



UNIVERSITAT DE
BARCELONA

Structural studies on *Porphyromonas gingivalis* proteins

Claus-A. Schmitz

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

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

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

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

Doctoral Thesis
DOCTORAL PROGRAM
HDK06 Biotechnologia

Structural studies on
Porphyromonas gingivalis
proteins

submitted by
Claus-A. Schmitz
from Bonn-Bad Godesberg
Germany

Barcelona 2022

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

Doctoral Thesis
DOCTORAL PROGRAM
HDK06 Biotechnologia

Structural studies on
Porphyromonas gingivalis
proteins

submitted by
Claus-A. Schmitz
from Bonn-Bad Godesberg
Germany

Barcelona 2022

This thesis was written in the research group and under supervision of Maria Solà - Structural Biology of Proteins Essential for Mitochondria and Pathogen Viability - Structural Mitolab, Structural Biology (SB) 'María de Maeztu' Unit of Excellence, Molecular Biology Institute of Barcelona, Barcelona, SPAIN.

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

PROGRAMA DE DOCTORAT
HDK06 Biotechnologia

TÍTOL DE LA TESI

Structural studies on
Porphyromonas gingivalis
proteins

Memòria presentada per Claus-A. Schmitz
per optar al títol de doctor per la universitat de Barcelona

Nom, cognoms i signatura del Director/s

Maria Solà Vilarrubias

Nom, cognoms i signatura del doctorand/a

Claus-A. Schmitz

Nom, cognoms i signatura del tutor/a

Josefa Badia Palacin

Claus-A. Schmitz, 2022.

Acknowledgments

Beginning I would like to thank my parents for their love and support throughout my academic journey, who allowed me to study, and for all they had to give up for themselves during the process, a support they did not have during my age.

I am also deeply grateful to the members of my PhD defense committee, Ignacio Fita, Roeland Boer, Alberto Marina Moreno, Xavier Gomis-Rüth and Joan Pous, who kindly agreed to reading and evaluation of my thesis.

I am grateful to the Ministry of Economy and Competitiveness (MINECO) and the Spanish National Research Council (CSIC) for their financial support, which allowed me to travel to Spain and work on this project.

I thank the staff and local contacts at the European Synchrotron Radiation Facility (ESRF) and ALBA synchrotron of Barcelona for their resources and assistance.

I owe the people and staff of the Institute of Molecular Biology of Barcelona (IBMB) for providing the resources, support and expertise I got throughout the project. Especially I want to thank Maria Solà who brought me to Barcelona, believing and trusting in me as a researcher, guiding and supporting me through all steps of this thesis. I am grateful for her feedback and that she always had time for my questions, organized seminars and meetings and the great teamwork we had in writing our paper. I want to thank Leonor Calopa for her outstanding help and assistance with all kinds of bureaucratic problems and procedures, be it with government institutions or internal at the IBMB. I want to thank Joan Pous, Ignacio Fita, Isabel Usón and Xavier Gomis-Rüth in sharing their expertise in protein X-ray crystallography. In particular I want to thank Isabel Usón for her help in solving the phase problem by SHELX_C/D/E and the sharing of computational resources.

I want to thank all my colleagues, not only from CRI7 but also CRI3, CRI4 and CRI6 for the mutual support of each other. Especially I want to mention Sébastien Lyonnais and Theodoros Goulas for all help and discussions we had about biochemistry, protein expression and purification, as well as Nicolas Soler for all kind of help and discussion related to bioinformatics.

From my group I want to thank Anna Cuppari, Arka Chakraborty, Arkadiusz Szura, Elena Ruiz, Nord Oliver Turro and Ramiro Illanes Vicioso for their help in the lab and with paperwork, moving inside Barcelona and not least for their collegiality during the shared time during and outside of work. I especially have to thank Anna Cuppari for starting the PorX project in our lab and proving that PorX crystallizes and it is a worthwhile project to work on and to Arka Chakraborty for setting up the initial crystallization screenings.

I am grateful to our collaborators from the Małopolska Center of Biotechnology, (Jagiellonian University, Krakow, Poland), Mariusz Madej and Jan Potempa, for supplying us with protein

samples, the enzymatic characterization, uncountable chats, emails, feedback and discussions about PorX function and of course the great collaboration in writing the publication.

I owe the University of Barcelona (UB) and Faculty of Pharmacy and Food Science, especially Josefa Badia Palacin for her guidance and bureaucratic assistance, help with the university and keeping deadlines, as well as Gemma for her time, help and feedback for the doctorate school and thesis requirements.

Outside of work I want to thank my former and current flatmates Marielou and Maikom for all kinds of personal support especially during and after the times of my COVID-19 infections while I wrote the publication and this thesis.

While it is not possible to mention everyone by name, I would like to thank everyone else who allowed, supported and assisted to my research and to my life during the years of my PhD.

Index

NOTES

Table of Content

1. Introduction.....	1
1.1 Porphyromonas gingivalis.....	1
1.2 Oral microflora.....	1
1.3 Subgingival biofilm formation.....	4
1.4 PG influence on the oral commensal-host-community.....	6
1.5 PG virulence factors and their effects on the host immune system.....	6
1.5.1 LPS.....	12
1.5.2 Fimbriae.....	13
1.5.3 TLR2-CXCR4-C5aR crosstalk.....	13
1.5.4 PG suppression of the adaptive immune response.....	15
1.5.5 Concentration dependency of PG virulence factors.....	15
1.5.6 Gingipains as the main immune modulator of PG.....	16
1.6 Periodontal disease progression and polymicrobial synergy.....	17
1.7 Bacterial secretion systems	18
1.8 Type IX secretion systems.....	19
1.8.1 T9SS transport	21
1.8.2 PG T9SS Regulation	23
1.9 Aims of this Project.....	24
2. Materials and Methods.....	1
2.1 Chemicals and Reagents.....	1
2.1.1 Chromatography columns and materials.....	1
2.1.2 Buffers and Solutions.....	1
2.1.3 Media.....	2
2.1.4 Plasmids.....	2
2.1.5 Primer.....	2
2.1.6 Bacterial strains.....	3
2.1.7 Informatics.....	3
2.1.8 Commercial Kits and tools.....	4
2.1.9 Equipment.....	4
2.1.10 Protein Samples	4
2.1.11 Crystallization Screens.....	5
2.2 Molecular biology.....	10
2.2.1 Polymerase Chain Reaction (PCR).....	11
2.2.2 DNA purification and analysis	11
2.2.3 Restriction digestion and ligations.....	12
2.2.4 Site-directed mutagenesis	12
2.2.5 DNA sequencing	12
2.3 Microbiological methods	13
2.3.1 Heat shock transformation	13
2.3.2 Recombinant protein expression	13
2.4 Protein Purification	14
2.4.1 Affinity chromatography.....	14
2.4.2 Ion-exchange Chromatography (IEC)	15
2.4.3 Size exclusion chromatography (SEC).....	15
2.5 Protein sample preparations	16
2.5.1 In vitro phosphorylation	16

