "High-affinity inhibitors of human NAD-dependent 15-hydroxyprostaglandin dehydrogenase: mechanisms of inhibition and structure-activity relationships." PLoS One 5(11): e13719.

Nordling, E., H. Jornvall and B. Persson (2002a). "Medium-chain dehydrogenases/reductases (MDR). Family characterizations including genome comparisons and active site modeling." Eur J Biochem 269(17): 4267-4276.

Nordling, E., B. Persson and H. Jornvall (2002b). "Differential multiplicity of MDR alcohol dehydrogenases: enzyme genes in the human genome versus those in organisms initially studied." Cell Mol Life Sci 59(6): 1070-1075.

Obeid, J. and P. C. White (1992). "Tyr-179 and Lys-183 are essential for enzymatic activity of 11 beta-hydroxysteroid dehydrogenase." Biochem Biophys Res Commun 188(1): 222-227.

Odermatt, A., P. Arnold, A. Stauffer, B. M. Frey and F. J. Frey (1999). "The N-terminal anchor sequences of 11beta-hydroxysteroid dehydrogenases determine their orientation in the endoplasmic reticulum membrane." J Biol Chem 274(40): 28762-28770.

Oerum S., M. Roovers, M. Leichsenring, C. Acquaviva-Bourdain, F. Beermann, C. Gemperle-Britschgi, A. Fouilhoux, A. Korwitz-Reichelt, H. J. Bailey, L. Droogmans, U. Oppermann, J. O Sass and W. W. Yue (2017). “Novel patient missense mutations in the HSD17B10 gene affect dehydrogenase and mitochondrial tRNA modification functions of the encoded protein.” Biochim Biophys Acta Mol Basis Dis. 1863(12):3294-3302.

Oerum S., M. Roovers, R. P. Rambo, J. Kopec, J. H. Bailey, F. Fitzpatrick, J. A. Newman, W. G. Newman, A. Amberger, J. Zschocke, L. Droogmans, U. Oppermann and W. W Yue (2018). “Structural insight into the human mitochondrial tRNA purine N1-methyltransferase and ribonuclease P complexes.” J Biol Chem. 293(33):12862-12876

Ofman, R., J. P. Ruiter, M. Feenstra, M. Duran, B. T. Poll-The, J. Zschocke, R. Ensenauer, W. Lehnert, J. O. Sass, W. Sperl and R. J. Wanders (2003). "2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency is caused by mutations in the HADH2 gene." Am J Hum Genet 72(5): 1300-1307.

**Ohno, S., K. Nishikawa, Y. Honda and S. Nakajin (2008). "Expression in *E. coli* and tissue distribution of the human homologue of the mouse Ke 6 gene, 17beta-hydroxysteroid dehydrogenase type 8." Mol Cell Biochem 309(1-2): 209-215.

Oppermann, U. (2006). "Type 1 11beta-hydroxysteroid dehydrogenase as universal drug target in metabolic diseases?" Endocr Metab Immune Disord Drug Targets 6(3): 259-269.

Oppermann, U., C. Filling, M. Hult, N. Shafqat, X. Wu, M. Lindh, J. Shafqat, E. Nordling, Y. Kallberg, B. Persson and H. Jornvall (2003). "Short-chain dehydrogenases/reductases (SDR): the 2002 update." Chem Biol Interact 143-144: 247-253.

Oppermann, U. C., C. Filling, K. D. Berndt, B. Persson, J. Benach, R. Ladenstein and H. Jornvall (1997). "Active site directed mutagenesis of 3 beta/17 beta-hydroxysteroid dehydrogenase establishes differential effects on short-chain dehydrogenase/reductase reactions." Biochemistry 36(1): 34-40.

Oppermann, U. C. and E. Maser (2000). "Molecular and structural aspects of xenobiotic carbonyl metabolizing enzymes. Role of reductases and dehydrogenases in xenobiotic phase I reactions." Toxicology 144(1-3): 71-81.

Park, S. W., H. S. Kim, M. S. Choi, W. J. Jeong, D. S. Heo, K. H. Kim and M. W. Sung (2011). "The effects of the stromal cell-derived cyclooxygenase-2 metabolite prostaglandin E2 on the proliferation of colon cancer cells." J Pharmacol Exp Ther 336(2): 516-523.

Pearl, F., A. Todd, I. Sillitoe, M. Dibley, O. Redfern, T. Lewis, C. Bennett, R. Marsden, A. Grant, D. Lee, A. Akpor, M. Maibaum, A. Harrison, T. Dallman, G. Reeves, I. Diboun, S. Addou, S. Lise, C. Johnston, A. Sillero, J. Thornton and C. Orengo (2005). "The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis." Nucleic Acids Res 33(Database issue): D247-251.

Peng, Y., J. Shi, X. Du, L. Wang, H. Klocker, L. Mo, Z. Mo and J. Zhang (2012). "Prostaglandin E2 induces stromal cell-derived factor-1 expression in prostate stromal cells by activating protein kinase A and transcription factor Sp1." Int J Biochem Cell Biol.

Perrakis, A., M. Harkiolaki, K. S. Wilson and V. S. Lamzin (2001). "ARP/wARP and molecular replacement." Acta Crystallogr D Biol Crystallogr 57(Pt 10): 1445-1450.

Persson, B., J. Hedlund and H. Jornvall (2008). "Medium- and short-chain dehydrogenase/reductase gene and protein families : the MDR superfamily." Cell Mol Life Sci 65(24): 3879-3894.

Persson, B., J. Jeffery and H. Jornvall (1991a). "Different segment similarities in long-chain dehydrogenases." Biochem Biophys Res Commun 177(1): 218-223.

Persson, B., Y. Kallberg, J. E. Bray, E. Bruford, S. L. Dellaporta, A. D. Favia, R. G. Duarte, H. Jornvall, K. L. Kavanagh, N. Kedishvili, M. Kisiela, E. Maser, R. Mindnich, S. Orchard, T. M. Penning, J. M. Thornton, J. Adamski and U. Oppermann (2009). "The SDR (short-chain dehydrogenase/reductase and related enzymes) nomenclature initiative." Chem Biol Interact 178(1-3): 94-98.

Persson, B., Y. Kallberg, U. Oppermann and H. Jornvall (2003). "Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs)." Chem Biol Interact 143-144: 271-278.

Persson, B., M. Krook and H. Jornvall (1991b). "Characteristics of short-chain alcohol dehydrogenases and related enzymes." Eur J Biochem 200(2): 537-543.

Persson, B., E. Nordling, Y. Kallberg, D. Lundh, U. C. Oppermann, H. U. Marschall and H. Jornvall (1999). "Bioinformatics in studies of SDR and MDR enzymes." Adv Exp Med Biol 463: 373-377.

Petersen, T. N., S. Brunak, G. von Heijne and H. Nielsen (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nat Methods 8(10): 785-786.

Phillips, K. and A. H. de la Pena (2011). "The combined use of the Thermofluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization." Curr Protoc Mol Biol Chapter 10: Unit10 28.

Pilka, E. S., F. H. Niesen, W. H. Lee, Y. El-Hawari, J. E. Dunford, G. Kochan, V. Wsol, H. J. Martin, E. Maser and U. Oppermann (2009). "Structural basis for substrate specificity in human monomeric carbonyl reductases." PLoS One 4(10): e7113.

Pollak, N., et al. (2007). "NAD kinase levels control the NADPH concentration in human cells." J Biol Chem 282(46): 33562-33571.

Portanova, J. P., Y. Zhang, G. D. Anderson, S. D. Hauser, J. L. Masferrer, K. Seibert, S. A. Gregory and P. C. Isakson (1996). "Selective neutralization of prostaglandin E2 blocks inflammation, hyperalgesia, and interleukin 6 production in vivo." J Exp Med 184(3): 883-891.

Pospiech, K., E. Pluciennik, A. K. Bednarek (2018). *WWOX* Tumor Suppressor Gene in Breast Cancer, a Historical Perspective and Future Directions. Front Oncol. 8:345

Prehn, C., G. Moller and J. Adamski (2009). "Recent advances in 17beta-hydroxysteroid dehydrogenases." J Steroid Biochem Mol Biol 114(1-2): 72-77.

Purdue, P. E., J. Allsop, G. Isaya, L. E. Rosenberg and C. J. Danpure (1991). "Mistargeting of peroxisomal L-alanine:glyoxylate aminotransferase to mitochondria in primary hyperoxaluria patients depends upon activation of a cryptic mitochondrial targeting sequence by a point mutation." Proc Natl Acad Sci U S A 88(23): 10900-10904.

Rao, S. T. and M. G. Rossmann (1973). "Comparison of super-secondary structures in proteins." J Mol Biol 76(2): 241-256.

Rappaport, N., N. Nativ, G. Stelzer, M. Twik, Y. Guan-Golan, T. I. Stein, I. Bahir, F. Belinky, C. P. Morrey, M. Safran and D. Lancet (2013). "MalaCards: an integrated compendium for diseases and their annotation." Database (Oxford) 2013: bat018.

Read, J. A., V. J. WINTER, C. M. Eszes, R. B. Sessions and R. L. Brady (2001). Structural basis for altered activity of M- and H-isozyme forms of human lactate dehydrogenase." Proteins 43(2): 175-185

Rattner, A., P. M. Smallwood and J. Nathans (2000). "Identification and characterization of all-trans-retinol dehydrogenase from photoreceptor outer segments, the visual cycle enzyme that reduces all-trans-retinal to all-trans-retinol." J Biol Chem 275(15): 11034-11043

Revilla, E., I Fabregat, C Santa María, A Machado (1987). "The NADPH-producing pathways (pentose phosphate and malic enzyme) are regulated by the NADPH consumption in rat mammary gland." Biochem Int. 14(5): 957-62.

Reyniers, E., P. Van Bogaert, N. Peeters, L. Vits, F. Pauly, E. Fransen, N. Van Regemorter and R. F. Kooy (1999). "A new neurological syndrome with mental retardation, choreoathetosis, and abnormal behavior maps to chromosome Xp11." Am J Hum Genet 65(5): 1406-1412.

Rheaume, E., J. Simard, Y. Morel, F. Mebarki, M. Zachmann, M. G. Forest, M. I. New and F. Labrie (1992). "Congenital adrenal hyperplasia due to point mutations in the type II 3 beta-hydroxysteroid dehydrogenase gene." Nat Genet 1(4): 239-245.

