DOCTOR OF PHILOSOPHY

The Development of Photo-crosslinkable Trapping Mutants as Tools to Investigate the Interactions of Protein Tyrosine Phosphatases

Pavic, Karolina

Award date:
2013

Awarding institution:
University of Dundee
European Molecular Biology Laboratory

DOCTOR OF PHILOSOPHY

The Development of Photo-crosslinkable Trapping Mutants as Tools to Investigate the Interactions of Protein Tyrosine Phosphatases

Karolina Pavic

2013

University of Dundee
European Molecular Biology Laboratory, Heidelberg

The Development of Photo-crosslinkable Trapping Mutants as Tools to Investigate the Interactions of Protein Tyrosine Phosphatases

Karolina Pavic

Ph.D

December, 2013

European Molecular Biology Laboratory, Heidelberg, Germany

and The University of Dundee

Declaration

I hereby declare that the following thesis is based on the results of investigations conducted by myself and that this thesis is of my own composition. Work other that my own is clearly indicated in the text by reference to the original publications. This dissertation has not in whole, or in part, been previously submitted for a higher degree.

Karolina Pavic

I certify that Karolina Pavic has spent the equivalent of eight terms in research work at the Genome Biology Unit, EMBL Heidelberg, and that she has fulfilled the conditions of the Ordinance General No. 14 of the University of Dundee and is qualified to submit the accompanying thesis in application for the degree of Doctor of Philosophy.

Dr. Maja Köhn

Acknowledgements

I would like to thank Dr. Maja Köhn for giving me an opportunity to carry out this research project in her laboratory. I am immensely grateful to her for providing constant support and patience and for sharing her scientific expertise throughout my PhD.

I am thankful to the members of my Thesis Advisory Committee (TAC) Prof. Stephen Keyse, Dr. Jeroen Krijgsveld and Dr. Edward Lemke, for their sound advice. I am especially grateful to Prof. Stephen Keyse for attending all annual TAC meetings at EMBL in Heidelberg and for all help in the preparation for the doctoral examination procedure at the University of Dundee.

I would like to thank all members of the Köhn group for truly pleasant working environment, understanding and support. Especially I am grateful to Dr. Pablo Rios, for sharing his great experience in the phosphatase field, unreserved input on the work presented here and for experimental assistance. I want to thank my desk neighbour Birgit Höger for keeping very nice atmosphere in our corner, and in our coffee breaks, and for the comments on the manuscript of the work presented in this thesis. I am thankful to Dr. Irina Cornaciu for helpful discussions, sharing her expertise in protein expression and purification and being my “agony aunt”.

I am grateful to Dr Edward Lemke and Christine Köhler for help with the initial experiments with the unnatural amino acids, and for generously sharing plasmids and advice in DNA cloning, and to Kristina Dzeyk from the Proteomics Core Facility for the sound patience and persistence in the analyses.

Last, I would like to thank my family for their love and support.

Contents

Declaration.....1

Acknowledgements.....2

Contents.....3

Abbreviations.....12

Abstract.....24

Chapter 1: Introduction.....26

1.1 Cell signaling and protein phosphorylation.....26

1.2 PPs.....27

Figure 1: PPs and PKs maintain homeostasis of protein phosphorylation.

1.3 PTPs.....29

Figure 2: Catalytic mechanism of PTPs.

Figure 3: Principle of regulation of PTPs through oxidation.

1.3.1 The Class I of PTPs.....35

1.3.1.1 DUSPs.....36

Figure 4: Classification of PTPs.

1.3.1.2 MKPs.....38

1.3.1.3 Atypical DUSPs.....40

1.4 VHR.....41

1.4.1 General properties of VHR.....41

1.4.2 Structural characteristics of VHR.....43

Figure 5: Comparing structure of VHR with classical PTPs PTP1B and *Yersinia* PTP.

1.4.3 Physiological substrates of VHR.....46

Figure 6: Summary of known Physiological substrates of VHR.

1.4.4 Regulation of VHR.....50

1.4.5 VHR in cancer.....51

1.5 The PRLs.....54
1.5.1 Prenylation of the PRLs.....55
1.5.2 Expression of the PRLs.....56
1.5.3 Subcellular localisation of the PRLs.....58
1.5.4 Structural characteristics of the PRLs.....60
Figure 7: Structure-based multiple sequence alignment of all three PRLs and the structurally most closely-related PTPs.
Figure 8: Solution structure of PRL-3 (PDB: 1R6H).
1.5.5 The PRLs in cancer.....69
1.5.6 Regulation of the PRLs by oxidation.....74
1.5.7 Other regulatory mechanisms of the PRLs: PRL-1 and PRL-2.....76
1.5.8 Interacting proteins and substrates of PRL-1 and PRL-2.....78
1.5.9 The PRLs in signaling: PRL-1 and PRL-2.....80
1.5.10 Regulation of PRL-3.....84
1.5.11 Interacting proteins and substrates of PRL-3.....88
1.5.12 PRL-3 in signaling.....95
Figure 9: Summary of the signaling pathways affectd by the PRLs and the outcome in cell migration and proliferation.
1.6 Challenges in identifying PTP substrates and interacting partners.....102
1.7 Expanded genetic code methodology.....103
Figure 10: A general principle of expanded genetic code methodology.
Figure 11: Difference in the expression of prokaryotic and eukaryotic tRNAs.
1.7.1 *p*BPA.....112
Figure 12: Principle of photo-induced cross-linking with *p*BPA.
1.8 Aims.....113

Chapter 2: Materials and Methods.....115

2.1 Materials and methods for Chapters 3.1-3.3.....115
2.1.1 Reagents.....115
2.1.2 Materials.....116
2.1.3 Instruments.....118
2.1.4 Miscellaneous.....119
2.1.5 Buffers, solutions and media.....120

2.1.5.1 TAE....120
2.1.5.2 SDS-PAGE running buffer....120
2.1.5.3 SDS-PAGE sample buffer....120
2.1.5.4 SDS (10%) stock solution)....121
2.1.5.5 Stacking buffer (4 x))....121
2.1.5.6 Resolving buffer (4 x))....121
2.1.5.7 Ammonium persulfate (10%))....121
2.1.5.8 Tris-HCl (1M, pH 7.5)....121
2.1.5.9 Phosphate buffer saline (PBS) 122
2.1.5.10 EDTA (0.5 M, pH 8.0) 122
2.1.5.11 Tris-buffered saline (TBS) (10 x) 122
2.1.5.12 TBS-Tween20 (TBS-T) 122
2.1.5.13 Staining solutions for protein gels....122
2.1.5.14 Staining nitrocellulose membranes....123
2.1.5.15 Protein transfer....123
2.1.5.16 Stripping buffer for immunoblots....123
2.1.5.17 Striping buffer for His Trap HP columns....123
2.1.5.18 Recharging buffer for His Trap HP columns....123
2.1.5.19 Eluting buffer for GSTrap FF columns....124
2.1.5.10 SOC medium....124
2.1.5.21 Media for cultivation of bacterial cultures....124
2.1.5.22 Lysis buffers for mammalian cell lines....125
2.1.6 DNA amplification by PCR....125
2.1.7 DNA ligation....126
2.1.8 Procedure for DNA isolation....126
2.1.9 Procedure for DNA propagation....126
2.1.10 Procedure for transformation of chemically competent strains of *Escherichia coli* (*E. coli*) 127
2.1.11 Transformation procedure for electrocompetent *E. coli* strains....127
2.1.12 Antibodies (ab) and Western blotting....128
2.1.13 Maintenance of mammalian cell lines....130
2.1.14 Removal of His_6-tag by GST-TEV 130
2.1.15 Intact protein sample analysis by LC-MS....131
2.1.16 Circular dichroism (CD) measurements....132

2.2 Materials and methods for Chapter 3.1.....133
2.2.1 Engineering of the PRL-3: D72A, I141Y, N142Y, N142A, Q145Y.....133
Figure 13: Plasmid map for the pET15b vector.
2.2.2 Large-scale bacterial expression of the PRL-3 variants from 2.2.1.....135
2.2.3 Large-scale purification of the PRL-3 variants from 2.2.2.....135
2.2.4 Characterisation of the catalytic activity for the PRL-3 variants I141Y, N142Y, N142A and Q145Y.....136
Figure 14: Principle of EnzCheck phosphatase assay kit.
2.2.5 Generating the PRL-3 variants for small-scale optimisation of conditions for expression with an unnatural amino acid: wtPRL-3, D72TAG, I141TAG, N142TAG.....138
Figure 15: Plasmid map for the pTXB3 vector.
2.2.6 Optimising expression and purification conditions with wtPRL-3 on a small scale.....140
2.2.7 Small-scale test-expression with *p*ACF.....141
Figure 16: Chemical structure of *p*ACF.
2.2.8 Small-scale test-expression with *p*BPA.....142
2.2.9 Generating pETM20/6His-3Flag-D72TAG, I141TAG and N142TAG.....142
Figure 17: Plasmid map for the pCMV-3Flag-1A vector.
Figure 18: Plasmid map for the pETM20 vector.
Figure 19: Summary of the primers used for generating PRL-3 variants expressed in pET15b, pTXB3-intein-6His and pETM20 vectors.
2.2.10 Generating the VHR variants for small-scale optimisation of conditions for expression with an unnatural amino acid: wtVHR, L25TAG, F68TAG, D92TAG, N163TAG, C124S F68TAG, C124S G161TAG.....146
2.2.11 Generating the VHR variants for large-scale bacterial over-expression: wtVHR, C124S, F68TAG and D92TAG.....149
Figure 20: Summary of the primers used for generating VHR variants expressed in pET15b, pTXB3-intein-6His and pETM20 vectors.
2.2.12 Large-scale bacterial expression of recombinant VHRs from 2.2.11.....151
2.2.13 Large-scale purification of recombinant VHRs from 2.2.12.....151

2.3 Materials and methods for Chapter 3.2.....152

2.3.1 Generating the VHR variants for large-scale bacterial over-expression: F68A, F68D, F68W, F68C, M69A and D92A.....152
2.3.2 Large-scale bacterial expression of recombinant VHRs from 2.3.1.....153
2.3.3 Large-scale purification of recombinant VHRs from 2.3.2.....153
2.3.4 Phosphatase activity measurements against *p*NPP.....153
2.3.5 Photo-cross-linking of recombinant VHRs with 345 nm light.....155
2.3.6 Cross-linking with glutaraldehyde.....156
2.3.7 Cross-linking with DSS.....157
Figure 21: Chemical structure of DSS.
2.3.8 Investigating dimerisation potential of the VHR variants with respect to oxidising conditions.....158
2.3.9 Size-exclusion chromatography.....158
2.3.10 LC-MS/MS analysis of the F68*p*BPA variant after exposure to 345 nm light.....159
Figure 22: Analysis of UV-induced intramolecular cross-linking for the F68*p*BPA variant of VHR.
2.3.11 Plasmids used for expression in mammalian COS-1 cells.....163
2.3.12 Transient transfection of COS-1 cells.....164
2.3.13 Immunoprecipitation after H_2O_2 stimulated COS-1 cells.....164
2.3.14 Cross-linking with PFA in intact COS-1 cells.....165
2.3.15 Dephosphorylation of Erk1/2 in COS-1 lysate by recombinant VHR.....165

2.4 Materials and methods for Chapter 3.3.....166
2.4.1 Generating pETM20/6His-3Flag-PRL-1.....166
2.4.2 Large-scale bacterial expression and purification of PRL-1.....167
2.4.3 Generating pETM20/6His-3Flag-wtPRL-3 and D72A.....167
2.4.4 Large-scale bacterial expression and purification of the $Flag_3$-tagged PRL-3 variants: wtPRL-3, D72A, D72*p*BPA.....168
2.4.5 Assessing catalytic activity of $Flag_3$-D72*p*BPA with OMFP.....168
2.4.6 PIP Strips assay.....168
2.4.7 Incubation of the PRL-3 variants (wt, D72A and D72*p*BPA) with HEK293 lysate.....169

Chapter 3: Results and Discussion.....172

3.1 Optimising conditions for expression of the PRL-3 and VHR variants with *p*BPA.....172

3.1.1 Selecting PRL-3 residues to be exchanged for *p*BPA 172

Figure 23: Selecting residues in PRL-3 to be exchanged for *p*BPA.

3.1.2 Selecting VHR residues to be exchanged for *p*BPA.....179

Figure 24: Selecting VHR residues to be exchanged for *p*BPA.

3.1.3 Generating the I141Y, N142Y, N142A and Q145Y mutants of PRL-3.....182

Table 1: Verifying the integrity of the PRL-3 variants (I141Y, N142Y, N142A and Q145Y) by determining full protein mass.

3.1.4 Characterisation of PRL-3: I141Y, N142Y, N142A and Q145Y.....185

Figure 25: Phosphatase activity of the PRL-3 variants against OMFP.

Figure 26: Phosphatase activity of the PRL-3 variants against $PI(3,4,5)P_3$.

3.1.5 Small-scale optimisation of conditions for bacterial over-expression with wtPRL-3.....190

Figure 27: Screening for the optimal *E. coli* strain and induction temperature by using pTXB3-intein-6His/wtPRL-3.

3.1.6 Small-scale optimisation of conditions for bacterial over-expression of the D72-, I141- and N142TAG variants of PRL-3 with *p*ACF.....194

Figure 28: Small-scale expression of the D72-, I141- and N142TAG variants of PRL-3, in pTXB3-intein-6His plasmid, with 1 mM *p*ACF.

3.1.7 Small-scale optimisation of conditions for bacterial over-expression of the D72-, I141- and N142TAG variants of PRL-3 with *p*BPA.....197

Figure 29: Small-scale test-expression of the PRL-3 variants, in pTXB3-intein-6His plasmid, with *p*BPA.

Figure 30: Small-scale expression of the PRL-3 D72*p*BPA mutant, under optimised conditions.

3.1.8 Small-scale expression of the VHR variants with *p*ACF and *p*BPA.....201

Figure 31: Small-scale expression of the VHR variants with *p*ACF at 2 mM final concentration in pTXB3-intein-6His plasmid.

Figure 32: Small-scale expression of the VHR variants with *p*BPA at the final concentration of 2.5 mM in pTXB3-intein-6His plasmid.

3.1.9 Creating the final *p*BPA-containing PRL-3 construct for biological exploration.....204

Figure 33: Comparing plasmids pTXB3-His_6 and pETM20 for the efficiency of expression of the *p*BPA variants of PRL-3 in *E. coli*.

3.1.10 Optimising concentration of *p*ACF for the I141- and N142TAG variants of PRL-3.....207

Figure 34: Optimising concentration of *p*ACF for expression with the I141- and N142TAG variants of PRL-3.

3.2: Applicability of the photo-cross-linkable VHR variants in investigating dimerisation of VHR as a potential novel regulatory principle.....210

3.2.1 Large-scale bacterial over-expression of wtVHR and all its variants generated.....210

Figure 35: Coomassie stained SDS-PAGE gel of wtVHR and all its variants generated.

Table 2: Analysis of the intact protein samples by LC-MS.

Table 3: Protein yields for the generated VHR variants.

3.2.2 Investigating cross-linking abilities of the VHR variants containing *p*BPA *in vitro*.....215

Figure 36: Reactive radius of benzophenones.

Figure 37: Examining the ability of the F68pBPA variant of VHR to form UV-induced dimers after 30 min exposure to 345 nm.

Figure 38: Assessing the effect of the active site Cys124 on the ability of VHR to form UV-induced dimers.

Figure 39: Examining the ability of the D92*p*BPA mutant of VHR to form photo-induced dimers.

3.2.3 Catalytic activity of the VHR variants against *p*NPP.....224

Table 4: Characterisation of the catalytic activity of the VHR variants.

Table 5: Assessing the effect of the general phosphatase inhibitor PS on the catalytic activity of the wtVHR.

3.2.4 Structural characterisation of the VHR variants by far-UV CD measurements.....230

Figure 41: Analysis of the secondary structure for the wtVHR and its variants by obtaining far-UV CD measurements.

Figure 42: Analysis of the secondary structure for M69A and for a series of the F68-mutants of VHR by obtaining far-UV CD measurements.

Figure 43: Comparing the secondary structure of the F68*p*BPA variant of VHR to the native protein by obtaining far-UV CD measurements.

Figure 44: Met69 of the variable insert segment forms hydrogen bond with Arg130 of the P-loop in VHR.

3.2.5 Investigating the self-association of the VHR variants with alternative cross-linking methods.....236

Figure 45: Investigating the potential of VHR to self-associate by chemical cross-linking.

Figure 46: Investigating if the potential of VHR to self-associate depends on the catalytic cysteine by chemical cross-linking.

3.2.6 Profiling the structural effect of position 68 on the self-association of VHR by chemical cross-linking with glutaraldehyde.....242

Figure 47: Profiling the effect of the structural and electrostatic properties of the residues incorporated at position 68, on the ability of VHR to self-associate, by chemical cross-linking with glutaraldehyde.

Figure 48: The control samples for Figure 47.

3.2.7 Investigating dimerisation potential of the VHR variants with respect to oxidising conditions.....246

Figure 49: Analysing dimerisation potential of the purified recombinant VHR variants with respect to oxidising conditions.

3.2.8 Investigating the dimerisation potential of VHR in cells.....250

Figure 50: Self-association of VHR in H_2O_2-stimulated COS-1 cells.

Figure 51: Assessing specificity of dimerisation to VHR, in cells.

3.2.9 Investigating the dimerisation potential of VHR in cells independently of the oxidative stress.....253

Figure 52: VHR can form dimers in cells in the absence of H_2O_2-induced oxidative stress.

3.2.10 Investigating functional implications of VHR dimerisation.....256

Figure 53: Assessing functional implications of VHR dimerisation against *p*NPP.

Figure 54: Assessing the effect of VHR dimerisation on the ability to dephosphorylate Erk1/2.

3.2.11 Analysis of VHR dimers after separation by size-exclusion chromatography.....261

Figure 55: Separation of F68*p*BPA dimer by size-exclusion chromatography.

Figure 56: Assessing catalytic activity of the F68*p*BPA monomer isolated by size-exclusion chromatography.

Figure 57: Relative catalytic activity of the irradiated F68*p*BPA with respect to the irradiated wtVHR.

3.2.12 Analysis of VHR dimers after separation by size-exclusion chromatography by LC-MS/MS.....266

Figure 58: Identification of the intramolecular cross-linking by LC-MS/MS.

Figure 59: Investigating the cross-linking ability of the monomeric fraction of F68*p*BPA isolated by size-exclusion chromatography.

Figure 60: Analysis of the intact sample of F68*p*BPA dimer by LC-MS.

3.3 Applicability of the photo-cross-linkable PRL-3 mutant in investigating protein-substrate and protein-protein interactions.....271

3.3.1 Assessing the cross-linking ability of the PRL-3 D72*p*BPA variant.....271

Figure 61: Assessing the activity of D72*p*BPA variant of PRL-3 with OMFP.

Figure 62: PRL-3 in lipid strips assay.

Figure 63: Assessing the effect of the triple mutation R157A R159A K161A on the lipid binding profile of PRL-3.

Figure 64: PRL-1 in lipid strips assay.

3.3.2 Assessing the applicability of PRL-3 D72*p*BPA with HEK293 cell lysate

Figure 65: Incubation of the PRL-3 variants with HEK293 lysate, with or without exposure to 345 nm light.....277

Chapter 4: Conclusions and outlook.....281

4.1 Critical comments of the preliminary experiments (Chapter 3.1 in the Results and discussion section).....281

4.2 Implications for VHR biology: findings enabled by application of the amber suppression methodology.....282

4.3 Applicability of the PRL-3 mutants containing photo-cross-linkable amino acid in investigating protein-substrate interactions.....288

4.4 General applicability of the amber supression methodology to develop (more) efficient substrate trapping mutants of PTPs.....290

Abbreviations

2D two dimensional

A absorbance

A adenine

Å Ångström

aaRS aminoacyl transfer ribonucleic acid synthetase

ab antibody

ADP adenosine diphosphate

AML acute myeloid leukemia

AMP adenosine monophosphate

AngII Angiotensin II

AOM azoxymethane

AP1 activating protein 1

APC antigen-presenting cell

Arf1 adenosine diphosphate ribosylation factor 1

ATF activating transcription factor

ATP adenosine triphosphate

B. stearothermophilus *Baccilus stearothermophilus*

bCdc42 brain specific splice variant of Cdc42

Bcl2 B-cell lymphoma 2

BMK big mitogen-activated protein kinase

BRCA1 breast cancer 1

BSA bovine serum albumine

bZIP basic leucine zipper

C carboxy-terminal

C cytosine

Ca^{2+} calcium

cAMP cyclic adenosine monophosphate

CAT chloramphenicol acetyltransferase

CBM carbohydrate-binding module

CD circular dichroism

CD45 cluster of differentiation 45

Cdc14 cell division cycle homologue 14

CDH22 cadherin 22

CDK cyclin-dependent kinase

C-ERMAD carboxy ezrin/radixin/moesin associated domain

C-H carbon-hydrogen bond

CH2 Cdc25 homology domain

CH_2 methenyl

CHO Chinese hamster ovary cells

CKI cyclin-dependent kinase inhibitor

COP1 caspase recruitment domain-containing protein 16

COS-1/COS-7 African green monkey kidney cells

CRE cyclic adenosine monophosphate responsive elements

CREB cyclic adenosine monophosphate response element binding protein

Csk carboxy-terminal Sarcoma kinase

c-Src cellular Sarcoma

Da dalton

DMEM Dulbecco’s modified Eagle’s medium

DNA deoxyribonucleic acid

DSS disuccinimidyl suberate

DTT dithiothereitol

DUPD1 dual specific phosphatase and pro isomerase domain containing 1

DUSP dual-specificity phosphatase

E. coli *Escherichia coli*

ECM extracellular matrix

EDTA ethylenediaminetetraacetic acid

EF-2 elongation factor 2

EGF epidermal growth factor

EGFP enhanced green fluorescent protein

EGFR epidermal growth factor receptor

Egr early growth response

eIF2α eukaryotic initiation factor 2α

EMT epithelial to mesenchymal transition

ER endoplasmic reticulum

ErbB2 erythroblastic leukemia viral oncogene homolog 2

Erk1/2 extracellular signal-regulated kinase 1/2

Erk5 extracellular signal-regulated kinase 5

ERM ezrin/radixin/moesin

ESI electrospray ionization

EST expressed sequence tags

F-actin filamentous actin

FAK focal adhesion kinase

FBS fetal bovine serum

FD fast digest

FKBP38 FK506 binding protein 38

FPR formyl peptide receptor

FT farnesyl transferase

G guanine

g gram

GAP guanosine triphosphate hydrolase activating protein

GATPT (glucosamine-aminoethoxy)triphenyltin

GDP guanosine disphosphate

GEF guanine nucleotide exchange factor

GFP green fluorescent protein

GFR growth factor receptor

GGT geranylgeranyl-transferase

GM growth medium

G-protein guanine nucleotide-binding protein

Grb2 growth factor receptor-binding protein 2

GSH glutathione

GSK3β glycogen synthase kinase β

GST Glutathione *S*-Transferase

GTP guanosine triphosphatae

GTPase guanosine triphosphate hydrolase

h hour

H_2O_2 hydrogen peroxide

HAD haloacid dehalogenase

HDAC4 histone deacetylase 4

HEK293 human embryonic kidney 293 cells

HeLa Henrietta Lacks cells

HEPES N-2-Hydroxyethylpiperazine-N'-2-ethane sulfonic acid

HePTP hematopoietic protein tyrosine phosphatase

HF high fidelity

HME human mammary epithelial

HNTG buffer containing HEPES, NaCl, Triton X and glycerol

HRP horseradish peroxidase

HS heat shock

HUVEC human umbilical vein endothelial cells

IFN interferon

IP immunoprecipitation

IPTG isopropyl-β-D-thiogalactoside

IR infrared

IR insulin receptor

JNK c-jun N terminal kinase

K_2HPO_4 dipotassium hydrogen phosphate

KAP kinase associated phosphatase

KCl potassium chloride

kDa kilodaltons

KGFR keratinocyte growth factor receptor

KH_2PO_4 potassium dihydrogen phosphate

KIM kinase interaction motif

KRT8 keratin 8

l liter

LB Luria-Bertani

LC liquid chromatography

LD Lafora disease

LMW low molecular weight

m meter

m mili

M molar

M. barkeri *Methanosarcina barkeri*

M. jannaschii *Methanococcus jannaschii*

MAPK mitogen-activated protein kinase

MAPKAPK mitogen-activated protein kinase activated protein kinase

MAPKK/MAPKK mitogen-activated protein kinase kinase

MAPKKK/MAPKKK mitogen-activated protein kinase kinase kinase

MARCKS myristoylated alanine-rich C kinase substrate

MDCK Madin-Darby canine kidney

mDia1 *Drosophila melanogaster* diaphanous 1

MDM2 mouse double minute 2 homologue

MDSP muscle-restricted dual-specificity phosphatase

MEF mouse embryonic fibroblast

MEF2C myocyte enhancer factor 2C

MEK mitogen-activated extracellular signal-regulated kinase

MEKK mitogen-activated extracellular signal-regulated kinase kinase

METC mitochondrial electron transport chain

$MgCl_2$ magnesium chloride

min minute

MKB mitogen-activated protein kinase binding

MKP mitogen-activated protein kinase phosphatase

MMP matrix metalloproteinase

mRNA messenger ribonucleic acid

MS mass spectrometry

N amino-terminal

n nano

Na_2HPO_4 disodium hydrogen phosphate

Na_3VO_4 sodium orthovanadate

NaCl sodium chloride

NEB New England Biolabs

N-ERMAD amino ezrin/radixin/moesin associated domain

NGF nerve growth factor

Ni nickel

NiSO4 nickel (II)- sulfate

NLS nuclear localisation signal

NMD nonsence-mediated decay

NMR nuclear magnetic resonance

NRPTP non-receptor-like protein tyrosine phosphatase

NSCLC non-small cell lung cancer

OMFP 3-*O*-methylfluorescein phosphate

OD optical density

p130Cas p130 Crk-associated substrate

*p*ACF *para*-Acetylphenylalanine

PAGE polyacrylamide gel electrophoresis

*p*BPA/*p*BzF *para*-Benzoyl-(L)-Phenylalanine

PBS phosphate buffer saline

PCBP1 poly-C-ribonucleic acid binding protein 1

PDB protein data bank

PDGFR platelet-derived growth factor receptor

pErk1/2 phosphorylated extracellular signal-regulated kinase1/2

PFA paraformaldehyde

PI(3,4,5)P2 phosphatidylinositol (3,4,5)-trisphosphate

PI(4,5)P2 phosphatidylinositol (4,5)-bisphosphate

PI3K phosphatidylinositol 3-kinase

PIC protease inhibitor coctail

PIEC phosphatase of regenerating liver-1 intron enhancer complex

PIP phosphoinositide

PIRH2 RCHY1 ring finger and CHY zinc finger domain containing 1

PK protein kinase

PL phospholipase

P-loop phosphate-binding loop

*p*NPP *para*-nitrophenylphosphate

p pico

PP protein phosphatase

PRL phosphatase of regenerating liver

PS PhosSTOP

pSer/pThr phosphorylated serine/threonine

PT prenyl-transferase

PTEN phosphatase and tensin homologue deleted on chromosome ten

PTK protein tyrosine kinase

PTM posttranslational modification

PTP protein tyrosine phosphatase

pTyr phosphorylated tyrosine

QCL QuickChangeLightning

RBD RhoA binding domain

REP Rab escort protein

RNA ribonucleic acid

ROCK Rho-coiled coil kinase

ROS reactive oxygen specie

rpm round per minute

RPTP receptor-like protein tyrosine phosphatase

Rsk ribosomal S6 kinase

rt room temperature

RTK receptor tyrosine kinase

s second

S. cerevisiae *Saccharomyces cerevisiae*

SAGE serial analysis of gene expression

Sap1 stomach-cancer associated protein tyrosine phosphatase

SAPK stress-activated protein kinase

SAPK stress-activated protein kinase

SDS sodium dodecylsulfate

SEM standard error of the mean

SFK Sarcoma family kinase

SFM serum-free medium

SH2 Src homology 2

SH3 Src homology 3

shRNA short hairpin ribonucleic acid

siRNA small interfering ribonucleic acid

Smad mothers against decapentaplegic homologue

SO_2H sulfinic acid

SO_3H sulfonic acid

SOH sulfenic acid

Sos son of sevenless

SPPS solid phase peptide synthesis

SRE serum response element

SRF serum response factor

STAT signal transducers and activators of transcription

STEP Striatal Enriched Phosphatase

TAE Tris-acetate-EDTA

TAG amber STOP codon

TB terrific broth

TBS tris buffer saline

TBST tris buffer saline with Tween20

TCEP Tris(2-carboxyethyl)phosphine hydrochloride

TCTP Translationally controlled tumor protein

TEV tobacco etch virus

TGFβ transforming growth factor beta

Tm melting temperature

TMDP testis and skeletal muscle specific dual-specificity phosphatase

tRNA transfer ribonucleic acid

TRP32 thioredoxin-related protein 32

Tyk-2 tyrosine kinase-2

U uracil

UAA ochre STOP codon

UPLC ultra performance liquid chromatography

UTR untranslated

UV ultraviolet

VEGF vascular endothelial growth factor

VH1 *Vaccinia* H1

VHR *Vaccinia* H1 related

VRK3 *Vaccinia*-related kinase 3

v-Src viral Sarcoma

wt wild type

Y2H yeast-two hybrid

α alpha

β beta

γ gamma

δ/Δ delta

ε epsilon

μ micro

Abstract

Abnormalities in the coordinated activities of protein phosphatases (PPs) and protein kinases (PKs) contribute to the development of many diseases. Phosphatase of regenerating liver (PRL)-3 and *Vaccinia* H1-Related (VHR) are two members of the protein tyrosine phosphatase (PTP) family shown to be involved in cancer. PRL-3 is a member of the PRL phosphatases containing a unique post-translationally modifiable prenylation CAAX motif at the carboxy (C)- terminal end. There is an immense body of evidence to support a role for PRL-3 in the development of various types of cancer and in progression to metastasic disease. However, many questions are still pending, especially with respect to the identity of physiological substrates, and interacting partners in general, of PRL-3. VHR is a model for a group of atypical dual specificity protein phosphatases (DUSPs) with a role in cell cycle progression. Only a few of VHR's physiological substrates have been reported to date and there are very few studies addressing its regulation and physiological role(s).

Generally, in order to isolate and identify transient phosphatase-substrate interactions, substrate-trapping mutants of PTPs are employed. Mutants which can function as substrate traps have the ability to recognise and bind substrates, yet they lack functionality of the key catalytic residues and cannot efficiently process the hydrolysis of the substrate. However, it is acknowledged that the efficiency of such standard substrate-trapping mutants of PTPs is low.

In this work, the expanded genetic code approach was applied to develop more efficient substrate trapping variants of the PTPs by incorporating the photo-cross-linkable amino acid *para*-benzoylphenylalanine (*p*BPA). The concept was

optimised for PRL-3 and VHR, and for both proteins, *p*BPA-containing variants were expressed at excellent yields and were highly purified.

By utilizing the photo-cross-linkable F68*p*BPA variant of VHR, dimerisation of VHR was detected in an *in vitro* ultraviolet (UV) exposure-mediated cross-linking assay. VHR dimerisation was further demonstrated to be a potential novel regulatory mechanism for VHR, having a negative effect on the catalytic activity of the protein. A specific region in VHR known as the variable insert segment was pinpointed as a region in the protein, which is either at the dimer interface or heavily contributing to dimeric association. Furthermore, the intrinsic ability of VHR to self-associate was also demonstrated by complementary methods.

For PRL-3 it was demonstrated that its D72*p*BPA variant could recognise and bind to lipids, with a stronger signal detected in the UV-exposed sample, and without altering the lipid binding profile with respect to the native protein. Lastly, the potential of exploiting photo-cross-linkable variants of the PTPs was also demonstrated by incubating PRL-3 variants, with or without selectively introduced *p*BPA, with mammalian cell lysates, followed by UV exposure. Western blot analysis detected new bands corresponding to covalently cross-linked PRL-3-protein complexes. Future work in our laboratory will follow up on these newly identified interactions.

1: Introduction

1.1 Cell signaling and protein phosphorylation

The term cell signaling can be referred to as a cell's ability to sense signals in the environments, intra- or extracellular, and to respond to them in an appropriate and rigorously regulated way (DeVita, Jr., Lawrence, Rosenberg et al., 2011; Neel and Tonks, 1997; Tonks, 2005). The responses are diverse and can include modulation of protein expression, subcellular localisation, conformation, stability or protein-protein interactions. Many cellular responses are a result of protein modifications at the post-translational level. Protein phosphorylation is a prevalent post-translational modification (PTM) with a pivotal role in the regulation of various biological processes including cell cycle progression, cell growth, differentiation, metabolism, motility, gene transcription, immune response or apoptosis (Burke and Zhang, 1998; Zhang, 2003a; Zhang, 2005). It is believed that protein phosphorylation accounts for modification of around 30% of proteins encoded in the human genome (Cohen, 2002). Mass spectrometry (MS) based study of temporal dynamics of phosphorylated proteins in Henrietta Lacks (HeLa) cells following epidermal growth factor (EGF) stimulation yielded several intriguing revelations (Olsen, Blagoev, Gnad et al., 2006). 2,244 phosphoproteins were identified, and over 90% of them had not been reported in SwissProt. Moreover, the authors noted that the abundance of phosphorylation on Tyr (pTyr) was around ten-fold higher than estimated in an earlier study. The difference was not just the result of a stimulus-triggered response, but also of the presence of pTyr modification on less abundant cellular proteins. Unraveling protein phosphorylation sites is expected to help in integrating cell signaling research efforts, with special

emphasis on cancer-linked aberrations (Julien et al., 2011; Motiwala and Jacob, 2006; Olsen, Blagoev, Gnad et al., 2006; van Huijsduijnen et al. 2002).

1.2 PPs

The homeostasis of protein tyrosine phosphorylation (Figure 1) in cellular milieu is achieved through complementary activities of PKs, which catalyse the transfer of a phosphoryl moiety donated by the adenosine triphosphate (ATP) to the hydroxyl group of a particular residue (tyrosine, serine or threonine) and PPs, with counteracting hydrolysis of the phosphate monoester finally resulting in the recuperated active enzyme and dissociation of the inorganic phosphate (Arantes, 2006; Elchebly et al., 1999; Jackson and Denu, 2001; Klaman et al., 2000; Kolmodin and Åqvist, 2001; Zhang, 1998). Perturbations in the balanced activities of PKs and PPs, yielding abnormalities in the phosphorylation level of protein tyrosine residues, have been linked to the aetiology of a plethora of human diseases, including cancer development and progression, diabetes or immune disorders (Barford, 1995; He et al., 2012; Hendrix, W.J.A.J. et al., 2013; Myers, M.P. et al., 1998; van Huijsduijnen et al., 2002; Julien et al., 2011; Tiganis and Bennett, 2007; Zhang, 2005).

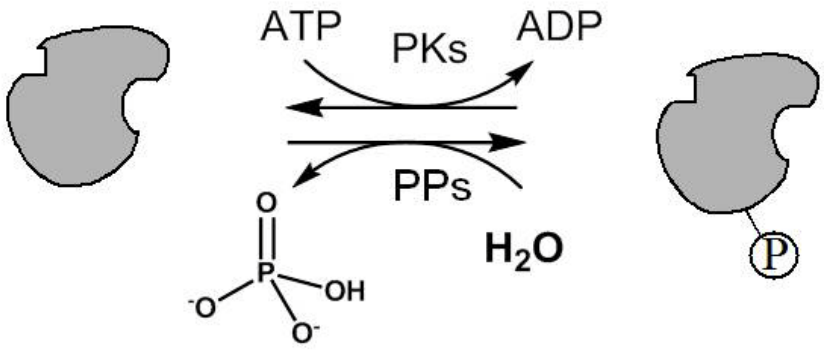


Figure 1: PPs and PKs maintain homeostasis of protein phosphorylation.

According to the substrate specificity, PPs can be divided into two large superfamilies, protein serine/threonine phosphatases (STPs) and protein tyrosine phosphatases (PTPs) (Denu et al., 1996a; Jackson and Denu, 2001; Zhang, 1998). In stark contrast to PKs, which developed from a common ancestor (Tonks, 2006), STPs and PTPs are characterised by diversities in sequence, fold or catalytic mechanism (Barford et al., 1998; Hoffman et al., 2004).

STPs function as metalloenzymes with a dinuclear bivalent metal ion center ($Fe^{2+/3+}$, Mn^{2+}, Zn^{2+}) at the active site, which is essential for both the enzymatic activity and maintenance of the structural integrity (Barford, 1995; Denu et al., 1996a; Hoffman et al., 2004). The metal ions are coordinated by an asparagine, two aspartates and three histidines, one of which acts as a general acid and helps to expel the leaving group of a substrate (Denu et al., 1996a). The catalytic mechanism of STPs consists of a single step. It involves a direct nucleophilic attack of a substrate by the activated water molecule. The water molecule is positioned optimally for catalysis by the two metal ions, which are also important for proper positioning of a substrate. It is interesting to note that fewer STPs (about 40) are opposed by as many as 428 STKs (Li et al., 2013). This apparent discrepancy in number can be assigned to the fact that STPs function *in vivo* as holoenzymatic complexes consisting of a catalytic subunit decorated by diverse regulatory or targeting subunits, thus generating combinatorial complexity.

1.3 PTPs

The research focus on tyrosine phosphorylation was stimulated by the finding that viral (v)-Src of Rous sarcoma virus could phosphorylate protein tyrosine residues (Cohen, 2002), thus implying involvement of tyrosine phosphorylation in growth regulation. PTP1B was the first PTP to be isolated, from human placenta, and characterised (Charbonneau et al., 1989; Cohen, 2002; Zhang, 1998). Comparing the sequences of the catalytic subunits of protein phosphatase (PP) 1, PP2A and PP2B/calcineurin, and to the available acid and alkaline phosphatases revealed PTP1B's membership to a novel class of phosphatases, further fueling complexity in the study of PTPs (Tonks, 2013). Many growth factor receptors (GFRs) were shown to have protein tyrosine kinase (PTK) activity (Cohen, 2002; Hunter, 2000; Zhang, 1998). Nevertheless, PTPs were long misconceptionally considered promiscuous housekeeping enzymes that merely served to reverse the activities of PTKs (Tonks, 2006; Zhang, 1998). They are sporting remarkable structural complexity and sophisticated substrate specificity (Andersen et al., 2001; Barr et al., 2009; Flint et al., 1997; He et al., 2012; Tiganis and Bennett, 2007; Tonks, 2006; Zhang, 2005). Historically, the discovery of PTKs dates almost a decade before PTPs (Cohen, 2002; Tiganis and Bennett, 2007). Signal transduction pathways were viewed as cascades igniting from the receptor PTKs which were associated with growth factors or peptide hormones (Hunter, 2000; Seger and Krebs, 1995; Zhang, 2005). This association led to receptor dimerisation and tyrosine transphosphorylation (Schlessinger, 1988), thus creating binding sites for proteins bearing Src homology-2 (SH2) domains (Hunter, 2000; Yaffe and Cantley, 1999). This, in turn, triggered activation of downstream targets, eventually eliciting a particular effect. Since the discovery of PTKs by Tony

Hunter and colleagues (Cohen, 2002; Eckhart et al., 1979; Julien et al., 2011), their preponderance in the regulation of fundamental cellular processes has been highlighted by the findings that aberrations in the PTK-linked signaling pathways provide a causal link to various diseases (Hunter, 2000; Lemmon and Schlessinger, 2010; Zhang, 1998). Consequentially, this urged the research focus on the enzymes with the counteracting activity, namely the PTPs. It is now well established that PTPs can act as positive or negative switches for signal transduction (Zhang, 2003b; Zhang, 2005).

The hallmark which defines PTPs is the presence of the consensus motif $HisCys(X)_5Arg(Ser/Thr)$ (where X represents any amino acid) in the catalytic domain (Denu et al., 1996a; Zhang, 1998; Zhang, 2005). This conserved amino acid signature is also termed the phosphate-binding loop, P-loop or the signature motif (Burke and Zhang, 1998; Jackson and Denu, 2001). The catalytic domain of the classical PTPs spans around 280 amino acid residues (Barr et al., 2009), and the sequence alignment identified ten conserved motifs which are essential for conserved structural fold, substrate recognition and catalytic mechanism (Andersen et al., 2001; Tiganis and Bennett, 2007).

The presence of the conserved signature motif was utilised in an attempt to provide systematic classification of PTPs (Alonso et al., 2004). Identification of 107 PTP-encoding genes in the human genome cemented the view of PTPs as significant contributors to the balance of the cellular pTyr level. A recent study (Li et al., 2013) reported 96 PTP-encoding genes in the human genome, focusing only on the catalytically active PTPs. The number of PTPs, in contrast to 90 genes encoding PTKs, is suggestive of their comparable complexity at structural and functional levels (Alonso et al., 2004; He et al., 2012; Li et al., 2013; Tonks, 2006). Whereas previous subdivision of PTPs into four classes

was solely based on the amino acid composition of the catalytic domains (Alonso et al., 2004), the recent study by Li et al. also included structural considerations (Li et al., 2013). The Classes I-III are cysteine-based PTPs (Alonso et al., 2004), whereas the representatives of the Class IV were re-classified into haloacid dehalogenase (HAD) superclass (Li et al., 2013).

The Class I of cysteine-based PTPs is the most abundant and will be discussed in the later section.

The basic catalytic mechanism of the PTPs encompasses two distinct enzymatic steps (shown in Figure 2) (Barford, 1995; Jackson and Denu, 2001; Kolmodin and Åqvist, 2001; Zhang, 2003a and b). It proceeds via catalytic cysteine and arginine residues, which belong to the consensus active site motif, and of an aspartate which resides on the flexible general acid/base loop (so-called WPD loop) separated from the signature motif in the primary sequence. The catalytic cysteine exists in the form of a thiolate ion under physiological conditions due to its unique microenvironment. In the first step of the hydrolytic reaction, it performs a nucleophilic attack on a phosphate ester of a substrate bound in the active site, leading to the formation of a cysteinylphosphate intermediate. The reaction is helped by the aspartate residue which acts as an acid and donates a proton to the oxygen of the leaving group. In the second step, the aspartate residue functions as a base and facilitates hydrolysis of the intermediate by extracting a proton from a water molecule, leading to a release of inorganic phosphate and recuperation of the active site cysteine. The conserved arginine residue of the PTP signature motif is important for binding of a substrate and stabilization of the phospho-enzyme intermediate by forming bidentate hydrogen bonds with two oxygens of the bound substrate. A conserved serine or threonine residue immediately following the conserved

arginine residue of the P-loop is important for facilitating the breakdown of a cysteinylphosphate intermediate in the second step of the catalytic reaction (Zhang, 2003a).

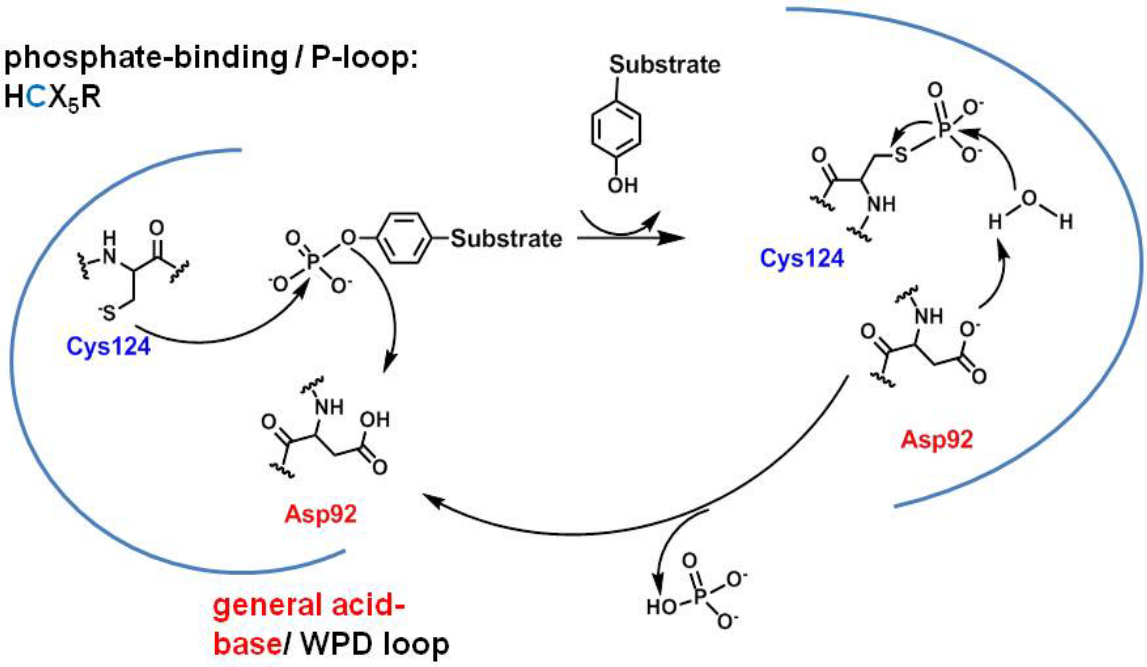


Figure 2: Catalytic mechanism of PTPs. Adapted from Tonks, 2006. The key catalytic resiudes are labeled based on VHR.

It has been well acknowledged that PTPs are susceptible to inactivation through oxidation of the catalytic cysteine in response to diverse stimuli (den Hertog et al., 2008; Meng et al., 2002; Tonks, 2006; Traore et al., 2008 and shown in Figure 3). PTPs are targets for reactive oxygen species (ROS) because they possess the catalytic cysteine residue that, under physiological conditions, exists in the form of a thiolate ion, which serves two functions- it enables it to function as a nucleophile in the first step of catalytic reaction and it renders it sensitive to oxidation (den Hertog, 2003; Meng et al., 2002; Miki and

Funato, 2012; Tonks, 2005). PTPs have different predispositions towards oxidative inactivation of their catalytic domains, which was demonstrated on RPTPα, whose membrane distal D2 domain proved to be more prone to oxidation-induced catalytic inactivation than D1 domain (den Hertog et al., 2008; Groen et al., 2005; Östman et al., 2011). It was also suggested that ROS could be generated in a spatio-restricted manner thus providing a mechanism for selective inhibition of the PTPs that are co-localised with the sites where ROS were generated (Böhmer et al., 2013).

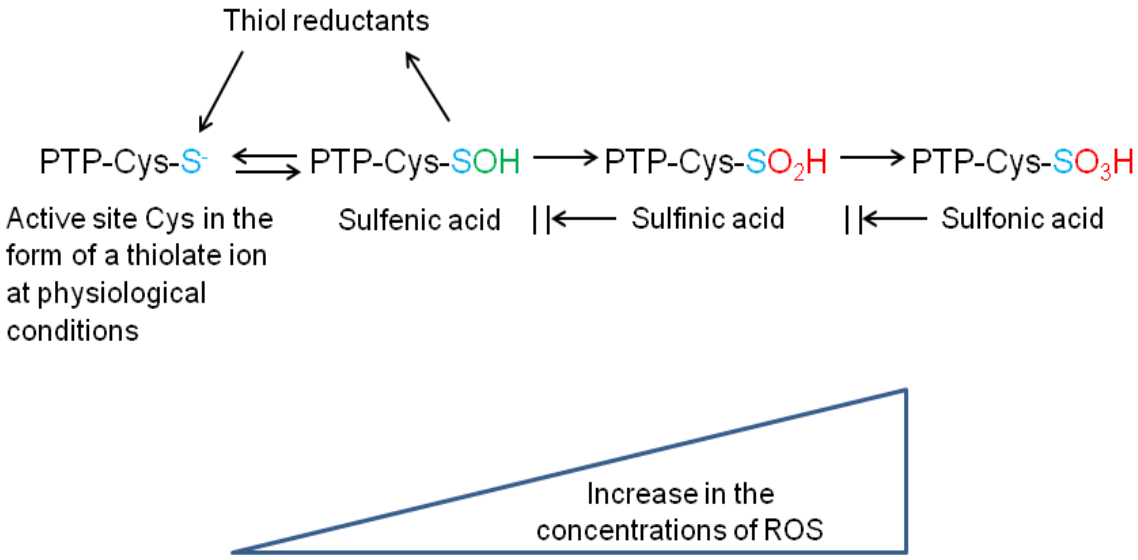


Figure 3: Principle of regulation of PTPs through oxidation. Active site Cys is present in the form of a thiolate ion at physiological conditions. In the presence of ROS, such as H_2O_2, it can be reversibly oxidised to sulfenic acid, thus rendering PTPs catalytically inactive. Sulfenic acid can be reduced back to the thiolate form through actions of thiol reductants, restoring the active form of the enzyme. Oxidations of cysteine residue to sulfinic and sulfonic acids are irreversible.

For oxidation to represent a general regulatory mechanism of PTPs, the oxidation of the catalytic cysteine must be reversible, which means that it must not proceed further than sulfenic acid (SOH-) (Finkel, 2012; Roos and Messens, 2011; Tonks, 2005; Tonks, 2006). Oxidations to sulfinic (SO_2H-) and sulfonic acid (SO_3H-) are considered irreversible, but there are several mechanisms for PTPs to prevent excessive oxidation. These include formation of cyclic sulphenamide, which induces conformational change preventing association of the phosphatase with a substrate and also exposing the oxidised residue to the reducing environment of the cell. Another mechanism is intramolecular disulfide bridge formation between the active site cysteine and proximal regulatory ("backdoor") cysteine (Buhrman, 2005; Funato and Miki, 2013; Östman et al., 2011; Poole and Nelson, 2008; Tonks, 2006). ROS for example, superoxide, hydroxyl radical and hydrogen peroxide (H_2O_2)) are generated by partial reduction of oxygen and are produced by molecules of oxygen catching electrons escaped from mitochondrial electron transport chain (METC), and also as a result of various environmental stresses, such as irradiative damage, heat-shock, infections or toxins (Freuhauf and Meyskens, Jr., 2007; Groen et al., 2005; Roos and Messens, 2011; Traore et al., 2008). Furthermore, ROS are not only regarded as cellular damaging agents, but are known to participate in intracellular signaling as second messengers generated at low levels after growth factor or cytokine stimulation (Finkel, 2012; Freuhauf and Meyskens, Jr., 2007; Gough and Cotter, 2011; Kamata et al., 2005; Meng et al., 2006; Miki and Funato, 2012; Rhee, 2006; Rhee et al., 2000; Roos and Messens, 2011; Sundaresan et al., 1995). ROS generated transiently in response to various stimuli impose their effect on downstream signaling by transiently inhibiting PTPs (Gough and Cotter, 2011; Karish and Neel, 2013;

Rhee, 2006; Traore et al., 2008). Moreover, H_2O_2 added exogenously was shown to induce signaling cascades resembling growth factor-induced activation (Denu and Tanner, 1998). In addition, H_2O_2 is the most commonly utilized ROS in investigating oxidation-susceptibility of PTPs, employed also in a study by Denu and Tanner focused on demonstrating reversible oxidation as a general regulatory principle of PTPs (Denu and Tanner, 1998; Östman et al., 2011).

1.3.1 The Class I of PTPs

The members of the Class I of PTPs (99) can be further subdivided into transmembrane receptor-like PTPs (RPTPs) (21), non-receptor-like PTPs (NRPTPs) (17) and *Vaccinia* H1 (VH1)-like phosphatases, which are also referred to as DUSPs (Alonso et al., 2004 and shown in Figure 4). Whereas NRPTPs consist of cytosolic proteins with a single catalytic domain and achieve diversity through domains flanking either side of the catalytic domain, RPTPs are transmembrane proteins. They contain various extracellular domains and transmembrane region, and 12 of them have a tandem of the cytoplasmic PTP domains (Majeti and Weiss, 2001; Tonks, 2006). Extracellular domains of the RPTPs display different motifs, which suggest their participation in intercellular communication or cell-matrix contacts (Barford, 1995; Böhmer et al., 2013). The PTP domains of RPTPs are referred to as D1 (membrane-proximal) and D2 (membrane-distal) (Tonks, 2006). Catalytic activity is almost exclusively confined to the D1 domain. The D2 domain is important for stabilization of the protein and for mediating protein interactions.

1.3.1.1 DUSPs

As mentioned previously, 61 DUSPs have been reported (Alonso et al., 2004). Excluding the conserved residues of the PTP-signature motif, they exhibit significant sequence divergence compared to the classical PTPs. Further subdivision into seven groups can be performed based on structure and homology between phosphatase domains. These groups include mitogen activated protein kinase (MAPK) phosphatases (MKPs) (11), atypical DUSPs (19), PRL phosphatases (3), slingshots (3), phosphatase and tensin homologues (PTENs) (5), myotubularins (16) and cell division cycle 14 homologues (Cdc14s) (4). As this thesis was focused on PRL-3 and VHR, the other DUSP families will not be discussed.

DUSPs constitute a heterogeneous group of the PTP superfamily (Tonks, 2013). Unlike classical PTPs, the more shallow and broad catalytic pocket of DUSPs yields more various substrate specificity (Jackson and Denu, 2001), rendering them capable to not only catalyze dephosphorylation of phosphotyrosine (pTyr), but also phosphoserine/threonine (pSer/pThr) residues, as well as nonpeptidic substrates, such as phosphoinositides (PIPs), messenger ribonucleic acid (mRNA) or glycans (Campbell et al., 2003; Gentry, Romá-Mateo et al., 2012; Leslie, Gray et al., 2000; Leslie et al., 2001; Li et al., 2013; Lokareddy et al., 2013; McParland, Varsano et al., 2011; Patterson et al., 2009; Zhang, 2003b).

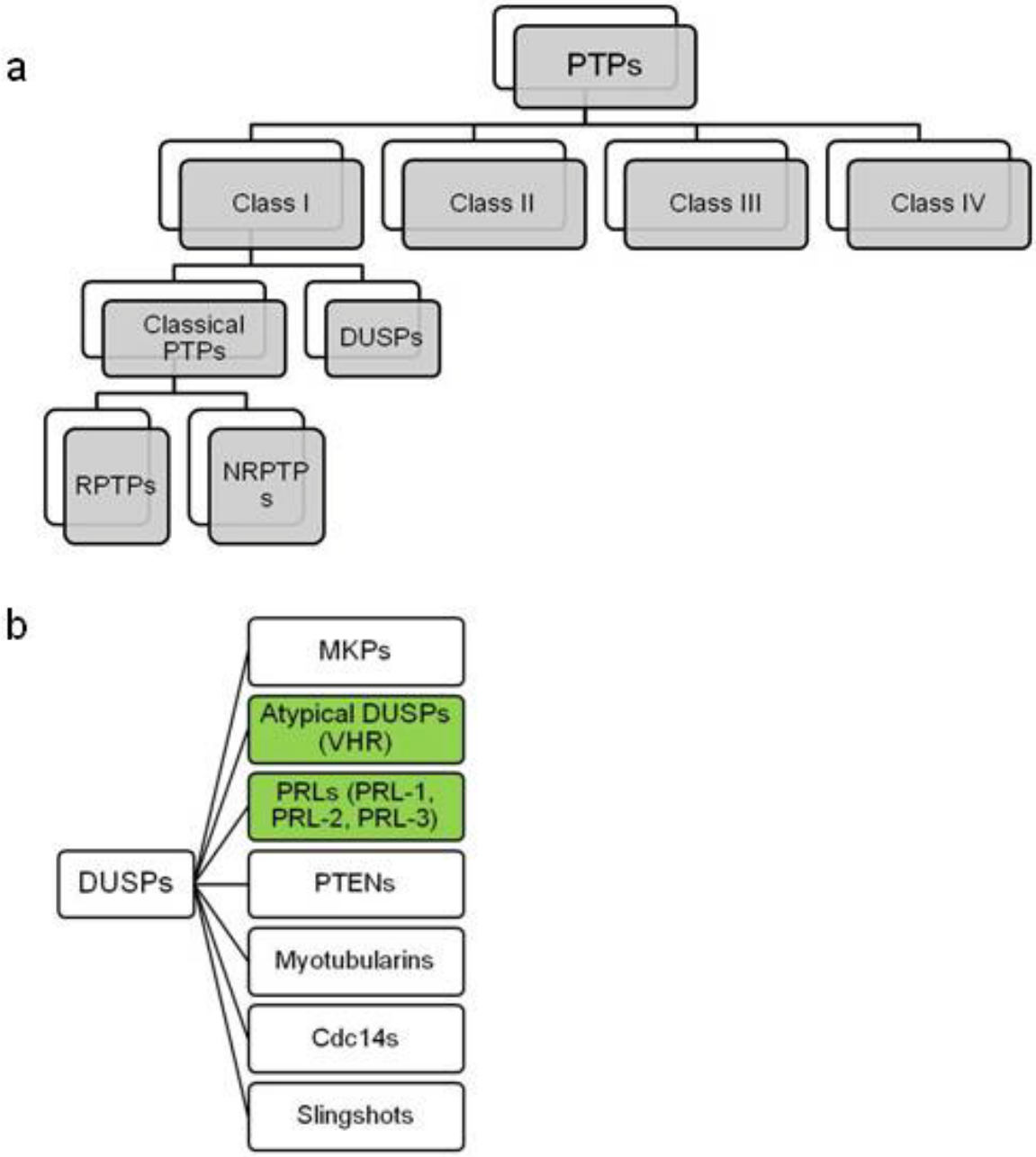


Figure 4: Classification of PTPs. (a) PTPs are divided into four classes, and classes I-III are cysteine-based. The Class I is further subdivided into classical PTPs, comprised of RPTPs and NRPTPs, and DUSPs. (b) DUSPs are further divided into seven subgroups based on structure and similarity of their phosphatase domains (adapted from Alonso et al., 2004).

1.3.1.2 MKPs

MKPs share little sequence similarity with the classical PTPs, but they have similar topology and employ the same catalytic mechanism (Patterson et al., 2009). MKPs represent a subgroup of DUSPs which can dephosphorylate MAPKs. The MAPKs represent a final constituent in conserved tripartite kinase signaling cascades, and are activated by the upstream dual-specificity kinases termed MAPK kinases (MAPKK, MAPKK or MEK), which, in turn, are activated by their upstream kinases (MAPKKK, MAPKKK or MEKK) (Caunt and Keyse, 2013). MAPKKs activate MAPKs by dual phosphorylation of both Tyr and Thr residues of the -ThrXTyr-motif in the kinase activation loop of the MAPKs (Katz et al., 2007; Theodosiou and Ashworth, 2002; Todd et al., 1999; Zhou et al., 2002). MAPKKK are activated by various upstream activators (Katz et al., 2007; Lahiny et al., 2010). The ultimate targets of the activated MAPKs cover a diverse portfolio of substrates involved in cellular processes as diverse as regulation of gene expression, proliferation, differentiation, immune response, cell survival versus programmed cell death or cell cycle progression (Patterson et al., 2009; Zhou et al., 2002). In mammals, there are four major families of MAPKs (Chang and Karin, 2001). These comprise extracellular signal-regulated kinase (Erk) 1/2 (Erk1 is also known as p44 MAPK or as MAPK1 and Erk2 as p42 MAPK or as MAPK2), p38 MAPKs (p38α, β, γ and δ), c-jun N-terminal kinases (JNKs; JNK 1-3) (also known as stress-activated protein kinase, SAPK) and Erk5, also known as Big MAP kinase (BMK) 1 (Caunt and Keyse, 2013; Theodosiou and Ashworth, 2002; Zhang and Liu, 2002). Distinct MAPK pathways are activated in response to a particular stimulus: Erk1/2 are commonly activated by growth or differentiation factors, Erk5 by stress cues as well as by growth factors, and the members of the JNK and p38 families by

inflammatory cytokines and various environmental stress conditions, such as oxidative stress, osmotic or heat shock or irradiation (Haagenson and Wu, 2010; Katz et al., 2007; Pumiglia and Decker, 1997; Raman et al., 2007; Robinson and Cobb, 1997; Todd et al., 1999; Wagner and Nebreda, 2009). MKPs can demonstrate activity against various MAPKs *in vitro*. However, the specificity of an MKP toward a specific MAPK is more distinct *in vivo* (Zhang et al., 2011; Zhou et al., 2002). MKPs can be further subdivided into three groups, based on their amino acid sequence similarity, subcellular localisation and substrate specificities (Dickinson and Keyse, 2006). These three groups contain ten catalytically active phosphatases, whereas DUSP24, also known as MK-STYX, resembles features of the MKPs, but it has naturally occurring mutation of the catalytic cysteine into glycine, thus rendering it a pseudophosphatase (Tonks, 2013; Wishart, 1995). The three subfamilies of the active MKPs include: nuclear MKPs, induced in response to mitogens and stress signals (MKP-1/DUSP1, MKP-2/DUSP4, DUSP2/PAC-1 and DUSP5), cytoplasmic MKPs, with activity toward Erk1/2 (MKP-3/DUSP6, MKP-4/DUSP9 and MKP-X/DUSP7), and MKPs with selectivity for JNK and p38 isoforms of MAPKs which can be localised in both nucleus or cytoplasm (DUSP8, MKP-5/DUSP10 and DUSP16) (Caunt and Keyse, 2013; Dickinson and Keyse, 2006). MKPs quench the activity of MAPKs by dephosphorylating the residues -ThrXTyr- in the activation loop, either or both (Caunt and Keyse, 2013). There is a number of phosphatases reported to downregulate MAPKs by dephosphorylation (Keyse, 2000; Saxena and Mustelin, 2000). They exhibit distinct subcellular localisation, substrate preference and regulation. Most of them are nuclear and silent until a proper signal is imposed. By controlling the magnitude and duration of the kinase activation, downregulation of a particular

MAPK through dephosphorylation fine-tunes cellular response to a particular stimulus (Dickinson and Keyse, 2006; Keyse, 2008).

All MKPs share common architectural features. They have a catalytic domain at the C-terminal end, with partial sequence similarity to the one of VH1 (Dickinson and Keyse, 2006; Theodosiou and Ashworth, 2002; Tonks, 2013). On the N-terminal end, they carry a domain with two regions which bear sequence similarity to the catalytic domain of Cdc25. Interestingly, the Cdc25 homology domain (CH2) is also pointing towards evolutionary link with the rhodanese-like enzymes (Alonso et al., 2004; Dickinson and Keyse, 2006). CH2 domain is not present in the atypical DUSPs, which are discussed in the next chapter. Further structural characteristics of the MKPs will not be discussed as MKPs were not the studied in this thesis.

1.3.1.3 Atypical DUSPs

Atypical DUSPs are a subfamily of DUSPs (Alonso et al., 2004). Although they have substrates amongst MAPKs, they show substrate specificity also outside of the family of the MAPKs (Bayón and Alonso, 2010). Compared to the classical MKPs, they lack the N-terminal MAPK-binding (MKB) domain and, almost exclusively, any other recognisable motifs and are smaller in size (Bayón and Alonso, 2010; Patterson et al., 2009; Wei, Ryu, et al., 2011).

In general, the data on the atypical DUSPs is very complex and conflicting (Bayón and Alonso, 2010; Patterson et al., 2009). Substrate specificity *in vivo* cannot be merely extrapolated from the *in vitro* studies, and quite often the observed function is dependent on the cell type or on a particular cell stimulus. In addition, some atypical DUSPs can act as scaffolds, or can require a yet unidentified component in the signal transduction pathways to elicit a particular

response. To shed light on the mechanisms behind the involvement of atypical DUSPs in various cellular processes, in normal or pathological states, more efficient substrate, and interacting partners, trapping mutants would need to be developed to complement approaches such as knockout mice models or MS based techniques, to enable identification of less abundant interacting partners.

1.4 VHR

1.4.1 General properties of VHR

VHR/DUSP3 is an atypical DUSP comprised of 185 amino acids organized in a single catalytic domain (Bayón and Alonso, 2010; Ishibashi et al., 1992; Yuvaniyama et al., 1996). It was identified by Ishibashi et al. by using expression cloning strategy (Ishibashi et al., 1992). The newly discovered enzyme bore resemblance to VH1, hence being named VH1-Related phosphatase. Apart from the residues of the PTP signature motif, it shares no discernible sequence homology to the classical PTPs (Schumacher et al., 2002). VHR demonstrated the ability to dephosphorylate artificial phosphatase substrate *p*NPP, but also tyrosine-phosphorylated insulin-, platelet-derived growth factor-, epidermal growth factor- and keratinocyte growth factor receptors (IR, PDGFR, EGFR and KGFR, respectively), RTKs, as well as serine-phosphorylated casein, classifying as a phosphatase with dual specificity (Denu et al., 1995a; Ishibashi et al., 1992). When nucleophilic Cys124 of the P-loop was exchanged to Ser, VHR completely lost its catalytic ability, thus indicating that a single active site governed its activity towards both classes of phosphorylated substrates (Ishibashi et al., 1992; Schumacher et al., 2002; Yuvaniyama et al., 1996; Zhou et al., 1994).

Being one of the smallest DUSPs, VHR emerged into a prototypic phosphatase used to extensively profile catalytic properties of DUSPs (Zhou et al., 1994). By using radioactively labeled pTyr-containing substrate, it was demonstrated that hydrolysis of a phosphate monoester proceeds via mechanism employed by the classical PTPs (Denu et al., 1995b; Zhou et al., 1994). In addition, the activity of VHR against pSer/pThr-containing peptides was around 300 times lower. VHR demonstrated up to three-fold preference for bisphosphorylated peptides over monophosphorylated ones (Luechapanichkul, Chen et al., 2013).

Moreover, the subcellular localisation of endogenous VHR was investigated in a panel of mammalian cell lines, and also in a set of human breast cancer cells (Todd et al., 1999). VHR was found to be mostly localised in the nucleus, with occasional cytoplasmic staining, and present at high levels. The protein level of VHR was found to fluctuate during cell cycle progression, being barely detectable in G1, peaking through S phase and being highest in G2/M phase (Rahmouni et al., 2006). The subcellular localisation of VHR was also found to change with the cell cycle progression. In interphase, VHR was mostly found in the nucleus, while in telophase it was concentrated between the daughter chromatids. When levels of VHR were reduced by small interfering (si) RNA, the cells arrested in G1/S and G2/M phases. These cells also had increased levels of $p21^{Cip/Waf1}$ and downregulated genes implicated in cell cycle regulation, deoxyribonucleic acid (DNA) replication, repair or transcription. Reduced level of VHR in G1 phase was linked to the growth-promoting role of MAPKs. It was previously recognised that in the presence of growth arresting stimuli, such as DNA damaging agents or phorbol esters, prolonged activation of Erk1/2 led to cell cycle arrest, whereas activation of JNK was reported to activate p53 and, subsequently, $p21^{Cip-Waf1}$ (Rahmouni et al., 2006; Tang et al.,

2002). Thus, cell cycle dependent levels of VHR would enable activation of MAPKs for only as long as necessary for normal cell cycle progression.

1.4.2 Structural characteristics of VHR

The report of the crystal structure of VHR listed several structural specificities of this phosphatase (Yuvaniyama et al., 1996). The overall protein fold was shown to be similar to the fold of the classical PTPs (Figure 5). Compared to the classical PTPs, the active site pocket of VHR was shallower (no more than 6Å in depth). The depth of the active site pocket was the result of differences in the architecture of the N-terminal α1- β1 loop, which was shown to be shorter in VHR than in the classical PTPs. This region was termed the “recognition region” due to its recognised role for substrate recognition. Next, the region defined by the β3-β7 sheets was called the “variable insert” segment. The corresponding region in the PTPs was shown to be almost four times longer. The amino acid composition of this region is diverse. The amino acid residues localised at the edges of the β3-β7 region were known to contribute to the integrity of the active site, but apart from Met69 in VHR, the functional significance of the other residues in this region has not been profiled. The carbonyl group of Met69 is forming a hydrogen bond with the conserved Arg130 of the P-loop.

The preference of VHR to dephosphorylate pTyr was explained by reporting the crystal structure of its catalytically inactive mutant C124S in complex with a peptide derived from the activation loop of p38 MAPK (Schumacher et al., 2002). The structure showed that a narrow entrance to the active site crevice in VHR was created by the bulky residues at the edges of the entrance, Glu126 and Tyr128. Most of the dual-specific MKPs have smaller hydrophobic residues

at these positions, Ala and Ile, respectively (Tonks, 2013). Interestingly, when those point-mutations were simultaneously generated in VHR and the activity of the double mutant was assessed against a peptide derived from the activation loop of JNK, but with both phosphorylated residues in the peptide being pThr, double mutant was nine-fold better than the native enzyme (Schumacher et al., 2002). Related to this, a recently reported crystal structure of DUSP27 revealed an analogous rationale for its preference for hydrolyzing pTyr over pThr/pSer (Lountos et al., 2011). In DUSP27, the sites of Glu126 and Tyr128 are occupied by Met149 and Arg151, respectively. In addition to DUSP27, DUSP13 was identified as a close structural homologue of VHR (Kim et al., 2007). The regions of the most prominent structural differences between VHR and DUSP27 are the N-terminal segment and the loop β3-α4 (Lountos et al., 2011). In VHR, the N-terminal helix is a part of the substrate-binding groove, while in DUSP27 it projects away from the active site, whereas the region corresponding to the loop β3-α4 is known as the “variable insert” segment (Yuvaniyama et al., 1996). The reported crystal structure of DUSP27 showed that the protein crystallized as a dimer through swapping of the N-terminal helices, similar to VH1 (Koksal *et al.*, 2009; Lountos *et al.*, 2011). Dimerisation mechanism is likely to be contributed by the extended conformation of the N-terminal helix. Similarly to VHR, DUSP27 prefers pTyr containing substrates (Friedberg *et al.*, 2007; Lountos *et al.*, 2011). It has narrow entrance to the active site pocket, dictated by the bulky Met149 and Arg151 (Lountos *et al.*, 2011). Intriguingly, the surface charge distribution in DUSP27 is reversed with respect to VHR (Friedberg *et al.*, 2007; Lountos *et al.*, 2011). All the aforementioned differences between VHR and DUSP27 are likely to result in different substrate portfolio, although the physiological substrates of DUSP27

are yet to be identified. Similarly, the most prominent structural differences between VHR and DUSP13 are in the α1-β1 loop, known as substrate recognition region in VHR, and in the β3-α4 loop, which is shorter in DUSP13 with two residues, corresponding to Phe68 and Met69 in VHR, missing (Kim et al., 2007; Yuvaniyama et al., 1996).

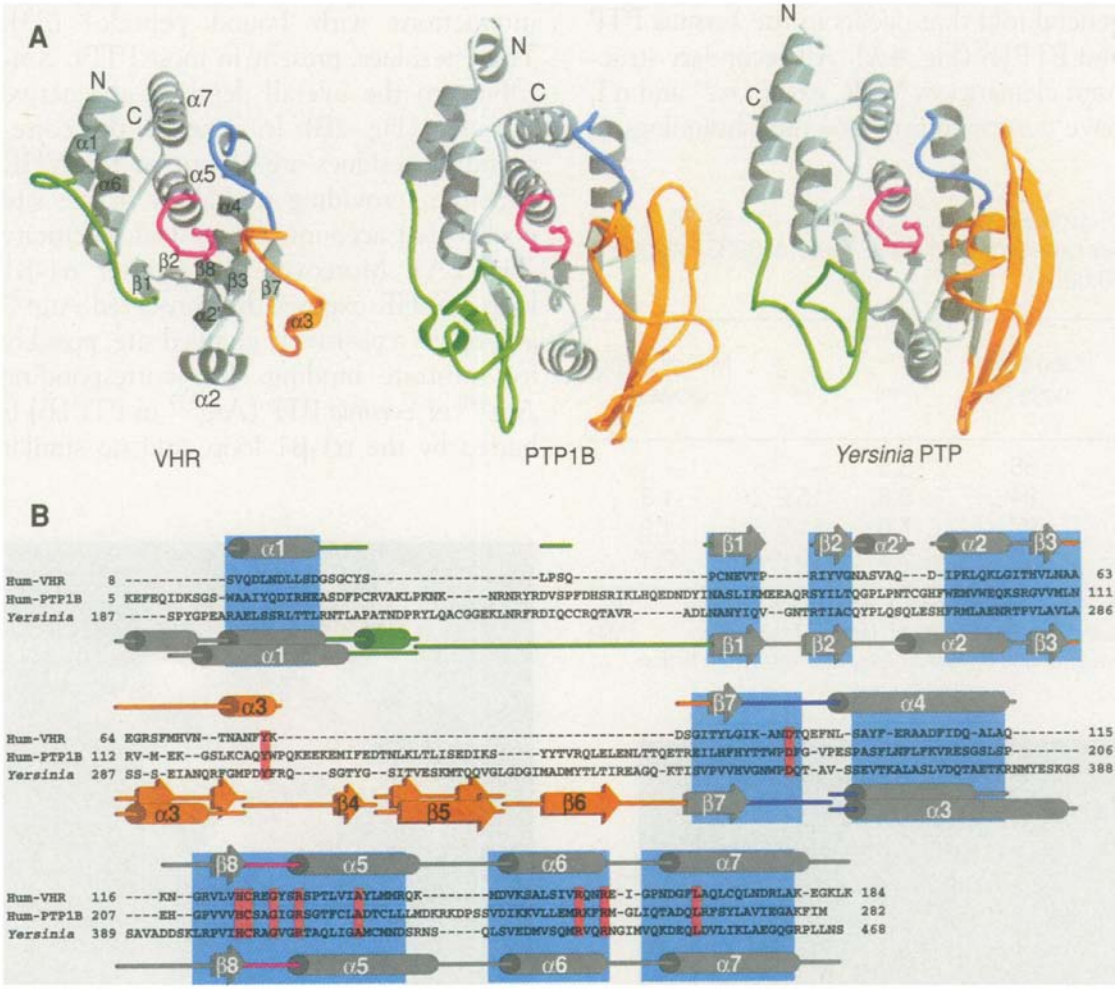


Figure 5: Comparing structure of VHR with classical PTPs PTP1B and *Yersinia* PTP. The overall fold of VHR resembles that of the classical PTPs, like indicated in panel A. “Recognition region” segment is pointed out in green and “variable insert” in orange. Panel B shows structure-based sequence alignment. Conserved residues are indicated in pink, helices are represented as cylinders, sheets as arrows and structurally equivalent regions in blue. Reprinted with permission from Yuvaniyama, J., Denu, J.M., Dixon, J.E. & Saper, M.A. Crystal structure of the dual specificity protein phosphatase VHR. *Science* **272**, 1328-1331 (1996).

1.4.3 Physiological substrates of VHR

Only few physiological substrates of VHR have been reported and these include MAPKs Erk1/2 and JNK, erythroblastic leukemia viral oncogene homolog 2 (ErbB2) and signal transducers and activators of transcription (STAT) 5 (Hoyt, Zhu, Cerignoli et al., 2007; Luechapanichkul, Chen et al., 2013; Todd et al., 1999; Todd et al., 2002; Wang, Yeh et al., 2011 and Figure 6).

Erk1/2 was identified as a physiological substrate of VHR by using its D92A mutant immobilized on an affinity column, which retained Erk1/2 from African green monkey kidney COS-1 cellular extracts enriched for tyrosine-phosphorylated proteins (Todd et al., 1999). Hence, D92A served as a substrate-trapping mutant of VHR. In general, substrate-trapping mutants are employed in order to isolate and identify transient phosphatase-substrate interactions, (Zhang, 2003a). Mutants which can function as substrate traps have the ability to recognise and bind substrates, yet they lack functionality of the key catalytic residues.

Purified recombinant VHR demonstrated the ability to dephosphorylate both recombinant Erk1/2 and Erk1/2 from the cellular extracts, with preference for Tyr185 residue from the activation motif of Erk1/2. Moreover, when the endogenous VHR was immunodepleted from COS-1 cell extracts, dephosphorylation of Erk1/2 was slowed down comparing to the mock-treated control, whereas immunodepletion had no effect on dephosphorylation of p38, confirming VHR as a regulator of Erk1/2. Under the conditions, no difference in the phosphorylation level of JNK was observed, whereas the level of phosphorylated p38 MAPK showed only small additional decrease in the

presence of the recombinant VHR. Unlike MKPs reported previously, the protein level of VHR was not fluctuating in response to different stimuli, such as EGF treatment or stimulation with different mitogens. Thus, VHR was suggested to maintain phosphorylated level of Erk1/2 at the basal level in the dormant state and by deactivating it after appropriate stimulation, by preferentially dephosphorylating Tyr185 residue of its activation loop.

It is interesting to note that activity of VHR against Erk1/2 was shown to be decreased in cultured human intestinal epithelial cells and in murine epithelia *in vivo* as a result of microbiota-mediated generation of ROS (Wentworth et al., 2011). The benefits of commensal bacteria on homeostasis of the intestine have been acknowledged. The aforementioned study helped elucidate a mechanism behind commensal bacteria-modulated manipulation of signaling in intestinal epithelial cells. According to it, mRNA and protein levels of VHR were upregulated upon stimulation with bacterial N-formyl-peptides. In addition, formyl peptide receptor (FPR), which recognises bacterial N-formyl-peptides, results in the transient generation of ROS. This enables a feedback loop between the Erk1/2 mediated proliferation stimulative effect and transient downregulation of the catalytic activity of Erk1/2-specific phosphatase VHR.

JNK was another representative of the MAPK family shown to be dephosphorylated by VHR (Todd et al., 2002). NIH3T3 mouse fibroblasts, transiently and stably transfected with wild type (wt) VHR or its inactive C124S mutant and subjected to various stress stimuli to activate JNK MAPK pathway, were shown to contain lower levels of phosphorylated JNK when wtVHR was overexpressed. Interestingly, it was demonstrated that association of c-Jun with JNK prevented VHR-mediated dephosphorylation of JNK, but had no negative effect on its activity against Erk1/2. This finding explained the aforementioned

observations when profiling the specificity of VHR against Erk1/2 and not seeing the effect on JNK (Todd et al., 1999). The work of Todd et al. offered an intriguing hypothesis to explain low levels of Erk1/2 found to associate with VHR, according to which VHR could be “masked” from activated Erk1/2 in a way comparable to the one of c-Jun and JNK (Todd et al., 2002).

STAT5 was identified as a VHR substrate outside of the MAPK family (Hoyt, Zhu, Cerignoli et al., 2007). STATs are a family of transcription factors that undergo phosphorylation-induced activation (Braunstein et al., 2003; Klemm et al., 1998; Olayioye et al., 1999). Activated STATs form homo- or heterodimers. Dimerisation is achieved by intermolecular association involving SH2 domain of one STAT monomer and pTyr residue in the C-terminal region of the second STAT monomer. Activated STATs translocate to the nucleus to activate transcription of the target genes regulating diverse cellular processes, such as immune response, apoptosis or growth suppression (Jardin and Sticht, 2012). The mechanism of STAT5 dephosphorylation by VHR was found to proceed in two steps (Hoyt, Zhu, Cerignoli et al., 2007). It requires phosphorylation of Tyr138 in VHR, which acts as a binding site for the SH2 domain in STAT5. The second step involves dephosphorylation of Tyr694 in the displaced C-terminal end of STAT5. This mechanism was also confirmed recently by Jardin and Sticht who applied molecular modeling and dynamics approaches to profile structural requirements for STATs dephosphorylation by VH1 and VHR (Jardin and Sticht, 2012). The overall conclusion was that the specificity for dephosphorylating a particular form of STATs is derived from the nature of the residues located at the interface of the STAT-dimers.

ErbB2 was identified as a novel substrate of VHR by using stably transfected non-small cell lung cancer (NSCLC) H1299 cells (Wang, Yeh et al., 2011). VHR

was shown to preferentially dephosphorylate Tyr992 residue of the receptor. VHR-mediated dephosphorylation of Tyr992 had a suppressive effect on the activation of PLCγ/PKC signaling pathway, whilst not affecting the activation of Src.