2.5.2 Beryllium fluoride induced PorX dimerization	16
2.5.3 Ultrafiltration.....	17
2.6 Protein analytics.....	17
2.6.1 SDS-PAGE.....	18
2.6.2 Spectrophotometry	18
2.6.3 SEC-Multi-Angel-Laser-Light-Scattering (MALLS).....	19
2.7 Crystallization	20
2.7.1 Crystallization strategy and screenings.....	20
2.7.2 Crystal manipulation, imaging, soaking and cryo cooling	21
2.7.3 Crystallization theory.....	21
2.8 X-ray diffraction	24
2.8.1 Experimental setup	25
2.8.2 Diffraction principles.....	27
2.8.3 Unit cells and asymmetric unit.....	30
2.8.4 Indexing.....	32
2.8.5 Space group determination.....	33
2.8.6 Friedel pairs.....	34
2.8.7 Data reduction and quality control	37
2.9 Phasing	38
2.9.1 The hand problem and phase improvement.....	40
2.9.2 Molecular Replacement (MR) and MR-SAD	41
2.9.3 Model building and refinement.....	42
2.10 Small-angle X-Ray scattering (SAXS)	43
2.10.1 SAXS-Theoretical background	44
2.10.2 Kratky plot & Porod equation.....	46
2.10.3 Guinier analysis.....	47
2.10.4 Molecular weight estimations.....	47
2.10.5 Distance distribution function $p(r)$	49
2.10.6 Structural reconstruction and validation.....	50
2.10.7 DENSS Electron density reconstruction.....	51
2.10.8 PglZ ortholog comparisons with BREX-proteins	52
3. Results	1
3.1 Preceding project status	1
3.2 Initial screenings	2
3.2.1 Scoring analysis.....	2
3.2.2 Structure determination by molecular replacement.....	2
3.2.3 Reproduction trials of initial crystals	4
3.3 Crystallization buffer optimizations.....	5
3.4 Crystallization strategy optimization.....	7
3.4.1 Semi-high throughput, complete datasets better than 3 Å ("JG5" and "SCN" data).....	8
3.4.2 Upscaling, sufficient anomalous scattering ("HB6" data).....	9
3.4.3 Soakings and co-crystallizations, "in situ dimerization", Opt_J, -K, -L, -M, K+/K++	10
3.5 PorX mutants crystallization.....	13
3.6 Crystallization trials of PorY, PG1660 and monomeric PorX	14
3.7 Structural results	14
3.7.1 Structural solution of "HB6" crystal	14
3.7.2 Full length PorX "JG5" structure	15
3.7.3 PorX receiver domain (RD).....	16

3.7.4 Helical bundle domain (HBD).....	17
3.7.5 PglZ domain	17
3.7.6 Substrate binding pocket.....	19
3.7.7 Dimerization interfaces.....	19
3.8 High resolution structure "HR"	21
3.9 "pGpG" structure, T272A mutant in complex with pGpG.....	25
3.10 Other crystallization results and structures	28
3.10.1 Native structures.....	28
3.10.2 "SCN" structure.....	28
3.10.3 GMP cocrystallization	29
3.10.4 GTPγS cocrystallization.....	29
3.10.5 PorX M94K/D104A/I129A mutant.....	29
3.10.6 PorX D58A mutant	29
3.10.7 PorX T272A mutant	30
3.10.8 T272A+c-di-GMP cocrystallization.....	30
3.11 Mutational studies.....	31
3.12 PorX dimerization analysis by SEC and SEC-MALLS.....	31
3.13 Small Angle X-ray scattering	33
3.13.1 Oligomeric state and data quality analysis	33
3.13.2 Electron density reconstructions by DENSS.....	37
3.13.3 SAXS modeling.....	39
3.13.4 SAXS model of PorX dimer.....	40
3.13.5 SAXS model of PorX monomer.....	41
3.14 Comparative homology.....	43
3.14.1 DALI analysis & homology search in the PDB.....	43
3.14.2 Substrate binding pocket comparisons.....	47
3.14.3 Structural comparisons to BREX PglZ orthologs.....	50
3.15 Enzymatic assays.....	55
3.15.1 Phosphorylation assays.....	55
3.15.2 Substrate cleavage assays.....	56
3.15.3 Ultrahigh performance liquid chromatography mass spectrometry (UHPLC-MS).....	57
4. Discussion.....	
4.1 Crystallization.....	1
4.1.1 Initial screenings.....	1
4.1.2 Buffer optimizations.....	2
4.1.3 Strategy optimizations.....	3
4.1.4 PorY, PG1660 and PorX monomer	4
4.2 PorX structural features.....	4
4.2.1 Receiver domain RD.....	4
4.2.2 Helical bundle domain (HBD).....	6
4.2.3 PglZ domain	6
4.3 Activation by interdomain crosstalk.....	9
4.4 Reaction pathway	11
4.5 Substrate specificity	11
4.6 Enzymatic function	13
4.7 Transcription regulation	14
4.8 Trouble shooting	17
4.8.1 Function of LH conformation in PorX.....	17
4.8.2 Enzymatic activities of monomeric PorX samples in enzymatic assays.....	17
4.8.3 SAXS sample concentrations and interparticle interactions.....	18

5. Prospects		
5.1 Crystallization.....		1
5.2 Comparative homology		1
5.3 Enzymatic activity and transcriptional regulation.....		2
5.4 PorX as a drug target.....		2
6. Conclusions		1
7. .Supplementary		
7.1 Literature Introduction.....		1
7.2 Literature Materials&Methods		8
7.3 Literature Results, Discussion, Prospects.....		12
7.4 Data, model and refinement statistics		14
7.5 PorX-structural subdivision.....		16
7.6 PorX-dimeric surface.....		17
7.7 SEC/SEC-MALLS of PorX and its mutants.....		18
7.8 Comparative homology		21
7.9 Enzymatic assays		27
7.10 Optimization Screens		30