Ricciotti, E. and G. A. FitzGerald (2011). "Prostaglandins and inflammation." Arterioscler Thromb Vasc Biol 31(5): 986-1000.

Rigas, B., I. S. Goldman and L. Levine (1993). "Altered eicosanoid levels in human colon cancer." J Lab Clin Med 122(5): 518-523.

Rimokh, R., M. Gadoux, M. F. Bertheas, F. Berger, M. Garoscio, G. Deleage, D. Germain and J. P. Magaud (1993). "FVT-1, a novel human transcription unit affected by variant translocation t(2;18)(p11;q21) of follicular lymphoma." Blood 81(1): 136-142.

Roe, C. R., D. S. Millington, D. L. Norwood, N. Kodo, H. Sprecher, B. S. Mohammed, M. Nada, H. Schulz and R. McVie (1990). "2,4-Dienoyl-coenzyme A reductase deficiency: a possible new disorder of fatty acid oxidation." J Clin Invest 85(5): 1703-1707.

Rossmann, M. G., D. Moras and K. W. Olsen (1974). "Chemical and biological evolution of nucleotide-binding protein." Nature 250(463): 194-199.

Rouviere, P. E. and R. S. Wolfe (1988). "Novel biochemistry of methanogenesis." J Biol Chem 263(17): 7913-7916.

Saluda-Gorgul, A., K. Seta, M. Nowakowska and A. K. Bednarek (2011). "WWOX oxidoreductase--substrate and enzymatic characterization." Z Naturforsch C 66(1-2): 73-82.

Sambrook, J., Fritsch, E.F., Mainiatis, T. (1982). "Molecular Cloning: A laboratory manual." Cold Spring Harbour Laboratory Press.

Sambrook, J. and Russell D,W., (2006). "Identification of Associated Proteins by Coimmunoprecipitation." Cold Spring Harbour Laboratory Press.

Savitsky, P., J. E. Bray, C. D. O. Cooper, B. D. Marsden, P. Mahajan, N. A. Burgess-Brown and O. Gileadi (2010). "Highthroughput production of human proteins for crystallization: The SGC experience". J Struct Biol 172(1): 3-13.

Schulz, J. M., A. L. Watson, R. Sanders, K. L. Ross, J. B. Thoden, H. M. Holden and J. L. Fridovich-Keil (2004). "Determinants of function and substrate specificity in human UDP-galactose 4'-epimerase." J Biol Chem 279(31): 32796-32803.

Schwarz, M., et al. (2000). "The bile acid synthetic gene 3beta-hydroxy-Delta(5)-C(27)-steroid oxidoreductase is mutated in progressive intrahepatic cholestasis." J Clin Invest 106(9): 1175-1184.

Schwartz, M. F. and H. Jornvall (1976). "Structural analyses of mutant and wild-type alcohol dehydrogenases from drosophila melanogaster." Eur J Biochem 68(1): 159-168.

Shafqat, N., et al. (2003). "Expanded substrate screenings of human and Drosophila type 10 17beta-hydroxysteroid dehydrogenases (HSDs) reveal multiple specificities in bile acid and steroid hormone metabolism: characterization of multifunctional 3alpha/7alpha/7beta/17beta/20beta/21-HSD." Biochem J 376(Pt 1): 49-60.

Shafqat, N., J. R. Muniz, E. S. Pilka, E. Papagrigoriou, F. von Delft, U. Oppermann and W. W. Yue (2013). "Insight into S-adenosylmethionine biosynthesis from the crystal structures of the human methionine adenosyltransferase catalytic and regulatory subunits." Biochem J 452(1): 27-36.

Shafqat, N., J. Shafqat, G. Eissner, H. U. Marschall, K. Tryggvason, U. Eriksson, F. Gabrielli, H. Lardy, H. Jornvall and U. Oppermann (2006). "Hep27, a member of the short-chain dehydrogenase/reductase family, is an NADPH-dependent dicarbonyl reductase expressed in vascular endothelial tissue." Cell Mol Life Sci 63(10): 1205-1213.

Shaw, J. P., F. Schwager and S. Harayama (1992). "Substrate-specificity of benzyl alcohol dehydrogenase and benzaldehyde dehydrogenase encoded by TOL plasmid pWW0. Metabolic and mechanistic implications." Biochem J 283 (Pt 3): 789-794.

Sheibanie, A. F., J. H. Yen, T. Khayrullina, F. Emig, M. Zhang, R. Tuma and D. Ganea (2007). "The proinflammatory effect of prostaglandin E2 in experimental inflammatory bowel disease is mediated through the IL-23-->IL-17 axis." J Immunol 178(12): 8138-8147.

Shen, H. B. and K. C. Chou (2007). "Hum-mPLoc: an ensemble classifier for large-scale human protein subcellular location prediction by incorporating samples with multiple sites." Biochem Biophys Res Commun 355(4): 1006-1011.

Shigehara, Y., S. Okuda, G. Nemer, A. Chedraoui, R. Hayashi, F. Bitar, H. Nakai, O. Abbas, L. Daou, R. Abe, M. B. Sleiman, A. G. Kibbi, M. Kurban and Y. Shimomura (2016). “Mutations in SDR9C7 gene encoding an enzyme for vitamin A metabolism underlie autosomal recessive congenital ichthyosis.” Hum Molec Genet 25: 4484-4493.

Skach, W. R. (2000). "Defects in processing and trafficking of the cystic fibrosis transmembrane conductance regulator." Kidney Int 57(3): 825-831.

Smilda, T., A. H. Kamminga, P. Reinders, W. Baron, J. E. van Hylckama Vlieg and J. J. Beintema (2001). "Enzymic and structural studies on Drosophila alcohol dehydrogenase and other short-chain dehydrogenases/reductases." J Mol Evol 52(5): 457-466.

Smith, G. K., D. S Duch, M. P. Edelstein and E. C. Bigham (1992) New inhibitors of sepiapterin reductase. Lack of an effect of intracellular tetrahydrobiopterin depletion upon in vitro proliferation of two human cell lines.” J Biol Chem. 267(8):5599-5607

Spelberg, J. H., J. E. van Hylckama Vlieg, L. Tang, D. B. Janssen and R. M. Kellogg (2001). "Highly enantioselective and regioselective biocatalytic azidolysis of aromatic epoxides." Org Lett 3(1): 41-43.

Starčević Š., P. Božnar, S. Turk, S. Gobec S and T. L. Rižner (2011) “Design and synthesis of substrate mimetics based on an indole scaffold: potential inhibitors of 17β-HSD type 1.” Horm mol Biol Clin Investig. 61):201-209

Stammers, D. K., J. Ren, K. Leslie, C. E. Nichols, H. K. Lamb, S. Cocklin, A. Dodds and A. R. Hawkins (2001). "The structure of the negative transcriptional regulator NmrA reveals a structural superfamily which includes the short-chain dehydrogenase/reductases." EMBO J 20(23): 6619-6626.

Stornaiuolo, M., L. Lotti, N. Borgese, M. Torrisi, G. Mottola, G. Martire and S. Bonatti (2003). “KDEL and KKXX retrival signals appended to the same reporter protein determine different trafficking between endoplasmic reticulum, intermediate compartment and golgi complex”. Mol Biol Cell 14 (3): 889-902

Strain-Damerall, C., P. Mahajan, O. Gileadi and N. A. Burgess-Brown (2014). “Medium-throughoput production of recombinant human proteins: Ligation-independent cloning”. Methods Mol Biol 1091: 55-72

Structural Genomics, C., et al. (2008). "Protein production and purification." Nat Methods 5(2): 135-146.

Su, Y., K. I. Varughese, N. H. Xuong, T. L. Bray, D. J. Roche and J. M. Whiteley (1993). "The crystallographic structure of a human dihydropteridine reductase NADH binary complex expressed in *Escherichia coli* by a cDNA constructed from its rat homologue." J Biol Chem 268(36): 26836-26841.

**Su, Z., R. Li, X. Song, G. Liu, Y. Li, X. Chang, C. Li and D. Huang (2012). "Identification of a novel isoform of DHRS4 protein with a nuclear localization signal." Gene 494(2): 161-167.

Subbaramaiah, K., L. R. Howe, X. K. Zhou, P. Yang, C. A. Hudis, L. Kopelovich and A. J. Dannenberg (2012). "Pioglitazone, a PPARgamma agonist, suppresses CYP19 transcription: evidence for involvement of 15-hydroxyprostaglandin dehydrogenase and BRCA1." Cancer Prev Res (Phila) 5(10): 1183-1194.

Sullivan, F. X., R. Kumar, R. Kriz, M. Stahl, G. Y. Xu, J. Rouse, X. J. Chang, A. Boodhoo, B. Potvin and D. A. Cumming (1998). "Molecular cloning of human GDP-mannose 4,6-dehydratase and reconstitution of GDP-fucose biosynthesis in vitro." J Biol Chem 273(14): 8193-8202.

Suter, B., X. Zhang, C. G. Pesce, A. R. Mendelsohn, S. P. Dinesh-Kumar, J. H. Mao (2015). “Next-Generation Sequencing for Binary Protein-Protein Interactions.” Front Genet. 6:346

Suzuki, Y., L. L. Jiang, M. Souri, S. Miyazawa, S. Fukuda, Z. Zhang, M. Une, N. Shimozawa, N. Kondo, T. Orii and T. Hashimoto (1997). "D-3-hydroxyacyl-CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase bifunctional protein deficiency: a newly identified peroxisomal disorder." Am J Hum Genet 61(5): 1153-1162.

Szyperski, T., S. Scheek, J. Johansson, G. Assmann, U. Seedorf and K. Wuthrich (1993). "NMR determination of the secondary structure and the three-dimensional polypeptide backbone fold of the human sterol carrier protein 2." FEBS Lett 335(1): 18-26.

Tai, H. H., C. M. Ensor, M. Tong, H. Zhou and F. Yan (2002). "Prostaglandin catabolizing enzymes." Prostaglandins Other Lipid Mediat 68-69: 483-493.

Tanaka, N., K. Aoki, S. Ishikura, M. Nagano, Y. Imamura, A. Hara and K. T. Nakamura (2008). “Molecular basis for peroxisomal localization of tetrameric carbonyl reductase.” Structure 16:388-397.