The first systematic profiling of combinatorial peptide libraries in order to illuminate substrate preferences of VHR *in vivo* was reported recently (Luechapanichkul, Chen et al., 2013). It was revealed that VHR could dephosphorylate two classes of pTyr containing peptides. Peptides from the Class I resembled the reported VHR substrates, while the members of the Class II constitute a potential pool of uninvestigated cellular substrates. Intriguingly, the latter ones exhibited distinct binding mode, with the consensus motif (V/A)P(I/L/M/V/F)$X_{1\text{-}6}$pY at the N-terminal end. Although screening PhosphoSite database did not identify any human protein matching this criteria, it is plausible that this motif can be generated through proteolytic action on the proteins which do contain the aforementioned consensus sequence internally.

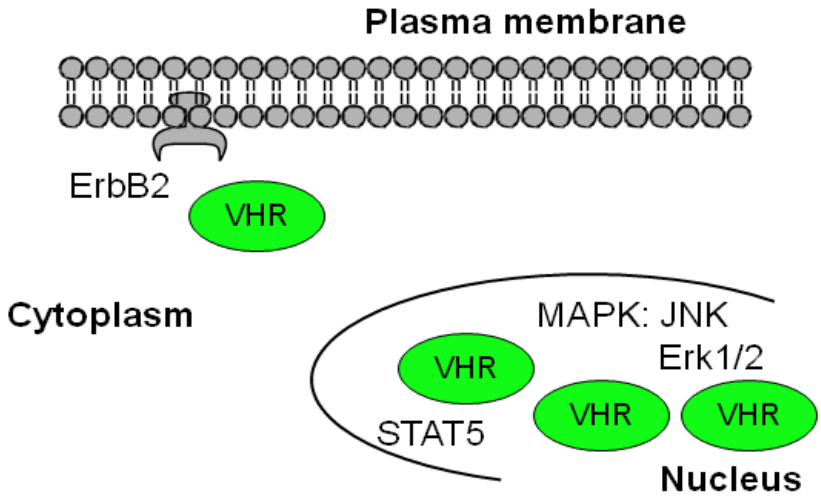


Figure 6: Summary of known physiological substrates of VHR. The figure was partially generated in ChemBioDraw.

1.4.4 Regulation of VHR

VHR expression is not triggered in response to various mitogenic stimuli (Todd et al., 1999). Related to this, it was reported that inactivation of Erk1/2 in EGF-stimulated cells does not cease after treatment with cycloheximide, which inhibits protein synthesis. This was noted to be in accordance with the earlier reports pointing towards a constitutively expressed phosphatase in the Erk1/2 pathway. According to them, downregulation of Erk1/2 occurred already 15 min after stimulation and independently of the de novo protein synthesis, whilst transcript for the inducible MKPs was present no sooner than 30 min after the stimulation. The aforementioned work of Todd et al. proposed a model according to which VHR could be kept inactive via transiently oxidised nucleophilic Cys124, due to the production of H_2O_2 as a part of “normal” receptor-mediated signaling events.

Reports on the regulatory principles of VHR are scarce. Two studies indicated that phosphorylation of Tyr138 is important for dephosphorylation of STAT5 in vivo, and for downregulation of Erk2 and JNK in T-cells (Alonso et al., 2003; Hoyt, Zhu, Cerignoli et al., 2007).

Phosphorylation of Tyr138, by ZAP-70 and tyrosine kinase (Tyk)-2 did not affect the catalytic activity of VHR *in vitro*, suggesting its likely regulatory role as a docking site for the SH2 domain of STAT5 (Hoyt, Zhu, Cerignoli et al., 2007), as already mentioned. Interestingly, Tyr138 phosphorylation was detected only in the catalytically inactive form of VHR, most likely indicating trans- and not auto-dephosphorylation of wtVHR, due to the localisation of Tyr138 being opposite to the active site pocket.

In T-cells, VHR was found to relocalize from diffused cytoplasmic distribution to the sites of the T-cell and antigen-presenting cell (APC) contact, after stimulation (Alonso et al., 2003). The time course of relocalisation coincided with the time after which downregulation of Erk1/2 and JNK was observed. For comparison, the effects of VH1 and VHR on T-cell activation was different; VH1 was a strong inhibitor (Koksal and Cingolani, 2011), converted to promote viral replication in the host cell by bypassing cellular antiviral response, and VHR acted like a negative regulator by decreasing the course of the MAPKs activation.

It was demonstrated that the activity of VHR against Erk1/2 can be enhanced by *Vaccinia*-related kinase (VRK) 3 (Kang and Kim, 2006). VRK3 imposed a negative effect on the activation of Erk1/2 by interacting with VHR, but independently of its kinase activity. The catalytic activity of VHR against *p*NPP and purified recombinant pErk2 was increased when it was overexpressed with VRK3. VRK3 was shown to be negative regulator of Erk1/2 signaling, while JNK and p38 MAPKs were not affected.

1.4.5 VHR in cancer

VHR has been linked to breast cancer (Hao and ElShamy, 2007). It was shown that expression of VHR was decreased upon overexpression of breast cancer 1 (BRCA1)-IRIS. BRCA1-IRIS is found overexpressed in different breast and ovarian cancer cell lines. It positively correlates with the expression of cyclin D1 which is implicated in cell growth regulation. Cyclin D1 expression is induced by binding of a transcription factor called activating protein (AP) 1, composed of c-Jun and Fos proteins, to the promoter region of cyclin D1. C-Jun is activated by JNK, which is reportedly downregulated by VHR (Todd et al., 2002).

Overexpression of VHR in human mammary epithelial (HME) cells reduced expression of cyclin D1, thus opening up possibilities for VHR inhibitors in the development of therapeutics targeting BRCA1-IRIS-mediated effects on cell proliferation (Hao and ElShamy, 2007). Next, it was demonstrated that mRNA levels of VHR were reduced in cancer tissues derived from NSCLC patients, and that nude mice injected with H1299 cells overexpressing VHR developed smaller tumours, thus linking VHR expression to suppression of cancer cell proliferation in NSCLC cells (Wang, Yeh et al., 2011).

In contrast to the above findings, VHR was shown to be overexpressed in prostate cancer as opposed to the normal prostate (Arnoldussen et al., 2008). This effect was linked to its ability to downregulate JNK thus preventing JNK-mediated apoptosis.

VHR was found overexpressed in several cervix cancer cell lines (Henkens et al., 2008). In those cell lines VHR was found in nucleus and in cytoplasm, whereas in normal cervix cells, it was localised in the cytoplasm. Increased levels of VHR were the result of increased protein stability because mRNA levels of VHR in normal and cancer cervix cells were comparable. In the light of the previously reported demonstration that VHR loss by siRNA caused cell senescence (Rahmouni et al., 2006), it was postulated that increased protein level of VHR in cervix cancer cells would enable the cells to proliferate (Henkens et al., 2008).

Development of small-molecule inhibitors of VHR is not just appealing from the stand of the therapeutic development, but also as a potential tool to manipulate physiological roles of VHR. The collaborative work of Osada and Sodeoka (Usui et al., 2001) resulted in the development of a potent competitive inhibitor

(RK-682) of VHR. The most promising aspect in VHR selective inhibitor development is by exploring its unique structural features. The aforementioned study showed that the inhibitor was binding to VHR as a dimer, potentiating association through extensive hydrophobic interactions. Arg158 was shown to be an important residue for inhibition. Although this residue is conserved in many PTPs (Yuvaniyama et al., 1996), its surrounding is markedly different. Moreover, Shi et al. docked over 80,000 compounds into the active site of VHR to isolate the most promising candidates (Shi et al., 2007). (Glucosamine-aminoethoxy) triphenyltin (GATPT) was identified as a competitive inhibitor, which caused accumulation of the phosphorylated Erk1/2 and JNK in HeLa cells due to inhibiting VHR's catalytic activity. The screening approach was marked as promising to increase the structural space of potential VHR inhibitors, and could be especially beneficial in the early phases of therapeutic development. The collaborative work of Mustelin and Tautz (Wu, Vossius, Rahmouni et al., 2009) was an extrapolation of the previous findings linking VHR to progression of cervical cancer (Henkens et al., 2008). The lead compounds were identified by chemical library screening. Docking of the most potent compound suggested that three hydrophobic patches proximal to the VHR's active site could be used for structural optimisation of the inhibitors, also demonstrated by the crystal structure of VHR in complex with the inhibitor SA3. Importantly, these hydrophobic areas are unique for VHR and comprised in part of the residues of the "recognition region" and "variable insert" segments in VHR (Yuvaniyama et al., 1996). SA3 compound was most efficient in decreasing proliferation rates of cervix cell lines HeLa and CaSki, thus proving that VHR could be exploited as a pharmaceutical target in treating cervix cancer.

1.5 The PRLs

The PRLs belong to a unique family of the VH1-like group of the Class I Cys-based PTPs (Alonso et al., 2004; Tonks, 2013). The PRL family of phosphatases consists of only three members, namely PRL-1, PRL-2 and PRL-3. The PRLs contain the CAAX motif (where C is cysteine, A is an aliphatic amino acid and X is any amino acid) at the C-terminal end, which is a unique feature amongst PTPs. They contain a consensus sequence for the PTPs, but apart from the PTP-signature motif, they lack recognisable domains which may partake in substrate recruitment (Bessette et al., 2007; Kozlov et al., 2004). In addition, as opposed to many PTPs, the PRLs carry an alanine in the position of the conserved serine in the signature motif. The PRLs are small phosphatases of about 20 kilodaltons (kDa) in size. PRL-1 and PRL-3 contain 173 amino acids, whereas PRL-2 contains 167 amino acids (Kozlov et al., 2004).

Rat PRL-1 was the first member of the PRL group of phosphatases to be identified as an immediate-early gene induced in regenerating liver after partial hepatectomy, and in insulin-treated H35 rat hepatoma cell line, where it was found to be consistently expressed (Diamond et al., 1994). Mouse PRL-2 and PRL-3 were discovered by searching the Expressed Sequence Tags (ESTs) database for homologues of PRL-1 (Zeng et al., 1998). Human PRL-1 and PRL-2 were also identified in an *in vitro* prenylation screen (Cates et al., 1996). The same study pinpointed PRL-1 and PRL-2 as oncogenic PTPs, because injecting nude mice with PRL-1 and PRL-2 stably expressed in D27 hamster pancreatic ductal epithelial cells caused tumour development.

1.5.1 Prenylation of the PRLs

As already mentioned, the PRLs are the only members of the PTPs to be modified by prenylation (Tonks, 2013). Protein prenylation is a type of PTM in which either farnesyl (15 carbon atoms) or geranylgeranyl (20 carbon atoms) isoprenoid units are covalently attached to a cysteine residue at the C-terminal end of proteins, aiding their localisation to the intracellular membrane compartments and also affecting protein-protein interactions (Gao et al., 2009; Si et al., 2001; Zhang and Casey, 1996). Protein prenylation proceeds via three distinct steps. In the first step, a lipid moiety is attached to a cysteine residue of the CAAX motif by prenyl-transferases (PTs). While PTs reside in the cytosol, the enzymes required for the following steps are membrane-associated (Choy et al., 1999). In the second step, prenylated CAAX motif is recognised by a prenyl-CAAX protease, which is localised on the endoplasmic reticulum (ER), releasing the -AAX residues (Gao et al., 2009). The final step involves methylation of the α-carboxyl group of the prenylated cysteine by a prenylcysteine carboxymethyltransferase.

In general, prenylated proteins are characterised either by the presence of the CAAX motif (also referred to as the CAAX-box) at their C-terminal end, or by the CCXX, XXCC or XCXC motifs (Gao et al., 2009; Si et al., 2001). The latter is mainly a feature of the Rab family of small guanosine triphosphate hydrolases (GTPases) (Zhang and Casey, 1996). Geranylgeranyl-transferase (GGT) II prenylates a substrate when it is bound to a carrier protein called Rab escort protein (REP) (Zeng et al., 2000). The former compiles proteins modifiable by farnesyl-transferases (FTs) and GGTs I and II (Gao et al., 2009). The type of lipid modification is determined by the identity of the amino acid in

the “X” position of the CAAX motif (Zeng et al., 2000). FTs prefer methionine, serine or glutamine, while GGT I prefers leucine.

In vitro prenylation of the purified recombinant PRLs showed that PRL-1 and PRL-2 could be modified by both farnesylation and geranylgeranylation, whereas PRL-3 was only farnesylated (Zeng et al., 2000). Interestingly, an earlier *in vitro* study of prenylated proteins showed that both PRL-1 and PRL-2 could be modified by farnesylation, whereas the signal for geranylgeranylation was much weaker (Cates et al., 1996). The study also identified FTs as the enzymes responsible for both lipid modifications. Recently, it was reported that PRL-3 transiently overexpressed in human embryonic kidney (HEK) 293 cells could also be modified by palmitoylation, whereas the same modification was not detected in PRL-1 and PRL-2 (Nishimura and Linder, 2013). Interestingly, the same study identified that brain specific splice variant of Cdc42 (bCdc42) with dual lipid modification exhibits distinct subcellular localisation pattern and functionalities compared to the variant that was shown to be only prenylated. In addition, bCdc42 which contained both prenyl and palmitoyl moieties was more prone to localisation to the plasma membrane. Given that PRL-1 and PRL-2 were shown to be only classically modified by prenylation, it would be interesting to see how this distinction with respect to PRL-3 reflects on functionalities.

1.5.2 Expression of the PRLs

Rat PRL-1 mRNA was found at high levels in muscle and brain (Diamond et al., 1994). Expression pattern of murine PRLs is different in various tissues (Zeng et al., 1998; Bessette et al., 2007). Mouse PRL-1 mRNA shows ubiquitous distribution, but is excluded from skeletal muscle and heart, where PRL-3 is

abundantly expressed (Lin et al., 2013). PRL-2 mRNA in mouse and rat is more widely expressed compared to the other PRL-counterparts. Human PRL-1 and PRL-2 mRNAs are ubiquitously expressed, and PRL-3 is highly expressed in skeletal muscle (Dumaual et al., 2006). In addition, PRL-3 was found expressed in developing blood vessels and erythrocytes at the mRNA and protein levels, but not in the corresponding adult counterparts (Guo et al., 2006). Furthermore, Zimmerman et al., found detectable levels of PRL-3 also in fetal intestine, adult heart, pancreas and a number of other organs, but not in the liver or kidney (Zimmerman et al., 2013). This data indicate that, in spite of high degree of amino acid sequence homology, PRLs could exhibit tissue-specific expression patterns, with interspecies variability.

Furthermore, both PRL-1 and PRL-3 were found expressed in the intestinal epithelial cells (Diamond et al., 1996; Zeng et al., 2000). They were found associated with the differentiated cells of the villus and not with the proliferating cells of the crypt. However, the work of Dumaual et al., that focused on the investigation of the expression pattern of PRL-1 and PRL-2 in normal human adult tissues, also detected PRL-1 mRNA in the cell of the crypt, thus suggesting that there might exist posttranscriptional regulatory mechanisms for the expression of PRL-1 mRNA (Dumaual et al., 2006). The early studies conducted on PRL-1 suggested that its role in the intestine might be linked to differentiation, and in liver to proliferation (Diamond et al., 1994). In addition, PRL-1 was also found expressed at the protein level in developing digestive tissues, where it was associated with differentiation (Kong et al., 2000). However, it cannot be excluded that the differential roles PRL-1 might have in regenerating liver and intestine were not the result either of diverse set of potential substrates present in different tissues, or of differences in the

amplitude and/or duration of the stimulus leading to distinct outcomes, like observed for MAPKs (Diamond et al., 1996).

1.5.3 Subcellular localisation of the PRLs

The data on the subcellular localiztion of the PRLs is not consistent (Fiordalisi et al., 2006; Song et al., 2009; Wang et al., 2002; Zeng et al., 2003).

Zeng et al. demonstrated that ectopically expressed PRL-3 localizes to the cytoplasmic site of the plasma membrane and to the punctate cytoplasmic structures resembling endosomes (Zeng et al., 2000). The protein could not be detected in the proliferating epithelial cells of the crypt in the small intestine but only in the villi, indicating possible role of PRL-3 in differentiation-linked processes. In addition, after the treatment with the FTase inhibitor, or by employing PRL-2 mutant lacking the consensus prenylation motif, the protein was relocalised to the nucleus thus demonstrating the importance of prenylation for subcellular localisation.

PRL-1 was identified as a phosphatase with predominantly nuclear localisation (Diamond et al., 1994). A later study (Wang et al., 2002) reported that localisation of PRL-1 was cell-cycle-dependent, by investigating endogenous phosphatase in HeLa cells. Generating green-fluorescent protein (GFP)-fused variants carrying different mutations in the CAAX motif demonstrated prenylation-dependent localisation to the ER in non-mitotic cells. In the mitotic phase, PRL-1 relocated to the centrosomes and it was also found to associate with the microtubules of the mitotic spindle. Moreover, it was demonstrated that cell cycle-dependent shuttling of PRL-1 between nucleus and cytoplasm had role in normal progression of the cell cycle. The catalytically deficient mutant D72A exhibited slower progression through the cell cycle compared to the

native protein or to the non-farnesylation mutant C170S. The C170S mutant demonstrated that farnesylation was important for regulating dynamics of the mitotic spindle as its expression caused defects in form of lagging chromosomes between two daughter nuclei. Therefore, several lines of evidence indicated that PRL-1 is important for normal progression of the cell cycle, and that this function depends on its phosphatase activity.

By using Chinese hamster ovary (CHO) cells stably transfected with enhanced GFP (EGFP)-fused PRL-3 variants, Zeng et al. demonstrated enrichment of the protein in the dynamic membrane structures, such as ruffles or protrusions (Zeng et al., 2003). Furthermore, the role of the constituting residues of the prenylation motif, which in PRL-3 has CCVM sequence, on the subcellular localisation, was also investigated (Song et al., 2009). The Cys170 of the CAAX-box proved essential for determining the localisation of PRL-3 in B16F1 murine myeloma cells. The ectopically expressed myc-tagged wtPRL-3 was diffused throughout the cytoplasm and at the plasma membrane.

The work of Fiordalisi et al. showed that in SW480 colon cancer cells transfected with GFP-fused PRL-1 and PRL-3 these phosphatases reside predominantly in the endomembranes, and that this localisation depends on the prenylation (Fiordalisi et al., 2006). The localisation of PRL-3 in human myeloma OH-2 cells was shown to be cell-cycle dependent, with the predominant nuclear localisation in G0/G1 phase of the cell cycle (Fagerli et al., 2008). Although this indicated its plausible participation in cell cycle progression, PRL-3 knockdown failed to induce cell cycle arrest.

Overall, the reported differences in the localisation of PRL-3 may be explained through the use of differentially-tagged constructs or to the cell-type or cell-cycle-dependent variations.

1.5.4 Structural characteristics of the PRLs

The three members of the PRL-family in humans share 76-87% of amino acid sequence identity (Fagerli et al., 2007; Zeng et al., 2003 and Figure 7).

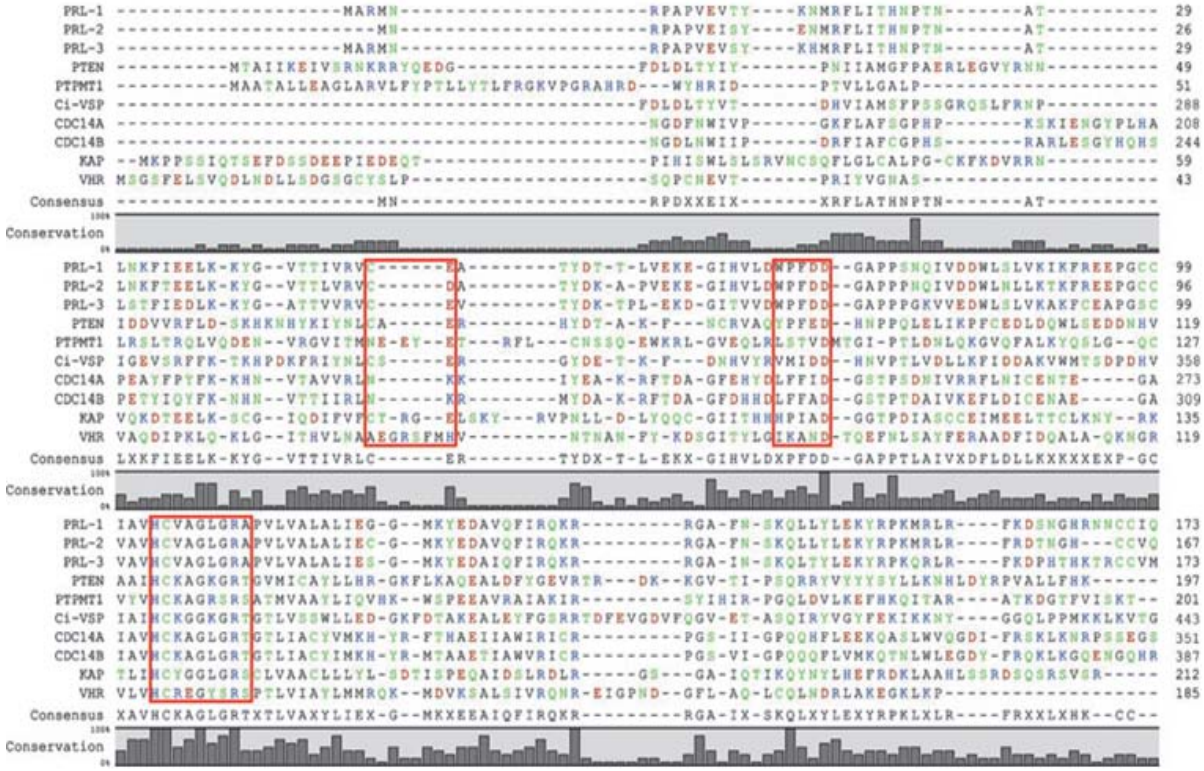


Figure 7: Structure-based multiple sequence alignment of all three PRLs and the structurally most closely-related PTPs. Ci-VSP and PTPMT1 show activity toward PI(4,5)P2 and PI(5)P, respectively. The PRLs are shown in full length; and the other PTPs in truncated versions according to relevance to THE sequence alignment with the PRLs. The amino acids are coloured by polarity: A, P, V, I, W, F, L, G, M (black); S, T, Y, N, C, Q (green); D, E (red); H, K, R (blue). The consensus residue and conservation rate at each position are shown below the sequences. An ambiguous residue is indicated as 'X'. The conserved 'CXnE' motif, the WPD-loop and the active site P-loop are indicated in red squares. With permission from Rios, P., Li, X. & Köhn, M. Molecular mechanisms of the PRL phosphatases. *FEBS J.* **280**, 505-524 (2013).

In addition, PRLs exhibit significant level of amino acid sequence homology between species: human and murine PRL-1 and PRL-2 are completely identical, while the PRL-3 phosphatases of the two species bear 96% sequence identity (Stephens et al., 2005). However, outside of its own family, the PRLs demonstrate quite low amino acid sequence homology in the regions excluding the residues of the PTP signature motif (Kozlov et al., 2004; Zeng et al., 1998). They display the highest homology to Cdc14p, a DUSP of *Saccharomyces cerevisiae* (*S. cerevisiae*) implicated in the regulation of cell cycle progression (Tonks, 2013; Zeng et al., 2000), and to PTEN, a DUSP with phosphoinositide phosphatase activity (Stephens et al., 2005). With Cdc14p and PTEN, the PRLs share around 20% and 17% sequence identity, respectively. The closest structural homologues of both PRL-1 and PRL-3 are VHR, PTEN and kinase-associated phosphatase (KAP) (Kozlov et al., 2004; Sun, Wang et al., 2005 and see Figure 7). A new study also recognised the structure-sequence-based similarity to the myotubularins, which are also phosphoinositide phosphatases (Li et al., 2013).

The solution structure of PRL-3 was solved by two independent studies (Kim et al., 2004; Kozlov et al., 2004). Kim et al. reported solution structure of PRL-3 solved by nuclear magnetic resonance (NMR) study (Kim et al., 2004). The structure was of the apo-form of the protein, comprising residues 1-162. The structural report showed that the loops forming the active site were significantly diverted from the catalytically most optimal conformation. In addition, the solution structure of PRL-3 solved by Kozlov et al., coupled to kinetic profiling of the site-specific mutants (C49A, D71A, D72A, A111S), shed light on their functional significance and helped to explain some structural distinctions of PRL-3 comparing to other DUSPs (Kozlov et al., 2004). The residues of the

PTP-signature motif are more hydrophobic in PRL-3 then in VHR or PTEN (Kozlov et al., 2004; Stephens et al., 2005). The active site pocket is broad and shallow due to the differences in the surrounding loops. The architecture of the active site pocket would suffice for binding of substrates with pSer/pThr or pTyr residues, or phosphoinositol headgroups (Al-Aidaroos and Zeng, 2010; McParland, Varsano et al., 2011). Cys104 serves as a nucleophile in the phospho-substrate hydrolysis (Kozlov et al., 2004). Asp72 acts as a general acid/base, and it is localised in the WPFDD loop. This loop is highly flexible. Upon binding of a substrate, it can undergo a large conformational change flipping over the active site entrance and optimising the relative positions of the key catalytic residues. Fine-tuning the arrangement of the catalytically important residues upon substrate binding has been acknowledged also for other representatives of the DUSPs, such as PTP1B, or a number of MKPs (Caunt and Keyse, 2013; Kozlov et al., 2004).

Cys49 was identified as a regulatory cysteine residue (Kozlov et al., 2004). Under less reducing conditions, it formed a disulfide bond with the catalytic Cys104. Intramolecular disulfide bond formation represents a feature PRLs share with PTEN. It is postulated that reversible disulfide bond formation observed between Cys49 and Cys104 in both PRL-1 and PRL-3 under oxidising conditions might provide a mechanism to protect the active site cysteine from the excessive oxidation (Sun, Wang et al., 2005).

Most of the conserved residues amongst PRLs are either important for catalysis or structural integrity. The highest conservation is around the active site pocket and the lowest one on the side opposite to it. For example, position 141 contains isoleucine in PRL-3 and phenylalanine in PRL-1 and PRL-2. This

difference could have a role in fine-tuning substrate specificity (Kozlov et al., 2004).

The PTP signature motif of PRL-3 has an untypical amino acid profile because the conserved arginine is followed by alanine, whereas in PRL-1 this position is occupied by valine (Kozlov et al., 2004). As it can be seen from the sequence profile of the consensus signature motif, most PTPs have serine or threonine residing in this position. Serine or threonine in this position is important in catalytic reaction by facilitating the breakdown of the phosphoenzyme intermediate (Sun, Luo et al., 2007; Zhang, 1998). Generating the A111S mutant of PRL-3 and examining its activity against 3-*O*-methylfluorescein phosphate (OMFP) showed an increase in the catalytic activity (Kozlov et al., 2004). Thus, the extremely low activity of PRL-3 observed against artificial phosphatase substrates or synthetic peptides could be attributed to the lack of residue with the hydroxyl group (Kozlov et al., 2004; McParland, Varsano et al., 2011). It is interesting to note that the A111S mutant was not active against phosphoinositides, unlike the wt protein, although it could not be determined if this was due to the observed structural perturbation of this variant or differences in the catalytic mechanism (McParland, Varsano et al., 2011).

For PRL-1, it was demonstrated that it forms trimers in two independently solved crystal structures, as well as in solution at higher concentrations, with the active site facing on the outside of the trimer (Jeong et al., 2005; Sun, Wang et al., 2005). The trimeric association was stabilised due to extensive hydrophobic and polar interactions, and almost 30% of each monomer was burried at the interface. Moreover, cross-linking with glutaraldehyde of the membrane fraction of human embryonic kidney (HEK) 293 cells transiently expressing wtPRL-1 confirmed the intrinsic ability of PRL-1 for trimers.

Prenylation of PRL-1 is important for localising the protein to the plasma membrane, and increase in the local concentration of PRL-1 could contribute to the trimer formation. Interestingly, farnesylation alone was not sufficient to keep the protein at the plasma membrane, because the mutant in which six positively charged residues were exchanged for alanine shifted to the cytoplasm (Sun, Wang et al., 2005). When only the first three residues of the polybasic stretch in PRL-1 were exchanged for alanine, only a fraction of the protein was found in the soluble fraction and the majority still demonstrated localisation pattern comparable to the wt protein, unlike to the case where all six basic residues were exchanged to alanine, as already mentioned (Sun, Luo et al., 2007). The synergism of both electrostatic and hydrophobic interactions for maintaining PRL-1 at the plasma membrane is not without precedence, and analogous requirement has already been noted for K-Ras or myristoylated alanine-rich C kinase substrate (MARCKS) (Murray et al., 1997; Sun, Wang et al., 2005). Furthermore, for PRL-1 it was demonstrated that trimerisation was necessary for PRL-1-mediated increase in cell proliferation and migration, alongside its phosphatase activity and prenylation (Sun, Luo et al., 2007). When mutations were introduced at the dimer interface in PRL-1 (T13F and G97R), trimerisation was reduced and the ability of PRL-1 to promote cell proliferation and migration was diminished. In addition, trimeric association of PRL-1 did not affect its catalytic performance against *p*NPP, and no cooperativity between the catalytic centers was observed. However, the effect might exist at the level of real physiological substrates. Furthermore, for PRL-3 it was also demonstrated that it had the ability to form trimers *in vitro* and in cells (Sun, Luo et al., 2007).

The region in the PRLs adjacent to the modification site for prenylation (residues 136-161) could serve as a bipartite NLS (Pascaru, Tanase et al.,

2009; Zeng et al., 2000). This polybasic region could therefore aid in the membrane tethering of the PRLs by interacting electrostatically with the negatively charged phospholipids and/or particular plasma membrane proteins (Sun, Wang et al., 2005; Zeng et al., 2000). Pascaru, Tanase et al. tried to elucidate the role of the residues in the polybasic stretch in PRL-3, with emphasis on their possible function as NLS (Pascaru, Tanase et al., 2009). They reported that removal of the first three residues, Lys-Arg-Arg, of the potential NLS, did not render the localisation pattern different than the one of the wt protein, and that only additional removal of the CAAX-box residues shifted the protein into the nucleus, thus doubting that polybasic region functions as an NLS. PRL-2 lacks three amino acid residues directly adjacent to the prenylation motif (Si et al., 2001; Sun, Wang et al., 2005). Drawing a parallel to the Rab GTPases, where divergent region at the C-terminal end dictates the interaction with the β-subunit of the GGT II, one could envision similar pattern in PRL-2. Moreover, only PRL-2 was shown to interact with the β-subunit of the GGT II (Si et al., 2001). Because PRL-1 and PRL-3 bear an insertion of three amino acids directly in front of the CAAX-box, this patch could play a role in determining specificity for a particular subcellular localisation. This case would not be without precedence, as it was already demonstrated that the Ras GTPases require a stretch of basic residues in addition to farnesylation in order to trigger localisation to the plasma membrane (Murray et al., 1997). Moreover, heterotrimeric guanine nucleotide-binding proteins (G-proteins) provide another example where multiple lipid modifications are needed for proper localisation to the plasma membrane in order to enable signal propagation to the downstream components (Takida and Wedegaertner, 2003). Modifications of these proteins include the attachment of a myristate (14 carbon

atoms) and/or a palmitate to an N-terminal glycine or cysteine residue, respectively, of the α-subunit, and by prenylation of the γ-subunit.

For PRL-1 it was demonstrated that the residues of the CAAX-box and polybasic region have an important role not only in shaping subcellular localisation but also in binding affinity to different PIPs of the plasma membrane (Bessette et al., 2008; Sun, Luo et al., 2007). However, no catalytic activity of PRL-1 against phospholipids was observed (Sun, Luo et al., 2007; Yu et al., 2007).

The residues at the C-terminal end of the PRLs could be important not only for subcellular localisation of the proteins, but also as a sophisticated regulatory mechanism of their catalytic activity. Pascaru, Tanase et al. investigated the catalytic activity of various forms of PRL-3 mutated in the residues of the CAAX-box or the polybasic stretch (Pascaru, Tanase et al., 2009). They determined that deletion of the CAAX-box residues increased the catalytic efficiency, hypothesizing that the hydrophobic interactions between the -VM residues of the CAAX-box could facilitate oligomeric association between PRL-3 monomers, with a derogative effect on the catalytic activity. This observation is opposite to the observation that trimerisation induced no effect on the catalytic efficiency of PRL-1 *in vitro* (Sun, Luo et al., 2007).

In general, the investigations of the structural determinants of the PRLs could be grouped into two categories. One encompasses the residues of the P-loop and of the WPD-loop directly involved in the dephosphorylation reaction, and the residue Cys49 implicated in the redox-mediated regulation of the catalytic activity of PRL-1 and PRL-3 through transient formation of a disulfide bond with the catalytic cysteine (Fiordalisi et al., 2006; Ishii et al., 2013; Jeong et al.,

2005; Kozlov et al., 2004; McParland, Varsano et al., 2011; Sun, Wang et al., 2005; Zeng et al., 2003). The second group of research interest spans across the residues of the C-terminal polybasic region, which were suggested to potentially function as a bipartite NLS (Pascaru, Tanase et al., 2009; Sun, Luo et al., 2007; Zeng et al., 2000) and the CAAX-box at the C-terminal end of the protein, which are important for localizing the protein to the plasma membrane and punctuate structures resembling endosomes (Fiordalisi et al., 2006; Pascaru, Tanase et al., 2009; Rios et al., 2013; Skinner et al., 2009; Song et al., 2009). The more intense research focus on the detailed analysis of the various mutants for PRL-2 is still pending.

Apart from the residues of these two categories, no other in-depth profiling of the other non-homologous residues amongst the PRLs has been reported. On exception is the recent discovery of Src-mediated phosphorylation of PRL-3 (Fiordalisi et al., 2013). PRL-3 was shown to be phosphorylated on Tyr53 by Src. This modification is required for the PRL-3-mediated promotion of cell invasion and activation of RhoC GTPase. Figure 8 provides a summary of key structural features of PRL-3, with pointed out the residues important for catalytic mechanism, redox regulation and sub-celullar localisation.

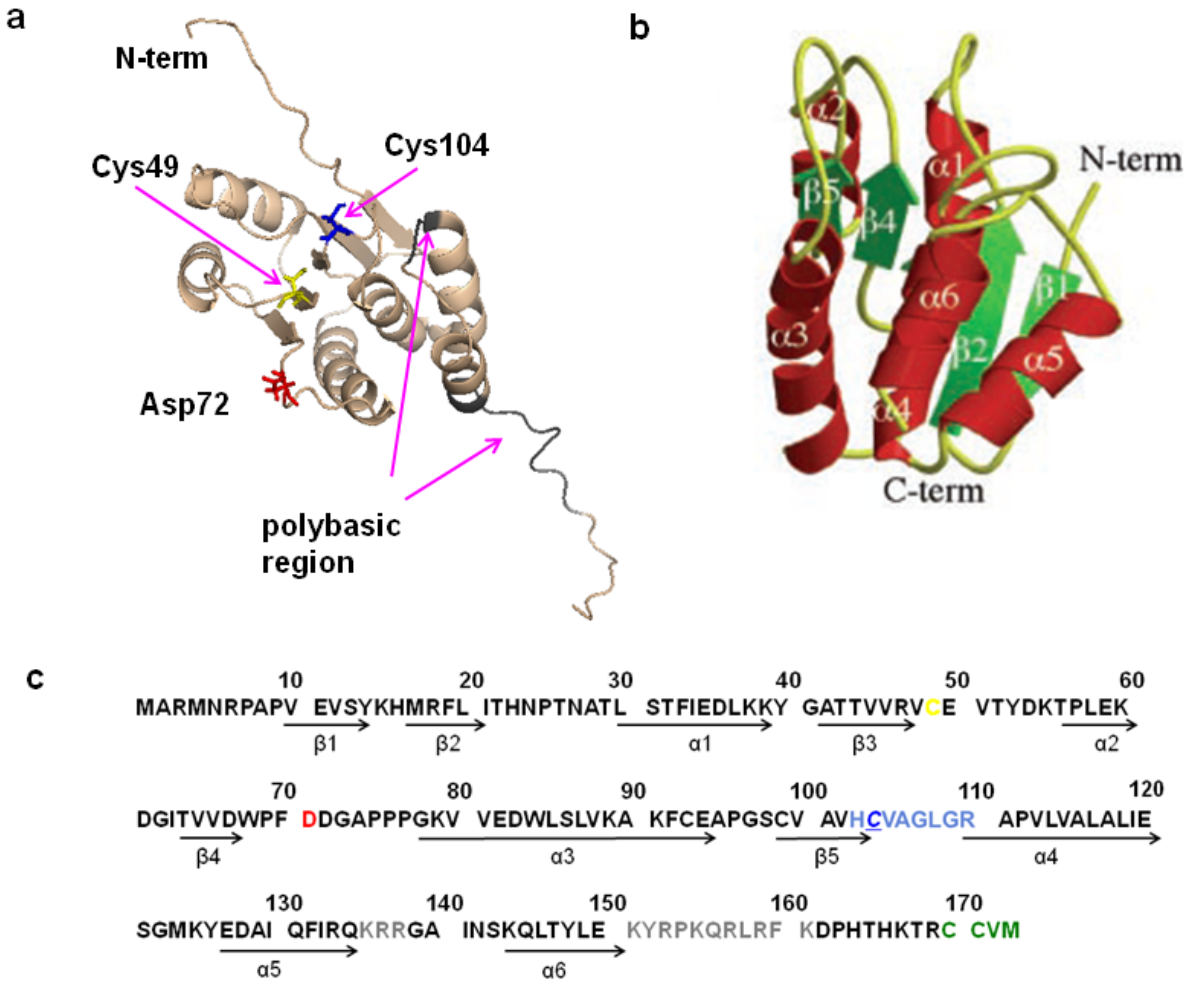


Figure 8: Solution structure of PRL-3 (PDB: 1R6H). (**a**) Structural representation of PRL-3 generated in Pymol with the most important structural features pointed out (see panel (**c**) for labeling). (**b**) Secondary structure elements of PRL-3. Reproduced from Kozlov, G. *et al.* Structural insights into molecular function of the metastasis-associated phosphatase PRL-3. *J. Biol. Chem.* **279**, 11882-11889 (2004) (no permission needed). (**c**) Sequence of PRL-3 with the elements of the secondary structure pointed out in the sequence. Highlighted are the following residues: Cys49 (yellow), general acid/base Asp72 (red), P-loop (blue, with catalytic Cys104 underlined), bi-partite polybasic region (grey) and CCVM of the CAAX-box (green).

1.5.5 The PRLs in cancer

Saha et al. applied the serial analysis of gene expression (SAGE) technology to identify consistent genetic alterations contributing to liver metastasis, as the primal cause of mortality of colorectal cancer patients (Saha et al., 2001). The PRL-3 gene was identified as the only one being highly expressed in all 18 colorectal cancer metastases. The same study demonstrated that the expression level of PRL-3 was not detectable in the normal epithelium, whilst reaching intermediate level in patients with an advanced disease state. In a later study, Saha et al. set another milestone by (i) demonstrating elevated expression levels of PRL-3 in metastases of colorectal cancer, regardless of the site of the metastatic lesions and by (ii) showing that levels of PRL-3 in liver metastases of different origins were low (Bardelli, Saha et al., 2003). Since then, the PRL family of phosphatases, and PRL-3 in particular, has received much research interest (Al-Aidaroos and Zeng, 2010; Rios et al., 2013). PRL-3 is regarded as a promising biomarker for monitoring tumour aggressiveness and as a predictive marker of colorectal cancer metastases to liver or lung, following surgical removal of the primary tumours (Al-Aidaroos and Zeng, 2010; Kato et al., 2004; Molleví et al., 2008). In addition, PRL-3 was found overexpressed in different types of tumours, including hepatocellular, nasopharyngeal, breast, gastric, lung or brain (Bardelli, Saha et al., 2003; Matsukwa et al., 2010; Parker et al., 2004; Zhao et al., 2008; Zhou, Wang et al., 2008). In addition, low levels of PRL-3 were detected in lung and liver metastases of primary cancers of non-colorectal origin (Bardelli, Saha et al., 2003; Stephens et al., 2005). Moreover, mRNA and protein levels of PRL-3 were increased in multiple myeloma cells, where depletion of PRL-3 reduced cell migration (Fagerli et al., 2008). The differences observed in the malignant

effect of PRL-3 in various systems studied could be an indication of a cell-type dependent effect of this phosphatase. PRL-3 mRNA was not detected in any of the 28 samples obtained from the pancreas cancer patients, whereas PRL-1 and PRL-2 were found upregulated in contrast to normal pancreas (Stephens et al., 2008). In addition, PRL-3 was shown to be upregulated in ovarian cancer as opposed to normal ovarian tissue, and to promote tumour development in mouse model, thus supporting a role of PRL-3 as an early prognostic marker in human ovarian cancer (Bessette et al., 2008; Liu, Al-Aidaroos et al., 2013). Downregulation of PRL-3 in ovarian cancer cells A2780 elevated protein and mRNA levels of integrin α2 via upregulation of transcription factors c-jun and c-fos and also reduced levels of adhesion protein paxillin. No effect on integrin β1, which was demonstrated to interact with PRL-3 (Tian et al., 2012), was observed thus suggesting that in ovarian cancer PRL-3 might downregulate integrin α2 thus contributing to cancer progression (Varner and Cheresh, 1996).

Furthermore, PRL-3 has been shown to have role in angiogenesis in several independent studies (Al-Aidaroos and Zeng, 2010; Bessette et al., 2007; Bessette et al., 2008; Guo et al., 2004; Guo et al., 2006). Guo et al. discovered that injection of CHO cells stably expressing wtPRL-3 into the tail vein of nude mice led to formation of vascularized tumours in the lung, whereas no metastases were detected for injection of C104S mutant (Guo et al., 2004). Moreover, sometimes the tumours were formed within the existing blood vessels. The authors suggested that PRL-3 expressing cancer cells might secrete proteases which would digest basal lamina of the surrounding blood vessels thus enabling invasion of the cancer cells into the circulatory system (Chambers et al., 2002; Folkman and Klagsbrun, 1987; Hanahan and Weinberg, 2000). Furthermore, additional study demonstrated that CHO cells

and human colon cancer DLD-1 cells expressing wtPRL-3 had the ability to recruit endothelial cells in an angiogenesis co-culturing assay *in vitro* (Guo et al., 2006). Moreover, when murine lung cancer cells were introduced subcutaneously into nude mice, only wtPRL-3 expressing cells demonstrated the ability to develop tumours with significant degree of vascularization. The mechanism that led to endothelial cells being recruited by PRL-3 expressing cells could involve either endothelial cells responding to specific surface factors on PRL-3 expressing cells, or the latter ones releasing diffusible angiogenesis-stimulating factors. Thus, these findings suggested that phosphatase activity of PRL-3 is an important factor in PRL-3-mediated role in angiogenesis, and that therapeutics which would inhibit it might find application in antiangiogenic therapy.

The role of the PRLs in promoting cell proliferation and metastases development was early recognised (Cates et al., 1996; Diamond et al., 1994; Zeng et al., 2003). Wang et al. investigated expression of PRL-1 mRNA in a broad panel of normal and cancer cell lines and demonstrated that PRL-1 mRNA was increased in a subset of cancer cells (Stephens et al., 2005; Wang et al., 2002). Furthermore, D27 hamster pancreatic epithelial cells overexpressing PRL-1 and PRL-2 demonstrated altered growth rates, characterised by shorter doubling time comparing to the vector control, and were able to induce tumour formation in nude mice (Cates et al., 1996).

In addition, nonmetastatic CHO cells stably overexpressing PRL-1 and PRL-3 exhibited enhanced cell motility and invasiveness compared to the control cells, and were also able to induce liver and lung metastases when injected in nude mice (Zeng et al., 2003). The transforming phenotype was dependent on the phosphatase activity, because the inactive C104S mutant of PRL-3 showed a

reduction in cell motility, albeit not at the level of the control cells. This result indicated that overexpression of the catalytically deficient mutant C104S could facilitate increase in cell motility in a manner independent of the catalytic activity (Bessette et al., 2007; Zeng et al., 2003). Furthermore, injection of B16F1 murine melanoma cells, with low metastatic potential, overexpressing wtPRL-3 and different mutants in the CAAX-box sequence, induced severe lung metastases in nude mice in a farnesylation-dependent manner (Song et al., 2009; Stephens et al., 2005). These results clearly indicated a causative role of the PRLs in promoting tumourigenesis. Recently, PRL-3 knockout mouse model was reported (Zimmerman et al., 2013). Mice lacking PRL-3 developed normally, although it was noticed that a number of male mice was little declined. Nevertheless, this was in contrast to the effect reported for PRL-2 knockout (Dong et al., 2012). It was demonstrated that treatment with a procarcinogen called azoxymethane (AOM) resulted in the immediately increased expression of PRL-3, whereas PRL-3 knockout mice exhibited over 50% reduction in tumour formation (Zimmerman et al., 2013). These findings indicated that PRL-3 could also be actively involved in the early stages of cancer. AOM acts as a DNA damaging agent, which induces expression of p53 and, subsequently, PRL-3, thus accounting for the observed early increase in PRL-3 expression following genotoxic stress. For comparison, PRL-2 knockout mice suffered growth retardation in embryos, due to impaired development of placenta, and also in adult mice (Dong et al., 2012; Dumaual et al., 2006). In PRL-2 knockout mice, the effect of PRL-2 deficiency was the result of reduced synthesis of glycogen (Dong et al., 2012). The level of active glycogen synthase kinase (GSK) 3*β* was lower due to decreased phosphorylation level of the enzyme responsible for its activation, Akt kinase, and increased level of

PTEN, which negatively regulates phosphatidylinositol 3-kinase (PI3K)/Akt pathway (Dong et al., 2012; Lee, Yang et al., 1999).

Although the causal role of PRL-3 in promoting metastasis and its potential to serve as a prognostic indicator of aggressiveness and clinical outcome for various types of tumours has been well documented, the data for PRL-1 and PRL-2 was limited (Al-Aidaroos and Zeng, 2010; Bessette et al., 2008). Recently, an extensive investigation of PRL-1 and PRL-2 distribution in normal and cancer human tissues of diverse origin, by *in situ* hybridization, was reported (Dumaual et al., 2012). It was shown that PRL-1 and PRL-2 mRNAs were expressed in both normal and cancer tissues examined. PRL-1 transcript was elevated in majority of breast, ovarian and renal cancer sample, and in all cancer samples derived from the bladder, colon, cervix, liver, lung, pancreas, prostate, skin, testis and stomach. Moreover, PRL-1 mRNA was also associated with all examined samples of the metastatic lesions to the liver and uterus. While the samples for PRL-1 showed variability between tissues and also within a particular tissue type, the levels of PRL-2 transcript were high in vast majority of the samples of normal or cancer origin. When correlating the transcript levels of the PRLs to the protein levels, it was shown that PRL-1 mRNA and protein levels exhibited similar relative expression level, whereas for PRL-2 this correlation was observed for only about half of the cases examined. Interestingly, Dumaual et al. showed that some cancer tissue samples examined had lower levels of PRL-1 and PRL-2 mRNAs comparing to their normal counterparts. PRL-1 was lower in all or majority of the cancer samples derived from the ovary, breast or lung, and PRL-2 in about half of the samples from kidney cancer, suggesting that in some tissues PRL-1 and PRL-2 may not have metastasis stimulatory role. Both PRL-1 and PRL-2 exhibited

mostly nuclear localisation, with occasional staining in the cytoplasm, in normal and cancer samples examined. In addition, PRL-2 was suggested as a marker for predicting breast cancer metastasis (Hardy et al., 2010). Its levels were found increased in various mammalian mammary cell lines. Moreover, primary breast tumours and lymph node metastases exhibited increased levels of PRL-2 transcript, and it was also shown that PRL-2 correlated with the course of breast cancer development. Furthermore, increased levels of PRL-2 were observed in cell lines derived from pancreatic and prostate cancers, and also in acute myeloid leukemia (AML) (Wang and Lazo, 2012).

1.5.6 Regulation of the PRLs by oxidation

Like already mentioned, the PRLs contain a regulatory cysteine residue at position 49 (Jeong et al., 2005; Kozlov et al., 1994). In general, the architecture of the active site of the PTPs renders them sensitive to inhibition due to susceptibility of the catalytic cysteine to oxidation (den Hertog et al., 2008; Funato and Miki, 2013; Tonks, 2006). Oxidation as a general regulatory principle of the PTPs has been addressed previously. It was demonstrated recently that all three PRLs can be oxidised in cells after H_2O_2 treatment (Ishii et al., 2013). Based on the crystal structure of PRL-1, generation of a disulfide bond between Cys49 and Cys104 would not cause significant structural perturbations of the protein, as the two cysteines are close in space and without obstructing structural elements between them (Jeong et al., 2005). Functional implications of the intramolecular disulfide bond formation in PRL-1 in cells were also investigated. Skinner et al. reported that PRL-1 in cells would be predominantly in the oxidised and, therefore, inactive form (Skinner et al., 2009). They determined that the reduction potential of the disulfide bond formed between Cys49 and the active site Cys104 was lower than the

reduction potential range in the normal cellular environments. In addition, the same study showed that the residues of the CAAX-box (which in PRL-1 has sequence CCIQ) influence the catalytic activity of the protein, possibly by causing conformational changes which would eventually affect the reduction potential of the disulfide bridge between the regulatory and the catalytic cysteines. The findings by Skinner et al. also helped to interpret earlier observations where cells expressing PRL-1 C170S mutant showed various mitotic defects (Wang et al., 2002).

Furthermore, it was shown that PRL-1 mRNA and protein expression increased in retinal cells following light-induced oxidative stress (Al-Aidaroos and Zeng, 2010; Yu et al., 2007). The activity of PRL-1 was modulated under oxidative conditions by disulfide bond formation between the conserved cysteines at positions 49 and 104. This phenomenon was observed *in vitro*, in cultured cells subjected to H_2O_2 and in rat retina after constant light exposure. The conversion into the reduced form of PRL-1 was found to be mediated specifically by the glutathione (GSH) system, because in the presence of the GSH-specific inhibitor, the extent of the oxidised PRL-1 in H_2O_2 treated cells increased. This study opened up the area of the possible involvement of PRL-1 in oxidative stress-mediated cellular responses.

Susceptibility of PRL-3 to oxidation was also demonstrated in an MS study (Orsatti et al., 2009a). Orsatti et al. showed that PRL-3 has two possible mechanisms of inactivation by oxidation. One included the already mentioned disulfide bridge formed between Cys104 and Cys 49, whereas the other one involved conversion of the catalytic cysteine to glycine. The latter inhibitory principle was not known amongst the PTPs.

Recently, thioredoxin-related protein (TRP) 32 was demonstrated to regulate reduction of PRL-3 after H_2O_2-induced oxidation in cultured cells (Ishii et al., 2013). TRP32 may be the key component in the regulation of the redox status of PRL-3 because knockdown of the endogenous TRP32 caused prolonged oxidation of the endogenous PRL-3. This study implicated that TRP32 could promote the causative role of PRL-3 in metastasis by regulating its catalytic ability.

1.5.7 Other regulatory mechanisms of the PRLs: PRL-1 and PRL-2

It was determined that the 5'- non-coding regions of the murine PRLs mRNAs share no more than 30% sequence identity, indicating possible differences at the gene level for regulation of the individual PRLs (Zeng et al., 1998). This observation is supported by varieties in tissue distributions of the individual PRLs, like already discussed. Furthermore, Zeng et al. applied the ScanProsie programme which identified potential phosphorylation sites in murine PRLs. Interestingly, PRLs exhibited differences in the number of motifs which could be phosphorylated by protein kinase C (PKC), casein kinase II or PTK (Stephens et al., 2005; Zeng et al., 1998). PRL-2 only contains two potential sites for phosphorylation by PTK. For PRL-1, one site for PKC and Casein kinase II were identified, whereas PRL-3 contained two and four sites, respectively, for these kinases. Further studies investigating more deeply whether phosphorylation could serve as a regulatory principle of PRLs and/or if differences in the possible phosphorylation sites by PKs would reflect in various regulatory principles of the PRLs have remained fairly scarce. Recently, Fiordalisi et al. reported significance of Src-mediated phosphorylation of Tyr53 for PRL-3 mediated effect in promoting enhanced invasiveness and motility in

SW40 colon adenocarcinoma cells, which is discussed in more detail in a later chapter (Fiordalisi et al., 2013).

PRL-1 was demonstrated to be regulated at the transcriptional level by the PRL-1 intron enhancer complex (PIEC), a developmentally regulated factor which binds to the first intron in the PRL-1 gene, and by an immediate-early gene encoded transcription factor called early growth response (Egr) factor 1, which was shown to bind to the P1 promoter of the PRL-1 gene (Al-Aidaroos and Zeng, 2010; Bessette et al., 2008; Peng et al., 1999). The expression level of Egr-1 and its binding to the P1 site of PRL-1 promoter were followed by an increase in the level of the PRL-1 transcript immediately after partial liver hepatectomy or following mitogen stimulation of NIH 3T3 mouse fibroblast cells. As the level of PRL-1 mRNA was found increased for up to 24 hours after hepatectomy, it was suggested that additional regulatory mechanisms exist, which were likely acting on the post-transcriptional level (Bessette et al., 2008; Peng et al., 1999). Moreover, Egr-1 did not resolve regulation of PRL-1 expression in tissues with normally high levels of the phosphatase nor in cancer cells implying the existence of tissue-specific regulatory principles (Peng et al., 1999).

Furthermore, both PRL-1 and PRL-3 were identified as p53 transcriptional targets (Al-Aidaroos and Zeng, 2010; Min, Kim et al., 2009; Stephens et al., 2005). p53 functions as a transcriptional regulator of many genes involved in various cellular functions, such as cell cycle progression, DNA repair or apoptosis, but it also regulates expression of ubiquitin ligases implicated in its own downregulation, which are encoded by RCHY1 ring finger and CHY zinc finger domain containing 1 (PIRH2), mouse double minute 2 homolog (MDM2)

and caspase recruitment domain-containing protein 16 (COP1) genes (Al-Aidaroos and Zeng, 2010; Min, Kim et al., 2009). One of the p53 targets is cyclin-dependent kinase (CDK) inhibitor (CKI) $p21^{Cip1/Waf1}$. Werner et al. identified that PRL-1 overexpressing cells had higher activity of CDK 2, higher level of cyclin A and reduced level of CKI $p21^{Cip1/Waf1}$, resulting in higher rate of cell proliferation (Werner et al., 2003). CDKs contain catalytic kinase and regulatory cyclin subunits (Song et al., 2001). Progress through the cell cycle is accompanied by periodic changes in the activity of CDK governed by the changes in the phosphorylation status of the complex and regulated synthesis/degradation of the constituting subunits. The work of Min, Kim et al. elucidated the mechanism governing PRL-1 mediated reduction in levels of $p21^{Cip1/Waf1}$ (Min, Kim et al., 2009). Overexpression of PRL-1 was shown to reduce the levels of p53 through two distinct pathways, both of which facilitated proteosomal degradation of p53 via ubiquitination associated with induction of PIRH2 and MDM2. In HeLa cells expressing PRL-1, the transcript levels of PIRH2 and Egr-1 transcription factor were elevated. In relation, Fiordalisi et al. already demonstrated that PRL-1 regulates expression of serum response element (SRE) transcription factor (Fiordalisi et al., 2006). Interestingly, Egr-1 was also known to positively regulate transcription of p53. The second pathway implicated in the downregulation of p53 involved Akt-mediated phosphorylation of MDM2 (Min, Kim et al., 2009). In conclusion, PRL-1 was added to the list of oncogenic transcriptional targets of p53.

1.5.8 Interacting proteins and substrates of PRL-1 and PRL-2

Only few substrates and interacting proteins for PRL-3 have been suggested (Al-Aidaroos and Zeng, 2010; Rios et al., 2013), thus making the investigation

of PRL-3-mediated effects on signaling pathways a notoriously difficult task. The data is even more obscure on PRL-1 and PRL-2.

Diamond et al. recognised that PRL-1 contains the consensus sequence motif for the PTPs (Diamond et al., 1994). They showed that PRL-1 can be phosphorylated by cellular (c)-Src *in vitro* and that phosphorylated PRL-1 could dephosphorylate itself and none of the control proteins. The link between trimer formation of PRL-1 and its autophosphatase activity has not been investigated (Jeong et al., 2005; Sun, Luo et al., 2007).

The only reported substrate of PRL-1 is activating transcription factor (ATF)-7, which is also known as ATF-5 or ATF-X (Al-Aidaroos and Zeng, 2010; Peters et al., 2001; Stephens et al., 2005). ATF-7 belongs to the family of transcription factors known as basic leucine zipper (bZIP) proteins. Its homodimer binds to cyclic AMP (cAMP) response elements (CRE). ATF-7 was identified as a PRL-1-interacting protein in a yeast-two hybrid (Y2H) screen (Peters et al., 2001). Subsequently it was demonstrated that it can be dephosphorylated by PRL-1 *in vitro*, albeit only partially. The work of Peters et al. suggested that PRL-1 and ATF-7 pair could represent phosphorylation-mediated activity regulation of a transcription factor, but also that ATF-7 could act as a scaffold for PRL-1. The functional significance of the ATF-7 dephosphorylation by PRL-1 has remained uninvestigated (Al-Aidaroos and Zeng, 2010).

Recently, p115 Rho GTPase-activating protein (GAP) was identified as a novel PRL-1 binding protein (Bai, Luo et al., 2011). p115 RhoGAP was identified by probing a phage display peptide library with recombinant PRL-1. A peptide which demonstrated binding to PRL-1 corresponded to the motif in the SH3 domain of p115 RhoGAP, and this domain was demonstrated to be

indispensable for its interaction with PRL-1. Furthermore, it was established that overexpression of p115 RhoGAP negatively correlated with PRL-1-mediated upregulation of the activation of Erk1/2 and RhoA, and cell migration (Bai, Luo et al., 2011; Fiordalisi et al., 2006; Luo et al., 2009). Binding of PRL-1 to the SH3 domain of p115 RhoGAP led to an increase in the level of active, GTP-bound, form of RhoA, whereas the activation of Erk1/2 was achieved by displacing p115 RhoGAP from association with the upstream kinase implicated in the activation of Erk1/2, MEKK1. These findings further corroborated the role of PRL-1 in promoting the acquisition of metastatic properties, such as enhanced cell motility and invasion.

GGT II was identified as a PRL-2 interacting protein in a Y2H screen (Al-Aidaroos and Zeng, 2010; Si et al., 2001; Stephens et al., 2005). Intriguingly, PRL-1 and PRL-3 were not found to interact with GGT II. The interaction between PRL-2 and GGT II was dependent on the prenylation status of PRL-2. In the presence of FT inhibitor or by using PRL-2 variant with deleted residues of the CAAX-box, the interaction was not detectable. PRL-2 did not act as a substrate for GGT II as no modification of PRL-2 by geranylgeranylation was detectable *in vivo*. In addition, PRL-2 was found to interact with the β-subunit of GGT II, which contains most of the catalytically important residues, but it displays activity only when in complex with the α-subunit (Si et al., 2001). Thus, an attractive hypothesis is that, by competing with the α-subunit for binding to the β-subunit of GGT II, PRL-2 could participate in the regulation of the activity of the GGT II. In addition, the constituting subunits of the GGT II heterodimer are not expressed equally in all the tissues, and the levels of PRL-2 and of the β-subunit of the GGT II were found to be high in muscle.

1.5.9 The PRLs in signaling: PRL-1 and PRL-2

Several signaling pathways in which PRL-1 and PRL-3 are engaged have been described, whereas investigation of PRL-2 is somewhat lagging behind (Al-Aidaroos and Zeng, 2010; Rios et al., 2013).

As mentioned previously, already the early studies of the PRLs implicated the involvement of this family of phosphatases in promoting cell proliferation (Diamond et al., 1994; Cates et al., 1996). Cells stably overexpressing PRL-1 and PRL-2 showed higher growth rates and in some cases elevated levels of PRL-1 resulted in multinuclear cell morphology. Diamond et al. showed that PRL-1 can be phosphorylated by Src *in vitro*. Src family kinases (SFKs) are involved in cell proliferation, migration or apoptosis (Achiwa and Lazo, 2007; Luo et al., 2009). In a recent study, Src and EGFR kinases were identified as the key upstream targets of H_2O_2-mediated increase in protein tyrosine phosphorylation (Chan et al., 2010). Interestingly, many of the proteins whose tyrosine phosphorylation level was elevated after treatment with H_2O_2 were found to be important for cell adhesion and regulation of cytoskeleton, whereas some were also known substrates of Src. The aforementioned study by Achiwa and Lazo investigated the effect of depletion of the expression of the endogenous PRL-1 in A549 lung cancer cell line by short hairpin RNA (shRNA). It demonstrated reduction in cell proliferation and increase in cell adhesion following reduction in the expression levels of PRL-1. The protein level of Src was decreased, but its phosphorylation status was not affected, and neither was the protein level of C-terminal Src kinase (Csk). Csk acts as a negative regulator of the SFKs, while the former determines if the protein is going to be in the active form. In relation, autophosphorylation of Tyr418 by SFKs leads to the activation, and phosphorylation of Tyr529 by Csk to the

inhibition of the Src's enzymatic activity (Luo et al., 2009). Also, the levels of p130 Crk-associated substrate (p130Cas) and of the active GTP-bound forms of Rac1 and Cdc42 GTPases were reduced concomitant with the reduction of the protein level of PRL-1 (Achiwa and Lazo, 2007). The members of the Rho family of GTPases are known key regulators of the actin cytoskeleton dynamics (Fiordalisi et al., 2006). In cells with reduced levels of both PRL-1 and Src, the focal adhesion turnover was slowed down (Achiwa and Lazo, 2007). Nakashima and Lazo investigated further the molecular mechanisms of the PRL-1-mediated effect on promoting cell proliferation, motility and invasion in A549 lung cancer cells (Nakashima and Lazo, 2010). Intriguingly, they observed that both the wtPRL-1 and its D72A mutant decreased the levels of the activated forms of Rac1 and Cdc42. Thus, it is possible that PRL-1 could exert its effect on modulating actin cytoskeleton dynamics in part independently of its catalytic activity, for example by protein-interaction-dependent mode of action. These observations were in slight discordance with the results reported by Fiordalisi et al., where colon cancer SW480 cells transiently expressing wtPRL-1 had elevated level of GTP-bound RhoA and RhoC and decreased level of Rac1, whereas the level of Cdc42 was not changed with respect to the control (Fiordalisi et al., 2006). This discrepancy could result from spatiotemporal oscillations in the activities of GTPases (Al-Aidaroos and Zeng, 2010).

It was also demonstrated that PRL-1 could promote cell motility and invasiveness by increasing expression of the matrix metalloproteinases (MMPs) (Luo et al., 2009). MMPs are hydrolytic enzymes which have an important role in facilitating an escape of cancer cells from the primary sites and promoting cell invasion through digestion of the extracellular matrix (ECM)

(Andl, 2010; Chambers et al., 2002). HEK293 cells stable overexpressing PRL-1 had elevated levels of activated Src, judged by the level of phosphorylated Tyr416 comparing to the control cells (Luo et al., 2009). Moreover, the levels of activated signaling components downstream of Src (p130Cas, focal adhesion kinase (FAK) and phosphorylated Erk1/2 (pErk1/2)) were found to be elevated upon PRL-1 expression. The increase in the active levels of the MMP2/9 was found to be dependent on the expression of PRL-1. Moreover, the stimulatory effect of PRL-1 on cell migration was diminished when activity or total protein levels of the MMP2/9 were reduced. In addition, the levels of the MMPs were increased as a result of the Src/Erk1/2 mediated upregulation of the key regulatory transcription factors AP1 and Sp1. Moreover, Luo et al. demonstrated that high levels of PRL-1 in invasive lung cancer cell lines A549 and H1299 promote cell invasiveness by operating through the aforementioned mechanisms.

Wang and Lazo profiled the role of PRL-2 in migration and invasion in lung cancer cells A549 and observed that knockdown of PRL-2 by shRNA decreased cell migration and invasion (Wang and Lazo, 2012). Analysis of the potential mechanism showed that there were no changes in, total or phosphorylated, levels of p53, Src, Csk and Akt. In addition, reduction of PRL-2 protein level resulted in decreased levels of scaffold protein p130Cas and vinculin, whereas levels of pErk1/2 and ezrin phosphorylated on Tyr146 were upregulated, which was comparable to the effect observed after reduction of PRL-1 in the same cell line (Achiwa and Lazo, 2007). Intriguingly, suppression of PRL-2 did not affect the level of ezrin phosphorylation on Thr567 and knockdown of PRL-1 did not impose any changes on the phosphorylation status of ezrin (Wang and Lazo, 2012).

1.5.10 Regulation of PRL-3

Overexpression of PRL-3 in colorectal cancer metastases to liver can in part be attributed to an increase in the PRL-3 gene copy number (Stephens et al., 2005). It was noted that the PRL-3 gene copy number was elevated in liver metastasis compared to the primary cancer sites, and in primary colorectal cancer with metastatic lesions to liver as opposed to metastasis-free states, and also in some myeloma cells (Al-Adaroos and Zeng, 2010). However, increase in gene copy number does not account for the observed overexpression of PRL-3 in other types of cancer (Bessette et al., 2008). Moreover, it was observed that increase in gene copy number does not always correlate with the elevated levels of PRL-3 mRNA and that in myeloma cells no connection between gene copy number and PRL-3 mRNA expression could be established (Al-Aidaroos and Zeng, 2010; Fagerli et al., 2008; Saha et al., 2001). Moreover, additional mechanisms of PRL-3 regulation have been discovered and involve regulation at the transcriptional and translational levels, as well as regulation via proteosomal degradation (Al-Aidaroos and Zeng, 2010; Rios et al., 2013). PRL-3 was found to be a transcriptional target of p53 in mouse and human cells, and a putative p53 binding site was identified in the PRL-3 gene (Basak et al., 2008; Rios et al., 2013). Interestingly, expression of exogenous PRL-3 in mouse embryonic fibroblasts (MEFs) caused cell cycle arrest in G1 phase in a manner similar to p53-induced arrest. Knockdown of PRL-3 by shRNA approach produced the same effect, which was additionally shown to depend on the intact p53 pathway. These observations were in contrast to the observed effect of PRL-3 overexpression on enhancing cell proliferation in various cancer cells and also in contrast to the observed normal development of the PRL-3 knockout mice (Zimmerman et al., 2013).

Concomitantly, Basak et al. demonstrated that PRL-3 was not able to induce cell cycle arrest in some human cancer cell lines, thus suggesting that the effect on cell cycle progression may be cell type specific (Basak et al., 2008). Overall, in intact MEFs, basal levels of PRL-3 could regulate progression through cell cycle. Increased levels of PRL-3 during cell cycle arrest, as a direct transcriptional target of p53, were accompanied by inactivation of PI3K/Akt pathway. In addition, in intact MEFs, knockdown of PRL-3 led to increase in the protein levels of p53. This effect was mediated by increased levels of a positive p53 regulator called $p19^{Arf}$, which was induced by the activation of p38 MAPK. Additionally, PRL-3 was shown to be involved in the negative regulation of protein levels of p53 by ubiquitin-mediated degradation, suggesting that in metastatic cells with elevated levels of PRL-3 these feedback loops are disrupted and/or different regulatory mechanism might apply (Al-Aidaroos and Zeng, 2010).

Furthermore, transcription of PRL-3 was shown to be activated by vascular endothelial growth factor (VEGF), involving the transcription factor called myocyte enhancer factor (MEF) 2C, in human umbilical vein endothelial cells (HUVEC), although the increase in the PRL-3 transcript was not reflected at the protein level (Rios et al., 2013; Xu, Cao et al., 2011). Both, VEGF and MEF2C have been described as important factors in angiogenesis, suggesting that the existence of a VEGF/MEF2C/PRL-3 pathway could exert similar angiogenesis-stimulatory effects.

Transforming growth factor (TGF) β suppressed expression of PRL-3 in a mothers against decapentaplegic homologue (Smad) 3-dependent way (Jiang, Liu, Rajput et al., 2011). For TGFβ, both metastasis inhibitory and stimulatory modes of action were reported (Wakefield and Roberts, 2002). Moreover, it

was demonstrated that expression of wtPRL-3 promoted survival of colon cancer cells by activating PI3K/Akt pathway, and either in the presence of the PI3K inhibitor, by using negative mutant of Akt or by knocking down expression of PRL-3, the PRL-3-mediated suppressive effect on apoptosis was reversed (Jiang, Liu, Rajput et al., 2011). This study did not disclose the mechanism behind the PRL-3-mediated increase in the level of activated Akt, but it suggested that the effect could involve stimulation of Akt-activating kinases. Considering that TGFβ mutations occur in up to 50% of colon cancer patients, this is likely to account for the increased expression of PRL-3 in colon cancer (Rios et al., 2013).

Transcription factor Snail was demonstrated to be a key factor in PRL-3-mediated progression of colorectal cancer to metastasis, albeit it is not known if there is a direct interaction between these two proteins (Rios et al., 2013; Zheng, Meng, Gao et al., 2011). Snail is a well recognised EMT-driving transcription factor, repressing expression of epithelial markers and promoting cell motility and invasiveness (Andl, 2010; Thiery and Sleeman, 2006). Moreover, colon cancer cells overexpressing PRL-3 showed upregulated levels of Snail in contrast to the controls where PRL-3 expression was knocked down (Liu, Zhou et al., 2009). Three potential Snail binding sites were identified in the promoter region of the PRL-3 gene, and after introducing mutations in these sites or when expression of Snail was knocked down, the transcriptional activity of the PRL-3 promoter and the protein levels of PRL-3 were reduced, respectively (Zheng, Meng, Gao et al., 2011). In addition, overexpression of Snail promoted cell proliferation, migration and colony formation of human colon cancer cells, whereas Snail knockdown had the reversed effect. In the context of PRL-3 knockdown, proliferation of colon cancer cells increased when

protein levels of Snail were upregulated. The link between PRL-3 and Snail, in the light of the demonstrated stimulatory effect both of them have on proliferation of human colon cancer cells, requires further analysis.

Expression of PRL-3 at the protein level can be negatively regulated by polyC-RNA-binding protein (PCBP) 1, which abrogates incorporation of PRL-3 mRNA into polyribosomes (Al-Aidaroos and Zeng, 2010; Wang et al., 2010a). PCBP1 was found to inhibit expression of PRL-3 by binding to the GC-rich region at the 5'-untranslated (UTR) of PRL-3 mRNA and it also suppressed the activity of Akt, without affecting PRL-3 mRNA levels. Moreover, it was shown that in human cancer samples of various origins the protein levels of PRL-3 and PCBP1 were inversely correlated. Clinical significance of the findings of Wang et al. was additionally validated by demonstrating that injection of human colon cancer cells with knockdown PCBP1 enhanced tumour development in mouse model (Wang et al., 2010a).

Furthermore, FKBP38 was identified as a PRL-3 interacting protein in a Y2H screen, and was established as a negative regulator of PRL-3 at the protein level (Choi, Min et al., 2011). It was demonstrated that FKBP38 promoted proteosomal degradation of PRL-3 and also inhibited PRL-3-mediated suppression of p53 activity and led to reduction in cell proliferation.

Regulation of PRL-3 by Src-mediated phosphorylation was reported recently (Fiordalisi et al., 2013). PRL-3 was found to be phosphorylated on Tyr53 *in vitro*. Moreover, phosphorylation of endogenous PRL-3 in colon cancer SW480 cells was also shown. In addition, modification of Tyr53 by phosphorylation proved required for PRL-3-mediated cell motility, invasion and activation of RhoC. Although no difference in the subcellular localisation was observed, the

authors speculated that phosphorylation could induce delicate changes in targeting PRL-3 to distinct membrane regions, albeit implications of the established PRL-3-Src link warrant further investigation.

1.5.11 Interacting proteins and substrates of PRL-3

There are several suggested protein and non-protein interacting partners of PRL-3 (integrin α1, keratin 8 (KRT8), elongation factor (EF) 2, cadherin 22 (CDH22), nucleolin, stathmin, peptidyl prolyl *cis/trans* isomerase FK506 binding protein (BP) 38 (FKBP38), phosphatidylinositol (4,5)-bisphosphate ($PI(4,5)P_2$), histone deacetylase (HDAC) 4, ADP-ribosylation factor (Arf) 1, and Src), whereas to date only few substrates are suggested (integrin β1 and ezrin) (Al-Aidaroos and Zeng, 2010; Choi, Min et al., 2011, Fiordalisi et al., 2013, Forte et al., 2008, Krndija et al., 2012, Liu, Zhou et al., 2009, McParland, Varsano et al., 2011, Mizuuchi et al., 2009, Orsatti et al., 2009b, Peng, Jin et al., 2006, Semba, Mizuuchi et al., 2010, Tian et al., 2012, Zheng et al., 2010; Zhou et al., 2011). Most of the PRL-3 interacting partners can be linked to the PRL-3-mediated role in cell migration and invasion.

Integrin α1 was identified as PRL-3 interacting protein by using Y2H approach and the interaction was also confirmed *in vivo* by co-immunoprecipitation from African green monkey kidney COS-7 cells transiently co-expressing both proteins (Al-Aidaroos and Zeng, 2010; Peng, Jin et al., 2006; Rios et al., 2013). Integrins are heterodimeric transmembrane receptors composed of α and β subunits, and containing extracellular domain, which binds to the ECM, transmembrane and cytoplasmic domains (Peng, Jin et al., 2006; Srichai and Zent, 2010). They are implicated in signaling originated by binding of the ECM components to the integrin receptors, which triggers recruitment of diverse

signaling molecules to the sites of FACs at the sites of cell-ECM contacts (Bessette et al., 2008). Integrins are important in regulating cell-cell and cell-ECM contacts, thus facilitating regulation of cell adhesion or motility (Plantefaber and Hynes, 1989). Unsurprisingly, aberrations in integrin-mediated signaling are a common feature of cancer cells (Al-Adaroos and Zeng, 2010; Bessette et al., 2008; Tian et al., 2012; Varner and Cheresh, 1996). Integrin α1, unlike integrin β1, does not contain any residues in its cytoplasmic domain that could be dephosphorylated by PRL-3. It was demonstrated that HEK293 cells stably overexpressing PRL-3 had reduced level of tyrosine-phosphorylated integrin β1, although, at the time, no direct interaction between PRL-3 and integrin β1 could be detected (Peng, Jin et al., 2006). Nevertheless, Tian et al. demonstrated that Tyr783 of integrin β1 was dephosphorylated by PRL-3 *in vitro* and also in cells after ectopic expression of PRL-3 (Tian et al., 2012). In the presence of inhibitors of PRL-3, the phosphorylation of this residue was enhanced. In addition, integrin α1 was shown to act as a negative regulator of the PRL-3-integrin β1 interaction because knockdown of integrin α1 by siRNA resulted in a decreased amount of pTyr783 of integrin β1 detected.

CDH22, a member of the cadherin family, was also identified in a Y2H screen and the interaction was further confirmed in colorectal cancer cells by co-immunoprecipitation (Liu, Zhou et al., 2009). Liu et al. noted that stable overexpression of PRL-3 in human colon cancer cells SW480 caused decrease in the levels of CDH22, epithelial (E) cadherin, which is a key component of the adherens junction complexes, and cytokeratin, thus illustrating the EMT-promoting role of PRL-3 (Bessette et al., 2008; Liu, Zhou et al., 2009). Moreover, PRL-3 overexpressing cells demonstrated elevated levels of Snail, a transcription factor already linked to downregulation of E-cadherin expression

(Liu et al., 2005). The PRL-3-CDH22 association was suggested to affect phosphorylation-dependent degradation of CDH22, and further contributed to strengthening the contribution of PRL-3 to EMT linked events (Liu, Zhou et al., 2009).

KRT8 was found to interact with PRL-3 by using proteomics approach in which human colon cancer SW480 cells transfected with wtPRL-3 and its C104S mutant were assessed for difference in the phosphorylation pattern by two-dimensional (2D) SDS-PAGE (Mizuuchi et al., 2009). KRT8 is a member of the intermediate filaments network (Khapare, Kundu, Sehgal et al., 2012) which is important for maintaining cell shape (Garrord and Chidgey, 2007). Furthermore, endogenous KRT8 was also shown to interact with ectopic C104S mutant of PRL-3 (Mizuuchi et al., 2009). KRT8 with decreased phosphorylation level at residues S73 and S431 was detected in human tissue samples from primary colon cancer and liver metastases. It was suggested that PRL-3 could be involved in the regulation of phosphorylation-dependent shuttling of KRT8 between soluble cytoplasmic population and polymerized filaments whose disintegration is stimulated by phosphorylation. PRL-3 could disrupt the balance by dephosphorylating S73 and S431 in KRT8, thus contributing to promotion of cell motility as a result of irregularities in the regulation of dis-assembly of intermediate filaments. To further support that PRL-3-mediated progression to metastasis could be helped by reducing phosphorylation of KRT8, it was also reported that decreased phosphorylation of KRT8 positively correlated with progression of oral squamous cell carcinoma (Alam et al., 2011). Furthermore, a recent study showed that loss of plakophilin3 was accompanied by an increase in the expression of PRL-3 and to the concomitant reduction in phosphorylated KRT8 (Khapare, Kundu, Sehgal et al., 2012).

Plakophilin3 is a desmosomal protein. Desmosomes represent adhesion complexes linked to intermediate filaments (Andl, 2010; Bass-Zubek et al., 2009; Garrod and Chidgey, 2007). Knockdown of plakophilin3 led to a decrease in cell-cell adhesion and increase in cell motility due to reduced desmosomes, in agreement with the earlier studies (Andl, 2010; Bass-Zubek et al., 2009; Kundu et al., 2008).

Stathmin/oncoprotein 18 was identified as a novel PRL-3 target in a proteomics study, which also reported a list of 39 proteins associated with PRL-3 (Zheng et al., 2009). In addition, Western blot analysis on tissue samples derived from colon cancer patients with or without metastasis, showed elevated levels of stathmin in primary tumours compared to the normal counterparts, but also increased stathmin expression in cases with metastasis as opposed to the samples of the primary cancer without metastasis. It was demonstrated that the interaction of stathmin and PRL-3 in colon cancer cells led to destabilization of microtubules. The mechanism was not clear, but a hypothesis was offered according to which PRL-3 could be involved in dephosphorylation of the key regulatory serine residues of stathmin.

Another proteomics-based investigation of PRL-3-mediated role in tumourigenesis was conducted recently (Chu et al., 2011). Applying 2D-electrophoresis based comparison of human colon cancer LoVo cells expressing wtPRL-3 with the control cells, it was established that 17 proteins were up- and ten downregulated upon PRL-3 expression. In addition, eight of these up- or downregulated proteins were shown to have metastasis-promoting functions. Translationally controlled tumour protein (TCTP) was shown to be elevated at the levels of mRNA and protein. TCTP is known to be upregulated in a number of cancers (Ma et al., 2010). Upon siRNA-mediated knockdown of

TCTP, proliferative, migratory and invasive abilities of LoVo cells expressing PRL-3 were diminished, but the mechanism of TCTP in the background of PRL-3 is yet to be disclosed (Chu et al., 2011).