Index of Figures

1.Introduction		
Figure 1.1: Sources of periodontal disease,		3
Figure 1.2: Successive microbiological biofilm formation within the oral cavity.....		5
Figure 1.3: Schematic representation of TLR2, TLR3 and TLR4 signaling		11
Figure 1.4: Inflammatory signaling of gingival tissue induced by LPS.....		12
Figure 1.5: TLR2-CXCR4-C5aR crosstalk.....		14
Figure 1.6: Transport, modification and regulatory components of the T9SS.....		22
2..Materials and Methods		
Figure 2.1: SEC-MALLS setup		19
Figure 2.2: Macromolecular crystallization phase diagram		22
Figure 2.3: General X-Ray diffraction setup		26
Figure 2.4: Geometrical explanation for Bragg's law.		28
Figure 2.5: Ewald's Sphere.....		29
Figure 2.6: Unit cells and asymmetric unit within a crystal lattice		30
Figure 2.7: Unit cell attributes		30
Figure 2.8: Miller plane intersection with the unit cell in 2 (A) and 3D (B)		32
Figure 2.9: Vector “Harker-” diagram of structure factor summation for inverse reflections ((h, k, l) and (-h, -k, -l)) and corresponding structure factor equations.....		36
Figure 2.10: Harker diagram for phasing by isomorphous replacement (A) or anomalous dispersion (B)		40
Figure 2.11: SAXS experimental setup		45
Figure 2.12: Kratky plot		46
Figure 2.13: Guinier plot		47
Figure 2.14: Distance distribution function of different molecule shapes		49
Figure 2.15: Schematic iterative electron density reconstruction steps used by DENSS.....		51

3. Results	
Figure 3.1: PorX dimer overall structure.....	15
Figure 3.2: Architecture of the receiver domain (RD) of PorX.	16
Figure 3.3: Architecture of the alkaline phosphatase-like PglZ domain of PorX and its two different conformations.	18
Figure 3.4: PorX-binding pocket (HH conformation).	19
Figure 3.5: Relative position of PorX receiver domains (RD) within the PorX dimer.....	20
Figure 3.6: PorX interaction areas of receiver domains (RD) and helical bundle domains (HBD). .	20
Figure 3.7: PorX interaction areas of PglZ domains in the PorX LH:HH dimer.	21
Figure 3.8: PorX phosphorylation sites bound to BeF3.	22
Figure 3.9: L α 3b- α 3c alternative conformations.....	23
Figure 3.10: Anomalous diffraction electron density maps for HR structure.....	24
Figure 3.11: PorX T272A-pGpG active site.....	26
Figure 3.12: PorX T272A-pGpG active site electron density.....	27
Figure 3.13: PorX SCN assymmetric unit	28
Figure 3.14: Active site GTP γ S cocrystallization (without ligand bound).....	29
Figure 3.15: D58A structure	30
Figure 3.16: SEC-MALLS elution profiles.....	32
Figure 3.17: SAXS data/fit for monomeric PorX.....	35
Figure 3.18: SAXS data/fit for dimeric PorX with BeF3 without zinc.....	35
Figure 3.19: SAXS data/fit for dimeric PorX with BeF3 and zinc.....	36
Figure 3.20: SAXS data/fit for monomeric PorX at low cententration without BeF3 but with zinc. 36	
Figure 3.21: SAXS data/fit for dimeric PorX at high cententration without BeF3 but with zinc.....	37
Figure 3.22: PorX SAXS-Dimer model. Superposition of crystal structure and DENSS electron density reconstruction.	41
Figure 3.23: PorX SAXS-Monomer model, crystal structure superposition and DENSS electron density reconstruction.	43
Figure 3.24: PorX active site and dimerization interface conservation throughout the alkaline phosphatase superfamily (APS).	44
Figure 3.25: PorX shares the APS two metal active site.	45
Figure 3.26: PorX_pGpG substrate binding pocket.	47
Figure 3.27: Substrate binding site of PorX.....	49
Figure 3.28: BREX1-6 primary sequence alignment.	53
Figure 3.29: Structural alignments of BREX1-6 models.	54
7. .Supplementary	
Figure 7.1: PorX-Domain and secondary structure border definitions.....	16
Figure 7.2: PorX-dimeric surfaces	17
Figure 7.3: Dimerization of PorX and its mutants after in vitro phosphorylation by AcP.....	18
Figure 7.4: PorX dimerization induced by zinc	19
Figure 7.5: PorX dimer stability test	19
Figure 7.6: PorX homologs (DALI) active site coordination (lower panels) and substrate binding pockets (upper panels).....	26
Figure 7.7: Phosphorylation assays of PorX and mutants	27
Figure 7.8: Phosphodiesterase assays	28
Figure 7.9: Positive substrate identifications of PorX by UHPLC-MS	28
Figure 7.10: Negative substrate identifications of PorX by UHPLC-MS	29

Index of Tables

1.Introduction.....	
TABLE 1.1: Overview on the most studied PG virulence factors.	10
TABLE 1.2: Essential proteins, complexes and their functions in the T9SS of PG.	21
2..Materials and Methods.....	
TABLE 2.1: Data collection setups.....	24
TABLE 2.2: Crystal systems	31
TABLE 2.3: The 14 Bravais lattices	31
TABLE 2.4: The 65 Sohncke space groups in protein crystallography	34
3. Results	
TABLE 3.1: Best conditions of initial crystallization screening in 100 nl drop sizes.	3
TABLE 3.2: Best Grid01_PEG_3350 crystals.....	4
TABLE 3.3: Best Opt_A crystals.....	6
TABLE 3.4: Best Opt_B crystals.....	6
TABLE 3.5: Best Opt_D crystals.....	6
TABLE 3.6: Best Opt_E crystals.....	7
TABLE 3.7: Best Opt_F crystals.....	7
TABLE 3.8: Best Opt_J crystals.....	8
TABLE 3.9: Best JG5-Optisalt-crystals.....	9
TABLE 3.10: Largest crystals (upper images) of PorX upscaling trials and their broken/ diffracted fragments (lower panel). (BeF3_48Ops_01, -02, -03, Opt_H and Opt_I)	10
TABLE 3.11: Crystals of dataset from co-crystallization and soaking experiments.	12
TABLE 3.12: Best crystals of PorX mutants with or without ligands.....	13
TABLE 3.13: PorX monomer crystallization.	14
TABLE 3.14: PorX-Mutants.....	31
TABLE 3.15: PorX-SAXS data analysis	34
TABLE 3.16: PorX-DENSS electron density reconstructions.....	38
7. .Supplementary.....	
TABLE 7.1: published structures	14
TABLE 7.2: supporting structures.....	15
TABLE 7.3: SEC-MALLS inicial quality assesment. As published[0].....	20
TABLE 7.4: TOP20 homolog proteins to PorX PglZ domain. Modified from publication[0]	21