**Tanaka, M., R. Bateman, D. Rauh, E. Vaisberg, S. Ramachandani, C. Zhang, K. C. Hansen, A. L. Burlingame, J. K. Trautman, K. M. Shokat and C. L. Adams (2005). "An unbiased cell morphology-based screen for new, biologically active small molecules." PLoS Biol 3(5): e128.

Tanaka, N., T. Nonaka, T. Tanabe, T. Yoshimoto, D. Tsuru and Y. Mitsui (1996). "Crystal structures of the binary and ternary complexes of 7 alpha-hydroxysteroid dehydrogenase from *Escherichia coli*." Biochemistry 35(24): 7715-7730.

Taneja, S. and F. Ahmad (1994). "Increased thermal stability of proteins in the presence of amino acids." Biochem J 303 (Pt 1): 147-153.

Tariq, M., Z. Azeem, G. Ali, M. S. Chishti and W. Ahmad (2009). "Mutation in the HPGD gene encoding NAD^+ dependent 15-hydroxyprostaglandin dehydrogenase underlies isolated congenital nail clubbing (ICNC)." J Med Genet 46(1): 14-20.

Tempel, W., et al. (2014). "Structural characterization of human cholesterol 7alpha-hydroxylase." J Lipid Res 55(9): 1925-1932.

Thatcher, D. R. (1980). "The complete amino acid sequence of three alcohol dehydrogenase alleloenzymes (AdhN-11, AdhS and AdhUF) from the fruitfly Drosophila melanogaster." Biochem J 187(3): 875-883.

Thoden, J. B. and H. M. Holden (1998). "Dramatic differences in the binding of UDP-galactose and UDP-glucose to UDP-galactose 4-epimerase from *Escherichia coli*." Biochemistry 37(33): 11469-11477.

Thoden, J. B., T. M. Wohlers, J. L. Fridovich-Keil and H. M. Holden (2000). "Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase." Biochemistry 39(19): 5691-5701.

Thompson, R. H. (1962). "Classification and nomenclature of enzymes." Science 137(3528): 405-408.

Timson, D. J. (2005). "Functional analysis of disease-causing mutations in human UDP-galactose 4-epimerase." FEBS J 272(23): 6170-6177.

Tonetti, M., L. Sturla, A. Bisso, U. Benatti and A. I. De Flora (1996). "Synthesis of GDP-L-fucose by the human FX protein." J Biol Chem 271(44): 27274-27279.

Upcroft, J. A. and P. Upcroft (1999). "Keto-acid oxidoreductases in the anaerobic protozoa." J Eukaryot Microbiol 46(4): 447-449.

Uppal, S., C. P. Diggle, I. M. Carr, C. W. Fishwick, M. Ahmed, G. H. Ibrahim, P. S. Helliwell, A. Latos-Bielenska, S. E. Phillips, A. F. Markham, C. P. Bennett and D. T. Bonthron (2008). "Mutations in 15-hydroxyprostaglandin dehydrogenase cause primary hypertrophic osteoarthropathy." Nat Genet 40(6): 789-793.

van den Bosch, B. J., M. Gerards, W. Sluiter, A. P. Stegmann, E. L. Jongen, D. M. Hellebrekers, R. Oegema, E. H. Lambrichs, H. Prokisch, K. Danhauser, K. Schoonderwoerd, I. F. de Coo and H. J. Smeets (2012). "Defective NDUFA9 as a novel cause of neonatally fatal complex I disease." J Med Genet 49(1): 10-15.

van Hylckama Vlieg, J. E., L. Tang, J. H. Lutje Spelberg, T. Smilda, G. J. Poelarends, T. Bosma, A. E. van Merode, M. W. Fraaije and D. B. Janssen (2001). "Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases." J Bacteriol 183(17): 5058-5066.

Vedadi, M., F. H. Niesen, A. Allali-Hassani, O. Y. Fedorov, P. J. Finerty, Jr., G. A. Wasney, R. Yeung, C. Arrowsmith, L. J. Ball, H. Berglund, R. Hui, B. D. Marsden, P. Nordlund, M. Sundstrom, J. Weigelt and A. M. Edwards (2006). "Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination." Proc Natl Acad Sci U S A 103(43): 15835-15840.

Venkatesan, R., S. K. Sah-Teli, L. O. Awoniyi, G Jiang, P. Prus, A. J. Kastaniotis, J. K Hiltunen, R. K. Wieranga and Z. Chen (2014). "Insights into mitochondrial fatty acid synthesis from the structure of heterotetrameric 3-ketoacyl-ACP reductase/3R-hydroxyacyl-CoA dehydrogenase." Nat Commun. 5:4805

Villarroya, A., E. Juan, B. Egestad and H. Jornvall (1989). "The primary structure of alcohol dehydrogenase from Drosophila lebanonensis. Extensive variation within insect 'short-chain' alcohol dehydrogenase lacking zinc." Eur J Biochem 180(1): 191-197.

Vogt, G. and P. Argos (1997). "Protein thermal stability: hydrogen bonds or internal packing?" Fold Des 2(4): S40-46.

Wahlberg, E., T. Karlberg, E. Kouznetsova, N. Markova, A. Macchiarulo, A. G. Thorsell, E. Pol, A. Frostell, T. Ekblad, D. Oncu, B. Kull, G. M. Robertson, R. Pellicciari, H. Schuler and J. Weigelt (2012). "Family-wide chemical profiling and structural analysis of PARP and tankyrase inhibitors." Nat Biotechnol 30(3): 283-288.

Winn, M. D., et al. (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr 67(Pt 4): 235-242.

Wan, Z. K., E. Chenail, H. Q. Li, C. Kendall, Y. Wang, S. Gingras, J. Xiang, W. W. Massefski, T. S. Mansour and E. Saiah (2011). "Synthesis of potent and orally efficacious 11beta-hydroxysteroid dehydrogenase type 1 inhibitor HSD-016." J Org Chem 76(17): 7048-7055.

Watmough, N. J. and F. E. Frerman (2010). "The electron transfer flavoprotein: ubiquinone oxidoreductases." Biochim Biophys Acta 1797(12): 1910-1916.

Webb, N. A., A. M. Mulichak, J. S. Lam, H. L. Rocchetta and R. M. Garavito (2004). "Crystal structure of a tetrameric GDP-D-mannose 4,6-dehydratase from a bacterial GDP-D-rhamnose biosynthetic pathway." Protein Sci 13(2): 529-539.

Weber, P. C. (1991). "Physical principles of protein crystallization." Adv Protein Chem 41: 1-36.

Wohlers, T. M., N. C. Christacos, M. T. Harreman and J. L. Fridovich-Keil (1999). "Identification and characterization of a mutation, in the human UDP-galactose-4-epimerase gene, associated with generalized epimerase-deficiency galactosemia." Am J Hum Genet 64(2): 462-470.

**Wohlers, T. M. and J. L. Fridovich-Keil (2000). "Studies of the V94M-substituted human UDPgalactose-4-epimerase enzyme associated with generalized epimerase-deficiency galactosaemia." J Inherit Metab Dis 23(7): 713-729.

Wu, X., P. Lukacik, K. L. Kavanagh and U. Oppermann (2007). "SDR-type human hydroxysteroid dehydrogenases involved in steroid hormone activation." Mol Cell Endocrinol 265-266: 71-76.

Xiao, H., V. Palhan, Y. Yang and R. G. Roeder (2000). "TIP30 has an intrinsic kinase activity required for up-regulation of a subset of apoptotic genes." EMBO J 19(5): 956-963.

Xiao, H., Y. Tao, J. Greenblatt and R. G. Roeder (1998). "A cofactor, TIP30, specifically enhances HIV-1 Tat-activated transcription." Proc Natl Acad Sci USA 95(5): 2146-2151.

Xie, Y. A., W. Lee, C. Cai, T. Gambin, K. Noupuu, T. Sujirakul, C. Ayuso, S. Jhangiani, D. Muzny, E. Boerwinkle, R. Gibbs, V. C. Greenstein, J. R. Lupski, S. H. Tsang and R. Allikmets (2014). “New syndrome with retinitis pigmentosa is caused by nonsense mutations in retinol dehydrogenase RDH11.” Hum Molec Genet 23: 5774-5780.

Yamamoto, H., A. Simon, U. Eriksson, E. Harris, E. L. Berson and T. P. Dryja (1999). "Mutations in the gene encoding 11-cis retinol dehydrogenase cause delayed dark adaptation and fundus albipunctatus." Nat Genet 22(2): 188-191.

**Yamaguchi, T., Y. Komoda and H. Nakajima (1994). "Biliverdin-IX alpha reductase and biliverdin-IX beta reductase from human liver. Purification and characterization." J Biol Chem 269(39): 24343-24348.

Yamanishi, Y., M. Hattori, M. Kotera, S. Goto and M. Kanehisa (2009). "E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs." Bioinformatics 25(12): i179-186.

Yang, J. and W. Zhang (2008). “WWOX tumour suppressor gene”. Histol Histopathol. 23(7): 877-882

Yang, S., R. F. Doolittle and P. E. Bourne (2005a). "Phylogeny determined by protein domain content." Proc Natl Acad Sci USA 102(2): 373-378.

**Yang, S. Y., X. Y. He and H. Schulz (2005b). "Multiple functions of type 10 17beta-hydroxysteroid dehydrogenase." Trends Endocrinol Metab 16(4): 167-175.

Yooseph, S. G., Sutton, D. B. Rusch, A. L. Halpern, S. J. Williamson, K. Remington, J. A. Eisen, K. B. Heidelberg, G. Manning, W. Li, L. Jaroszewski, P. Cieplak, C. S. Miller, H. Li, S. T. Mashiyama, M. P. Joachimiak, C. van Belle , J. M. Chandonia, D. A. Soergel, Y. Zhai, K. Natarajan, S. Lee, B. J. Raphael, V, Bafna, R. Friedman, S. E. Brenner, A. Godzik, D. Eisenberg, J. E. Dixon, S.S Taylor, R. L. Strausberg, M. Frazier and J. C. Venter (2007). “The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families”. Plos Biol 5(3):e16

Young, D. S. (1977). "Classification of enzymes and current status of enzyme nomenclature and units." Ann Clin Lab Sci 7(2): 93-98.