Nucleolin was identified as a novel PRL-3 interacting protein by applying MS analysis of immunoprecipitates from colon cancer SW480 cells expressing wtPRL-3 and its catalytically inactive mutant C104S (Semba, Mizuuchi et al., 2010). Nucleolin is a known nucleolar protein (Mongelard and Bouvet, 2007). The discrepancy between the previously reported subcellular distribution of PRL-3 and nucleolar protein being identified as a putative interacting partner of PRL-3 was explained by assessing the distribution pattern of nucleolin after ectopic expression of PRL-3 (Semba, Mizuuchi et al., 2010). Interestingly, only overexpression of wtPRL-3 was able to show an increase in the nucleolar accumulation of nucleolin, suggesting that PRL-3 could have role in regulating phosphorylation status and dis-assembly of nucleolin, whereas the inactive C104S mutant just trapped nucleolin in the cytoplasm. Nucleolin was known to be phosphorylated by CK2, which promotes its degradation. The mechanism of the role nucleolin might have in promoting proliferation of colon cancer cells was hypothesized based on its previously reported functions (Erard et al., 1988; Semba, Mizuuchi et al., 2010). Intriguingly, nucleolin was reported to increase stability of B-cell lymphoma (Bcl) 2 mRNA, thus promoting cancer development (Sengupta, Bandyopadhyay et al., 2004).

Recently, it was demonstrated that PRL-3 could act as a phosphatidylinositol-5-phosphatase by dephosphorylating PI(4,5)P_2, and also phosphatidylinositol (3,4,5)-trisphosphate (PI(3,4,5)P_3) to a smaller extent, *in vitro* (McParland, Varsano et al., 2011). By using HEK293 stably transfected with wtPRL-3 and its different variants (D72A, C104S and A111S) in a cell migration and wound

healing assay, it was demonstrated that not only wtPRL-3 but also its D72A variant exhibited migratory and invasive abilities. No *in vivo* activity of PRL-3 against $PI(4,5)P_2$, or phosphoinositides in general, has yet been demonstrated (Rios et al., 2013). The consequences of the PRL-3-mediated downregulation in the levels of $PI(4,5)P_2$ are only hypothesised, but there is a strong body of literature to support that $PI(4,5)P_2$ levels need to be under rigorous control and that aberrations of the regulatory mechanisms could contribute to cancer development (Bunney and Katan, 2010; McParland, Varsano et al., 2011; Mitra et al., 2005; Raucher et al., 2000; Redfern et al., 2008).

Ezrin was suggested as a substrate for PRL-3 after it was demonstrated that HCT116 colon cancer cells stably expressing only wtPRL-3 and not its catalytically deficient mutants sported distinct phosphorylation pattern (Forte et al., 2008). Ezrin was shown to be dephosphorylated on Thr567 by PRL-3. Ezrin belongs to the ezrin/radixin/moesin (ERM) family of proteins (Neisch and Fehon, 2011). The members of the ERM family act as linkers between plasma membrane and actin cytoskeleton (Zhu et al., 2010). They bind to filamentous (F) actin and plasma membrane via their C- and N-terminal associated domains (C-ERMAD and N-ERMAD), respectively. Ezrin is maintained in an inactive (“closed”) conformation through the intramolecular association of the C- and N-termini, which prevents its association with F-actin and reduces the affinity for the plasma membrane. The active form of ezrin is important for signaling affecting cell adhesion or motility (Barret et al., 2000; Fievet et al., 2004; Rios et al., 2013). The switch between the closed and open conformation of ezrin is regulated by the phosphorylation status of Thr567, with phosphorylation disrupting intramolecular association (Forte et al., 2008). Overexpression of PRL-3 would therefore increase the content of inactive

dephosphorylated ezrin. This, in turn, might be reflected on the signaling components affected by ezrin, promoting cancer development by, for example, modulating the activity of Src or of the small GTPases Rho and Rac, for which the changes in the activities in the context of the exogenous PRL-3 expression had been reported (Fiordalisi et al., 2006; Forte et al., 2008; Liang et al., 2008; Wang et al., 2007a). In the light of the recently reported ability of PRL-3 to dephosphorylate $PI(4,5)P_2$ *in vitro*, PRL-3 would affect ezrin-mediated signaling by both, depleting $PI(4,5)P_2$ from the plasma membrane and by dephosphorylating ezrin on Thr567, thus affecting two requirements for ezrin to assume active conformation at the plasma membrane (Fievet et al., 2004; Rios et al., 2013).

A later study by the same group (Orsatti et al., 2009b) demonstrated, by using 2D SDS-PAGE, that phosphorylation levels of ezrin and EF-2 were changed in HCT116 colon cancer cells overexpressing PRL-3. Interestingly, EF-2 was also identified among 39 proteins associated with PRL-3 in a proteomics-based study of PRL-3 knockdown in SW480 colon cancer cells (Zheng et al., 2009). EF-2 is involved in mediating peptide chain elongation step during protein synthesis (Wang and Proud, 2002) and it is in its active form when dephosphorylated (Orsatti et al., 2009b; Ryazanov, 2002). Moreover, its phosphorylation on Thr56 acts as a general regulator of protein synthesis because phosphorylated form diminished it from binding to the ribosome (Orsatti et al., 2009b). No direct interaction between PRL-3 and EF-2 has been demonstrated (Rios et al., 2013).

Recently, Arf1 was identified as a novel interacting partner of PRL-3, and it was shown to colocalize with PRL-3 at the Golgi and endosomes (Krndija et al., 2012). Moreover, a motif within PRL-3 (MKYE) absolutely essential for its

interaction with Arf1 was identified. The Arf family of proteins belongs to the small GTPases implicated in membrane trafficking and their unique feature comparing to the other representatives of the small GTPases is an N-terminal helix detrimental for their binding to the plasma membrane (de Curtis and Meldolesi, 2012). Knockdown of Arf1 and structurally and functionally related Arf3 diminished PRL-3-mediated promotion of cell migration (Krndija et al., 2012). In addition, expression of wtPRL-3, but not of its catalytically inactive (C104S) or prenylation deficient (ΔCAAX) mutants increased the level of the active GTP-bound form of Arf1. Dynamic changes at the plasma membrane are coordinated by the activities of small GTPases (de Curtis and Meldolesi, 2012). PRL-3 was demonstrated to increase turnover of integrin α5 in a manner dependent on the activity of Arf1, thus stimulating cell migration (Krndija et al., 2012).

1.5.12 PRL-3 in signaling

Signaling pathways mediated by PRL-3 promote cancer progression and metastasis by enhancing cell proliferation, migration and invasion, mediating EMT switch, potentiating angiogenesis or inhibiting apoptosis (Al-Aidaroos and Zeng, 2010; Bessette et al., 2007; Bessette et al., 2008). The key signaling effects exerted by PRL-3 will be discussed.

In the light of the previously acknowledged expression of the PRL-3 transcript in skeletal and cardiac muscle (Zeng et al., 1998), one early report addressed its possible physiological role and reported that it mediated Angiotensin-II (AngII) signaling (Bessette et al., 2007; Matter et al., 2001). AngII is involved in hyperthrophy of cardiomyocytes (Matter et al., 2001). It was demonstrated that HEK293 cells overexpressing wtPRL-3 had lower levels of phosphorylated

p130Cas, which is phosphorylated after AngII stimulation-mediated release of intracellular calcium (Ca^{2+}). p130Cas is a scaffold protein with diverse functional domains, important for regulating cell proliferation, motility and apoptosis (Bessette et al., 2007; O'Neill et al., 2000; Wei et al., 2004). It is phosphorylated by FAK and Src in response to various stimuli, and increase in the phosphorylated level of p130Cas positively correlates with cell migration and invasion (Bessette et al., 2007; Cabodi et al., 2006; Defilippi et al., 2006). The reduction in the phosphorylated level of p130Cas could not be attributed to the direct action of PRL-3 on p130Cas. Decrease in the phosphorylated and total levels of p130Cas was also observed after depletion of PRL-3 by siRNA, with more striking effect when all three PRLs were knockdown (Daouti et al., 2008). Moreover, thienopyridone, a small molecule identified by chemical library screening, demonstrated selective inhibition of the PRLs, compared to 11 other phosphatases belonging to different classes tested (Al-Aidaroos and Zeng, 2010; Daouti et al., 2008). Treatment with thienopyridone caused inhibition of anchorage-independent growth of tumour cells and induced anoikis. Anoikis is a type of apoptosis caused by disrupted cell-ECM interactions and is caused by proteolytic cleavage of p130Cas (Daouti et al., 2008; Wei et al., 2004). Interestingly, this effect of PRL-3 inhibitor involved a novel mechanism independent of the p53 pathway.

PRL-3 was found to stimulate EMT (Al-Aidaroos et al., 2010; Wang et al., 2007a). The EMT is characterised by loss of adhesion of epithelial cells through disruption of the assembly and stability of the adherens junction complexes, thus leading to an increase in motility and gain of ability to extravagate into the surrounding tissue (Andl, 2010; Thiery and Sleeman, 2006). CHO cells stably expressing wtPRL-3 exhibited decreased levels of vinculin and, total and

phosphorylated, paxillin. Paxillin is a scaffold protein which binds to FAK, which is an essential component in promoting integrin and growth factor mediated signaling (Andl, 2010; Yano et al., 2004). Vinculin is an adapter protein localised to the sites of cell-matrix and cell-cell contacts and its loss reduces cell adhesion and increased motility (Carisey & Ballestrem, 2011). In addition, in colorectal cancer DLD-1 cells stably expressing wtPRL-3, the expression levels of epithelial markers E-cadherin, integrin β3 and γ-catenin were reduced, and the levels of mesenchymal markers Snail and fibronectin were increased (Wang et al., 2007a). E-cadherin is known to interact with a number of growth factor receptors preventing their prolonged activation and to suppress expression of the proteolytic MMPs (Andl, 2010). Moreover, the cytoplasmic domain of cadherins is engaged in interactions with β-catenin, which links components of adherens junctions to actin cytoskeleton. PRL-3 was shown to promote EMT transition by upregulating the PI3K/Akt pathway (Wang et al., 2007a), which is known to be an important factor in promoting cell proliferation and anti-apoptosis (Cully et al., 2006). The study of Wang et al. suggested that PRL-3 functions upstream of PI3K, because its effect on promoting EMT was not observed after treatment with the inhibitor of PI3K. In addition, PRL-3 exerted its effect by downregulating PTEN at the post-translational level. The mechanism could involve β-catenin and MAGI-2 dependent localisation of PTEN to the plasma membrane (Subauste et al., 2005). Furthermore, in CHO and DLD-1 cells expressing wtPRL-3, the expression of GTPases RhoA and Rac1 was downregulated (Wang et al., 2007a). It was suggested that this could decrease the stability of adherens junctions, which would agree with the previously reported observation of the effect of RhoA on the increase of the amount of vinculin in focal adhesions in Madin-Darby canine kidney (MDCK)

cells (Kotani et al., 1997). However, in the other study it was demonstrated that human colon cancer SW480 cells expressing wtPRL-3 had elevated level of active forms of RhoA and RhoC, reduced level of Rac1, while Cdc42 remained unaffected (Fiordalisi et al., 2006). Moreover, the ability of PRL-3 to promote cell migration and invasion was abrogated in the presence of the upstream kinase activator of RhoA called Rho-coiled coil kinase (ROCK), suggesting that PRL-3 acts upstream of RhoA in a manner dependent on the activity of ROCK. The opposite effect of PRL-3 expression on the activation of Rho GTPases observed in these two studies could reflect the involvement of PRL-3 in focal adhesion turnover by promoting both their disassembly and assembly (Bessette et al., 2008; Fiordalisi et al., 2006; Wang et al., 2007a). Interestingly, Fiordalisi et al. dissected features dependent on the phosphatase activity of PRL-3, such as invasion, motility and regulation of GTPases, from activation of the transcription factor called serum response element (SRE). This was only partially reduced in case of the catalytically inefficient mutants of PRL-3 (D72A and C104A). The authors speculated that the latter could be the result of the substrate-trapping effect of D72A and C104A mutants, which might preserve certain substrate-interactions sufficiently long to elicit a particular effect. In the light of a complex crosstalk existing between integrins and cadherins, it was reported that activation of integrins led to downregulation of E-cadherin and to activation of RhoA and Rac1 GTPases, perturbing cadherin-based cell-cell contacts, which is in accordance with the results obtained in the study of Fiordalisi et al. (Fiordalisi et al., 2006; Thiery and Sleeman, 2006).

Recently, it was reported that expression of PRL-3 at the level of mRNA and protein was high in a set of lung cancer cell lines tested (Jian et al., 2012). In accordance with this observation, knockdown of PRL-3 by siRNA caused

changes in cell morphology, yielding reduction in the extent of protrusive membrane structures, and reduced the activity of RhoA. In addition, knockdown of PRL-3 and downregulation of active RhoA led to reduction in levels of the protein called mammalian homologue of *Drosophila melanogaster* diaphanous (mDia1). mDia1 regulates polymerisation of actin after it was released from the autoinhibitory conformation by binding of the active RhoA to its RhoA binding domain (RBD) (Brandt et al., 2007; Jian et al., 2012). This study provided another demonstration that PRL-3 exerts its effect on cell motility and invasiveness by affecting components that are important for the maintenance and organization of the actin cytoskeleton.

PRL-3 was linked to integrin/Src signaling (Peng et al., 2009; Rios et al., 2013; Tian et al., 2012). Important downstream effectors of integrin signaling are FAK and Src (Bessette et al., 2008; Thiery and Sleeman, 2006). In addition, Src was reported to promote EMT by internalizing E-cadherin in a manner dependent on the phosphorylation by FAK, whereas FAK recruits scaffold proteins p130Cas and paxillin in focal adhesions (Andl, 2010; Thiery and Sleeman, 2006). Moreover, FAK-Src association is important in regulating activities of GAPs and GEFs, which catalyze inactivation of GTPases by hydrolyzing GTP to GDP and activation of GTPases by releasing bound GDP and enabling binding of GTP, respectively (Al-Aidaroos et al., 2010; Fiordalisi et al., 2006). It was demonstrated that HEK293 cells stably expressing PRL-3 had reduced levels of Csk, which acts as a negative regulator of Src by phosphorylating Tyr527 residue in the tail region of Src (Bessette et al., 2008; Liang et al., 2008; Rios et al., 2013). Upon expression of PRL-3, the pTyr527 levels were elevated, thus indicating increased activity of Src (Liang et al., 2007). However, the reduction in the level of pTyr527 of Src was not the result of direct dephosphorylation by

PRL-3 (Liang et al., 2008). Instead, PRL-3 expressing cells had reduced levels of Csk as a result of global reduction of protein synthesis caused by increase in phosphorylation of eukaryotic initiation factor (eIF) 2α. The general suppression of translation was confirmed in colon cancer SW480 cells naturally expressing PRL-3. Moreover, PRL-3 was shown to lead to activation of Erk1/2 in colon cancer LoVo cells through integrin β1 (Peng et al., 2009). This was in agreement with an earlier study, where expression of PRL-3 downregulated phosphorylation of integrin β1 and upregulated activation of Erk1/2 (Peng, Jin et al., 2006). Knockdown of integrin β1 by siRNA diminished PRL-3-mediated increase in the level of activated Erk1/2 and also abrogated PRL-3-triggered motility and invasion in cells and in mouse model (Peng et al., 2009). It was demonstrated that PRL-3 promotes cell invasion by contributing to aberrant regulation of MMP2 through decreased expression of the MMP inhibitor called TIMP2. This could provide another link of PRL-3-mediated EMT switch, because it was reported that E-cadherin suppresses expression of MMP-2, reducing invasive ability of the cells (Andl et al., 2010). Signaling pathways affected by PRL-3 are shown in Figure 9.

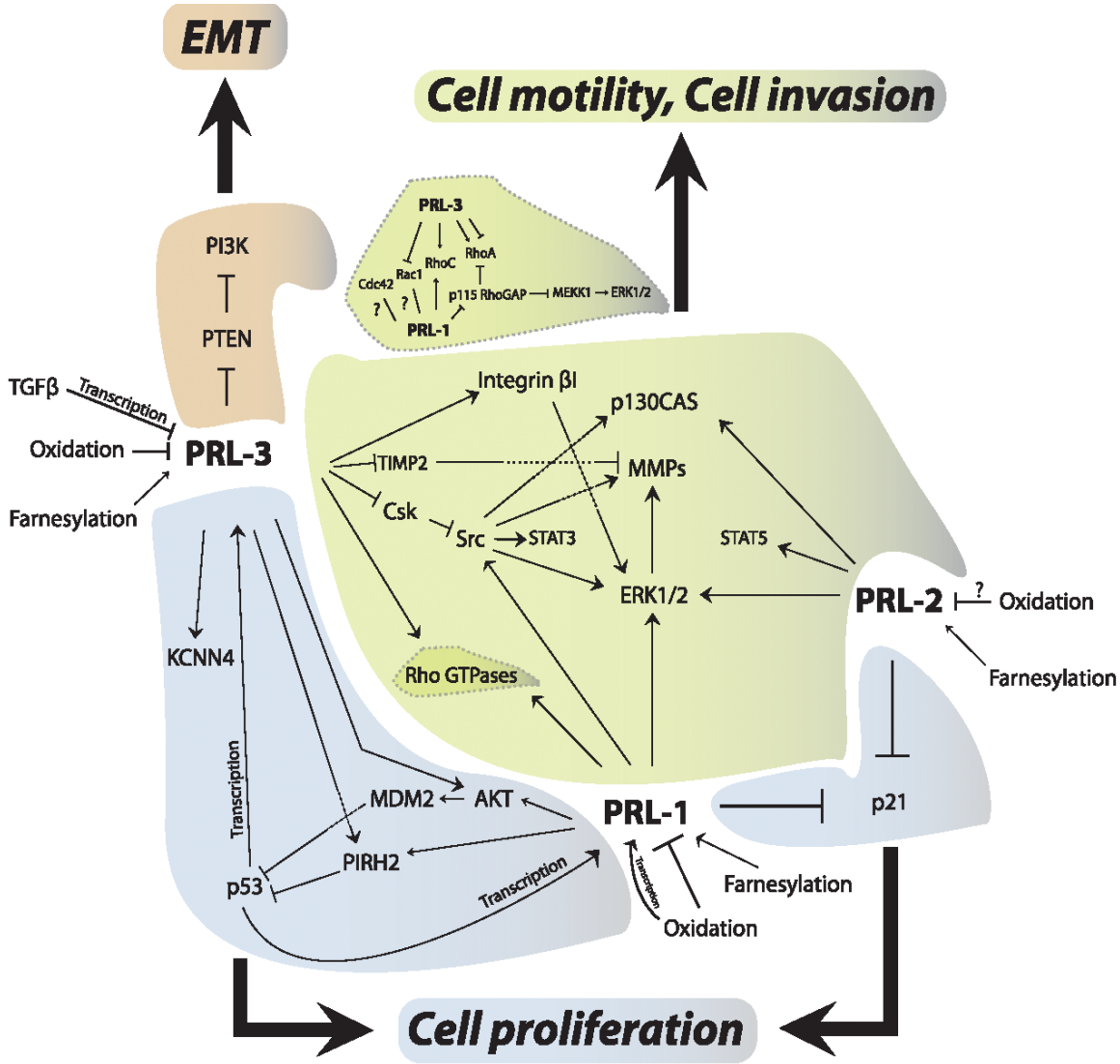


Figure 9: Summary of the signaling pathways affected by the PRLs and the outcome in cell migration and proliferation. Arrows indicate positive regulation; crossed lines indicate negative regulation and question marks indicate either not yet understood or not studied processes. With permission from Rios, P., Li, X. & Köhn, M. Molecular mechanisms of the PRL phosphatases. *FEBS J.* **280**, 505-524 (2013).

1.6 Challenges in identifying PTP substrates and interacting partners

Considering the importance of protein phosphorylation in regulating diverse cellular processes, identifying interacting partners of the PPs could unarguably contribute to establishing a link between a particular phosphatase and its contribution to a signaling outcome. It has been already mentioned that substrate-trapping mutants of the PTPs are commonly used in studying interactions of the PTPs (Zhang, 2003a). The most commonly applied substrate trapping variants of the PTPs involve mutations of the residues which are essential for the efficient binding and hydrolysis of a substrate, such as the catalytic cysteine of the P-loop or the general acid/base aspartate residue of the WPD-loop (Blanchetot et al., 2005; Zhang, 2003a). Cysteine of the P-loop is commonly substituted for a serine and aspartate of the WPD-loop for an alanine. Notwithstanding, the double mutant simultaneously bearing both of these mutations may not have enhanced substrate-trapping ability, as witnessed for PTP1B. In spite of resulting in catalytically deficient enzymes which would be expected to stabilize even transient interactions, this approach has not been widely efficient in elucidating substrate portfolio of the PTPs (Flint et al., 1997; Zhang, 2003a). In addition to the mutations of the catalytically essential cysteine and aspartate residues, further mutations which would reduce the catalytic efficiency of the PTPs, without having negative effect on their substrate-recognition ability, were also tried (Blanchetot et al., 2005), for example mutation of the conserved glutamine residue of the signature motif to an alanine or mutation of the catalytic cysteine to an aspartate. Substrate-trapping mutants of the PTPs were also combined with other approaches, for example yeast two-hybrid (Y2H) or MS (Tiganis & Bennett, 2007). In general, substrate-trapping PTP mutants still have considerable caveats, for example (i)

they are restricted to the catalytically important residues, (ii) validating the interactions within the cell or (iii) difficulties in identifying less abundant substrates. Screening various mutants for different phosphatases can be regarded as laborious, so more efficient approaches in investigating the interactions of the PTPs are needed.

1.7 Expanded genetic code methodology

Expanded genetic code methodology constitutes a mighty tool for probing even weak and transient protein interactions by using genetically encoded photo-cross-linkable amino acids, thus bringing advantage over more conventional approaches which rely on Y2H screens or chemical cross-linking reagents (Ai, Shen, et al., 2011). Modification of the amino acid building blocks enable manipulation of protein structure, which not only provide a mean to gain insight into functional specificities of proteins or probe for protein interactions and associated biological pathways, but also create proteins with novel properties (Wang et al., 2009; Xie and Schultz, 2005). Development of an expanded genetic code provides a useful and powerful toolkit for studying proteins in prokaryotes and eukaryotes *in vitro* and *in vivo* (Xie and Schultz, 2006). This method enables an unnatural amino acid to be genetically encoded into protein concomitantly with the course of protein translation with high fidelity and efficiency. Over 80 unnatural amino acids have been incorporated into proteins to date in prokaryotes and eukaryotes, with diverse side-chain moieties as well as photo-excitable, redox-reactive and photocaged amino acids, clearly providing a sophisticated tool for controlled manipulation of structural and functional properties of proteins (Chen et al., 2007; Lemke et al., 2007; Liu and Schultz, 2010; Strømgaard et al., 2004; Wu et al., 2004; Zheng et al., 2013). These include, for example, amino acids which can facilitate investigation of the

PTMs or which contain chemical moieties that are modifiable *in vitro* or in cells, thus expanding the range of reactive side chain groups in amino acid. Moreover, these amino acids can serve as a platform for selective chemical modifications thus giving rise to, for example, fluorescent, photo-cleavable, NMR or infrared (IR) probes or redox-sensitive reagents, or they can facilitate investigation of structural and conformational properties of proteins by X-ray (Liu and Schultz, 2010; Strømgaard et al., 2004; Xie and Schultz, 2006).

There are several methods to introduce unnatural amino acids into proteins (Chin et al., 2003; Hendrickson et al., 2004; Liu and Schultz, 2010; Muir, 2003; Sakamoto et al., 2002; Suchanek et al., 2005; Strømgaard et al., 2004; Thibodeaux et al., 2010; Wang et al, 2009; Xie and Schultz, 2006), but in general they are not as advantageous as co-translational introduction of an unnatural amino acid. These additional approaches include chemical derivatisations of the modifiable amino acid side chains in proteins (thiol moieties of cysteines or ε-amino groups of lysines), solid-phase peptide synthesis (SPPS), ligation approaches using protein semi-synthesis, residue-specific insertion of an unnatural amino acid by using bacterial strains auxotrophic for this amino acid, site-specific incorporation of an unnatural amino acid by chemically acylating a suppressor tRNA and then adding it either to the cell extracts or to the *Xenopus* oocytes and co-translational insertion of an unnatural amino acid. Since exploting those altarnative approaches was not the subject of this thesis, they will only be refered to briefly. Obviously, chemical modifications of the existing reactive amino acid side chains are restricted by the limited choice of the modifiable side chain groups. Next, SPPS is limited to peptides and small proteins with up to 100 residues in size. Size-imposed restrictions can be circumvented by protein ligation methods where at

least two peptide fragments are ligated to produce full-length protein, but plausible limitations include, for example, restrictions on selecting the ligation site(s) or possible necessity to denature and refold the protein. Furthermore, residue-specific integration of an unnatural amino acid by employing a bacterial strain which is auxotrophic for the desired unnatural amino acid offers the advantage of being able to exploit an unmodified protein synthesis machinery, but is restricted to the unnatural amino acids which are structurally resembling to the canonical amino acids and it results in global replacement of the natural amino acid with its close structural analogue. Last, chemical acylation of a nonsense suppressor tRNA is limited by the need to continuously deliver tRNA with stoichiometric acylation, which can then be used *in vitro* or in microinjected cells, providing restrains on the overall protein amount which can be generated. It is unarguable that the ability to introduce unnatural amino acids co-translationally at defined sites in proteins opens up plethora of opportunities for controlled manipulations and investigations, *in vitro* and *in vivo*, in the field of protein biology.

The unnatural amino acid must be delivered in response to a unique STOP or extended codon, although missing or rare codons were also exploited (Strømgaard et al., 2004; Wang et al., 2009). Moreover, a combination of STOP codon and frameshift mutation was used to introduce two different unnatural amino acids simultaneously in the same protein in *E. coli* and in mammalian cells and Köhrer et al. combined amber and ochre (UAA) STOP codons to insert two different unnatural amino acid in mammalian cells (Anderson et al., 2004 PNAS; Köhrer et al., 2003; Rodriguez et al., 2006; Xie and Schultz, 2005).

Incorporation of an unnatural amino acid at defined position in a protein requires an orthogonal pair comprised of a transfer RNA (tRNA) and the adjacent aminoacyl tRNA synthetase (aaRS) specific for the unnatural amino acid (Chin, 2011; Hino et al., 2006; Liu and Schultz, 2010; Xie and Schultz, 2006 and Figure 10). The amber STOP codon (TAG) is the most frequently used as codon assigned to the unnatural amino acid because (i) it is the least used to terminate protein synthesis of the essential genes in *E. coli* and in yeast, (ii) it was reported that amber suppression can naturally occur in mammalian cells and (iii) it was demonstrated that amber suppression in *E. coli* with the canonical amino acids did not considerably hinder growth rate of the host organism, thus allowing an assumption that similar could be expected for amber suppression with the unnatural amino acids (Beier and Grimm, 2001; Hino et al., 2006; Noren et al., 1989; Wang et al., 2009; Xie and Schultz, 2005).

An orthogonal tRNA/aaRS pair developed for the unnatural amino acid must not cross-react with any of the endogenous elements of the protein synthesis machinery. Orthogonality means that the unnatural amino acid must be incorporated at defined site in response to a unique codon, which must not encode any of the canonical amino acids. Moreover, the tRNA specific for the unnatural amino acid must be optimised in such a manner that it does not recognise any of the 20 common amino acids, and that it is not recognised by any of the endogenous aaRSs. In addition, its cognate aaRS must not aminoacylate any of the endogenous tRNAs. Furthermore, this aaRS must aminoacylate only its corresponding tRNA and only with the unnatural amino acid. Last, the unnatural amino acid must be stable when added to the growth medium, must be efficiently transported into the cytoplasm and stable to the hydrolytic actions of the cellular proteases. Liu and Schultz investigated the

efficiency of cellular uptake of 138 unnatural amino acids in *E. coli* and concluded that the highest cellular toxicity was obse rved when the unnatural amino acids closely resembled the canonical ones or when they contained reactive functional groups (Liu and Schultz, 1999).

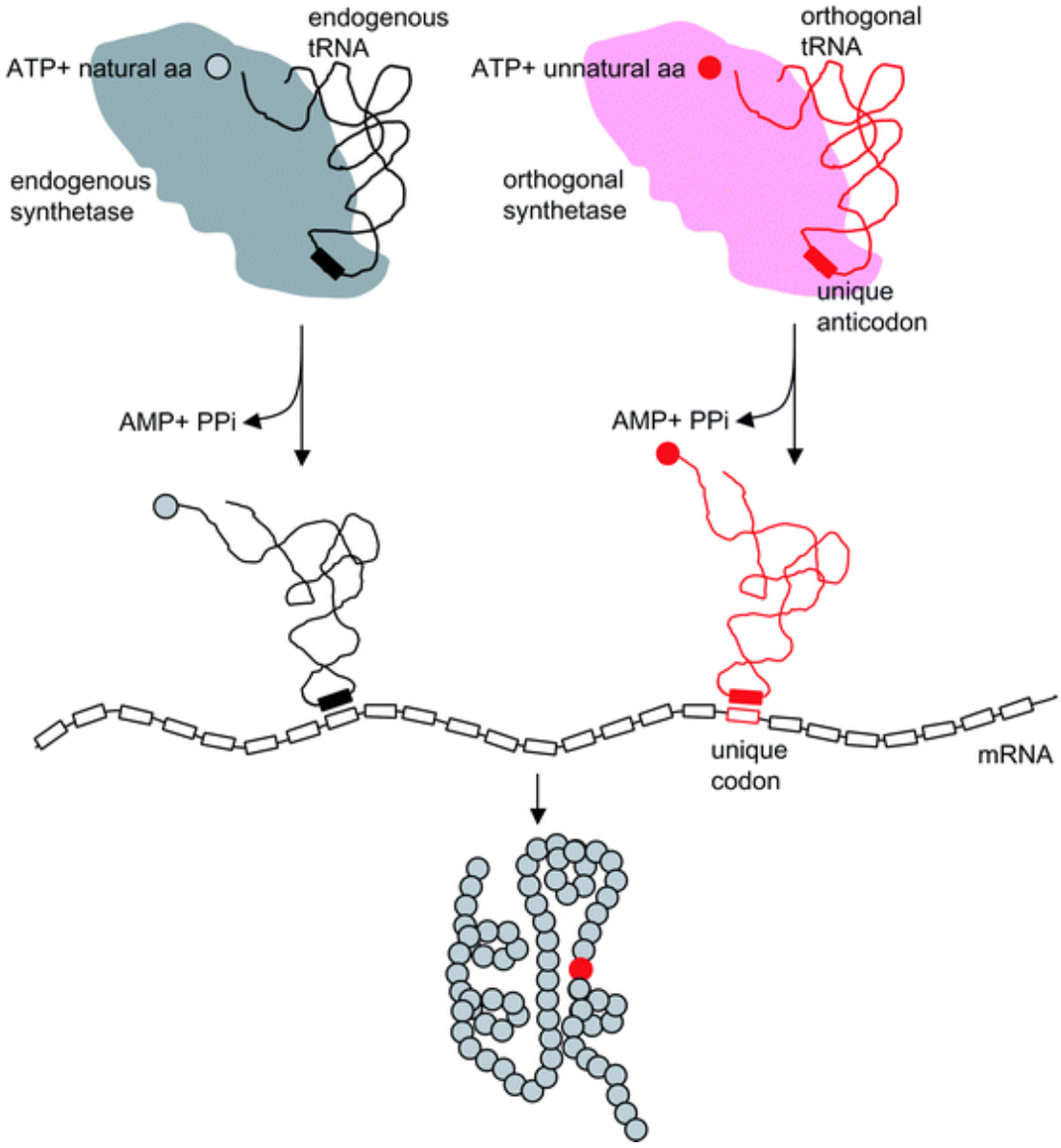


Figure 10: A general principle of expanded genetic code methodology. To introduce unnatural amino acids in vivo, the orthogonal aaRS acylates orthogonal tRNA with an unnatural amino acid, which is then inserted at the position specified by the unique codon (reproduced from Wang and Schultz, 2002, with permission of the Royal Society of Chemistry (RSC) on behalf of the European Society for Photobiology, the European Photochemistry Association and the RSC).

In general, the most challenging aspect in developing an orthogonal tRNA/aaRS pair, specific for the unnatural amino acid, refers to the orthogonality of the tRNA and its cognate aaRS (Liu and Schultz, 2010). A pair of tRNA/aaRS, which would be specific for an unnatural amino acid, must be imported from a different organism, providing there is no cross-reactivity with the elements of the endogenous protein synthesis machinery, thus avoiding cumbersome approach of generating a novel, functional and orthogonal pair of tRNA/aaRS (Chin et al., 2003; Liu and Schultz, 2010; Wang et al., 2000; Wang et al., 2001; Wang et al., 2009; Xie and Schultz, 2006).. The development of an orthogonal pair of tRNA/aaRS specific for the photo-cross-linkable *p*BPA was not the subject of this thesis, but the development of such an orthogonal pair is provided in the following chapters. The tyrosyl-tRNA/RS (Tyr-tRNA/TyrRS) pair from archaebacterium *Methanococcus jannaschii* provided an orthogonal pair for use in *E. coli* (Wang et al., 2000). It was demonstrated that *M. jannaschii* TyrRS does not aminoacylate tRNAs from *E. coli*, whereas *M. jannaschii* Tyr-tRNA provided a poor substrate for *E. coli* aaRSs (Wang et al., 2000, Wang et al., 2001). In addition, aaRSs in *M. jannaschii* do not have a proofreading mechanism and should therefore not interfere with the delivery of the unnatural amino acid ligated to the tRNA specific for it when applied in *E. coli*. To further increase selectivity of the imported tRNA/aaRS pair for the unnatural amino acid and/or reduce reactivity of the endogenous translational elements for any of the imported components for protein synthesis, the imported tRNA/aaRS pair was subjected to mutagenesis of the nucleotides that are not directly in contact with the *M. jannaschii* TyrRS and to the subsequent rounds of positive and negative selection. In brief, the crystal structure of the homologous TyrRS from *Baccilus stearothermophilus* provided structural rationale for introducing

alterations in order to increase selectivity of the novel tRNA/aaRS orthogonal pair. Next, a library of *M. jannaschii* aaRS mutants was transformed in *E. coli* cells which contained a nonsense codon introduced in the gene encoding chloramphenicol acetyltransferase (CAT). By allowing cells to grow in the media supplied with both the unnatural amino acid and the appropriate selection antibiotic, survived cells enabled selection of the aaRSs which can be charged with any amino acid thus suppressing the nonsense codon in the CAT gene and enabling survival. In the next round, the survived cells from the first, positive, selection were grown without the addition of the unnatural amino acid to the medium, thus eliminating aaRSs which could be charged with the canonical amino acids. For fine-tuning selectivity of the tRNAs, a library of the tRNA mutants was first subjected to the negative selection, based on the suppression of a nonsense codon introduced in the gene encoding toxic product called barnase, thus eliminating the tRNAs which can be aminoacylated also by the endogenous aaRSs of *E. coli*. The second round of positive selection was based on the nonsense suppression in the gene encoding β-lactamase, thus selecting the tRNAs which were not aminoacylated by the aaRSs from *E. coli*. In addition, more orthogonal tRNA/aaRS pairs for application in *E. coli* were developed, thus increasing diversity of the unnatural amino acids which could be incorporated into proteins by expanded genetic code methodology (Xie and Schultz, 2005). To extrapolate the application of the expanded genetic code to the eukaryotic cells, *S. cerevisiae* was selected as the initial organism from which a novel pair of orthogonal tRNA/aaRS could be developed (Chin et al., 2003; Liu and Schultz, 2010; Wang et al., 2009). The above described two-step selection procedure was not applicable for developing an orthogonal pair of tRNA/aaRS in eukaryotic host because of low

efficiency of transformation, thus making generation of big mutant tRNA and aaRS libraries impractical. Furthermore, an additional obstacle is the existence of the nonsense-mediated decay (NMD) in eukaryotes (Amrani et al., 2006; Wang et al., 2009). NMD accelerates degradation of the mRNAs containing premature STOP codons, and the effect is more pronounced if the premature STOP codons were to reside closer to the 5'-end of an mRNA.

In addition, pairs of tRNAs/aaRSs developed in *E. coli*, albeit not cross-reacting with eukaryotic protein synthesis elements, were limited in the final protein yield due to differences in the mechanisms for transcription and processing of the tRNAs (Liu et al., 2007; Wang et al., 2009). Namely, in the eukaryotes, a tRNA is transcribed from the promoter, referred to as the A- and B-box, which is integrated within the tRNA gene, whereas *E. coli* contains the promoter, only the A-box, upstream of the tRNA encoding gene (Figure 11). In contrast, *B. stearothermophilus* Tyr-tRNA contains A- and B-boxes, it is not recognised and charged by the mammalian aaRSs and, together with the *E. coli* TyrRS, represents an orthogonal pair for use in yeast and mammalian cells (Chin et al., 2003; Wang et al., 2009). In addition, expression of functional prokaryotic tRNAs in mammalian cells was advanced by using external mammalian promoters which are positioned upstream of the coding region and do not require any intragenic elements, such as A- and B-box, thus enabling unobstructed transcription and generation of functional prokaryotic tRNAs (Liu and Schultz, 2010; Wang, et al., 2009).

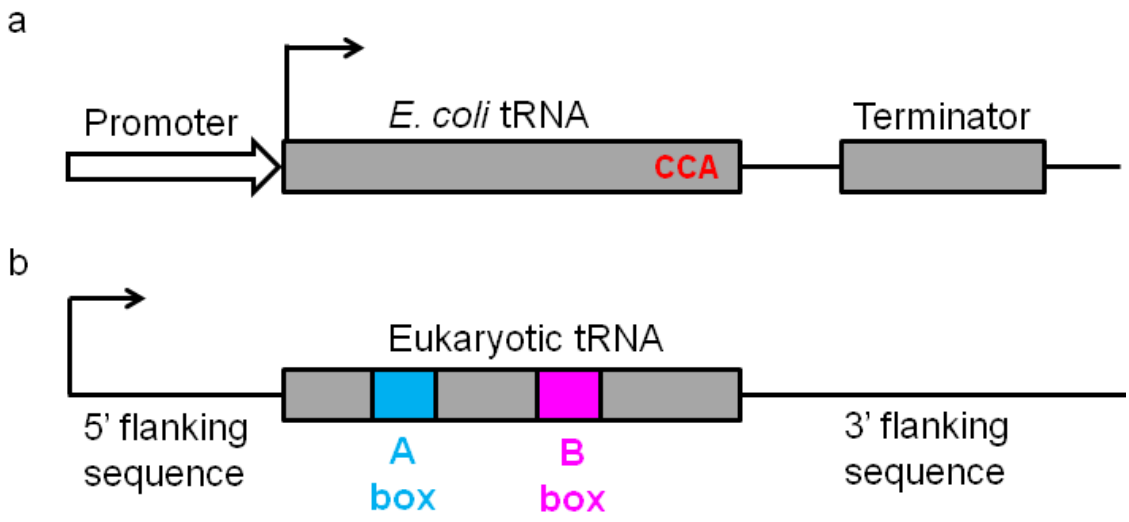


Figure 11: Difference in the expression of prokaryotic and eukaryotic tRNAs. (**a**) Genes encoding tRNAs in *E . coli* are transcribed through promoters located upstream of the gene. (**b**) In eukaryotes, promoters are within the geen (A and B box). Adapted from Wang et al., 2009.

The spectrum of applications of the expanded genetic code is vast, and the unnatural amino acids have been successfully incorporated into proteins in CHO and HEK293 mammalian cells, and even in primary neurons (Liu and Schultz, 2010; Wang et al., 2007b; Wang, et al., 2009). One of the more attractive aspects of this methodology involves incorporation of the unnatural amino acids with chemical moieties that could serve as a reactive handle for further derivatization. The carbonyl group represents one such moiety. Photo-excitable amino acid *para*-benzoylphenylalanine (*p*BPA) is very appealing because its carbonyl group enables cross-linking of proteins through highly reactive diradical-based intermediate which is generated transiently after exposure to 350-365 nm light (Hino et al., 2005; Tanaka et al., 2008).

1.7.1 *p*BPA

The synthesis of *p*BPA was reported by in 1973, but it was only in 1986 that Kauer *et al.* described its successful incorporation into peptides and subsequent photo-induced cross-linking of 17mer calmodulin-binding peptide with calmodulin upon exposure to 350 nm light (Kauer *et al.*, 1986). Since then, the benzophenones became wildly attractive choice for photo-labeling reagents due to several distinct advantages when comparing properties of *p*BPA to the other photo-probes, such as diazirirines or aryl azides, which were also genetically encoded in prokaryotes and eukaryotes and used for cross-linking (Ai, Shen et al., 2011; Chin et al., 2002a; Chin et al., 2003; Chou et al., 2011; Dormán and Prestwich, 1994; Haberkant and van Meer, 2009; Liu et al., 2007; Tanaka et al., 2008; Tippmann, Liu et al., 2007).

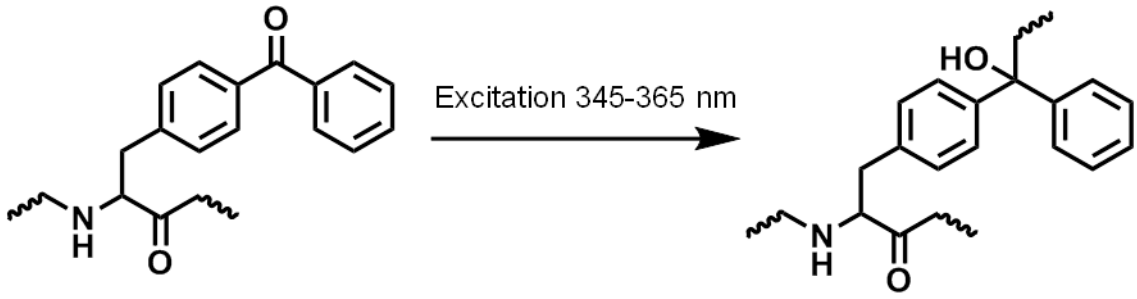


Figure 12: Principle of photo-induced cross-linking with *p*BPA. Structures were generated with ChemBioDraw.

Benzophenones are chemically stable (Dormán and Prestwich, 1994). Upon exposure to 350-360 nm light, they preferentially cross-link with the proximal C-H bonds (Tanaka et al., 2008 and Figure 12). Initially, the reactive radius of the benzophenone moiety was approximated as a sphere with 3.1 Å radius measured from the oxygen of the carbonyl group (Dormán and Prestwich, 1994). The reactivity of the benzophenone group and, thus, the efficiency of the

cross-linking depend on the geometric relationship between the carbonyl group of the benzophenone and a proximal C-H bond, and also on their electrochemical environment (Dormán and Prestwich, 1994; Sato et al., 2011; Wittelsberger et al., 2008). Benzophenones exhibit enhanced reactivity toward electron rich tertiary atoms, present, for example in leucine or valine, but also toward methenyl (CH_2) groups close to a heteroatom, such as those present in methionine, arginine or lysine (Dormán and Prestwich, 1994). Importantly, *p*BPA has been genetically encoded for expression in prokaryotes, yeast and mammalian cells and also widely used for photo-induced cross-linking, employing the expanded genetic code methodology or incorporation into peptides, proving especially beneficial for studying membrane-associated components and transient protein interactions or PTMs (Alexander, Gerauer, et al., 2009; Chen et al., 2007; Chin et al., 2002a and b; Chin and Schultz, 2002; Chin, 2011; Chou et al., 2011; Forné et al., 2012; Hino et al., 2005; Kage et al., 1996; Krishnamurthy, Dugan et al., 2011; Liu et al, 2007; Majmudar et al., 2009; Musial-Siwek et al., 2007; O'Neil et al., 1989; Ryu and Schultz, 2006; Sato et al., 2011; Uezu et al., 2012; Völkert et al., 2003; Wang et al., 2010b; Xie and Schultz, 2005; Xie and Schultz, 2006; Ye, Köhrer et al., 2008).

1.8 Aims

The ultimate aim of the work described in this thesis was to show if nonsense suppression strategy utilizing the amber STOP codon could be applied to develop substrate -trapping mutants of PTPs, which would be more efficient in trapping potential interacting partners than conventional substrate-trapping mutants of PTPs, which contain mutations mostly of the catalytically important residues. The model phosphatases with which this strategy was to be developed were PRL-3 and VHR. These phosphatases were selected due to (i)

the inherent difficulties of traditional methods to identify their interacting partners and (ii) their link to cancer.

To achieve this, the first aims were to show that *p*BPA could be selectively incorporated, *in vivo*, in the selected PTPs. For this, proper positions for the introduction of *p*BPA had to be found in PRL-3 and in VHR and the conditions for the over-expression in *E. coli* and purification had to be optimised. Screening for the most optimal expression conditions was to be done on a small-scale bacterial culture (50 ml) and then extrapolated to at least five times higher culture volume.

Next, the aim was to characterize the generated *p*BPA-containing variants for their cross-linking ability *in vitro*, and to interpret the findings by using more conventional mutants as controls. Characterisation of the cross-linking abilities of the *p*BPA-containing proteins was to be done on a small scale to fit 96 well-plate setup and to minimise the amount of the *p*BPA-containing proteins needed.

Finally, the aim was to extrapolate if this tool could be used in profiling interactions of the PTPs in cell lysates. This was to be achieved by using PTP variants where catalytically essential residue (general acid/base aspartate) was replaced with *p*BPA (D72*p*BPA mutant of PRL-3). This aim will require substantial optimisation- for example, the amounts of the bait protein and the lysate used, duration of pre-incubation of the bait with the lysate prior to UV-exposure, duration of UV-exposure and optimisation of the purification conditions.

2. Materials and methods

2.1 Materials and methods for Chapters 3.1-3.3

2.1.1 Reagents

para-Benzoyl-L-Phenylalanine (*p*BPA) was from PepTech (Massachusetts, USA). *p*-Acethylphenylalanine hydrochloride (*p*ACF) was from Sinon Pharma (China). 4-nitrophenyl phosphate disodium salt hexahydrate (*p*NPP) and sodium orthovanadate (Na_3VO_4) were from Sigma (Steinheim, Germany). 3-*O*-Methylfluorescein phosphate cyclohexylammonium salt (OMFP) and Glutaraldehyde solution Grade I, 50% in H_2O were purchased from Sigma-Aldrich. D-*myo*-Phosphatidylinositol 3,4,5-trisphosphate ($PI(3,4,5)P_3$) was from Echelon Biosciences (USA). Disuccinimidyl suberate (DSS) was from Thermo Scientific (Rockford, IL, USA). N-2-Hydroxyethylpiperazine-N'-2-ethane sulfonic acid (HEPES) was purchased from Biomol GmbH (Hamburg, Germany). Sodium dodecylsulfate (SDS) was purchased from SERVA Electrophoresis GmbH (Heidelberg, Germany). 30% Acrylamide/Bis Solution, 37.5:1 (2.6% C) was from BIO-RAD (Hercules, CA, USA).

Fetal bovine serum (FBS), penicillin/streptomycin and Opti-MEM I + GlutaMAX-I were from Gibco, Life Technologies (Paisley, UK). L-glutamine (200 mM stock) and Trypsin-EDTA 1 x solution were from Sigma. Fugene HD was purchased from Promega (Madison, USA). Complete, ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail (PIC) tablets and PhosSTOP (PS) phosphatase inhibitor coctail tablets were

purchased from Roche (Mannheim, Germany). H_2O_2 was 30% solution from Fluka (Steinheim, Germany). Acetonitrile and formic acid (99%), used for liquid chromatography (LC)-MS analysis, were both Ultra LC MS (ULC MS), from Biosolve.

For amplification and site-directed mutagenesis (SDM) by polymerase chain reaction (PCR), the following enzymes were used: Phusion DNA Polymerase (New England Biolabs, NEB), Phire Hot Start II DNA Polymerase, Dream Taq Green PCR Master Mix, Maxima Hot Start PCR Master Mix (all from Thermo Scientific), Ampli Taq Gold 360 DNA Polymerase (Life Technologies) and QuickChangeLightning (QCL) SDM kit (Agilent Technologies, USA).

dNTP Mix 10 mM each was from Thermo Scientific. All regular or high fidelity (HF) restriction enzymes used were from NEB, and all fast digest (FD) ones were from Thermo Scientific. DpnI and Tango buffer (10 x) were from Thermo Scientific.

T4 DNA Ligase was from Thermo Scientific or from Life Technologies.

Other reagents (used for preparing different buffers and solutions listed in the section 2.1.5, agarose, L-(+)-Arabinose, antibiotics, reducing agents L-Glutathione and DL-Dithiothreitol (DTT), detergents Nonidet P 40 Substitute (NP40) and TWEEN 20) and paraformaldehyde (PFA) were from Merck (Darmstadt, Germany) or from Sigma-Aldrich.

2.1.2 Materials

Buffers for protein purification and growth media (GM) for cultivation of mammalian cell lines were filtered by using Millipore Express PLUS 0.22 µm

filters, and for sterilization of buffers and media on a scale to 20 ml, Millex-GP 0.22 µm filters, both from Merck Millipore (Cork, Ireland), were used.

For small scale protein purifications, Ni-NTA Superflow and polypropylene columns from Qiagen (Hilden, Germany) were used. His-Select Nickel Affinity Gel was from Sigma.

For purification of recombinant PRL-3 and VHR variants, HisTRAP HP (1 ml) and GSTrap FF (1 ml) columns from GE Healthcare (Uppsala, Sweden) were used.

Protein samples were resolved by using precast Nu Polyacrylmide gel electrophoresis (PAGE) 4-12% Bis-Tris Gels (1.0 mm thick) from Novex by Life Technologies.

Protein samples were dialysed by using Spectra/Por Dialysis Membranes with molecular weight cut-off (MWCO) 3,000 or 6-8,000 (3K or 6-8K) from Spectrum Laboratories, Inc., USA.

Protein samples were concentrated by using Ultra Centrifugal Filters-0.5 ml, with 3K or 10K MWCO from Amicon (Cork, Ireland).

PIP Strips were from Echelon Biosciences.

The following films were used: Kodak BioMax MR (high resolution radioisotope and chemiluminescent) from Sigma-Aldrich, and Amersham Hyperfilm ECL (high performance chemiluminescence film) from GE Healthcare (Buckinghamshire, UK).

The following Western blotting detection reagents were used: ECL Western Blotting Detection Reagents (GE Healthcare, Buckinghamshire, UK) and

Western Lightning Plus-ECL Enhanced Chemiluminescence Substrate (Perkin Elmer, USA).

2.1.3 Instruments

DNA amplification PCR was conducted on a MJ Research Peltier Thermal Cycler PTC-200 from Biozym Diagnostik GmbH (Oldendorf, Germany).

DNA bands in agarose gels were visualized with a Dual-intensity transilluminator from Herolab (Wiesbaden, Germany).

DNA and protein concentrations were measured on a NanoDrop 1,000 Spectrophotometer from PEQLAB Biotechnologie GmbH (Erlangen, Germany).

Optical density (OD) of growing bacterial cultures was measured by Ultrospec 2,100 *pro* UV/Visible Spectrophotometer from Amersham Biosciences.

To mechanically disrupt the cells by sonication, Branson Sonifier W-250 Ultrasonic Disintegrator with a small microtip was used.

To load the samples onto the HisTRAP HP and GSTrap FF columns, LKB Pump P-1 from Pharmacia (Sweden) was used.

Proteins were purified with Åktapurifier 10 system (GE Healthcare). The system contains a P-903 pump, an M-925 mixer, an UV-900 detector and a fraction collector Frac 950, and is operated with the UNICORN 5.20 software.

Activity assays with *p*NPP or OMFP were conducted in a 96-well plate setup in a microplate reader Tecan Saphire 2 (Tecan, Austria).

Photo-induced cross-linking was conducted by using 1,000 watt (W) mercury lamp equipped with 20CGA-345 filter with a 345 ± 5 nm cut-on/cut-off wavelength, 1 mm thick and dimensions 50.8 x 50.8 mm (Newport, USA).

For lipid-binding assay, photo-induced cross-linking was performed with Spectrolinker UV Crosslinker Model XL-1500 A/F from Spectroline (Westbury, USA), by using an optimal cross-linking mode which provides UV energy dosage of 120 mJ/cm^2.

The proteins were transferred from polyacrylamide gels to nitrocellulose membranes by using iBlot Dry Blotting System with iBlot Gel Transfer Stacks (both from Novex by Life Technologies).

2.1.4 Miscellaneous

DNA primers were purchased from Eurofins MWG GmbH (Ebersberg, Germany).

DNA sequencing was done by GATC Biotech AG (Konstanz, Germany).

Purification of DNA was done by using the following commercial kits: QIAquick PCR Purification Kit, QIAquick Gel Extraction Kit (both from Qiagen, Hilden, Germany) or PureLink PCR Purification Kit (Life Technologies).

For isolation of DNA, the following kits were used: Quick Lyse Miniprep Kit, QIA prep Spin Miniprep Kit or QIA filter Plasmid Maxi Kit (all from Qiagen).

For removal of the N-terminal His_6-tag in the proteins, Glutathione *S*-Transferase fused tobacco etch virus (GST-TEV) protease (stock at 1 mg/ml) provided from the Protein Expression and Purification Core Facility at EMBL Heidelberg was used.

GeneRuler 1 kb DNA Ladder, used as DNA reference standard and 6 x TriTrack DNA Loading Dye were from Thermo Scientific.

Blue Eye Prestained Protein Marker from Jena Bioscience (Jena, Germany), Spectra Multicolor Broad/High Range Protein Ladder from Thermo Scientific and BenchMark Prestained Protein Ladder from Novex by Life Technologies were used as protein size references.

EnzCheck Phosphate Assay Kit was from Molecular Probes.

The following centrifuges were used: centrifuge 5415 R (for 1.5 and 2 ml tubes), with up to 13,200 rpm, from Eppendorf (Hamburg, Germany) and Sorvall RC6, with up to 22,000 rpm, from Thermo Scientific.

The following film development system was used: Kodak RP X-OMAT Processor Model M6B, from Kodak GmbH (Stuttgart, Germany).

2.1.5 Buffers, solutions and media

2.1.5.1 TAE

As running buffer for agarose gels, Tris-acetate-EDTA (TAE) buffer was used. It contained 40 mM Tris base, 20 mM glacial acetic acid and 1 mM EDTA; pH value is 8.1-8.3.

2.1.5.2 SDS-PAGE running buffer

As SDS-PAGE running buffer, NuPAGE MOPS SDS (20 x) from Novex by Life Technologies was used.

2.1.5.3 SDS-PAGE sample buffers

SDS-PAGE sample buffer (2 x) was prepared by mixing 2.0 ml of stacking gel buffer (4 x) pH 6.8, 1.6 ml of glycerol, 3.2 ml of 10% (w/v) SDS, 0.8 ml of β-mercaptoethanol and 0.4 ml of 1% (w/v) bromphenol blue. Non-reducing SDS-PAGE sample buffer (2 x) did not contain β-mercaptoethanol. SDS-PAGE

sample buffer (6 x) contained 7 ml of stacking gel buffer (4 x) pH 6.8, 2.6 ml of glycerol, 1 g of DTT, 400 µl of 10% (w/v) SDS and 60 µl of 1% (w/v) bromphenol blue.

2.1.5.4 SDS (10%) stock solution

SDS (10%) stock solution was prepared by dissolving 10 g of SDS in 0.8 l of ddH_2O and than topping up the final volume to 1 l with ddH_2O.

2.1.5.5 Stacking gel buffer (4 x)

Stacking gel buffer (4 x) was prepared by dissolving 60.5 g of Tris-base in 1 l of ddH_2O and adjusting the pH value to 6.8.

2.1.5.6 Resolving gel buffer (4 x)

Resolving gel buffer (4 x) was prepared by dissolving 181.5 g of Tris base in 1 l of ddH_2O and adjusting the pH value to 8.8.

2.1.5.7 Ammonium persulfate (10%)

To make 10% (w/v) solution of ammonium persulfate, 1 g of a substance was dissolved in 10 ml of ddH_2O.

2.1.5.8 Tris-HCl (1M, pH 7.5)

1 M Tris-HCl pH 7.5 was prepared by EMBL's media kitchen. Tris base (121 g) was dissolved in 800 ml of H_20, and then the pH value was adjusted to 7.5 with concentrated HCl. The final volume was topped up to 1 l with H_2O prior to autoclaving.

2.1.5.9 Phosphate Buffer Saline (PBS)

PBS was prepared by EMBL's media kitchen and it contained 137 mM sodium chloride (NaCl), 2.7 mM potassium chloride (KCl), 2 mM potassium dihydrogen phosphate (KH_2PO_4), 10 mM disodium hydrogenphosphate (Na_2HPO_4), with the pH value adjusted to 7.2.

2.1.5.10 EDTA (0.5 M, pH 8.0)

0.5 M EDTA pH 8.0 was prepared by EMBL's media kitchen. Disodium-EDTA dihydrate (186.1 g) was dissolved in 700 ml of H_20, and then the pH value was adjusted to 8.0 with 10 M NaCl. The final volume was topped up to 1 l with H_2O prior to autoclaving.

2.1.5.11 Tris buffered saline (TBS) (10 x)

TBS 10 x stock solution contained 0.5 M Tris-HCl (pH 7.5) and 1M NaCl. TBS 1 x solution was prepared by diluting the stock solution in ddH_2O.

2.1.5.12 TBS-Tween20 (TBS-T)

TBS-T buffers were prepared by adding an appropriate volume of Tween-20 to a 1 x solution of TBS in order to obtain 0.05% or 0.1% (both v/v).

2.1.5.13 Staining solutions for protein gels

Protein gels were regularly stained with an aqueous solution prepared from Coomassie Brilliant Blue G-250 at 0.008% (w/v) and 35 mM HCl.

Alternatively, colloidal Coomassie Blue G-250 stain was utilized and was prepared according to the reported protocol (Candiano, Bruschi, et al., 2004). In brief, the dye was made by sequentially adding the components listed below to an aqueous solution: 10% (v/v) phosphoric acid, 10% (w/v) ammonium

sulfate, 0.12% (w/v) Coomassie Blue G-250 and 20% (v/v) anhydrous methanol. The solution was kept light protected at rt. The gels were incubated with a dye either for a couple of hours at rt or overnight, either at rt or at 4°C, with gentle agitation, and then destained in ddH_2O.

2.1.5.14 Staining nitrocellulose membranes

Ponceau S staining solution was used to assess the efficiency of protein transfer. It contained 0.1% (w/v) Ponceau powder and 5% (v/v) acetic acid in aqueous solution.

2.1.5.15 Protein transfer

Buffer for protein transfer for Western blot analyses contained 48 mM Tris base, 39 mM glycine, 0.37% (v/v) SDS (from 10% stock solution) and 20% (v/v) methanol.

2.1.5.16 Stripping buffer for immunoblots

As stripping buffer for immunoblots, 0.1 M glycine pH 2.8 was used.

2.1.5.17 Stripping buffer for His Trap HP columns

The columns were stripped by washing with stripping buffer as recommended by the manufacturer, and the stripping buffer contained 20 mM sodium phosphate, 0.5 M NaCl, 50 mM EDTA (pH 7.4).

2.1.5.18 Recharging buffer for His Trap HP columns

The columns were recharged by loading 0.5 ml of 0.1 M nickel (II)- sulfate in ddH_2O.

2.1.5.19 Eluting buffer for GSTrap FF columns

GSTrap columns were eluted with at least 10 mM GSH, pH 8.

2.1.5.20 SOC medium

SOC medium was prepared by EMBL's media kitchen and it contained 2% (w/v) of bacto tryptone, 0.5% (w/v) of yeast extract, 5 M NaCl, 1 M magnesium chloride ($MgCl_2$) 1 M KCl, 1 M magnesium sulfate and 20% (v/v) glucose, with the pH value adjusted to 7.2.

2.1.5.21 Media for cultivation of bacterial cultures

For propagation, bacterial strains were incubated in autoclaved Luria-Bertani (LB) or Terrific Broth (TB)-FB broths, supplemented with the appropriate antibiotic(s), as determined by the plasmid vectors and bacterial strains used. Both broths were prepared by EMBL's media kitchen. LB medium (1 l) contained: 10 g of bacto tryptone, 5 g of yeast extract and 5 g of NaCl, with the pH value adjusted to 7.2. TB-FB medium (1 l) contained phosphate stock and extract stock at 1:9 ratio (v/v). Phosphate stock buffer contained 0.16 M KH_2PO_4 and 0.72 M dipotassium hydrogen phosphate (K_2HPO_4) (both from Merck). Extract stock (0.9 l) contained 12 g bacto tryptone, 24 g yeast extract and 4 ml glycerol. The final pH value was adjusted to 7.2.

The antibiotics were used at the concentrations indicated: ampicillin (Amp) 50 or 100 µg/ml, kanamycin (Km) 50 µg/ml, chloramphenycol (Cm) 33 µg/ml and tetracycline (Tet) 2.5 µg/ml.

LB agarose plates were prepared by EMBL's media kitchen. They contained 1.5% (w/v) bacto agar and the following antibiotics at the concentrations indicated: Amp 100 µg/ml, Km 30 µg/ml, Cm 33 µg/ml, combination of Amp and

Cm at 100 and 33 µg/ml, respectively, and combination of Amp and Tet at 100 and 10 µg/ml, respectively.

2.1.5.22 Lysis buffers for mammalian cell lines

The buffer used to lyse COS-1 cells contained 20 mM Tris-HCl (pH 7.4), 137 mM NaCl, 10% (v/v) glycerol, 1% (v/v) NP-40, 1 mM EDTA and PIC tablet prepared according to the manufacturer's instructions and used at 1:25 dilution.

For interactome analysis involving different variants of PRL-3, HEK293 cells were lysed in the HNTG buffer containing 20 mM HEPES (pH 7.5), 150 mM NaCl, 1% (v/v) Triton X-100, 10% (v/v) glycerol and PIC (as above).

2.1.6 DNA amplification by PCR

The DNA templates were commonly used at the final amount of 20 or 100 ng for 50 µl PCR mix, or accordingly downscaled for 25 µl PCR mix.

For 50 µl PCR reaction mix with Phire Hot Start II DNA Polymerase or Phusion DNA Polymerase, 0.5 µl of the enzyme was used; when using QCL-SDM kit, 25 µl reaction mixes were made. Where required, dNTPs were added extra from the stock with 10 mM concentration of each dNTP to a final concentration of 0.2 mM. The forward and reverse primers were used at the final concentration of 0.2 µM. When using QCL-SDM kit, the primer concentration was 0.125 µM.

The PCR conditions were adjusted depending on the amplification enzyme and the set of sense and antisense primers used. The annealing temperature was adjusted accordingly, and in general was set to about 4°C below the melting temperature (Tm) of the cognate primer set; the Tm was approximately calculated by using the equation: (G+C) x 4 + (A+T) x 2. Amplification step was

generally conducted for 25-30 cycles, with the final extension step at 72°C of at least 10 min.

For DpnI digestion, 1 µl of the enzyme was added directly to the PCR mixture and incubated for 2 h at 37°C, after which the enzyme was heat-inactivated by incubating at 80°C for 20 min. When using QCL-SDM kit, DpnI enzyme supplied with the kit was used.

2.1.7 DNA ligation

For ligation, 1 µL of T4 DNA Ligase was added to the digested DNA insert and vector mixed at 3:1 (v/v) ratio. Ligation was conducted by incubating the ligation mixtures at rt for up to 5 h, or by overnight incubation at 16°C, and then transforming into chemically competent *E. coli* strain.

2.1.8 Procedure for DNA isolation

For isolation of DNA, the colonies were inoculated into LB medium (around 3 ml) supplemented with the appropriate antibiotics and incubated at 37°C with vigorous agitation. For isolation and purification of DNA, commercially available kits were used according to the manufacturer's instructions.

2.1.9 Procedure for DNA propagation

For propagation of DNA, a single colony was inoculated into LB medium (200-400 ml) supplemented with the appropriate antibiotics and incubated for at least 16 h at 37°C with vigorous agitation. For isolation and purification of DNA, commercially available kits were used according to the manufacturer's instructions.

2.1.10 Procedure for transformation of chemically competent strains of *E. coli*

For propagation of DNA, chemically competent DH5α and TOP10 *E. coli* strains were used, and for expression by using chemically competent cells, BL21 DE3 strain.

DNA (1-5 µl) was added to the thawed cells (at least 50 µl) and incubated on ice for 30 min. HS was applied for approximately 90 s at 42°C, after which the cells were recovered on ice for 2 min. Next, 300 µl of LB medium was added, followed by incubation for about 45-60 min with vigorous agitation at 37°C before plating on LB-agar with appropriate antibiotic selection. The plates were incubated overnight at 37°C or for approximately 48 h at rt.

At least 20 µl of the PCR mix was used in transforming no less than 50 µl of the thawed cells. For ligation mixture, 10 µl of the ligation reaction was used for 100 µl of the cell suspension. For propagation of DNA, 1-5 µl of DNA was used for 50-150 µl volume of the cell suspension.

For BL21 DE3 strain, the same procedure was used, with 5 µl of DNA added to 100 µl of the thawed cells.

2.1.11 Transformation procedure for electrocompetent *E. coli* strains

Electrocompetent *E. coli* strains BL21 AI and BL21 RIL were obtained from the laboratory of Dr. Edward Lemke (Structural and Computational Unit, EMBL Heidelberg). Plasmid encoding PRL-3 or VHR variant with a specific position exchanged for the amber STOP codon, and plasmid encoding tRNA/aaRS specific for *p*ACP (pEVOLV-ACP) or for *p*BPA (pEVOLV-BzF), which were originally obtained from Peter Schultz (The Scripps Research Institute,

California), were co-transformed by electroporation into bacterial strain BL21 AI or BL21 RIL by using *E. coli* gene pulser cuvette and MicroPulser. Typically, 2-4 ng of each plasmid was mixed with 50 µl of the thawed cells prior to transferring the mixture to a pulser cuvette. After a pulse, the cells were recovered in 200 µl of the SOC medium warmed to rt and incubated for around 45-60 min at 37°C with horizontal agitation.

2.1.12 Antibodies (ab) and Western blotting

The following primary ab were used: anti-Flag mouse monoclonal ab (mAb) at 1:10,000 (M2, Sigma), anti-c-Myc mAb at 1:500 (9E10, Santa Cruz Biotechnology), rabbit polyclonal ab (rAb) specific to DUSP3/VHR and pErk1/2 (Thr202/Tyr204), and monoclonal rAb specific to total Erk1/2 (137F5) at 1,000 (all from Cell Signaling Technology). Polyclonal rAb specific to VHR was kindly provided by Dr. Rafael Pulido (Centro de Investigación Principe Felipe) and was used at 1:5,000. Anti-PRL-3 ab (1E7) was used at 1:400.

The following secondary ab were used: goat anti-mouse IgG-peroxidase conjugate (Sigma) and donkey anti-rabbit IgG-horseradish peroxidase (HRP) conjugate (GE Healthcare). Dynabeads protein G was used for immunoprecipitation (IP) (Novex by Life Technologies).

Anti-Flag ab was used according to the slightly modified general protocol enclosed in the manufacturer's instruction sheet. In brief, the membrane was blocked with TBS containing 3% (w/v) nonfat milk for 60 min at rt and then rinsed two times for 5 min with TBS. The primary ab was diluted in blocking solution and was incubated with the membrane for 60 min. The membrane was rinsed as above before adding the secondary ab diluted 1:10,000 in blocking

buffer and incubating for 30-60 min. The membrane was rinsed with TBST-0.05% about eight times for 20 min prior to detection.

For immunoblot analysis using anti-c-Myc ab, the general protocol provided by the manufacturer was followed, with slight modifications. Briefly, the membrane was rinsed for 5 min with TBS, followed by blocking in TBST-0.05% supplemented with 5% (w/v) nonfat milk for 60 min. The membrane was rinsed with TBST-0.05% three times for 15 min. Incubation with the primary ab diluted 1:500 in blocking buffer was conducted for 3 h at rt. After washing the membrane as above, the secondary ab was added at 1:2,000 in blocking buffer and incubated for 60 min. The membrane was washed as above prior to visualization.

Western blots with ab against VHR and, total or pErk1/2 (all from Cell Signaling) were conducted like suggested by the manufacturer. The appropriate secondary ab was used at 1:5,000 in blocking buffer for 60 min.

For Western blot analysis using polyclonal anti-VHR rAb provided by Dr. R. Pulido, the exact procedure described for anti-Flag ab was applied. Anti-VHR ab and the appropriate secondary ab were used at 1:5,000 in blocking buffer for 60 min at rt.

For Western blot using ab against PRL-3 (1E7), the general procedure used for anti-Flag ab was used, but the primary ab was used at 1:400 dilution in TBS with 3% (w/v) BSA for 3 h at rt. The secondary ab was used at 1:10,000 in TBS with 3% (w/v) BSA for 45 min-1 h at rt.

The membranes were stripped by incubating in 0.1 M glycine pH 2.8 for at least 60 min at rt with mild agitation, followed by extensive rinsing with TBS prior to reprobing.

2.1.13 Maintenance of mammalian cell lines

COS-1 cells were a generous gift from Prof. Dr. Antje Gohla (Rudolf Virchow Center/DFG Research Center for Experimental Biomedicine, University of Würzburg, Germany), and HEK293 cells were obtained from the laboratory of Dr. Carsten Schultz (EMBL Heidelberg).

HEK293 and COS-1 cells were cultured at 37°C in a humidified incubator under an atmosphere of 5% CO_2 and passaged 2-3 times a week. Heracell240i and 240 CO_2 incubators used were from Thermo Scientific. The cells were grown in Dulbecco's modified Eagle's medium (DMEM), prepared by EMBL's media kitchen, containing glucose at 4.5 g/l ("high glucose") and additionally supplemented with 10% (v/v) FBS, 1% (v/v) penicillin/streptomycin and 2 mM L-glutamine.

The components of the starving medium were as listed above, but omitting FBS.

Trypsin-EDTA solution 1 x was used for detachment of cells from the culturing plates.

2.1.14 Removal of His_6-tag by GST-TEV

Removal of His_6-tag in the recombinant proteins was done at least overnight at 4°C: (i) by adding the protease to a protein sample in solution; (ii) by adding the protease to a protein sample during dialysis or (iii) by concentrating the protein sample to a volume bellow 1 ml, loading it onto HisTRAP column manually and then loading the protease. To remove the protease, GSTrap column was usually coupled directly below the HisTRAP column, and Åktapurifier was used to elute the sample. Alternatively, the sample was first eluted from HisTRAP

column, concentrated to a volume bellow 1 ml and then passed over GSTrap column. The amount of the protease needed was initially calculated according to the equation: 0.06 mg of protease per 1 mg of protein to be digested (which had been used in the lab previously). Regularly, the amount of the protease used was reduced by about 20% from the calculated one to (i) ease the regeneration of the columns after purification and to (ii) reduce the amount of the protease purchased. Reduction in the amount pf the protease had no effect on the efficiency of digestion.

2.1.15 Intact protein sample analysis by LC-MS

All sample analyses by LC-MS(/MS) were performed by the Proteomics Core Facility at EMBL Heidelberg, and the protocols for sections 2.1.15, 2.3.10 and 2.4.7 were provided by Kristina Dzeyk.

Protein samples (around 1 mg/ml) were acidified using 1% formic acid solution and transferred into vials prior to LC-MS analysis. Desalting and protein separation were carried out using an Acquity Ultra Performance Liquid Chromatography (UPLC) system (Waters) fitted with a C4 column (2.1 mm x 15 cm, 5 µm particle size). The column was maintained at a constant temperature (40°C) throughout. The outlet of the column was coupled directly to a Q-Toff II mass spectrometer (Waters) using the standard electrospray ionization (ESI) in positive ion mode. Solvent A was water, 0.1% formic acid, and solvent B was acetonitrile with 0.1% formic acid. The samples (between 1-20 µl) were loaded onto the column and desalted for 5 min at a flow rate of 0.2 ml/min, 4% B. The proteins were eluted from a column with a constant flow of 0.2 ml/min. During the elution step, the percentage of solvent B increased in linear fashion from 5 to 25% in 1 min, and then increased to 80% in the further 11 min. On the Q-

Toff, a spray voltage of 3.5 kV was applied, with a cone voltage of 35 V and extraction cone at 10 V. Collision energy of 8 eV was used, with argon in the collision cell. The desolvation temperature was set to 320°C, with a source temperature of 120°C. Data were acquired in continuum mode, over a mass range 500-3500 m/z, with a scan time of 0.5 and interscan delay of 0.1 s. Data were externally calibrated against a reference standard of intact myoglobin, acquired immediately after sample data acquisition. Spectra across the protein chromatographic peak(s) were summed and intact mass was calculated using the MaxEnt1 maximum entropy algorithm (Waters/Micromass) to give the zero charge deconvolution molecular weight.

2.1.16 Circular dichroism (CD) measurements

CD measurements were done with assistance of Prof. Vladimir Rybin (Biophysical support, Protein Expression and Purification Core Facility, EMBL Heidelberg, Germany) as described in McParland, Varsano et al., 2011. In brief, the samples were dialysed into 20 mM Tris-HCl (pH 7.4), 50 mM NaCl and 1 mM DTT or β-mercaptoethanol. Protein concentration was measured on a NanoDrop 1000 Spectrophotometer, before and after dialysis, and occasionally also verified by SDS-PAGE. In general, at least 0.5 ml of the sample, with minimal protein concentration of 0.2 mg/ml, was needed. The spectra were recorded on Jasco J-715 spectropolarimeter. The instrument was purged with nitrogen for at least 15 min before the measurements. The spectra were acquired across far-UV range (200-250 nm) (value for each wavelength represented an average of three scans). The spectra were measured at 20°C with a path length of 0.1 cm. The scan speed was 10 nm/min. As a control, a sample without the protein (blank) was also measured to check for background absorption. The data were analysed by using GraphPad 5.0 software.

2.2 Materials and methods for Chapter 3.1

2.2.1 Engineering of PRL-3: D72A, I141Y, N142Y, N142A and Q145Y

The mutants of PRL-3 were generated by one-step SDM procedure with Phusion DNA Polymerase and were verified by DNA sequencing (PCR-based SDM is used to change the nucleotides encoding a particular amino acid in the native protein in such a way that this amino acid gets replaced with the amino acid of choice). The primers used are listed below in sense and antisense order. Vector pET15b/wtPRL-3 was used as a template to generate the D72A mutant of PRL-3 by using the following set of primers: 5'-CGTTTGACGCTGGGGCGCC-3' and 5'-GGCGCCCCAGCGTCAAACG-3'. To obtain I141Y, N142Y, N142A and Q145Y, pET15b/PRL-3 ΔCAAX was used as a template. The following pairs of primers were used: I141Y, 5'-CGCGGAGCCTACAACAGCAAGC-3' and 5'-GCTTGCTGTTGTAGGCTCCGCG-3'; N142Y, 5'-GGAGCCATCTACAGCAAGCAGC-3' and 5'-GCTGCTTGCTGTAGATGGCTCC-3'; N142A, 5'-GGAGCCATCGCCAGCAAGCAGC-3' and 5'-GCTGCTTGCTGGCGATGGCTCC-3' and Q145, 5'-AACAGCAAGTACCTCACCTAC-3' and 5'-GTAGGTGAGGTACTTGCTGTT-3'. Plasmid map for vector pET15b is shown in Figure 13 and the summary of the vectors used is given in Figure 19.

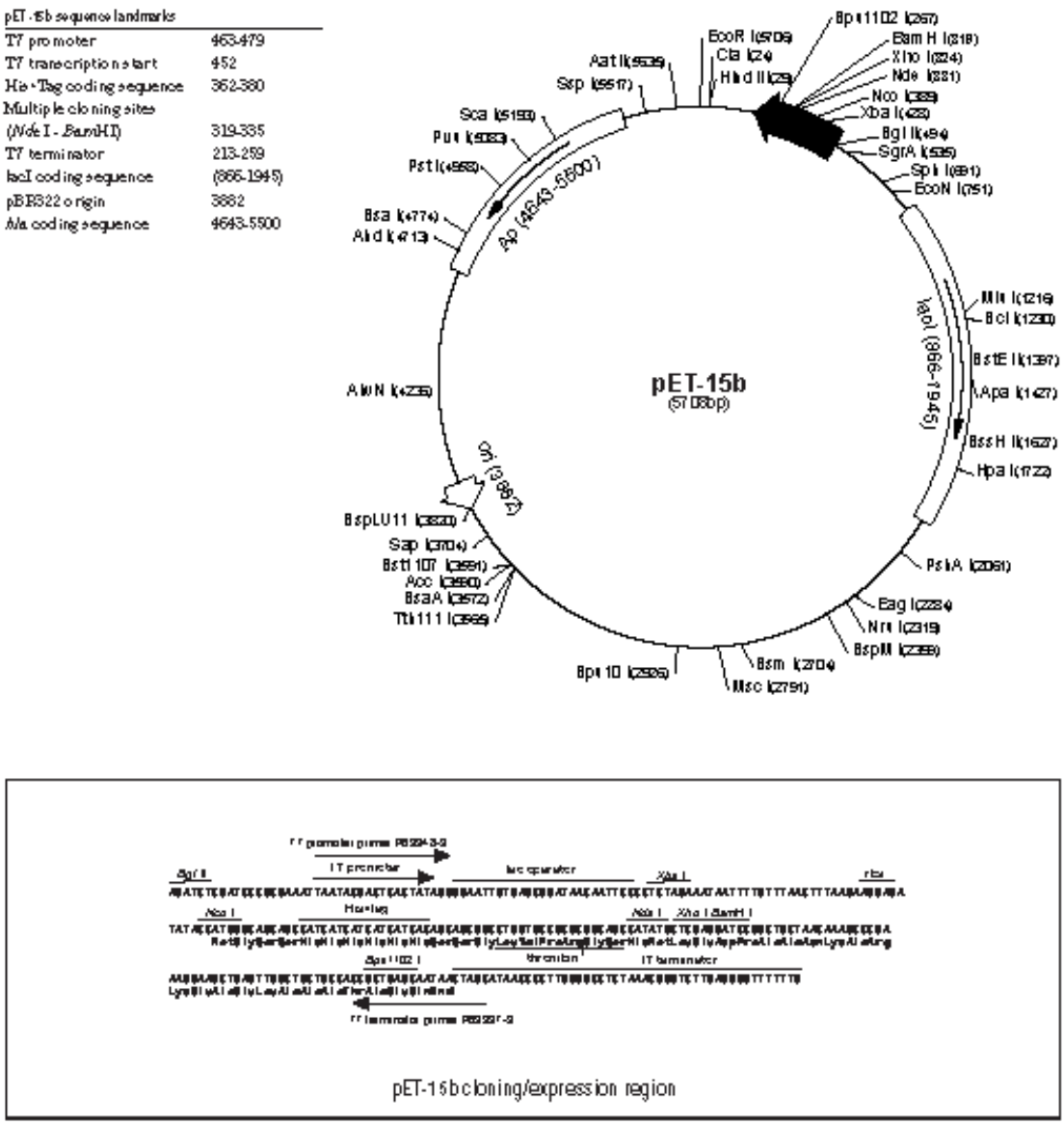


Figure 13: Plasmid map for the pET15b vector. The plasmid was originally from Novagen and is about 5.7 kDa in size. The plasmids encoding full length human PRL-3 protein and the variant with the deleted residues of the C-terminal CAAX-box were generated previously in the lab. The proteins were expressed with the thrombin-cleavable 6His-tag at the N-terminal end. Sequence recognisable by thrombin (LeuValProArg$\downarrow$GlySer, with the arrow indicating cleavage point) was separated by three amino acids from the 6His-tag (SerSerGly), which was preceeded by four amino acids (MetGlySerSer). The sequences encoding all PRL-3 variants were inserted into NdeI-XhoI restriction sites of the vector and in all the cases the 6His-tag was not removed by thrombin cleavage.