List of Abbreviations

AC	adenylate cyclase	PgIZ	phage growth limitation factor Z
AcP	acetyl-phosphate	PI3K	Phosphoinositide-3-kinase
APS	alkaline phosphatase superfamily; except Introduction anionic- lipopolysaccharid	PKA	protein kinase A
ASU	asymmetric unit	PRR	pattern recognition receptor
BREX	Bacteriophage-Exclusion	RD	receiver domain
C5aR	C5a receptor	RgpA / RgpB	arginine specific gingipains type A / B
CBASS	cyclic oligonucleotide-based anti- phage signaling systems	RhoA	Ras homolog family member A
CBP	CREB-binding protein	RR	resonse regulator
CBP	CREB-binding protein	SeMet	seleno-methionine
cOA	cyclic oligoadenylate	SHK	sensor histidine kinase
CREB	cAMP response element binding protein	SS	secretion systems
CSF(s)	colony stimulating factor(s)	T9SS	Type-9-secretion-system
CTD	C-terminal signaling domain	TCS	two component system
GCF	gingival crevicular fluid	TLR(s)	toll-like-receptor(s)
HBD	helical bundle domain	TNF	tumor necrosis factor
IL	interleukin		
IM	inner membrane		
INF	interferon		
Kgp	lysine specific gingipains		
LPS	lipopolysaccharides		
Mw/MWs	molecular weight/-s		
NFκB	nuclear factor kappa light chain enhancer of activated B-cells		
OM	outer membrane		
OMV(s)	outer membrane vesicle(s)		
PAMP	pathogen-associated molecular pattern		
PAR	protease activated receptors		
PDB	Protein Data Bank		
PG	Porphyromonas gingivalis		

Abstract

In this work we characterize the structure function relationship of PorX, an essential protein for the regulation of transcription of the type 9 secretion system (T9SS) of the human pathogen *Porphyromonas gingivalis* (PG), responsible for its virulence.^[1]

PG is a member of the oral biofilm community and acts as the keystone pathogen for periodontal diseases.^[2] Even though its main habitat is the oral cavity it is also able to infect the inner organs such as brain, liver, coronary artery.^{[3], [4], [5]} PG has been associated to a series of human systemic diseases such as cancer^[6], atherosclerosis^{[7], [8]} or cardiovascular disease^{[9] [10]} Also it rises the mortality during viral respiratory infections.^[11] As those diseases affect the fast majority of the human population and PG's virulence depends on PorX, it makes a valuable target for scientific investigation as a potential drug target to improve human health and life span.

We solved the crystal structure of dimeric PorX by X-ray crystallography, with and without substrates and characterized its domains, active site and substrate binding pocket. By small angle X-ray scattering we characterized PorX in solution, monomeric and dimeric, and we investigated the mechanism of dimerization and thus PorX activation by structure based mutational studies, phosphorylation assays, SEC and SEC-MALLS.

The enzymatic active effector domain of PorX belongs to the PglZ family, which is widespread among bacteria^[12], and was previously uncharacterized in structure and function. Therefore we compared it to its closest known structures and enzyme classes, as well as the models of other known PglZ subfamily homologs among BREX proteins. We found out that PglZs are phosphodiesterases of the alkaline phosphatase superfamily. This motive folds upon dimerization and thus activates the protein, while in the monomer, in solution, it is disordered.

By enzymatic assays we determined PorX specific hydrolytic activity against linear and cyclic oligonucleotides as well as linear dinucleotides, while we did not found any activity against cyclic mono- or dinucleotides, mono- or dinucleotide polyphosphates. Therefore we have shown that PorX is likely involved in oligonucleotide signaling in PG and, by PorX inhibition based on our structures, it might be possible to render PG avirulent by designing specific PorX inhibitors binding to its active site.

Literature Abstract

- 1: Veith, Paul D.; Glew, Michelle D.; Gorasia, Dhana G.; Reynolds, Eric C., *Type IX secretion*, Molecular microbiology 106, 1, p35–53, 2017.
- 2: Hajishengallis, G.; Lamont, R. J., *Beyond the red complex and into more complexity*, Molecular oral microbiology 27, 6, p409–419, 2012.
- 3: Olsen, Ingar; Yilmaz, Özlem, *Modulation of inflammasome activity by Porphyromonas gingivalis in periodontitis and associated systemic diseases*, Journal of oral microbiology 8, p30385, 2016.
- 4: Huck, Olivier; You, Jian; Han, Xianxian; Cai, Bin; Panek, James; Amar, Salomon, *Reduction of Articular and Systemic Inflammation by Kava-241 in a Porphyromonas gingivalis-Induced Arthritis Murine Model*, Infection and immunity 86, 9, 2018.
- 5: Jia, Lu; Han, Nannan; Du, Juan; Guo, Lijia; Luo, Zhenhua; Liu, Yi, *Pathogenesis of Important Virulence Factors of Porphyromonas gingivalis via Toll-Like Receptors*, Frontiers in cellular and infection microbiology 9, p262, 2019.
- 6: Gholizadeh, Pourya; Eslami, Hosein; Yousefi, Mehdi; Asgharzadeh, Mohammad; Aghazadeh, Mohammad; Kafil, Hossein Samadi, *Role of oral microbiome on oral cancers, a review*, Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 84, p552–558, 2016.
- 7: Reyes, Leticia; Herrera, David; Kozarov, Emil; Roldá, Silvia; Progulske-Fox, Ann, *Periodontal bacterial invasion and infection*, Journal of periodontology , 84(4 Suppl), pS30-50, 2013.
- 8: Huck, O; Saadi-Thiers, K; Tenenbaum, H; Davideau, JL; Romagna, C; Laurent, Y; Cottin, Y; Roul, JG., *Evaluating periodontal risk for patients at risk of or suffering from atherosclerosis*, Archives of cardiovascular diseases 104, 5, p352–358, 2011.
- 9: Chistiakov, Dimitry A.; Orekhov, Alexander N.; Bobryshev, Yuri V., *Links between atherosclerotic and periodontal disease*, Experimental and molecular pathology 100, 1, p220–235, 2016.
- 10: Fiorillo, Luca; Cervino, Gabriele; Laino, Luigi; D'Amico, Cesare; Mauceri, Rodolfo; Tozum, Tolga Fikret; Gaeta, Michele; Cicciù, Marco, *Porphyromonas gingivalis, Periodontal and Systemic Implications-A Systematic Review*, Dentistry journal 7, 4, 2019.
- 11: Bao, Lirong; Zhang, Cheng; Dong, Jiajia; Zhao, Lei; Li, Yan; Sun, Jianxun, *Oral Microbiome and SARS-CoV-2*, Frontiers in microbiology 11, p1840, 2020.
- 12: Goldfarb, T; Sberro, H; Weinstock, E; Cohen, O; Doron, Sy et al., *BREX is a novel phage resistance system widespread in microbial genomes*, The EMBO journal 34, 2, p169–183, 2015.

German Abstract

In dieser Doktorarbeit beschreiben wir die Struktur und Funktion von PorX, einem essentiellen Protein des Typ-9-Sekretionssystems von *Porphyromonas gingivalis* (PG).^[1] PG ist Teil des dentalen Biofilms und gilt als Hauptauslöser für Gingivitis und Parodontitis.^[2] PG befällt nicht nur orales Gewebe, sondern befällt auch u.a. Gehirn, Leber und Koronararterien.^{[3], [4], [5]} Es ist an einer Vielzahl von systemischen Erkrankungen beteiligt, darunter Krebs^[6] Arteriosklerose^{[7], [8]} und Herz-Kreislauf-Erkrankungen.^{[9] [10]} Da die Gesamtheit dieser Krankheiten die Mehrheit der menschlichen Bevölkerung betrifft, ist die Erforschung von PorX ein lohnendes Forschungsprojekt, nicht zuletzt, da das Ziel der Inhibierung ein lohnendes Folgeziel für Medikamentenentwicklung darstellt, welche Allgemeingesundheit und Lebensspanne erhöhen werden.