Yu, C. S., Y. C. Chen, C. H. Lu and J. K. Hwang (2006). "Prediction of protein subcellular localization." Proteins 64(3): 643-651.

Yuan, Z. and R. D. Teasdale (2002). "Prediction of Golgi Type II membrane proteins based on their transmembrane domains." Bioinformatics 18(8): 1109-1115.

Yue, W. W., M. Hassler, M. S. Roe, V. Thompson-Vale, Vivienne And L. H. Pearl (2007) **Ins**ights Into Histone Code Syntax From Structural And Biochemical Studies Of CARM1 Methyltransferase." Embo J 26: 4402-4412

Zhang, J., A. M. Dean, F. Brunet and M. Long (2004). "Evolving protein functional diversity in new genes of Drosophila." Proc Natl Acad Sci U S A 101(46): 16246-16250.

Zhang, J., T. D. Osslund, M. H. Plant, C. L. Clogston, R. E. Nybo, F. Xiong, J. M. Delaney and S. R. Jordan (2005). "Crystal structure of murine 11 beta-hydroxysteroid dehydrogenase 1: an important therapeutic target for diabetes." Biochemistry 44(18): 6948-6957.

Zhang, J., H. Yang, M. Long, L. Li and A. M. Dean (2010). "Evolution of enzymatic activities of testis-specific short-chain dehydrogenase/reductase in Drosophila." J Mol Evol 71(4): 241-249.

**Zhang, M., P. Hu and J. L. Napoli (2004). "Elements in the N-terminal signaling sequence that determine cytosolic topology of short-chain dehydrogenases/reductases. Studies with retinol dehydrogenase type 1 and cis-retinol/androgen dehydrogenase type 1." J Biol Chem 279(49): 51482-51489.

Zhang, Q., D. W. Piston and R. H. Goodman (2002). "Regulation of corepressor function by nuclear NADH." Science 295(5561): 1895-1897.

Zhang, Z. and A. G. Marshall (1998). "A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra." J Am Soc Mass Spectrom 9(3): 225-233.

Zhao, Y., J. Zhang, H. Li, Y. Li, J. Ren, M. Luo and X. Zheng (2008). "An NADPH sensor protein (HSCARG) down-regulates nitric oxide synthesis by association with argininosuccinate synthetase and is essential for epithelial cell viability." J Biol Chem 283(16): 11004-11013.

Zheng, X., X. Dai, Y. Zhao, Q. Chen, F. Lu, D. Yao, Q. Yu, X. Liu, C. Zhang, X. Gu and M. Luo (2007). "Restructuring of the dinucleotide-binding fold in an NADP(H) sensor protein." Proc Natl Acad Sci U S A 104(21): 8809-8814.

Zhou, H., L. Sun, J. Li, C. Xu, F. Yu, Y. Liu, C. Ji and J. He (2013). "The crystal structure of human GDP-L-fucose synthase." Acta Biochim Biophys Sin (Shanghai) 45(9): 720-725.

Ziegler, M. (2000). "New functions of a long-known molecule. Emerging roles of NAD in cellular signaling." Eur J Biochem 267(6): 1550-1564.

APPENDIX

Table 1 – Tablulation of human SDR family genomic Information – details of HUGO ID, Synonyms, chromosome location, SDR nomenclature, activity or function and substrate class.

	HUGO ID	Synonyms	Full name	Location	SDR nomenclature	Activity /Function	Substrate class	Reference
1	BDH1	BDH	3-hydroxybutyrate dehydrogenase, type 1	3q29	SDR9C1	(R)-3-hydroxybutanoate	Fatty acids and lipids	(Green et al. 1996)
2	BDH2	DHRS6	3-hydroxybutyrate dehydrogenase, type 2	4q24	SDR15C1	(R)-3-hydroxybutanoate	Fatty acids and lipids	(Guo et al. 2006)
3	BLVRB	FLR	biliverdin reductase B (flavin reductase (NADPH))	19q13.1-q13.2	SDR43U1	Flavins	Amino acid derivatives	(Yamaguchi et al. 1994)
4	CBR1		carbonyl reductase 1	21q22.1	SDR21C1	Quinones, prostaglandins, ,xenobiotics	Carbonyl xenobiotics, Retinoids	(Tanaka et al. 2005)
5	CBR3		carbonyl reductase 3	21q22.2	SDR21C2	4-benzoylpyridine and menadione	Carbonyl xenobiotics	(Miura et al. 2008)
6	CBR4	FLJ14431	carbonyl reductase 4	4q32.3	SDR45C1	Quinones	Carbonyl xenobiotics	(Endo et al. 2008)
7	DCXR	DCR, KIDCR	dicarbonyl/L-xylulose reductase	17q25.3	SDR20C1	Carbonyl xenobiotics and xylitol	Carbonyl xenobiotics	(Nakagawa et al. 2002; Matsunaga et al. 2006)
8	DECR1	DECR	2,4-dienoyl CoA reductase 1, mitochondrial	8q21.3	SDR18C1	2,4-dienoyl-CoA	Fatty acids and lipids	(Alphey et al. 2005)
9	DECR2	PDCR	2,4-dienoyl CoA reductase 2, peroxisomal	16p13.3	SDR17C1	2,4-dienoyl-CoA	Fatty acids and Lipids	(De Nys et al. 2001)
10	DHRS1	FLJ25430, MGC20204	dehydrogenase/reductase (SDR family) member 1	14q11.2	SDR19C1	Unknown	Orphan	
11	DHRS11	MGC4172	dehydrogenase/reductase (SDR family) member 11	17q12	SDR24C1	Unknown	Orphan	
12	DHRS12	FLJ13639	dehydrogenase/reductase (SDR family) member 12	13q14.3	SDR40C1	Unknown	Orphan	
13	DHRS13	MGC23280	dehydrogenase/reductase (SDR family) member 13	17q11.2	SDR7C5	Unknown	Orphan	
14	DHRS2	HEP27	dehydrogenase/reductase (SDR family) member 2	14q11.2	SDR25C1	3,4-Hexanedione, 2,3-Heptanedione and 1-Phenyl-1,2-propanedione	Carbonyl xenobiotic	(Gabrielli et al. 1995; Shafqat et al. 2006; Hawkins et al. 2007)
15	DHRS3	RDH17, retSDR1, Rsdr1, SDR1	dehydrogenase/reductase (SDR family) member 3	1p36.1	SDR1	All-trans-retinal	Retinoid	(Haeseleer et al. 1998)
16	DHRS4	FLJ11008, SDR-SRL	dehydrogenase/reductase (SDR family) member 4	14q11.2	SDR-SRL	3-keto steroids, alkyl phenyl ketones and α-dicarbonyl	Carbonyl xenobiotics, steroid and retinoid	(Matsunaga et al. 2006; Matsunaga et al. 2008; Su et al. 2012)
17	DHRS4L2		dehydrogenase/reductase (SDR family) member 4 like 2	14q11.2	SDR25C3	Unknown	Orphan	
18	DHRS7	retDSR4	dehydrogenase/reductase (SDR family) member 7	14q23.1	SDR34C1	Nicotine-derived nitrosamine ketone (NNK)	Carbonyl xenobiotics	
19	DHRS7B	CGI-93, DKFZp566O084, MGC8916	dehydrogenase/reductase (SDR family) member 7B	17p12	SDR32C1	Unknown	Orphan	
20	DHRS7C		dehydrogenase/reductase (SDR family) member 7C	17p13.1	SDR32C2	Unknown	Orphan	
21	DHRS9	3alpha-HSD, RDH15, RDHL, RETSDR8	dehydrogenase/reductase (SDR family) member 9	2q31.1	SDR9C4	3-alpha-hydroxysteroids,	Steroid	(Chetyrkin et al. 2001)
22	DHRSX	DHRS5X, DHRS5Y, DHRSXY, DHRSY	dehydrogenase/reductase (SDR family) X-linked	Xp22.33 and Yp11	SDR46C1	Unknown	Orphan	
23	FAR1	MLSTD2, FLJ22728	fatty acyl CoA reductase 1	11p15.2	SDR10E1	Long-chain-fatty acyl-CoAs	Faty acid and lipid	(Cheng and Russell 2004)

Table 1 – Tablulation of human SDR family genomic Information (continued)