2.2.2 Large-scale bacterial expression of the PRL-3 variants from 2.2.1

All the variants of PRL-3 were cloned in pET15b plasmid, which expresses proteins of interest as His_6-tagged fusion proteins. Typically, 0.4-1 l of LB medium was used for large-scale expression. The medium was supplemented with Amp at 100 µg/ml final concentration. After inoculating the medium with the overnight culture of the BL21 DE3 *E. coli* strain, the cultures were incubated at 37°C and 160-200 rpm until optical density measured at 600 nm (OD_{600}) reached 0.6-0.9. Protein expression was induced with 0.1 mM IPTG for 3 h at 37°C; the cells were harvested and the pellets stored at -20°C until purification.

2.2.3 Large-scale purification of the PRL-3 variants from 2.2.2

The pellet was resuspended in an appropriate volume of lysis buffer A containing 50 mM Tris-HCl (pH 7.4), 500 mM NaCl, 20 mM imidazole, 4 mM DTT and PIC (1 x). The cells were lysed by sonication at 4°C (20 rounds of 20 pulses). The lysate was cleared by centrifugation at 4°C and 18,000 *g* for at least 30 min. The cleared lysate was applied to a HisTrap HP column equilibrated in buffer A, by using a peristaltic pump at a flow-rate of around 1 ml/min. The column was washed with buffer B (the same composition as for buffer A, but without PIC). The protein was eluted with gradient of imidazole from 20 mM to 500 mM, by using an Åktapurifier. Purified His_6-tagged proteins were dialysed against 50 mM Tris-HCl (pH 7.4), 500 mM NaCl and 4 mM DTT.

2.2.4 Characterisation of the catalytic activity for PRL-3: I141Y, N142Y, N142A and Q145Y

Preliminary assessments of the catalytic activity for the PRL-3 variants with OMFP and PI(3,4,5)P_3 were done as described in McParland, Varsano et al., 2011. The final volume of the assay reaction was 100 µl. The assays were performed in 96-well plate format in a microplate reader Tecan Saphire 2. The assays conducted without the addition of enzyme (blank) were included. A single triplicate set was measured for each sample. In both assays, wtPRL-3 was used as a comparison for the activity of the mutants. In addition to the wtPRL-3, PTP1B was used as a positive control with I141Y, N142Y, N142A and Q145Y mutants of PRL-3 in OMFP assay. Absorbance (A) over time was plotted by using GraphPad 5.0 software. The values obtained for the triplicate for every sample tested were plotted with the values obtained for the corresponding blank. The data was shown as end-point measurement, plotting absorbance values at the end of the assay.

For OMFP assay, the reaction was monitored as an increase in absorbance at 450 nm, over time (2 hours). All PRL-3 variants were assayed at 6 µM concentration and the concentration of OMFP used was 600 µM. Assays were performed at 37°C in 40 mM Tris-HCl (pH 6.2), 150 mM NaCl and 4 mM DTT; assay conditions had been optimised in the lab previously (McParland, Varsano et al., 2011).

For assay with PI(3,4,5)P_3, commercially available EnzCheck phosphatase assay kit was used. The reaction was monitored by measuring absorbance at 360 nm, over time (2 hours). All PRL-3 variants were assayed at 6 µM concentration and the concentration of PI(3,4,5)P_3 used was 100 µM. Assays were performed at 37°C in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM

$MgCl_2$ and 4 mM DTT; assay conditions had been optimised in the lab previously (McParland, Varsano et al., 2011). The assay kit contains 20 x buffer (the composition is listed above), the substrate 2-amino-6-mercapto-7-methyl-purine riboside (MESG) and purine nucleoside phosphorylase (PNP). MESG is converted by PNP to ribose-1 phosphate and 2-amino-6-mercapto-7-methyl-purine ribose in the presence of a free phosphate (released in the enzymatic reaction), with the accompanying change in the absorption at 360 nm (shown in Figure 14).

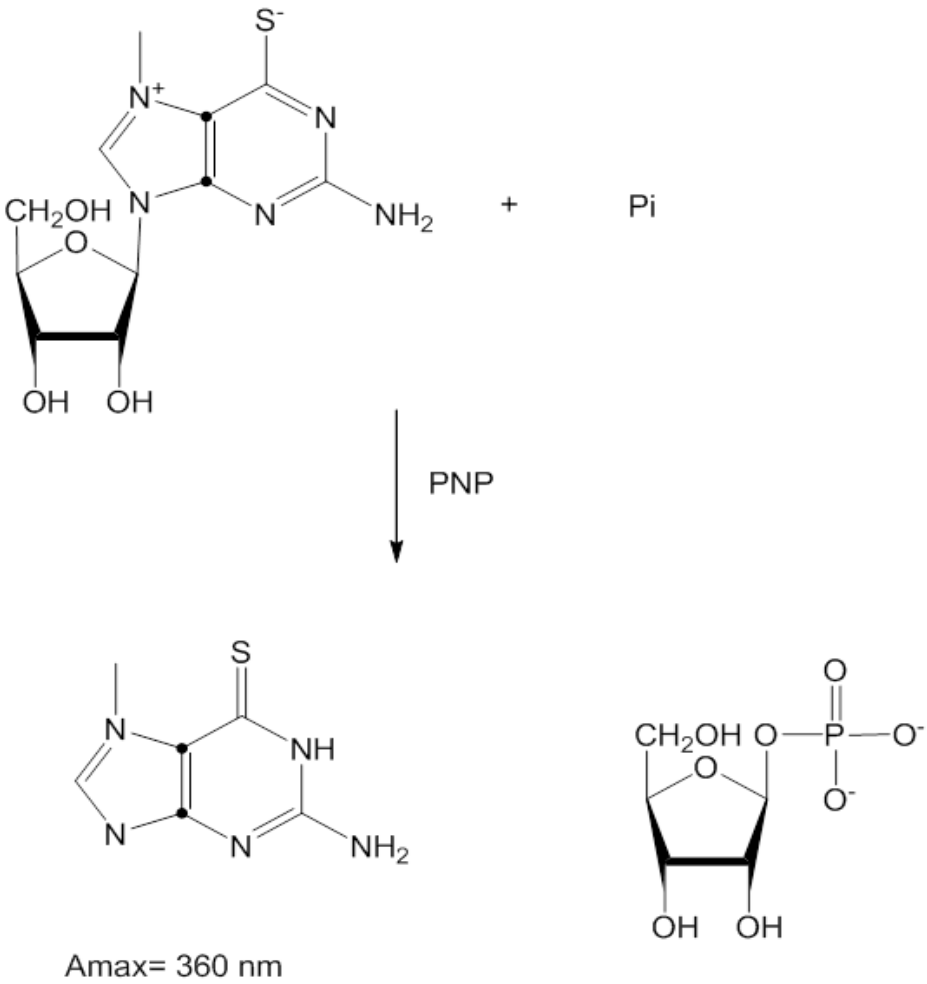


Figure 14: Principle of EnzCheck phosphatase assay kit. The scheme was adapted from the product datasheet provided by the manufacturer and the structures were generated using ChemBioDraw.

2.2.5 Generating the PRL-3 variants for small-scale optimisation of conditions for expression with an unnatural amino acid: wtPRL-3, D72TAG, I141TAG, N142TAG

pTXB3-His_6 plasmid (map shown in Figure 15) was obtained from the laboratory of Dr. E. Lemke. It expresses protein of interest as a C-terminal intein-His_6-tagged fusion protein. Phusion DNA Polymerase was used for PCR amplifications. To generate pTXB3-His_6/wtPRL-3, wtPRL-3 was cloned into NcoI-SapI sites of pTXB3-His_6 plasmid by using pTE15b/wtPRL-3 as a template and the following pair of primers: 5'-GTTGTTCCATGGCTCGGATGAACCGCC-3' and 5'-GGTTGGTTGCTCTTCCGCACATAACGCAGCACCGGGTC-3'. Digestion of the PCR product was conducted with NcoI/SapI enzymes in NEB4 buffer. For wtPRL-3, PCR sample (10 µl) was first digested with NcoI (1 µl) overnight at 37°C, followed by inactivation of the enzyme for 20 min at 65°C. Next, digestion with SapI (2.5 µl) was conducted for at least 1 h at 37°C, followed by inactivation as above, prior to purification from the agarose gel using commercially available kits. Digestion of pTXB3-His_6 was conducted as described above.

D72TAG, I141TAG and N142TAG were obtained by using pTXB3-His_6/wtPRL-3 as a template and the following pairs of sense and antisense primers: 72, 5'-GCCGTTTGACTAGGGGGCGCC-3' and 5'-GGCGCCCCCTAGTCAAACGGC-3'; 141, 5'-CGCGGAGCCTAGAACAGCAAGC-3' and 5'-GCTTGCTGTTCTAGGCTCCGCG-3' and 142, 5'-GGAGCCATCTAGAGCAAGCAGC-3' and 5'-GCTGCTTGCTCTAGATGGCTCC-3. The summary of the vectors used is given in Figure 19.

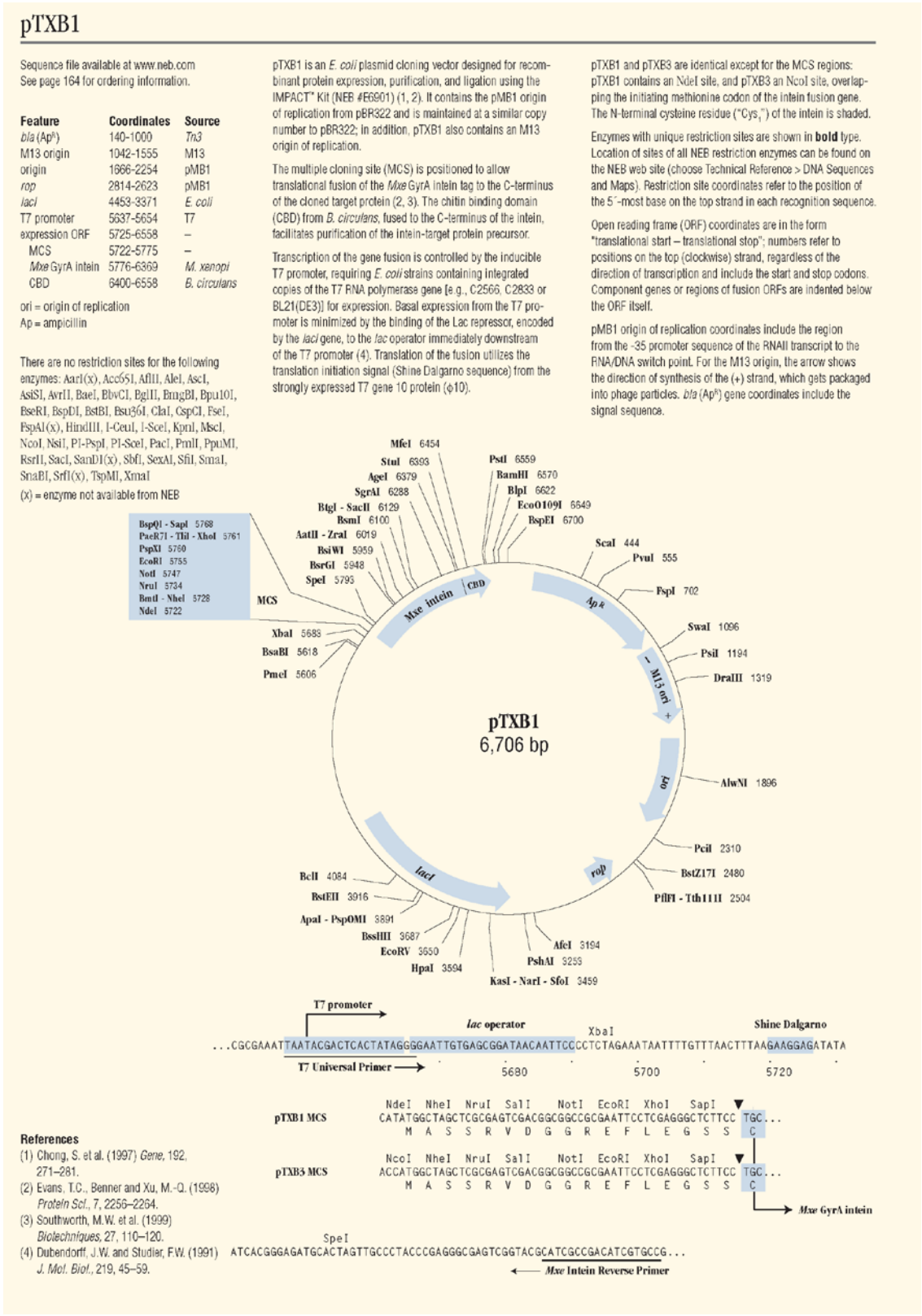


Figure 15: Plasmid map for the pTXB3 vector. The plasmid is commercially available from Life Technologies (former Invitrogen). The plasmid was obtained from the laboratory of Edward Lemke, and it had been modified so that the protein of interest is expressed with the 6His-tag at the C-terminal end instead of the chitin-binding domain. The sequences encoding all PRL-3 and VHR variants were inserted into NcoI-SapI restriction sites of the vector.

2.2.6 Optimising expression and purification conditions with wtPRL-3 on a small scale

Plasmid pTXB3/His_6-wtPRL-3 was transformed by electroporation into BL21 AI and RIL strains of *E. coli* and into chemically competent BL21 DE3 strain.

For small-scale test-expression experiments, 50 ml of LB or TB medium was used, supplemented with the appropriate antibiotics at the following final concentrations: Amp 50 µg/ml, Cm 33 µg/ml and Tet 2.5 µg/ml. After inoculating GM with the overnight culture, the cultures were incubated at 37°C and 160-200 rpm until OD_{600} reached 0.6-0.8. Protein expression was induced with 1 mM IPTG and/or 0.02% (w/v) L-(+)-Arabinose (from 20% (w/v) stock) for 4 h at 37°C and for 6 h at 25°C; the cells were harvested and the pellets stored at -20°C until purification. The non-induced control samples were incubated for 4 h at 37°C.

Purification under native and denaturing conditions was tried.

For purification under native conditions, lysis buffer A contained 50 mM Tris-HCl (pH 7.4), 0.5 M NaCl, 5 mM imidazole, 0.2 mM *tris* (2-carboxyethyl) phosphine (TCEP) and PIC (1 x final concentration). For purification under denaturing conditions, the lysis buffer contained PBS, 4 M urea, 0.25 M NaCl, 5 mM imidazole, 0.2 mM TCEP and PIC, with pH value of 8.5. For both conditions, the imidazole concentration in washing and eluting buffer was 10 and 300 mM, respectively.

The pellet was resuspended in 1 ml of the lysis buffer A and transferred into an Eppendorf tube. The cells were lysed by sonication at 4°C (1 round of 30 pulses). The lysate was cleared by centrifugation at 4°C and 13,200 rpm for 45 min.

The cleared lysate was incubated with His-Select Nickel Affinity Gel; about 100-150 µl of the beads were used for each 1 ml sample of the cleared lysate. The beads were prepared by transferring the appropriate volume into a 15 ml tube and first rinsing in ddH_2O, and then in washing buffer B. The lysate was allowed to incubate with the beads for 1-2 h on a roller at 4°C.

Ni-NTA Superflow and polypropylene columns were used for purification. Before applying the sample, they were first rinsed with one bed volume of ddH_2O, and then with washing buffer B. After applying the lysate, the beads were rinsed with one bed-volume of buffer B. The bound material was eluted by applying 300 µl of eluting buffer C. For analysis by SDS-PAGE, 30 µl of each sample was withdrawn. For intein cleavage, β-mercaptoethanol was added to the remaining samples to reach 50 mM final concentration and left to incubate at 37°C for at least 16 h, with moderate agitation.

2.2.7 Small-scale test-expression with *p*ACF

For small-scale test-expression experiments with the unnatural amino acids, 50 ml of TB-FB medium was used, supplemented with the antibiotics like in 2.2.6.

After inoculating the TB-FB medium with the overnight culture of the BL21 AI strain, the cultures were incubated at 37°C and 160-200 rpm until OD_{600} reached about 0.4-0.5. Then, *p*ACF (structure shown in Figure 16) was added from the stock prepared by dissolving an appropriate amount of the amino acid in ddH_2O; the stock concentration was usually 10 x. *p*ACF was added about 20 min before induction; protein expression was induced for 5-6 h at 25°C by adding 1 mM IPTG and 0.02% (w/v) L-(+)-Arabinose (from 20% (w/v) stock).

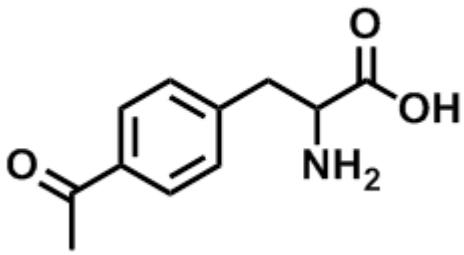


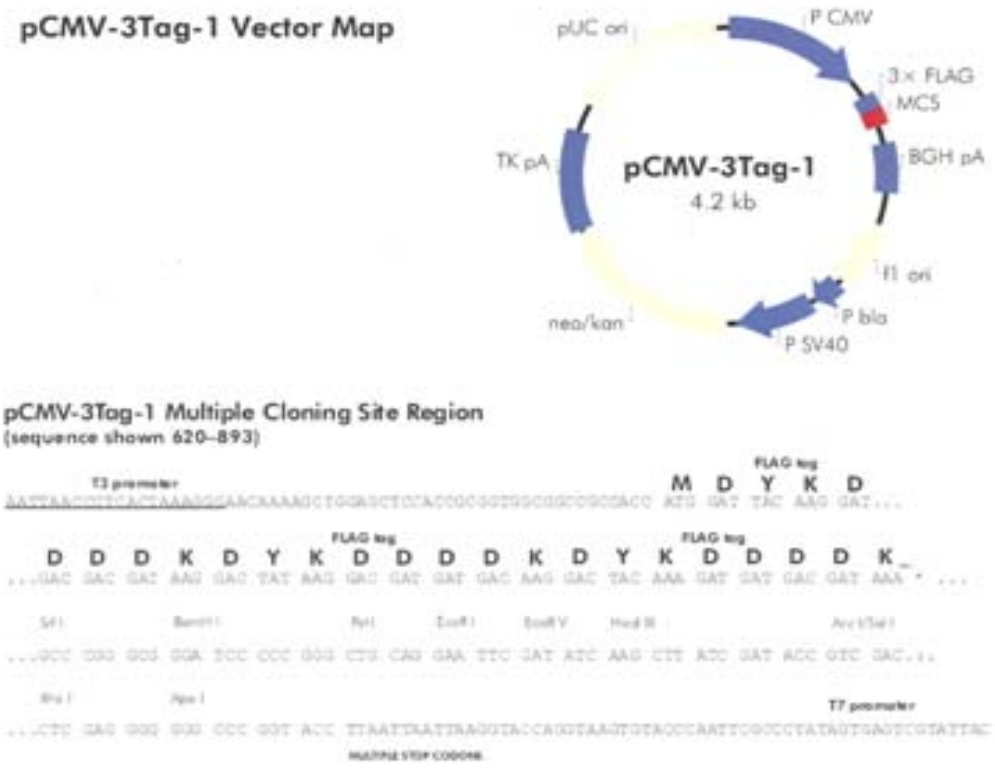
Figure 16: Chemical structure of *p*ACF. The strcuture was generated by using ChemBioDraw.

2.2.8 Small-scale test-expression with *p*BPA

*p*BPA was added either after inoculating the TB-FB medium with the overnight culture of the BL21 AI strain or about 20 min before induction; *p*BPA was prepared by dissolving an appropriate amount in the small volume of 2 M NaOH to reach the desired final concentration of the amino acid, and then filtering it into the medium. The pH value of the medium remained 7-8. The flasks were kept protected from a direct light exposure throughout. The cultures were incubated at 37°C and 160-200 rpm until OD_{600} reached 0.6-0.9. Protein expression was induced with 1 mM IPTG and 0.02% (w/v) L-(+)-Arabinose (from 20% (w/v) stock) for 6 h at 25°C and then overnight at 18°C.

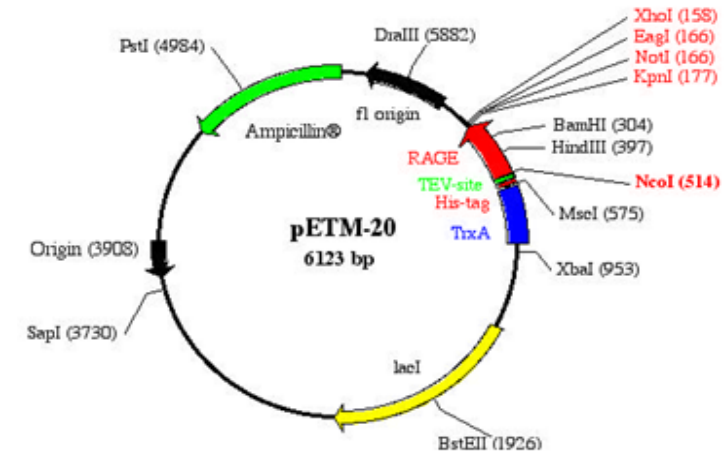
2.2.9 Generating pETM20/6His-3Flag-D72TAG, I141TAG and N142TAG

pETM20/D72TAG was obtained by using pCMV-3Flag-1A/D72TAG as DNA template and Phusion DNA Polymerase to amplify PRL-3-encoding region with the same set of primers containing sites for NcoI/XhoI enzymes like above for wtPRL-3. Plasmid maps for pCMV-3Flag and pETM20 vectors are given in Figures 17 and 18, respectively. PCR was digested with regular NcoI/XhoI enzymes in 2 x Tango buffer for 3 h at rt.



Feature	Nucleotide Position
CMV promoter	1–602
T3 promoter and T3 primer binding site [5' AATTAACCCTCACTAAAGGG 3']	620–639
3× FLAG tag	682–753
multiple cloning site	754–828
T7 promoter and T7 primer binding site [3' CGGGATATCACTCAGCATAATG 5']	872–893
BGH polyA signal	908–1134
f1 origin of ss-DNA replication	1273–1579
bla promoter	1604–1728
SV40 promoter	1748–2086
neomycin/kanamycin resistance ORF	2121–2912
HSV-thymidine kinase (TK) polyA signal	2916–3371
pUC origin	3500–4167

Figure 17: Plasmid map for the pCMV-3Flag-1A vector. The plasmid was originally from Agilent Technologies, but was already available in the lab. The plasmid is about 4.2 kDa in size. It expresses a protein with a triple Flag-tag fused onto the N-terminal end. The plasmids encoding all full length human PRL-3 and VHR variants were inserted into the BamHI-EcoRI restriction sites of the vector.



pETM-20

```
     T7 promoter  -->          Lac operator          XbaI
GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAAT
CTTTAATTATGCTGAGTGATATCCCCTTAACACTCGCCTATTGTTAAGGGGAGATCTTTA

                  rbs
AATTTTGTTTAACTTTAAGAAGGAGATATACATATGAGCGATAAAATTATTCACCTGACT
TTAAAACAAATTGAAATTCTTCCTCTATATGTATACTCGCTATTTTAATAAGTGGACTGA
                                 METSerAspLysIleIleHisLeuThr

                                                  His-tag
GACGACAGT...333bp....GGTTCTGGCCATATGCACCATCATCATCATCATTCTTCT
CTGCTGTCA...TrxA.....CCAAGACCGGTATACGTGGTAGTAGTAGTAGTAAGAAGA
AspAspSer...111aa....GlySerGlyHisMetHisHisHisHisHisHisSerSer

          TEV-site      NcoI
GGTGAGAATCTTTATTTTCAG GGCGCCATGGGTGCTCAAAAC...303bp...GTAGAT
CCACTCTTAGAAATAAAAGTC CCGCGGTACCCACGAGTTTTG....RAGE...CATCTA
GlyGluAsnLeuTyrPheGln|GlyAlaMETGlyAlaGlnAsn...101aa...ValAsp

                  NotI
                  EagI
              KpnI          XhoI      C-His-tag
TCTGCCTCTGAATAAGGTACCGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGAT
AGACGGAGACTTATTCCATGGCGCCGGCGTGAGCTCGTGGTGGTGGTGGTGGTGACTCTA
SerAlaSerGlu***
```

Figure 18: Plasmid map for the pETM20 vector. The plasmid was originally from Protein Expression and Purification Core Facility (EMBL Heidelberg). The plasmid is about 6.1 kDa in size. The initial 123 amino acid, including 111 residues of the Trx chaperone domain, are separated by five amino acids from the following His-tag. Recognition site for the Tobacco Etch Virus (TEV) (GluAsnLeuTyrPheGln↓Gly, with the arrow indicating the cleavage point) is separated by a three-residue-linker (SerSerGly) from the 6His-tag. The plasmids encoding all full length human PRL-3 and VHR variants were inserted into the NcoI-XhoI restriction sites of the vector. The proteins were expressed with the cleavable 6His-tag fused at the N-terminal end, and the 6His-tag was regularly removed by GST-TEV protease.

pCMV-3Flag-1A/D72TAG was generated by using pTXB3-His_6/D72TAG as DNA template, with Phusion DNA Polymerase and the following pair of primers with recognition sites for BamHI/EcoRI restriction enzymes: 5'-CGGGATCCATGGCTCGGATGAACCGCC-3' and 5'-CCGCCGGAATTCTCACATAACGCAGCACCGGG-3'. PCR (30 µl) was digested with regular BamHI/EcoRI enzymes (0.5 µl each) at 37°C overnight.

The same set of primers was used to generate pCMV-3Flag-1A/I141 (or N142) TAG.

For the I141 and N142 mutants (in pCMV-3Flag-1A vector), the second NcoI site (detected between the $Flag_3$-tag and the initial amino acid) was annulled after two rounds of SDM. In the first round, the following pair of primers was used: 5'-GCGGGATCTATGGCTCGGATGAACCGCC-3' and 5'-GAGCCATAGATCCCGCCCGGGCTTTATC-3'. The positive constructs from the first round of SDM were used as DNA template in the second round, with QCL-SDM and the following pair of primers: 5'-GCGGGAAGTATGGCTCGGATGAACCGCC-3' and 5'-GAGCCATACTTCCCGCCCGGGCTTTATC-3'. Positive samples are hereafter referred to as pCMV-3Flag-1A/I141 (or N142) TAG_QCL-SDM. The positive samples were used as a template to amplify PRL-3-encoding region by using the following set of primers containing sites for NcoI/XhoI enzymes: 5'-AATTAACCCTCACTAAAGGG-3' and 5'-CCGCCGCTCGAGTCACATAACGCAGCACCGGG-3'. PCR products were digested with FD NcoI/XhoI and 2 x Tango buffer for 3 h at rt. The summary of the primers used is given in Figure 19.

Primer sequence (5'-3')	Comments
CGCGGAGCCTAC AACAGCAAGC	F. PRL-3 I141Y
GCTTGCTGTTGTAGGCTCCGCG	R. PRL-3 I141Y
GGAGCCATCTACAGCAAGCAGC	F. PRL-3 N142Y
GCTGCTTGCTGTAGATGGCTCC	R. PRL-3 N142Y
AACAGCAAGTACCTCACCTAC	F. PRL-3 Q145Y
GTAGGTGAGGTACTTGCTGTT	R. PRL-3 Q145Y
GGAGCCATCGCCAGCAAGCAGC	F. PRL-3 N142A
GCTGCTTGCTGGCGATGGCTCC	R. PRL-3 N142A
CGCGGAGCCTAGAACAGCAAG	F. PRL-3 I141TAG
GCTTGCTGTTCTAGGCTCCGCG	R. PRL-3 I141TAG
GGAGCCATCTAGAGCAAGCAGC	F. PRL-3 N142TAG
GCTGCTTGCTCTAGATGGCTCC	R. PRL-3 N142TAG
GCCGTTTGACTAGGGGGCGCC	F. PRL-3 D72TAG
GGCGCCCCCTAGTCAAACGGC	R. PRL-3 D72TAG
GTTGTTC↓CATGGCTCGGATGAACCGCC	F. PRL-3 NcoI
GGTTGGTTGCTCTTCC↓GCACATAACGCACCGGGTC	R. PRL-3 SapI
CGG↓GATCCATGGCTCGGATGAACCGCC	F. PRL-3 full length BamHI
CCGCCGG↓AATTCTCACATAACGCAGCACCGGG	R. PRL-3 full length EcoRI
CCGCCGC↓TCGAGTCACATAACGCAGCACCGGG	R. PRL-3 full length XhoI

Figure 19: Summary of the primers used for generating PRL-3 variants expressed in pET15b, pTXB3-intein-6His and pETM20 vectors. The list refers to sections 2.2.1, 2.2.5 and 2.2.9. The sequences subjected to modifications through SDM are indicated in red. For primers introducing recognition site for restriction enzymes, the recognition site is indicated by an arrow. In the comments column, “F.” indicates forward and “R.” reverse set of primers.

2.2.10 Generating VHR variants for small-scale optimisation of conditions for expression with an unnatural amino acid: wtVHR, L25TAG, F68TAG, D92TAG, N163TAG, C124S F68TAG, C124S G161TAG

Vectors pGEX4T1 encoding human wtVHR and its catalytically inactive mutant C124S as the N-terminal GST-fusion proteins were generously provided by Dr. R. Pulido. The pairs of primers in the sections below are listed sense and antisense, respectively.

Phusion DNA Polymerase was used to exchange Phe68 for the amber STOP codon, TAG, by using pTXB3-His_6/wtVHR as the DNA template, and the following pair of primers: 5'-CAGGTCCTAGATGCACGTCAACACCAATGCCAAC-3' and 5'-GTGCATCTAGGACCTGCCCTCAGCC-3'.

The following pair of primers was used to substitute Asp92 for the amber STOP codon by using pTXB3-His_6/wtVHR and Phusion DNA Polymerase: 5'-GCCAACTAGACACAGGAGTTCAACCTGAGCG-3' and 5'-CTGTGTCTAGTTGGCCTTGATGCCCAGG-3'.

N163TAG mutant was generated from pTXB3-His_6/wtVHR as the template by using Phusion DNA Polymerase and the following pair of primers: 5'-GGCCCCTAG GATGGCTTCCTGGCCCAG-3' and 5'-GCCATCCTAGGGGCCGATCTCACGGTTC-3'; pTXB3-His_6/wtVHR was obtained from pTXB3-His_6/L25TAG by using the following pair of primers: 5'-GCTACAGCCTTCCGAGCCAGC-3' and 5'-GCTGGCTCGGAAGGCTGTAGC-3'. To generate pTXB3-His_6/L25TAG, VHR L25TAG was cloned into NcoI-SapI sites of pTXB3-His_6 vector by using the following pair of primers: 5'-GTTGTTCCATGTCGGGCTCGTTCGAGC-3' and 5'-GGTTGGTTGCTCTTCCGCAGGGTTTCAACTTCCCCTCCT-3'; amplification was conducted using Dream Taq Green PCR Master Mix. VHR mutant with Leu25 exchanged for the amber STOP codon was obtained from pGEX4T1/wtVHR by using Phusion DNA Polymerase and the following pair of primers: 5'-GCTACAGCTAGCCGAGCCAGC-3' and 5'-GCTGGCTCGGCTAGCTGTAGC-3'.

The C124S G161TAG mutant was generated from pGEX4T1/C124S by using Ampli Taq Gold 360 DNA Polymerase and the following pair of primers: 5'-CCGTGAGATCTAGCCCAACGATG-3' and 5'-CATCGTTGGGCTAGATCTCACGG-3'. Figure 15 shows the plasmid map for pTXB3 vector and all primers used are summarised in Figure 20.

Cloning the VHR variants in pTXB3-His_6 vector by using NcoI-SapI restriction sites was extremely laborious and subject to constant optimisation. Regular SapI and FD SapI (LguI) were tried, and also regular and HF NcoI, and FD NcoI. The protocols which worked are outlined: (i) Digestion of pTXB3-His_6 vector by using NcoI-SapI enzymes was conducted by incubating an appropriate amount of vector (4 µg) diluted in NEB4 buffer at 1 x final concentration and supplemented with 10% (v/v) BSA first with SapI (2 µl) for 2 h at 37°C, followed by heat-inactivation for 20 min at 65°C; the volume of the digestion mix was 50 µL. The sample was purified by using PCR Purification Kit (Life Technologies). In the final step, 35 µL of 2.5 mM Tris pH 7.9 preheated to 65°C was used to elute DNA. The second digestion step was conducted according to the outline described above, but with HF NcoI (2 µl). (ii) In a total volume of the digestion mix set to 20 µL, 1 µg of pTXB3-His_6 vector was incubated with FD SapI (LguI) (1 µl) for 30 min at 37°C, then heat-inactivated for 5 min at 65°C before adding FD NcoI (1 µl) for 30 min at 37°C, which was heat-inactivated for 15 min at 65°C before purifying from the agarose gel.

The digestion of the PCR products with introduced cleavage sites for NcoI-SapI was performed in a similar way; regular NcoI and SapI enzymes were used, and also prolonged incubation times were allowed.

2.2.11 Generating the VHR variants for large-scale bacterial over-expression: wtVHR, C124S, F68TAG and D92TAG

Plasmid vector pETM20 was used to over-express recombinant human wtVHR and all its variants as N-terminal $Flag_3$-tagged fusion proteins. The mutant VHRs were generated by one-step SDM procedure and were verified by DNA sequencing.

The following pair of primers with introduced cleavage sites for BamHI-EcoRI restriction enzymes was used to amplify the fragments encoding wtVHR, D92TAG (both from pTXB3-His_6), F68TAG and C124S (both from pGEX4T1) by using Phusion DNA polymerase: 5'-GTTGTTGGATCCATGTCGGGCTCGTTCGAGC-3' and 5'-CGGAATTCTCAGGGTTTCAACTTCCCCTC-3'. Then, the PCR fragments were cloned into BamHI-EcoRI sites of pCMV-3Flag-1A vector, thus generating VHRs with the N-terminal $Flag_3$-tag. The following pair of primers with introduced cleavage sites for NcoI-XhoI enzymes was used to amplify the fragments encoding wtVHR, C124S, F68TAG and D92TAG with the N-terminal $Flag_3$-tag by using Phire Hot Start II DNA Polymerase: 5'-AATTAACCCTCACTAAAGGG-3' and 5'-CCGCCGCTCGAGTCAGGGTTTCAACTTCCCCTC-3'. Next, the PCR fragments were cloned into NcoI-XhoI sites of pETM20 vector. Figure 18 shows the plasmid map for pETM20 vector and all primers used are summarised in Figure 20. Digestion of the PCR products was conducted with FD BamHI and EcoRI enzymes in FD Green Buffer at 1 x final concentration for at least 1 h at 37°C. For the C124S mutant of VHR, digestion with FD BamHI-EcoRI was conducted with Tango Buffer at 1 x final concentration. Digestion of the PCR products with FD NcoI and XhoI enzymes was done in Tango Buffer at 2 x final

concentration for about 1 h at 37°C. Plasmids pCMV-3Flag-1A and pETM20 were digested using FD BamHI-EcoRI and NcoI-XhoI, respectively, in FD Green Buffer diluted to 1 x final concentration, at 37°C. Regularly, the amounts of the insert-free plasmid and the PCR products set for digestion were 200 ng and 1 µg, respectively, for which 1 µl of each of the FD enzymes was used. Minimal duration of digestion was 15 min at 37°C.

Primer sequence (5'-3')	Comments
GCTACAGCTAGCCGAGCCAGC	F. VHR L25TAG
GCTGGCTCGGCTAGCTGTAGC	R. VHR L25TAG
GCTACAGCCTTCCGAGCCAGC	F. VHR TAG 25L
GCTGGCTCGGAAGGCTGTAGC	R. VHR TAG 25L
CAGGTCCTAGATGCACGTCAACACCAATGCCAAC	F. VHR F68TAG
GTGCATCTAGGACCTGCCCTCAGCC	R. VHR F68TAG
GCCAACTAGACACAGGAGTTC AACCTCAGC G	F. VHR D92TAG
CTGTGTCTAGTTGGCCTTGATGCCCAGG	R. VHR D92TAG
GGCCCCTAGGATGGCTTCCTGGCCCAG	F. VHR N163TAG
GCCATCCTAGGGGCCGATCTCACGGTTC	R. VHR N163TAG
CCGTGAGATCTAGCCCAACGATG	F. VHR G161TAG
CATCGTTGGGCTAGATCTCACGG	R. VHR G161TAG
GCCAACGCCACACAGGAGTTCAACCTCAG	F. VHR D92A
CTGTGTGGCGTTGGCCTTGATGCCCAGG	R. VHR D92A
GCAGGTCCTGGATGCACGTCAACACCAATGC	F. VHR F68W
CGTGCATCCAGGACCTGCCCTCAGCCG	R. VHR F68W
GCAGGTCCGACATGCACGTCAACACCAATGC	F. VHR F68D
CGTGCATGTCGGACCTGCCCTCAGCCG	R. VHR F68D
GGTCCTTCGCGCACGTCAACACCAATGCCAAC	F. VHR M69A
GACGTGCGCGAAGGACCTGCCCTCAGC	R. VHR M69A
GCAGGTCCTGCATGCACGTCAACACCAATGC	F. VHR F68C
CGTGCATGCAGGACCTGCCCTCAGCCG	R. VHR F68C
GTTGTTC↓CATGTCGGGCTCGTTCGAGC	F. VHR NcoI
GGTTGGTTGCTCTTCC↓GCAGGGTTTCAACTTCCCCTCCT	R. VHR SapI
GTTGTTG↓GATCCATGTCGGGCTCGTTCGAGC	F. VHR BamHI
CGG↓AATTCTCAGGGTTTCAACTTCCCCTC	R. VHR EcoRI
CCGCCGC↓TCGAGTCAGGGTTTCAACTTCCCCTC	R. VHR XhoI

Figure 20: Summary of the primers used for generating VHR variants expressed in pTXB3-intein-6His and pETM20 vectors. The list refers to sections 2.2.10 and 2.2.11. The sequencess subjected to modifications through SDM are indicated in red. For primers introducing recognition site for restriction enzymes, the recognition site is indicated by an arrow. In the comments column, “F.” indicates forward and “R.” reverse set of primers.

2.2.12 Large-scale bacterial expression of recombinant VHRs from 2.2.11

Typically, 0.5-1 l of TB-FB medium was used for large-scale expression of VHR variants. The general outline was as described in section 2.2.8. Expression of the D92*p*BPA mutant was induced at 20°C, for the same total time.

For expression of the wtVHR and its C124S variant, TB-FB medium was supplemented with Amp at 50 µg/ml. Protein expression was induced with 1 mM IPTG for 4 h at 37°C.

2.2.13 Large-scale purification of recombinant VHRs from 2.2.12

The pellet (for wtVHR and F68*p*BPA) was resuspended in an appropriate volume of the lysis buffer A containing 50 mM Tris-HCl (pH 7.4), 500 mM NaCl, 10 mM imidazole, 2 mM DTT and PIC (1 x). The cells were lysed by sonication at 4°C (20 rounds of 30 pulses). The lysate was cleared by centrifugation at 4°C and 12,300 rpm for at least 30 min. The cleared lysate was applied to a HisTrap HP column equilibrated in buffer A, by using a peristaltic pump at a flow-rate of around 1 ml/min. The column was washed with buffer B (the same composition as for buffer A, but without PIC). The protein was eluted with 100% buffer C containing 500 mM imidazole by using an Åktapurifier. Purified His_6-TEV-$Flag_3$-VHR proteins were dialysed against 50 mM Tris-HCl (pH 7.4), 500 mM NaCl and 0.1% (v/v) β-mercaptoethanol. The remaining VHR variants were purified by using the following buffers: buffer A containing 50 mM Na_2HPO_4 (pH 6.4), 200 mM NaCl, 10 mM imidazole, 2 mM DTT and PIC (1 x) and were finally dialysed as above.

The N-terminal His_6-tag was removed by adding GST-TEV protease as described in section 2.1.14 and by incubating at 4°C for at least 24 h.

2.3 Materials and methods for Chapter 3.2

2.3.1 Generating the VHR variants for large-scale bacterial over-expression: F68A, F68D, F68W, F68C, M69A and D92A

Plasmid vector pETM20 was used to over-express all the VHR variants as N-terminal $Flag_3$-tagged fusion proteins. The mutant VHRs were generated by one-step SDM procedure and were verified by DNA sequencing.

The D92A mutant was obtained by using pETM20/wtVHR as the DNA template and the following primers: 5'-GCCAACGCCACACAGGAGTTCAACCTCAG-3' and 5'-CTGTGTGGCGTTGGCCTTGATGCCCAGG-3', and by using Phusion DNA polymerase. Substitution of Phe68 with alanine, aspartate, tryptophan or cysteine, and substitution of Met69 with alanine was conducted by using QCL-SDM kit, employing pETM20/3Flag-wtVHR as the DNA template and the following pairs of the forward and reverse oligonucleotide primers: F68A, 5'-GCAGGTCCGCCATGCACGTCAACACCAATGC-3' and 5'-CGTGCATGGCGGACCTGCCCTCAGCCG-3'; F68D, 5'-GCAGGTCCGACATGCACGTCAACACCAATGC-3' and 5'-CGTGCATGTCGGACCTGCCCTCAGCCG-3'; F68W, 5'-GCAGGTCCTGGATGCACGTCAACACCAATGC-3' and 5'-CGTGCATCCAGGACCTGCCCTCAGCCG-3'; F68C, 5'-GCAGGTCCTGCATGCACGTCAACACCAATGC-3' and 5'-CGTGCATGCAGGACCTGCCCTCAGCCG-3'; M69A, 5'-GGTCCTTCGCGCACGTCAACACCAATGCCAAC-3' and 5'-GACGTGCGCGAAGGACCTGCCCTCAGC-3'. The summary of the vectors

used is given in Figure 20. Digestion of PCR products was conducted as described in section 2.2.11.

2.3.2 Large-scale bacterial expression of recombinant VHRs from 2.3.1

Typically, 0.5-1 l of TB-FB medium was used for large-scale expression of VHR variants. The general outline was as described in section 2.2.8. For expression of the VHR variants F68A, F68D, F68W, F68C, M69A and D92A, TB-FB medium was supplemented with Amp at 50 µg/ml. Protein expression was induced with 1 mM IPTG for 4 h at 37°C.

2.3.3 Large-scale purification of recombinant VHRs from 2.3.2

The general outline was as described in section 2.2.13. The following buffers were used: buffer A containing 50 mM Na_2HPO_4 (pH 6.4), 200 mM NaCl, 10 mM imidazole, 2 mM DTT and PIC (1 x). Purified proteins were finally dialysed exactly as described in section 2.2.13. The N-terminal His_6-tag was removed by adding GST-TEV protease as described in 2.1.14 and by incubating at 4°C for at least 24 h.

2.3.4 Phosphatase activity measurements against *p*NPP

Measurements of the catalytic activity for all the VHR variants generated were performed by Pablo Rios according to the procedure below.

Catalytic activity measurements of the wtVHR and its variants were conducted with *p*NPP. The reaction was monitored as an increase in absorbance at 405 nm, over time. All the VHR variants were assayed at 500 nM concentration and the concentrations of *p*NPP used were 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 80 and 100 mM. Assays were performed at 37°C in 25 mM HEPES (pH 7.5), 124.5 mM NaCl, 2.5 mM EDTA and 2 mM DTT, in a final volume of 100 µl in 96-well

plate format, in a microplate reader Tecan Saphire 2. The reactions were initiated by adding 10 µl *p*NPP, with a multichannel pipette, at the final concentrations indicated above. For every substrate concentration tested, assays conducted without the addition of enzyme (blank) were included. All assays were conducted in triplicate. The minimum of two triplicate sets were conducted for each sample. The mean of the triplicate was subtracted to the mean of its corresponding blank, for every substrate concentration tested. The linear range of every graph, plotted as absorbance over time, was defined in order to calculate the slope values, which were further transformed to picomol (pmol) of product per minute. The data was fitted to the Michaelis-Menten equation to determine the Km value by using the GraphPad 5.0 software. The k_{cat} values were determined manually by dividing the V_{max} value by the enzyme concentration.

In the regular activity assays, *p*NPP was used at 10 mM final concentration and VHR variants at 0.5 mM. The general conditions applied were as described above. Absorbance over time was plotted by using GraphPad 5.0 software. The data was shown as end-point measurement, for the time-point indicated.

For assessing catalytic activity of the monomeric form of F68*p*BPA isolated by size-exclusion chromatography after UV-cross-linking, 0.5 µM of the protein was incubated with 10 mM *p*NPP, as described above. wtVHR was included as a control. Three triplicates were conducted for each sample.

The GraphPad 5.0 software was also used to perform the statystical analysis. A Student's upaired t-test was used to calculate the statystical significance and the asteriscs indicate ***$p<0.001$.

2.3.5 Photo-cross-linking of recombinant VHR with 345 nm light

Photo-cross-linking experiments with purified recombinant VHRs (*in vitro*) or with PRL-3s (with cell lysate from HEK293 cells) were performed in the following setups: (a) 96-well plate (Costar UV plates with UV transparent flat bottom from Corning), (b) multidish 24-well or (c) 35 x 10 mm cell culture dish (both Nunc dishes from Thermo Fisher Scientific), as specified in the appropraite chapters in the Results and discussion section.

(a) For *in vitro* cross-linking, the proteins were diluted in a buffer containing 50 mM Tris-HCl (pH 7.4) and 150 mM NaCl to reach an appropriate protein concentration. The final reaction volume was 50 µl. The samples were kept on ice during 30 min exposure to 345 nm light. The irradiation was conducted at a distance of around 40 cm from the lamp.

PS was prepared according to the manufacturer's instruction and was diluted to a final 1 x concentration. The samples were allowed to incubate with PS on ice for 15 min prior to starting irradiation. PS is a broad-range phosphatase inhibitor. Its composition is not made available, but the tablet was handled according to the datasheet available from the manufacturer, according to which one tablet dissolved in 1 ml of H_2O represents 10 x stock.

In all experiments, the non-irradiated control samples were kept on ice.

Immediately after termination of 345 nm light exposure, the samples were recovered (40 µl) into a clean Eppendorf tube. An aliquot of each sample was withdrawn, mixed with reducing 2 x SDS-PAGE sample buffer and heated at 95°C for 5 min. Equal amounts of proteins (250 ng) were resolved by reducing 4-12% SDS-PAGE and analysed by Western blot (anti-VHR).

(b) In an attempt to discover the identity of the cross-linked peptides (after photo-induced cross-linking of F68*p*BPA variant of VHR), a sample containing F68*p*BPA variant of VHR was concentrated to 150 µl and 2.33 mg/ml. The sample was then subjected to 345 nm light for 30 min, as described in (a) above. After irradiation, 130 µl of the sample was recovered and diluted to 0.7 mg/ml with a buffer containing 50 mM Tris-HCl (pH 7.4), 500 mM NaCl and 0.1% (v/v) β-mercaptoethanol. An aliquot of the sample was mixed with reducing 6 x SDS-PAGE sample buffer and 17.5 µg of protein was resolved by reducing 4-12% SDS-PAGE and allowed to stain by colloidal Coomassie Blue G-250 stain prior to excising the bands of interest (see section 2.3.10).

2.3.6 Cross-linking with glutaraldehyde

Glutaraldehyde was prepared at 0.05% (v/v) in 20 mM HEPES pH 7.4-7.5. Proteins were first prepared at 500 µg/ml by diluting from the stocks into dialysis buffer containing 50 mM Tris-HCl (pH 7.4), 500 mM NaCl and 0.1% (v/v) β-mercaptoethanol. Next, the appropriate protein concentrations (50-250 µg/ml) were achieved by diluting in the buffer containing glutaraldehyde so that to have the final 0.025% (v/v) concentration of glutaraldehyde. The final reaction volume was 50 µl. Cross-linking was performed at 37°C for 30 min. The reaction was quenched by adding 2.5 µl of 1 M Tris-HCl pH 7.5 and by incubating on ice for 5 min. Next, 250 ng of proteins were resolved by SDS-PAGE and further analysed by Western blotting (anti-VHR). The control samples were prepared by diluting the stocks in 20 mM HEPES pH 7.4-7.5 and were kept on ice for the duration of the cross-linking termination step.

As for the choice of the chemical cross-linking reagents, glutaraldehyde is a dialdehyde with five carbon atoms spanning chain-length shown to have the

highest cross-linking reactivity, when compared to formaldehyde or other dialdehydes with up to six carbon atoms in the chain (Migneault et al., 2004). Furthermore, complexes generated by glutaraldehyde are seemingly more thermostable, unlike formaldehyde-generated cross-linking products, which can be reversed if subjected to incubation temperatures above 65°C (Klockenbusch and Kast, 2010). The spacer chain of a cross-linking reagent represents an important feature when it comes to setting distance limitations assessable by a specific cross-linker (Back et al., 2003; Leitner et al., 2010).

2.3.7 Cross-linking with DSS

DSS (structure shown in Figure 21) was stored at 4°C. The product was handled according to the manufacturer's instructions. The cross-linker was always used fresh prepared. The vial was equilibrated to rt before use. An Eppendorf tube was flushed with argon prior to weighing the cross-linker. Next, DSS was dissolved in dry DMSO so that the final concentration of the cross-linker was 25 mM. The protein samples were diluted in 20 mM HEPES pH 7.4-7.5. DSS was added to the protein samples so that the final concentration of the cross-linker was 2.5 mM. The reactions were conducted on a 50 µl scale. The reaction mixtures were incubated at rt for 60 min. The reaction was quenched at rt for 15 min by adding Tris-HCl pH 7.5 from 1 M stock solution to a final 20 mM. The control samples were treated the same way, but without DSS. Next, a sample of the reaction mixture was mixed with a 2 x SDS-PAGE sample buffer and heated at 95°C for 5 min. For immunoblot analysis (anti-VHR), the samples containing 250 ng of protein were resolved by SDS-PAGE.

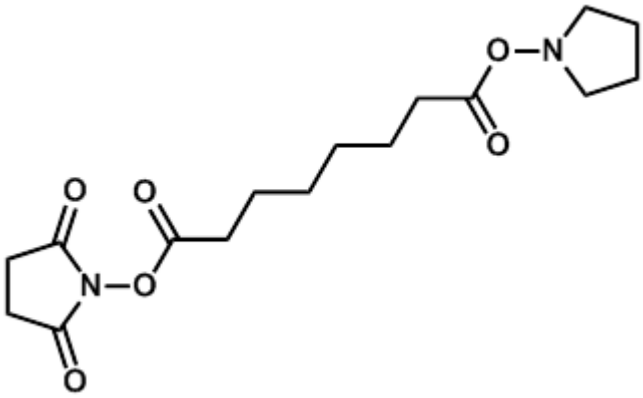


Figure 21: Chemical structure of DSS. DSS is a water-insoluble homo-bi-functional chemical cross-linking reagent. The structure was generated by using ChemBioDraw.

2.3.8 Investigating dimerisation potential of the VHR variants with respect to oxidising conditions

The samples were dialysed into a buffer containing 50 mT Tris-HCl (pH 7.4) and 500 mM NaCl, and with or without 0.1% (v/v) β-mercaptoethanol. The samples were mixed with 2 x SDS-PAGE sample buffer with or without the reducing agent. The samples with reducing agent were incubated at 95°C for 5 min, and without it for 30 min at 37°C. The samples were analysed by SDS-PAGE under standard or non-reducing conditions.

2.3.9 Size-exclusion chromatography

$Flag_3$-F68*p*BPA variant of VHR was subjected to 345 nm light for 30 min by aliquoting it into a 96-well plate setup, with the protein concentration set to about 20 mg/ml (550 µl protein sample). The recovered sample (around 400 µl) was subjected to several rounds of centrifugation (5 min) at 4°C and 4,000 rpm in order to remove any precipitates, before diluting it about 15-fold and setting

dialysis into a fresh buffer containing 50 mM Tris-HCl (pH 7.4), 500 mM NaCl and 0.1% (v/v) β-mercaptoethanol in order to check protein concentration before subjecting the sample to separation by size-exclusion chromatography. Purification was done at the Protein Expression and Purification Core Facility at EMBL Heidelberg. The sample was loaded at 0.5 ml/min on a 16/60 Superdex 75 pg column, and the buffer used was 50 mM Tris-HCl (pH 7.4), 500 mM NaCl and 0.1% (v/v) β-mercaptoethanol. Collected fractions were concentrated (0.22 µm Millex-GP filters), and the concentration was measured by NanoDrop 1000 Spectrophotometer and additionally validated by SDS-PAGE. Fraction enriched for the dimer was subjected for full protein mass determination. Isolated fractions enriched for the monomeric of the protein were pulled and assayed for the catalytic activity against *p*NPP.

2.3.10 LC-MS/MS analysis of the F68*p*BPA variant after exposure to 345 nm light

The analyses were done by Kistina Dzeyk from the Proteomics Core Facility (EMBL Heidelberg). Protein bands were manually excised from the gel and transferred to an Eppendorf tube (0.5 ml). In-gel tryptic digestion was performed as described previously (Rosenfeld et al., 1992). Briefly, gel bands were cut into cubes (1 mm x 1 mm), reduced with 10 mM DTT and alkylated with 55 mM iodoacetamide. Enzymatic digestion with trypsin (Promega) was then performed for 18 h at 37ºC. After acidification to eliminate trypsin activity, peptides were extracted from the gel pieces and concentrated using a SpeedVac. Prior to analysis by LC-MS/MS, peptides were reconstituted in 10 µL 0.1% (v/v) formic acid.

Peptides were separated using the nanoAcquity UPLC system (Waters) fitted with a trapping (nanoAcquity Symmetry C_{18}, 5µm, 180 µm x 20 mm) and an analytical column (nanoAcquity BEH C_{18}, 1.7 µm, 75 µm x 200 mm). The outlet of the analytical column was coupled directly to an LTQ Orbitrap Velos (Thermo Fisher Scientific) using the Proxeon nanospray source. Solvent A was water, 0.1% (v/v) formic acid and solvent B was acetonitrile, 0.1% (v/v) formic acid. The samples (5 µL) were loaded with a constant flow of solvent A at 15 µL/min onto the trapping column. Trapping time was 1 minute. Peptides were eluted via the analytical column at a constant flow of 0.3 µL/min. During the elution step, the percentage of solvent B increased in a linear fashion from 3% to 40% B in 30 min. The peptides were introduced into the mass spectrometer via a Pico-Tip Emitter 360 µm OD x 20 µm ID; 10 µm tip (New Objective) and a spray voltage of 2.1 kV was applied. The capillary temperature was set at 300°C. Full scan MS spectra with mass range 300-1,700 *m/z* were acquired in profile mode in the FT with resolution of 30,000. The ion accumulation time was set at maximum of 500 ms with limitation of 10^6 ions. The 15 most intense ions from the full scan MS were subjected to collision induced dissociation in the LTQ as soon as the precursor intensity exceeded the threshold of 1,000 counts. Normalized collision energy of 40% was used, and the fragmentation was performed after accumulation of 3 x 10^4 ions or after filling time of 50 ms for each precursor ion (whichever occurred first). MS/MS data were acquired in centroid mode. Charge state screening was enabled and only doubly and triply charged precursor ions were selected for MS/MS. The dynamic exclusion list was restricted to 500 entries with maximum retention period of 30 s and relative mass window of 7 ppm. For internal mass calibration, a lock mass correction using a background ion (*m/z* 445.12003) was applied.

Software Max Quant (version 1.0.13.13) was used for generating a peak list, which was converted to mgf format (Cox and Mann, 2008). The MS/MS spectra were searched against Uniprot Human database (86,940 entries), using MASCOT (version 2.2.07) (Matrix Science). Enzyme specificity was set to trypsin and a maximum of two missed cleavages. Cleavage after arginine or lysine, if followed by proline, was allowed. Carbamidomethylation on cysteines was set as fixed modification, methionine oxidation and *p*BPA modification on phenylalanine were included as variable modifications. The search was performed with an initial mass tolerance of 10 ppm for the precursor ion and 0.5 Da for the fragment spectra. Peptides with Mascot score above 20 were reported.

For the analysis of intramolecular cross-linking (see also Figure 22 below), a sequence of VHR was created without methionine on the position 69 and imported in user specific database called UserDB. For the search, a new modification with a mass of 217.0561 Da was predefined in Mascot science server. The modification encloses methionine (131.041 Da) with *p*BPA modification of 104.026Da and without water (18.011Da). The raw-data spectrum shows peaks for the ion series denoted in the table. The colour of each mass value in the table represents whether an ion was matched (red) or not (black) to the theoretically calculated spectrum.

#	b	b^{++}	b^*	b^{*++}	b^0	b^{0++}	Seq.	y	y^{++}	y^*	y^{*++}	y^0	y^{0++}	#
1	88.039	44.523			70.029	35.518	S							12
2	452.164	226.586			434.153	217.580	F	1571.710	786.359	1554.683	777.845	1553.699	777.353	11
3	589.223	295.115			571.212	286.110	H	1207.585	604.296	1190.559	595.783	1189.575	595.291	10
4	688.291	344.649			670.281	335.644	V	1070.527	535.767	1053.500	527.254	1052.516	526.762	9
5	802.334	401.671	785.308	393.157	784.324	392.665	N	971.458	486.233	954.432	477.719	953.448	477.227	8
6	903.382	452.195	886.355	443.681	885.371	443.189	T	857.415	429.211	840.389	420.698	839.405	420.206	7
7	1017.425	509.216	1000.398	500.703	999.414	500.211	N	756.368	378.687	739.341	370.174			6
8	1088.462	544.735	1071.435	536.221	1070.451	535.729	A	642.325	321.666	625.298	313.153			5
9	1202.505	601.756	1185.478	593.243	1184.494	592.751	N	571.287	286.147	554.261	277.634			4
10	1349.573	675.290	1332.547	666.777	1331.563	666.285	F	457.245	229.126	440.218	220.613			3
11	1512.636	756.822	1495.610	748.309	1494.626	747.817	Y	310.176	155.592	293.150	147.078			2
12							K	147.113	74.060	130.086	65.547			1

Figure 22: Analysis of UV-induced intramolecular cross-linking for the F68*p*BPA variant of VHR. After dissociation of the peptide at the amide bond, the C- and N- terminal fragments are formed. The N-terminal fragments are called “b” ions and the corresponding C-terminal fragments are called “y” ions. The nomenclature of the ions is originally described in Roepstorff & Fohlman, 1984. For the analysis of intramolecular cross-linking, a sequence of VHR was created without methionine on the position 69 and imported in user specific database called UserDB. For the search, a new modification with a mass of 217.0561 Da was predefined in Mascot science server. The modification encloses methionine (131.041 Da) with *p*BPA modification of 104.026Da and without water (18.011Da). The raw-data spectrum shows peaks for the ion series denoted in the table. The colour of each mass value in the table represents whether an ion was matched (red) or not (black) to the theoretically calculated spectrum.

2.3.11 Plasmids used for expression in mammalian COS-1 cells

For transient transfection into COS-1 cells, pCMV plasmids encoding wtVHR, and PRL-3 ΔCAAX, all expressed as the N-terminal $Flag_3$-tagged fusion proteins were used; the plasmids encoding PRL-3 ΔCAAX was generated previously in our laboratory by Giulia Varsano.

Dr. Stephen Keyse (CR UK Stress Response Laboratory, Biomedical Research Centre, Dundee, UK) kindly provided modified pSG5 plasmid which expresses protein of interest as the C-terminal myc-tagged fusion protein. First, pSG5 vector was digested with FD EcoRI-XhoI enzymes to remove the existing insert. The sequence encoding wtVHR was amplified from pETM20 vector (100 ng) by using the following pair of primers introducing EcoRI and XhoI restriction sites at the 5'- and 3'-end, respectively, 5'-GTTGTTGAATTCATGTCGGGCTCGTTCGAGC-3' and 5'-CCGCCGCTCGAGGGGTTTCAACTTCCCCTCC-3', and by using Maxima Hot Start PCR Master Mix, like suggested in the product datasheet, and with the initial denaturation step conducted for 15 min at 95°C.

Digestion of both pSG5/myc and PCR product encoding wtVHR with introduced cleavage sites for EcoRI-XhoI was performed with FD enzymes for 2 h at 37°C, after which they were purified from the agarose gel. pSG5/myc was then subjected to a test re-ligation reaction (10 µl) by incubating with T4 DNA ligase (1 µl) for 15 min at rt, followed by transformation into chemically competent TOP10 *E. coli* strain (100 µl) and gave no colonies after overnight incubation at 37°C.

Ligation reaction (10 µl) with EcoRI-XhoI digested pSG5/myc plasmid and insert was performed using T4 DNA Ligase, with an approximate 4:1 molar ratio, for 2 h at rt, and then transformed into TOP10 *E. coli* strain. Test digestion using FD EcoRI-XhoI rendered an insert of the appropriate size. The insertion was checked by DNA sequencing.

2.3.12 Transient transfection of COS-1 cells

For transient transfection, COS-1 cells were seeded in 100-mm cell culture dishes (Corning), so that the confluence of 50-70% would be achieved after incubation for 16-24 h. Fugene HD transfection reagent was used at 6:1 ratio, diluted in Opti-MEM I + GlutaMAX-I (325 µl per 6.5 µg DNA). A mixture was allowed to incubate for 10-15 min at rt and then was added to the cells dropwise. The cells were typically assayed 24-30 h following transfection.

2.3.13 Immunoprecipitation after H_2O_2 stimulated COS-1 cells

Transfected cells were placed in a serum-free medium (SFM) for 2 h. The cells were then treated with 10 mM H_2O_2 for 60 min at 37°C. Next, the cells were rinsed twice with cold PBS and lysed in 1 ml of ice-cold buffer described under 2.1.13. The cells were incubated with the lysis buffer on ice for 10 min, after which the lysate was cleared by centrifugation at 13,200 rpm for 6 min at 4°C. To the cleared cellular extracts, anti-Flag (6 µg) ab was added and incubated for 15-20 h at 4°C on a roller. Next, 40 µl of Dynabeads Protein G was added to each sample and incubated under conditions listed above for further 6 h. After rinsing the beads three times with stock of the lysis buffer (no PIC), the bound material was first eluted by applying 25 µl of 2 x SDS-PAGE sample buffer, without reducing agent, and incubating at 65°C for 10 min. The second elution was performed at 95°C for 5 min.

2.3.14 Cross-linking with PFA in intact COS-1 cells

The cells (60 mm) were transfected with the plasmid pCMV/$Flag_3$-wtVHR. After 24 h, the growth medium was removed and the cells were rinsed twice with pre-warmed PBS. Next, they were incubated with 0.8% (w/v) PFA in PBS at 37°C, after which the cells were rinsed twice with cold PBS. The control dish was incubated in the GM only, at 37°C. The cells were lysed in ice-cold buffer (178.6 µl) containing 20 mM Tris-HCl (pH 7.4), 137 mM NaCl, 10% (v/v) glycerol, 1% (v/v) NP-40, 1 mM EDTA and PIC (1 x). The cells were incubated with the lysis buffer on ice for 10 min, after which the lysate was cleared by centrifugation at 13,200 rpm for 6 min at 4°C. The aliquots were withdrawn, mixed with 2 x SDS-PAGE sample buffer, without reducing agent, and incubated at 60°C for 20 min or at 95°C for 5 min. To the pellets left after lysate clarification, 80 µl of the 2 x SDS-PAGE sample buffer, as above, was added and incubated at 95°C for 5 min. 20 µl of the lysate (about 5.6% of the total lysate) and 10 µl of the resuspended pellets (12.5%) was resolved by 4-12% SDS-PAGE and probed with an anti-Flag ab.

2.3.15 Dephosphorylation of Erk1/2 in COS-1 lysate by recombinant VHR

COS-1 cells were seeded in 100-mm plates and grown to approximately 80-90% confluence. Next, they were placed in SFM for 16-24 h. To inactivate endogenous phosphatases and activate Erk1/2, the cells were subjected to a 30 min treatment at 37°C with 1 mM Na_3VO_4 and 1 mM H_2O_2, which were mixed directly into the SFM. After the treatment, the cells were rinsed twice with cold PBS and lysed in a buffer (500 µl) described under 2.1.13. The cells were scraped, transferred into an Eppendorf tube and disrupted by sonication for 3 s on ice. The extracts were cleared by centrifugation at 13,200 rpm for 10 min at

4°C. The final volume of dephosphorylation reaction was 100 µl. The wtVHR and its F68*p*BPA variant, both with or without prior subjection to 345 nm light for 30 min at 250 µg/ml, were added to have a final protein concentration of 500 nM. The dephosphorylation reaction was conducted at 30°C. At appropriate time points (0, 30 and 90 min), the aliquots were withdrawn and the reactions were quenched by adding 2 x SDS-PAGE sample buffer followed by heating at 95°C for 5 min. Next, 20 µl of each sample was resolved by SDS-PAGE and analysed by Western blotting (anti-VHR, total and pErk1/2).

2.4 Materials and methods for Chapter 3.3

2.4.1 Generating pETM20/6His-3Flag-PRL-1

Hot Start Master Mix was used. PRL-1 was sub-cloned from pETM20 vector into BamHI-EcoRI sites of pCMV-3Flag-1A by using the following pair of sense and antisense primers, respectively: 5'-CGGGATCCATGGCTCGGATGAACCGCC-3' and 5'-CCGCCGGAATTCTTATTGAATGCAACAGTTGTTTC-3'. To generate pETM20/$Flag_3$-PRL-1, the following pair of primers was used: 5'-AATTAACCCTCACTAAAGGG-3' and 5'-CCGCCGCTCGAGTTATTGAATGCAACAGTTGTTTC-3'. Digestion (200 ng DNA) with FD BamHI/EcoRI and NcoI/XhoI enzymes (1 µl each) was done in a FD Green buffer, for at least 30 min at 37°C.

2.4.2 Large-scale bacterial expression and purification of PRL-1

Expression of PRL-1 was induced for 4 h at 37°C and then overnight at 18°C, in 1 l of TB-FB medium. The purification setup was as outlined in section 2.3.3, with the following buffers: lysis and washing buffer A contained 50 mM Tris-HCl (pH 7.4), 200 mM NaCl, PIC, 5% (v/v) glycerol, 1 mM TCEP and 10 mM imidazole and eluting buffer B had 500 mM imidazole. For removal of the His_6-tag with GST-TEV, the protein was concentrated so that 0.8 ml of the sample was loaded onto the HisTRAP column. The sample was finally stored in 50 mM Tris-HCl (pH 7.4), 200 mM NaCl and 4 mM DTT.

2.4.3 Generating pETM20/6His-3Flag-wtPRL-3 and D72A

To annul the second NcoI recognition site identified in pCMV-3Flag-1A/wtPRL-3 (between the $Flag_3$-tag and the initial amino acid), the following set of primers was used with QCL-DSM kit: 5'-GCGGGAAGTATGGCTCGGATGAACCGCC-3' and 5'-GAGCCATACTTCCCGCCCGGGCTTTATC-3'. The construct is hereafter referred to as pCMV-3Flag-1A/wtPRL-3_QCL-SDM.

pCMV-3Flag-1A/wtPRL-3_QCL-SDM was used as DNA template with Phusion DNA Polymerase and the same set of primers containing sites for NcoI/XhoI listed in section 2.2.9.

pETM20/D72A was generated by using pETM20/wtPRL-3 as DNA template Phusion DNA Polymerase and the following set of primers: 5'-GTTTGACGCTGGGGCGCCCCCGCCC-3' and 5'-CGCCCCAGCGTCAAACGGCCAGTCCAC-3'.

2.4.4 Large-scale bacterial expression and purification of the $Flag_3$-tagged PRL-3 variants: wtPRL-3, D72A, D72*p*BPA

The general protocols for purification were as outlined in section 2.3.3. To purify His_6-TEV-$Flag_3$-tagged wtPRL-3 and its D72*p*BPA mutant, the same buffer conditions like for wtVHR and its F68*p*BPA mutant were applied. Occasionally, DTT (2-4 mM) was used in the final buffer. D72A mutant was purified with buffer conditions used for the other VHR variants and was finally dialysed against 20 mM Tris-HCl (pH 7.4), 200 mM NaCl and 0.1% (v/v) β-mercaptoethanol.

2.4.5 Assessing catalytic activity of $Flag_3$-D72*p*BPA with OMFP

The assay was performed as generally described in section 2.2.4. Protein was assayed at 6 μM and OMFP at 50 μM final concentration.

2.4.6 PIP Strips assay

For PIP strips assay, all proteins were in buffer containing 50 mM Tris-HCl (pH 7.4), 500 mM NaCl and 0.1% (v/v) β-mercaptoethanol.

The strips were blocked in TBST-0.1% with 3% (w/v) fatty acid-free BSA (5 ml), for 60 min at rt with gentle agitation. The blocking buffer was discarded and the membrane was rinsed tree times in TBST-0.1% (5 ml), for 5 min each. Next, the proteins were incubated at the final concentration of 0.5 μg/mL in blocking buffer (5 ml) for 30 min at rt, with gentle agitation. The proteins were discarded, the membrane was rinsed three times, as above, and then the appropriate primary ab was added. Anti-Flag or anti-PRL-3 ab were used at 1:5,000 (for 60 min) or 1:400 (for 3 h), respectively. Next, the membrane was rinsed three times, as above, and then the appropriate secondary ab was added at 1:5,000

dilution in blocking buffer and incubated for 60 min. The membrane was rinsed three times, as above, prior to visualization, first with ECL Western Blotting Detection Reagents and then with Western Lightning Plus-ECL Enhanced Chemiluminescence Substrate.

Sample containing D72*p*BPA variant of PRL-3 was incubated with the strip for 10 min prior to exposure to 365 nm, which was conducted for the total time of 5 min (two irradiation intervals of 2.5 min with a 2.5 min break). After irradiation, the sample continued to be incubated with the strip until total incubation time of 30 min. The non-irradiated control sample of D72*p*BPA was kept light protected throughout the assay.

For control without protein, after initial blocking of the membrane, 5 ml of the fresh blocking buffer was added to it and incubated as above.

Purified PRL-3_3A was obtained from Irina Cornaciu and the construct was generated by Miriam Bru Roig, from the Köhn group.

PRL-1 was assayed under two different conditions- in the first assay, protein at 0.5 µg/mL was incubated for 30 min, and in the second assay, the protein was at 5 µg/mL and was incubated with the strip for 60 min.

2.4.7 Incubation of the PRL-3 variants (wt, D72A, D72*p*BPA) with HEK293 lysate

All variants of PRL-3 were in the form of His_6-TEV-$Flag_3$-fusion proteins. For each sample, two 14 cm dishes of fully confluent HEK293 cells were set. Each dish was lysed in 1 ml of lysis buffer described under section 2.1.5.22, scraped and transferred into an Eppendorf tube and then sonicated for 3 s with 20% duty cycle. The lysate was cleared by centrifugation for 15 min at 4°C and

13,200 rpm. To 2 ml of the cleared lysate, 200 µg of the bait was added and incubated on a roller at 4°C for 2 h. Then, the samples were transferred into a 35 mm dish and either exposed to 345 nm light for 30 min or incubated on ice, light-protected, for the same time. Next, the samples were recovered into an Eppendorf tube and subjected to IP using Ni-affinity gel, overnight at 4°C.

The beads were prepared as follows: 200 µl of the beads slurry was pipetted for each sample and rinsed twice with the stock solution of the lysis buffer (no PIC). The samples were centrifuged for 1 min at 4°C and 7,300 rpm. After the IP, the supernatant was recovered and the beads were rinsed twice as above and then one last time for 5 min on a roller at 4°C.

The bound material was eluted first (E1) by adding 100 µl of 2 x SDS PAGE sample buffer, incubating for 10 min at 65°C and 500 rpm and then transferring into a clean tube. The sample was additionally cleared off the residual beads by centrifuging at maximal speed at 4°C for a couple of min. The second elution (E2) was done for 5 min at 95°C and 500 rpm. The samples were analysed by Western blotting (anti-PRL-3), and also subjected for LC-MS/MS analysis, which was done by Kristina Dzeyk from the Proteomics Core Facility (EMBL Heidelberg). The samples (E1) were resolved by loading 60 µl on self-made 8% SDS-PAGE. To prepare 20 ml of the resolving gel, the following components were mixed in the order indicated: ddH_2O (9.3 ml), 1.5 M Tris-HCl pH 8.8 (5.0 ml), 30% acrylamide mix (5.3 ml), 10% SDS (0.2 ml), 10% ammonium persulfate (0.2 ml) and TEMED (0.012 ml). To prepare 10 ml of the stacking gel, the following components were mixed in the order indicated: ddH_2O (6.8 ml), 1.0 M Tris-HCl pH 6.8 (1.26 ml), 30% acrylamide mix (1.66 ml), 10% SDS (0.1 ml), 10% ammonium persulfate (0.1 ml) and TEMED (0.01 ml). The gel was run at about 100 V, stained in colloidal Coomassie overnight and

destained in ddH_2O. For each sample (irradiated and non-irradiated samples of D72A and D72*p*BPA and non-irradiated sample of wtPRL-3), the gel-bands were cut out in three ranges of molecular weight (I ~41-53 kDa; II ~70-93 kDa and III ~130-243 kDa). The raw data obtained from Orbitrap-MS/MS were processed with software MaxQuant, version 1.1.13, which was used for filtering of the data and creating .mgf files, needed for searching in MASCOT version 2.2.03 (Matrix Science). The data were searched against Uniprot_Human database with the following settings: Enzyme specificity was set to trypsin and a maximum of one missed cleavage was allowed. Cysteine carbamidomethylation was used as a fixed modification and methionine oxidation as a variable one. The mass error tolerance for the full scan MS spectra was set at 10 ppm and for the MS/MS spectra at 0.5 Da. The result files (.dat files) were loaded into Scaffold (version 3.00.06) and identifications with a minimum number of peptides per protein of two and 95% confidence in peptide ID were reported. The data were exported to excel. The protein IDs were sorted by molecular weight (biggest to smallest).

3. Results and discussion

3.1 Optimising conditions for expression of the PRL-3 and VHR variants with *p*BPA

3.1.1 Selecting PRL-3 residues to be exchanged for *p*BPA

Selecting PRL-3 residues which would be considered suitable to be exchanged for the photo-cross-linkable amino acid *p*BPA was done by in-depth analysis of the solution structure of PRL-3 (Kozlov et al., 2004) and its comparison to the crystal structure of highly homologous PRL-1 (Jeong et al., 2005; Sun, Wang et al., 2005).

For PRL-1, the structures of both the wt protein and its C104S mutant with the abrogated catalytic activity, in complex with a sulfate ion, were available. The sulfate ion acts as a mimic of a phosphate moiety of a substrate, thus enabling us to analyze what amino acid residues in a protein are important for association with a substrate. The solution structure of PRL-3, described by Kim et al., included the first 162 amino acid residues, whereas the earlier reported solution structure of PRL-3 by Kozlov et al., did not include only the residues of the C-terminal CCAX-box (Kim et al., 2004; Kozlov et al., 2004).

Moreover, whilst PRL-1 crystallized as a trimer, PRL-3 was found to be monomeric. However, the intrinsic ability of PRL-3 to form trimers was demonstrated by performing cross-linking experiments with glutaraldehyde (Sun, Luo et al., 2007). Therefore, at present it can be hypothesized that the monomeric state of PRL-3 revealed in both of its NMR-based solution structure

reports (Kim et al., 2004; Kozlov et al., 2004) could be a result of employed experimental conditions, but also of different tendencies of various PRLs to form dimers or higher-order oligomers, namely the requirement of distinct structural regions. Even PRL-1, for which trimer formation proved essential for exhibiting proper membrane localisation and the stimulatory effect on cell proliferation and migration, showed to be monomeric in solution at concentrations under 3 mg/ml and to contain a mixture of different-state oligomers at higher protein concentrations (Jeong et al., 2005; Sun, Luo et al., 2007). It is interesting to note that several residues of the trimer interface in PRL-1 (Thr13, Arg18, Glu36 and Gly97) were already investigated for their contribution to the trimer formation (Sun, Luo et al., 2007; Sun, Wang et al., 2005). Thr13 in PRL-1 corresponds to serine in both, PRL-2 and PRL-3, and all other residues (Arg18, Glu36 and Gly97) are found conserved across all three PRLs. Whether PRL-2 also has the intrinsic ability to form higher-order oligomers, *in vitro* and/or in cells, has not been investigated yet.

Therefore, it could have been postulated that exchanging analogous residues in PRL-3, namely Ser13/Arg18/Glu36 and/or Gly97, for photo-cross-linkable amino acid *p*BPA, might be beneficial for shedding light on the investigative efforts tackling oligomerization potential of the PRLs. However, when Gly97 in PRL-1 was substituted with Arg, and Arg18 and Glu36 with alanine, the ability of PRL-1 to form trimers was greatly diminished in comparison to the wt protein. Therefore, selecting what residues in the native protein to mutate requires extreme caution. All mutations at the trimer interface described included substitutions of the residues in the native protein with amino acids of considerably different properties, both structural and electrochemical. *p*BPA is hydrophobic and sterically demanding and it is reasonable to believe that

exchanging any of the residues which were reported to be at the trimer interface in PRL-1 might have a similar derogative effect in PRL-3.

Therefore, the attention was focused on the region proximal to the entry into the active site crevice of PRL-3. The three members of the PRLs are highly homologous, showing a high level of sequence identity as well as sequence homology between species (Jeong et al., 2005; Kozlov et al., 2004; Rios et al., 2013; Zeng et al., 1998; Zeng et al., 2003). The surfaces around the active site are highly conserved between PRL-1 and PRL-3 and the regions of the most prominent sequence divergence were found to be localised on the face opposite to the active site pocket, whereas the residues around the P-loop were highly conserved (Jeong et al., 2005; Kozlov et al., 2004). It was postulated that differences in the composition of the amino acids of the loops surrounding the active site pocket are likely to contribute to the distinctions in the portfolio of the interacting partners/substrates amongst the PRLs (Kozlov et al., 2004). Comparing the amino acid sequences of PRL-1 and PRL-3, there are only 14 non-homologous residues between the two phosphatases (Jeong et al., 2005). Excluding the residues of the polybasic region at the C-terminal end of the proteins, the other non-homologous residues are mostly found at the end regions of the loops connecting the secondary structural elements. As for the region in the immediate vicinity to the entrance into the active site pocket, one distinct amino acid resides at position 141, which is occupied by isoleucine in PRL-3 and by phenylalanine in both PRL-1 and PRL-2. In addition, as this residue is localised in the connecting loop between the helices α5 and α6, at the molecular surface, exchanging it for *p*BPA was not considered likely to generate obstructions to the access of a substrate or a PRL-3-interacting partner, in general, to the relevant residues of the active site pocket. The

second two residues, Asn142 and Gln145, were selected based on the same rationale. Unlike position 141, the residues Asn142 and Gln145 are conserved amongst all three PRLs (Sun, Wang et al., 2005). Furthermore, Asn142 and Gln145 offered the advantage of both being localised in the flexible loop. In addition, the general acid/base Asp72 was also selected to be exchanged for *p*BPA. In addition to its functional implications, Asp72 is located on the WPD-loop, which in PRL-3 is very flexible (Kozlov et al., 2004). In contrast, in the structure of PRL-1 in complex with a sulfate ion (Jeong et al., 2005), WPD loop resumes closed conformation, flipping over the entrance to the active site cleft. Therefore, introducing photo-cross-linkable amino acid at the position of the generally conserved aspartate residue would provide further derivatization of the most frequently targeted residues to generate substrate-trap variants of PTPs.

Figure 23 summarises the positioning of the selected residues Ile141, Asn142 and Gln145, first opted to be substituted for tyrosine, within the native PRL-3 protein. The position of the catalytically essential residues, Cys104 and Asp72 (Kozlov et al., 2004), is also pointed out, for comparison of the relative orientations.