Durch Röntgenkristallographie haben wir die dimere Kristallstruktur von PorX aufgeklärt, sowohl ohne, als auch im substratgebundenen Zustand, in Komplex mit pGpG. Wir beschreiben die einzelnen PorX Domänen und ihr Zusammenwirken, das katalytische Zentrum sowie die Substratbindestellen. Durch Röntgenkleinwinkelstreuung haben wir die Struktur von PorX in Lösung analysiert, als Monomer sowie Dimer. Daraufhin und durch strukturbasierte Mutationsstudien haben wir den Mechanismus der Aktivierung und Dimerisierung von PorX erklärt und durch Phosphorylierungsanalysen, Größenausschlusschromatographie (SEC) und SEC-MALLS (*multi-angle-laser-light-scattering*) bestätigt. Die enzymatisch aktive C-terminale Domäne von PorX gehört zur PglZ-Familie, dessen Struktur und Funktion, trotz der relativen Häufigkeit in Prokaryonten,^[12] bisher unaufgeklärt war. Daher wurde in dieser Arbeit die Struktur und das katalytische Zentrum von PorX mit seinen nächsten Homologen verglichen, sowie mit bekannten PglZ Klassen der BREX Systeme. Als Ergebnis konnten wir PorX-PglZ als Phosphodiesterase der Superfamilie der alkalinen Phosphatasen (APS) identifizieren. Sowohl das aktivitätsdeterminierende katalytische Zentrum, wie auch das bisher unbeschriebene Dimerisierungsmotiv, beteiligt an enzymatischer Aktivierung, sind zwischen BREX-PglZs hoch konserviert. In PorX faltet sich dieses Motiv während des Dimerisierungsvorganges und ist entscheidend für die Metallkoordinierung des katalytischen Zentrums. Weiterhin konnten wir über enzymatische Analysen die Substratspezifität für PorX aufklären und zeigen, dass PorX spezifisch lineare und zyklische Oligonukleotide zu Nukleotidmonophosphaten hydrolysiert. Somit ist die Funktion von PorX an der Regulation von Oligonukleotidsignaltransduktion in PG beteiligt, und durch unsere Kristallstrukturen ist es möglich, dass, durch spezifisches Design von PorX-Inhibitoren, in Zukunft die Virulenz von PG verhindert werden kann.

Literature German Abstract

- 1: Veith, Paul D.; Glew, Michelle D.; Gorasia, Dhana G.; Reynolds, Eric C., *Type IX secretion*, 2017.
- 2: Hajishengallis, G.; Lamont, R. J., *Beyond the red complex and into more complexity*, 2012.
- 3: Olsen, Ingar; Yilmaz, Özlem, *Modulation of inflammasome activity by Porphyromonas gingivalis in periodontitis and associated systemic diseases*, 2016.
- 4: Huck, Olivier; You, Jian; Han, Xianxian; Cai, Bin; Panek, James; Amar, Salomon, *Reduction of Articular and Systemic Inflammation by Kava-241 in a Porphyromonas gingivalis-Induced Arthritis Murine Model*, 2018.
- 5: Jia, Lu; Han, Nannan; Du, Juan; Guo, Lijia; Luo, Zhenhua; Liu, Yi, *Pathogenesis of Important Virulence Factors of Porphyromonas gingivalis via Toll-Like Receptors*, 2019.
- 6: Gholizadeh, Pourya; Eslami, Hosein; Yousefi, Mehdi; Asgharzadeh, Mohammad; Aghazadeh, Mohammad; Kafil, Hossein Samadi, *Role of oral microbiome on oral cancers, a review*, 2016.
- 7: Reyes, Leticia; Herrera, David; Kozarov, Emil; Roldá, Silvia; Progulske-Fox, Ann, *Periodontal bacterial invasion and infection*, 2013.
- 8: Huck, O; Saadi-Thiers, K; Tenenbaum, H; Davideau, JL; Romagna, C; Laurent, Y; Cottin, Y; Roul, JG., *Evaluating periodontal risk for patients at risk of or suffering from atherosclerosis*, 2011.
- 9: Chistiakov, Dimitry A.; Orekhov, Alexander N.; Bobryshev, Yuri V., *Links between atherosclerotic and periodontal disease*, 2016.
- 10: Fiorillo, Luca; Cervino, Gabriele; Laino, Luigi; D'Amico, Cesare; Mauceri, Rodolfo; Tozum, Tolga Fikret; Gaeta, Michele; Cicciù, Marco, *Porphyromonas gingivalis, Periodontal and Systemic Implications-A Systematic Review*, 2019.
- 11: Goldfarb, T; Sberro, H; Weinstock, E; Cohen, O; Doron, Sy et al., *BREX is a novel phage resistance system widespread in microbial genomes*, 2015.

Summary

We investigated the structure function relationship of PorX by combined biophysical experiments and enzymatic assays. As a regulator of the type 9 secretion system (T9SS) of the human pathogen *Porphyrromonas gingivalis* (PG), PorX is responsible for PG virulence^[1] and as drug target could be the key in disease prevention.

We started with crystals of wild type PorX dimers, which were unsuitable for structure solution as well as further PG proteins for potential cocrystallization. We successfully optimized the crystallization conditions for selenomethionine derivative PorX by adding beryllium fluoride as phosphate mimic, zinc for higher occupancies in the active site, as well as additives and optimizing buffers, concentrations, crystallization- and crystal manipulation techniques.

By X-ray diffraction and single anomalous dispersion we solved several structures of dimeric PorX with and without substrate (pGpG) up to a resolution of 1.9 Å, characterized its domains, active site and substrate binding pocket.

The PorX N-terminal domain is a typical response regulator receiver domain (RD) of the *OmpR*^[13] subfamily. PorX has the characteristic $\alpha_5\beta_5$ Rossmann-like fold, conserved phosphorylation site and the YT pair responsible for conformational change during phosphorylation which leads to dimerization. A helical bundle domain (HBD) connects the RD to the C-terminal enzymatic active PglZ domain which was previously uncharacterized in structure and function. The PglZ domain has the characteristic α/β -fold of the alkaline phosphatase superfamily (APS), the metal coordination sphere of phosphodiesterase, and an extra domain capping the active site (CAP^[14]).^{[15], [16]}

By small-angle X-ray scattering we investigated PorX dimers in solution, which did not significantly differed from the crystal structure, and PorX monomers showing significant conformational changes. We build and validated a SAXS model of monomeric PorX with a χ^2 of 1.15 and explained the mechanism of PorX dimerization: Phosphorylation leads to RD dimerization, their conformational changes lead to an exchange of surfaces between RDs and HBDs, which then bring both PglZs into contact, followed by PglZ dimerization. The newly formed PglZ dimeric interface (DI) folds into the catalytic active APS bi-metallo active site, and the capping subdomain opens the substrate binding pocket due to interaction with DI. These we confirmed by structure based mutational studies, SEC-MALLS as well as phosphorylation assays, SEC and enzymatic assays performed by our collaboration partners.