24	**FAR2**	**MLSTD1, FLJ10462**	**fatty acyl CoA reductase 2**	**12p11.23**	**SDR10E2**	**Long-chain-fatty acyl-CoAs**	**Fatty acid and lipid**	**Cheng and Russell 2004)**
25	FASN	FAS	fatty acid synthase	17q25	SDR27X1	acetyl-CoA and malonyl-CoA	Fatty acid and lpid	(Jayakumar et al. 1995; Chi et al. 2006)
26	GALE		UDP-galactose-4-epimerase	1p36-p35	SDR1E1	UDP-glucose	Nucleotide sugar	(Wohlers and Fridovich-Keil 2000)
27	GMDS	GMD	GDP-mannose 4,6-dehydratase	6p25	SDR3E1	GDP-D-mannose	Nucleotide sugar	(Sullivan et al. 1998)
28	HPGD		hydroxyprostaglandin dehydrogenase 15-(Roe et al.)	4q34-q35	SDR36C1	PGE2	Fatty acid and lipid	(Clish et al. 2000; Cho et al. 2005)
29	HSD11B1	HSD11, HSD11B	hydroxysteroid (11-β) dehydrogenase 1	1q32-q41	SDR26C1	Cortiol, 7-ketosterol, Carbonyl xenobiotics,	Carbonyl xenobiotics, steroids	(Odermatt et al. 1999; Matsunaga et al. 2006)
30	HSD11B1L	SCDR10	hydroxysteroid (11-β) dehydrogenase 1-like	19p13.3	SDR26C2	Unknown	Orphan	
31	HSD11B2		hydroxysteroid (11-β) dehydrogenase 2	16q22	SDR9C3	Cortisone	Steroid	(Odermatt et al. 1999)
32	HSD17B1	EDH17B2, EDHB17	hydroxysteroid (17-β) dehydrogenase 1	17q11-q21	SDR28C1	17β-estradiol and DHT	Steroid	(Breton et al. 1996; Gangloff et al. 2003)
33	HSD17B10	HADH2, ABAD, CAMR, ERAB, MHBD, MRPP2	hydroxysteroid (17-β) dehydrogenase 10	Xp11.2	SDR5C1	3-hydroxyacyl-CoA, estradiol, 5α-androstanediol	Fatty acid and lipis, steroid	(He et al. 2003; Yang et al. 2005)
34	HSD17B11	DHRS8, PAN1B, RetSDR2	hydroxysteroid (17-β) dehydrogenase 11	4q22.1	SDR16C2	5 α-androstane-3 α, 17 β-diol	Steroid	(Chai et al. 2003)
35	HSD17B12	KAR	hydroxysteroid (17-β) dehydrogenase 12	11q11	SDR12C1	Estrone, long chain 3-ketoacyl-CoAs	Fatty acid and lipid, steroid	(Moon and Horton 2003; Luu-The et al. 2006)
36	HSD17B13	SCDR9	hydroxysteroid (17-β) dehydrogenase 13	4q22.1	SDR16C3	Unknown	Orphan	
37	HSD17B14	DHRS10, retSDR3	hydroxysteroid (17-β) dehydrogenase 14	19q13.33	SDR47C1	Oestradiol, 5-androstene-3-β,17-β-diol	Steroid	(Lukacik et al. 2007)
38	HSD17B2		hydroxysteroid (17-β) dehydrogenase 2	16q24.1-q24.2	SDR9C2	17β-estradiol, testosterone and DHT	Steroid	(Wu et al. 1993; Breton et al. 1996)
39	HSD17B3		hydroxysteroid (17-β) dehydrogenase 3	9q22	SDR12C2	Androstenedione	Steroid	(Geissler et al. 1994; Lindqvist et al. 2001)
40	HSD17B4	β-hydroxyacyl dehydrogenase, β-keto-reductase, D-3-hydroxyacyl-CoA dehydratase, D-bifunctional protein, peroxisomal multifunctional protein 2	hydroxysteroid (17-β) dehydrogenase 4	5q2	SDR8C1	straight-chain and 2-methyl-branched-chain fatty acids	Fatty acids and Lipids	(Jiang et al. 1997)
41	HSD17B6	3-hydroxysteroid epimerase, RODH	hydroxysteroid (17-β) dehydrogenase 6 homolog (mouse)	12q13	SDR9C6	5-α-androstan-3-alpha,17-β-diol , 17β-estradiol, all-trans-retinol	Steroids	(Huang and Luu-The 2000; Chetyrkin et al. 2001; Huang and Luu-The 2001)
42	HSD17B7	PRAP	hydroxysteroid (17-β) dehydrogenase 7	1q23	SDR37C1	zymosterone	Steroid	(Marijanovic et al. 2003)
43	HSD17B8	D6S2245E, H2-KE6, HKE6, KE6, RING2, FABGL	hydroxysteroid (17-β) dehydrogenase 8	6p21.3	SDR30C1	17β-estradiol, testosterone	Steroid	(Ohno et al. 2008; Chen et al. 2009)
44	HSD3B1	HSD3B, HSDB3	hydroxy-delta-5-steroid dehydrogenase, 3 β- and steroid delta-isomerase 1	1p12	SDR11E1	3-β-hydroxy-Delta(5)-steroid	Steroid	(Lorence et al. 1990)
45	HSD3B2		hydroxy-delta-5-steroid dehydrogenase, 3 β- and steroid delta-isomerase 2	1p12	SDR11E2	3-β-hydroxy-Delta(5)-steroid	Steroid	(Rheaume et al. 1991)
46	HSD3B7	C(27)-3β-HSD	hydroxy-delta-5-steroid dehydrogenase, 3 β- and steroid delta-isomerase 7	16p11.2	SDR11E3	7-α-hydroxylated sterols	Steroid	(Schwarz et al. 2000)
47	HSDL1		hydroxysteroid dehydrogenase like 1	16q24	SDR12C3	Inactive	Regulatory	(Meier et al. 2009)
48	HSDL2		hydroxysteroid dehydrogenase like 2	9q32	SDR13C1	Unknown	Orphan	
49	HTATIP2	CC3, FLJ26963, TIP30	HIV-1 Tat interactive protein 2, 30kDa	11p15.1	SDR44U1	HIV-1 TAT-interacting protein	Regulatory	(Xiao et al. 1998; King and Shtivelman 2004)
50	KDSR	FVT1, DHSR,	3-ketodihydrosphingosine reductase	18q21	SDR35C1	3-ketodihydrosphingosine	Fatty acid and lipids	(Kihara and Igarashi 2004)

Table 1 – Tablulation of human SDR family genomic Information (continued)

51	**MAT2B**		**methionine adenosyltransferase II, β**	**5q34-q35**	**SDR23E1**	**Non-catalytic regulatory subunit of S-adenosylmethionine synthetase 2 (MAT2A)**	**Regulatory**	**(LeGros et al. 2000)**
52	NDUFA9	NDUFS2L, CI-39k	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9, 39kDa	12p13.3	SDR22E1	Accessory subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase (Complex I), that is believed not to be involved in catalysis	Regulatory	(Yamaguchi et al. 1994; Murray et al. 2003)
53	NMRAL1		NmrA-like family domain containing 1	16p13.3	SDR48A1	Redox sensor protein that regulates ASS1 activity	Regulatory	(Zhao et al. 2008)
54	NSDHL	H105e3	NAD(Roe et al.) dependent steroid dehydrogenase-like	Xq28	SDR31E1	Lanosterol	Steroid	(Konig et al. 2000)
55	PECR	HSA250303, TERP	peroxisomal trans-2-enoyl-CoA reductase	2q35	SDR29C1	***trans***-2-enoyl-CoAs	Fatty acid and lipid	(Das et al. 2000)
56	QDPR	DHPR, PKU2	quinoid dihydropteridine reductase	4p15.31	SDR33C1	5,6,7,8-tetrahydropteridine	Vitamin and amino acid derivatives	(Lockyer et al. 1987)
57	RDH10		retinol dehydrogenase 10 (all-trans)	8q21.11	SDR16C4	All-trans-retinol	Retinoid	(Wu et al. 2002)
58	RDH11	ARSDR1, MDT1	retinol dehydrogenase 11 (all-trans/9-cis/11-cis)	14q24.1	SDR7C1	9-cis, 13-cis and all-trans-retinal	Retinoid	(Kedishvili et al. 2002)
59	RDH12	FLJ30273, LCA13	retinol dehydrogenase 12 (all-trans/9-cis/11-cis)	14q24.1	SDR7C2	9-cis and all-trans-retinol	Retinoid	(Haeseleer et al. 2002)
60	RDH13		retinol dehydrogenase 13 (all-trans/9-cis)	19q13.42	SDR7C3	all-***trans***-retinaldehyde, all-***trans***-retinol	Retinoid	(Belyaeva et al. 2008)
61	RDH14	PAN2	retinol dehydrogenase 14 (all-trans/9-cis/11-cis)	2p24.2	SDR7C4	9-cis and all-trans-retinol	Retinoid	(Belyaeva and Kedishvili 2002)
62	RDH16	RODH-4	retinol dehydrogenase 16 (all-trans)	12q13.3	SDR9C8	all-trans-retinol, 13-cis-retinol, androstanediol and androsterone	Retinoid	(Gough et al. 1998)
63	RDH5	RDH1, HSD17B9	retinol dehydrogenase 5 (11-cis/9-cis)	12q13-q14	SDR9C5	11-cis retinol	Retinoid	(Mertz et al. 1997; Zhang et al. 2004)
64	RDH8	PRRDH	retinol dehydrogenase 8 (all-trans)	19p13.2	SDR28C2	All-trans-retinol	Retinoid	(Rattner et al. 2000)
65	SDR16C5	RDHE2	short chain dehydrogenase/reductase family 16C, member 5	8q12.1	SDR16C5	All-trans-retinol	Retinoid	(Lee et al. 2009)
66	SDR39U1	C14ORF124	short chain dehydrogenase/reductase family 39U, member 1	14q12	SDR39U1	Unknown	Orphan	
67	SDR42E1	HSPC105	short chain dehydrogenase/reductase family 42E, member 1	16q23.3	SDR42E1	Unknown	Orphan	
68	SDR9C7	RDHS, SDR-O	short chain dehydrogenase/reductase family 9C, member 7	12q13.3	SDR9C7	all-trans-retinal	Retinoid	(Gough et al. 1998)
69	SPR		sepiapterin reductase (7,8-dihydrobiopterin:$NADP^+$ oxidoreductase)	2p14-p12	SDR38C1	Sepiapterin	Vitamin and amino acid derivative	(Fujimoto et al. 2002)
70	TGDS	TDPGD	TDP-glucose 4,6-dehydratase	13q32.1	SDR2E1	dTDP-glucose	Nucleotide sugar	Inferred from electronic annotation
71	TSTA3	FX, "GDP-L-fucose synthase", P35B,	tissue specific transplantation antigen P35B	8q24.3	SDR4E1	GDP-L-fucose	Nucleotide sugar	(Tonetti et al. 1996)
72	UXS1	FLJ23591	UDP-glucuronate decarboxylase 1	2q12.2	SDR6E1	UDP-glucuronorate	Nucleotide sugar	(Bakker et al. 2009)
73	WWOX	FOR	WW domain containing oxidoreductase	16q23.3-q24.1	SDR41C1	Negative regulation with Glycogen synthase kinase 3***β***	Orphan	

Table 2 – Human SDR family experimental and predicted subcellular localization and targeting signals