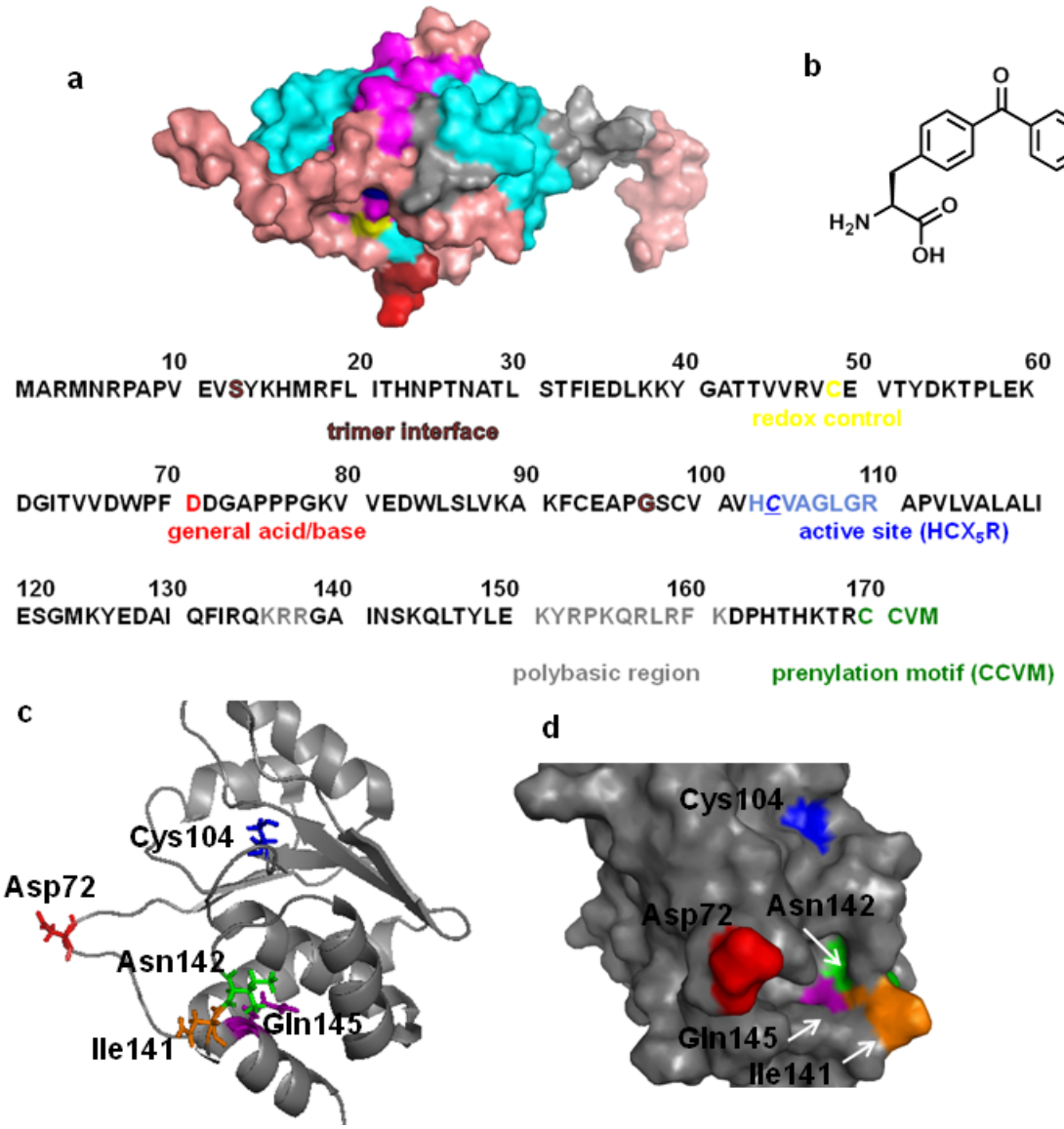


Figure 23: Selecting residues in PRL-3 to be exchanged for *p*BPA. (**a**) Solution structure of PRL-3 (PDB entry: 1R6H) shown as surface representation with colouring based on the secondary structure (helices in blue, sheets in magenta and loops in light pink). Residues important for catalysis, regulation by redox and subcellular localisation, highlighted in the sequence of PRL-3, are pointed out: catalytic Cys104 (blue), regulatory Cys49 (yellow), general acid/base Asp72 (red), and bi-partite polybasic region (grey). The residues of the CAAX box are not present in the structure. (**b**) Chemical structure of *p*BPA, generated in ChemBioDraw. (**c**) Overall solution structure of PRL-3 (PDB entry: 1R6H) presented as a ribbon diagram, with the residues selected to be first exchanged for tyrosine shown as sticks. (**d**) Surface representation of the zoomed-in P-loop of the native PRL-3. For (**c**) and (**d**), the residues are coloured as indicated: Asp72 (red), Cys104 (blue), Ile141 (orange), Asn142 (green) and Gln145 (purple). All structures were generated using Pymol.

To assess if introduction of bulky *p*BPA would be likely to cause structural perturbations, first all three residues (Ile141, Asn142 and Gln145) were decided to be exchanged for tyrosine. When focusing on the set of canonical amino acids, tryptophan is likely to best mimic *p*BPA in terms of hydrophobicity and size of the side chain residue (O'Neill et al., 1989). However, *p*BPA is known to exhibit cross-linking efficiency which is dependent on the geometric constrains, imposed by the distance of the carbonyl moiety of *p*BPA from the surrounding amino acid residues and the rotation angle around the Cα-Cβ bond of *p*BPA, and the electrochemical properties of a potential cross-linking partner carrying the proximal C-H bond to which *p*BPA could covalently cross-link (Dormán and Prestwich, 1994; Tanaka et al., 2008; Wittelsberger et al., 2008). Due to the rotation around the Cα-Cβ bond, *p*BPA has certain geometric plasticity, in contrast to tryptophan. Therefore, trying to mimic *p*BPA by generating tryptophan mutants first might not have rendered a desirable outcome in terms of generating protein variants which would be structurally non-perturbed and still capable to recognise and/or dephosphorylate their putative interacting partners. Therefore, generating PRL-3 variants with tyrosine exchanging residues at positions 141, 142 and 145 was marked as a more appropriate approach. Phenylalanine, albeit smaller and perhaps less suitable than tryptophan in terms of size-comparability to *p*BPA, could better mimic *p*BPA-mediated interactions due to the aforementioned geometric flexibility of *p*BPA. For example, tryptophan might impose a barricade to the entry into the active site pocket due to its fused ring system. On the other hand, the accessibility into the active site crevice for an analogous phenylalanine mutant, carrying a single ring in the side chain group, might be more reflective of the pocket

accessibility for *p*BPA variant, with higher geometric flexibility enabled by the rotation around its Cα-Cβ bond, as mentioned previously.

Furthermore, there is a wealth of literature data to support that exchanging different amino acid residues for *p*BPA reflects in the cross-linking efficiency, namely that some positions are more efficient than the other (Chin and Schultz, 2002; Krishnamurthy, Dugan et al., 2011; Majmudar et al., 2009; O'Neill et al., 1989; Sato et al., 2011; Uezu et al., 2012; Wittelsberger et al., 2006). Therefore, selecting several residues in PRL-3 to be exchanged for *p*BPA was strongly advised. Moreover, it was also noted previously that different *p*BPA-carrying mutants of the same protein exhibited differences in terms of expression efficiency and the total protein yield obtained (Krishnamurthy, Dugan et al., 2011; Sato et al., 2011). In addition, another advantage of selecting several amino acid residues across a protein to be substituted for photo-cross-linkable *p*BPA lies in the opportunity to possibly obtain high-resolution structural information of the two interacting proteins, pinpointing structural regions important for the interaction to occur and amino acid interactions critical to maintain the assembly (Krishnamurthy, Dugan et al., 2011; Forné et al., 2012; Musial-Siwek et al., 2007; O'Neill et al., 1989; Wittelsberger et al., 2006a and b; Wittelsberger et al., 2008). It is obvious that such interpretations of the cross-linking experimental data absolutely requires bioinformatics tools to help identify cross-linked peptides, alongside high resolution tandem MS, like elegantly demonstrated in the recent work of Forné et al., who, by generating a novel software tool termed "Crossfinder", were able to propose a structural model of chromatin remodeling enzyme called ISWI (Forné et al., 2012).

3.1.2 Selecting VHR residues to be exchanged for *p*BPA

VHR (Denu et al., 1995a; Ishibashi et al., 1992) was selected as another representative of the atypical DUSPs (Alonso et al., 2004) in order to demonstrate (i) the general applicability of the expanded genetic code toolkit in developing more efficient substrate traps of PTPs and (ii) to provide a possible quality control for the trapping efficiency of PRL-3.

VHR was known to be one of the closest homologues of PRL-3 and PRL-1, considering the structural similarity (Jeong et al., 2005; Kozlov et al., 2004; Stephens et al., 2005). Structure-based sequence comparison showed that PRLs did not share more than 20% of the amino acid sequence identity with the other members of the classical PTPs (Kozlov et al., 2004). VHR has been widely exploited as a prototypic DUSP in detailed profiling of the catalytic principles of the DUSPs (Denu et al., 1995b; Denu et al., 1996a and b; Hengge et al., 1996; Peters et al., 1998; Zhou et al., 1994), but investigation of its physiological interacting partners or regulatory mechanisms are still sparse (Alonso et al., 2003; Hoyt, Zhu, Cerignoli et al., 2007; Kang and Kim, 2006; Rahmouni et al., 2006; Todd et al., 1999; Todd et al., 2002; Wang et al., 2011; Zheng et al., 2013).

To select residues in VHR which might be suitable for introduction of *p*BPA, detailed analysis of the available structural information was conducted (Schumacher et al., 2002; Wu et al., 2009; Yuvaniyama et al., 1996). The structural data was available for the native protein in complex with a sulfate ion (PDB: 1VHR), catalytically deficient C124S mutant in complex with a peptide derived from the activation loop of p38 MAPK (PDB: 1J4X) and the wt protein in complex with a small molecule inhibitor termed SA3 (PDB: 3F81). In addition,

being one of the structurally most similar PTPs to the members of the PRL family of phosphatases, some comparison with respect to the structural and kinetic characteristic were already made (Jeong et al., 2005; Kozlov et al., 2004; Rios et al., 2013).

Asp92 in VHR serves as a general acid/base (Denu et al., 1995b; Yuvaniyama et al., 1996), therefore representing the functional equivalent to Asp72 in PRL-3. Next, according to the structure-based sequence alignment, Asn163 in VHR represents an equivalent of Asn142 in PRL-3. According to the crystal structure of native VHR with a small molecule inhibitor SA3 (Wu et al., 2009), Leu25 is positioned proximal to the inhibitor. From the perspective of “closed” protein conformation (substrate-bound), Leu25 was regarded as a suitable residue to be tried for substitution with *p*ACF and/or *p*BPA. In addition, Leu25 is a constituent of a surface-exposed hydrophobic patch, which was shown to be a significant component of the substrate-binding region (Wu et al., 2009). Moreover, Leu25 is a constituent of the N-terminal α1-β1 loop (“recognition region”) which is implicated in the formation of the substrate-binding groove. (Yuvaniyama et al., 1996). Phe68, located on the protruding β3-β4 loop of the “variable insert” segment lining the entrance into the active site crevice (Wu et al., 2009; Yuvaniyama et al., 1996) offered an advantage of being structurally similar to *p*ACF and *p*BPA. In addition, it is positioned in the flexible loop, thus minimizing the likelihood that its replacement with either of the unnatural amino acids of interest would cause structural perturbations. With the exception of Met69, the function of the other residues of the “variable insert” section is still elusive (Yuvaniyama et al., 1996). Given the position of Phe68 relative to the active site pocket, exchanging this residue for *p*BPA might help investigating potential involvement of the “variable insert” segment in protein interactions.

Gly161, due to its positioning with respect to the active site pocket, could be regarded as potential negative control when trying to filter out the VHR substrates from its interacting partners.

The residues selected for further optimisation in VHR are shown in Figure 24.

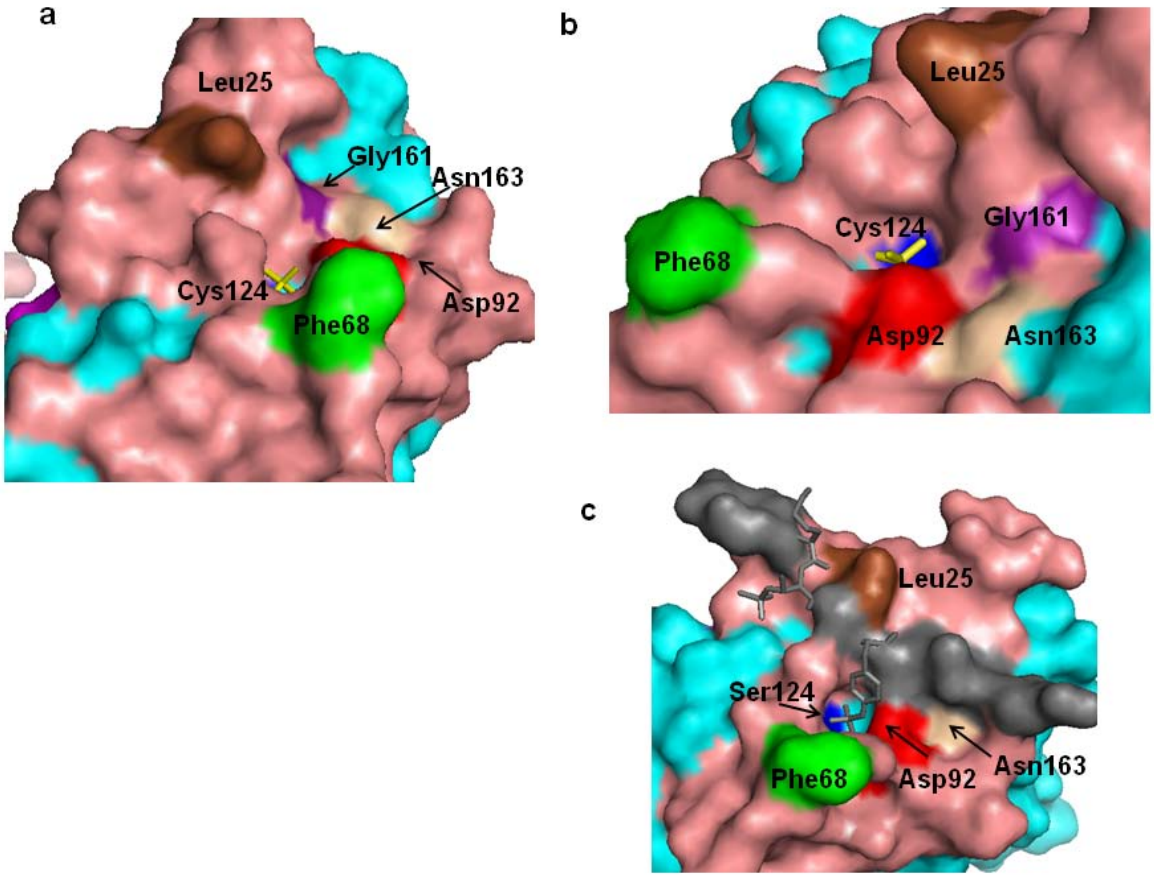


Figure 24: Selecting VHR residues to be exchanged for *p*BPA. (**a**) Surface representation of the crystal structure of VHR (PDB: 1VHR) with colouring based on secondary structure (helices in cyan, sheets in magenta and loops in light pink) and residues selected to be exchanged for the amber STOP codon coloured as indicated: Leu25 (brown), Phe68 (green), Asp92 (red), Cys124 (blue) with the bound sulfate ion shown as a stick representation (yellow), Gly161 (purple) and Asn163 (wheet). (**b**) Zoom-in of (**a**). (**c**) Crystal structure of the catalytically inactive VHR mutant C124S, with peptide derived from p38 MAPK (shown in grey) bound in the active site (PDB:1J4X). The structures were generated using Pymol.

3.1.3 Generating the I141Y, N142Y, N142A and Q145Y mutants of PRL-3

As discussed in detail in section 3.1.1, the residues that were identified as potentially suitable for bearing a bulky photo-cross-linkable amino acid *p*BPA, were first substituted for tyrosine. The general acid/base Asp72 was not mutated to tyrosine because it is located on a flexible loop which might enable introduction of a bulkier residue without compromising structural integrity of the protein. The proteins were over-expressed in *E. coli* as N-terminal His_6-tagged fusion proteins, which enabled affinity-based purification. The protein yields are listed in Table 1, and were calculated as mg of protein obtained from one l of a bacterial culture.

The size of the employed N-terminal His_6-tag is not likely to cause functional impairment of the tagged protein. The reasoning for the N-terminal positioning of the affinity tag is justified by the reported findings on the importance of the residues of the C-terminal end in PRL-3, and PRLs in general, on the subcellular localisation and catalytic activity of the proteins and also on the importance the subcellular localisation pattern imposes on the related signaling pathways (Al-Aidaroos and Zeng, 2010; Pascaru, Tanase et al., 2009; Rios et al., 2013; Skinner et al., 2009; Zeng et al., 2000).

Table 1: Verifying the integrity of the PRL-3 variants (I141Y, N142Y, N142A and Q145Y) by determining full protein mass. The procedure is described in detail in section 2.1.15. The proteins were expressed from 0.4-1 l of LB medium, and protein expression was induced with 0.1 mM IPTG for 3 h at 37°C. The details of the large-scale expression and purification of the aforementioned variants of PRL-3 are provided in section 2.2 and 2.3, respectively.

PRL-3	Protein yield (mg of protein/ l of bacterial culture)	Molecular mass calculated (Da)	Molecular mass obtained (Da)
I141Y	9.1	20,024.80	20,025.00
N142Y	18.98	20,023.85	20,022.43
N142A	25.75	19,930.34	19,932.33
Q145Y	25.85	20,009.82	20,008.51

It can be concluded that all variants of PRL-3 generated expressed at satisfactory and comparable yields, with the exception of I141Y. In addition, only the fourth attempt to purify this mutant was successful. Although the obstacle could in part be attributed to the technical aspect, namely the reduced binding efficiency of Ni HisTRAP affinity columns after a couple of consecutive rounds of purification without stripping and recharging the columns, it could also be indicative of differences in the stability and/or solubility existing between distinct variants of PRL-3. Like mentioned previously, the amino acid composition of the position 141 in PRLs is distinct, with PRL-3 on the one side, having isoleucine, and PRL-1 and PRL-2, on the other one, both carrying phenyalalanine (Jeong et al., 2005; Kim et al., 2004). Thus, exchanging one of only 14 of the non-homologous residues between the PRLs (Jeong et al., 2005) might compromise structural integrity of the protein, disrupting proper folding.

No data is yet available to support a hypothesis that a dramatically reduced yield of the I141Y mutant of PRL-3 in contrast to the other mutants generated, could be a direct implication of the importance of this residue for structural integrity of the PRLs. The key structural characteristics of PRL-3 that are the object of more intense research focus are discussed in section 1.5.4.

Position 142 in PRL-3 corresponds to a potentially promising residue in a more general concept of obtaining substrate-trapping variants of PTPs with better substrate-trapping efficiency (Blanchetot et al., 2005; Tiganis and Bennet, 2007; Zhang, 2003a). The most commonly utilized substrate-trapping mutants of the PTPs include variants with manipulated residues which are acknowledged to be indispensable for efficient hydrolysis of a substrate (for example, exchanging the active site cysteine for a serine or exchanging the aspartate of the general acid loop to an alanine or a glutamine) (Blanchetot et

al., 2005). However, additional PTP variants which could serve as substrate traps were also designed (Blanchetot et al., 2005; Zhang, 2003a). For example, for PTP1B, the successful approach included the Q262A mutant, for which crystal structure of a cysteinylphosphate intermediate was obtained. By comparing the solution structure of PRL-3 with the crystal structure of PTP1B, Asn142 could be equivalent to Asn262. Moreover, like stated previously, the position of Asn142 is conserved amongst the PRLs (Kozlov et al., 2004). Thus, Asn142 in PRL-3 could be lucrative for identifying novel interacting partners of all PRLs as well as for investigating differences amongst them. The N142A mutant of PRL-3 was also generated in the context of the wt background. It was expressed and purified under the conditions used for the previously generated tyrosine-mutants of PRL-3.

3.1.4 Characterisation of PRL-3: I141Y, N142Y, N142A and Q145Y

The next step was to profile the newly generated mutants of PRL-3 to check if any of the mutations introduced disrupted catalytic performance of the variants.

Catalytic performance was assessed in preliminary experiments against the artificial phosphatase substrate OMFP, and then against PI(3,4,5)P_3, as our laboratory discovered the activity of PRL-3 against PIPs (McParland, Varsano et al., 2011; Rios et al., 2013). In all activity assays with OMFP, PTP1B was used as a positive control, because PRL-3 was known to have low catalytic efficiency against artificial phosphatase substrates or synthetic peptides *in vitro* (Kozlov et al., 2004; McParland, Varsano et al., 2011). PTP1B regularly showed about ten times higher activity against OMFP than PRL-3 (it was excluded from the plotting as the focus was on the PRL-3 mutants and PTP1B was used only as an indication that the assay in general worked). The data

from both activity assays was regarded as a rough indication if the mutations introduced affected the catalytic activity of the enzyme. However, performing the activity assay once makes any firm statement about the catalytic performance of the mutants unadvisable as well as stayistical analysis. Comparing all the mutants generated, the catalytic activity of the Q145Y mutant was the lowest in both assays (Figures 25A and 26A). Unlike positions 141 and 142, which reside in the loop connecting the helices α5 and α6 proximal to the entrance into the active site pocket, Gln145 is located within the α6 helix, pointing inwards. Therefore, it is likely that exchanging Gln for sterically more demanding Tyr obstructs access of a substrate into the active site. *p*BPA, being even bulkier, was regarded as likely causing even more destructive effect on the accessibility into the catalytic pocket, so the Gln145 mutant was removed from further screen.

Interestingly, the I141Y mutant exhibited a better catalytic performance than the wt protein in both activity assays (Figures 25B and 26B). As stated previously, this position is non-homologous amongst the PRLs, and suggested to possibly govern substrate specificity (Kozlov et al., 2004). The generated I141Y mutant could have positive effect on substrate binding through hydrogen bonding-mediated interactions enabled via an introduced hydroxyl moiety in the side chain of the tyrosine. These interactions could lead to more optimal positioning of a substrate with respect to the key catalytic residues, thus having positive outcome on the enzyme turnover.

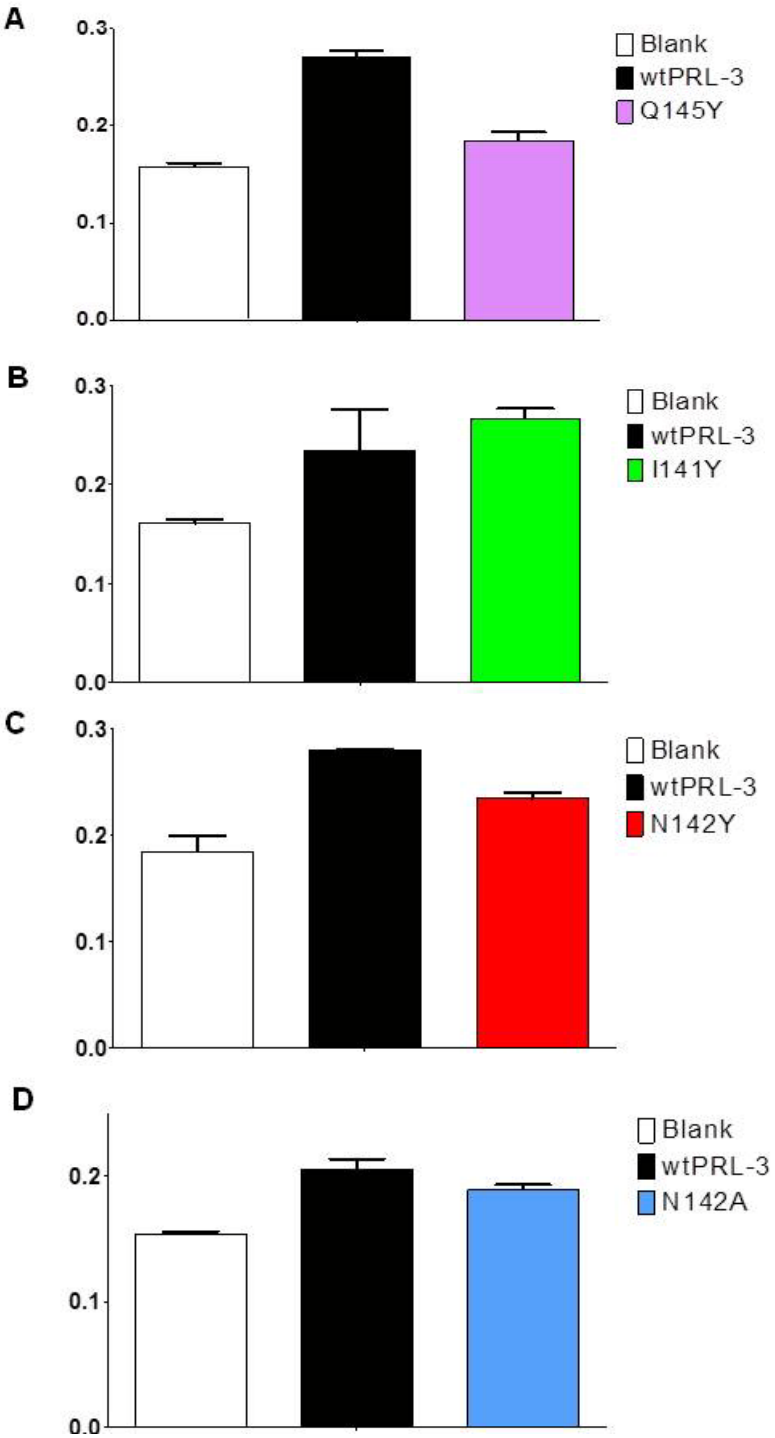


Figure 25: Phosphatase activity of the PRL-3 variants against OMFP. The protocol is described in section 2.2.4. The reaction was monitored as an increase in absorbance at 450 nm, over time (2 hours). All PRL-3 variants were assayed at 6 μM and OMFP was used at 600 μM. Assays were performed according to the conditions had been optimised in the lab previously (McParland, Varsano et al., 2011). The data was plotted using GraphPad Prism 5.0 and shown as end-point measurement (strength of signal as absorbance at 450 nm plotted on y-axis), with standard error of the mean (SEM).

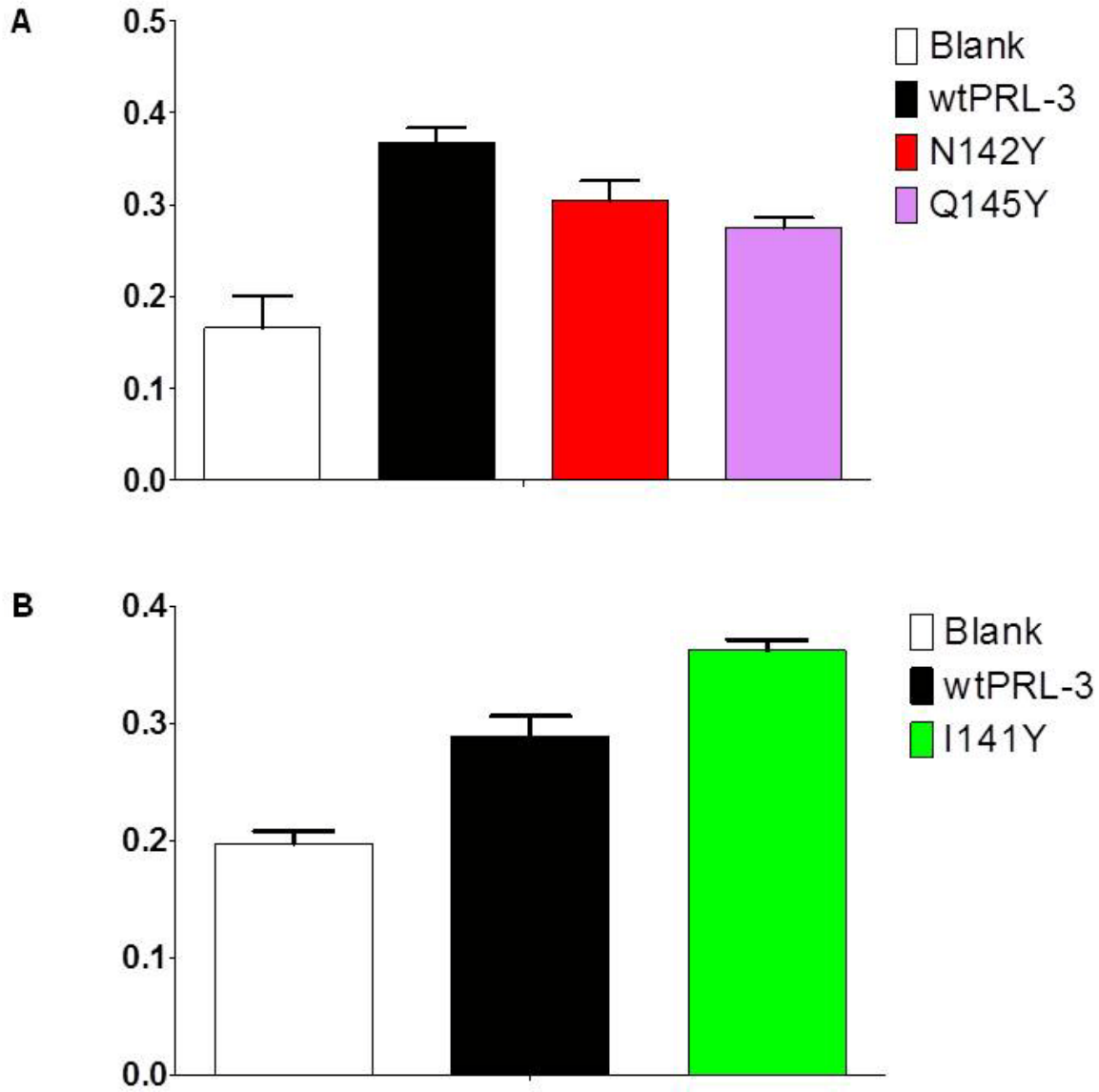


Figure 26: Phosphatase activity of the PRL-3 variants against PI(3,4,5)P_3. The protocol is described in section 2.2.4. Commercially available EnzCheck phosphatase assay kit was used. The reaction was monitored by measuring absorbance at 360 nm, over time (2 hours). All PRL-3 variants were assayed at 6 µM concentration and the concentration of PI(3,4,5)P_3 used was 100 µM. Assays were performed at 37°C in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM $MgCl_2$ and 4 mM DTT; assay conditions had been optimised in the lab previously (McParland, Varsano et al., 2011). The data was plotted using GraphPad Prism 5.0 and shown as end-point measurement (strength of signal as absorbance at 360 nm plotted on y-axis), with SEM.

Alignments of PRL-3 with its closest structural homologues, in the lack of a structure of PRL-3 in complex with a ligand, mostly rely on extrapolating observations for PRL-1 in the analogous procedures (Rios et al., 2013). For PRL-3, substrate-binding pocket is shallower and more hydrophobic when compared to its closest structural homologues VHR and PTEN (Kozlov et al., 2004; Stephens et al., 2005). The patch containing residues Ile141 and Asn142 in PRL-3 is not homologous with Cdc14 or PTEN (Kim et al., 2004). It is reasonably to hypothesize that the location of both Ile141 and Asn142 in the rim at the entrance into the crevice of the active site pocket would put them in good position to mediate governing of a substrate into the catalytic cleft. From the general concept of the project, this would make those two residues well suited for substitution with the photo-cross-linkable unnatural amino acid.

Finally, concluding from the catalytic performance with OMFP of both PRL-3 mutants generated for position 142, this residue, albeit suggested to be an equivalent of Gln262 in PTP1B, failed to match the general criteria for developing substrate trapping mutants of the PTPs (Blanchetot et al., 2005; Zhang, 2003a) as it showed a very similar catalytic activity compared to the wtPRL-3 (Figures 25C, 25D and 26A). N142Y mutant of PRL-3 also showed similar albeit somewhat reduced activity against PI(3,4,5)P_3 (Figure 26A). In a PTP variant supposed to act as a substrate trap, catalytic activity is anticipated to be greatly reduced compared to the unaltered protein, trapping a substrate in the active site pocket for sufficiently long to be isolated and profiled. However, it cannot be excluded that the catalytic profile of a PRL-3 variant carrying mutation at position 142 would not be shifted into that direction when combined with the additional mutation of some of the key catalytic residues (Asp72 or Cys104). Therefore, position 142 in PRL-3 was still regarded as promising for

introduction of a photo-cross-linkable amino acid, but the trial to further profile it as an improved PTP substrate trap in analogy to Q262A mutant of PTP1B, containing only the canonical amino acids, was abandoned.

3.1.5 Small-scale optimisation of conditions for bacterial over-expression with wtPRL-3

To optimize conditions for bacterial over-expression on a small scale, selected PRL-3 variants were expressed as C-terminal intein-His_6-tagged fusion proteins by using pTXB3-His_6 plasmid. Opting for a plasmid which would express PRL-3 variants with an affinity tag at the C-terminal end would enable purification of only the full length protein, and not of the truncated variant which is also produced by employing the amber suppression technology. Therefore, removal of the intein-coupled His_6-tag fused at the C-terminal end of PRL-3 variants by β-mercaptoethanol provided an additional means to check for the expression efficiency of different variants. Potential bacterial contaminants which eventually co-purified with the recombinant protein would not be expected to be affected by β-mercaptoethanol.

The native protein was considered the best starting point to optimize conditions for expression of different variants with selected residues exchanged for the amber STOP codon. Figure 27 summarises the efforts to find the most appropriate experimental conditions, namely the adequate *E. coli* strain and induction temperature. Several trial experiments where expression of pTXB3-intein-6His/wtPRL-3 in different *E. coli* strains and in LB and TB-FB media was induced at 37°C for 4 h and/or the following purification steps under native conditions were not successful and no protein of the expected size was observed by SDS-PAGE analysis.

This was surprising as previous large-scale expression of different N-terminally tagged variants of PRL-3 (D72A, I141Y, N142Y, N142A and Q145Y) did not show any insurmountable difficulties in terms of protein stability, showing that the expression was generally feasible. Furthermore, concentration of imidazole in the lysis and washing buffers was not exceeding 10 mM, and therefore did not hinder binding of the protein to the Ni-containing affinity gel. Another plausible explanation of the failed trail expression and purification experiments could be that bacterial proteins were also present alongside protein of interest. This could represent a problem if the amount of co-purified bacterial proteins exceeded the binding capacity of the affinity gel and/or if the amount of the recombinant protein expressed was low, which would render detection even more difficult. Therefore, purification under denaturing conditions was tried to rule out the possibility that the negative result was the outcome of inefficient solubility of the fused protein.

However, large-scale purification under denaturing conditions would require refolding of the protein *in vitro*, and examining the efficiency of the refolding by performing activity assay. Reductions in the catalytic performance with respect to the non-treated protein sample would not be straightforward to explain, because they could be attributed to incomplete refolding, which would still preserve a denatured fraction of the protein, but could also indicate that the protein was not folded correctly. Thus, purification of the PRL-3 variants under denaturing conditions would be used as the last resort, albeit it was acceptable for small-scale test-expression experiment to rule out the aforementioned obstacles.

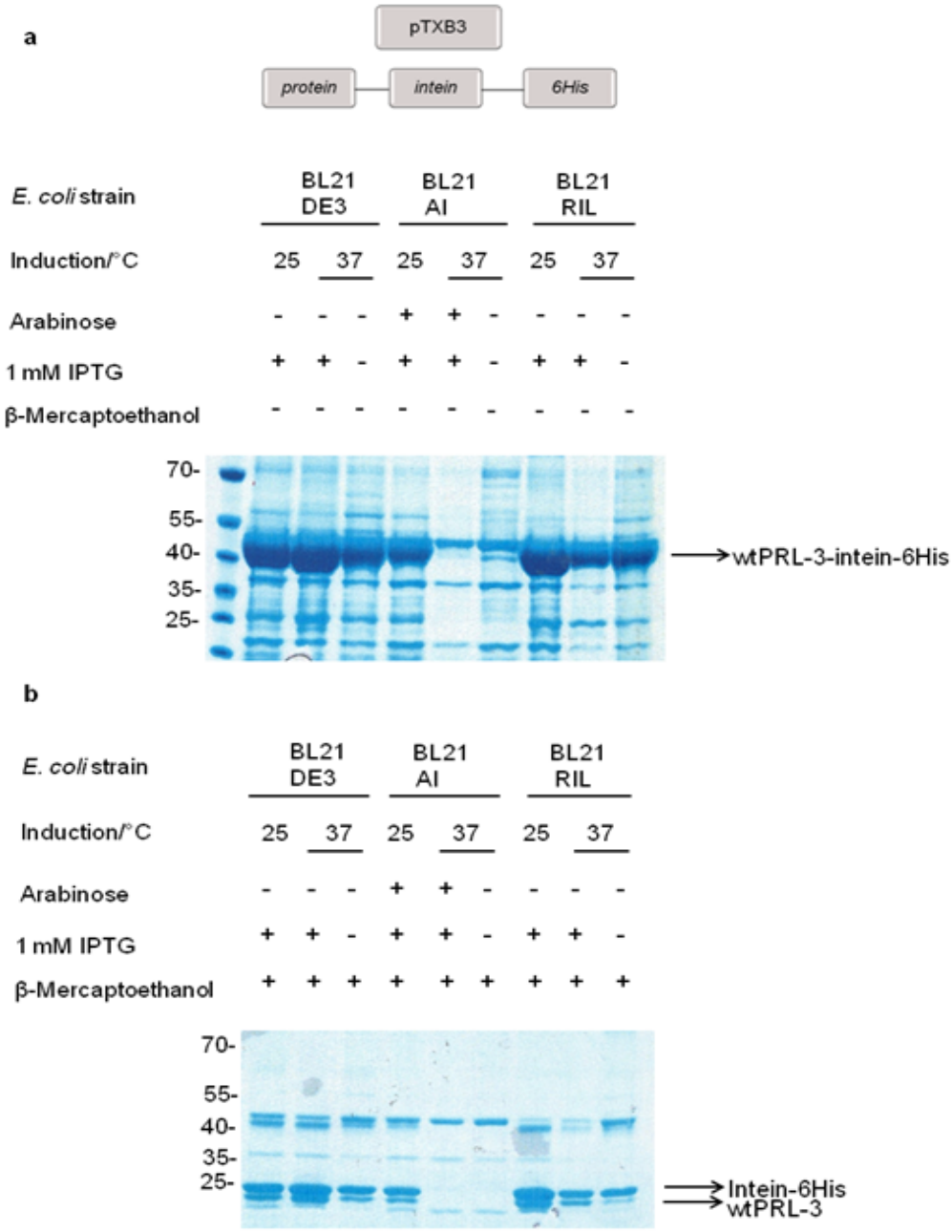


Figure 27: Screening for the optimal *E. coli* strain and induction temperature by using pTXB3-intein-6His/wtPRL-3. (a) Chemically competent BL21 DE3 and electrocompetent BL21 AI and RIL strains were usedProtein expression was induced for 6 h at 25°C or for 3 h at 37°C in the TB-FB medium. The non-induced samples were incubated for 3 h at 37°C. The samples were purified under denaturing conditions. An aliquot of the purified sample was resolved by denaturing 4-12% SDS-PAGE, stained in Coomassie and destained in ddH_2O. (b) Cleavage with 50 mM β-mercaptoethanol was done for 48 h at 4°C after which the samples were resolved by SDS-PAGE, like under (a), without purification. The identity of the other co-purified proteins was not determined.

As indicated on Figure 27, when BL21 AI strain was used, and protein expression induced at 25°C, expression of the fusion PRL-3 protein could be detected by SDS-PAGE analysis. Moreover, the non-induced control sample and the sample of BL21 AI strain with protein expression induced at 37°C exhibited the same background level and lack of discernible expression of the fused wtPRL-3.

In stark contrast were the results obtained with the remaining two strains, chemically competent BL21 DE3 and electrocompetent BL21 RIL. For both of them, the expression of the fused wtPRL-3 protein seemed to be even more efficient than in BL21 AI strain, there was considerable level of “leaky” expression in the non-induced control samples, especially for BL21 DE3 strain.

In the pTXB3-His_6 plasmid, the expression of the gene encoding protein of interest is under control of T7 RNA polymerase promoter, which is predominantly expressed upon induction with IPTG. It has been noted that even under repressive conditions some T7 RNA polymerase could be expressed, thus driving expression of the protein of interest (Gräslund et al., 2008). In the BL21 AI strain, the gene encoding T7 RNA polymerase is under control of arabinose-inducible promoter. From the results presented in Figure 27, it can be concluded that the basal expression of T7 RNA polymerase was low in BL21 AI cells.

BL21 AI electrocompetent strain of *E. coli* was selected as the most optimal for further experiments. TB-FB was selected as the preferred GM.

3.1.6 Small-scale optimisation of conditions for bacterial over-expression of the D72-, I141- and N142TAG variants of PRL-3 with *p*ACF

After having established the basic experimental conditions, namely the strain of *E. coli*, the choice of GM for culturing of bacteria and the induction temperature, by using wtPRL-3, the next step was to translate the setup into optimisation of conditions for expression of PRL-3 variants with the selected positions exchanged for the amber STOP codon.

*p*ACF was considered an excellent way to fine-tune the conditions for the expression of PRL-3 variants with the unnatural amino acid before transferring to *p*BPA. *p*ACF was genetically encoded for expression in *E. coli*, *S. cerevisiae* and mammalian cells (Chin et al., 2003; Liu et al., 2007; Ryu and Schultz, 2006; Ye, Köhrer et al., 2008) and is inexpensive. It contains a carbonyl moiety which acts as a reactive handle to allow additional chemical modifications (Liu and Schultz, 2010). Both *p*ACF and *p*BPA are phenyalalanine derivatives, with *p*BPA having additional phenyl ring comparing to *p*ACF. A critical thing to be optimised was the concentration of the unnatural amino acid which would allow for the significant protein expression level.

In addition to PRL-3 variants with positions 141 and 142 exchanged for the amber codon, additional variant carrying TAG codon at the position of the general acid/base Asp72 was generated. For PRL-3, it was demonstrated that D72A mutant still displayed some catalytic activity in an *in vitro* assay with PIPs $PI(3,4,5)P_3$ and $PI(4,5)P_2$, while showing no activity against robust artificial substrate OMFP (McParland, Varsano et al., 2011). The residual phosphatase activity of PTPs with D→A mutation has been acknowledged (Blanchetot et al., 2005). The similar effect of a mutation of the conserved aspartate was also

reported for PTEN. Although PTEN contains conserved signature motif characteristic for the PTPs, it was shown that it does not employ exactly the same mechanism (Xiao et al., 2007). Generating D92A mutant of PTEN did not significantly affect the phosphoenzyme intermediate formation with $PI(3,4,5)P_3$. Whether lipid and protein phosphatase activities in PRL-3 might occur via distinct mechanistic principles has not been addressed further. The D72*p*ACF variant of PRL-3 expressed with 1 mM final concentration of *p*ACF (Figure 28). The expression of N142 mutant was taken as satisfactory, although the level of the full length protein, after cleavage with β-mercaptoethanol, was not as good as for the D72 mutant. I141 mutant was expressed at the very low level. In addition, the expression with 2 mM *p*ACF was also tried. In this case, I141 mutant of PRL-3 showed an increase in the expression compared to the expression with 1 mM *p*ACF, but the total level of the protein detected remained lower than for the remaining two PRL-3 variants screened, D72- and N142TAG. Therefore, at that stage, all the mutants were carried out onto the next stage of experiments with *p*BPA.

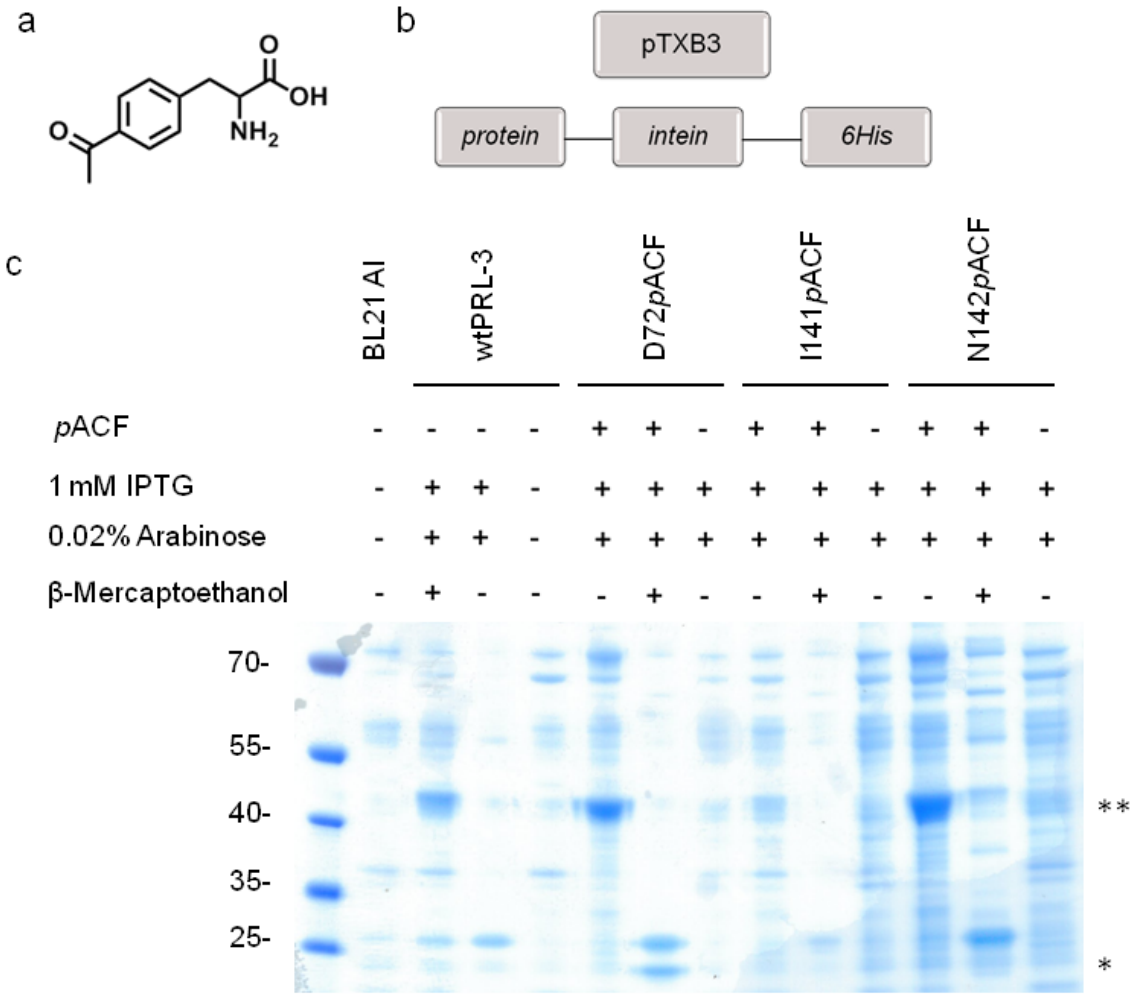


Figure 28: Small-scale expression of the D72-, I141- and N142TAG variants of PRL-3, in pTXB3-intein-6His plasmid, with 1 mM *p*ACF. (a) Chemical structure of *p*ACF generated in ChemBioDraw. (b) Plasmid pTXB3-intein-6His expresses proteins with intein-6His-fused to the C-terminal protein end. (c) Protein expression of wtPRL-3 and of its *p*ACF-containing variants was induced for 6 h at 25°C. The non-induced sample of the wt protein and the sample of BL21 AI cells were incubated for 4 h at 37°C. The samples were purified under native conditions. Cleavage with 50 mM β-mercaptoethanol was done overnight at 37°C. An aliquot of the purified sample was resolved by denaturing 4-12% SDS-PAGE. The gel was placed in Coomassie dye, put in a microwave for about 10 s and then left to stain at rt, followed by destaining in ddH_2O. The band at approximately 20 kDa (*) corresponds to the full length protein and at about 45-50 kDa (**) to the intein fused form; the identity of the other co-purified proteins was not determined.

3.1.7 Small-scale optimisation of conditions for bacterial over-expression of D72-, I141- and N142TAG variants of PRL-3 with *p*BPA

*p*BPA is known to be chemically stable (Chin et al., 2002a) and, like most of the unnatural amino acids, is readily taken up by prokaryotic or eukaryotic cells (Liu and Schultz, 2010).

In order to try to maximize the yield of the full length protein with selectively introduced unnatural amino acid, conditions for induction of protein expression were varied, as described for Figure 29. The I141TAG mutant demonstrated poor suppression efficiency with *p*BPA, like in the case of *p*ACF, thus ruling out that the low yield of the full length protein detected was the effect of the nature of the unnatural amino acid (Hino et al., 2012). The N142TAG mutant showed a good expression level, like marked with a red circle mark in Figure 29, albeit the concentration of *p*BPA in the medium had to be increased to 2.5 mM. In case of Asn142 mutant, likely additional advantage was allowing prolonged induction time. Lowering temperature to 18°C for the overnight incubation reduced the speed of protein production and allowed proper protein folding (Gräslund et al., 2008) and showed to be beneficial for the expression of the N142TAG variant.

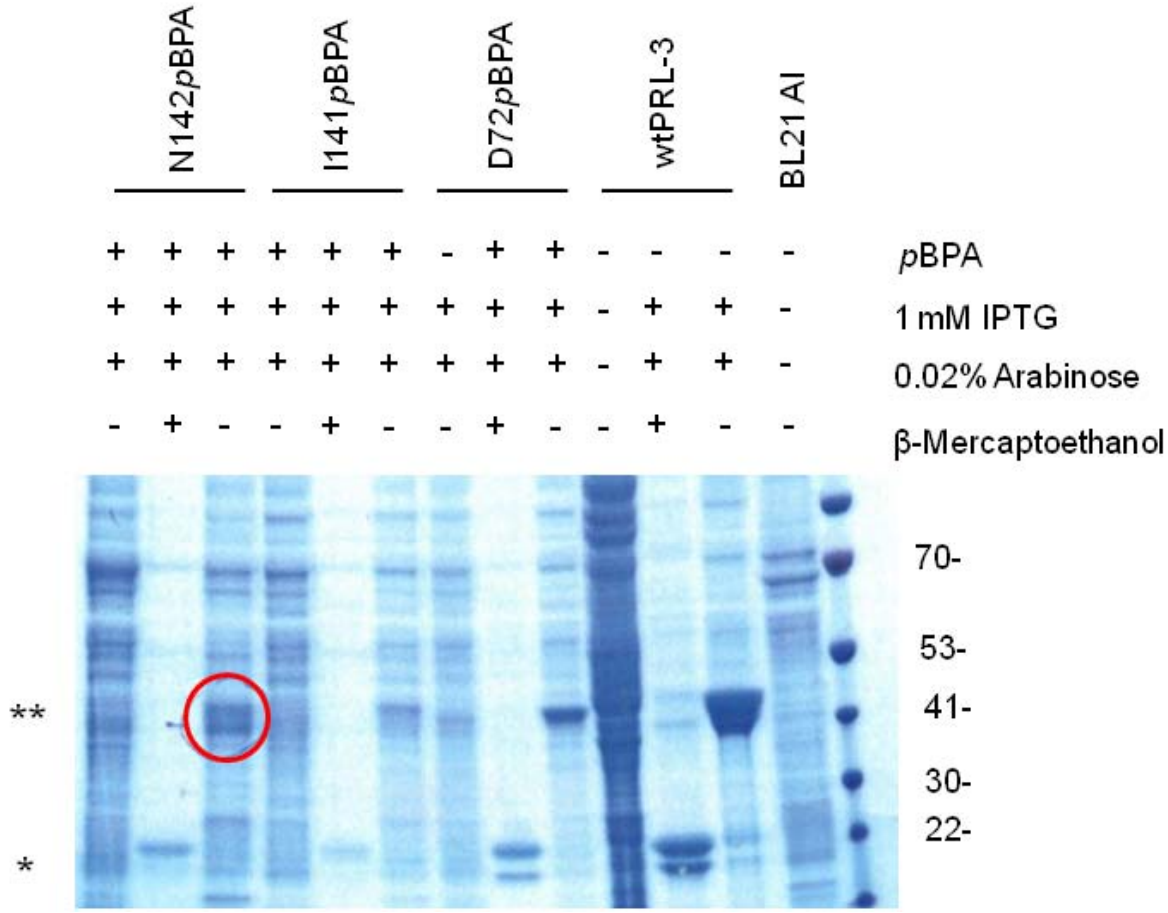


Figure 29: Small-scale test-expression of the PRL-3 variants, in pTXB3-intein-6His plasmid (see Figure 28b), with *p*BPA. BL21 AI cells were used as a background control. The D72*p*BPA mutant was expressed with 1 mM final concentration of *p*BPA, filtered into the TB-FB GM at the beginning of the experiment. For the I141- and N142*p*BPA mutants, *p*BPA was added to the GM about 20 min before induction at the final concentration of 2.5 mM. For the D72- and I141*p*BPA mutants, expression was induced for 6 h at 25°C; for the N142*p*BPA mutant, it was induced for 6 h at 25°C and then overnight at 18°C. Expression of the wtPRL-3 was induced for 4 h at 37°C. The samples were purified under native conditions. An aliquot of each sample was taken for SDS-PAGE analysis and to the rest of the purified sample β-mercaptoethanol was added. Intein-6His cleavage was conducted for 36 h at 37°C, after which 20 µl of the final mixture, without additional purification, was resolved by denaturing 4-12% SDS-PAGE. The band at about 20 kDa (*) corresponds to the full length protein and at about 45-50 kDa (**) to the intein fused form. The identity of the other co-purified proteins was not determined.

The D72TAG variant of PRL-3 was further subjected to a trial where protein expression was further allowed to proceed by overnight incubation at 18°C to assess if prolonged incubation time would negatively affect protein stability or, like in case of the N142 mutant, improve the overall yield of the full length protein produced.

Like indicated in Figure 30, the mutant exhibited excellent expression efficiency, without observable readthrough of the STOP codon in the control sample in the absence of the addition of *p*BPA into the GM. In addition, the amount of the full length protein detected after prolonged induction time was higher (Figure 30) than after only 6 h incubation at 25°C (Figure 29). No observable effect on the overall purity of the sample was detected.

The D72 mutant was the only variant of PRL-3 for which excellent yield of *p*BPA-containing protein was detected already at low level of *p*BPA in the medium (1 mM). It was noted previously that different unnatural amino acids have different optimal concentration ranges for successful protein expression and that higher concentrations of the unnatural amino acid in the GM could negatively affect cell growth (Hino et al., 2012). Moreover, the extensive study for the unnatural amino acid uptake from the medium, which profiled a total of 138 amino acids, showed that the highest toxicity was linked to amino acids with reactive functional groups and to those which were structurally similar to the canonical ones (Liu and Schultz, 1999). *p*BPA was not in that screen and for *p*ACF no toxicity was reported.

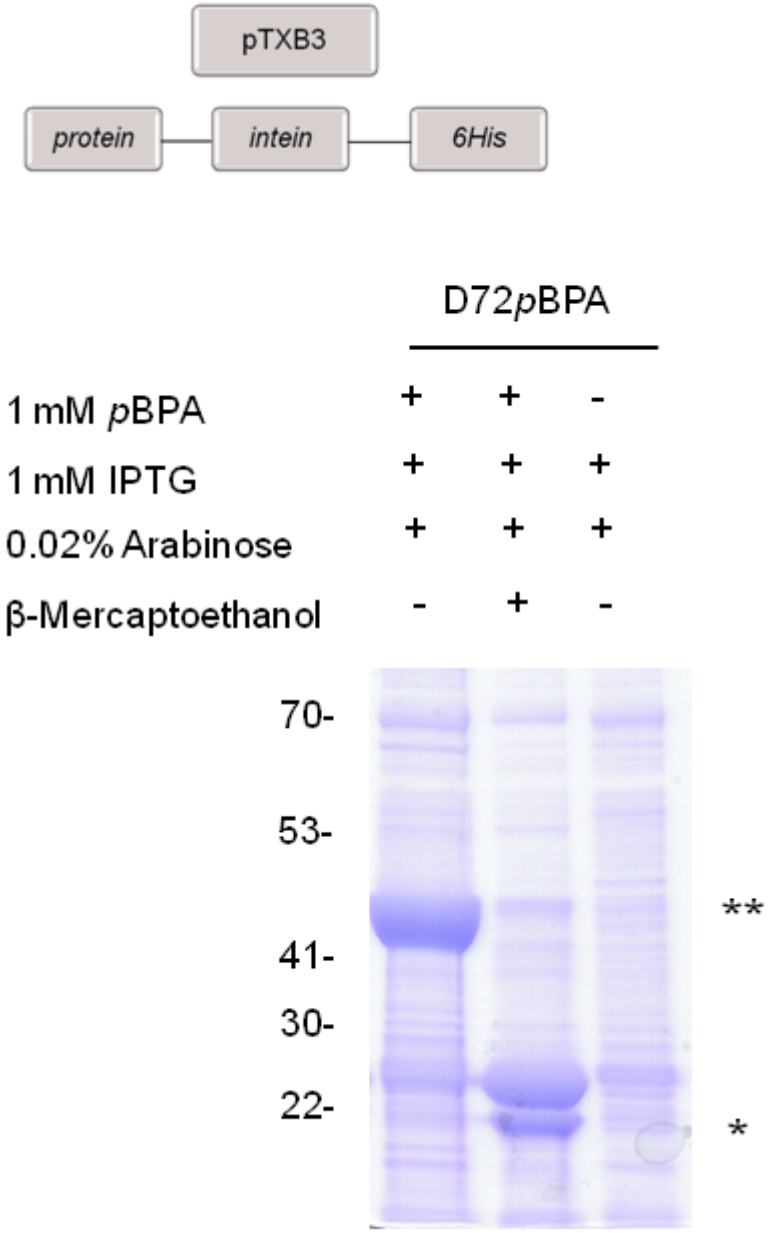


Figure 30: Small-scale expression of the PRL-3 D72*p*BPA mutant, in pTXB3-intein-6His plasmid, under optimised conditions. *p*BPA was filtered into the TB-FB GM at the beginning of the experiment, to reach the final concentration of 1 mM. Protein expression was induced for 6 h at 25°C and then overnight at 18°C. Intein cleavage was done with final concentration of 50 mM β-mercaptoethanol overnight at 37°C. 20 µl of the final mixture, without additional purification, was resolved by denaturing 4-12% SDS-PAGE, stained in Coomassie and destained in ddH_2O. The band at about 20 kDa (*) corresponds to the full length protein and at about 45-50 kDa (**) to the intein fused form. The identity of the other co-purified proteins was not determined.

3.1.8 Small-scale expression of VHR variants with *p*ACF and *p*BPA

For small-scale expression and purification experiments with different variants of VHR, the protocols developed with PRL-3 were successfully translated. A trial experiment with 1 mM *p*ACF and the L25 mutant of VHR was not clearly indicative of the successful expression of the full-length protein, so the concentration of the amino acid was increased to 2 mM, and all the variants were subjected to these conditions. As shown in Figure 31, the F68 and N163 mutants exhibited the highest expression level. According to the previously elaborated observations for PRL-3, it can be hypothesized that introducing a bulky and hydrophobic amino acid at positions 25 and 161 does not act in favour of proper protein folding, although both amino acids are located in the loop regions, which was anticipated to allow for some flexibility in accommodating a sterically more demanding unnatural amino acids. This seems to be even more likely cause in case of the L25 mutant, which is, as already mentioned, part of a surface-exposed hydrophobic patch, so introduction of *p*ACF is not expected to negatively affect electrostatic charge distribution in the region.

Nevertheless, all variants were also tried in the trial experiment with *p*BPA (Figure 32). Based on the results obtained with *p*ACF, no lower concentrations than 2.5 mM of *p*BPA were screened. Unlike the generally satisfactory results obtained with *p*ACF for all the mutants of VHR screened, only the F68 mutant, and to a lesser extent the L25 and G161 mutants, exhibited good expression level with *p*BPA. Taking into consideration the position of these residues, F68 was selected for the large-scale expression in *E. coli*. In addition, the D92 mutant also demonstrated expression level comparable to F68 mutant, under the same conditions, and was selected for up-scaled expression.

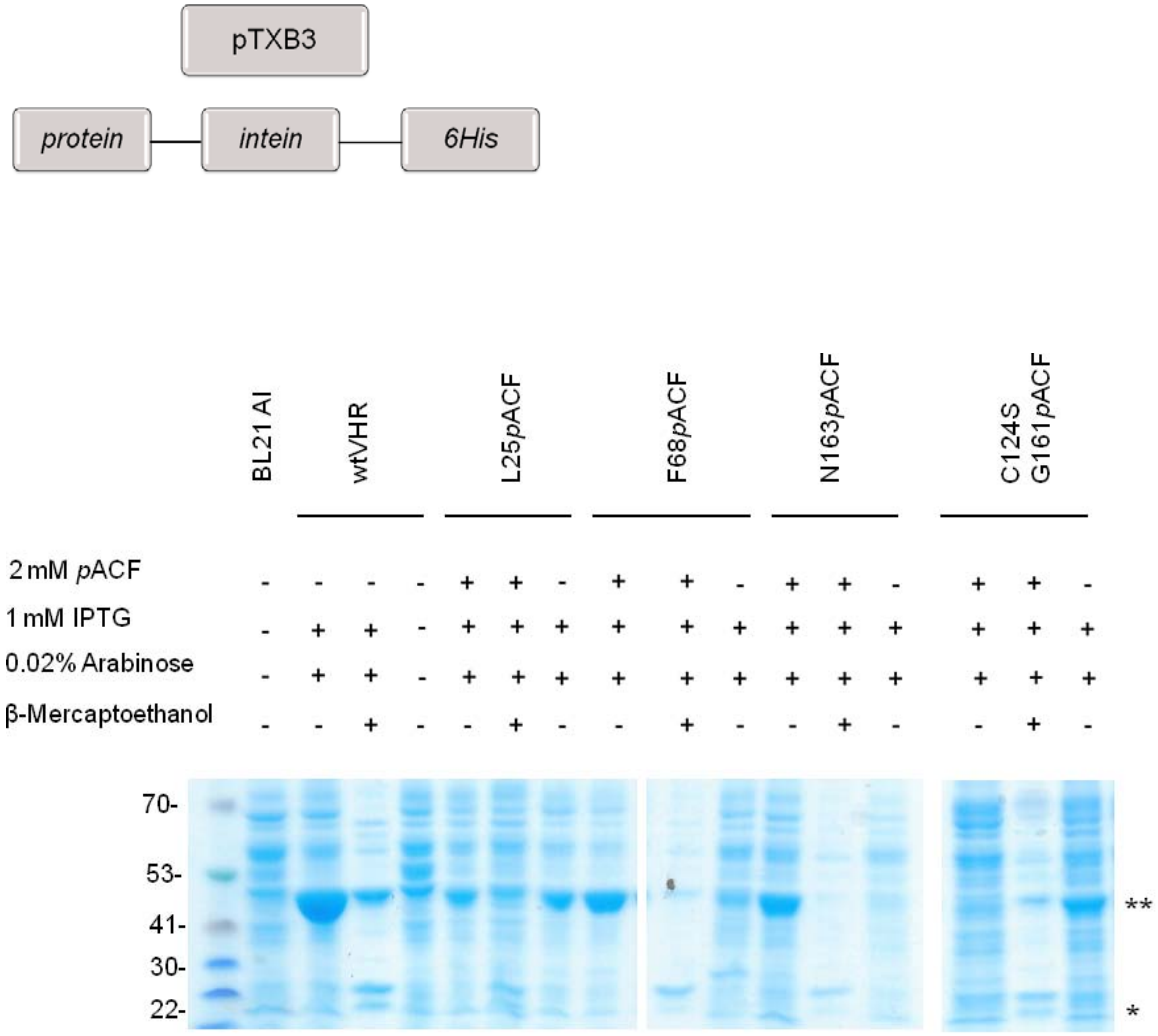


Figure 31: Small-scale expression of the VHR variants with *p*ACF at 2 mM final concentration in pTXB3-intein-6His plasmid. *p*ACF was added to the medium about 20 min before induction. Protein expression was induced for 6 h at 25°C. The samples were purified under native conditions. An aliquot of each sample was taken for SDS-PAGE analysis and to the rest of the purified sample β-mercaptoethanol was added. Intein cleavage was done with final concentration of 50 mM β-mercaptoethanol for 48 h at 37°C. 20 µl of the final mixture, without additional purification, was resolved by denaturing 4-12% SDS-PAGE. The band at approximately 50 kDa (**) corresponds to the full length fusion protein and at about 20 kDa to the full length protein after β-mercaptoethanol-induced intein cleavage (*).The identity of the other co-purified proteins was not determined.

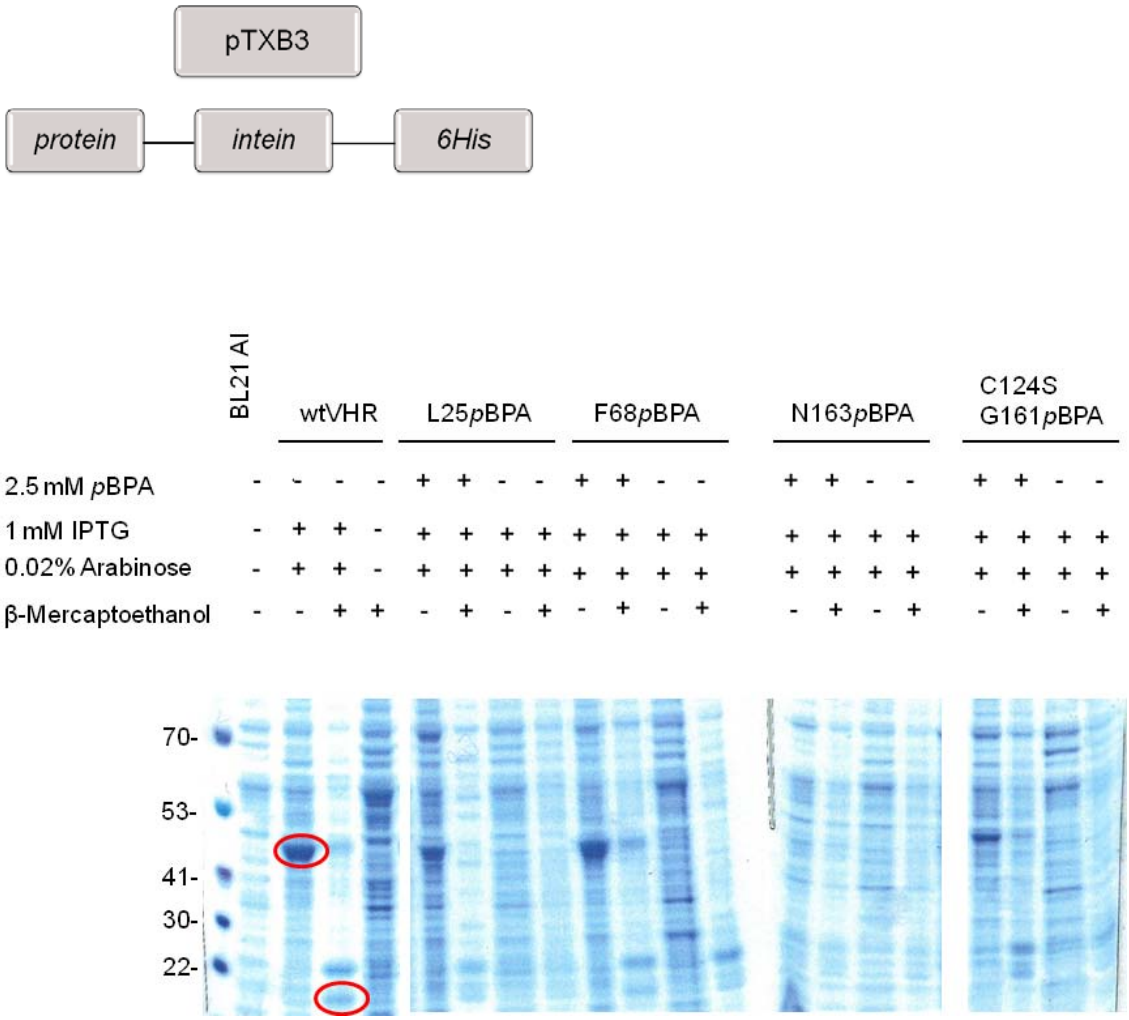


Figure 32: Small-scale expression of the VHR variants with *p*BPA at the final concentration of 2.5 mM in pTXB3-intein-6His plasmid. *p*BPA was filtered into the into the TB-FB GM at the beginning of the experiment. Protein expression was induced for 6 h at 25°C and then overnight at 18°C. The samples were purified under native conditions. An aliquot of each sample was taken for SDS-PAGE analysis and to the rest of the purified sample β-mercaptoethanol was added. Intein cleavage was done with final concentration of 50 mM β-mercaptoethanol for 48 h at 37°C. 20 µl of the final mixture was resolved by denaturing 4-12% SDS-PAGE. The band at approximately 20 kDa corresponds to the full length protein and at about 50 kDa to the intein fused form- the areas are marked with red circles for the wtVHR. The identity of the other co-purified proteins was not determined.

3.1.9 Creating the final *pBPA*-containing PRL-3 construct for biological exploration

The final construct for large-scale bacterial expression in *E. coli* needed to carry a tag for pull-down experiments. The available ab against PRL-3 are not suited for this, particularly when PRL-3 is cross-linked to another protein, which might potentially obscure ab recognition site. Whereas the pTXB3 construct has the advantage of removal of the truncated protein from the full length form, carrying the C-terminal intein-His_6 tag, its use for pull-downs faced several problems, such as (i) instability of intein and its potential cleavage when used for pull-downs with the cell lysate, (ii) importance of the C-terminal region in the PRLs, which was already addressed in detail (Pascaru, Tanase et al., 2009; Rios et al., 2013; Skinner et al., 2009; Song et al., 2009) or (iii) plethora of literature evidence for studying the PRLs, *in vitro* and in cells, by utilizing the N-terminal tagging approach (Daouty et al., 2008; Fiordalisi et al., 2006; Hardy et al., 2010; Liu, Al-Aidaroos et al., 2013; Song et al., 2009; Wang and Lazo, 2012; Zeng et al., 2000).

The pETM20 plasmid was compared with pTXB3-intein-6His, with the D72 *p*BPA variant of PRL-3. The result is shown in Figure 33. In pETM20, the D72 mutant of PRL-3 was expressed with an N-terminal $Flag_3$-tag. Generating an analogous construct in the pTXB3/His_6 plasmid was not successful because adding N-terminal tag to the protein generated a second recognition site for the NcoI restriction enzyme, after the nucleotide sequence encoding the affinity tag and immediately preceding the sequence for the starting amino acid of the protein. Because the PRL-3-encoding sequence was cloned into the NcoI/SapI sites in the pTXB3/His_6 plasmid, only the final construct without the $Flag_3$-tag was detected. Nevertheless, the trial experiment was conducted.

Like indicated previously, the yield obtained for D72*p*BPA, by using pTXB3/His_6 plasmid, was very satisfactory, and the level of the full length protein obtained was comparable to the level detected for the wt protein (Figures 30 and 33). It could not be excluded that, by placing the affinity tag at the C-terminal end of PRL-3, would not have negative effect on its proper functioning. In addition, considering the possible downstream applications of the unnatural amino acid-containing variants of the recombinant PRL-3, such as microinjection into the cells or incubation with the cell lysate followed by photo-induced cross-linking, the N-terminal $Flag_3$-tag was ultimately regarded as the most desirable, rendering the pETM20 plasmid of choice. Furthermore, like shown in Figure 33, the amount of the truncated form of the protein expressed when using the pETM20 plasmid was much less compared to the full length protein, so pETM20 plasmid was selected for further experiments.

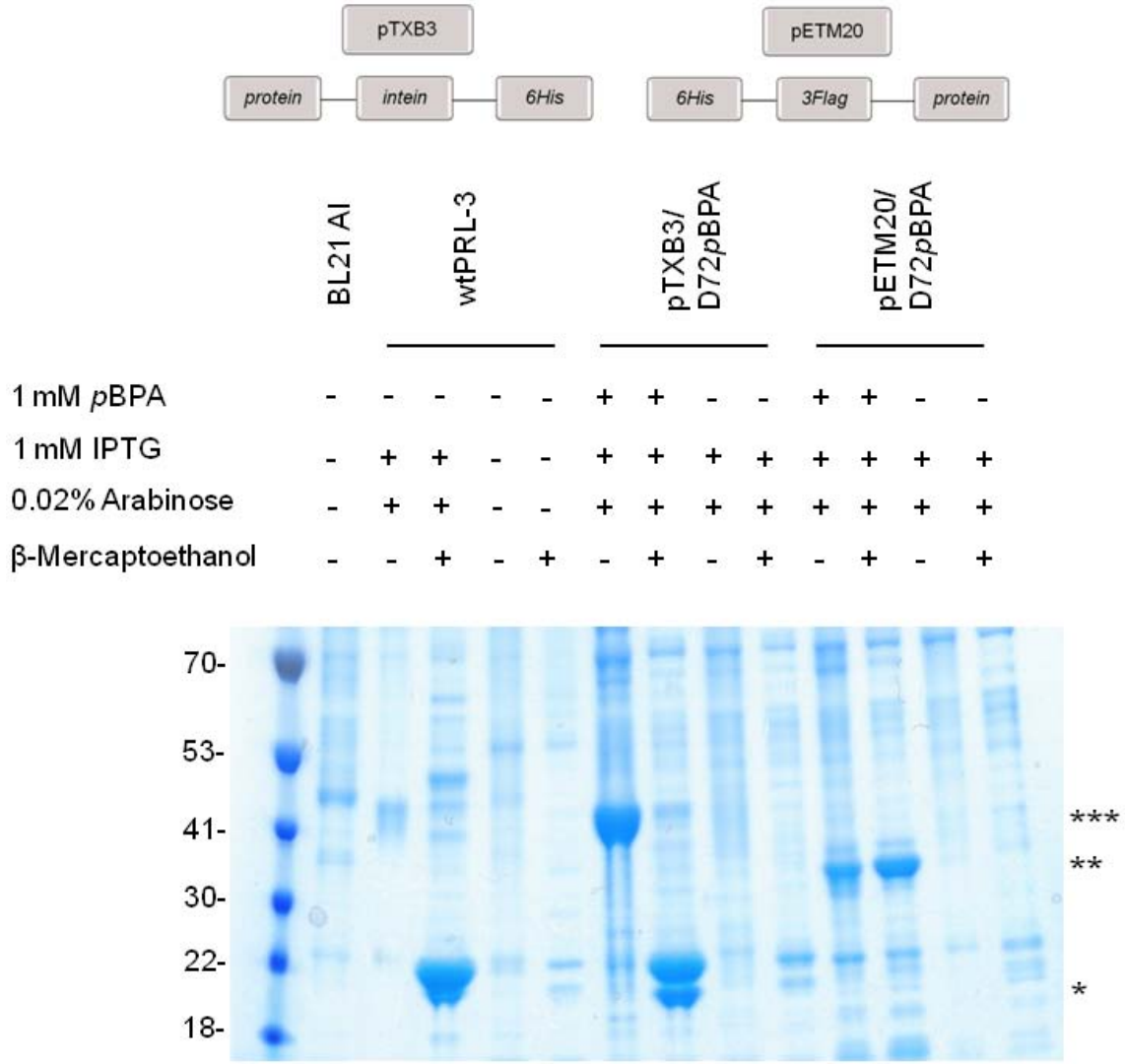


Figure 33: Comparing plasmids pTXB3-His_6 and pETM20 for the efficiency of expression of the *p*BPA variants of PRL-3 in *E. coli*. Protein expression was induced for 6 h at 25°C and then overnight at 18°C. In the pETM20 vector, D72*p*BPA mutant was expressed with an N-terminal $Flag_3$-tag. wtPRL-3 was expressed in the pTXB3-His_6 vector. The samples were purified under native conditions. An aliquot of each sample was taken for SDS-PAGE analysis. The rest of the purified sample was subjected to intein cleavage by adding β-mercaptoethanol to a final concentration of 50 mM. The band at about 45 kDa corresponds to the intein-fused form (***). When pETM20 vector was used, the band at about 40 kDa corresponds to the full length protein (**). The band at about 20 kDa corresponds to the full length protein (*). The identity of the other co-purified proteins was not determined.

3.1.10 Optimising concentration of *p*ACF for the I141- and N142TAG variants of PRL-3

Unlike fine-tuning the conditions for expression of PRL-3 variants with unnatural amino acid introduced at position 72, optimisation with the remaining two mutants proved to be much more cumbersome. For both, I141- and Asn142TAG, plasmids encoding both variants as N-terminal His_6-TEV-$Flag_3$-tagged fusion proteins were generated and were subjected to another round of small-scale expression with *p*ACF. The concentration of up to 5 mM *p*ACF was screened. As shown in Figure 34, for I141*p*ACF, the predominantly prematurely truncated form, where the amber STOP codon resulted in the termination of protein synthesis, was detected by SDS-PAGE analysis. For the N142*p*ACF variant, the yield of the full length protein increased concomitantly with the increase of the concentration of *p*ACF in the medium.

It has been reported that the efficiency of the amber STOP codon suppression depends on the sequence context (Miller and Albertini, 1983). Efficiency of suppression of a nonsense codon was shown to depend on the composition of the nucleotides surrounding the STOP codon, and mainly on the first codon following a nonsense codon. Extensive investigation of the efficiency of the suppression of the amber STOP codon concluded that the codons followed on their 3'-side by uracil (U) or by cytosine (C) were poorly suppressed, unlike the codons followed by adenine (A) or guanine (G) which showed good suppression. However, there were also some exceptions noted, because STOP codons that were prior to CUG or CUC proved to be well suppressed. It is possible that the suppression efficiency could be affected by the secondary structure of the mRNA containing a nonsense codon, namely that the structure could be such that translation termination factors would not find it accessible

easily. In the case of PRL-3 variants, the most optimal sequence context existed.

Next, not all positions in the protein are equally amenable to substitution with an unnatural amino acid (Farrell et al., 2005; Krishnamurthy, Dugan et al., 2011; Sato et al., 2011; Uezu et al., 2012). Predicting which amino acid residues in the native protein would be well suited to accommodate *p*ACF and/or *p*BPA can be helped by the available structural information, wicht was also the strategy applied here. It has been acknowledged that the efficiency of incorporation of an unnatural amino acid in response to the amber codon varies with the identity of the unnatural amino acid and with the position of the amino acid to be substituted in the gene (Hino et al., 2012). Moreover, the efficiency of the amber suppression with *p*ACF was reported to be better than with *p*BPA because of the higher activity of its adjacent aaRS.

Since we observed that the yield of the full length protein obtained with I141*p*ACF variant was low, the mutant was abandoned at this stage. The N142TAG variant of PRL-3 was selected to be subjected to further trial experiments with *p*BPA. However, large-scale expression of N142*p*BPA, with 2.5 mM concentration of *p*BPA, was extremely laborious and not efficient- the yield after the first purification step was about ten times lower. An additional drawback was the necessity to finally purify the sample by size-exclusion chromatography, to try to separate the prematurely truncated variant from the full length protein, and the pending optimisation of the conditions to obtain the separation. Therefore, the attention for biological exploration was focused on the D72*p*BPA variant of PRL-3.