Comparing PglZ subfamily homologs, we found out that the active center as well as its DI is

conserved with most BREX-PglZs, involved in phage defense^[17]. In PorX the DI folds upon dimerization and thus activates the protein, which thereby is likely also true for BREX-PglZs.

Our collaborators determined the enzymatic activity by enzymatic assays confirming PorX as phosphodiesterase and not a phosphatase even though APS active sites often have promiscuous activity^{[18], [19]}. Phosphodiesterase substrate affinities are regulated by the substrate binding pockets, so in order to narrow down potential substrates to screen, we compared PorX to its closest known structures and enzyme classes. As results we found the most similar substrate cavities in ectonucleotide pyrophosphatases which cleave linear and cyclic dinucleotides as well as mono- and dinucleotide polyphosphates. This led to cocrystallization trials yielding our pGpG bound structure with the PorX substrate binding pocket only partial filled.

This led us screen activity against a large polynucleotide library and product detection using ultra-high performance liquid-chromatography, coupled with mass spectrometry (UHPLC-MS). As result we found out PorX cleaves linear and cyclic oligonucleotides as well as linear dinucleotides. We did not found any activity against cyclic mono- or dinucleotides, mono- or dinucleotide polyphosphates. Therefore it is likely that PorX is involved in oligonucleotide signaling in PG and by designing of PorX PglZ inhibitors it might be possible to stop PG T9SS transcriptional activation and render PG avirulent.

Literature Summary

- 1: Veith, Paul D.; Glew, Michelle D.; Gorasia, Dhana G.; Reynolds, Eric C., "*Type IX secretion*", Molecular microbiology 106, 1, p35–53, 2017.
- 2: Gao, Rong; Bouillet, Sophie; Stock, Ann M., Structural Basis of Response Regulator Function, 2019
- 3: Kim, A; Benning, MM; OkLee, S; Quinn, J Martin, BM; Holden, HM; Dunaway-Mariano, D, "*Divergence of chemical function in the alkaline phosphatase superfamily*", Biochemistry 50, 17, p3481–3494, 2011.
- 4: Galperin, Michael Y.; Koonin, Eugene V., "*Divergence and convergence in enzyme evolution*", The Journal of biological chemistry 287, 1, p21–28, 2012.
- 5: Sunden F, AlSadhan I, Lyubimov A, Doukov T, Swan J, Herschlag D., "*Differential catalytic promiscuity of the alkaline phosphatase superfamily bimetallo core reveals mechanistic features underlying enzyme evolution.*", J Biol Chem. 292, 51, p20960–20974, 2017.
- 6: Chaudhary, Kulbhushan, "*BacteRiophage EXclusion (BREX)*", Journal of cellular physiology 233, 2, p771–773, 2018.
- 7: Mohamed, MF.; Hollfelder, F., "*Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer*", Biochimica et biophysica acta 1834, 1, p417–424, 2013.
- 8: Pabis, Anna; Kamerlin, Shina Caroline Lynn, "*Promiscuity and electrostatic flexibility in the alkaline phosphatase superfamily*", Current opinion in structural biology 37, p14–21, 2016.

Introduction

This image shows a full page of white paper with horizontal dotted lines. The lines are evenly spaced and run across the width of the page, providing a guide for handwriting practice. There are no margins, text, or other markings on the page.

1. Introduction

1.1 Porphyromonas gingivalis

Porphyromonas gingivalis (PG) is a non motile, anaerobe, gram negative bacterium belonging to the Bacterioides phylum.^[1] It is a member of the oral biofilm community and acts as the keystone pathogen for periodontal diseases.^[2] These include “gingivitis”, a mild inflammation, common in most of the human adult population^[3] and the more severe chronic inflammation “periodontitis”^[4] which affects 11.2% of the world’s population.^[5] In contrast to dominant pathogens, which suppress and displace other species in their environment, PG stays at low abundance in the gingival biofilm during periodontal disease progression, where it remodels the hosts immune system and disturbs the homeostasis of the oral microbiome.^[6]

Further PG itself is able to settle outside the oral cavity, colonizing the brain, liver, coronary artery and placenta^{[7],[8],[4]} and it has been associated to systemic diseases such as oral cancer^[9], rheumatic arthritis^[10], atherosclerosis^{[11],[12]}, cardiovascular disease^[13], coronary heart disease^[14], Diabetes^[15], Alzheimer disease^[16], pancreatic cancer^[17] and AIDS progression^[18].^[19] Other systemic inflammatory diseases can be enhanced by PG's immune modulatory lipopolysaccharides, which are found in the blood of 50 % of periodontitis patients^[20] and there is evidence for PG being involved directly and indirectly (by enhancing the bacterial load of oral microbiota) to be responsible for mortality by coinfections of the lung in severe cases of SARS (the Severe Acute Respiratory Syndrome), the current COVID-19 (coronavirus disease 2019) pandemic and H₁N₁ influenza.^[21]

1.2 Oral microflora

Up to 1,179 different taxa have been shown to colonize the human oral cavity.^[22] Some of them coevolved together with their hosts and established a stable relationship in which commensals and potential pathogenic bacteria live in homeostasis.^{[23],[24]} Among these colonizers there are ca. 50 predominant bacterial species, yeasts, fungi, protozoa and mycoplasma.^{[25],[22]} The host contributes to this community supplying nutrition, buffering the environment and fighting exogenous species with leukocytes, macrophages and secretion of antimicrobial factors.^{[26],[27],[24]} Balance in composition and amounts of these colonizers with the host establish a symbiotic host-commensal relationship,^{[28],[26]} while an alteration of these levels can lead to virulence factor expression of pathobionts, bacterial overgrowth and the arise of several diseases^{[29],[30]}.^{[6],[26],[24]} Most of the oral microbes grow within the oral cavity in a planktonic lifestyle^{[31],[27]} and some even invade and grow

within the epithelial tissues or intracellularly^[24]. However, periodontal diseases are associated with biofilm formation, as are 60 %^[32] of all human bacterial infections. ^{[26],[24]}