HUGO ID	SDR Cluster	Subcellular Localization					Target Signal			
		Experimental	Hum-mPloc	MultiLoc	Cello	WoLF PSORT	Signal Sequence	Position	Type	Signal Source
BDH1	C2	Mito	Mito	Mito	Mito	Mito	MLATRLSRPLSRLP GKTLSACDRENGA RRPLLLGSTSFIPIGR RTY	N-term	MTS	Mitoprot
BDH2	C1	Cyto	Mito	Cyto	Cyto	Mito				
BLVRB	Aytpical	Cyto	Cyto	Perox	Cyto	Cyto				
CBR1	C1	Cyto	Cyto	Cyto	Cyto	Cyto				
CBR3	C1	Cyto	Cyto	Mito	Cyto	Mito				
CBR4	C1	Mito	Mito	Mito	Mito	Cyto				
DCXR	C1	Cyto	Cyto	Mito	Mito	Mito				
DECR1	C1	Mito	Mito	Mito	Mito	Mito	MKLPARVFFTLGSR LPCGLAPRRFFSYG TKILYQ	N-term	MTS	Mitoprot
DECR2	C1	Perox	Perox	Perox	PM	Cyto	AKL	C-term	PTS1	PTS1 predictor
DHRS1	C1	Undefined	ER	Cyto	Mito	Extr				
DHRS2	C1	Mito/ Nucl/ Perox	Nucl	Mito	Mito/ Nucl	Mito/ Nucl	MSSTGIDRKGVLA NRVAVVTGSTSGIG FAI	N-term	Mitochondrial	(Deisenroth et al. 2010)
							RRLARDGAHVVISS RK	N-term	NLS	(Filling et al. 2001)
							STRL	C-term	PTS1	PTS1 predictor
							LTRTLALEL	Central region	NES	NetNES predictor
DHRS3	C3	ER	ER	ER	PM	PM	MVWKRLGALVMF PLQMIYLVVKA	N-term	Signal anchor	(Deisenroth et al. 2011)
DHRS4	C1	Perox	Perox	Mito	Mito	Mito	SRL	C-term	PTS1	PTS1 predictor
DHRS4L2	C1	Undefined	Mito	Mito	Mito	Mito	MARLLGLCAWARK SVRLASS	N-term	MTS	Mitoprot
							RKKRK	C-term	NLS	MultiLoc
DHRS7	C3	ER	ER	ER	PM	ER	MNWELLLWLLVLC ALLLLLVQLLRFLRA	N-term	Signal anchor	SignalP/Phobius
							KTKHD	C-term	ER retention	Manual
DHRS7B	C3	Undefined	Mito	Perox	PM	PM	MVSPATRKSLPKVK A	N-term	MTS	MitoProt
DHRS7C	C3	ER	ER	PM	PM	ER	MGVMAMLMLPLL LLGISG	N-term	Signal peptide	SignalP/Phobius
							KEKLN	C-term	ER retention	Manual
DHRS9	C2	ER	ER	ER	Cyto	ER	MLFWVLGLLILCGF LWT	N-term	Signal peptide	SignalP/Phobius
DHRS11	C1	Undefined	Mito	Mito	Mito	Extr	MARPGMERWRDR LALVTGASGGIGAA VARA	N-term	Signal anchor	SignalP/Phobius
DHRS12	C2	Undefined	Cyto	Cyto	Cyto	Mito				
DHRS13	C2	Undefined	Mito	ER	PM	Extr	MEALLLGAGLLLGA YVLVYYNLKA	N-term	Signal anchor	SignalP/Phobius
DHRSX	C2	Undefined	Mito	Mito	Mito	Extr	MSPLSAARAALRVY AVGAAVILAQLLRR CRGG	N-term	MTS	Mitoprot
FAR1	Aytpical	Perox	ER/Perox/ Mem	Perox	PM	Cyto				
FAR2	Aytpical	Perox	ER	Perox	PM	Cyto				
FASN	Aytpical	Cyto	Cyto	Cyto	Cyto	Cyto				
GALE	Extended	Undefined	Lysosome/ Mito	Cyto	Cyto	Cyto				
GMDS	Extended	Cyto	Cyto	Perox	Cyto	Mito/ Cyto				
HPGD	C1	Cyto	Mito	Cyto	Cyto	Cyto				
HSD3B1	Extended	ER/ Mito	ER	Perox	PM	PM	KSKTQ	C-term	ER Retention	Manual
HSD3B2	Extended	ER/ Mito	ER	Perox	PM	PM	KSKTQ	C-term	ER Retention	Manual
HSD3B7	Extended	ER	ER	Cyto	PM	ER/ Mito				
HSD11B1	C3	ER	ER	Extr	PM	Pero	LLPILGLFMAYYYS AN	N-term	Signal anchor	(Odermatt et al. 1999)
HSD11B1L	C3	Undefined	ER	ER	Mito	Extr	MKVLLLTGLGALFF A	N-term	Signal anchor	SignalP/Phobius

Table 2 – Human SDR family experimental and predicted subcellular localization and targeting signals (continued)

HSD11B2	**C3**	**ER**	**ER**	**Lyso**	**PM**	**Extr**	**SGGAWLLVAARALLQLL**	**N-term**	**Signal anchor**	**(Odermatt et al. 1999)**
HSD17B1	C2	Cyto	Cyto	PM	Cyto	Mito	LPLLRMRL	C-term	NES	NetNES predictor
HSD17B2	C2	ER	ER	Lyso	PM	PM	FFSDTAWICLAVPTVLCGTVF	N-term	Signal anchor	SignalP/Phobius
							KKKAT	C-term	ER retention	Filling et al. 2001)
HSD17B3	C3	ER	ER	ER	PM	Extr				
HSD17B4	C3	Perox	Perox	Mito	Mito	Cyto	AKL	C-term	PTS1	PTS1 predictor
HSD17B6	C2	ER	Centriole/ER	PM	PM	Pero	MWLYLAAFVGLYYLLHW	N-term	Signal anchor	SignalP/Phobius
HSD17B7	C2	ER	ER	Perox	PM	Mito				
HSD17B8	C1	Mito	Cyto/ Mito	Mito	PM	Mito				
HSD17B10	C1	Mito	Mito	Mito	Mito	Mito	MAACRSVKGLVAVITG	N-term	Non-cleavable MTS	Filling et al. 2001)
HSD17B11	C3	Undefined	Cyto/ER	ER	PM	Mito	MKFLLDILLLLPLLIVCSL	N-term	Signal peptide	SignalP/Phobius
							KMKAQ	C-term	ER retention	Manual
HSD17B12	C3	ER	ER	Perox	PM	Cyto	KTKKN	C-term	ER retention	Uniprot
HSD17B13	C3	Undefined	Cyto/ER	ER	PM	ER	MNIILEILLLLITIIYSYLE	N-term	Signal peptide	SignalP/Phobius
HSD17B14	C1	Cyto	Mito	Perox	Cyto	Mito				
HSDL1	C3	Mito	ER	Perox	PM	PM	MAAVDSFYLLYREIARSCNCYMEALALVGAWYTARKSITICDFYSLIRLHFIPRLGSRADLIKQYGRWAVVSGATDGIGKA	N-term	Mitochondrial translocation	(Meier et al. 2009)
HSDL2	C1	Perox	Mito/ Perox	Mito	Cyto	Cyto	ARL	C-term	PTS1	PTS1 predictor
HTATIP2	Atypical	Cyto/ Nucl	Cyto/ Nucl	Cyto	Cyto	Cyto				
KDSR	C3	ER	ER	ER	PM	ER	MLLLAAAFLVAFVLLLYMVSPLISP	N-term	Signal anchor	SignalP/Phobius
MAT2B	Extended	Undefined	Nucl	Cyto	Cyto	Cyto				
NDUFA9	Atypical	Mito	Mito	Mito	Mito	Mito	MAAAQSRVVRVLSMSRSAITAIATSVCHGPPCRQ	N-term	MTS	Mitoprot
NMRAL1	Atypical	Cyto	Nucl	Cyto	Cyto	Cyto	IELTLRL	N-term	NES	(Zhang et al. 2012)
NSDHL	Extended	ER	ER	Perox	Cyto	Cyto	ALLLSLLVMVI	C-term	NES	NetNES predictor
PECR	C1	Perox	Perox	Mito	Mito	Cyto	AKL	C-term	PTS1	PTS1 predictor
QDPR	Atypical	Undefined	Cyto	Mito	Mito	Cyto				
RDHE2	C3	ER	ER	ER	PM	Extr	KKKL	C-term	ER retention	Manual
RDH5	C2	ER	ER/Mem	Lyso	PM	Extr	MWLPLLGALLWAVLWLLRDRQSLP	N-term	Signal peptide	SignalP/Phobius
RDH8	C2	Membrane of photoreceptor cells	Cyto/ Mem	Perox	PM	Extr				
RDH10	C3	ER	Cyto/ Mito	ER	PM	Extr	IVVEFFVVTFKVLWAFVLAAA	N-term	Signal anchor	SignalP/Phobius
RDH11	C2	ER	ER	PM	Extr	Extr	MVELMFPLLLLLLPFLLYMA A	N-term	Signal peptide	SignalP/Phobius
RDH12	C2	ER	ER	ER	Mito	Extr	MLVTLGLLTSFFSFLYMVAPSIRKFFAGGVCRTNV	N-term	Signal peptide	SignalP/Phobius
RDH13	C3	Mito	Mito	ER	Mito	Extr	MSRYLLPLSALGTVAGAAVLLKDYVTGGACPSKATIPGKTVIVTGANTGIGKQTALELARR	N-term	MTS	Mitoprot
RDH14	C2	ER	ER	Mito	Mito	Mito	MAVATAAAVLAALGGALWLAA	N-term	Signal anchor	SignalP/Phobius
RDH16	C2	ER	ER	ER	PM	Extr				
SDR9C7	C2	Cyto	Mito	Perox	Mito	Cyto				

Table 2 – Human SDR family experimental and predicted subcellular localization and targeting signals (continued)

SDR39U1	Atypical	Undefined	Mito	Perox	Mito	Nucl	LPFRLGL	Central	NES	NetNES predictor
SDR42E1	Extended	Undefined	ER	Perox	PM	Pero				
SPR	C1	Cyto	Cyto	Extr	Cyto	Mito				
TGDS	Extended	Undefined	Mito	Perox	Cyto	Mito				
TSTA3	Extended	Undefined	Cyto	Perox	Cyto	Cyto				
UXS1	Extended	Golgi	Mem	Golgi	Mito	Mito	MVSKALLRLVSAVN RRRMK LLLGIALLAYVASV WG	N-term	Signal Anchor	SignalP/P hobius
WWOX	C2	Cyto./ Golgi/Nucl	Nucl	Cyto	Extr	Cyto/ Nucl	GKRKRV	N-term	NLS	Uniprot

Table 3 – Human SDR proteins with transmembrane regions and associated subcellular localisation.