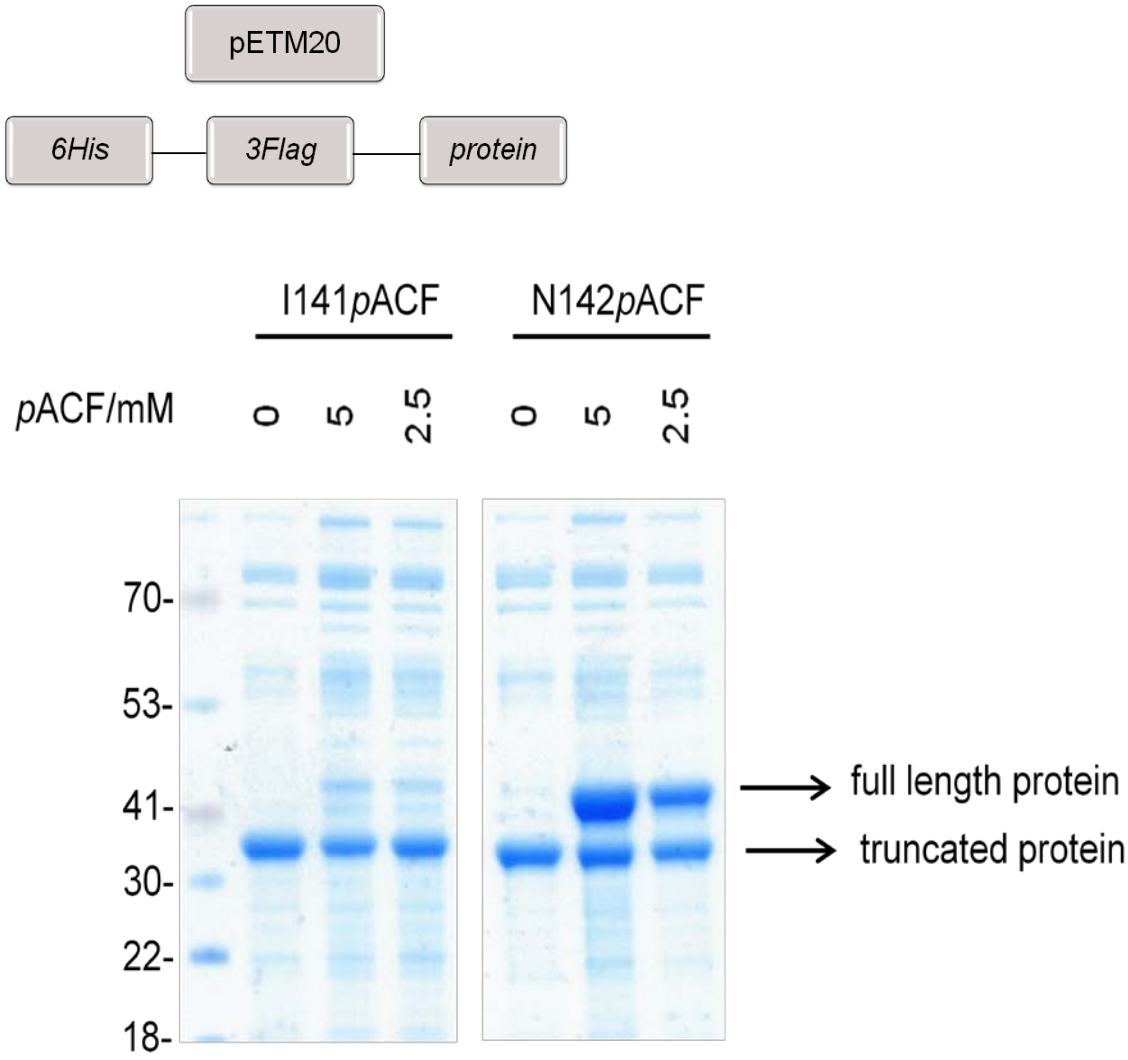


Figure 34: Optimising concentration of *p*ACF for expression with the I141- *and* N142TAG variants of PRL-3. Protein expression on a small scale was induced for 5 h at 25°C. *p*ACF was used at 2.5 or 5 mM final concentration. The samples were purified under native conditions. An aliquot of each sample was resolved by 4-12% SDS-PAGE gel, stained in Coomassie dye and destained in ddH_2O. The band at approximately 41 kDa corresponds to the expected size of the full length protein and at about 35 kDa to the variant where the amber STOP codon terminated protein synthesis. The identity of the other co-purified proteins was not determined. Blue Eye was used as the MW Protein Ladder.

3.2: Applicability of the photo-cross-linkable VHR variants in investigating dimerisation of VHR as a potential novel regulatory principle

3.2.1. Large-scale bacterial over-expression of wtVHR and all its variants generated

For large-scale bacterial expression of VHR, the pETM20 plasmid was selected. For VHR, the positioning of the affinity tag, which would be used in the pull-down experiments, at the N- or at the C-end of the protein was not at the level of importance discussed previously for PRL-3 (Chapter 3.1.9), as no sequence or structural determinants contributing to the specific subcellular localisation of VHR are known. The expression conditions initially applied for PRL-3 were taken and basically no modifications in the protocol were required.

As indicated in Figure 35, all the variants were purified to a high purity. Like indicated previously, all the VHR variants were expressed as N-terminal $Flag_3$-tag fusion proteins and purified to a high purity. Typically 0.5-1 l scale of GM was prepared. Verification of all the variants generated by determining full mass of protein samples in solution and summary of the protein yields are provided in Tables 2 and 3, respectively. By using the pETM20 vector, a form of the protein where protein synthesis terminated at the position of the amber STOP codon, was also generated. Considering that for both of the *p*BPA-containing variants of VHR the truncated form does not encompass the residues important for catalysis, it was not regarded necessary to try to separate the truncated form from the full-length protein by size-exclusion chromatography.

The following variants of VHR were selected for large-scale bacterial over-expression: native VHR, its photo-cross-linkable variants F68*p*BPA and

D92*p*BPA, catalytically inactive C124S, catalytically deficient D92A (Denu et al., 1995b; Ishibashi et al., 1992; Schumacher et al., 2002; Zhou et al., 1994) and a set of variants with systematically mutagenised position 68 (F68A, F68D, F68W and F68C) and the M69A mutant. The last set of mutants (F68A, F68D, F68W, F68C and M69A) was generated with the aim of intensive investigation of one of the most distinct structural features of VHR phosphatase, namely the "variable insert" segment.

Considering that the F68*p*BPA variant of VHR showed promising results in expression and purification experiments conducted on a large-scale (see Figure 40), the advantages of utilizing an expanded genetic code approach were diverted from the initial attempt to identify (novel) interacting partners and/or substrates of VHR to providing the first in-depth investigation of the "variable insert" region in this phosphatase.

In general, all the variants were expressed and purified with extremely satisfactory yields, which are summarised in Table 3. For simplification, the reported yields, calculated as mg of protein obtained from 1 l of bacterial culture, were calculated for the first purification step (HisTRAP)- regularly, not the whole batch obtained after the first HisTRAP was set for removal of the His_6-tag by GST-TEV and the sample-to-sample variations in the efficiency of the recovery of the final protein sample, after the second HisTRAP and GSTrap, were to be expected. Nevertheless, although for large-scale purification of the VHR variants the conditions initially applied for PRL-3 were taken, basically no modifications in the protocol were required.

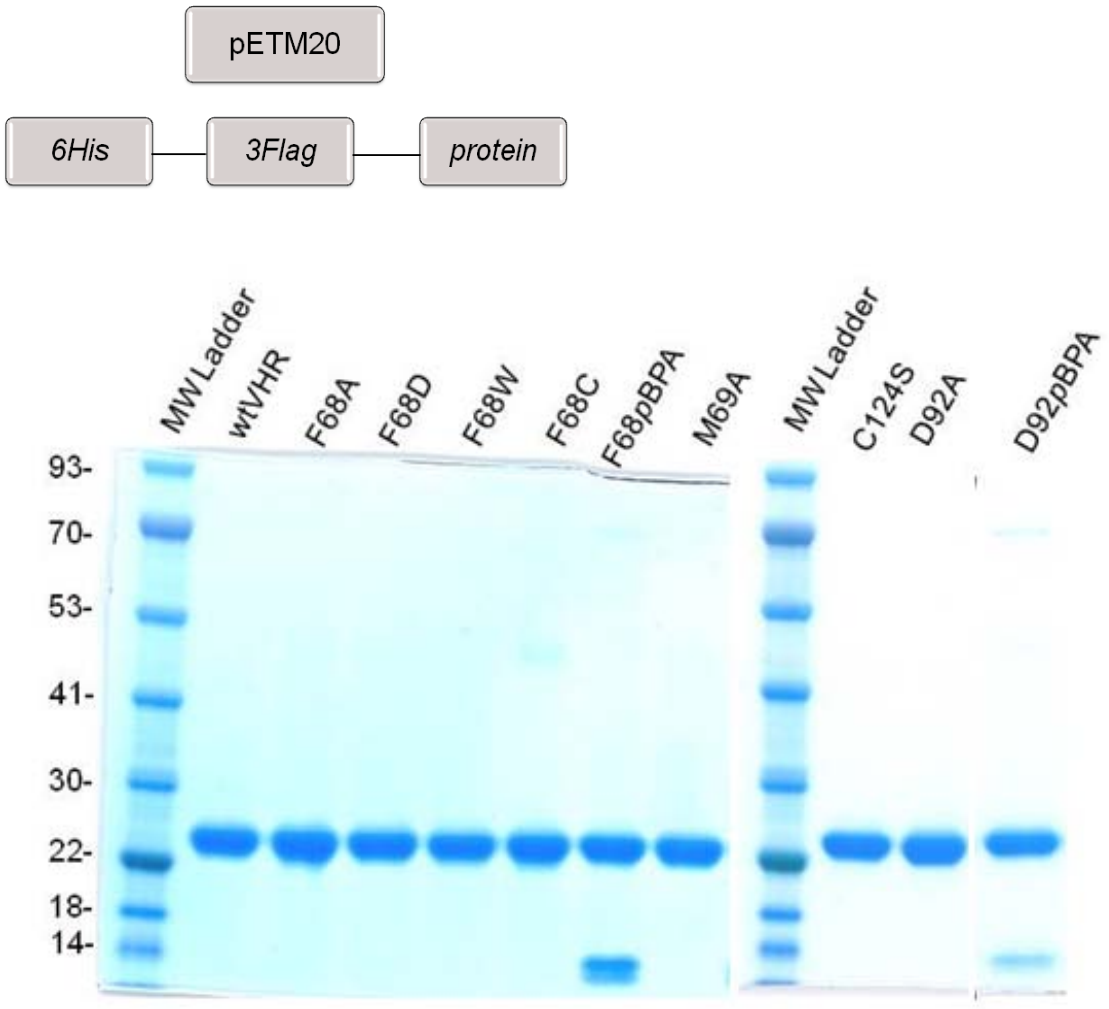


Figure 35: Coomassie stained SDS-PAGE gel of the wtVHR and all its variants generated. Plasmid pETM20 was used to over-express all the VHR variants as N-terminal $Flag_3$-tagged fusion proteins, according to the detailed protocol described in sections 2.2.12 and 2.2.13. The variants of VHR with the amino acids at the indicated positions exchanged for the amber STOP codon, TAG, were expressed by growing bacterial cultures in the presence of 2.5 mM *p*BPA. 2 µg of each protein was resolved by 4-12% SDS-PAGE gel, stained in Coomassie and destained in ddH_2O. The band at approximately 23 kDa corresponds to the expected size of the full length protein. The band at approximately 14 kDa corresponds to the truncated version where the amber STOP codon resulted in the termination of protein synthesis**.** Sometimes the faint band at about 70 kDa was visible in some samples, the identity of which has not been determined. Blue Eye was used as the MW Protein Ladder.

Table 2: Analysis of the intact protein samples by LC-MS. The protocol is described in section 2.1.15. The analysis was done by using an Acquity UPLC system (Waters) fitted with a C4 column coupled to a Q-Toff II mass spectrometer (Waters) using the ESI source in positive ion mode. Data were acquired over a mass range 500-3,500 m/z, with a scan time of 0.5 s. Data were calibrated externally against intact myoglobin as a reference standard, acquired immediately after sample data acquisition. Intact mass was calculated using the MaxEnt1 maximum entropy algorithm (Waters/Micromass).

VHR	Molecular mass expected (kDa)	Molecular mass obtained (kDa)	[Molecular mass expected (kDa) $+H^+$]$+x_{\beta\text{-mercaptoethanol}}$
Wild type	24.16	24.47	4 units of β-mercaptoethanol
M69A	24.10	24.10	
F68A	24.09	24.17	1 unit of β-mercaptoethanol
F68C	24.12	24.12	
F68D	24.13	24.13	
F68W	24.20	24.20	
F68*p*BPA	24.27	24.5	2 units of β-mercaptoethanol
D92*p*BPA	24.30	24.46	2 units of β-mercaptoethanol
D92A	24.12	24.20	1 unit of β-mercaptoethanol
C124S	24.15	24.23	1 unit of β-mercaptoethanol

Table 3: Protein yields for the generated VHR variants. The yields were calculated as mg of protein obtained from 1 l of bacterial culture. The calculations were done for the protein samples after the first purification step by using HisTRAP. The calculations for the yield after the following purification steps (removal of the His_6-tag by GST-TEV and the subsequent removal of the protease by the second HisTRAP and GSTrap columns), for which only a limited amount of the protein after the first purification step was used, were done by extrapolating the cleaving and recovery efficiencies to the whole batch, and checking protein concentration by measuring absorbance at 280 nm and resolving the samples by SDS-PAGE.

VHR	Protein yield (mg of protein/ l of bacterial culture)	Protein yield (after 2nd HisTRAP and GSTrap/ l of bacterial culture)
Wild type	81.8	14.4
M69A	40.4	11.2
F68A	22.5	5.6
F68C	46.5	13.1
F68D	36.1	9.4
F68W	34.7	10.2
F68*p*BPA	73.6	8.3
D92*p*BPA	50.4	4.4
D92A	47.7	9.6
C124S	32	1.8

3.2.2 Investigating cross-linking abilities of VHR variants containing *p*BPA *in vitro*

In the set of preliminary experiments, which involved incubation of the F68*p*BPA variant of VHR with the cell lysate from HEK293 cells followed by exposure to 365 nm UV light, Western blot analysis (anti-Flag) of the immunoprecipitates (performed by using anti-Flag affinity gel) detected the bands which in size corresponded to the dimeric complexes of VHR. Analogous experiments without UV exposure showed no dimer formation (not shown because these preliminary experiments were not followed upon, as the focus of the project shifted into investigating dimerisation of VHR *in vitro*; they are accordingly not described in the methods, but was found necessary to mention them to explain how investigation of VHR homodimers started in the first place). VHR dimerisation has not been previously reported. In the light of these surprising results, an extensive investigation was designed to address apossibility of dimerisation as a novel feature linked to VHR. Therefore, to further characterize the potential of VHR to self-associate, a series of *in vitro* experiments was conducted.

First, the dependence of the dimeric complex formation on the concentration of the protein was evaluated. For this purpose, the native protein and its F68*p*BPA variant were exposed to 345 nm light for 30 minutes, on ice, spanning protein concentration from 50 to 500 µg/mL. The subsequent Western blot analysis (anti-VHR) of the recovered samples detected the bands corresponding to the cross-linked complexes of VHR already at 50 µg/mL, as indicated in Figure 37 (the band at about 53 kDa). There was proportional increase of the covalently cross-linked complexes detected following the increase of the protein concentration. Based on the results shown in Figure 37, the protein

concentration range between 50 and 250 µg/mL was selected as the range of interest. In addition, this range enabled keeping the reaction scale at 50 µL, thus minimizing the amount of the protein samples required.

Intriguingly, the F68*p*BPA variant of VHR was found to also covalently cross-link to the truncated product formed when the amber STOP codon resulted in the termination of protein translation instead of *p*BPA incorporation (Figure 37) (the band at about 40 kDa). This enthralling observation led to the indication that the region of the protein corresponding to the truncated product might bear sufficient structural requirements to drive intermolecular association between two monomeric units of VHR.

As indicated already in Figure 35, in the purified samples of both of the *p*BPA-containing variants of VHR the truncated products, with the protein translation aborted at the position of the amber STOP codon, were also generated, notwithstanding at a different extent. For the D92*p*BPA variant, the proportion of the prematurely truncated form was regularly lower than for F68*p*BPA. These truncated products were stable in the cells and, due to the N-terminal tagging, detected in the purified samples. At that point no further separation by size-exclusion chromatography was rendered necessary because this fragment was not anticipated to impose interference with phosphatase activity assessments. This is due to neither of the truncated products generated for both *p*BPA-containing variants included the residues of the canonical PTP signature motif or the residues detrimental for hydrolytic activity of the enzyme.

It was regarded as reasonable not to perform exposures to 345 nm light for longer than 30 minutes to (i) prevent possible damage of the protein and/or cross-linked complexes which might result in yield reduction and to (ii) avoid

the formation of larger complexes with ambiguous solubility/stability properties that may generate even more artefacts, for example by modifying recognition by the specific antibody. Moreover, in the preliminary time-screen experiment, employing 365 nm light, the complex corresponding in size to the dimer was visible already after 5 min of UV exposure (not shown). In all the experiments, the samples were kept on ice for the duration of exposure to 345 nm light and the box was at about 40 cm distance from the UV lamp, to try to minimize the destructive effect such exposure might have on the protein samples. To control for the potential damaging effect of the UV light, wtVHR was constantly utilised as a control. As indicated in Figure 37, no bands in the lower molecular weight range were detected, indicating that under this setup no damaging effect of UV light on the proteins was exerted. In addition, another set of controls regularly used included the non-irradiated set of F68*p*BPA-containing samples. Lack of appearance of the bands in the higher molecular weight range for these samples indicated that the formation of the observed intermolecular complexes dependent on (i) the presence of *p*BPA in the protein and (ii) on the exposure of only *p*BPA-containing protein variants to 345 nm UV light.

From the results presented in Figure 37, it can be seen that increasing protein concentration from 250 to 500 µg/mL had no significant effect on improving the yield of the covalently cross-linked dimers formed. Based on this observation, it would be reasonable to assume that the likely cause might be intramolecular cross-linking. Like already mentioned, the reactivity of the photo-excitable *p*BPA is affected by the geometrical constraints (Tanaka et al., 2008). This refers to not just the distance from the proximal C-H bonds to which *p*BPA could covalently cross-link, but also to the relative orientation the amino acid assumes within the peptide (Wittelsberger et al., 2008 and see Figure 36). To

account for negligible increase in the dimers formed after photo-induced cross-linking, it could be hypothesized that *p*BPA incorporated at position 68 assumes such an orientation that intramolecular reaction is more favoured then intermolecular, under the experimental conditions applied.

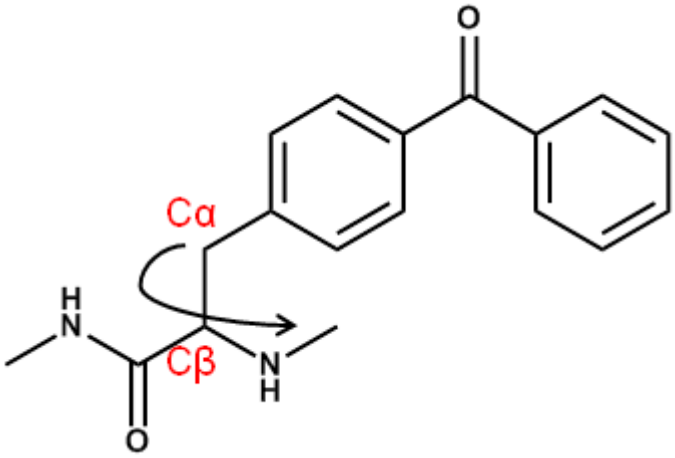


Figure 36: Reactive radius of benzophenones. Initially, reactive radius of pBPA was considered a sphere about 3Å in radius, but rotation about Cα-Cβ bond increases reactive space.

Therefore, in order to investigate if dimerisation was mediated by the cysteine residue of the catalytic site, an analogous experiment was performed after pre-incubating the samples with the cocktail of the general inhibitors of the phosphatases known for its broad spectrum phosphatase inhibition, PS (Figure 38). PS was used with the F68*p*BPA variant of VHR, but also with the wt protein as a control. Moreover, the C124S mutant, which has diminished catalytic activity, was used as additional control. All negative controls were clean, as expected- there were no bands in the higher molecular weight range detected for the non-irradiated control samples, and also no bands at the lower molecular weight range in case of the irradiated samples (Figure 38).

A possible mechanism through which the intermolecular association in VHR could be mediated might involve dimeric association through disulfide bond formation. Cys124 is placed at the base of the active site crevice and, from merely structural considerations, the intermolecular association through Cys124-mediated disulfide bond formation was not regarded as likely. There was a work reported previously which investigated the mechanism commensal bacteria use to activate Erk1/2 pathway of the host (Wentworth et al., 2011). In this study it was shown that treatment of cells expressing wtVHR or its C124S mutant with H_2O_2 led to the formation of higher molecular weight complexes only in case of the native protein. These complexes were assigned to be disulfide-bond mediated complexes formed through the active site cysteine.

The cross-linking pattern (Figure 38) obtained for the F68*p*BPA variant of VHR, with or without treatment with PS, indicated that the observed self-association was not significantly impacted by the availability of the active site cysteine.

s

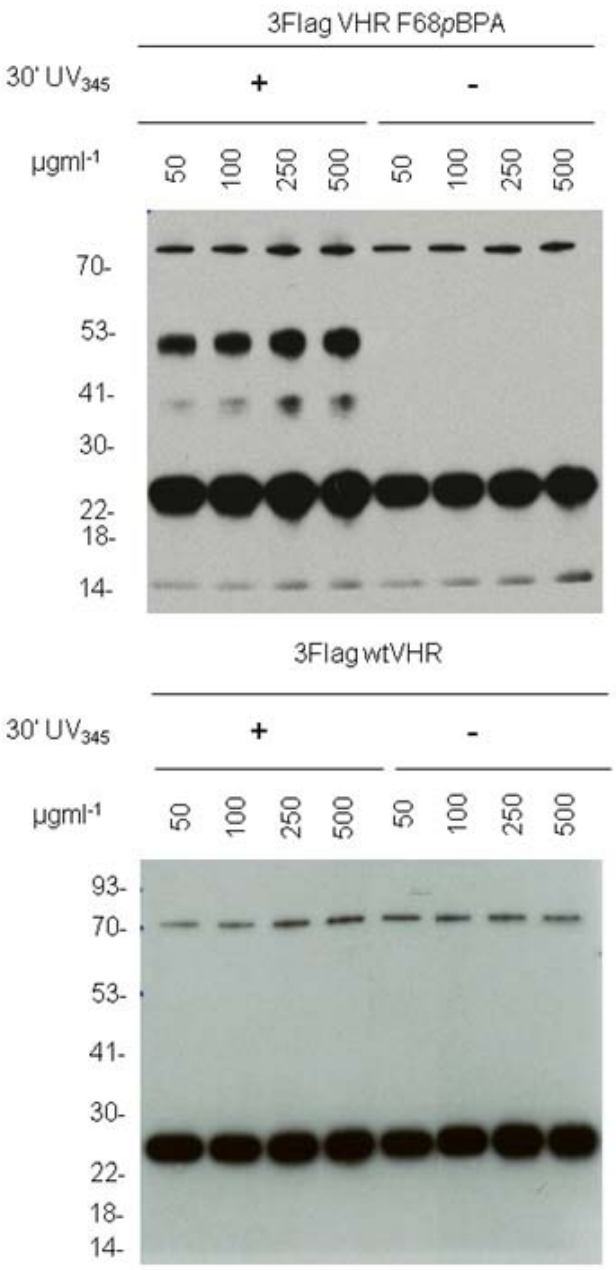


Figure 37: Examining the ability of the F68*p*BPA variant of VHR to form UV-induced dimers after 30 min exposure to 345 nm. Non-irradiated F68*p*BPA and wtVHR, with or without UV-exposure, were used as controls. The protocol is described in section 2.3.5.The band at about 53 kDa corresponds to a covalently cross-linked dimer and at about 40 kDa to a cross-link of the monomer to a product with protein synthesis terminated at position of the amber codon. Monomers run as a band at about 25 kDa. The band at about 70 kDa was occasionally observed after purification of the recombinant protein (see Figure 35). The analysis was done by Western blot (anti-VHR). The results are representative of two experiments.

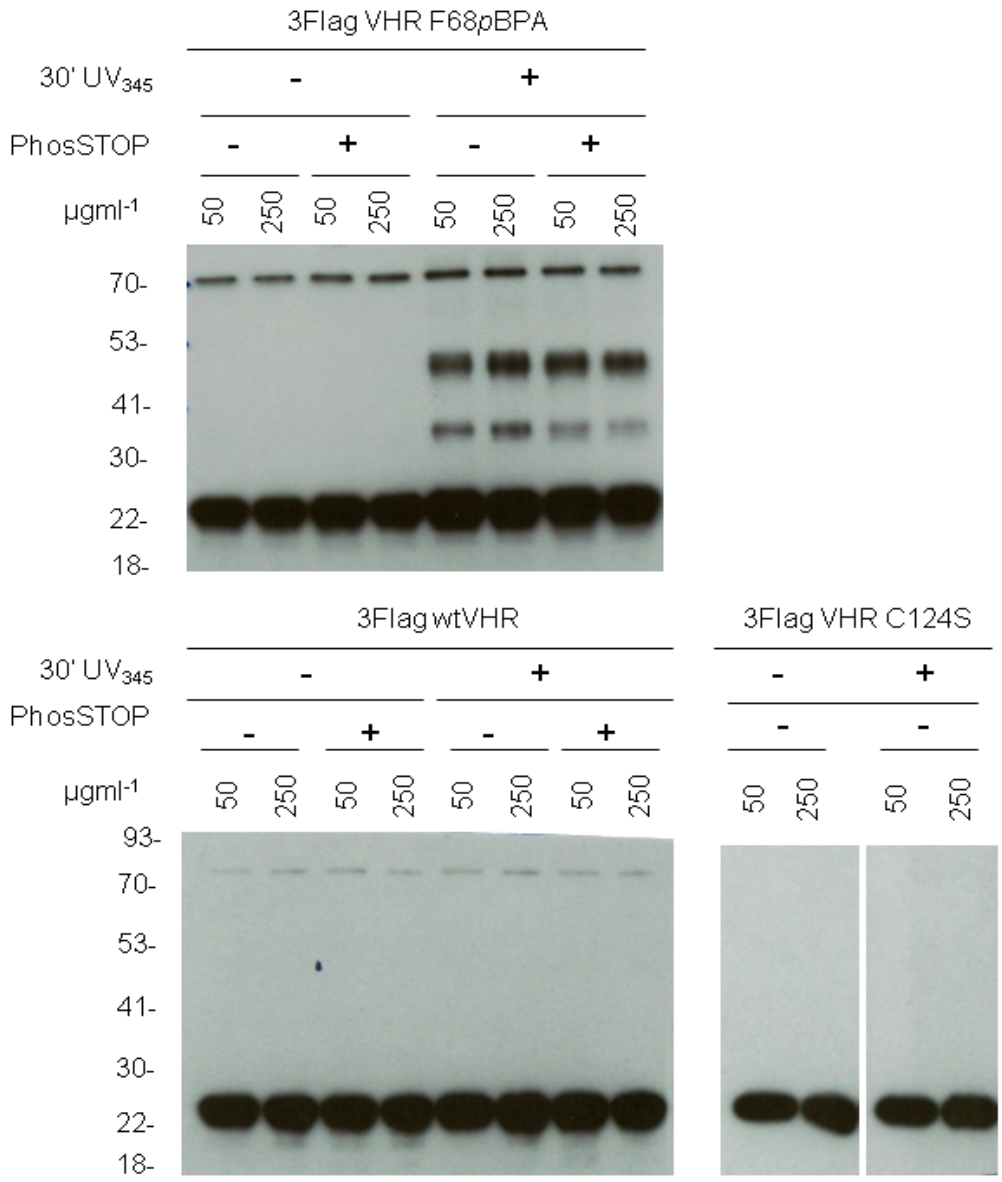


Figure 38: Assessing the effect of the active site Cys124 on the ability of VHR to form UV-induced dimers. This was investigated by (i) using robust phosphatase inhibitor PS and (ii) by employing the catalytically inactive mutant C124S. The protocol is described in section 2.3.5. Reactions were done in a 96-well plate setup and the final volume was 50 µl. Recovered samples were mixed with 2 x SDS-PAGE sample buffer prior to resolving 250 ng of protein by reducing 4-12% SDS-PAGE. The analysis was done by Western blot (anti-VHR). The band at about 53 kDa corresponds to a covalently cross-linked dimer, at about 40 kDa to a cross-link of the monomer to a truncated product and at about 25 kDa to a monomer. The band at about 70 kDa was occasionally observed after purification of the recombinant protein. The results are representative of at least two experiments.

It was acknowledged before that for a photo-cross-linkable amino acid incorporated at different positions within the same protein, different efficiencies in the UV-induced cross-linking were observed (Hino et al., 2005; Krishnamurthy, Dugan et al., 2011; Sato et al., 2011). Therefore, the D92*p*BPA variant of VHR was subjected to an analogous set of experiments to check if there would be any differences in the generated cross-linking pattern. The result is presented in Figure 39. Exposure of the D92*p*BPA variant of VHR to 345 nm light resulted in the formation of a band corresponding in size to a dimeric complex, just like previously shown for the F68*p*BPA variant. Moreover, for the D92*p*BPA variant there was no band, at least at a detectable level, corresponding in size to a cross-link formed between the full length protein and the product terminated at the position of the amber codon. Like shown in Figure 15, the amount of the truncated product present in the purified sample of the D92*p*BPA variant was less than in the case of the F68*p*BPA variant, which could be a contributing factor to the lack of covalently cross-linked complexes formed and/or detected with this fragment. Another plausible explanation could be the orientation of *p*BPA at position 92, which could be such that cross-linking to the full length protein might be preferred.

In summary, the variants of VHR with *p*BPA selectively introduced at positions 68 and 92 showed the ability to form dimers after exposure to 345 nm light, *in vitro*, thus underscoring the importance of evaluating diverse positions for replacement with *p*BPA.

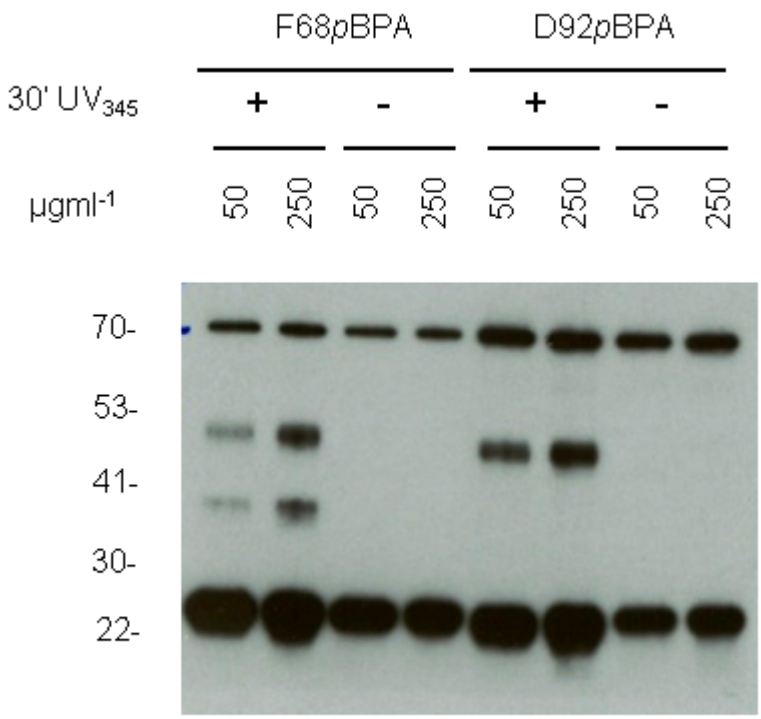


Figure 39: Examining the ability of the D92*p*BPA mutant of VHR to form photo-induced dimers. Shown are the dimer-forming yields of the purified recombinant $Flag_3$-tagged F68*p*BPA and D92*p*BPA variants upon 30 min exposure to 345 nm light and across protein concentration range of 50-250 µg/ml. Irradiation was performed according to the protocol described in detail in section 2.3.5. Reactions were done in a 96-well plate setup and the final volume was 50 µl. Recovered samples were mixed with 2 x SDS-PAGE sample buffer prior to resolving 250 ng of protein by reducing 4-12% SDS-PAGE. The analysis was done by Western blot (anti-VHR). The band at about 53 kDa corresponds to a covalently cross-linked dimer, at about 40 kDa to a cross-link of the monomer to a truncated product and at about 25 kDa to a monomer. The band at about 70 kDa was occasionally observed after purification of the recombinant protein. Proteins were detected by Western blotting using anti-VHR antibody. The results are representative of two experiments.

3.2.3 Catalytic activity of the VHR variants against *p*NPP

To assess if, and how, the introduced mutations affected the catalytic performance of the recombinant VHR variants, an extensive kinetic profiling was conducted, according to the specifications described in section 2.3.4. The hydrolytic activity was assessed against the robust non-specific phosphatase substrate *p*NPP. All the variants were at the concentration of 500 nM, whereas the concentration of the substrate spanned across 0.5 to 100 mM. The pH value of the reaction buffer was set to 7.5. The calculated kinetic parameters k_{cat} and Km are summarised in Table 4.

The D92A and C124S mutants of VHR was omitted from the kinetic profiling because it had already been reported that substitution at this position entirely abolishes phosphatase activity (Denu et al., 1995b; Ishibashi et al., 1992; Zhou et al., 1994). These early studies addressing catalytic specificities of VHR as well as the dephosphorylation mechanism employed helped to establish that (i) VHR has only one catalytic centre governing dephosphorylation of both pTyr and pSer/pThr-containing substrates and that (ii) VHR employs the same catalytic principle applied by the Tyr-specific PTPs (Denu et al., 1995a and b; Denu et al., 1996a; Ishibashi et al., 1992; Zhou et al., 1994).

For VHR, it was acknowledged that the enzyme displays the most efficient catalytic performance at the pH value 6 of the reaction buffer (Denu et al., 1995a; Hengge et al., 1996). The activity was found to decrease at the elevated levels of the pH value (Denu et al., 1995a). The pH value for the data presented in Tables 4 and 5 was 7.5, which could contribute in accounting generally lower catalytic activity demonstrated by the wtVHR in contrast to the previously reported values (Denu et al., 1995a; Denu et al., 1996b). Other

differing factors were different composition of the reaction buffers used and also different ways of analyzing the data.

Drawing conclusions from the data presented in Table 4, it can be generally stated that replacing Phe68 in the native VHR with alanine, aspartate, cysteine, tryptophan or photo-cross-linking *p*BPA resulted in no drastic alterations in the catalytic efficiency. The Km values of all the aforementioned variants were within 1-1.2 fold of the value obtained for the wtVHR.

On the other hand, when Met69 of the native VHR was replaced with alanine, the Km value approximately doubled. Based on the known role of Met69 in contributing to the maintenance of the proper orientation of the P-loop by anchoring the conserved Arg130 through hydrogen bonding (Yuvaniyama et al., 1996), it might be that generating the M69A substituent dislocates the most optimal relative position of the residues of the P-loop. Moreover, the value of the second order rate constant k_{cat}/Km obtained for the M69A mutant was about 3.2 times lower comparing to the wtVHR. Its lower value with respect to the values obtained for the other VHR variants subjected to the screen could be attributed by the already mentioned effect the substitution exerted on the Km value because the catalytic efficiency, as indicated by the catalytic turnover k_{cat} value, was only about 1.6 times lower.

As for the variants generated for position 68, the values of the kinetic parameters k_{cat} and k_{cat}/Km exhibited only low reductions with respect to the wtVHR.

In the native protein, Asp92 acts as a general acid/base catalyst in the hydrolytic reaction leading to the formation of a phosphoenzyme intermediate and the following intermediate hydrolysis (Denu et al., 1996b). Substituting

Asp92 with Asn, which mimics it in size but is robbed off the ability to act as an acid/base catalyst in the dephosphorylation reaction, significantly lowered the values of the kinetic parameters k_{cat}, Km and the k_{cat}/Km (Denu et al., 1996b; Kim et al., 2001). In contrast, the derogative effect of the D92*p*BPA substitution on the catalytic performance of this VHR variant was not as striking as for the mentioned D92N substituent. For the D92*p*BPA variant, the catalytic efficiency (indicated by k_{cat} value) was only around 25 times lower, whereas the values reported previously for the D92N mutant were up to almost 120 times lower (Denu et al., 1996b; Kim et al., 2001). The less drastic effect of the D92*p*BPA substitution is also reflected in the value of the second order rate constant which is only slightly lower (about 1.4 times) than for the wtVHR. Therefore, based on the results presented in Table 4 for the D92*p*BPA variant of VHR, it could be postulated that the carbonyl moiety in the side chain of *p*BPA could in part rescue the lack of the putative general acid/base aspartate.

In summary, the current study provides the first extensive kinetic investigation of the generally nonconserved residues in VHR, Phe68 and Met69, located in spatial proximity to the phosphosubstrate-binding cradle. As shown in Table 4, for all VHR variants generated by site-directed replacement of Phe68 in the native protein with alanine, aspartate, tryptophan, cysteine or photo-activatable *p*BPA, the values of the kinetic parameters were not substantially affected. Therefore, it could be concluded that the overall catalytic performance of the enzymes was virtually not altered.

In slight contrast was catalytic efficiency of M69A variant, which exhibited a more pronounced shift in the values of kinetic parameters Km and k_{cat}/Km when compared to the wtVHR and also the other variants generated for the position 68. Hence, it could be reasoned that substituting methionine for a

residue that lacks hydrogen bonding capacity disrupts maintenance of the proper geometry of the active site cleft via dislocation of the concise hydrogen bonding array required to fine tune relative positions and optimal distance of the key catalytic residues thus indirectly engaging in promoting efficient turnover.

Table 4: Characterisation of the catalytic activity of the VHR variants. The general protocol is outlined in section 2.3.4. Kinetic parameters for the wtVHR and its variants were determined by using *p*NPP as a substrate. The proteins were assayed at 500 nM and concentrations of *p*NPP used were 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 80 and 100 mM. The assays were conducted at 37°C in a buffer containing 25 mM HEPES (pH 7.5), 124.5 mM NaCl, 2.5 mM EDTA and 2 mM DTT, in 96-well plate format, in a microplate reader Tecan Saphire 2. The minimum of two triplicate sets of experiments was conducted for each sample. The Km was determined using the GraphPad 5.0 software.

VHR	**k_{cat}**	**Km**	**k_{cat}/Km**
	s^{-1}	mM	$M^{-1}s^{-1}$
Wild type	43.85 ± 5.16	21.38 ± 6.42	2.14 ± 0.40
M69A	26.44 ± 3.35	43.35 ± 13.92	656.3 ± 0.29
F68A	28.21 ± 5.56	20.66 ± 7.52	1.41 ± 0.24
F68C	39.71 ± 1.83	25.47 ± 0.87	1.56 ± 0.12
F68D	26.82 ± 0.62	22.49 ± 6.35	1.24 ± 0.32
F68W	36.15 ± 24.42	21.60 ± 5.72	1.58 ± 0.71
F68*p*BPA	40.04 ± 5.43	21.84 ± 8.54	2.09 ± 0.99
D92*p*BPA	1.74 ± 0.01	10.48 ± 1.11	1.56 ± 0.12

Earlier studies also showed that the entire phosphatase activity of the wtVHR was utterly abrogated in the presence of a competitive inhibitor such as arsenate or vanadate (Zhou et al., 1994), consistent with a model of a single catalytic site accounting for the dual-specificity profile. The phosphatase inhibitor employed for the data presented in Table 5 was a broad-range inhibitor called PhosSTOP. According to the product specifications, PS demonstrated various efficiencies in inactivating diverse set of PPs screened. To examine the effects it imposes on the catalytic activity of VHR, a set of experiments was done where the native protein was pre-incubated with PS prior to adding *p*NPP (Table 5) and then compared to the performance of the untreated control.

By comparing with the values of the kinetic parameters obtained for the wtVHR, it can be seen that the wtVHR treated with PS demonstrated only about 1.2 times reduction in the catalytic turnover. The exhibited approximately 2.1 lower value of the second-order rate constant k_{cat}/Km is the result of about 2.2 times higher value of the Km parameter, reflecting binding of a substrate. Whether further increase in the blocking of the catalytic activity of the native VHR could be achieved by prolonging incubation times of the enzyme with the inhibitor was not tested.

Table 5: Assessing the effect of the general phosphatase inhibitor PS on the catalytic activity of the wtVHR. The general protocol is outlined in section 2.3.4. The samples were pre-incubated with PS for 15 min prior to measurements. PS is a broad range phosphatase inhibitor, the composition of which is not speciifed, so the concentartions of its components are also not known. It was prepared according to the instructions from the manufacturer and was used at 1 x final concentration (from 10 x stock). The activity was examined with *p*NPP as a substrate. The protein concentration was 500 nM and the concentrations of *p*NPP used were 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 80 and 100 mM. The assays were conducted at 37°C in a buffer containing 25 mM HEPES (pH 7.5), 124.5 mM NaCl, 2.5 mM EDTA and 2 mM DTT, in 96-well plate format, in a microplate reader Tecan Saphire 2. Two triplicate sets of experiments were conducted for each sample. The Km was determined using the GraphPad 5.0 software.

VHR	**k_{cat}**	**Km**	**k_{cat}/Km**
	s^{-1}	mM	$M^{-1}s^{-1}$
Wild type	47.80 ± 0.84	14.96 ± 1.54	2.87 ± 0.10
Wild type + PhosSTOP	39.90 ± 4.11	32.37 ± 11.29	1.34 ± 0.25

3.2.4 Structural characterisation of the VHR variants by far-UV CD measurements

Structural integrity of the VHR variants was investigated by obtaining fa-UV CD spectra (Ranjbar and Gill, 2009). Prior to CD measurements, the protein concentration was determined by measuring A280 with a NanoDrop Spectrophotometer. However, sometimes there was a discrepancy between the measured concentration value and the corresponding spectra obtained, possibly due to dilution effect when preparing the measuring cuvette for various samples. This concern was also noted when acquiring the CD spectra for PRL-3 variants (D72A, I141Y, N142Y, N142A and Q145Y; data not shown). In addition, the CD spectra presented earlier for PRL-3 (McParland, Varsano et al., 2011) support the above stated difficulties in obtaining the correct protein concentration value by using NanoDrop. Therefore, plotting the raw data was a more favourable option then normalizing them to a fixed value, whereas it still provided a good qualitative indication of the proper protein folding.

When comparing the spectra acquired for all the VHR variants with the CD measurements obtained for the native protein, only minor modifications of the CD spectra were observed (Figures 41-43). All the variants were in good agreement with the spectra acquired for the native protein. In addition, the spectrum for D92*p*BPA compared well to the one for D92A (Figure 41), as expected due to the positioning of this residue in the general acid/base loop. Acquiring far-UV CD spectra (190-250 nm) can give information about secondary structure of proteins because every element of the secondary structure (alpha-helices, beta-sheets or random coils) give characteristic shape and magnitude in the spectra (Ranjbar and Gill, 2009), and, like shown in

Figures 41-43, for all VHR variants the obtained CD spectra showed comparable secondary structures.

Moreover, the variants generated for position 68 all aligned well when mutually compared (Figure 42). In the native protein, Phe68 is located on the surface-exposed hydrophobic patch (Wu et al., 2009; Yuvaniyama et al., 1996), which prompted a hypothesis that certain plasticity to accomodate residues of diverse structural and electrochemical properties might be reasonable to expect. In additional support of this hypothesis, even the F68*p*BPA variant of VHR exhibited excellent alignment with the spectra acquired for the native protein (Figure 43). Furthermore, when substituting Met69, known to form hydrogen bond with a conserved Arg130 of the P-loop (Yuvaniyama et al., 1996), for alanine did not seem to cause structural perturbations (Figures 42 and 44).

Therefore, it can be overall concluded that none of the introduced mutations disrupted proper folding of the mutant forms of the phosphatase, as judged by comparing the profile of the spectra with the results obtained for the wtVHR.

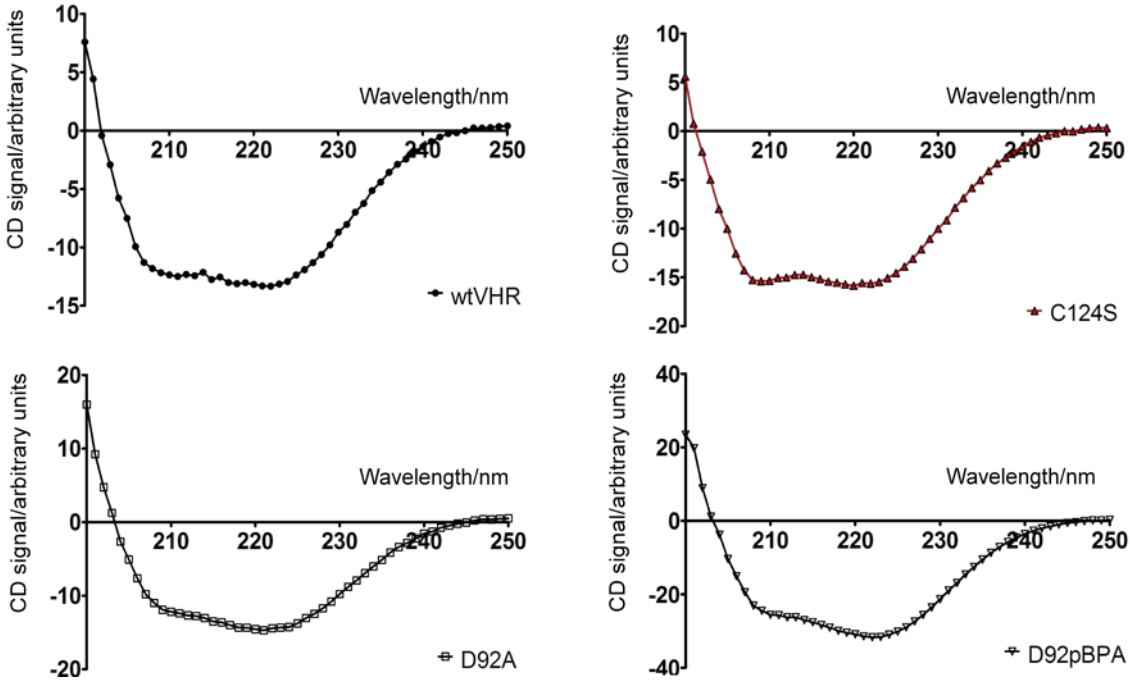


Figure 41: Analysis of the secondary structure for the wtVHR and its variants by obtaining far-UV CD measurements. The detailed protocol is outlined in section 2.1.16. The proteins were dialysed against 20 mM Tris-HCl (pH 7.4), 50 mM NaCl and 1 mM β-mercaptoethanol and the concentration was measured by NanoDrop Spectrophotometer. The CD spectra were acquired across 200-250 nm range. The spectra obtained for all the mutant forms were compared with the one for the native protein. The spectrum for D92pBPA was also assessed with respect to a conventional substrate-trap D92A. The raw data was plotted using GraphPad Prism 5.0 software.

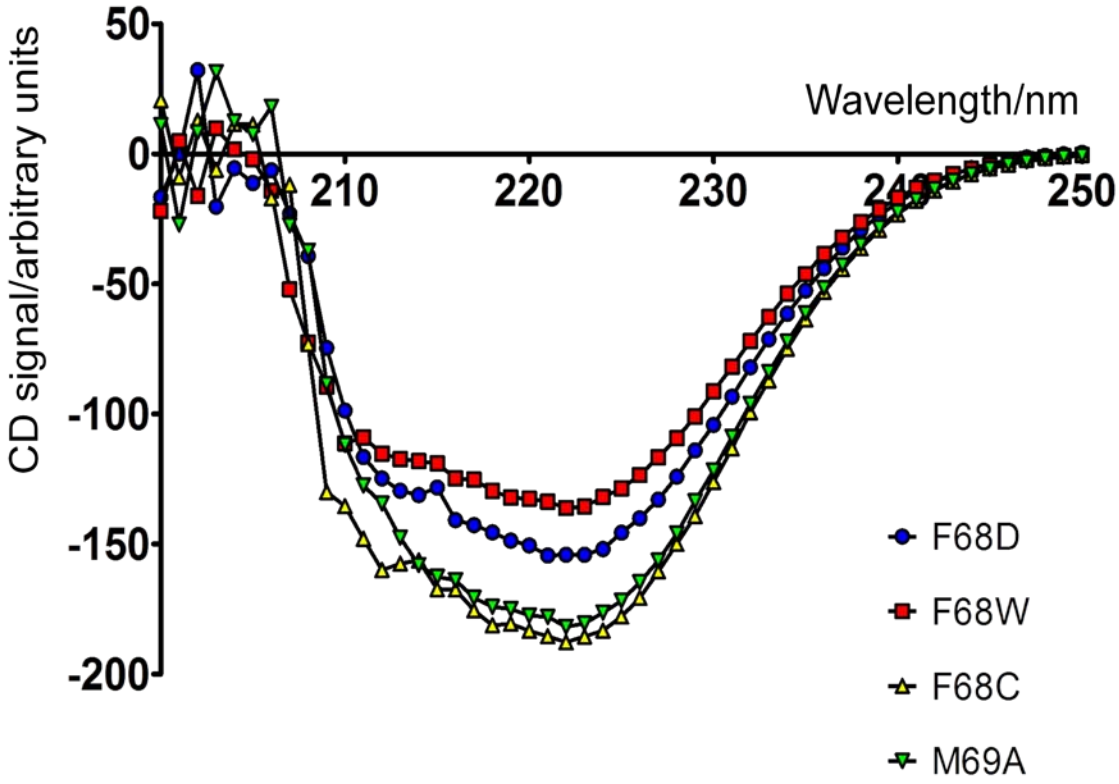


Figure 42: Analysis of the secondary structure for M69A and for a series of the F68-mutants of VHR by obtaining far-UV CD measurements. The detailed protocol is outlined in section 2.1.16. The proteins were dialysed against 20 mM Tris-HCl (pH 7.4), 50 mM NaCl and 1 mM β-mercaptoethanol and the concentration was measured by a NanoDrop Spectrophotometer. The CD spectra were acquired across 200-250 nm range. The raw data was normalized to the highest concentration value amongst the samples measured (0.47) and the normalized values were plotted using GraphPad Prism 5.0 software.

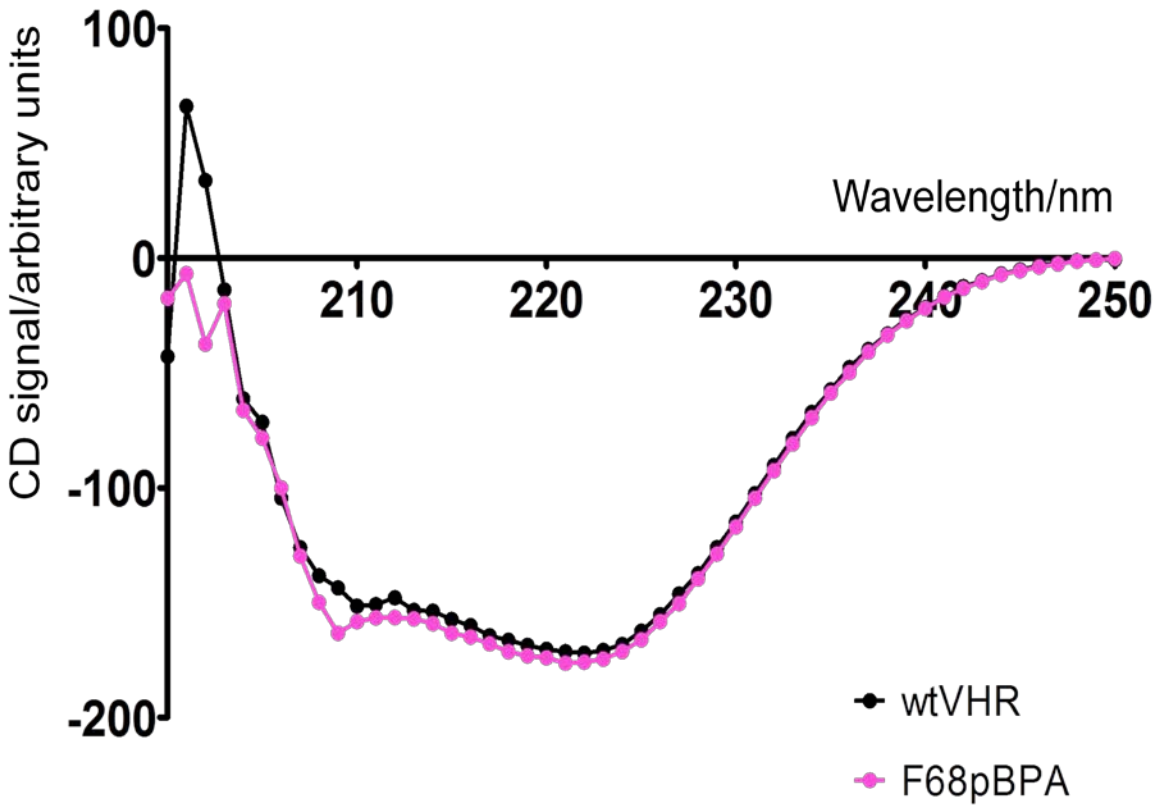


Figure 43: Comparing the secondary structure of the F68*p*BPA variant of VHR to the native protein by obtaining far-UV CD measurements. The detailed protocol is outlined in section 2.1.16. The proteins were dialysed against 20 mM Tris-HCl (pH 7.4), 50 mM NaCl and 1 mM β-mercaptoethanol and the concentration was measured by a NanoDrop Spectrophotometer. The CD spectra were acquired across 200-250 nm range. The spectra for wtVHR and for its F68*p*BPA mutant were acquired in two independent measurements so the raw data was normalized to 1 and the normalized values were plotted using GraphPad Prism 5.0 software.

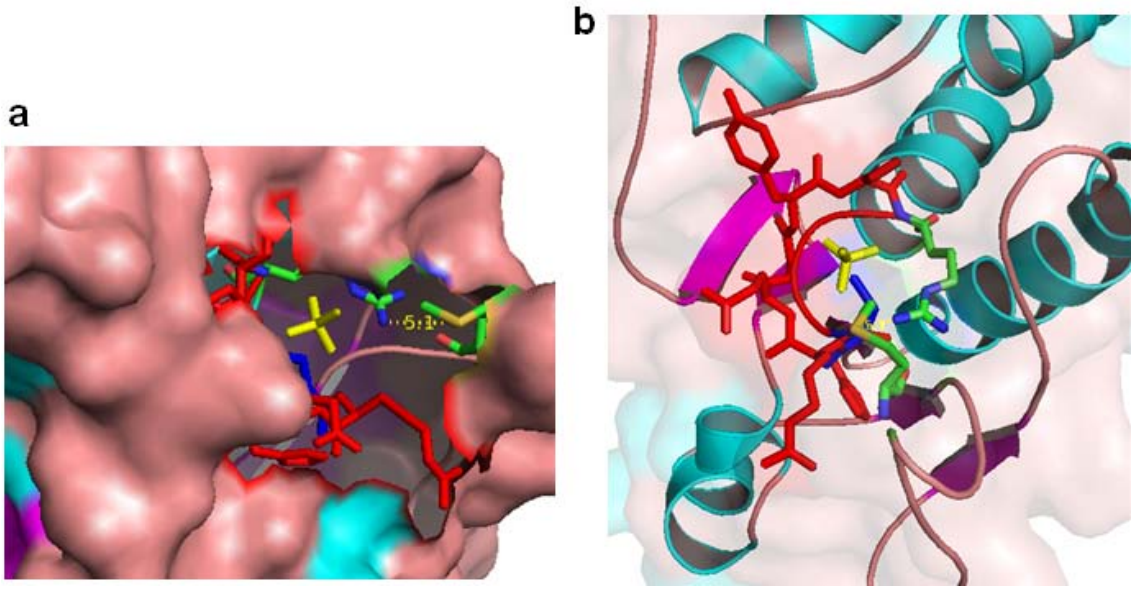


Figure 44: Met69 of the variable insert segment forms hydrogen bond with Arg130 of the P-loop in VHR. (**a**) Zoom-in into the active site pocket of VHR. The structure of VHR (PDB: 1VHR) is shown as surface representation with colouring based on the secondary structure elements (helices in cyan, sheets in magenta and loops in light pink). Sulfate ion bound in the active site is shown as stick (yellow). The residues of the P-loop (from His123 to Arg130) are shown as sticks (Cys124 is blue and other residues from 123-129 red). Met69 and Arg130 are coloured based on atoms (carbon-green, hydrogen-grey, nitrogen-blue, oxygen-red and sulfur-orange). Distance between Met69 and Arg130 was measured to be about 5 Å. Arg130 is at the end of the P-loop and hydrogen bond with Met69 helps to keep it in the optimal position (Yuvaniyama et al., 1996). (**b**) Representation like under (**a**), but the surface spanning residues of the P-loop (from His123 to Arg130) is rendered transparent. All structures were generated using Pymol.

3.2.5 Investigating the self-association of the VHR variants with alternative cross-linking methods

The next task was to further inspect if VHR has an intrinsic ability to form dimers and/or higher order oligomers. To probe the propensity of VHR for dimerisation without using genetically encoded site-specific incorporation of photo-cross-linkable amino acid, more conventional cross-linking approaches were employed. Despite the vastly recognised attraction of this benzophenone-based derivative in probing stably or transiently associated interacting partners via photo-induced cross-linking (Hino et al., 2005; Kage et al., 1996; Krishnamurthy, Dugan et al., 2011; Majmudar et al., 2009; Musial-Siwek et al., 2007; Sato et al., 2011; Uezu et al., 2012; Ye, Köhrer et al., 2008), there are several issues to be considered: (i) It has been demonstrated that placing *p*BPA at the protein-protein binding interface hindered cross-linking, although meeting predicted geometric requirements for detection of the interaction (for example, sufficient distance between the carbonyl group of *p*BPA and preferentially reactive proximal C-H bonds (Sato et al., 2011). (ii) Photoaffinity scanning experiments provided evidence that *p*BPA exhibits more reactivity towards methionine residue in the target domain than for any other amino acid (Tanaka et al., 2008; Wittelsberger et al., 2006a and b; Wittelsberger et al., 2008). Moreover, introduction of single methionine mutations into the N-terminal extracellular domain of the parathyroid hormone receptor (PTHR1) revealed that *p*BPA, which was introduced selectively at two positions in the docking region of the PT hormone (PTH), previously implied in the receptor activation, cross-linked to methionine residues spanning a range of up to 11 amino acids in the receptor (Wittelsberger et al., 2006b). The subsequent work (Wittelsberger et al., 2008) strongly suggested to relax the distance constraint

between the carbonyl moiety of *p*BPA and the proximal C-H bond to at least 10 Å when conducting computer simulations aiming at obtaining the structure of the interacting interface. This opposed a postulation deduced from a detailed mechanism describing the photochemical reactivity of the benzophenone-based photo-cross-linkers (Dormán and Prestwich, 1994), according to which the mentioned distance is not to exceed 3.1 Å if abstraction of hydrogen in the initial reaction step is to take place. Hence, opting for a more traditional approach of probing the oligomerization state of proteins by using the cross-linking reagents glutaraldehyde and DSS to investigate VHR dimerisation was a necessary complementary approach.

First, wtVHR and its F68*p*BPA variant were subjected to cross-linking with glutaraldehyde. The range of the protein concentrations screened was the same like in the photo-induced cross-linking. As indicated in Figure 45a, the monomeric form of the proteins runs as a band at apparent molecular weight of approximately 20 kDa. A second population at about 40 kDa clearly corresponds to the size of a dimer. Incubation with glutaraldehyde also led to the detection of bands corresponding in size to higher-order oligomers. Moreover, in case of the F68*p*BPA variant of VHR, a smear of aggregated proteins was detected at larger molecular masses. Consistent with the results obtained when exposing *p*BPA-variants of VHR to 345 nm light, the bands which size could be attributed to the cross-linking with the truncated product terminated at the position of the amber STOP codon were also detected. Therefore, the results for the F68*p*BPA variant of VHR obtained through two different cross-linking approaches correlated nicely. The control samples, incubated without glutaraldehyde, exhibited no bands in the higher molecular weight range (Figure 45b). To ensure that the protein concentration was proper

and that the equal protein amounts were employed in the reactions, the aliquots of the samples were resolved by SDS-PAGE and stained with colloidal Coomassie (Figure 45c).

Finding the most optimal experimental conditions to boost cross-linking efficiency without accompanying increase in the heterogeneity of the generated conjugates is still largely based on the “trial-and-error” principle (Klockenbusch and Kast, 2010; Leitner et al., 2010). Glutaraldehyde, although wildly applicable as an effective cross-linking reagent, bears somewhat controversial chemistry- up to 13 different forms of this reagent have been identified, depending on the plethora of the reaction conditions, such as pH, temperature or concentration, and even eight different reaction mechanisms in which these diverse forms could engage (Migneault et al., 2004). Unlike genetically encoded photo-activatable amino acids, glutaraldehyde exhibits non-selective reactivity, thus generating a number of intra- and intermolecular connections. Considering debatable features of glutaraldehyde, cross-confirmation of the intrinsic ability of VHR to dimerize was sought by using another type of chemical cross-linking reagent, DSS, which has six carbon atoms in the spacer chain spanning residues over 11 Å. In general, succinimidyl esters classify amongst frequently used conventional chemical cross-linkers targeting primary amino group of lysine residues, and the N-terminal end of the protein (Leitner et al., 2010; Trakselis et al., 2005). wtVHR was used at the concentrations of 50 and 500 µg/mL and was incubated with or without the presence of DSS in the reaction buffer. The subsequent analysis of the reaction mixture by Western blotting (Figure 45d) demonstrated that only when incubation was conducted with DSS, VHR monomers crooss-linked into homodimers, with enhancement in the cross-linking yield being concomitant with the abundance of the protein in the

sample. This result was consistent with the combined results obtained in the previous cross-linking reactions.

Collectively with the results obtained for profiling of the F68*p*BPA mutant of VHR, it could be postulated that regulated variations in the amino acid composition with respect to the native protein might trigger discrete conformational changes with derogative effect on the dimer assembly and/or stabilization. As a result, two reactive moieties may not be optimally positioned to each other what might be reflected in different efficiency of the covalent capture via cross-linking.

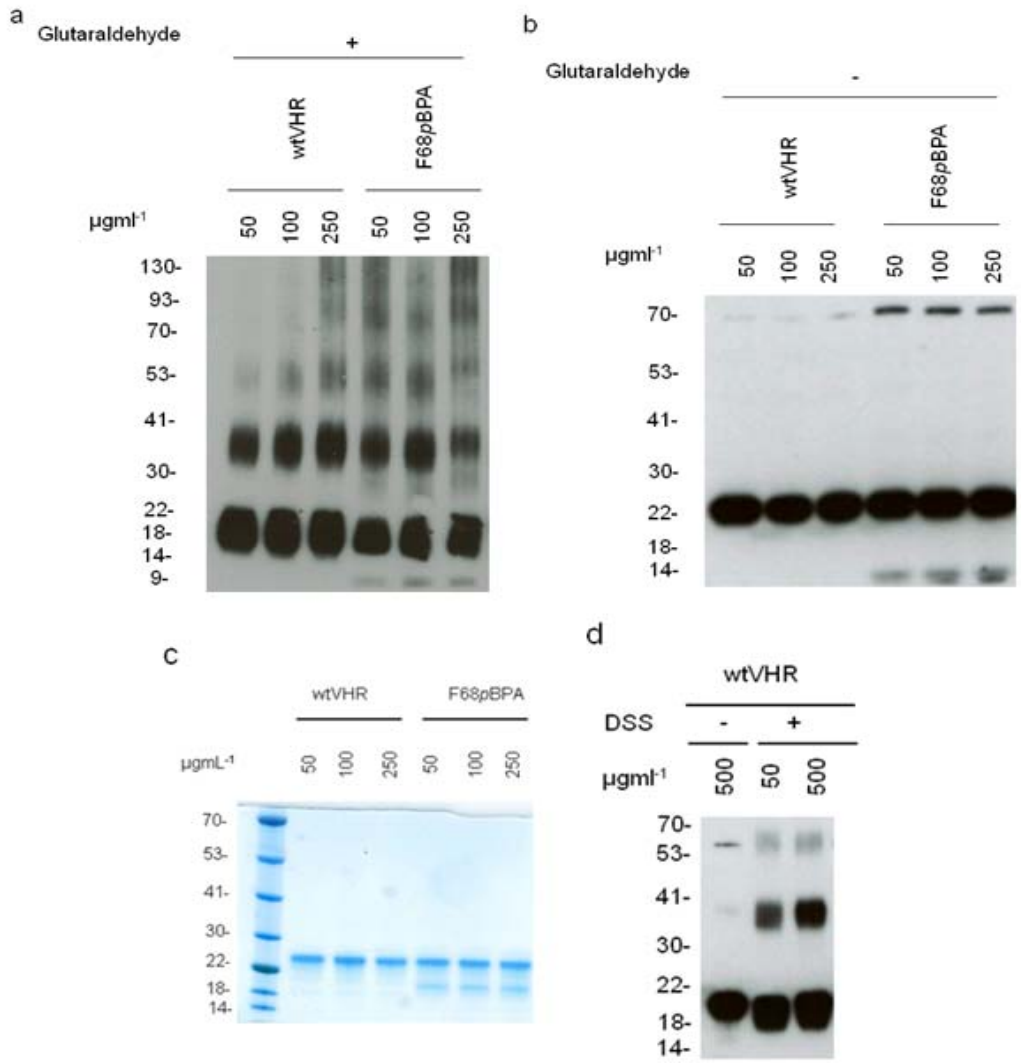


Figure 45: Investigating the potential of VHR to self-associate by chemical cross-linking. (**a**) Cross-linking was done with glutaraldehyde, which was used at 0.025% (v/v). The monomeric forms run as a band at apparent MW of approximately 20 kDa. A second population at about 40 kDa corresponds to the size of the dimer. The faint band at about 30 kDa, present only in the glutaraldehyde-treated sample of the $Flag_3$-tagged F68*p*BPA variant, matched the MW of the crosslinked complex between the full length and the truncated form. Bands at higher MW (in the range 53-130 kDa) correspond to the higher order oligomers and are the result of the non-specific cross-linking by GA. (**b**) The control samples were incubated without glutaraldehyde. (**c**) Estimating that equal protein amounts were used for cross-linking by resolving 1 µg of the protein samples (without glutaraldehyde) by 4-12% SDS-PAGE and staining with colloidal Coomassie. (**d**) Cross-linking of wtVHR with DSS showed dimer formation. The cross-linked species were detected by Western blotting (anti-VHR). The results reflect the minimum of two independent experiments.

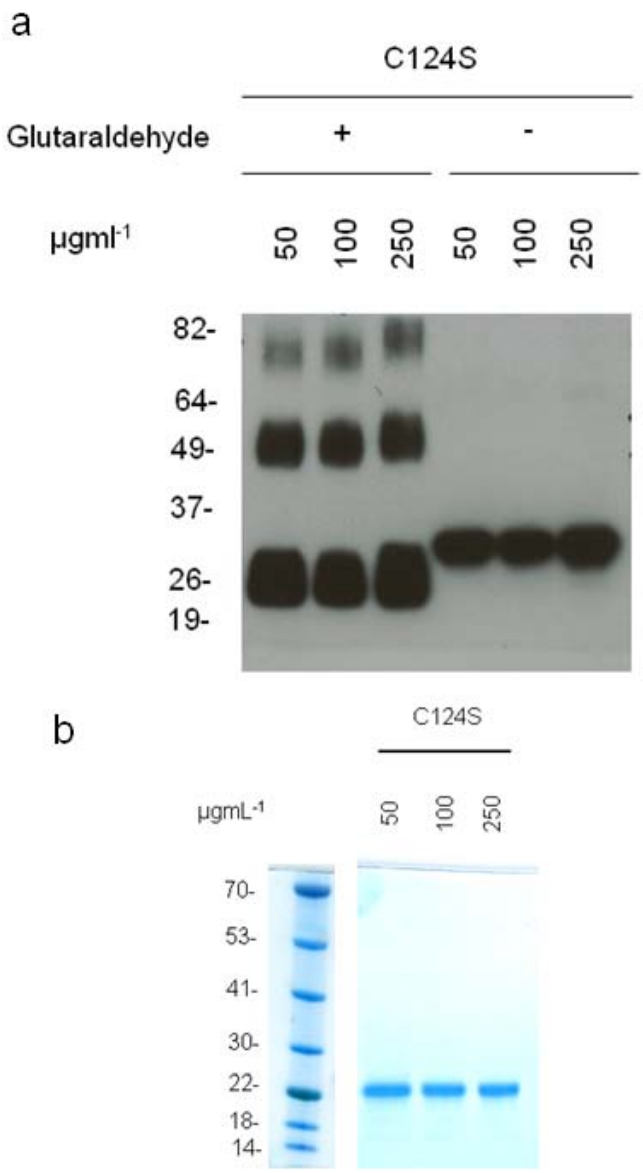


Figure 46: Investigating if the potential of VHR to self-associate depends on the catalytic cysteine by chemical cross-linking. (**a**) Glutaraldehyde was used from 0.05% (v/v) stock to top up to 50 µl reaction volume. The monomeric forms run as a band at apparent MW of approximately 20 kDa. A second population at about 40 kDa corresponds to the size of the dimer. Band at about 80 kDa correspond to the size of the trimer and are the result of the non-specific cross-linking by GA. The control samples were incubated without glutaraldehyde. The cross-linked species were detected by Western blotting (anti-VHR). The results reflect the minimum of two independent experiments.
(**b**) Estimating that equal protein amounts were used for cross-linking by resolving 1 µg of the protein samples (without glutaraldehyde) by 4-12% SDS-PAGE and staining with colloidal Coomassie.

3.2.6 Profiling the structural effect of position 68 on the self-association of VHR by chemical cross-linking with glutaraldehyde

The next major focus was to systematically profile a series of mutants generated for the position of Phe68 in order to rigorously investigate the impact of steric and/or electrostatic features of this patch on the establishment of VHR homodimers. The variants of VHR with Phe at position 68 replaced by alanine, aspartate, tryptophan or cysteine, in addition to the previously obtained photo-crosslinkable *p*BPA-variant, were used. Next, by subjecting the samples to the cross-linking with glutaraldehyde across a broad protein concentration range, it was assessed if any discernible effect on the dimer formation could be detected. All newly generated Phe68-variants of VHR were able to form dimers, and higher order oligomers, upon incubation with glutaraldehyde (Figure 47). Exchanging a hydrophobic aromatic residue in the native protein (phenylalanine) for a small one (alanine) or a polar, non-charged one (cysteine) led to a more pronounced effect on the self-association of the protein than exchanging it for an aliphatic charged residue (aspartate) possibly by enabling closer association of the monomeric units, whereas the cross-linking profile for the wtVHR and its F68D and F68W variants was comparable (Figures 45 and 47). One could envision several mechanisms through which the latter could disrupt intermolecular association, for example by exerting electrostatic repulsion or by favouring the formation of salt bridges and/or polar interactions with the proximal residues that would obliterate fixing of the most optimal orientation of the monomeric units of VHR. In addition, the M69A variant also exhibited cross-linking pattern similar to the one obtained for the wtVHR. Hence, its predominant role seems to be the stabilization of the active site (Yuvaniyama et al., 1996).

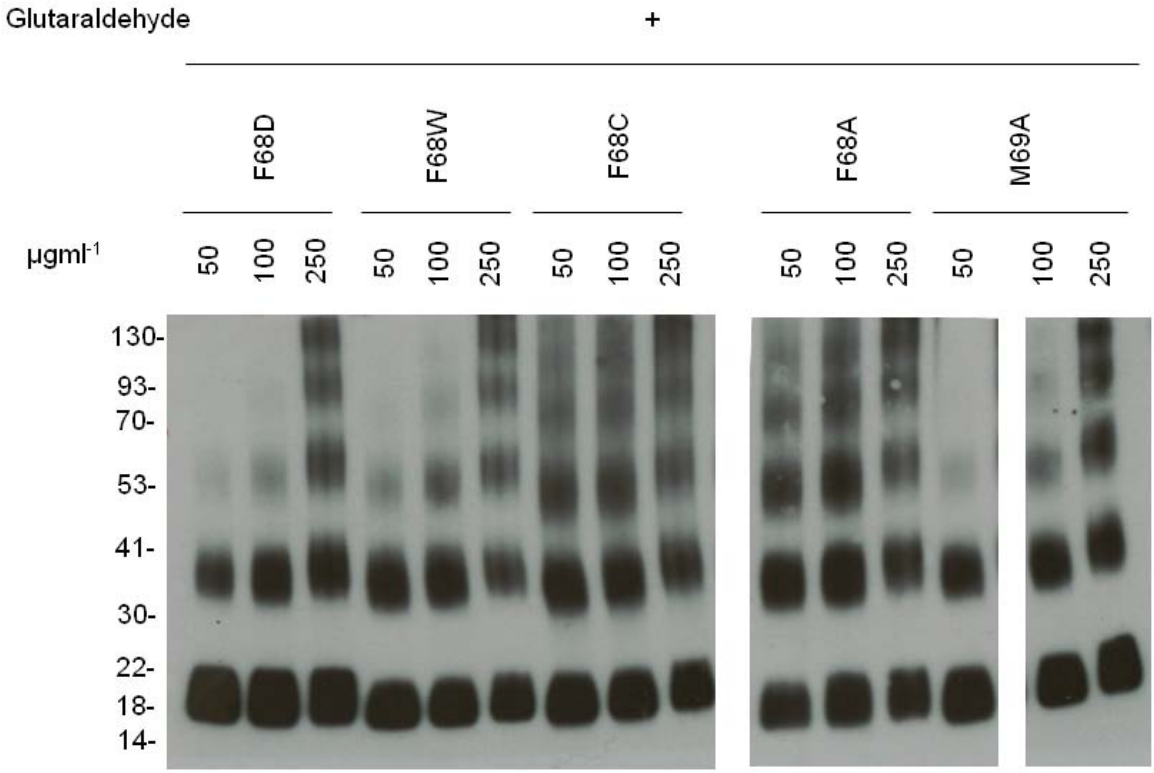


Figure 47: Profiling the effect of the structural and electrostatic properties of the residues incorporated at position 68, on the ability of VHR to self-associate, by chemical cross-linking with glutaraldehyde. Phe68 was systematically replaced with alanine, aspartate, tryptophan and cysteine. The experiment was conducted according to the protocol described in section 2.3.6. The samples were incubated with 0.025% (v/v) of glutaraldehyde in 20 mM HEPES pH 7.4 for 30 min at 37°C and the final reaction volume of 50 µl. All proteins were used at three different concentrations (50, 100 and 250 µg/ml). The M69A variant was also analysed when profiling the involvement of the "variable insert" segment in dimerisation of VHR. The control samples are shown in Figure 48. The monomeric forms run as a band at apparent MW of approximately 20 kDa. A second population at about 40 kDa corresponds to the size of the dimer in the glutaraldehyde treated samples. Bands at higher MW (in the range 53-130 kDa) correspond to the higher order oligomers and are the result of the non-specific cross-linking by GA. The result is representative of two experiments. Western blot was done by using anti-VHR ab.

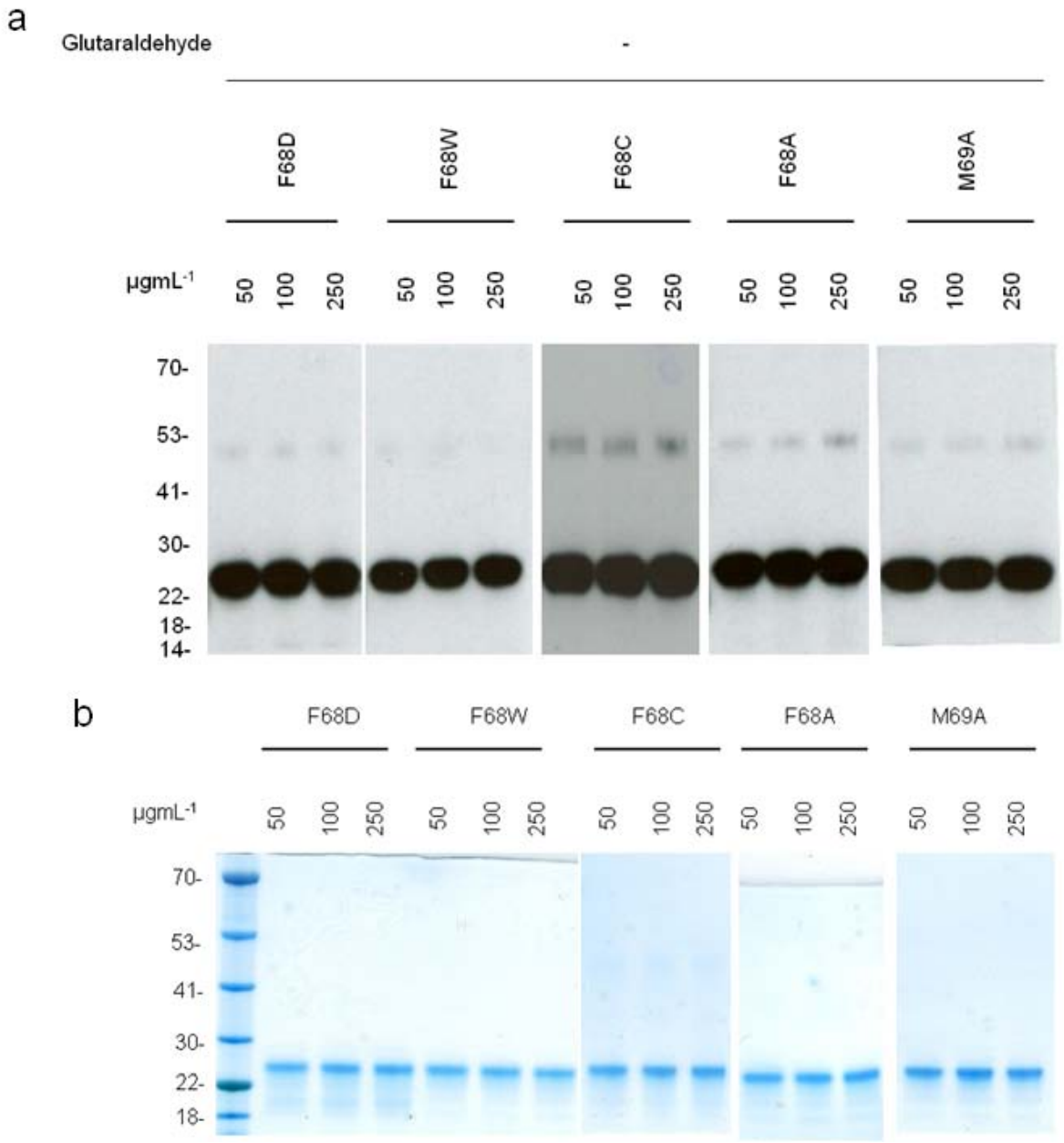


Figure 48: The control samples for Figure 47. They were processed in the same way, but without glutaraldehyde. For panel (**a**), the effect of the amino acid substitution on the potential of VHR to self-associate was checked by Western blot (anti-VHR). The results represent the minimum of two independent experiments. (**b**) To estimate that equal amounts of VHR variants were used for cross-linking, an 1 µg of of each sample (without glutaraldehyde) was resolved by 4-12% SDS-PAGE and allowed to stain with colloidal Coomassie overnight, with gentle agitation at room temperature, prior to distaining in ddH_2O.