Biofilms are complex multispecies communities enclosed in an extracellular matrix ^[23], which provides bacteria within with higher immunity against the host's immune system as well as against antibiotics^[33]. To build up a bacterial biofilm, bacteria attach to the host's surface either directly or by coadherence via cell-cell interactions ^{[33],[34],[35]} and secrete polymeric substances^[24], including proteins and exopolysaccharides. The resulting extracellular matrix provides mechanical stability to the attached cells (5-35 % (v/v) of the biofilm), prevents the hosts immune cells from entering it,^[33] as well as builds a physical diffusion barrier for regulatory glycoproteins, albumins and lipids from the host's saliva and gingival crevicular fluid (GCF) as well as from other microbiota outside the biofilm. ^{[36],[27]} To share the nutrition and space, the different species within the biofilm develop utilize intra- and inter-species communications^[23] including specific cell-to-cell recognitions (coaggregation)^{[23],[37]}, quorum sensing^{[38],[39]}, horizontal gen transfers^[40], signal molecule communication and metabolic crossfeeding^[41]. ^{[23],[24],[26]}

Within the oral cavity two distinct types of biofilm exist and are classified as supra- (also referred as "marginal biofilms") and subgingival biofilms (also referred as "plaque"). ^{[23],[27]} The supragingival biofilms are localized above the gingival border, are more exposed to oxygen and thus composed mainly of gram-positive, saccharolytic facultative aerobe bacteria.^[27] They are responsible for caries and calculus formation. ^[27] On the other hand, the subgingival biofilms are localized between tooth and the gingivae, favor asaccharolytic, obligate anaerobic bacteria and are responsible for gingivitis and periodontitis ^[27].

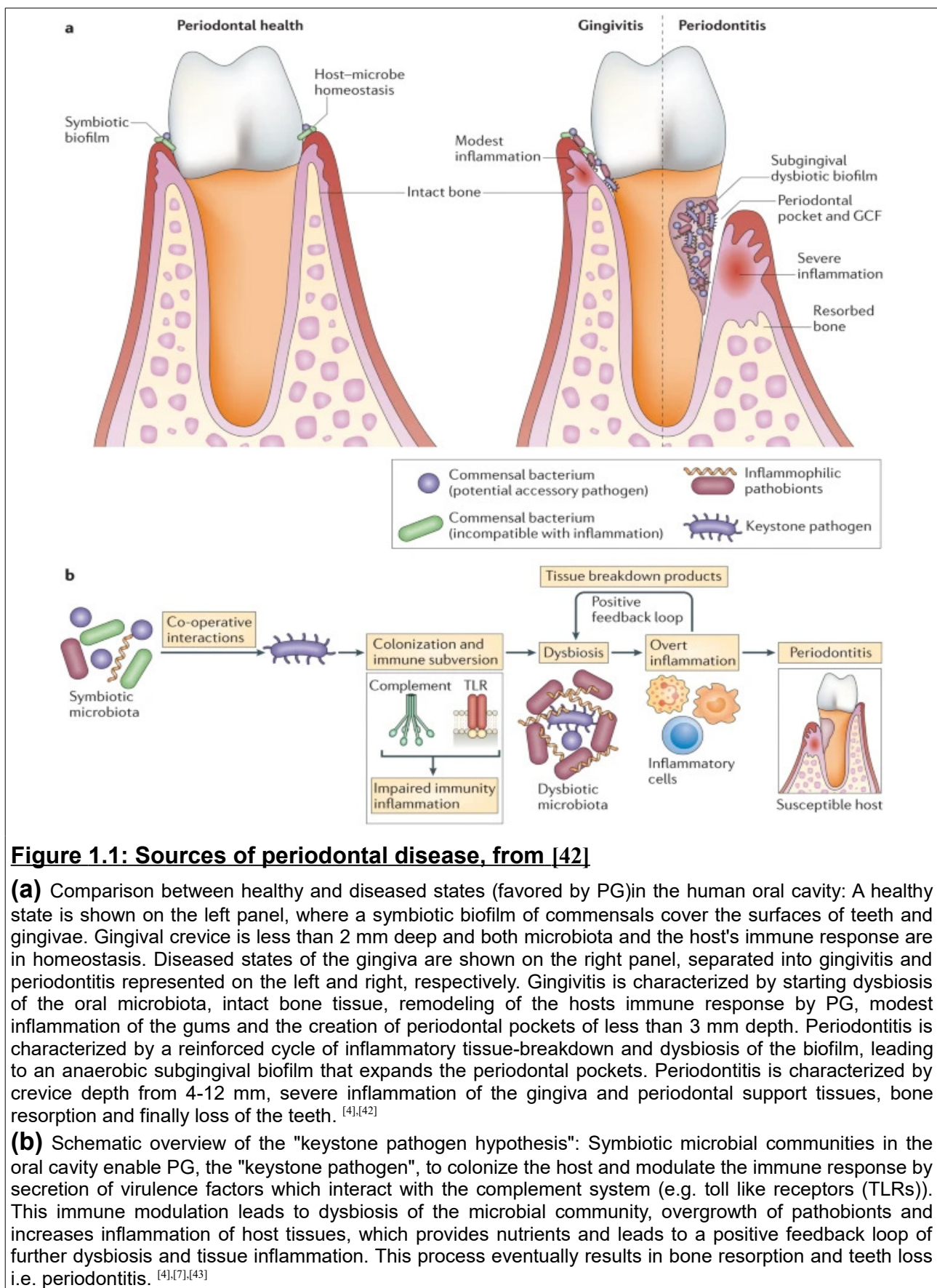


Figure 1.1: Sources of periodontal disease, from [42]

(a) Comparison between healthy and diseased states (favored by PG) in the human oral cavity: A healthy state is shown on the left panel, where a symbiotic biofilm of commensals cover the surfaces of teeth and gingivae. Gingival crevice is less than 2 mm deep and both microbiota and the host's immune response are in homeostasis. Diseased states of the gingiva are shown on the right panel, separated into gingivitis and periodontitis represented on the left and right, respectively. Gingivitis is characterized by starting dysbiosis of the oral microbiota, intact bone tissue, remodeling of the host's immune response by PG, modest inflammation of the gums and the creation of periodontal pockets of less than 3 mm depth. Periodontitis is characterized by a reinforced cycle of inflammatory tissue-breakdown and dysbiosis of the biofilm, leading to an anaerobic subgingival biofilm that expands the periodontal pockets. Periodontitis is characterized by crevice depth from 4-12 mm, severe inflammation of the gingiva and periodontal support tissues, bone resorption and finally loss of the teeth. [4],[42]

(b) Schematic overview of the "keystone pathogen hypothesis": Symbiotic microbial communities in the oral cavity enable PG, the "keystone pathogen", to colonize the host and modulate the immune response by secretion of virulence factors which interact with the complement system (e.g. toll like receptors (TLRs)). This immune modulation leads to dysbiosis of the microbial community, overgrowth of pathobionts and increases inflammation of host tissues, which provides nutrients and leads to a positive feedback loop of further dysbiosis and tissue inflammation. This process eventually results in bone resorption and teeth loss i.e. periodontitis. [4],[7],[43]