HUGO ID	SDR Cluster	Uniprot predicted transmembrane region	AA Position	Subcellular localization	Membrane Association
DHRS3	C3	LVMFPLQMIYLV VKAAVGLVL	9-24	ER	Yes
		IVCLNSVLALSAIPGAIDYCT	170-190		
		AFAFMESLTLGLLDCPGVSAT	195-215		
		AVQLNQALLLLPWTMHALVIL	253-273		
HSD17β11	C3		4-23	n.d	
RDH10	C3	IVVEFFVVTFKVLWAFVLAAA	2-23	ER	Yes
SDR16C5	C3	LFIFLGKSLFSLLEAMIFAL L	11-31	ER	Yes
		L LYFMMFLKSFLPLKTGLLIA	270-290		
KDSR	C3	GYMLSALTCGMAPVTSITEGL	271-291	ER	Yes
		VVTMGLFRTIALFYLGSFDSI	294-314		
HSD17β3	C3	VLEQFFILTGLLVCLACLAKC	4-24	ER	Yes
HSD17β12	C3	ALPAAGFLYWVGAGTVAYLAL	4-24	ER	Yes
		GAILNISSGSGMLPVPLLTIY	182-202		
		GYLIHALMGS IISNLPSWIYL	271-291		
HSD11β1	C3	LLPILGLFMAYYYYSAN	8-24	ER	Yes
HSD11β1L			1-15	n.d	
DHRS7C	C3		4-23	ER	
DHRS7β	C3		18-38	n.d	
DHRS7	C3		4-23	ER	
HSD17β7	C2	LTYGILPPFIWTLLMPAILLL	230-250	ER	Yes
DHRSX	C2		98–115	ER	
RDH14	C2		4–26	ER	
			161–178		
			260–276		
RDH13	C2		243–261	Mitochondria	
RDH12	C2		5–22	ER	
			120–137		
RDH11	C2	MVELMFPLLLLLLPFLLYMA A	1-21	ER	Yes
RDH8	C2	VLVNNAGMGLVGPLEGLSLAA	85-106	Photoreceptor outer segments	Yes
		IVVISSVMGLQGVIFNDVYAA	137-157		
		LAIQLLQFNIFISLVEPGPVV	169-189		
HSD11β2	C2		2-24	ER	Yes
HSD17β2	C2	FFSDTAWICLAVPTVLCGTVF	4-24	ER	Yes
DHRS9	C2		136–152	ER	
			159–175		
RDH5	C2		24–41	ER	Yes
			104–121		
			132–151		
			158–177		
RDH16	C2		130–152	ER	Yes
		LLYLPMSYMPTFLVDAIMYWV	289-309		
HSD17β6	C2		133–152	ER	
			159–178		
FAR2	Aytpical	NIHYLFNTALFLIAWRLLIA	465-484	Peroxisome	Yes
		NVWFFIVSFCYKFLSYFRAS	491-510		
FAR1	Aytpical	IRYGFNTILVILIWRIFI	466-483	Peroxisome	n.d
SDR42E1	Extended	LPLTLVYCFAFLTEMVHFILG	282-302	n.d	n.d
		GLLVFLLIIAVLMWLPSSVIL	371-391		
HSPC105	Extended		282–302		
			371–391		
NSDHL	Extended	WVAYYLALLLSLLVMVISPVI	298-318	ER	Yes

Table 3 – Human SDR proteins with transmembrane regions and associated subcellular localisation. (continued)

HSD3β7	**Extended**	**LLPYWLLVFLAALNALLQWLL**	**289-309**	**ER**	**Yes**
		PLVLYAPLLN PYTLAVANTTF	311-331		
HSD3β1	Extended	LSLMYWIGFLLEIVSFLLRPI	288-308	ER /Mitochondria	Yes
HSD3β2	Extended	LTLMYWIGFLLEVVSFLLSPI	287-307	ER/Mitochondria	Yes

Table 4 – Human SDRs tested against known substrate/s for thermal stability in DSF ligand library.

SDR	Substrate/s	T_m shift °C
BDH2	(R)-3-hydroxybutyrate acetoacetate	b.t b.t
CBR1	1,2-napthoquinone 1,4 benzoquinone 1,4-chresequinone 1,4-napthoquinone 2,6-dimethylquinone 9,10-phrenanthrenequinone Coenzyme Q10 Droperidol Daunorubicin Isatin Menadione PGE2	b.t b.t b.t b.t b.t b.t b.t b.t b.t b.t b.t b.t
CBR3	1,2-napthoquinone 9,10-phrenanthrene quinone Isatin Menadione	b.t b.t b.t b.t
DCXR	1-Phenyl-1,2-propanedione 2,3-Heptanedione 3,4-Hexanedione Isatin	b.t b.t b.t b.t
DHRS2B	1-Phenyl-1,2-propanedione 2,3-Heptanedione 3,4-Hexanedione	b.t b.t b.t
DHRS4	5α-DHT 9,10-phrenanthrenequinone All-trans-retinal Allopregane-3,20 dione Isatin Menadione Pyrimidine-4-aldehyde	b.t b.t b.t b.t b.t b.t b.t
GALE	UDP-glucose UDP-galactose UDP-GalNAc UDP-GlcNac	b.t/1.7 NAD/H b.t/1.6 NAD/H b.t/b.t b.t/2.7 NAD/H
HPGD	PGE2	5.1/b.t NAD/H
HSD11B1	4-nitroacetophenone 7-ketosterol Cortisol Ketoprofen Menadione	b.t b.t b.t b.t b.t/1.4 NADP/H
HSD17B10	2-methyl-3-hydroxybutyryl-CoA 5α-androstane, 3β,17β-diol 5α-DHT Acetoacetyl-CoA Chenodeoxycholic acid Cholic acid Cortisol Cortisone Corticosterone Dehydrocorticosterone Dihydroandrosterone Estradiol Estrone Testosterone	b.t 1.9/b.t b.t b.t b.t b.t b.t b.t b.t b.t 5.4/b.t NAD/H 3.2/b.t NAD/H b.t/b.t 1.5/b.t NAD/H
HSD17B12	acetoacetyl-CoA Estradiol Estrone	b.t b.t b.t
HSD17B4	Estradiol Estrone	b.t b.t

Table 4 – Human SDRs tested against known substrate/s for thermal stability in DSF ligand library (continued).

HSD17B8	**2-methyl-3-hydroxybutyryl-CoA** **Androstanedione** **Androstenedione** **Dihydrostestosterone** **DL-hydroxybutyryl-CoA** **Estradiol** **Estrone** **Malonyl-CoA** **Testosterone**	**b.t** **b.t** **b.t** **b.t** **b.t** **b.t** **b.t** **b.t** **b.t**
PECR	Acetyl-CoA Butyryl-CoA Crotonyl-CoA Decanoyl-CoA Hexanoyl-CoA Lauroyl-CoA Myristoyl-CoA Oleoyl-CoA Palmitoyl-CoA Palmitoleoyl-CoA Stearoyl-CoA	**b.t** **b.t** **b.t** **5.7/2.5** **3.5/b.t** **5.1/1.1** **b.t** **b.t** **b.t** **1.1/3.0** **b.t**
SPR	Sepiapterin Menadione	b.t/b.t b.t/b.t
UXS1	UDP-glucuronorate	b.t/ 7.9 NAD/H

Table 5 – Thermal stability of SDRs with compounds classes in DSF ligand library and a comparison to their known substrate class.

SDR HUGO ID	COMPOUND CLASSES	Thermal shift within substrate class
BDH2	Lipid	Yes
CBR1	Carbonyl Xenobiotic Steroids	Yes Yes
CBR3	Carbonyl Xenobiotic Steroid	No No
DCXR	Carbonyl Xenobiotic Sugar	Yes No
DHRS2B	Carbonyl Xenobiotic	Yes
DHRS4	Carbonyl Xenobiotic Retinoid	Yes No
GALE	Sugar	Yes
GMDS	Sugar	Yes
HPGD	Fatty acid and Lipid	Yes
HSD11B1	Carbonyl xenobiotics Steroid	Yes Yes
HSD17B10	Fatty acid and Lipid Steroid	Yes Yes
HSD17B12	Fatty acid and Lipid Steroid	Yes Yes
HSD17B4	Fatty acid and Lipid	No
HSD17B8	Steroid	Yes
PECR	Fatty acid and Lipid	Yes
QDPR	Vitamin and amino acid derivatives	No
SPR	Vitamin and amino acid derivatives	No
TSTA3	Nucleotie sugar	Yes
UXS1	Nucleotie sugar	Yes

Table 6 – Human SDRs with known inhibitors in DSF igand library.

SDR	Inhibitor	T_m Shift °C
CBR1	Genestein Quercetin Triclosan	b.t/2.4 NADP/H 2.1/5.9 NADP/H 2.1/b.t NADP/H
HPGD	Celecoxcib Diclofenac Fenoprofen Ketoprofen Indomethacin Niflumic acid Pioglitazone Rosiglitazone Sulindac	b.t/b.t NAD/H 1.4/b.t NAD/H 3.8/b.t NAD/H 3.2/b.t NAD/H 2.1/b.t NAD/H 1.9/b.t NAD/H 5.2/1.5 NAD/H 2.7/b.t NAD/H 1.5/b.t NAD/H
HSD11B1	CBX Enoxolone	7.0/6.6 NADP/H 6.4/5.7 NADP/H
HSD17B4	Gossypol	b.t/1.4 NAD/H
HSD17B10	Dilsulfiram Hydroquinone Quercetin	4.3/2.7 NAD/H b.t/b.t NAD/H 9.4/b.t NAD/H
MAT2B	Resvetrol	1.8/1.8 NADP/H
SPRA	N-acteyl Serotonin	b.t/b.t

Table 7 – Mechanism of action for selected inhibitors and selectivity of human 15-PGDH A) PGE2 titration B) NAD titration c) profiling of key inhibitors against related dehydrogenase/reductase enzymes.