Figure 48b demonstrates that protein concentrations were equal.

The control samples for the reaction with glutaraldehyde were incubated without the cross-linking reagent in the reaction buffer (Figure 48a) and exhibited no oligomer formation. It might be worth noting that in the control samples, F68C variant showed band at the expected size of a dimer, presumably due to disulfide bond association which could not be reversed upon boiling of the samples. Association through disulfide bond formation which was not completely reversible under reducing conditions was also reported for laforin (Liu, Wang et al., 2006). Intermolecular association through disulfide bond formation has been regarded as the most frequent principle governing intermolecular association (Nardozza et al., 2012). Moreover, the ROSs were recognised to mediate intermolecular association amongst PTPs through disulfide bond formation (Blanchetot et al., 2002; Deb et al., 2011; Sun, Wang et al., 2005).

3.2.7 Investigating dimerisation potential of the VHR variants with respect to oxidising conditions

As a result from the previous experiments (shown in the Chapter 3.2.6), the next step was to investigate whether generating F68C mutant renders the protein more susceptible to intermolecular disulfide bond formation.

The samples of the wtVHR and its F68C and F68*p*BPA variants were examined for dependence of the oligomerization state on the oxidising conditions. This was conducted by dialyzing the samples in the buffer without the reducing agent and by running SDS-PAGE under standard or non-reducing conditions. For the latter, the samples were incubated with 2 x SDS-PAGE sample buffer which did not contain any reducing agent, and the samples were not boiled at 95°C, but just incubated for 30 min at 37°C. The result is shown in Figure 49. Arrows indicate the position of the apparent molecular weight of the dimer, triggered by oxidative conditions-mediated disulfide bond formation. Moreover, in the lane for the samples of F68C and F68*p*BPA variants ran under non-reducing SDS-PAGE conditions, at least two bands could be distinguished corresponding to the oxidised (faster migrating) and reduced form.

As indicated in Figure 49, the F68C mutant showed an increased potential for self-association via intermolecular disulfide-bond formation than the wt protein. Under reducing SDS-PAGE conditions, hardly any homodimers could be detected anymore, suggesting that the self-association of all VHR variants examined was reversible. Together with the results obtained in the cross-linking experiments, this observation strongly suggests that spatial proximity of two VHR molecules is such that the distance between the C_α atoms of the cysteine residues is within the range of 4.4-5.9 Å which can be spanned via disulfide

bond (Sun, Wang et al., 2005). Strikingly, the F68*p*BPA mutant exhibited a non-reducing SDS-PAGE pattern more resembling that of the F68C variant than the one of the wild type protein. “Variable insert” region contains a small hydrophobic and protruding patch of non-conserved amino acid composition (Yuvaniyama et al., 1996). As stated previously, Phe68 is a part of this island. Exchanging phenylalanine for a more hydrophobic *p*BPA is likely to not only further increase hydrophobicity of the rim, but also to increase the extent of the hydrophobic interactions, thus further stabilizing intermolecular self-association of VHR. In addition to the catalytically active cysteine of the P-loop, VHR has three more cysteine residues at positions 22, 30 and 171. Careful structural examination would support that the localisation and solvent exposure of Cys22 might suffice for an increase in disulfide mediated self-association observed for F68*p*BPA variant of VHR (Figure 49), possibly enabled by the stabilizing effect the F68*p*BPA substitution might have, like just discussed.

For VHR it was reported previously that treatment with H_2O_2 triggered transient conversion of the catalytic site Cys124 to sulfenic acid (Denu and Tanner, 1998). The same study reported that conversion of sulfenic acid to non-reducible forms, namely to sulfinic and sulfonic acid, was not detected in VHR even under stringent oxidising conditions. The authors pointed out that the lack of excessive oxidation in VHR might be due to the active site Cys124 being protected from the solvent exposure. It is tempting to hypothesize that the shielding mechanism might be provided by the proximal self-associated unit of VHR.

In addition, in PTPs there is a diversity of mechanisms reported which serve to protect the catalytic cysteine from excessive irreversible oxidation and to recuperate the catalytically active form (Bonham and Vacratsis, 2009; Buhrman

et al., 2005; Chen et al., 2009; Östman et al., 2011; Seth and Rudolph, 2006; Silver et al., 2013; Sun, Wang et al., 2005; Tonks, 2005). They span from the intramolecular disulfide bond formation, like in case of the PRLs (Sun, Wang et al., 2005) or PTEN (Lee, Yang et al., 2002), across cyclic sulphenamide, like for PTP1B (Tonks, 2005), to disulfide switches, reported for MKP-3 (Seth and Rudolph, 2006), disulfide bond formation between the backdoor cysteines, in case of SHPs (Chen et al., 2009) or more custom protective mechanisms, like reported for DUSP12, also known as human orthologue of yeast (hYVH1) (Bonham and Vacratsis, 2009), which was shown to form disulfide bonds within catalytic and zinc binding domains. Dimerisation could be an interesting mechanism to protect PTPs from irreversible oxidation.

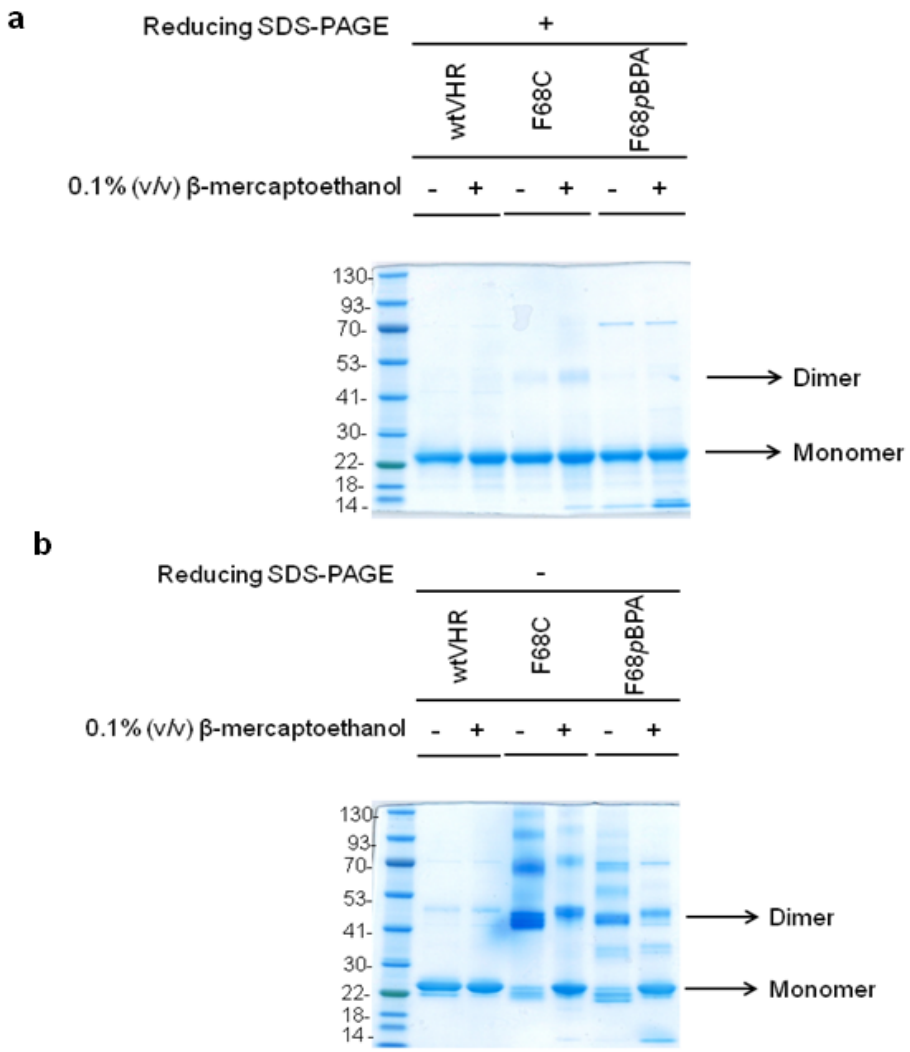


Figure 49: Analysing dimerisation potential of the recombinant VHR variants with respect to oxidising conditions. The procedure is described in section 2.3.8. 3 µg of the wtVHR, F68C and F68*p*BPA, stored with or without the presence of 0.1% (v/v) β-mercaptoethanol, were resolved by 4-12% SDS-PAGE under non-reducing or standard conditions and stained with colloidal Coomassie (panels (**b**) and (**a**), respectively). Monomers run as a band at about 23 kDa. Under non-reducing SDS-PAGE (**b**) additional band probably corresponding to the sulfenic acid-form is also observed. Dimers run at about 50 kDa and are mostly completely reversible under reducing SDS-PAGE conditions (**a**). The bands of higher MW (from 53 kDa) especially prominent in the sample of the F68C, and to a somewhat lesser extent to the F68*p*BPA, variants under non-reducing SDS-PAGE (**b**) most likely correspond to different disulfide-bond-based associations. The band at about 70 kDa, visible in the sample of F68*p*BPA variant, was detected regularly after purification. The result is representative of at least two independent experiments.

3.2.8 Investigating the dimerisation potential of VHR in cells

The findings from the cross-linking experiments prompted an investigation if the homodimeric association of VHR could be detected in cells. For that purpose, mammalian COS-1 cells were transfected with myc- and $Flag_3$-tagged VHR constructs. By IP with an anti-Flag ab and the subsequent probing of the immunoblots with an anti-myc ab, co-IP of $Flag_3$-tagged VHR was demonstrated. Interestingly, self-association of VHR was detected only after oxidative stress stimulus induced by treating the cells with 10 mM H_2O_2 for 60 min at 37°C prior to analysis (Figure 50), supporting the hypothesis that dimerisation could be a consequence of oxidative stress, as implicated from the previous results. To exclude the possibility that the detected self-association of VHR was the result of the selected experimental setup, 3Flag-PRL-3 ΔCAAX was utilized as a control. The ΔCAAX variant of PRL-3 had deleted four residues of the CAAX-box that are important for localizing the protein to the plasma membrane and structures resembling endosomes (Rios et al., 2013; Zeng et al., 2000). Moreover, for PRL-1 it was shown that trimerisation is required for PRL-1-mediated cell growth and migration (Sun, Luo et al., 2007), whereas the same study also demonstrated the intrinsic ability of PRL-3 to trimerize. 3Flag-PRL-3 ΔCAAX was transfected with myc-tagged VHR into COS-1 cells. As shown in Figure 51, 3Flag-PRL-3 ΔCAAX was not co-immunoprecipitated when performing IP of the COS-1 lysates with an anti-myc ab, confirming the specificity of the interaction to VHR.

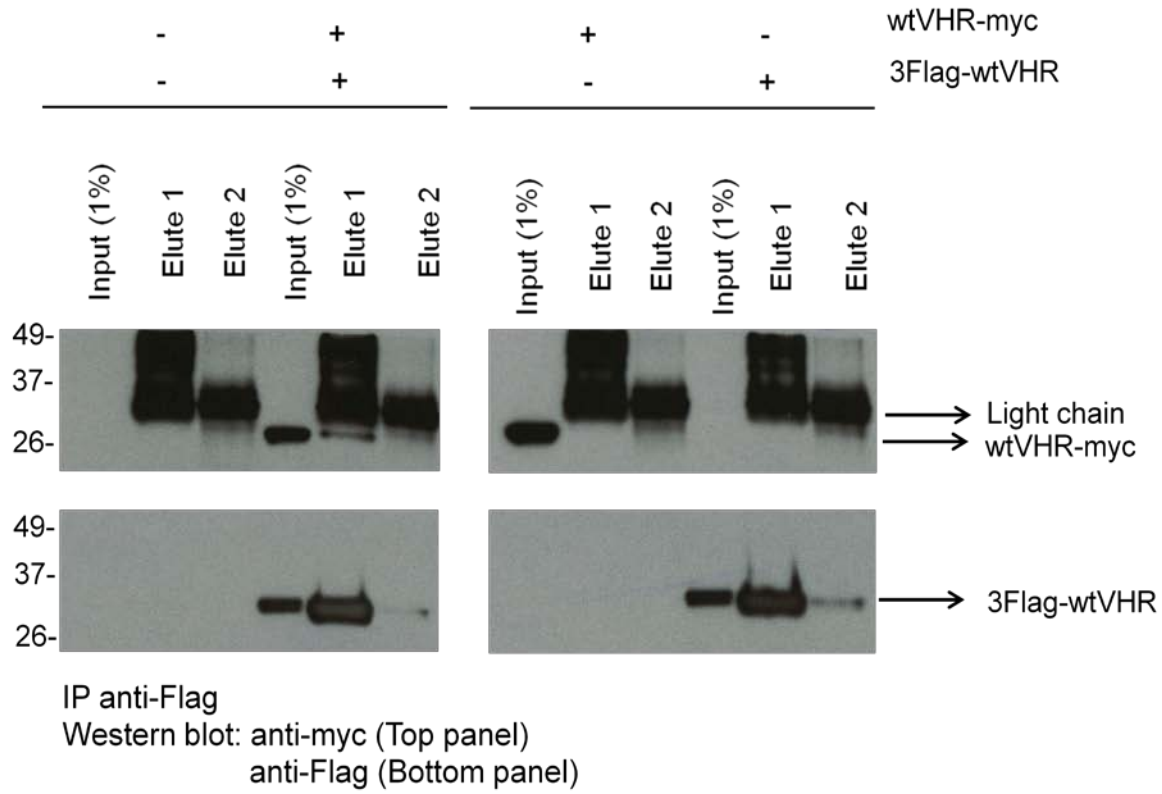


Figure 50: Self-association of VHR in H_2O_2-stimulated COS-1 cells. COS-1 cells were transiently co-transfected with myc- and $Flag_3$-tagged wtVHR. The analysis was conducted 24 h post-transfection. The cells were first placed in SFM for 2 h at 37°C, followed by stimulation with 10 mM H_2O_2 for 1 h at 37°C prior to lysis. IP was done by using ab against the Flag-tag. Elute 1 was done by applying the non-reducing 2 x SDS-PAGE sample buffer to the samples and incubating for 10 min at 65°C and Elute 2 by incubating for 5 min at 95°C. Immunoblots were probed for VHR dimerisation by using anti-myc ab. Afterwards, the membrane was stripped and re-blotted using ab against the Flag-tag. The detailed protocols are described in sections 2.3.11 and 2.3.12. The image is representing the result of two independent experiments.

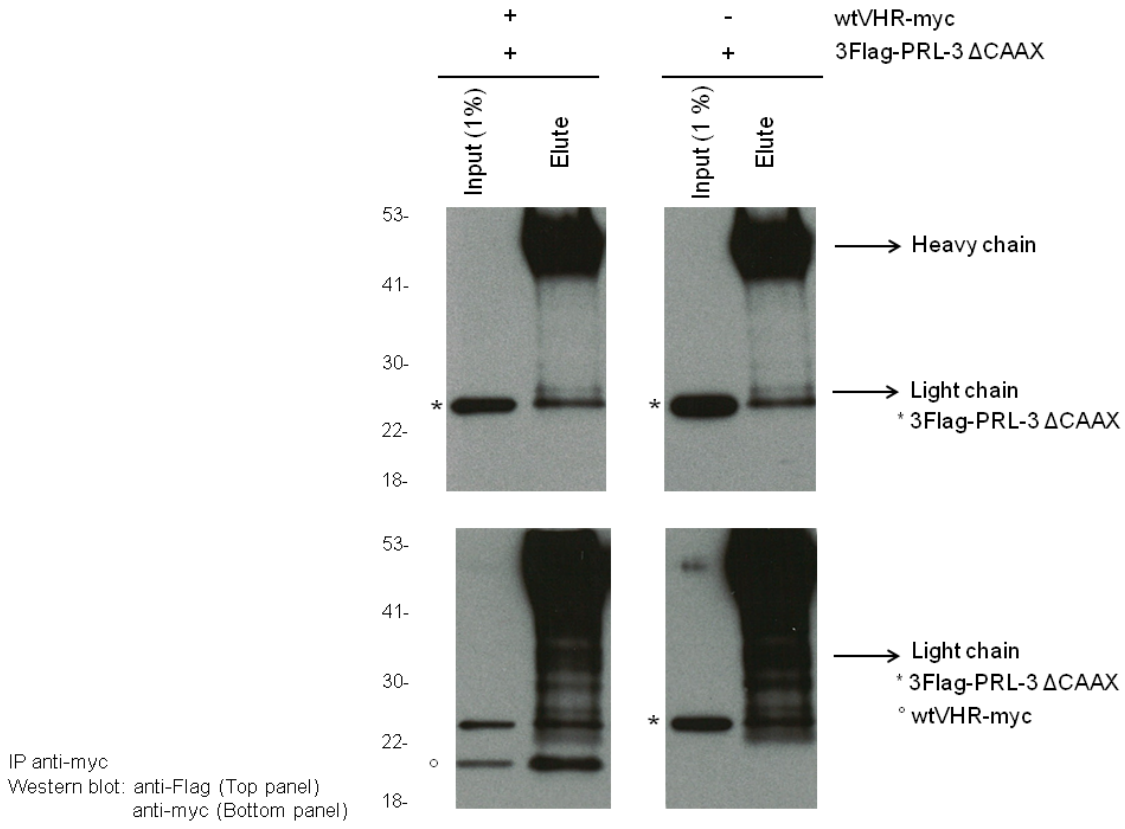


Figure 51: Assessing specificity of dimerisation to VHR, in cells. The detailed protocol is described in sections 2.3.11 and 2.3.12. COS-1 cells were co-transfected with plasmids pSG5/wtVHR-myc and pCMV-3Flag-1A/PRL-3 ΔCAAX. About 24 h after transfection, the cells were treated with 10 mM H_2O_2. IP of the COS-1 lysate was conducted overnight at 4°C by using anti-myc ab. The first analysis was done by using ant anti-Flag ab (top panle). After that the filter was stripped and re-blotted with an anti-myc ab (bottom panel) (however, the presence of a band corresponding in size to the 3Flag-tagged PRL-3 in the bottom panel (marked with °) clearly indicates that stripping of the filter was not absolutely efficient or cross-reactivity of the ab). The bound material was eluted by applying the non-reducing 2 x SDS-PAGE sample buffer to the samples and by incubating for 10 min at 65°C. The band below the arrow indicating the light chain is non-specific and was observed in the IP experiments using anti-myc ab. Band at about 23 kDa corresponds to the 3Flag-tagged PRL-3 (marked with *) and at about 20 kDa to the myc-tagged VHR (marked with °). The image represents the result of two independent experiments.

3.2.9 Investigting the dimerisation potential of VHR in cells independently of the oxidative stress

When it comes to detecting protein-protein interactions *in vivo*, PFA is the rather appealing choice of chemical cross-linker. It is a nonspecific cell-permeable monofunctional reagent with the length of the spacer arm reportedly spanning 2.3-2.7 Å, thus enabling it to capture even transiently tethered spatially proximal complexes (Klockenbusch and Kast, 2010; Vasilescu et al., 2004). Therefore, formaldehyde was selected as a suitable reagent to probe if self-association of VHR in cells could be detected also in the absence of H_2O_2-induced oxidative stress. COS-1 cells were transfected with myc-tagged wtVHR and $Flag_3$-tagged C124S mutant, as it had been shown previously that self-association of VHR did not depend on the availability of the catalytic Cys124 (Figures 38 and 45). To avoid losing the read-out (Klockenbusch and Kast, 2010) of the *in vivo* association of two differentially tagged VHR constructs, the samples were split in two and subjected to heating at 60°C for the total of 20 min and, alternatively, to heating at 95°C for 5 min, prior to being resolved by SDS-PAGE and analysed by Western blotting using ab against the affinity tags. Figure 52 shows that the self-association of VHR was stable under the conditions applied and no reversibility of the cross-links was observed when the samples were under more stringent temperature conditions. The control dish treated in an analogous fashion but with formaldehyde-free PBS resulted in detection of no complexes of higher molecular masses. Interestingly, in the samples incubated with PFA for 2.5 h, a reduction in the yield of the cross-linked complexes was observed, compared to the sample with 30 min incubation with PFA. This was in accordance with the previously acknowledged

issues that ought to be considered in the reactions with chemical cross-linkers (Klockenbusch and Kast, 2010). For example, the cross-linked complexes may not be soluble or they may be so conformationally distorted so as to be rendered non-recognisable by the ab.

Capturing two differentially tagged VHR variants by means of covalent bond formation induced via chemical cross-linking reagents is a direct and persuasive indication of their spatial proximity. Therefore, it is possible that self-association of VHR can be stabilized through oxidative stress and/or by using already short-span cross-linking reagents, chemical or genetically encoded, as demonstrated.

.

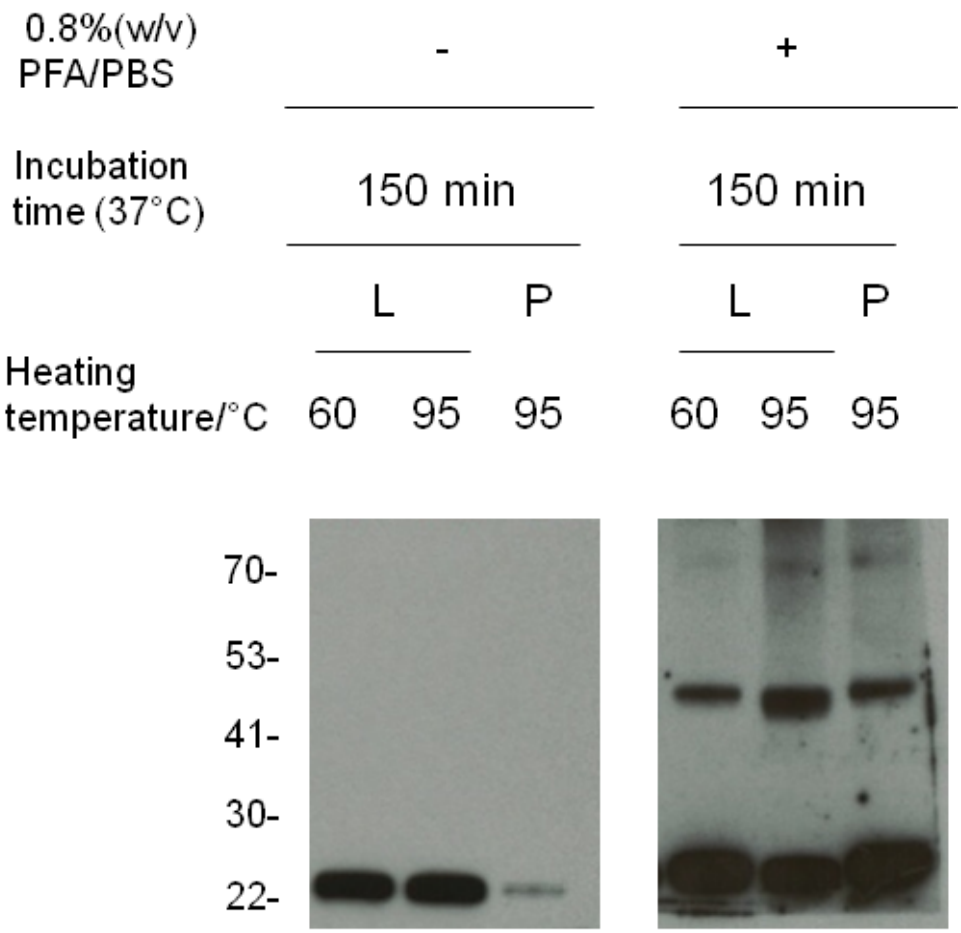


Figure 52: VHR can form dimers in cells in the absence of H_2O_2-induced oxidative stress. Investigating VHR dimerisation in COS-1 cells in the absence of H_2O_2-induced oxidative stress was done by cross-linking with PFA, according to the protocol in section 2.3.13. COS-1 cells (60 mm) were transfected with pCMV-3Flag-1A/wtVHR and after 24 h probed for dimeric association of VHR inside the cell. The cells were treated with 0.8% (w/v) PFA/PBS at 37°C and then lysed. Control sample was incubated only in the GM only for the total of 150 min. In order to preserve the cross-linked complexes, the aliquots of the soluble fraction (L) were mixed with 2 x SDS-PAGE sample buffer, without reducing agent, and heated at 60°C for 20 min or at 95°C for 5 min. To ensure that the complexes are not in the pellet due to extensive, non-specific cross-linking, the pellet (P) was resuspended in 80 µl of non-reducing 2 x SDS-PAGE sample buffer and heated at 95°C for 5 min. 20 µl of the lysate (about 5.6% of the total lysate) and 10 µl of the resuspended pellets (12.5%) were resolved by 4-12% SDS-PAGE and probed with an anti-Flag antibody. The result is indicative of two independent experiments.

3.2.10 Investigating functional implications of VHR dimerisation

To assess the effect of the dimer formation on the enzymatic activity of the protein, the native VHR and its F68*p*BPA mutant, both at at 250 µg/mL, were subjected to 345 nm light for 30 min. Recovered samples were then subjected directly to the activity measurement by employing *p*NPP at 10 mM concentration as a substrate (sections 2.3.5 and 2.3.4, respectively). The protein concentration was adjusted so that to match the conditions from the kinetic profiling experiments (results of which were presented in Table 4 and Table 5 in section 3.2.2). As indicated in Figure 53, it was observed that the sample of the 68*p*BPA variant containing covalently cross-linked dimer exhibited reduced catalytic efficiency with respect to the non-irradiated sample. For the wtVHR, there was no difference in the catalytic activity against *p*NPP regardless of its exposure to the UV light. This observation indicated that exposure to 345 nm light did not hamper the catalytic performance of the protein, under the experimental setup applied, and, moreover, did not result in any damage to the protein. The latter reflected the results from the previous experiments (Figures 37-39). In addition, the efficiency of the non-irradiated sample of the F68*p*BPA variant in dephosphorylating *p*NPP was not altered compared to the native protein. Moreover, the level of the activity demonstrated for the non-irradiated sample of the F68*p*BPA variant well reflected the results obtained in the in-depth kinetic profiling conducted for all VHR variants (Table 4).

In order to obtain a more physiologically relevant model to further evaluate if self-association of VHR might have significant effect in modulating its catalytic

activity in cells, the levels of dephosphorylated Erk1/2 in COS-1 cell extracts after incubation with recombinant VHRs were assessed. It has been demonstrated that Erk1/2 are authentic substrates of VHR *in vitro* and *in vivo* (Todd et al., 1999). To inhibit the endogenous phosphatases, the approaches suggested in Blanchetot et al. proved successful (Blanchetot et al., 2005). By treating the cells with a mixture of sodium orthovanadate and H_2O_2, the endogenous phosphatases should be inhibited and the effect of dephosphorylation of a targeted substrate from the cell lysate would be the result of the added recombinant protein.

Overall, reduction in the catalytic performance, excluding a derogative effect of the UV exposure on the protein stability, is clearly indicative of the non-availability of the active site pocket to recognise and/or accommodate and dephosphorylate the substrate. Like indicated previously when explaining plausible lack of increase in the dimers formed when increasing the concentration of the F68*p*BPA variant from 250 to 500 µg/ml (Figure 37), one likely explanation might also be attributed to the intramolecular cross-linking. In the light of the results presented in Figures 53 and 54, the possibility remains that the observed reduction in the activity is also, at least in part, resulting from obstructions to the entrance into the crevice of the active site due to intramolecular cross-linking.

The native protein and its F68*p*BPA variant were subjected to 30 min exposure to 345 nm light, at 250 µg/ml protein concentration, as described in section 2.3.5.

As indicated in Figure 54, in the buffer control samples no significant dephosphorylation of Erk1/2 was observed, thus suggesting that the

endogenous VHR and other Erk1/2-directed phosphatases had been transiently inhibited by pre-treatment of the cells with a mixture of 1 mM orthovanadate and 1 mM H_2O_2 for 30 min at 37°C, whilst simultaneously enriching for the pool of tyrosine-phosphorylated proteins (Blanchetot et al., 2005). Exposure of the wt protein to 345 nm had no detrimental effect on the protein stability or catalytic activity, as demonstrated by probing with the anti-VHR and anti-pErk1/2 ab, respectively. Interestingly, the level of phosphorylated Erk1/2 in the sample incubated with F68*p*BPA variant after its exposure to 345nm light was almost at the level of the buffer control and in stark contrast to its non-irradiated homologue, which in turn exhibited catalytic efficiency at the level of the wt protein. Based on the results presented in Figures 53 and 54, it can be postulated that the reduction in the catalytic activity of the irradiated F68*p*BPA variant arose from its UV exposure. One possible explanation would be a hypothesis addressing the possible topology of the dimeric association. Accordingly, the covalent capture of the two monomeric units into a dimer might curtail efficient catalytic performance by providing sterical hindrance to the active site pocket.

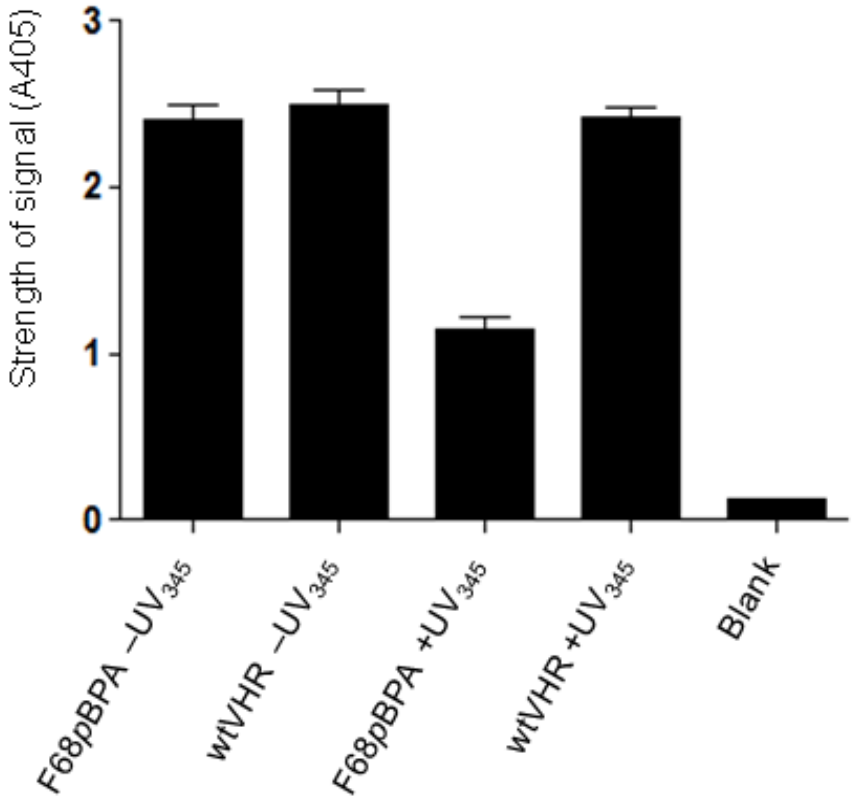


Figure 53: Assessing functional implications of VHR dimerisation against *p*NPP. The protocols are described in sections 2.3.4 and 2.3.5. The effect of dimer formation on the catalytic efficiency of the F68*p*BPA mutant of VHR was investigated by employing *p*NPP as a substrate. The protein samples were subjected to 345 nm light for 30 min at 250 µg/ml concentration in a 96 well-plate format, followed by dilution to 500 nM before subjecting to the activity assay with *p*NPP at 10 mM. The data were analysed using GraphPad Prism 5.0 software. For clearer representation, the result is shown as an end-point measurement for a discrete time point of 46 min. Three sets of triplicate were analysed for each sample.

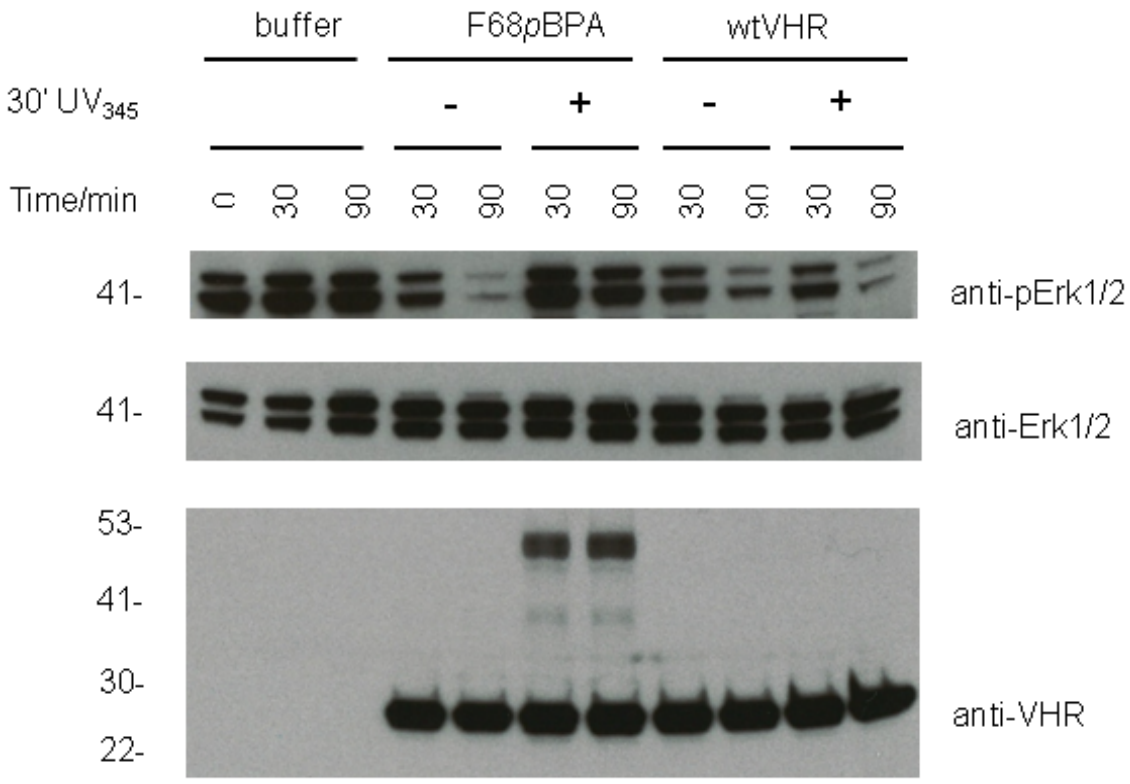


Figure 54: Assessing the effect of VHR dimerisation on the ability to dephosphorylate Erk1/2. The detailed protocol is described in section 2.3.14. In brief, the cells were serum-starved for 16-24 h then treated for 30 min at 37°C with a mixture of 1 mM orthovanadate and 1 mM H_2O_2 in an SFM to inhibit the endogenous phosphatases. Lysates were incubated with 500 nM concentration of recombinant VHRs (with or without prior exposure to 345 nm light at 250 µg/ml), or an equal volume of buffer, for 0, 30 and 90 min at 30°C. Time-dependent reduction of pErk1/2 in COS-1 extracts was analayzed with anti-pErk 1/2 ab (upper panel). Anti-Erk 1/2 (middle panel) and anti-VHR (lower panel) ab were used to assess equivalent amounts of lysates and bait used, respectively. The result is representative of three independent experiments.

3.2.11 Analysis of VHR dimers after separation by size-exclusion chromatography

Two attempts to separate the covalently cross-linked dimer of F68*p*BPA variant of VHR by size-exclusion chromatography were conducted and none was successful in isolating pure complex of a dimer. Based on the size difference between the non-cross-linked monomeric form and the covalently cross-linked dimer of the F68*p*BPA variant of VHR, it was surprising that the separation did not succeed. This was attributed to a possible high mutual affinity of VHR entities present in the sample. However, in the second attempt, several fractions enriched for the dimer were obtained (Figure 55). In addition, pure sample of the monomeric form of the protein was also isolated. This enabled examining previously mentioned hypothesis of the intramolecular cross-linking in (i) stalling more efficient yield in the photo-induced cross-linking and (ii) causing reduced catalytic performance of the irradiated sample of F68*p*BPA (Figures 53 and 54). Figure 56 demonstrates that the fraction of the isolated monomeric form of F68*p*BPA also showed a reduced catalytic efficiency in comparison to the wt protein. The wtVHR was not irradiated in this case as it had been demonstrated previously that irradiation had no derogative effect on its stability or catalytic performance (Figures 53 and 54). When comparing the results of the irradiated sample of F68*p*BPA (data shown in Figure 53) (the sample was assigned as “Mixture” in Figure 57) with the monomeric form isolated by size exclusion (Figure 56), and presenting the data as relative activity with respect to the wt, it can be seen that the sample of the F68*p*BPA_Mixture shows further reduction in the catalytic efficiency in

comparison to the irradiated monomeric form separated by size-exclusion (Figure 57).

Therefore, the additional protein forms present in the mixture sample (dimer and the cross-link of the full length protein to the variant truncated at the position of the amber STOP codon) could cause a further reduction in the catalytic efficiency. For example, because the self-association tendency appeared to be such that the sample was not able to be purified by size-exclusion chromatography, the truncated form might linger proximal to the monomeric form in a way that would block access of a substrate into the active site pocket. A reasonable explanation would be that the conformation of the covalently cross-linked dimer might be such that the entry into the crevice of the active site is not accessible.

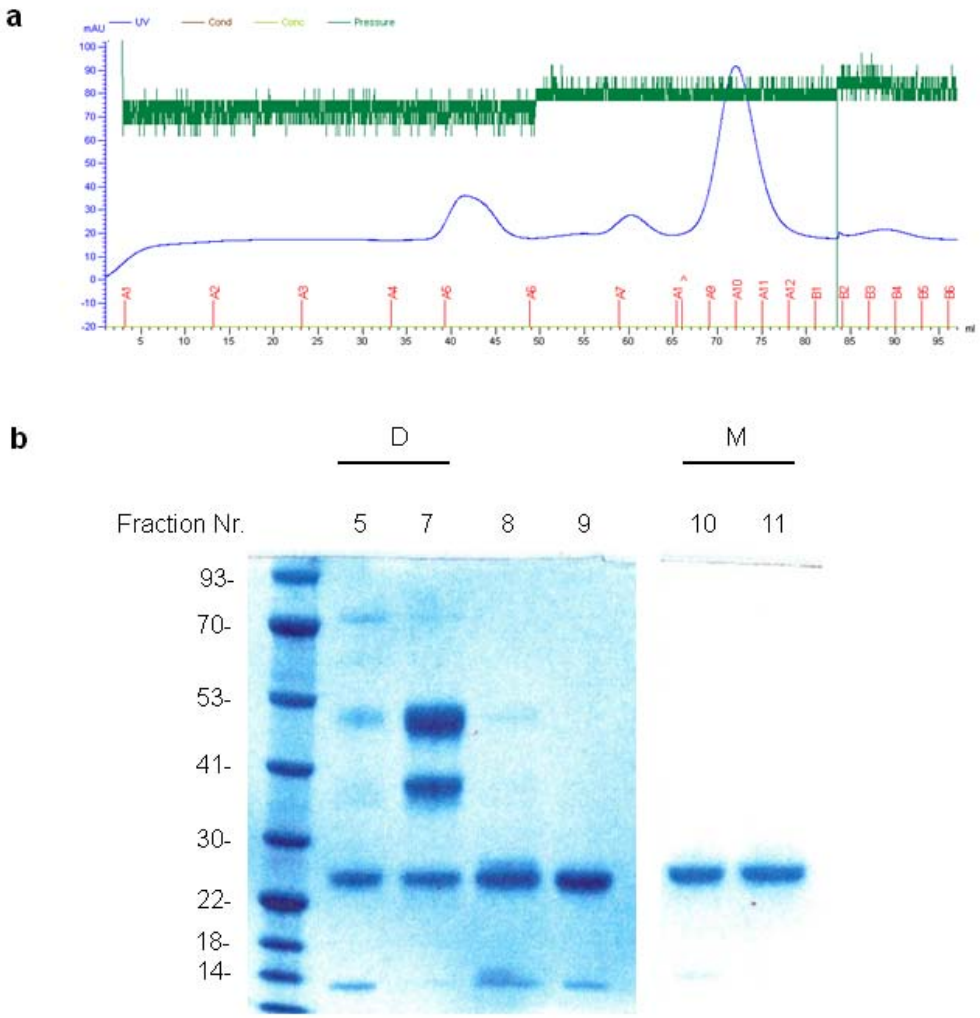


Figure 55: Separation of F68*p*BPA dimer by size-exclusion chromatography. Detailed protocol is described iin section 2.3.9. The sample was subjected to 345 nm light for 30 min in 96-well plate setup, with the protein concentration set to about 20 mg/ml. The recovered sample was diluted about 15-fold and subjected to separation by size-exclusion. (**a**) The sample was loaded at 0.5 ml/min on a 16/60 Superdex 75 prep grade column, and the buffer used was 50 mM Tris-HCl (pH 7.4), 500 mM NaCl and 0.1% (v/v) β-mercaptoethanol. (**b**) The collected fractions were concentrated (0.22 µm Millex-GP filters), and the concentration was measured by NanoDrop 1000 Spectrophotometer and additionally validated by resolving 2 µg of the protein by denaturing 4-12% SDS-PAGE and staining with colloidal Coomassie. Both fractions enriched for the monomeric form were combined (sample assigned as M), and used in the activity assays against *p*NPP. The fraction enriched for the dimer (sample assigned as D) was subjected to the determination of the full protein mass by LC-MS.

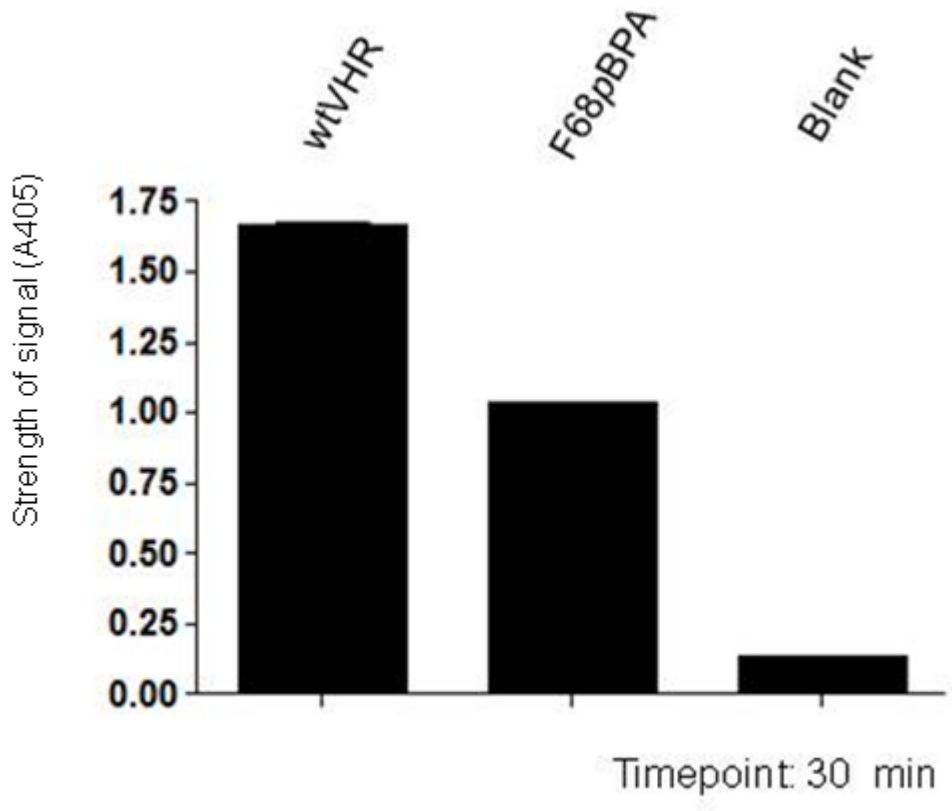


Figure 56: Assessing catalytic activity of the F68*p*BPA monomer isolated by size-exclusion chromatography. Figure 55 contains more details on the sample purification. wtVHR was not exposed to 345 nm light, and the sample of the F68*p*BPA variant was subjected to size-exclusion after UV exposure. The isolated fractions containing the pure monomer (assigned as M) were pulled and assayed against *p*NPP at 10 mM (section 2.3.4). The protein concentration was 500 nM. By denaturing SDS-PAGE analysis, it was validated that equal amounts of the proteins were used in the assays. The product of the reaction was measured for 2 h, acquiring a measurement every 2 min. A discrete time-point (30 min) was selected for a better representation. Three sets of triplicate were analysed and the figure corresponds to a representative result of three experiments. The data was analysed using GraphPad 5 software.

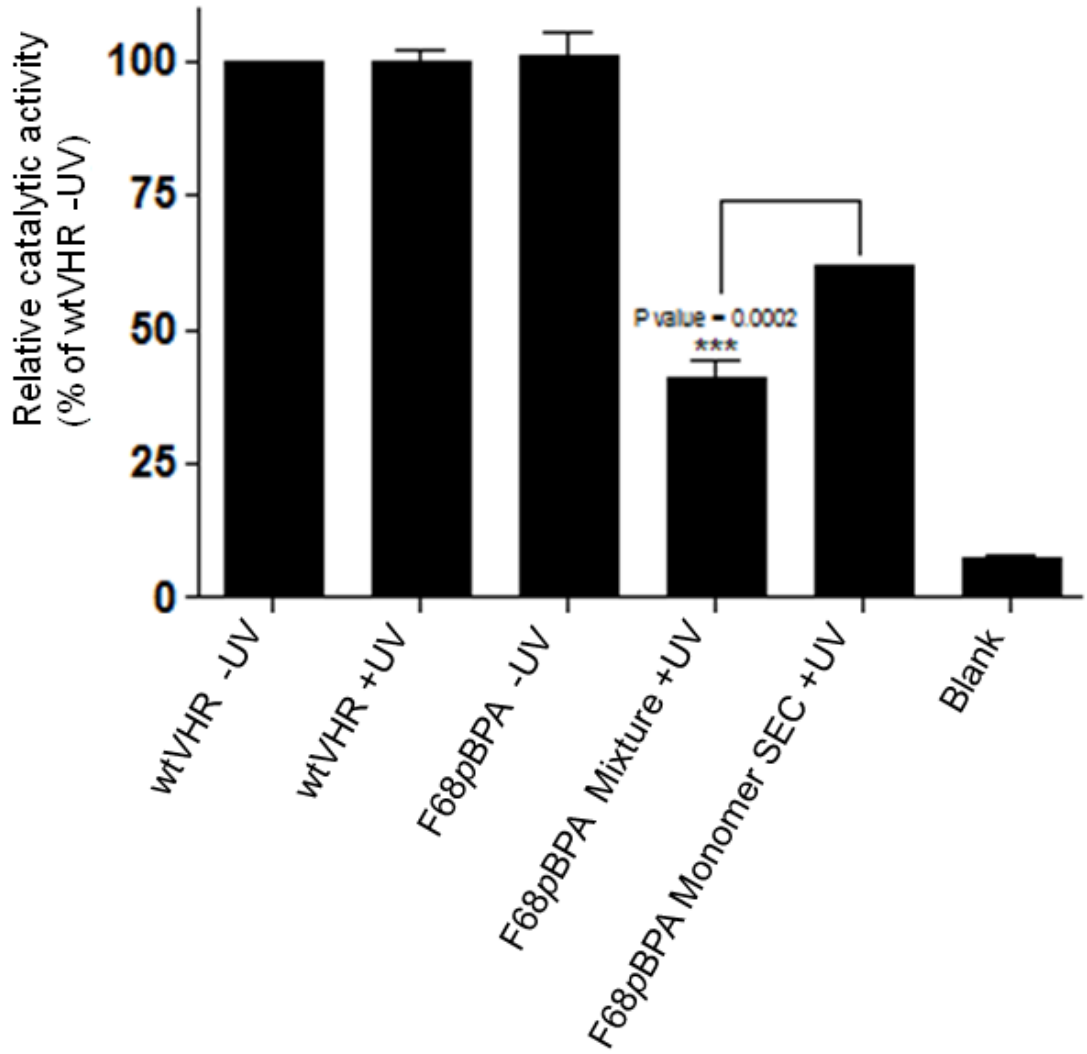


Figure 57: Relative catalytic activity of the irradiated F68*p*BPA with respect to the irradiated wtVHR. The result of the kinetic measurement was originally plotted as absorbance versus time. Significant statistical differences are shown between the F68*p*BPA monomer (assigned as fraction M after size-exclusion, Figure 55) and the mixture sample (containing monomer, dimer and cross-link of monomer to the truncated form of the protein). The data was analysed using GraphPad 5 software which was also used to perform the statystical analysis. A Student's upaired t-test was used to calculate the statystical significance and the asteriscs indicate ***p<0.001.

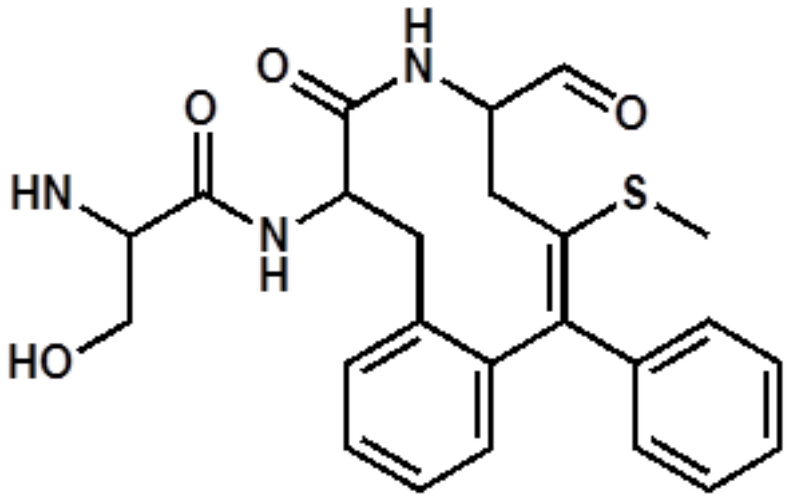
3.2.12 Analysis of VHR dimers after separation by size-exclusion chromatography by LC-MS/MS

Significant effort was put to try to obtain the identity of the cross-linked peptide from two covalently cross-linked monomeric units of VHR. According to the procedure outlined in section 2.3.5b, after irradiating the sample of the F68*p*BPA variant for 30 min at 345 nm, the bands corresponding in size to the molecular weight of the monomer, cross-link of the full length protein to the truncated variant and the band of the cross-linked dimer were excised from the gel and subjected to analysis by LC-MS/MS, according to the procedure summarized in section 2.3.9. The analysis was also tried for the sample of the fraction D (Figure 55). Repeatedly, the peptide containing F68*p*BPA (S*p*BPAMHVNTNANFYK) could not be identified. Digestion with chymotrypsin was also tried, but without success. Identifying the mass corresponding to the cross-linked peptide was not straightforward because of the complexity considering (i) two molecules of *p*BPA, (ii) possible missed trypsin cleavage sites (because the conformation of the cross-linked complex(es) might be such that the cleaving site may not be accessible to trypsin or (iii) possible modifications (dehydration, deamidation, oxidation). The data had to be analysed manually.

Whereas identifying a cross-linked peptide for the dimer was not successful, LC-MS/MS analysis revealed the intramolecular cross-link in the irradiated sample of F68*p*BPA, in the excised gel band corresponding to the monomer. As shown in Figures 22 and 58, the modification corresponding to the cross-link of F68*p*BPA to the neighbouring Met69 was identified. This finding nicely corroborated with the hypothesis of intramolecular cross-linking as source of reduced catalytic efficiency of the monomeric sample isolated by size-exclusion

chromatography. In addition, this sample was subjected to another round of the UV exposure, as described in section 2.3.5, but the yield of the formed dimers was bearly detectable (Figure 59). This indicated that F68*p*BPA was not available for cross-linking any more. Therefore, it is likely that additional intramolecular cross-links exist, in addition to the one we managed to identify.

Finally, the fraction D enriched for the dimer (Figure 55) was subjected to full mass determination (described in section 2.1.15). Figure 60 shows that the mass corresponding to the covalently cross-linked dimer of the F68*p*BPA variant of VHR was obtained. Whereas the identification of the cross-linked amino acids in detail were not possible, this results proves that the covalent VHR dimer exists.



$$MW_{calculated} = 452.164$$

$$MW_{found} = 452.164$$

Figure 58: Identification of the intramolecular cross-linking by LC-MS/MS. The sample was irradiated at 2.33 mg/ml. The recovered sample was diluted to 0.7 mg/ml, mixed with 6 x SDS-PAGE sample buffer, 17.5 µg were resolved by SDS-PAGE, stained with colloidal Coomassie and the bands of interest were excised. The analysis was done as described in section 2.3.10. A peptide corresponding to the intramolecular cross-linking of 68*p*BPA to Met69 was found (452.164) (see Figure 22). The above structure represents a possible fragment matching the found MW.

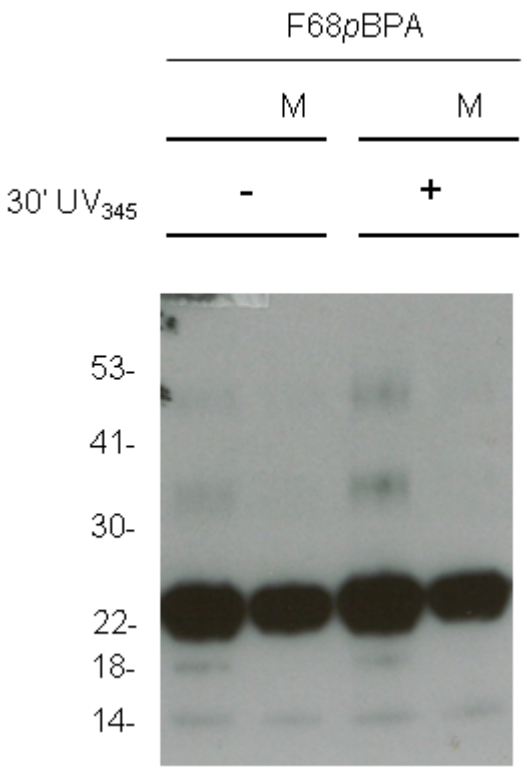


Figure 59: Investigating the cross-linking ability of the monomeric fraction of F68*p*BPA isolated by size-exclusion chromatography. Details to sample purification are described in section 2.3.8. The fraction enriched for the monomeric form of the protein (see Figure 55) was subjected to another round of exposure to 345 nm light at the concentration of 250 µg/ml, as described in section 2.3.5. To check for cross-links, 250 ng of protein was resolved by denaturing 4-12% SDS-PAGE and analysed by Western blotting (anti-VHR). The non-irradiated samples were used as controls. The samples in lanes 1 and 3 correspond to a different batch of F68*p*BPA for which purification by size-exclusion was attempted after UV-induced cross-linking, and were used just as a comparison. Monomers run at about 23 kDa. A faint band at about 35 kDa corresponds to a cross-link between monomer and a truncated form. A faint band at about 50 kDa corresponds to covalently cross-linked two monomers. A faint band at about 14 kDa corresponds to a truncated form (containing 3Flag VHR 1-67) and at about 18 kDa (only visible in lanes 1 and 3) to a likely degradation product due to longer sample storage. The result is representative of two experiments.

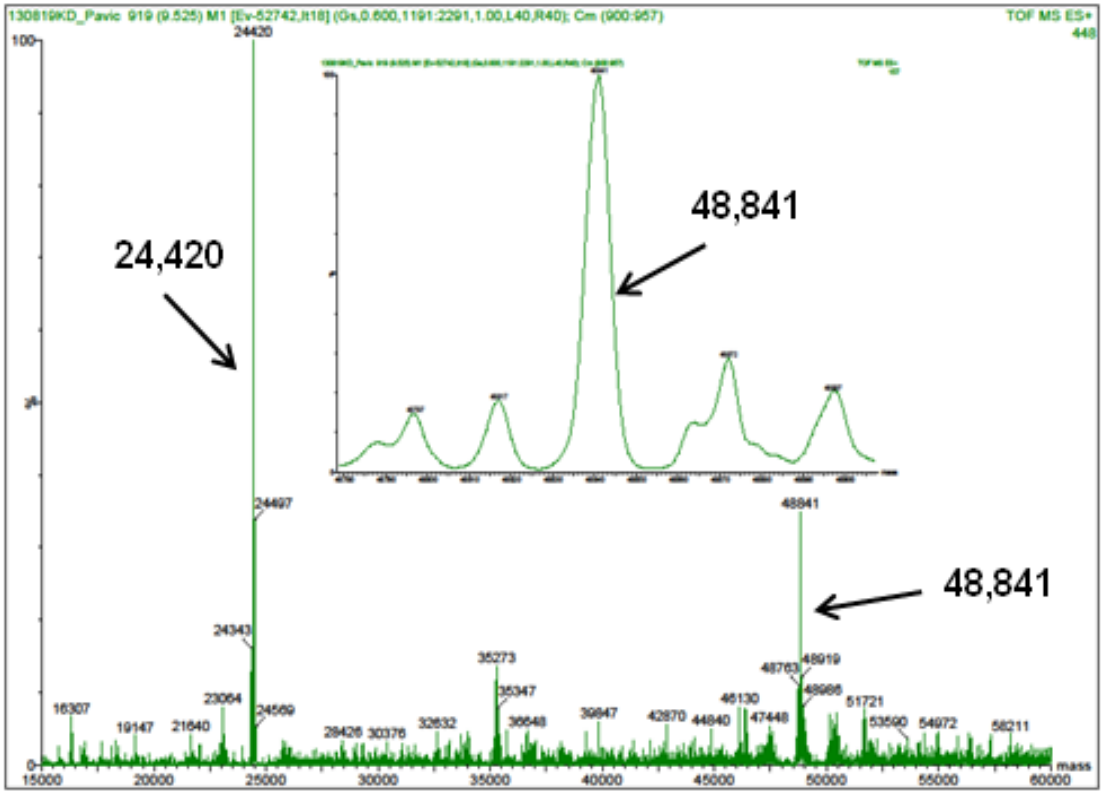


Figure 60: Analysis of the intact sample of F68*p*BPA dimer by LC-MS. The solution sample of fraction D (Figure 55) enriched for the dimer was submitted for determination of full protein mass. The value obtained (48,841 Da) corresponds to the expected value, presented as an adduct with β-mercaptoethanol which was used as a reducing agent in the storing buffer. The mass of 24,420 Da corresponds to the monomer and the mass of 35,273 Da to the crosslinked heterodimer of full length and truncated version, also containing β-mercaptoethanol as adduct. The detailed protocols are provided in sections 2.1.15, 2.3.5 and 2.3.8.

3.3 Applicability of the photo-cross-linkable PRL-3 mutant in investigating protein-substrate and protein-protein interactions

3.3.1 Assessing the cross-linking ability of the PRL-3 D72*p*BPA variant

The D72*p*BPA variant of PRL-3 was expressed on a big scale in *E. coli* and purified as described in section 2.4.4. To probe how exchanging the conserved Asp72 of the WPD loop affected the catalytic performance of the enzyme, activity assay against artificial phosphatase substrate OMFP was performed (Figure 61). The D72*p*BPA variant of PRL-3 exhibited reduced catalytic activity compared to the native protein. This result was expected based on the previous results for the activity assays obtained for the wtPRL-3 and its D72A variant against OMFP in our lab (McParland, Varsano et al., 2011). The D72A variant of PRL-3 was inactive against OMFP, but still showed activity against PIPs.

Investigation of the cross-linking ability of D72*p*BPA was conducted by performing a lipid-binding assay by using the lipid strips, according to the protocol described in section 2.4.6. This assay was selected because the preceeding attempts to cross-link D72*p*BPA to a soluble PI(4,5)P_2, di-C8 PI(4,5)P_2, were not successful in demonstrating the cross-linking by MS analysis. Because the earlier results from our group (McParland, Varsano et al., 2011) demonstrated that the wtPRL-3 and to some extent its substrate-trapping variant D72A, have the ability to recognise and dephosphorylate PI(4,5)P_2 and PI(3,4,5)P_3, it can be hypothesized that the lack of success in identifying the cross-linking by MS is a result of (i) too transient interaction between the protein and the lipid or (ii) because the orientation of *p*BPA is not the most optimal to covalently capture the lipid under the experimental conditions applied.

In addition to the wtPRL-3 and its photo-excitable variant D72*p*BPA, the standard substrate-trap D72A was also used. Furthermore, to control for the non-specific reactivity of the ab, the sample without any protein was also included. PRL-1 was selected because it had already been profiled for its preference for binding to different membrane phospholipids (Sun, Luo et al., 2007). A variant of PRL-3 with the triple mutation in the polybasic region (R157A R159A K161A) (referred to as PRL-3_3A) was assayed to compare its binding preferences with the one reported for PRL-1, for which the residues of the polybasec region were reported to be crucial for binding to PIPs (Sun, Luo et al., 2007). All constructs but PRL-3_3A were expressed and purified as N-terminal $Flag_3$-tagged fusion proteins.

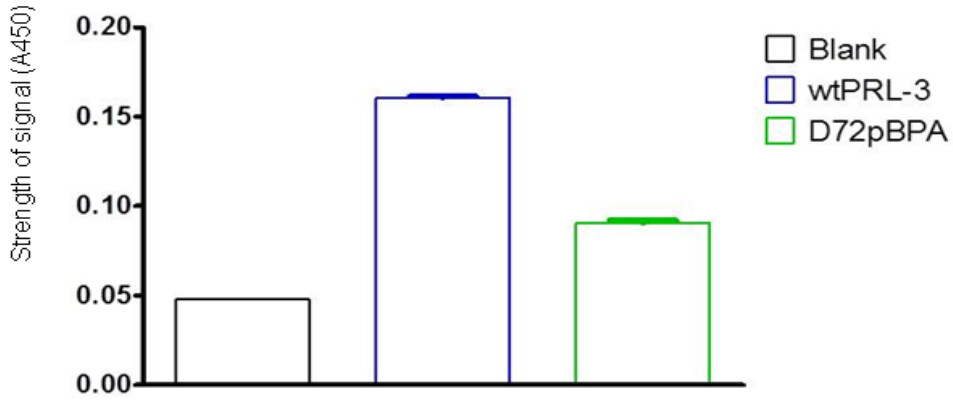


Figure 61: Assessing the activity of the D72*p*BPA variant of PRL-3 with OMFP. *E. coli* over-expression and purification, and the assessment of the catalytic activity against OMFP were conducted as described in detail in sections 2.4.5 and 2.2.4, respectively. Protein was assayed at 6 µM and OMFP at 50 µM final concentration. Reaction volume was 100 µl. wtPRL-3 was used for comparison. The “blank” sample contained all the components but not the enzyme. The graph reflects the results of two experiments. The data was analysed by using GraphPad Prism 5.0 software and represented as an end point measurement, at the final point of the assay with SEM.

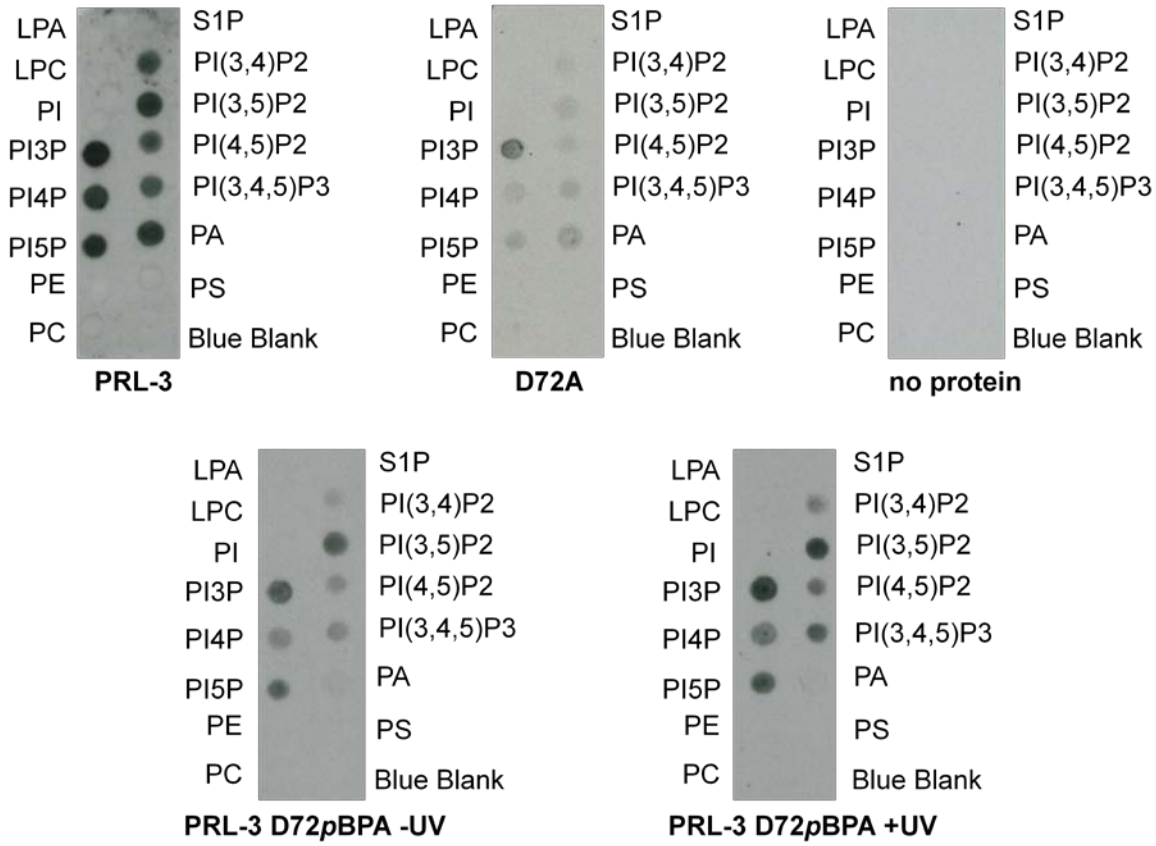


Figure 62: PRL-3 in lipid strips assay. Detailed description of the protocol is provided in section 2.4.6. All proteins were expressed as N-terminal $Flag_3$-tagged fusion proteins and were incubated with the strips at 0.5 µg/mL for 30 min in TBS-T (0.1%) buffer containing 3% (w/v) fatty acid-free BSA. The control sample contained no protein and was treated in the same way described above. D72*p*BPA sample was incubated with the strip for 10 min prior to exposure to 365 nm, which was conducted for the total time of 5 min (two intervals of 2.5 min, with 2.5 min break in-between). The non-irradiated control sample of D72*p*BPA was kept light protected throughout duration of the assay. Protein-lipid interactions were probed using anti-Flag ab and visualized with ECL Western Blotting Detection Reagents (developing time 1 min).

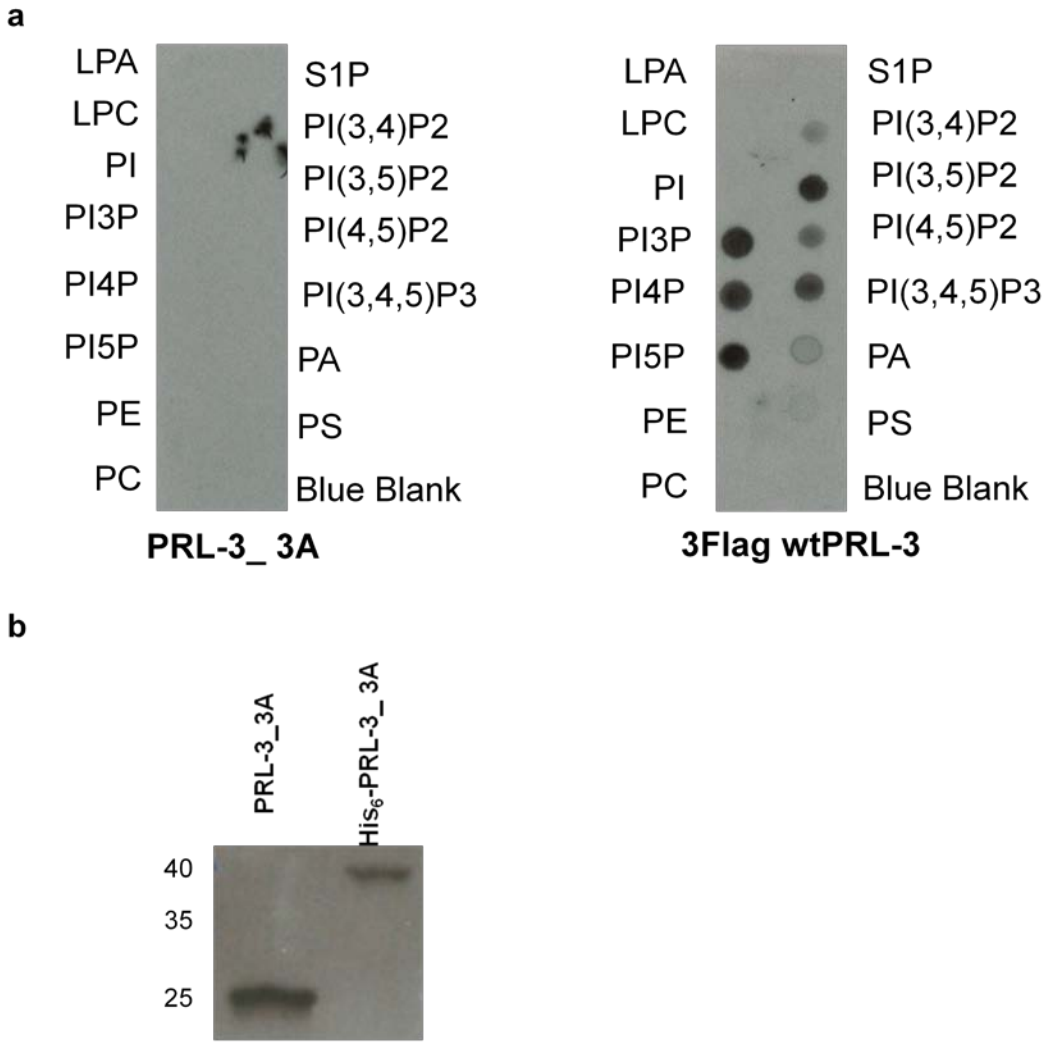


Figure 63: Assessing the effect of the triple mutation R157A R159A K161A on the lipid binding profile of PRL-3. (**a**) Detailed description of the protocol is provided in section 2.4.6. All proteins were incubated with the strips at 0.5 µg/mL for 30 min in TBS-T (0.1%) buffer containing 3% (w/v) fatty acid-free BSA. Protein-lipid interactions were probed using anti-PRL-3 ab and visualized with Western Lightning Plus-ECL Enhanced Chemiluminescence Substrate (developing time 5 min). (**b**) Assessing if the triple mutation had effect on the recognition by the PRL-3 ab by resolving 500 ng of the proteins, with or without the N-terminal His_6-tag, by 4-12% SDS-PAGE and probing with anti-PRL-3 ab.

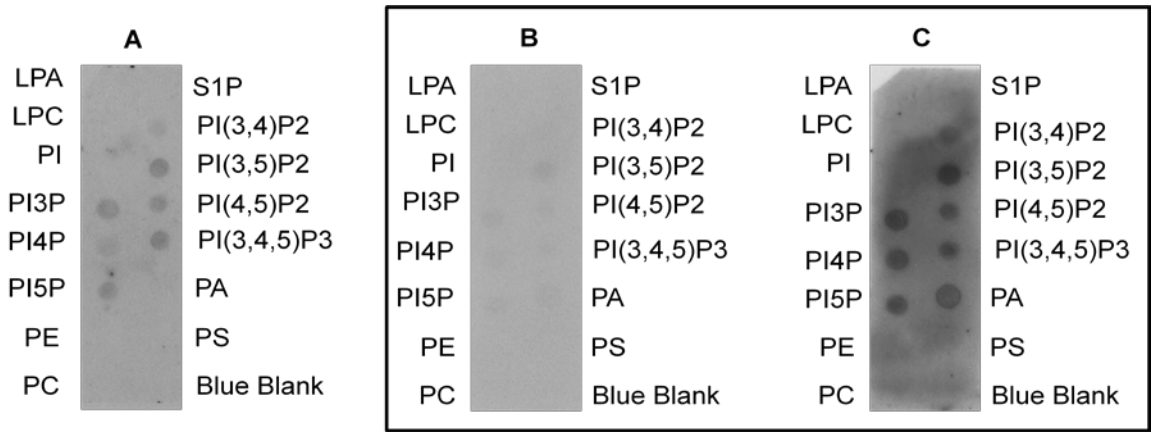


Figure 64: PRL-1 in lipid strips assay. A detailed description of the protocol is provided in section 2.4.6. PRL-1 was expressed as N-terminal $Flag_3$-tagged fusion proteins and was assayed under two different conditions of protein concentration and total incubation time. Strip A was incubated with the protein at 0.5 µg/mL for 30 min in TBS-T (0.1%) buffer containing 3% (w/v) fatty acid-free BSA and strips B and C were incubated with the protein at 5 µg/mL for 60 min in the same buffer as above. Protein-lipid interactions were probed using anti-Flag ab and visualized with ECL Western Blotting Detection Reagents (PIP strip B, developing time 1 min) and Western Lightning Plus-ECL Enhanced Chemiluminescence Substrate (PIP strips A and C; developing times 15 min and 5 min, respectively). The results indicates that, compared to PRL-3, PRL-1 shows much less lipid-binding affinity, as tenn times more protein and double the incubation time were needed to detect weak signal (A).

The preliminary results obtained in the lipid binding assays are summarised in Figures 62-64. From 62, it can be seen that all PRL-3 variants screened seemed to prefer PIPs phosphorylated at the position 3 and/or 5 of the inositol ring (PI3P, PI5P and PI(3,5)P_2), whereas seemingly showing the weakest preference for binding to PI(3,4)P_2. Furthermore, wtPRL-3 (Figures 62 and 63) and PRL-1 (Figure 64) also showed non-specific binding to phosphatidic acid (PA). Binding of the D72A variant to PA was much weaker compared to the wtPRL-3, but the same effect was also observed for the other lipids on the strip (Figure 62). The D72*p*BPA variant showed bearly detectable binding to PA (Figure 62). Possible explanation for these differences in the non-specific binding to PA could be that the *p*BPA moiety imposes sterical effect on the binding to different lipids, but it has no negative effect on the recognition and binding to PIPs, which were demonstrated to be pysiologically relevant for PRL-3 (McParland, Varsano et al., 2011). Furthermore, two different ab were used in probing for binding of different PRL-3 variants to lipids in lipid strips, anti-Flag (Figures 62 and 64) and anti-PRL-3 (Figure 63). As shown in Figure 63, binding of wtPRL-3 to PA looks much weaker with anti-PRL-3 ab. Because the control without the protein and analysed with anti-Flag ab showed no signal (Figure 62), it can be hypothesized that non-specific binding of PRL-3 to PA might occur via different mechanism than its binding to PIPs and that these differences in the binding mechanism might have the effect on the recognition of the protein in the suubsequent analyses by different ab.

It was reported previously that PRL-3 dephosphorylated PI(4,5)P_2, and to a smaller extent PI(3,4,5)P_3, *in vitro* (McParland, Varsano et al., 2011). It was also reported that PRL-1 and apparently PRL-3 showed the same lipid binding preferences in lipid strips assay (Sun, Luo et al., 2007). It would be interesting

to probe if there is a correlation between lipid binding and catalytic activity for PRL-3 like it was shown for PTEN and myotubularin, both of which are PIP phosphatases (Campbell et al., 2003). Furthermore, the basic residues in PRL-3 seem to be important for binding of PRL-3 to PIPs, because PRL-3_3A exhibited no detectable binding (Figure 63), even after long exposure times. This result correlates with the reported binding profile for PRL-1 with various mutations in the polybasic region (Sun, Luo et al., 2007). However, this study reported that PRL-1 and, apparently, PRL-3 showed the strongest preference for binding to monophosphorylated PIPs, and the weakest one for binding to $PI(3,4,5)P_3$, which is not really correlating with the results here. As indicated in Figure 62, the D72*p*BPA variant of PRL-3 after exposure to 365 nm light for the total time of 5 min showed a stronger signal than its non-irradiated control. The mutant showed largely the same binding profile like the native protein. These results indicated that the introduction of the photo-cross-linkable amino acid did not negatively affect the ability of PRL-3 to recognise and bind to PIPs.

3.3.2 Assessing the applicability of PRL-3 D72*p*BPA with HEK293 cell lysate

To probe if the photo-cross-linkable variant of PRL-3 could be used to capture (novel) interacting partners of this phosphatase, photo-induced cross-linking following incubation with the lysate from HEK293 cells was conducted. The detailed description of the experimental conditions, including the data analysis, is provided in section 2.4.7. In addition to the photo-cross-linkable variant of PRL-3 (used with or without exposure to 345 nm light for 30 min), the native protein and its D72A variant were utilized as controls. The wtPRL-3 was used without UV exposure, and its D72A variant with and without irradiation. The selected protein controls were expected to show if there would be difference in

the portfolio of the identified interacting partners following the pull-down with His-Select Nickel Affinity Gel for various PRL-3 variants and if there would be a UV-mediated derogative effect on the stability of the generated cross-links. Of note, various experimental conditions for assessing the applicability of the photo-cross-linkable variant of PRL-3 on a broader scale were tried (for example, the amount of protein used as bait, pre-clearing of the lysate, pre-incubation prior to UV-exposure, pull-down of the associated or cross-linked complexes, buffer and washing conditions). In the irradiated sample, additional bands corresponding to the complexes generated by covalent cross-linking of *p*BPA moiety to the proximal C-H bonds were detected (Figure 65), whereas the intensity of some bands in the irradiated sample was intensified with respect to the control. However, when 60 µl of the elute for each sample was resolved by 8% SDS-PAGE and left to stain with colloidal Coomassie stain, the differences from the Western blot did not reflect in the differences of the stained gel. Moreover, for all five samples the gel bands of three ranges of molecular weight were excised and subjected to MS analysis with the aim of identifying (novel) interacting partners and/or potential substrates of PRL-3. However, the retrieved list of identified proteins bore no striking differences amongst all the samples analysed.

Considering this, the fact that colloidal Coomassie stain and the subsequent MS analysis failed to demonstrate noticeable differences in the profile of identified interacting proteins for the diferent PRL-3 variants screened could be an indication that (i) the introduced amino acid substitution had no effect on the ability of PRL-3 variants to recognise the intrinsic interacting partner(s) or that (ii) the yield of the cross-linked complexes generated after UV exposure was not high enough to render distinctions visible in the colloidal Coomassie

staining of SDS-PAGE gel. To enrich the UV-exposed sample for the covalently cross-linked complexes and, therefore, increase the odds of identifying novel interesting hits in the following MS analysis, the sample was subjected to pull-down with His-Select Nickel Affinity Gel. However, this attempt was not completely successful. As evident from the Western blot analysis using anti-PRL-3 ab, protein stock of the D72*p*BPA variant used as bait also contained a fraction without the His_6-tag at the N-terminal end. This cleaved fraction also bound to the affinity beads and co-purified with the His_6-tagged variant.

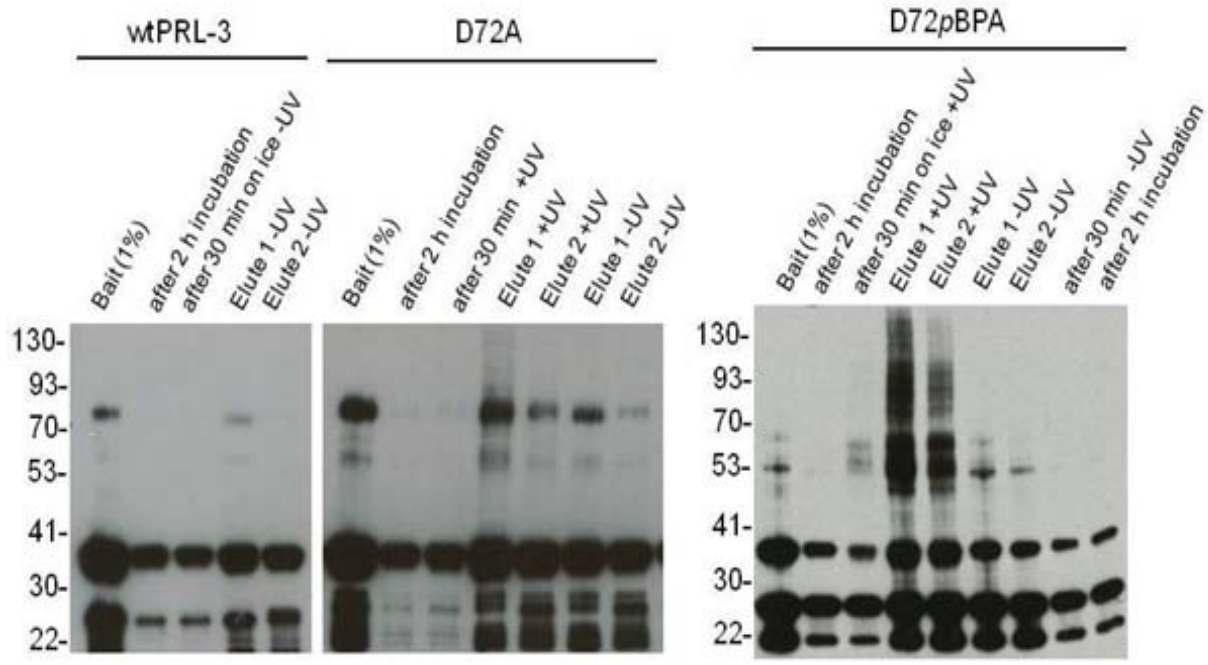


Figure 65: Incubation of the PRL-3 variants with HEK293 lysate, with or without exposure to 345 nm light. The protocol is described in section 2.4.7. PRL-3 variants were used as bait at 200 µg, and were incubated with 1 ml of lysate from HEK293 cells (2 x 14 cm dishes were used for each sample) for 2 h at 4°C on a roller. The samples were irradiated for 30 min at 345 nm, and recovered samples were subjected to pull-down with Ni-affinity gel (200 µl slurry/sample) overnight at 4°C. After washing, the bound material was eluted by applying 100 µl of 2 x SDS-PAGE sample buffer and by incubating first for 10 min at 65°C and 500 rpm. The first elute was recovered and the fresh portion of sample buffer was added and incubated for 5 min at 95°C and 500 rpm. 2.5 µl of each elute was resolved by SDS-PAGE, and 5 µl of all other samples withdrawn. Western blot was done by using anti-PRL-3 ab. By loading the sample of pure bait for each protein (1%), it can be excluded what impurities are derived from the bait-sample. In the UV-irradiated sample of D72*p*BPA, additional bands showing up in the higher MW range (above 41 kDa) correspond to the covalently cross-linked complexes (they are also missing from the non-irradiated sample of D72*p*BPA). The cut-out gel bands were subjected to analysis by MS, but the identity of proteins corresponding to the individual bands visible by Western blot at higher MW at present cannot be assigned.

Chapter 4: Conclusion and outlook

4.1 Critical comments of the preliminary experiments (Chapter 3.1 in the Results and discussion section)

Overall, this section can be viewed as the weekest section of the thesis. This can in part be attributed to the learning process, as most of the experiments were conducted with next to nothing hands-on experience in molecular biology techniques in general.

Generating mutants of PRL-3 with selected residues (Ile141, Asn142 and Gln145) exchanged first for tyrosine and assessing their suitability for the folowing substitution with an unnatural amino acid based on the (i) CD spectra of the mutants with respect to the native protein and (ii) their catalytic performance might be better substituted by designing an analogous set of experiments involving *p*ACF. *p*ACF might be more similar to *p*BPA in terms of hidrophobicity. It may not be the most optimal choice due to the sterical bulk of *p*BPA, but it could still be regarded as more suitable than substitutions of the selected residues with any of the canonical amino acids (for example, introducing tryptophane might impose more sterical constrain due to its structure comprised of two fused rings, whereas in *p*BPA the two phenyl rings are separated by a carbonyl moiety). Furthermore, for the assessment of the catalytic activity, a more optimised set of assay conditions would need to be applied. The strength of the signal, measured as absorbance at a particular wavelength (as determined by the phosphatase substrate used in the assay) was very weak even for the wt protein, making any strong conclusions about the behaviour of the mutants, with respect to the native protein, more difficult.

Another weak spot in this section is the data analysis. Like already mentioned, no miningful conclusion can be drawn by conducting an assay once. Thus, it would be advisable to use a more suitable activity assay and to generate several replicates for each of the mutants investigated.

4.2 Implications for VHR biology: findings enabled by application of the amber suppression methodology

The results presented in the Chapter 3.2 provided a clear demonstration of dimerisation as a potential novel regulatory feature of the VHR phosphatase. This finding was enabled by selective incorporation of photo-cross-linkable *p*BPA. Unlike chemical cross-linkers, use of photo-excitable *p*BPA also enabled investigation of how dimerisation affects the catalytic activity of VHR.

Oxidative stress could have a stabilizing effect on the intermolecular self-association of VHR, which in turn might impose steric hindrance to the active site cleft, thus providing a mechanism to protect catalytically essential cysteine of the P-loop from irreversible oxidation of sulfenic acid to sulfinic or sulfonic acid. There is a growing number of reports elucidating the mechanisms through which H_2O_2, and ROS in general, modulate cellular signaling (Adler et al., 1999; den Hertog et al., 2008; Gough and Cotter, 2011). Moreover, H_2O_2 has been demonstrated to not only induce a mitogen-activated signaling cascade, but to also be required, or participate, in normal receptor-mediated signal transduction (Denu and Tanner, 1998; Todd et al., 1999). Self-association of VHR to form dimers would also result in a high local concentration of the enzymatically active form after release of the oxidative stress, which could act rapidly to elicit a specific cellular response after the dimeric complex is disassembled into

monomers. Further investigations involving, for example, structural analysis of the native VHR protein with catalytic cysteine oxidised to sulfenic acid or of the covalent complex of the homodimer, are needed to evaluate: (i) if oxidation of the active site cysteine to sulfenic acid causes local structural changes or affects electrostatic interactions which may have positive implications for the dimer assembly; (ii) how self-association of VHR influences the overall structure of the complex; (iii) how the cellular redox-state and extracellular stimuli-induced production of ROS ultimately inflict a network of interactions due to spatio-temporally and mechanistically distinct regulatory principles of PTPases. The accumulating knowledge of the mechanisms through which ROS, such as H_2O_2, finely modulate signaling cascades demonstrates an immense complexity and transform the view of H_2O_2 as a malignant molecule (Adler et al., 1999; den Hertog et al., 2008; Denu and Tanner, 1998; Gough and Cotter, 2011; Meng et al., 2006). H_2O_2 is acknowledged as a secondary messenger that mediates signal transducing cascades triggered in response to a broad range of extracellular stimuli. Moreover, H_2O_2 production needs to be under strict spatio-temporal control in order to maintain homeostatic levels of ROS and redox balance. The unique microenvironment of the conserved active site cysteine residue decreases its p*Ka* for more than 3 pH units than the pH value of a normal cysteine, so that it exists in the form of a thiolate ion under physiological conditions (Denu and Tanner, 1998). This renders it more sensitive to H_2O_2 than the protonated thiol would be, thus making PTPs appealing targets for redox-mediated regulation of cell signaling. Furthermore, formation of dimers, and higher order oligomers, amongst PTPs in response to H_2O_2-imposed oxidative stress has been reported previously (Blanchetot et al., 2002; Deb et al., 2011; Liu, Wang et al., 2006; Nardozza et al., 2012; Östman

et al., 2011). For example, it was demonstrated that the endogenous SHP-2 could dimerize, with concomitant decrease in the phosphatase activity, and that the monomer-dimer equilibrium was modulated by the redox status of the cell and by growth factors, thus proposing a novel model for redox-based structure-activity regulation (Nardozza et al., 2012). Next, it was proposed that H_2O_2 induced inactivation of RPTPα by oxidising the catalytically active cysteine residue in D1 domain, thus allowing propagation of downstream signaling, and by inducing conformational changes in D2 domain, with stabilizing effect on the dimer assembly (Blanchetot et al., 2002). Recently, the membrane-associated isoform of the neuron-specific Striatal Enriched Phosphatase (STEP), $STEP_{61}$, was shown to dimerize via intermolecular disulfide bond formation, with the considerable increase in the order of the oligomers detected after H_2O_2-induced oxidative stress and the accompanying reduction of the catalytic activity (Deb et al., 2011). VHR is classified as atypical DUSP, a subgroup of DUSPs with a growing interest of research community in deciphering the principle of their role and regulation under physiological and pathological conditions (Patterson et al., 2009). However, the current knowledge is still overwhelming with conflicting data, especially regarding discrepancy on substrate specificity of DUSPs towards MAPKs *in vitro* and *in vivo*. Additional layer of complexity is underlined by a conclusion that atypical DUSPs may exert stimuli-dependent-, cell-type-, or context-specific effects on MAPKs, have a broader repertoire of substrates or that they can act as scaffolds or require additional binding partner(s) within a particular signaling cascade. Rapidly increasing data on structural characterisation of atypical DUSPs suggests a potential for vastly different means of regulation of their biological functions. Moreover, according to the “fingerprint”-based computational analysis, aiming to classify atypical DUSPs

with respect to their amino acid sequence, two subfamilies of DUSPs emerged (Romá-Mateo et al., 2007). Interestingly, VHR was classified as a member of atypical DUSP I subfamily, alongside, amongst others, DUSP13a, DUSP13b, DUSP26 or DUSP27, for all of which dimer-forming propensity has been demonstrated (Kim et al., 2007; Lokareddy et al., 2013; Lountos et al., 2011). VHR was identified as the closest structural homologue of DUSP27 and testis and skeletal muscle specific DUSP (TMDP), also known as DUSP13b, both of which (i) do not have an N-terminal MKB domain, (ii) cannot be aligned with VHR in the region of β3-α4 loop and (iii) were demonstrated to form dimers. In TMDP, β3-α4 is shorter, rendering surface around the active site more planar when compared with VHR (Kim et al., 2007; Lountos et al., 2011; Patterson et al., 2009). Together with the α1-β1 loop, known as “substrate recognition region” in VHR, being displaced further from the active site, these differences contribute to the broader active site pocket in TMDP and to its equal preference for pTyr and pSer/pThr substrates. Intriguingly, the dimer-driving interactions were not sufficient to maintain the association of TMDP in solution. There is a number of reports on the attempts to identify features common to protein-protein interfaces or trying to address properties of homo- or heterodimeric protein assemblies by computational analysis (Dey et al., 2010; Ozbabacan et al., 2011). In summary, one could substantiate that the subunit interfaces are more hydrophobic than the surface accessible areas, those of weak dimers being slightly more polar, but still with the similar amino acid composition and hydrogen bonding capacity. Why weak dimers are weak can be explained probably by buried surface area alone, where homodimers bury less than 1000 Å, resulting in fewer stabilizing interactions. In TMDP, the dimers have interface of 841 Å, making inability to preserve self-association in solution plausible (Kim

et al., 2007). Another contributing factor is the quality of the atomic packing (Dey et al., 2010). Namely, loose contacts will result in a more flexible geometry of the dimeric association, more hydrated protein surface and less than optimal interactions comparing to the close-packed interface.

In summary, based on the current study, dimerisation could represent a novel regulatory mechanism of VHR. In a non-stimulated state, VHR might be present as a monomer, or loosely packed dimer, with the overall catalytic activity dependent on the mutual orientation of the self-associating units. It is known that the conserved cysteine residue of the P-loop of VHR is sensitive to H_2O_2-induced oxidative inhibition (Denu and Tanner, 1998). Here, it was clearly demonstrated that oxidative stress has a stabilizing effect on the interacting two monomers of VHR, plausibly, but not necessarily, via disulfide bond assembly. We were able to demonstrate self-association of VHR in cells by co-IP only after H_2O_2-induced stress, but also after treatment with formaldehyde and in the absence of oxidative stress.

When analysing the effect of the reducing conditions on the assembly of the homomeric VHR dimers, the profile of the F68*p*BPA mutant of VHR, when compared to the native protein and to the F68C mutant, was quite surprising. Simultaneously, this analysis helped to postulate one possible suggestion how dimeric association of VHR might come to occur, namely through a disulfide bond-based association that would be enabled by hydrophobic interactions of the rim spanning residues of the “variable insert” segment. In order to validate this hypothesis further, candidate cysteine residues in VHR might be exchanged by SDM to a residue that lacks dislfide-bonding capacity. Assuming that those substitutions would not have any negative effect on the stability of the mutants or on their catalytic behaviour, a set of *in vitro* experiments might

be conducted to examine the profile those mutants would show when analysed under reducing or native SDS-PAGE conditions.

In the future, it will be interesting to further dissect physiological implications of diverse regulatory principles of VHR as well as to see if this mode of regulation could be extrapolated to other atypical DUSPs, with special emphasis on probing non-conserved structural features. For example, amongst atypical DUSPs that share significant sequence similarity with VHR are DUSP27 and DUSP13b (Friedberg et al., 2007; Kim et al., 2007; Lountos et al., 2011). Moreover, VHR was also identified as a close structural homologue of MKP2 (Jeong, Jung et al., 2009). For the former two phosphatases, dimer-forming ability was demonstrated, whereas for MKP-2, it was shown that it could form a 24-mer assembly. It is interesting to note that the segment corresponding to the “variable insert” region in VHR regularly cannot be perfectly aligned with the corresponding segment in those three atypical DUSPs. This observation at present might further (i) strengthen our observation of VHR dimerisation as a potential novel regulatory mechanism and (ii) support the hypothesis that combining detailed structural investigation, with special focus on the uinque residues or regions in the protein, with expanded genetic code methodology might accelerate our understanding of the aforementioned unique structural characteristics.

Furthermore, investigating dimerisation at the level of endogenous protein represents another challenge, which is currently being undertaken in our lab.

Finally, identifying if, and how, distinct stimuli, or PTMs, could affect dimer topology, stability or the outcome of downstream signaling cascades would further clarify the patho-physiological role of VHR. In addition, any knowledge

sprouting from such investigations can be useful for potential design of compounds that could promote or inhibit dimeric association of VHR.

4.3 Applicability of the PRL-3 mutants containing photo-cross-linkable amino acid in investigating protein-substrate interactions

At present, it is not established if D72*p*BPA with and without the N-terminal His_6-tag have the same cross-linking efficiencies. Considering that the tag is positioned far from the region bearing cross-linkable amino acid, interferences with the substrate recognition and/or binding to the active site crevice are however not likely. Importantly, control pull-down without protein might help filter out the non-specific interactions.

The introduction of amino acid substitutions with respect to the native protein, with the aim of expanding the portfolio of known interacting partners of a protein, must not distort naturally existing interactions. In order to have a protein variant that would serve as a “substrate trap”, the longevity of the substrate-protein association needs to be prolonged. Moreover, protein variants bearing mutations that retain the ability to bind substrates, but have hindered catalytic processing of a transient cysteinylphosphate intermediate, were broadly used in studying phosphatase-substrate interactions (Blanchetot et al., 2005; Zhang, 2003).

The C104S variant of PRL-3 was shown to be misfolded (McParland, Varsano et al., 2011), whereas the same study demonstrated still some *in vitro* activity od the D72A variant against PIPs. This residual acctivity observed for both D72 variants of PRL-3 (A and *p*BPA) could result in protein-substrate or protein-protein interactions being too transient. To potentially overcome this, additional

combinations of mutations might be explored, combining for example mutations of the conserved catalytic residues with a photo-cross-linkable amino acid introduced proximal to the entry into the active site pocket. Moreover, if the focus of investigation are potential interacting partners of PRL-3, it might be

beneficial to introduce photo-cross-linkable amino acid at positions in the protein that are not directly involved in the catalytic mechanism.

Moreover, the approach might also be applicable to investigate possible PTMs and/or to capture PKs that are responsible for phosphorylation of specific residues in PRL-3. It was noted previously that, when the sequences of all three PRLs were compared, several potential consensus sites for modification by phosphorylation were suggested (Zeng et al., 1998). Further investigation of (i) potential modification by phosphorylation in PRL-3, and especially in PRL-1 and PRL-2 and of (ii) possible functional implication such modification might have is lacking. Moreover, it would be interesting to extrapolate the approach to PRL-1 and PRL-2. The residues which could be substituted for a photo-cross-linkable amino acid might be divided into two categories. One might comprise residues that are known to be involved in the catalytic mechanism, in a manner similar to the one used for PRL-3 in the course of this project. It is known that PRLs share high degree of amino acid sequence identity and that the most prominent differences in the sequences can be mapped onto the face opposite to the active site pocket (Kozlov et al., 2004). The other set of residues that could be replaced with an unnatural amino acid could comprise the aforementioned unique residues amongst the PRLs. It would be interesting to investigate if this approach would render differences (and if so, to what extent) amongst the portfolio of the photo-induced cross-linked proteins.

In the course of this PhD, these issues could not be further addressed since the focus was directed towards detailed investigations of VHR.

4.4 General applicability of the amber supression methodology to develop (more) efficient substrate trapping mutants of PTPs

Selective incorporation of the photo-cross-linkable amino acid *p*BPA into VHR enabled investigation of dimerisation as a potential regulatory mechanism of this phosphatase. It is important to note that without *p*BPA-containing mutants of VHR this observation would not be rendered likely, because all the results generated for VHR in the course of this project point towards transient homodimeric assembly.

Importantly, the effect of dimerisation on the catalytic activity of VHR could also be investigated, which would not be possible by using chemical cross-linkers due to their non-specific and extensive cross-linking, like it was demonstrated in the cross-linking experiments with glutaraldehyde in the course of this project. Moreover, another obstacle in using chemical cross-linking reagents is need for extensive optimisation. This includes, for example, fine-tuning protein concentration, type and concentration of the cross-linker, duration of the cross-linking reaction, temperature and buffer conditions or the conditions for the subsequent analysis by, for example, Western blotting.

Selective incorporation of *p*BPA could also be a valuable tool in investigating (substrate) interactions of PRL-3. The D72*p*BPA variant of PRL-3 was able to recognise and bind to PIPs, with the increase in the signal clearly detected after exposure of the sample to 365 nm light for only 5 min. Moreover, the D72*p*BPA variants showed reduced level of the nonspecific binding to PA. This could be

due to loss of determinants necessary to recognise PA, but also due to sterical reasons as a result of selective incorporation of *p*BPA. Considering the demonstrated success of the D72*p*BPA variant in forming UV-induced covalent complexes with either already reported substrates (PIPs) or upon incubation with the cell lysate, it might also be successful in structural studies as a result of the stabilizing effect of the protein-substrate association upon UV exposure. The D72*p*BPA variant could be a promising tool in expanding the known interacting space of PRL-3. Subsequently, the signaling networks affected by PRL-3 might be constructed, especially considering the advantages of photo-cross-linkable benzophenone moiety.

Through the successful application of the expanded genetic code toolkit, the advantages of exploiting the technology of selective incorporation of photo-activatable non-canonical amino acid *p*BPA were clearly demonstrated. Moreover, uniting the expanded genetic code toolkit with the more conventional approaches of investigating transient protein-protein interactions by means of diverse chemical cross-linking reagents was successful.

Based on the study presented here, it is reasonable to assume that the selective incorporation of photo-cross-linkable amino acids would be a useful tool in investigating interactions of PTPs in general. These investigations would be greatly supported by the availabilities of protein structures to speed up the selection of the amino acid residues that would be most suitable to be exchanged with a photo-cross-linkable amino acid. In addition, other photo-cross-linkable amino acids, such as the recently reported aliphatic diazirine (Chou et al., 2011), which might circumvent the disadvantage of sterical bulk of *p*BPA and which have the photo-excitable moiety at the end of a longer chain, might also be explored.

References

Achiwa, H. & Lazo, J.S. PRL-1 tyrosine phosphatase regulates c-Src levels, adherence, and invasion in human lung cancer cells. *Cancer Res.* **67**, 643-650 (2007).

Adler, V., Yin, Z., Tew, K.D. & Ronai, Z. Role of redox potential and reactive oxygen species in stress signaling. *Oncogene* **18**, 6104-6111 (1999).

Ai, H., Shen, W., Sagi, A., Chen, P. R. & Schultz, P. G. Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. *ChemBioChem* **12**, 1854-1857 (2011).

Al-Aidaroos, A.Q.O. & Zeng, Q. PRL-3 phosphatase and cancer metastasis. *J. Cell. Biochem.* **111**, 1087-1098 (2010).

Alam, H. *et al.* Loss of keratin 8 phosphorylation leads to increased tumor progression and correlates with clinico-pathological parameters of OSCC patients. *PLoS One* **6**, e27767 (2011).

Alexander, M., Gerauer, M. *et al.* Mapping the isoprenoid binding pocket of PDEδ by a semisynthetic, photoactivatable N-Ras lipoprotein. *ChemBioChem* **10**, 98-108 (2009).

Alonso, A. *et al.* Tyrosine phosphorylation of VHR phosphatase by ZAP-70. *Nat. Immunol.* **4**, 44-48 (2003).

Alonso, A. *et al.* Protein tyrosine phosphatases in the human genome. *Cell* **117**, 699-711 (2004).

Ambrogelly, A., Palioura, S. & Söll, D. Natural expansion of the genetic code. *Nat. Chem. Biol.* **3**, 29-35 (2007).

Amrani, N., Sachs, M.S. & Jacobson, A. Early nonsense: mRNA decay solves a translational problem. *Nat. Rev. Mol. Cell Biol.* **7**, 415-425 (2006).

Andersen, J.N. *et al.* Structural and evolutionary relationships among protein tyrosine phosphatase domains. *Mol. Cell. Biol.* **21**, 7117-7136 (2001).

Anderson, J.C., Wu, N. *et al.* An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. USA* **101**, 7566-7571 (2004).

Andl, C. D. The Misregulation of cell adhesion components during tumorigenesis: overview and commentary. *J. Oncol.* **2010**, Article ID 174715 (2010).

Arantes, G.M. Free-energy profiles for catalysis by dual-specificity phosphatases. *Biochem. J.* **399**, 343-350 (2006).

Arnoldussen, Y.J. *et al.* The mitogen-activated protein kinase phosphatase *Vaccinia* H1-related protein inhibits apoptosis in prostate cancer cells and Is overexpressed in prostate cancer. *Cancer Res.* **68**, 9255-9264 (2008).

Back, J.W., de Jong, L., Muijsers, A.O. & de Koster, C.G. Chemical cross-linking and mass spectrometry for protein structural modeling. *J. Mol. Biol.* **331**, 303-313 (2003).

Bai, Y., Luo, Y. et al. PRL-1 protein promotes ERK1/2 and RhoA protein activation through a non-canonical interaction with the Src homology 3 domain of p115 Rho GTPase-activating protein. *J. Biol. Chem.* **286**, 42316-42324 (2011).

Bardeli, A. et al. PRL-3 expression in metastatic cancers. *Clin. Cancer Res.* **9**, 5607-5615 (2003).

Barford, D. Protein phosphatases. *Curr. Opin. Struct. Biol.* **5**, 728-734 (1995).

Barford, D., Das, A.K. & Egloff, M.-P. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 133-164 (1998).

Barr, A.J. *et al.* Large-scale structural analysis of the classical human protein tyrosine phosphatome. *Cell* **136**, 352-363 (2009).

Basak, S. *et al.* The metastasis-associated gene *Prl-3* is a p53 target involved in cell-cycle regulation. *Mol. Cell* **30**, 303-314 (2008).

Bass-Zubek, A. E., Godsel, L. M., Delmar, M. & Green, K. J. Plakophilins: multifunctional scaffolds for adhesion and signaling. *Curr. Opin. Cell Biol.* **21**, 708-716 (2009).

Bayón, Y. & Alonso, A. Atypical DUSPs: 19 phosphatases in search of a role. *Emerging Signaling Pathways in Tumor Biology.* **661**, 185-208 (2010).

Beier, H. & Grimm, M. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res.* **29**, 4767-4782 (2001).

Bessette, D.C., Wong, P.C.W. & Pallen, C.J. PRL-3: A metastasis-associated phosphatase in search of a function. *Cells Tissues Organs* **185**, 232-236 (2007).

Bessette, D.C., Qiu, D. & Pallen, C.J. PRL PTPs: mediators and markers of cancer progression. *Cancer Metastasis Rev.* **27**, 231-252 (2008).

Blanchetot, C., Tertoolen, L.G.J. & den Hertog, J. Regulation of receptor protein-tyrosine phosphatase alpha by oxidative stress. *EMBO J.* **21**, 493-503 (2002).

Blanchetot, C., Chagnon, M., Dubé, N., Hallé, M. & Tremblay, M.L. Substrate-trapping techniques in the identification of cellular PTP targets. *Methods* **35**, 44-53 (2005).

Bonham, C. A. & Vacratsis, P. O. Redox regulation of the human dual specificity phosphatase YVH1 through disulfide bond formation. *J. Biol. Chem.* **284**, 22853-22864 (2009).

Böhmer, F., Szedlacsek, S., Tabernero, L., Östman, A. & den Hertog, J. Protein tyrosine phosphatase structure-function relationships in regulation and pathogenesis. *FEBS J.* **280**, 413-431 (2013).

Brandt, D.T. *et al.* Dia1 and IQGAP1 interact in cell migration and phagocytic cup formation. *J. Cell Biol.* **178**, 193-200 (2007).

Braunstein, J., Brutsaert, S., Olson, R. & Schindler, C. STATs dimerize in the absence of phosphorylation. *J. Biol. Chem.* **278**, 34133-34140 (2003).

Buhrman, G., Parker, B., Sohn, J., Rudolph, J & Mattos, C. Structural mechanism of oxidative regulation of the phosphatase Cdc25B via an intramolecular disulfide bond. *Biochemistry* **44**, 5307-5316 (2005).

Burke, Jr. & Zhang, Z.-Y. Protein-tyrosine phosphatases: structure, mechanism, and inhibitor discovery. *Biopolymers.* **47**, 225-241 (1998).

Cabodi, S. *et al.* p130Cas as a new regulator of mammary epithelial cell proliferation, survival, and HER2-neu oncogene-dependent breast tumorigenesis. *Cancer Res.* **66**, 4672-4680 (2006).

Campbell, R.B., Liu, F. & Ross, A.H. Allosteric activation of PTEN phosphatase by phosphatidylinositol 4,5-bisphosphate. *J. Biol. Chem.* **278**, 33617-33620 (2003).

Candiano, G. *et al.* Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**, 1327-1333 (2004).

Carisey, A. & Ballestrem, C. Vinculin, an adapter protein in control of cell adhesion signalling. *Eur. J. Cell Biol.* **90**, 157-163 (2011).

Cates, C.A. et al. Prenylation of oncogenic human PTP_{CAAX} protein tyrosine phosphatases. *Cancer Lett.* **110**, 49-55 (1996).

Caunt, C.J. & Keyse, S.M. Dual-specificity MAP kinase phosphatases (MKPs): shaping the outcome of MAP kinase signalling. *FEBS J.* **280**, 489-504 (2013).

Chambers, A.F., Groom, A.C. & MacDonald, I.C. Dissemination and growth of cancer cells in metastatic sites. *Nature Rev. Cancer* **2**, 563-572 (2002).

Chan, H.-L. *et al.* Major role of epidermal growth factor receptor and Src kinases in promoting oxidative stress-dependent loss of adhesion and apoptosis in epithelial cells. *J. Biol. Chem.* **285**, 4307-4318 (2010).

Chang, L. & Karin, M. Mammalian MAP kinase signalling cascades. *Nature* **410**, 37-40 (2001).

Charbonneau, H. *et al.* Human placenta protein-tyrosine-phoshatase: amino acid sequence and relationship to a family of receptor-like proteins. *Proc. Natl. Acad. Sci. USA* **86**, 5252-5256 (1989).

Chen, S., Schultz, P. G. & Brock, A. An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **371**, 112-122 (2007).

Chen, C.-Y., Willard, D. & Rudolph, J. Redox regulation of SH2-domain-containing protein tyrosine phosphatases by two backdoor cysteines. *Biochemistry* **48**, 1399-1409 (2009).

Chin, J.W., Martin, A.B., King, D.S., Wang, L. & Schultz, P.G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **99**, 11020-11024 (2002a).

Chin, J.W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* **124**, 9026-9027 (2002b).

Chin, J.W. & Schultz, P.G. In vivo photocrosslinking with unnatural amino acid mutagenesis. *ChemBioChem* **11**, 1135-1137 (2002).

Chin, J.W. *et al*. An expanded eukaryotic genetic code. *Science* **301**, 964-966 (2003).

Chin, J.W. Reprogramming the genetic code. *EMBO J.* **30,** 2312-2324 (2011).

Choi, M.-S. *et al.* The essential role of FKBP38 in regulating phosphatase of regenerating liver 3 (PRL-3) protein stability. *Biochem. Biophys. Res. Commun.* **406**, 305-309 (2011).

Choy, E. *et al.* Endomembrane trafficking of Ras: the CAAX motif targets proteins to the ER and Golgi. *Cell* **98**, 69-80 (1999).

Chou, C., Uprety, R., Davis, L., Chin, J.W. & Deiters, A. Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chem. Sci.* **2**, 480-483 (2011).

Chu, Z.-H. *et al.* Proteomic analysis identifies translationally controlled tumor protein as a mediator of phosphatase of regenerating liver-3-promoted proliferation, migration and invasion in human colon cancer cells. *Chin. Med. J.* **124**, 3778-3785 (2011).

Cohen, P. The origins of protein phosphorylation. *Nat Cell Biol* **4**, E127-E130 (2002).

Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).

Cully, M., You, H., Levine, A. J. & Mak, T. W. Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis. *Nat. Rev. Cancer* **6**, 184-192 (2006).

Daouti, S. *et al.* A selective phosphatase of regenerating liver phosphatase inhibitor suppresses tumor cell anchorage-independent growth by a novel mechanism involving p130Cas cleavage. *Cancer Res.* **68**, 1162-1169 (2008).

Deb, I., Poddar, R. & Paul, S. Oxidative stress-induced oligomerization inhibits the activity of the non-receptor tyrosine phosphatase STEP61. *J. Neurochem.* **116**, 1097-111 (2011).

de Curtis, I. & Meldolesi, J. Cell-surface dynamics-how Rho GTPases orchestrate the interplay between the plasma membrane and the cortical cytoskeleton. *J. Cell Sci.* **125**, 4435-4444 (2012).

den Hertog, J. Regulation of protein phosphatases in disease and behaviour. *EMBO Rep* **4**, 1027-1031 (2003).

den Hertog, J., Ostman, A. & Böhmer, F.-D. Protein tyrosine phosphatases: regulatory mechanisms. *FEBS J.* **275**, 831-847 (2008).

Denu, J.M., Zhou, G., Wu, L., Zhao, R., Yuvaniyama, J., Saper, M.A. & Dixon, J.E. The purification and characterisation of a human dual-specific protein tyrosine phosphatase. *J. Biol. Chem.* **270**, 3796-3803 (1995a).

Denu, J.M., Zhou, G., Guo, Y. & Dixon, J.E. The catalytic role of aspartic acid-92 in a human dual-specific protein-tyrosine-phosphatase. *Biochemistry* **34**, 3396-3403 (1995b).

Denu, J. M., Stuckey, J. A., Saper, M. A. & Dixon, J. E. Form and function in protein dephosphorylation. *Cell* **87**, 361-364 (1996a).

Denu, J.M., Lohse, D.L., Vijayalakshmi, J., Saper, M. & Dixon, J.E. Visualization of intermediate and transition-state structures in protein-tyrosine phosphatase catalysis. *Proc. Natl. Acad. Sci. USA* **93**, 2493-2498 (1996b).

Denu, J.M. & Tanner, K.G. Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulfenic acid intermediate and implications for redox regulation. *Biochemistry* **37**, 5633-5642 (1998).

DeVita, Jr., V.T., Lawrence, T.S. & Rosenberg, S.A. Cancer: Principles & practice of oncology. Primer of the molecular biology of cancer. (2011).

Dey, S., Pal, A., Chakrabarti, P. & Janin, J. The subunit interfaces of weakly associated homodimeric proteins. *J. Mol. Biol.* **398**, 146-160 (2010).

Diamond, R.H., Cressman, D.E. & Laz, T.M. PRL-1 , a unique nuclear protein tyrosine phosphatase, affects cell growth. *Mol. Cell. Biol.* **14**, 3752-3762 (1994).

Diamond, R.H. *et al.* Expression of PRL-1 nuclear PTPase is associated with proliferation in liver but with differentiation in intestine. *Am. J. Physiol. Gastrointest. Liver Physiol.* **271**, G121-G129 (1996).

Dickinson, R.J. & Keyse, S.M. Diverse physiological functions for dual-specificity MAP kinase phosphatases. *J. Cell Sci.* **119**, 4607-4615 (2006).

Dong, Y. *et al.* Phosphatase of regenerating liver 2 (PRL2) is essential for placental development by down-regulating PTEN (phosphatase and tensin homologue deleted on chromosome 10) and activating Akt protein. *J. Biol. Chem.* **287**, 32172-32179 (2012).

Dormán, G. & Prestwich, G.D. Benzophenone photophores in biochemistry. *Biochemistry* **33**, 5661-5673 (1994).

Dumaual, C.M., Sandusky, G.E., Crowell, P.L. & Randall, S.K. Cellular localisation of PRL-1 and PRL-2 gene expression in normal adult human tissues. *J. Histochem. Cytochem.* **54**, 1401-1412 (2006).

Dumaual, C.M. *et al.* Tissue-specific alterations of PRL-1 and PRL-2 expression in cancer. *Am. J. Transl. Res.* **4**, 83-101 (2012).

Eckhart, W. *et al.* An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. *Cell* **18**, 925-933 (1979).

Elchebly, M. *et al.* Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* **283**, 1544-1548 (1999).

Fagerli, U.-M. *et al.* Overexpression and involvement in migration by the metastasis-associated phosphatase PRL-3 in human myeloma cells. *Blood* **111**, 806-815 (2008).

Farrell, I.S., Toroney, R., Hazen, J.L., Mehl, R.A. & Chin, J.W. Photo-cross-linking interacting proteins with a genetically encoded benzophenone. *Nature Methods* **2**, 377-384 (2005).

Finkel, T. From Sulfenylation to Sulfhydrations: What a Thiolate Needs to Tolerate. *Sci. Signal.* **5**, pe10 (2013).

Fiordalisi, J.J., Keller, P.J. & Cox, A.D. PRL tyrosine phosphatases regulate rho family GTPases to promote invasion and motility. *Cancer Res.* **66**, 3153-3161 (2006).

Fiordalisi, J. J., Dewar, B. J., Graves, L. M., Madigan, J. P. & Cox, A. D. Src-mediated phosphorylation of the tyrosine phosphatase PRL-3 is required for PRL-3 promotion of Rho activation, motility and invasion. *PLoS One* **8**, e64309 (2013).

Flint, A.J., Tiganis, T., Barford, D. & Tonks, N.K. Development of “substrate-trapping” mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc. Natl. Acad. Sci. USA* **94**, 1680-1685 (1997).

Folkman, J. & Klagsbrun, M. Angiogenic factors. *Science* **235**, 442-447 (1987).

Forné, I., Ludwigsen, J., Imhof, A., Becker, P.B. & Mueller-Planitz, F. Probing the conformation of the ISWI ATPase domain with genetically encoded photoreactive crosslinkers and mass spectrometry. *Mol. Cell. Proteomics* **11**, M111.012088 (2012).

Fruehauf, J.P. & Meyskens, F.L. Reactive oxygen species: a breath of life or death? *Clin. Cancer Res.* **13**, 789-794 (2007).

Funato, Y. & Miki, H. Reversible oxidation of PRL family protein-tyrosine phosphatases. *Methods* **65**, 184-189 (2014).

Gao, J., Liao, J. & Yang, G.-Y. CAAX-box protein, prenylation process and carcinogenesis. *Am. J. Transl. Res.* **1**, 312-325 (2009).

Garrod, D. & Chidgey, M. Desmosome structure, composition and function. *Biochim. Biophys. Acta* **1778**, 572-587 (2008).

Gentry, M. S., Romá-Mateo, C. & Sanz, P. Laforin, a protein with many faces: glucan phosphatase, adapter protein, et alii. *FEBS J.* **280**, 525-537 (2013).

Gough, D.R. & Cotter, T.G. Hydrogen peroxide: a Jekyll and Hyde signaling molecule. *Cell death & disease* **2**, e213 (2011).

Gräslund, S. et al. Protein production and purification. *Nature Methods*, **5**, 135-146 (2008).

Groen, A. *et al.* Differential oxidation of protein-tyrosine phosphatases. *J. Biol. Chem.* **280**, 10298-10304 (2005).

Guo, K. *et al.* Catalytic domain of PRL-3 plays an essential role in tumor metastasis. *Cancer Biol. Ther.* **3**, 945-951 (2004).

Guo, K. *et al.* PRL-3 initiates tumor angiogenesis by recruiting endothelial cells In vitro and In vivo. *Cancer Res.* **6**, 9625-9635 (2006).

Haagenson, K.K. & Wu, G.S. Mitogen activated protein kinase phosphatases and cancer. *Cancer Biol. Ther.* **9**, 337-340 (2010).

Haberkant, P. & van Meer, G. Protein-lipid interactions: paparazzi hunting for snap-shots. *Biol. Chem.* **390**, 795-803 (2009).

Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).

Hao, L. & ElShamy, W.M. BRCA1-IRIS activates cyclin D1 expression in breast cancer cells by downregulating the JNK phosphatase DUSP3/VHR. *Int. J. Cancer* **121**, 39-46 (2007).

Hardy, S., Wong, N. N., Muller, W. J., Park, M. & Tremblay, M. L. Overexpression of the protein tyrosine phosphatase PRL-2 correlates with breast tumorf formation and progression. *Cancer Res.* **70**, 8959-8967 (2010).

He, R., Zeng, L.-F., He, Y., Zhang, S. & Zhang, Z.-Y. Small molecule tools for functional interrogation of protein tyrosine phosphatases. *FEBS J.* **280**, 731-750 (2013).

Hendriks, W.J.A.J. *et al.* Protein tyrosine phosphatases in health and disease. *FEBS J.* **280** 708-730 (2013).

Hengge, A.C., Denu, J.M. & Dixon, J.E. Transition-state structures for the native dual-specific phosphatase VHR and D92N and S131A mutants. Contributions to the driving force for catalysis. *Biochemistry* **35**, 7084-7092 (1996).

Henkens, R. *et al.* Cervix carcinoma is associated with an up-regulation and nuclear localisation of the dual-specificity protein phosphatase VHR. *BMC Cancer* **8**, 147-155 (2008).

Hino, N. *et al.* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nature Methods* **2**, 3-8 (2005).

Hino, N., Hayashi, A., Sakamoto, K. & Yokoyama, S. Site-specific incorporation of non-natural amino acids into proteins in mammalian cells with an expanded geentic code. *Nature Protocols* **1**, 2957-2962 (2006).

Hino, N., Sakamoto, K. & Yokoyama, S. Unnatural amino acids: methods and protocols. *Methods in molecular biology* (editors: Pollegioni, L. & Servi, S.) **794**, 215-228 (2012).

Hoffman, B.T., Nelson, M.R., Burdick, K. & Baxter, S.M. Protein tyrosine phosphatases: strategies for distinguishing proteins in a family containing multiple drug targets and anti-targets. *Curr. Pharm. Des.* 10, 1161-1181 (2004).

Hoyt, R., Zhu, W., Cerignoli, F. *et al.* Cutting edge: selective tyrosine dephosphorylation of interferon-activated nuclear STAT5 by the VHR. *J. Immunol.* **179**, 3402-3406 (2007).

Hunter, T. Signaling-2000 and beyond. *Cell* **100**, 113-127 (2000).

Ishibashi, T., Bottaro, D.P., Chan, A., Miki, T. & Aaronson, S.A. Expression cloning of a human dual-specificity phosphatase. *Proc. Natl. Acad. Sci. USA* **89**, 12170-12174 (1992).

Ishii, T., Funato, Y. & Miki, H. Thioredoxin-related protein 32 (TRP32) specifically reduces oxidised phosphatase of regenerating liver (PRL). *J. Biol. Chem.* **288**, 7263-7270 (2013).

Jackson, M.D. & Denu, J.M. Molecular reactions of protein phosphatases-insights from structure and chemistry. *Chem. Rev.* **101**, 2313-2340 (2001).

Jardin, C. & Sticht, H. Identification of the structural features that mediate binding specificity in the recognition of STAT proteins by dual-specificity phosphatases. J. Biomol. Struct. Dyn. **29**, 777-792 (2012).

Jeong, D.G. *et al.* Trimeric structure of PRL-1 phosphatase reveals an active enzyme conformation and regulation mechanisms. *J. Mol. Biol.* **345**, 401-413 (2005).

Jeong, D.G. *et al.* Crystal structure of the catalytic domain of human MKP-2 reveals a 24-mer assembly. *Proteins* **76**, 763-767 (2009).

Jian, M. *et al.* Downregulating PRL-3 inhibit migration and invasion of lung cancer cell via RhoA and mDia1. *Tumori* **98**, 370-376 (2012).

Jiang, Y., Liu, X.-Q., Rajput, A. *et al.* Phosphatase PRL-3 is a direct target of TGFβ in colon cancer metastasis. *Cancer Res.* **71**, 234-244 (2011).

Julien, S.G., Dubé, N., Hardy, S. & Tremblay, M.L. Inside the human cancer tyrosine phosphatome. *Nat. Rev. Cancer* **11**, 35-49 (2011).

Kage, R., Leeman, S. E., Krause, J. E., Costello, C. E. & Boyd, N. D. Identification of methionine as the site of covalent attachment of a p-Benzoyl-Phenylalanine-containing analogue of substance P on the substance P (NK-1) Receptor. *J. Biol. Chem.* **271**, 25797-25800 (1996).

Kamata, H. *et al.* Reactive oxygen species promote TNFα-induced death and sustained JNK activation by inhibiting MAP kinase phosphatases. *Cell* **120**, 649-661 (2005).

Kang, T.-H. & Kim, K.-T. Negative regulation of ERK activity by VRK3-mediated activation of VHR phosphatase. *Nat. Cell Biol.* **8**, 863-869 (2006).

Karisch, R. & Neel, B.G. Methods to monitor classical protein-tyrosine phosphatase oxidation. *FEBS J.* **280**, 459-475 (2013).

Kato, H. *et al.* High expression of PRL-3 promotes cancer cell motility and liver metastasis in human colorectal cancer: a predictive molecular marker of metachronous liver and lung metastases. *Clin. Cancer Res.* **10**, 7318-7328 (2004).

Katz, M., Amit, I. & Yarden, Y. Regulation of MAPKs by growth factors and receptor tyrosine kinases. *Biochim. Biophys. Acta* **1773**, 1161-1176 (2007).

Kauer, J.C., Erickson-Viitanen, S., Wolfe Jr., H.R. & DeGrdao, W.F. p-Benzoyl-L-phenylalanine, a new photoreactive amino acid. *J. Biol. Chem.* **261**, 10695-10700 (1986).

Keyse, S.M. Protein phosphatases and the regulation of mitogen-activated protein kinase signalling. *Curr. Opin. Cell Biol.* **12**, 186-192 (2000).

Keyse, S.M. Dual-specificity MAP kinase phosphatases (MKPs) and cancer. *Cancer Metastasis Rev.* **27**, 253-261 (2008).

Khapare, N., Kunde, S.T., Sehgal, L. *et al.* Plakophilin3 loss leads to an increase in PRL3 levels promoting K8 dephosphorylation, which is required for transformation and metastasis. *PlosONE* **7**, e38561 (2012).

Kim, J.H., Shin, D.Y., Han, M.H. & Choi, M.U. Mutational and kinetic evaluation of conserved His-123 in dual specificity protein-tyrosine phosphatase vaccinia H1-related phosphatase: participation of Tyr-78 and Thr-73 residues in tuning the orientation of His-123. *J. Biol. Chem.* **276**, 27568-27574 (2001).

Kim, K.-A. *et al.* Structure of human PRL-3, the phosphatase associated with cancer metastasis. *FEBS Lett.* **565**, 181-187 (2004).

Kim, S.J. *et al.* Crystal structure of human TMDP, a testis-specific dual specificity protein phosphatase: implications for substrate s pecificity. *Proteins* **245**, 239-245 (2007).

Klaman, L.D. *et al.* Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice. *Mol. Cell. Biol.* **20**, 5479-5489 (2000).

Klemm, J.D., Schreiber, S.L. & Crabtree, G.R. Dimerisation as a regulatory mechanism in signal transduction. *Annu. Rev. Immunol.* **16,** 569-592 (1998).

Klockenbusch, C. & Kast, J. Optimisation of formaldehyde cross-linking for protein interaction analysis of non-tagged integrin beta1. *J. Biomed. Biotechnol.* **2010**, Article ID 927585 (2010).

Koksal, A.C., Nardozzi, J.D. & Cingolani, G. Dimeric quaternary structure of the prototypical dual specificity phosphatase VH1. *J. Biol. Chem.* **284**, 10129-10137 (2009).

Koksal, A.C. & Cingolani, G. Dimerisation of Vaccinia Virus VH1 is essential for dephosphorylation of STAT1 at tyrosine 701. *J. Biol. Chem.* **286**, 14373-14382 (2011).

Kolmodin, K. & Åqvist, J. The catalytic mechanism of protein tyrosine phosphatases revisited. *FEBS Lett.* **498**, 208-213 (2001).

Kong, W., Swain, G.P., Li, S. & Diamond, R.H. PRL-1 PTPase expression is developmentally regulated with tissue-specific patterns in epithelial tissues. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G613-G621 (2013).

Kotani, H., Takaishi, K., Sasaki, T. & Takai, Y. Rho regulates association of both the ERM family and vinculin with the plasma membrane in MDCK cells. *Oncogene* **14**, 1705-1713 (1997).

Kozlov, G. *et al.* Structural insights into molecular function of the metastasis-associated phosphatase PRL-3. *J. Biol. Chem.* **279**, 11882-11889 (2004).

Köhrer, C., Yoo, J.-H., Bennett, M., Schaack, J. & RajBhandary, U. L. A possible approach to site-specific insertion of two different unnatural amino acids into proteins in mammalian cells via nonsense suppression. *Chem. Biol.* **10**, 1095-1102 (2003).

Krishnamurthy, M. *et al.* Caught in the act: Covalent cross-linking captures activator-coactivator interactions *in vivo*. *ACS Chem. Biol.* **6**, 1321-1326 (2011).

Krndija, D. *et al.* The phosphatase of regenerating liver 3 (PRL-3) promotes cell migration through Arf-activity-dependent stimulation of integrin α5 recycling. *J. Cell Sci.* **125**, 3883-3892 (2012).

Kundu, S. T. *et al.* Plakophilin3 downregulation leads to a decrease in cell adhesion and promotes metastasis. *Int. J. Cancer* **123**, 2303-2314 (2008).

Lahiry, P., Torkamani, A., Schork, N.J. & Hegele, R.A. Kinase mutations in human disease: interpreting genotype-phenotype relationships. *Nat. Rev. Genet.* **11**, 60-74 (2010).

Lee, J.-O. *et al.* Crystal structure of the PTEN tumor suppressor: Implications for its phosphoinositide phosphatase activity and membrane association. *Cell* **99**, 323-334 (1999).

Lee, S.-R., Yang, K.-S. *et al.* Reversible inactivation of the tumor suppressor PTEN by H_2O_2. *J. Biol. Chem.* **277**, 20336-20342 (2002).

Leitner, A. *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **9**, 1634-1649 (2010).

Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Na.t Chem. Biol.* **3**, 769-772 (2007).

Lemmon, M.A. & Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **141**, 1117-1134 (2010).

Leslie, N.R., Gray, A., Pass, I., Orchiston, E.A. & Downes, C.P. Analysis of the cellular functions of PTEN using catalytic domain and C-terminal mutations: differential effects of C-terminal deletion on signalling pathways downstream of phosphoinositide 3-kinase. *Biochem. J.* **346**, 827-833 (2000).

Leslie, N.R. *et al.* Targeting mutants of PTEN reveal distinct subsets of tumour suppressor functions. *Biochem. J.* **357**, 427-435 (2001).

Li, X., Wilmanns, M., Thornton, J. & Köhn. Elucidating human phosphatase-substrate networks. *Sci. Signal.* **6**, rs1-rs10 (2013).

Liang, F. *et al.* PRL3 promotes cell invasion and proliferation by down-regulation of Csk Leading to Src activation. *J. Biol. Chem.* **28**, 5413-5419 (2007).

Liang, F. *et al.*Translational control of C-terminal Src kinase (Csk) expression by PRL3 phosphatase *J. Biol. Chem.* **283**, 10339-10346 (2008).

Lin, M.-D. *et al.* Expression of phosphatase of regenerating liver family genes during embryogenesis: an evolutionary developmental analysis among Drosophila, amphioxus, and zebrafish. *BMC Dev. Biol.* **13**, 18 (2013).

Liu, D.R. & Schultz, P.G. Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. USA* **96**, 4780-4785 (1999).

Liu, Y.-N. *et al.* Regulatory mechanisms controlling human E-cadherin gene expression. *Oncogene* **24**, 8277-8290 (2005).

Liu, Y., Wang, Y., Wu, C., Liu, Y. & Zheng, P. Dimerisation of Laforin is required for its optimal phosphatase activity, regulation of GSK3beta phosphorylation, and Wnt signaling. *J. Biol. Chem.* **281**, 34768-34774 (2006).

Liu, W., Brock, A., Chen, S., Chen, S. & Schultz, P.G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nature Methods* **4**, 239-244 (2007).

Liu, Y. *et al.* PRL-3 promotes epithelial mesenchymal transition by regulating cadherin directly. *Cancer Biol. Ther.* **8**, 1352-1359 (2009).

Liu, C.C. & Schultz, P.G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413-444 (2010).

Liu, H., Al-Aidaroos, A.Q.O. *et al.* PRL-3 suppresses c-Fos and integrin α2 expression in ovarian cancer cells. *BMC Cancer* **13**, 80 (2013).

Lokareddy, R.K., Bhardwaj, A. & Cingolani, G. Atomic structure of dual-specificity phosphatase 26, a novel p53 phosphatase. *Biochemistry* **52**, 938-948 (2013).

Lountos, G.T., Tropea, J.E. & Waugh, D.S. Structure of human dual-specificity phosphatase 27 at 2.38 Å resolution. *Acta crystallog.* **67**, 471-479 (2011).

Luechapanichkul, R. *et al.* Specificity profiling of dual specificity phosphatase VHR reveals two distinct substrate-binding modes. *J. Biol. Chem.* **288**, 6498-6510 (2013).

Luo, Y., Liang, F. & Zhang, Z.-Y. PRL1 promotes cell migration and invasion by increasing MMP2 and MMP9 expression through Src and ERK1/2 pathways. *Biochemistry* **48**, 1838-1846 (2009).

Majeti, R. & Weiss, A.. Regulatory mechanisms for receptor protein tyrosine phosphatases. *Chem. Rev.* **101**, 2441-2448 (2001).

Majmudar, C.Y. *et al.* Impact of nonnatural amino acid mutagenesis on the in vivo function and binding modes of a transcriptional activator. *J. Am. Chem. Soc.* **131**, 14240-14242 (2009).

Matsukawa, Y. *et al.* Constitutive suppression of PRL-3 inhibits invasion and proliferation of gastric cancer cell in vitro and in vivo. *Pathobiology* **77**, 155-162 (2010).

Matter, W.F. *et al.* Role of PRL-3, a human muscle-specific tyrosine phosphatase, in angiotensin-II signaling. *Biochem. Biophys. Res. Commun.* **283**, 1061-1068 (2001).

McParland, V., Varsano, G. *et al.* The metastasis-promoting phosphatase PRL-3 shows activity toward phosphoinositides. *Biochemistry* **50**, 7579-7590 (2011).

Meng, T.-C., Fukada, T. & Tonks, N.K. Reversible oxidation and inactivation of protein tyrosine phosphatases *in vivo*. *Mol. cell* **9**, 387-399 (2002).

Meng, T., Lou, Y., Chen, Y., Hsu, S. & Huang, Y. Cys-oxidation of protein tyrosine phosphatases: its role in regulation of signal transduction and its involvement in human cancers. *J. Cancer Mol.* **2**, 9-16 (2006).

Migneault, I., Dartiguenave, C., Bertrand, M. J. & Waldron, K. C. Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *Biotechniques* **37**, 790-802 (2004).

Miki, H. & Funato, Y. Regulation of intracellular signalling through cysteine oxidation by reactive oxygen species. *J. Biochem.* **151**, 255-261 (2012).

Miller, J.H. & Albertini, A.M. Effects of surrounding sequence on the suppression of nonsense codons. *J. Mol. Biol.* **164**, 59-71 (1983).

Min, S.-H. *et al.* New p53 target, phosphatase of regenerating liver 1 (PRL-1) downregulates p53. *Oncogene* **28**, 545-554 (2009).

Mizuuchi, E., Semba, S., Kodama, Y. & Yokozaki, H. Down-modulation of keratin 8 phosphorylation levels by PRL-3 contributes to colorectal carcinoma progression. *Int. J. Cancer* **124**, 1802-1810 (2009).

Mollevi, D.G. *et al.* PRL-3 is essentially overexpressed in primary colorectal tumours and associated with tumour agresiveness. *Br. J. Cancer* **99**, 1718-1725 (2008).

Motiwala, T. & Jacob, S.T. Role of protein tyrosine phosphatases in cancer. *Prog. Nucleic Acid Res. Mol. Biol.* **81**, 297-329 (2006).

Muir, T.W. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72**, 249-289 (2003).

Murray, D., Ben-Tal, N., Honig, B. & McLaughlin, S. Electrostatic interaction of myristoylated proteins with membranes: simple physics, complicated biology. *Structure* **5**, 985-989 (1997).

Musial-Siwek, M., Rusch, S.L. & Kendall, D.A. Selective photoaffinity labeling identifies the signal peptide binding domain on SecA. *J. Mol. Biol.* **365**, 637-648 (2007).

Myers, M.P., Pass, I., Batty, I.H., Van der Kaay, J., Stolarov, J.P., Hemmings, B.A., Wigler, M.H., Downes, C.P. & Tonks, N.K. The lipid phosphatase activity of PTEN is critical for its tumor supressor function. *Proc Natl Acad Sci USA* **95**, 13513-13518 (1998).

Nakashima, M. & Lazo, J. S. Phosphatase of regenerating liver-1 promotes cell migration and invasion and regulates filamentous actin dynamics. *J. Pharmacol. Exp. Ther.* **334**, 627-633 (2010).

Nardozza, A.P. *et al.* Reactive oxygen species and epidermal growth factor are antagonistic cues controlling SHP-2 dimerisation. *Mol. Cell. Biol.* **32**, 1998-2009 (2012).

Neel, B.G. & Tonks, N.K. Protein tyrosine phosphatases in signal transduction. *Curr. Opin. Cell Biol.* **9**, 193-204 (1997).

Nishimura, A. & Linder, M.E. Identification of a novel prenyl and palmitoyl modification at the CaaX motif of Cdc42 that regulates RhoGDI binding. *Mol. Cell. Biol.* **33**, 1417-1429 (2013).

Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182-188 (1989).

Olayioye, M.A., Beuvink, I., Horsch, K., Daly, J.M. & Hynes, N.E. ErbB Receptor-induced Activation of Stat Transcription Factors Is Mediated by Src Tyrosine Kinases. *J. Biol. Chem.* **274**, 17209-17218 (1999).

Olsen, J.V *et al.* Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648 (2006).

O' Neil, K.T., Erickson-Viitanen, S. & DeGrado, W.F. Photolabeling of calmodulin with basic α-helical peptides containing *p*-benzoylphenylalanine. *J. Biol. Chem.* **25**, 14571-14578 (1989).

O'Neill, G. M., Fashena, S. J. & Golemis, E. a. Integrin signalling: a new Cas(t) of characters enters the stage. *Trends Cell Biol.* **10**, 111-119 (2000).

Orsatti, L., Innocenti, F., Surdo, P. Lo, Talamo, F. & Barbato, G. Mass spectrometry study of PRL-3 phosphatase inactivation by disulfide bond formation and cysteine into glycine conversion. *Rapid Commun. Mass Spectrom.* **23**, 2733-2740 (2009a).

Orsatti, L. *et al.* 2-D difference in gel electrophoresis combined with Pro-Q diamond staining: a successful approach for the identification of kinase/phosphatase targets. *Electrophoresis* **30**, 2469-2476 (2009b).

Östman, A., Frijhoff, J., Sandin, A. & Böhmer, F.-D. Regulation of protein tyrosine phosphatases by reversible oxidation. *J. Biochem.* **150**, 345-356 (2011).

Ozbabacan, S. E. A., Engin, H. B., Gursoy, A. & Keskin, O. Transient protein-protein interactions. *Protein Eng. Des. Sel.* **24**, 635-648 (2011).

Parker, B.S. *et al.* Alterations in vasccular gene expression in invasive breast carcinoma. *Cancer Res.* **64**, 7857-7866 (2004).

Pascaru, M., Tanase, C. *et al.* Analysis of molecular determinants of PRL-3. *J. Cell. Mol. Med.* **13**, 3141-3150 (2009).

Patterson, K.I., Brummer, T., Brien, P.M.O. & Daly, R.J. Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochem. J.* **489**, 475-489 (2010).

Peng, Y., Du, K., Ramirez, S., Diamond, R.H. & Taub, R. Mitogenic up-regulation of the PRL-1 protein-tyrosine phosphatase gene by Egr-1. *J. Biol. Chem.* **274**, 4513-4520 (1999).

Peng, L., Jin, G. *et al.* Identification of integrin α1 as an interacting protein of protein tyrosine phosphatase PRL-3. *Biochem. Biophys. Res. Commun.* **342**, 179-183 (2006).

Peng, L. *et al.* PRL-3 promotes the motility, invasion, and metastasis of LoVo colon cancer cells through PRL-3-integrin beta1-ERK1/2 and-MMP2 signaling. *Mol. Cancer* **8**, 110-123 (2009).

Peters, G.H., Frimurer, T.M. & Olsen, O.H. Electrostatic evaluation of the signature motif (H/V)CX5R(S/T) in protein-tyrosine phosphatases†. *Biochemistry* **37**, 5383-5393 (1998).

Peters, C.S. *et al.* ATF-7, a novel bZIP protein, interacts with the PRL-1 protein tyrosine phosphatase. *J. Biol. Chem.* **276**, 13718-13726 (2001).

Plantefaber, L. C. & Hynes, R. O. Changes in integrin receptors on oncogenically transformed cells. *Cell* **56**, 281-290 (1989).

Poole, L.B. & Nelson, K.J. Discovering mechanisms of signaling-mediated cysteine oxidation. *Curr. Opin. Chem. Biol.* **12**, 18-24 (2008).

Pumiglia, K.M. & Decker, S.J. Cell cycle arrest mediated by the MEK/mitogen-activated protein kinase pathway. *Proc. Natl. Acad. Sci. USA* **94**, 448-452 (1997).

Rahmouni, S. *et al.* Loss of the VHR dual-specific phosphatase causes cell-cycle arrest and senescence. *Nat. Cell Biol.* **8**, 524-531 (2006).

Raman, M., Chen, W. & Cobb, M. H. Differential regulation and properties of MAPKs. *Oncogene* **26**, 3100-3112 (2007).

Ranjbar, B. & Gill, P. Circular dichroism techniques: biomolecular and nanostructural analyses- a review. *Chem. Biol. Drug. Des.* **74**, 101-120 (2009).

Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173**, 530-540 (2011).

Rhee, S.G., Bae, Y.S., Lee, S.-R. & Kwon, J. Hydrogen Peroxide: A key messenger that modulates protein phosphorylation through cysteine oxidation. *Sci. Signal.* **2000**, pe1–pe1 (2000).

Rhee, S.G. H_2O_2, a necessary evil for cell signaling. *Science* **312**, 1882-1883 (2006).

Rios, P., Li, X. & Köhn, M. Molecular mechanisms of the PRL phosphatases. *FEBS J.* **280**, 505-524 (2013).

Robinson, M.J. & Cobb, M.H. Mitogen-activated protein kinase pathways. *Curr. Opin. Cell Biol.* **9**, 180-186 (1997).

Rodriguez, E.A., Lester, H.A. & Dougherty, D.A. *In vivo* incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci. USA* **103**, 8650-8655 (2006).

Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601 (1984).

Romá-Mateo, C., Ríos, P., Tabernero, L., Attwood, T.K. & Pulido, R.A novel phosphatase family, structurally related to dual-specificity phosphatases, that displays unique amino acid sequence and substrate specificity. *J. Mol. Biol.* **374**, 899-909 (2007).

Roos, G. & Messens, J. Protein sulfenic acid formation: from cellular damage to redox regulation. *Free Radic. Biol. Med.* **51**, 314-326 (2011).

Rosenfeld, J., Capdevielle, J., Guillemot, J.C. & Ferrara, P. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal Biochem.* **203**, 173-179 (1992).

Ryazanov, A. G. Elongation factor-2 kinase and its newly discovered relatives. *FEBS Lett.* **514**, 26-29 (2002).

Ryu, Y. & Schultz, P.G. Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*. *Nature Methods* **3**, 263-265 (2006).

Saha, S., Bardelli, A. *et al.* A phosphatase associated with metastasis of colorectal cancer. *Science* **294**, 1343-1346 (2001).

Sakamoto, K. *et al.* Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* **30**, 4692-4699 (2002).

Sato, S. *et al.* Crystallographic study of a site-specifically cross-linked protein complex with a genetically incorporated photoreactive amino acid. *Biochemistry* **50**, 250-257 (2011).

Saxena, M. & Mustelin, T. Extracellular signals and scores of phosphatases: All roads lead to MAP kinase. *Semin. Immunol.* **12**, 387-396 (2000).

Schlessinger, J. Signal transduction by allosteric receptor oligomerization. *Trends Biochem. Sci.* **13**, 443-447 (1988).

Schumacher, M. *et al.* Structural basis for the recognition of a bisphosphorylated MAP kinase peptide. *Biochemistry* **41**, 3009-3017 (2002).

Schlessinger, J. Signal transduction by allosteric receptor oligomerization. *TIBS* **13**, 443-447 (1988).

Semba, S., Mizuuchi, E. & Yokozaki, H. Requirement of phosphatase of regenerating liver-3 for the nucleolar localisation of nucleolin during the progression of colorectal carcinoma. *Cancer Sci.* **101**, 2254-2261 (2010).

Seger, R. & Krebs, E.G. The MAPK signaling cascade. *FASEB J.* **9**, 726-735 (1995).

Seth, D. & Rudolph, J. Articles Redox regulation of MAP kinase phosphatase 3 †. *Biochemistry* **45**, 8476-8487 (2006).

Shi, Z. *et al.* Identification of a potent inhibitor of human dual-specific phosphatase, VHR, from computer-aided and NMR-based screening to cellular effects. *ChemBioChem* **8**, 2092-2099 (2007).

Silver, D. M. *et al.* Insight into the redox regulation of the phosphoglucan phosphatase SEX4 involved in starch degradation. *FEBS J.* **280**, 538-548 (2013).

Skinner, A. L., Vartia, A. A., Williams, T. D. & Laurence, J. S. Enzyme activity of phosphatase of regenerating liver is controlled by the redox environment and its C-terminal residues. *Biochemistry* **48**, 4262-4272 (2009).

Song, H. *et al.* Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2. *Mol. Cell* **7**, 615-626 (2001).

Song, R. *et al.* Phosphatase of regenerating liver-3 localizes to cyto-membrane and is required for B16F1 melanoma cell metastasis. *PLoS One* **4**, e4450 (2009).

Srichai, M. B. & Zent, R. Integrin structure and function. *Cell-extracellular matrix interactions in cancer* (editors: Zent, R. & Pozzi, A.) doi:10.1007/978-1-4419-0814-8 (2010).

Stephens, B. J., Han, H., Gokhale, V. & Von Hoff, D. D. PRL phosphatases as potential molecular targets in cancer. *Mol. Cancer Ther.* **4**, 1653-1661 (2005).

Strømgaard, A., Jensen, A.A. & Strømgaard, K. Site-specific incorporation of unnatural amino acids into proteins. *ChemBioChem* **5**, 909-916 (2004).

Subauste, M. C., Nalbant, P., Adamson, E. D. & Hahn, K. M. Vinculin controls PTEN protein level by maintaining the interaction of the adherens junction protein β-catenin with the scaffolding protein MAGI-2. *J. Biol. Chem.* **280**, 5676-5681 (2005).

Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nature Methods* **2**, 261-268 (2005).

Sun, J.-P., Wang, W.-Q. *et al.* Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation and tumor invasion. *Biochemistry* **44**, 12009-12021 (2005).

Sun, J.-P., Luo, Y. *et al.* Phosphatase activity, trimerisation, and the C-terminal polybasic region are all required for PRL1-mediated cell growth and migration. *J. Biol. Chem.* **282**, 29043-29051 (2007).

Sundaresan, M., Yu, Z.-X., Ferrans, V.J., Irani, K. & Finkel, T. Requirement for generation of H_2O_2 for platelet-derived growth factor signal transduction. *Science* **270**, 296-299 (1995).

Takida, S. & Wedegaertner, P. B. Heterotrimer formation, together with isoprenylation, is required for plasma membrane targeting of Gβγ. *J. Biol. Chem.* **278**, 17284-17290 (2003).

Tang, D. *et al.* ERK activation mediates cell cycle arrest and apoptosis after DNA damage independently of p53. *J. Biol. Chem.* **277**, 12710-12717 (2002).

Tanaka, Y., Bond, M.R. & Köhler, J.J. Photocrosslinkers illuminate interactions in living cells. *Mol. Biosyst.* **4**, 473-480 (2008).

Theodosiou, A. & Ashworth, A. MAP kinase phosphatases. *Genome Biol.* **3**, reviews3009.1-3009.10 (2002).

Thibodeaux, G. N. *et al.* Transforming a pair of orthogonal tRNA-aminoacyl-tRNA synthetase from *Archae* to function in mammalian cells. *PLoS One* **5**, e11263 (2010).

Thiery, J. P. & Sleeman, J. P. Complex networks orchestrate epithelial-mesenchymal transitions. *Nat. Rev. Mol. Cell Biol.* **7**, 131-142 (2006).

Tian, W. *et al.* Phosphatase of regenerating liver-3 directly interacts with integrin β1 and regulates its phosphorylation at tyrosine 783. *BMC Biochem.* **13**, 22 (2012).

Tiganis, T. & Bennett, A.M. Protein tyrosine phosphatase function: the substrate perspective. *Biochem. J.* **402**, 1-15 (2007).

Tippmann, E. M., Liu, W., Summerer, D., Mack, A. V & Schultz, P. G. A genetically encoded diazirine photocrosslinker in *Escherichia coli*. *ChemBioChem* **8**, 2210-2214 (2007).

Todd, J.L., Tanner, K.G. & Denu, J.M. Extracellular signal regulated kinases (ERK) 1 and ERK2 are authentic substrates for the dual-specificity protein-tyrosine phosphatase VHR. *J. Biol. Chem.* **274**, 13271-13280 (1999).

Todd, J.L., Rigas, J.D., Rafty, L.A. & Denu, J.M. Dual-specificity protein tyrosine phosphatase VHR down-regulates c-Jun N-terminal kinase (JNK). *Oncogene* **21**, 2573-2583 (2002).

Tonks, N.K. Redox redux: revisiting PTPs and the control of cell signaling. *Cell* **121**, 667-670 (2005).

Tonks, N.K. Protein tyrosine phosphatases: from genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* **7**, 833-846 (2006).

Tonks, N.K. Protein tyrosine phosphatases-from housekeeping enzymes to master regulators of signal transduction. *FEBS J.* **280**, 346-378 (2013).

Trakselis, M. A., Alley, S. C. & Ishmael, F. T. Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. *Bioconjug. Chem.* **16**, 741-750 (2005).

Traore, K. *et al.* Redox-regulation of Erk1/2-directed phosphatase by reactive oxygen species: role in signaling TPA-induced growth arrest in ML-1 cells. *J. Cell. Physiol.* **216**, 276-285 (2008).

Uezu, A. *et al.* Modified SH2 domain to phototrap and identify phosphotyrosine proteins from subcellular sites within cells. *Proc. Natl. Acad. Sci. USA* **109**, E2929-E2938 (2012).

Usui, T. *et al.* Design and synthesis of a dimeric derivative of RK-682 with increased inhibitory activity against VHR, a dual-specificity ERK phosphatase: implications for the molecular mechanism of the inhibition. *Chem. Biol.* **8**, 1209-1220 (2001).

van Huijsduijnen, R. H., Bombrun, A. & Swinnen, D. Selecting protein tyrosine phosphatases as drug targets. *Drug Discov. Today* **7**, 1013-1019 (2002).

Varner, J. A & Cheresh, D. A. Integrins and cancer. *Curr. Opin. Cell Biol.* **8**, 724-730 (1996).

Vasilescu, J., Guo, X. & Kast, J. Identification of protein-protein interactions using *in vivo* cross-linking and mass spectrometry. *Proteomics* **4**, 3845-3854 (2004).

Völkert, M. *et al.* Synthesis and biological activity of photoactivatable N-Ras peptides and proteins. *J. Am. Chem. Soc.* **125**, 12749-12758 (2003).

Wagner, E.F. & Nebreda, A.R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat Rev Cancer* **9**, 537-549 (2009).

Wakefield, L. M. & Roberts, A. B. TGF-beta signaling: positive and negative effects on tumorigenesis. *Curr. Opin. Genet. Dev.* **12**, 22-29 (2002).

Wang, L, Magliery, T.J., Liu, D.R. & Schultz, P.G. A new functional supressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.* **122**, 5010-5011 (2000).

Wang, L., Brock, A., Herberich, B. & Schultz, P.G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498-500 (2001).

Wang, J., Kirby, C. E. & Herbst, R. The tyrosine phosphatase PRL-1 localizes to the endoplasmic reticulum and the mitotic spindle and is required for normal mitosis. *J. Biol. Chem.* **277**, 46659-46668 (2002).

Wang, L. & Proud, C. G. Regulation of the phosphorylation of elongation factor 2 by MEK-dependent signalling in adult rat cardiomyocytes. *FEBS Lett.* **531**, 285-289 (2002).

Wang, L. & Schultz, P.G. Expanding the genetic code. *Chem. Commun.* **1**, 1-11 (2002) http://dx.doi.org./10.10339/B108185N.

Wang, H. *et al.* PRL-3 down-regulates PTEN expression and signals through PI3K to promote epithelial-mesenchymal transition. *Cancer Res.* **67**, 2922-2926 (2007a).

Wang, W. *et al.* Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat. Neurosci.* **10**, 1063-1072 (2007b).

Wang, Q., Parrish, A.R. & Wang, L. Expanding the genetic code for biological studies. *Chem. Biol.* **16**, 323-336 (2009).

Wang, H. *et al.* PCBP1 suppresses the translation of metastasis-associated PRL-3 phosphatase. *Cancer Cell* **18**, 52-62 (2010a).

Wang, F., Robbins, S., Guo, J., Shen, W. & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in *Mycobacterium tuberculosis*. *PLoS One* **5**, e9354 (2010b).

Wang, J.-Y. *et al.* Vaccinia H1-related phosphatase is a phosphatase of ErbB receptors and is down-regulated in non-small cell lung cancer. *J. Biol. Chem.* **286**, 10177-10184 (2011).

Wang, Y. & Lazo, J. Metastasis-associated phosphatase PRL-2 regulates tumor cell migration and invasion. *Oncogene* **31**, 818-827 (2012).

Wei, L., Yang, Y., Zhang, X. & Yu, Q. Cleavage of p130Cas in anoikis. *J. Cell. Biochem.* **91**, 325-335 (2004).

Wei, C.H. *et al.* Crystal structure of a novel mitogen-activated protein kinase phosphatase, SKRP1. *Proteins* **79**, 3242-3246 (2011).

Wentworth, C.C., Alam, A., Jones, R.M., Nusrat, A. & Neish, A.S. Enteric commensal bacteria induce extracellular signal-regulated kinase pathway signaling via formyl peptide receptor-dependent redox modulation of dual specific phosphatase 3. *J. Biol. Chem.* **286**, 38448-38455 (2011).

Werner, S.R. *et al.* Enhanced cell cycle progression and down regulation of p21 Cip1/Waf1 by PRL tyrosine phosphatase. *Cancer Lett.* **202**, 201-211 (2003).

Wishart, M.J., Denu, J.M., Williams, J.A. & Dixon, J.E. A single mutation converts a novel phosphotyrosine binding domain into a dual-specificity phosphatase. *J. Biol. Chem.* **270**, 26782-26785 (1995).

Wittelsberger, A. *et al.* The mid-region of parathyroid hormone (1-34) serves as a functional docking domain in receptor activation. *Biochemistry* **45,** 2027-2034 (2006a).

Wittelsberger, A., Thomas, B.E., Mierke, D.F. & Rosenblatt, M. Methionine acts as a “magnet” in photoaffinity crosslinking experiments. *FEBS Lett.* **580**, 1872-1876 (2006b).

Wittelsberger, A., Mierke, D.F. & Rosenblatt, M. Mapping ligand-receptor interfaces: approaching the resolution limit of benzophenone-based photoaffinity scanning. *Chem. Biol. Drug Des.* **71**, 380-383 (2008).

Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A Genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126**, 14306-14307 (2004).

Wu, S. *et al.* Multidentate small-molecule inhibitors of vaccinia H1-related (VHR) phosphatase decrease proliferation of cervix cancer cells. *J. Med. Chem.* **52**, 6716-6723 (2009).

Xiao, Y. *et al.* PTEN catalysis of phospholipid dephosphorylation reaction follows a two-step mechanism in which the conserved aspartate-92 does not function as the general acid- mechanistic analysis of a familial Cowden disease-associated PTEN mutation. *Cell. Signal.* **19**, 1434-1445 (2007).

Xie, J. & Schultz, P.G. Adding amino acids to the genetic repertoire. *Curr. Opin. Chem. Biol.* **9**, 548-554 (2005).

Xie, J. & Schultz, P.G. A chemical toolkit for proteins-an expanded genetic code. *Nature Rev. Mol.Cell Biol.* **7**, 775-782 (2006).

Xu, J. *et al.* VEGF promotes the transcription of the human PRL-3 gene in HUVEC through transcription factor MEF2C. *PLoS One* **6**, e27165 (2011).

Yaffe, M.B. & Cantley, L.C. Grabbing phosphoproteins. **402**, 30-31 (1999).

Yano, H. *et al.* Roles played by a subset of integrin signaling molecules in cadherin-based cell-cell adhesion. *J. Cell Biol.* **166**, 283-295 (2004).

Ye, S., Köhrer, C. *et al.* Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis. *J. Biol. Chem.* **283**, 1525-1533 (2008).

Yu, L. *et al.* Oxidative stress-induced expression and modulation of Phosphatase of regenerating liver-1 (PRL-1) in mammalian retina. *Biochim. Biophys. Acta* **1773**, 1473-1482 (2007).

Yuvaniyama, J., Denu, J.M., Dixon, J.E. & Saper, M.A. Crystal structure of the dual specificity protein phosphatase VHR. *Science* **272**, 1328-1331 (1996).

Zeng, Q., Hong, W. & Tan, Y. H. Mouse PRL-2 and PRL-3, two potentially prenylated protein tyrosine phosphatases homologous to PRL-1. *Biochem. Biophys. Res. Commun.* **244**, 421-427 (1998).

Zeng, Q. *et al.* Prenylation-dependent association of protein-tyrosine phosphatases PRL-1, -2, and -3 with the plasma membrane and the early endosome. *J. Biol. Chem.* **275,** 21444-21452 (2000).

Zeng, Q. *et al.* PRL-3 and PRL-1 promote cell migration, invasion, and metastasis. *Cancer Res.* **63**, 2716-2722 (2003).

Zhang, F.L. & Casey, P.J. Protein prenylation: molecular mechanisms and functional consequences. *Annu. Rev. Biochem.* **65**, 241-269 (1996).

Zhang, Z.-Y. Protein-tyrosine phosphatases: biological function, structural characteristics, and mechanism of catalysis. *Crit. Rev. Biochem. Mol. Biol.* **33**, 1-52 (1998).

Zhang, W. & Liu, H.T. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res* **12**, 9-18 (2002).

Zhang, Z.-Y. Mechanistic studies on protein tyrosine phosphatases. *Prog. Nucleic Acid Res. Mol. Biol.* **73**, 171-220 (2003a).

Zhang, Z.-Y. Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Acc.Chem. Res.* **36**, 385-392 (2003b).

Zhang, Z.-Y. Functional studies of protein tyrosine phosphatases with chemical approaches. *Biochim. Biophys. Acta* **1754**, 100-107 (2005).

Zhang, Y.-Y., Wu, J.-W. & Wang, Z.-X. Mitogen-activated protein kinase (MAPK) phosphatase 3-mediated cross-talk between MAPKs ERK2 and p38α *J. Biol. Chem.* **286**, 16150-16162 (2011).

Zhao, W.-B., Li, Y., Liu, X., Zhang, L.-Y. & Wang, X. Evaluation of PRL-3 expression, and its correlation with angiogenesis and invasion in hepatocellular carcinoma. *Int. J. Mol. Med.* **22**, 187-192 (2008).

Zheng, P. *et al.* Stathmin, a new target of PRL-3 identified by proteomic methods, plays a key role in progression and metastasis of colorectal cancer. *J. Proteome Res.* **9**, 4897-4905 (2010).

Zheng, P. *et al.* Snail as a key regulator of PRL-3 gene in colorectal cancer. *Cancer Biol. Ther.* **12**, 742-749 (2011).

Zheng, Y., Lv, X. & Wang, J. A genetically encoded sulfotyrosine for VHR function research. *Protein Cell* **4**, 4-7 (2013).

Zhou, G., Denu, J.M., Wu, L. & Dixon, J.E. The catalytic role of Cys124 in the dual specificity phosphatase VHR. *J. Biol. Chem.* **269**, 28084-28090 (1994).

Zhou, B., Wang, Z.-X., Zhao, Y., Brautigan, D.L. & Zhang, Z.-Y. The specificity of extracellular signal-regulated kinase 2 dephosphorylation by protein phosphatases. *J. Biol. Chem.* **277**, 31818-31825 (2002).

Zhou, J., Wang, S., Lu, J., Li, J. & Ding, Y. Over-expression of phosphatase of regenerating liver-3 correlates with tumor progression and poor prognosis in nasopharyngeal carcinoma. *Int. J. Cancer* **124**, 1879-1886 (2009).

Zhou, J. *et al.* PRL-3, a metastasis associated tyrosine phosphatase, is involved in FLT3-ITD signaling and implicated in anti-AML therapy. *PLoS One* **6**, e19798 (2011).

Zhu, L., Crothers, J., Zhou, R. & Forte, J. G. A possible mechanism for ezrin to establish epithelial cell polarity. **299**, C431-C443 (2010).

Zimmerman, M.W., Homanics, G.E. & Lazo, J.S. Targeted deletion of the metastasis-associated phosphatase Ptp4a3 (PRL-3) suppresses murine colon cancer. *PLoS One* **8**, e58300 (2013).