A)

Compound	Compound structure	Conc. [nM]	Titration of PGE2				Inhibition Mode
			K_m [µM]	V_{max} [µmol/(min x mg)]	K_{cat} [1/sec]	K_{cat}/K_m [x10^6]	
13		10	3.7 ±0.5	21.3	10.3 ±0.3	2.8	Uncompetitive
		20	2.8 ±0.4	12.6	3.1 ±0.2	2.1	
72		4	5.6 ±1.1	22.8	11.1 ±0.5	2.0	Noncompetitive
		10	6.4 ±0.6	15.4	7.5 ±0.5	1.2	
61		10	7.5 ±0.5	29.0	14.1 ±0.2	1.8	Competitive
		50	13.5 ±1.3	25.6	12.4 ±0.3	0.9	

B)

Compound	Conc. [nM]	Titration of NAD				Inhibition Mode
		K_m [µM]	V_{max} [µmol/(min x mg)]	K_{cat} [1/sec]	K_{cat}/K_m [x10^6]	
13	10	10.7 ±0.6	17.5	8.5 ±0.1	0.8	Noncompetitive
	20	8.9 ±0.8	8.2	4.0 ±0.1	0.4	
72	4	n.d	n.d	n.d	n.d	N/A
	10	n.d	n.d	n.d	n.d	
61	10	17.0 ±0.8	21.1	10.3 ±0.1	0.6	Noncompetitive
	50	11.1 ±1.2	10.2	5.0 ±0.1	0.4	

C)

Compound	IC50/nm			
	HPGD	ALDHA1	HADH2	HSD17B4
13	56	36000	>57500	>57500
72	89	>57500	>57500	>57500
61	141	>57500	n.d	n.d

Table 8 – Crystallographic data collection of SDRs

Parameter	DHRS4	$NMRAL1_{NFL}$	$NMRAL1_{ZZ0}$	HSDL2
PDB ID	3O4R	2WM3	2WMD	3KVO
Space group	P 21 21 21	P 41 21 2	P2 21 21	P 21 21 21
Wavelength	Single	Single	Single	Single
Unit cell a, b, c	60.07, 132.2, 133.8	68.7, 68.7, 183.8	57.7, 79.7, 86.3	61.2, 78.3, 121.8
Unit cell α β γ	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range Å	29.9-1.7	55.05-1.85	50.0-2.0	48.07-2.25
Number of reflections (observed)	110448	38676	27651	28262
Number of unique reflections				17872
R_{merge}	0.07	0.12	0.1	0.895
$\langle I \rangle / \langle \sigma i \rangle$	2.16	1.98	1.89	2.23
Completeness (%)	98.6	99.9	100.0	99.5
Anomalous redundancy	98.6	99.9	100.0	99.5
Refinement				
***R* factor**	0.210	0.166	0.202	0.199
***R* free**	0.250	0.197	0.232	0.247
Rmsd bond lengths (å)	0.50	0.79	0.85	0.60
Rmsd bond angles (^{o})	0.60	078	0.82	0.66

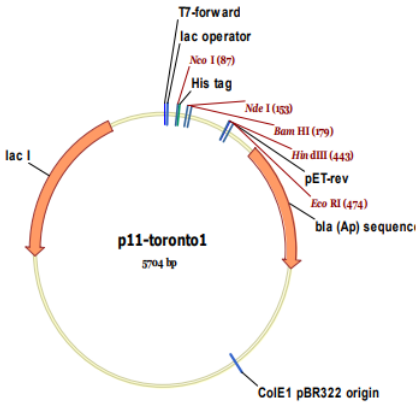


Figure 1. The p11 vector used for cloning of SDR constructs. This is a pET derived expression vector with His_6 tag in 22-aa N-terminal fusion peptide, with TEV protease cleavage site. Depicted in the figure are the genes for Lac I promoter and Ampicillin resistance gene, respectively (orange arrows), the main sites important for cloning (black denotations), and the sites of restriction enzymes (maroon denotations). https://thesgc.org/sites/default/files/toronto_vectors/p11.pdfc

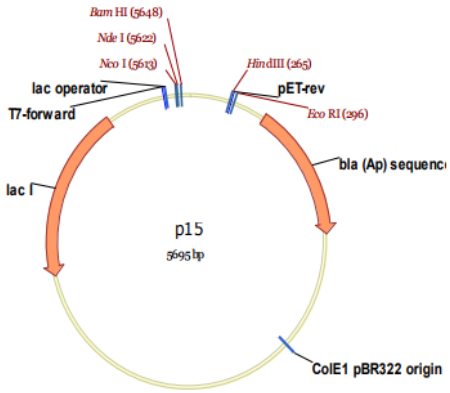


Figure 2: The p15 vector used for cloning of SDR constructs. This is a pET derived expression vector with C-terminal His6 tag preceded with a TEV protease cleavage site. Depicted in the figure are the genes for Lac I promoter and Ampicillin resistance gene, respectively (orange arrows), the main sites important for cloning (black denotations), and the sites of restriction enzymes (maroon denotations). https://www.thesgc.org/sites/default/files/oxford_vectors/p15.pdf

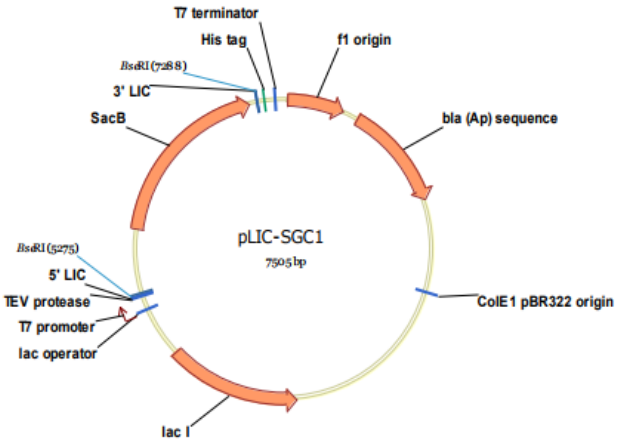


Figure 3: The pLIC-SGC1 vector used for cloning of SDR constructs. pET derived expression vector with His_6 tag in 23-aa N-terminal fusion peptide His_6 tag in 22-aa N-terminal fusion peptide, with a TEV protease cleavage site. Depicted in the figure gene for Lac I promoter, respectively (orange arrows), the main sites important for cloning (black denotations), including LIC cloning, and a "stuffer" fragment that incorporates the SacB gene for negative selection on 5 % sucrose, and the sites of restriction enzymes (maroon denotations). https://www.thesgc.org/sites/default/files/oxford_vectors/pLIC-SGC1.pdf

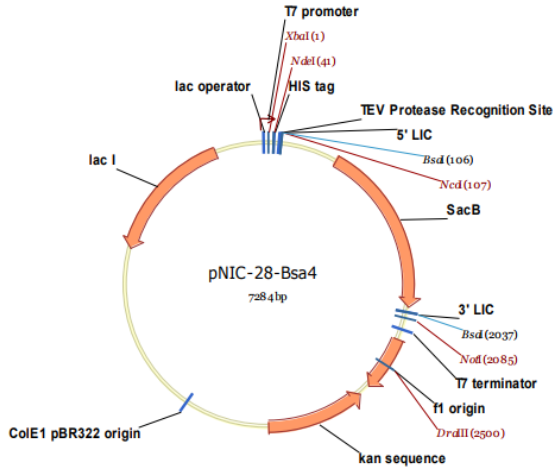


Figure 4: The pNIC-28-Bsa4 vector used for cloning of SDR constructs. This is pET derived expression vector with C-terminal His6 tag and FLAG tag, preceded by a TEV protease cleavage site. Depicted in the figure gene for Lac I promoter and Kanamycin resistance gene, respectively (orange arrows), the main sites important for cloning (black denotations), including LIC cloning, and a "stuffer" fragment that incorporates the SacB gene for negative selection on 5 % sucrose, and the sites of restriction enzymes (maroon denotations) (Savitsky et al 2010) https://www.thesgc.org/sites/default/files/oxford_vectors/pNIC28-Bsa4t.pdf

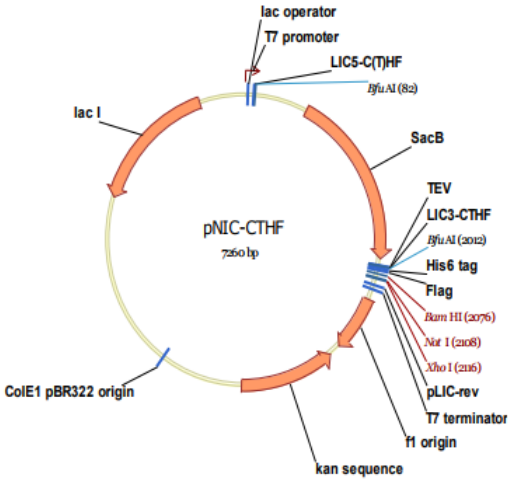


Figure 5: The pNIC-CTHF vector used for cloning of SDR constructs. This is pET derived expression vector with N-termina His6 tag and FLAG tag, preceded by a TEV protease cleavage site. Depicted in the figure are the genes for Lac I promoter and kanamycin resistance gene, respectively (orange arrows), the main sites important for cloning (black denotations), and the sites of restriction enzymes (maroon denotations). https://www.thesgc.org/sites/default/files/oxford_vectors/pNIC-CTHF.pdf

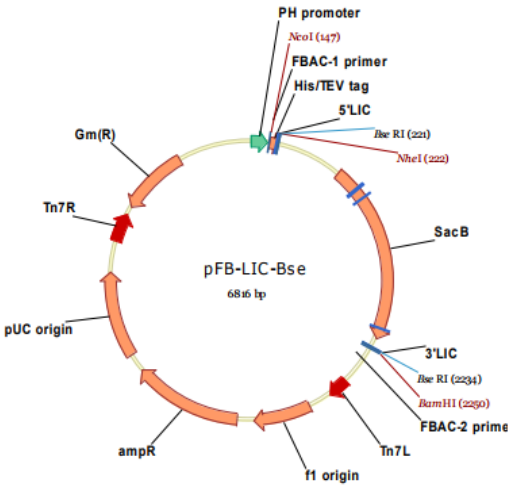


Figure 6: The pFB-Lic-Bse vector used for cloning of SDR constructs. This is a Baculovirus transfer vector with His_6 tag in 22-aa N-terminal fusion peptide, with TEV protease cleavage site. Include sites for LIC cloning and a "stuffer" fragment that inclides the SacC gene, allowing negative selection of transformed bacteria on 5 % sucrose. https://www.thesgc.org/sites/default/files/oxford_vectors/pFB-LIC-Bse.pdf