1.3 Subgingival biofilm formation

Biofilm formations can be subdivided into four stages^[33]: First early colonizers attach to the host's dental surface and establish initial colonies, followed by the second stage of coadherence of secondary colonizers. In a third stage secondary colonizers attract late colonizers and in a final stage, microorganisms start to disperse away from the mature biofilm.^{[26],[33],[44]}

Each type of oral microorganism involved in biofilm formation has a specific set of adhesins and receptors that mediate coaggregation and growth through communication with the other microorganisms and the host.^[37] Through this communication, the different species colonize preferred sites within the biofilm architecture, and the diversity of the biofilm composition increases during maturation.^{[23],[37]}

After tooth-brushing (or similar mechanical removal of the oral microflora) glycoproteins from the host's saliva will cover the exposed teeth and gingiva surfaces within a few minutes. These glycoproteins are recognized from early colonizers by specific receptors such as mucin-, amylase-, statherin-, agglutinin receptors or receptors binding to proline- or phosphate rich protein sequences.^{[23],[24]} Four hours after cleaning, the initial biofilm consists mainly of *Streptococci*, among them *S. oralis*, *S. mitis*, *S. gordonii*, *S. cristatus*, *S. sanguinis*, *S. intermedius*^{[45],[46],[24]} which build up between 60 and 90 %^[47] of the initial biofilm. These organisms are interrelated with a stable commensal-host relationship and oral health.^{[25],[48],[24]} Other microorganisms known to participate in the early colonization are species of the genera *Actinomyces*, *Veillonella*, *Capnocytophaga*, *Eikenella*, *Haemophilus*, *Prevotella* and *Propionibacteria*.^{[23],[24]} The growth of the saccharolytic aerobic early colonizers then lowers the redox potential of the biofilm environment which allows proliferation of anaerobic bacteria.^[27] The moderately pathogenic "bridging colonizers"^[49] such as *Fusobacterium nucleatum* attach to the early colonizers and coaggregate with the late colonizers via specific saccharid- or adhesin receptors.^[23] When the redox potential is low enough the late colonizers of the "red complex" are recruited. This complex consists of PG, *Treponema denticola* and *Tannerella forsythia* and is correlated with periodontal diseases.^{[50],[49]} These species proliferate in the anaerobic subgingival biofilms, where *T. denticola* is more aerotolerant and also grows in the outer layer of gingival pockets whereas PG is restricted to the inner layer^{[51],[37]}. These anaerobic bacteria are recruited from the oral environment, where they were able to survive but not proliferate for several hours at interproximal spaces under ambient air.^{[52],[37]} PG is usually attached via *S. gordonii* and *F. nucleatum*.^[53] *T. denticola* is also attached by *F. nucleatum* and stimulates PG cell growth and vice versa.^[23]

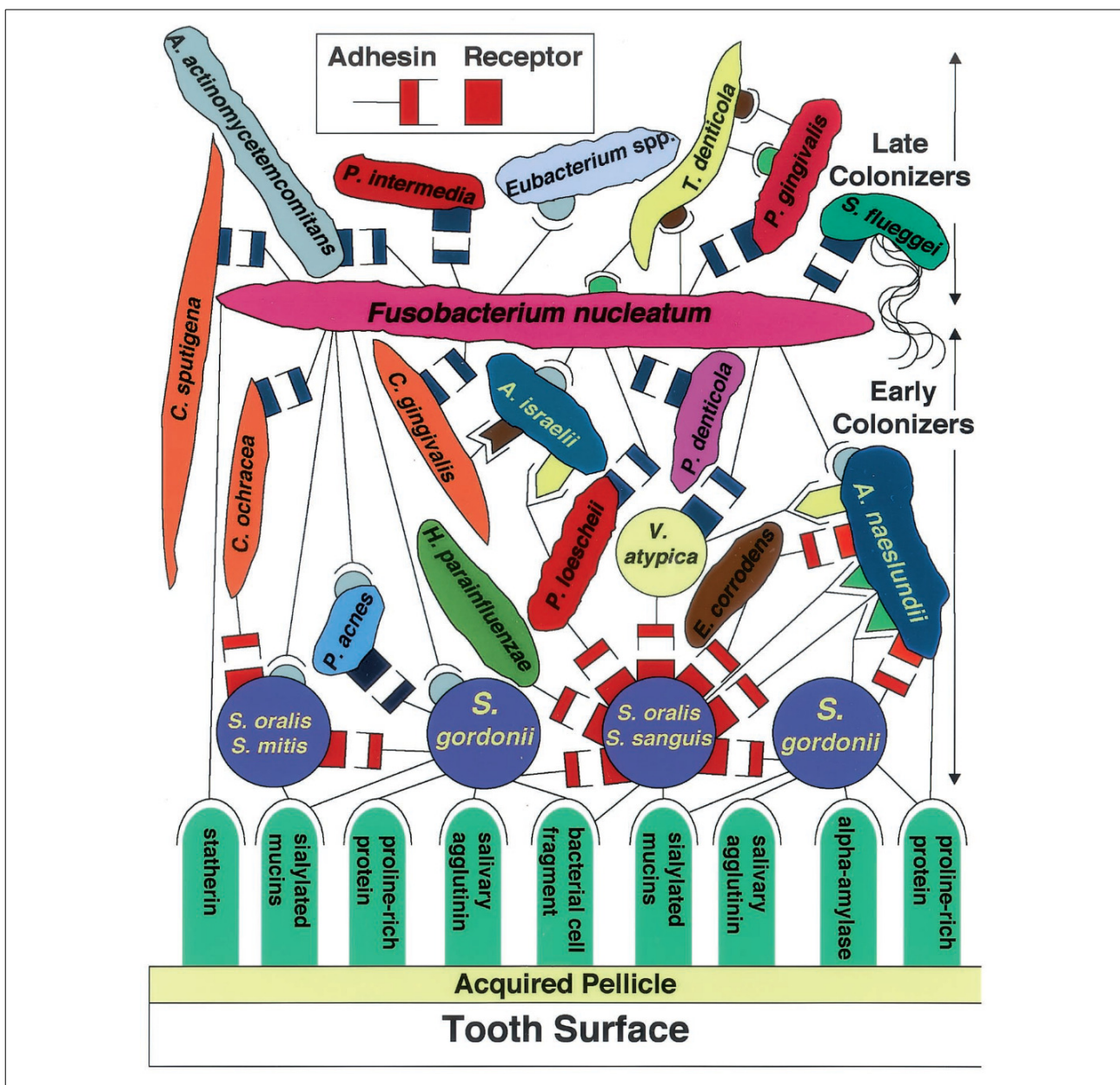


Figure 1.2: Successive microbiological biofilm formation within the oral cavity, from [23]:

Primary colonizers (represented as blue circles) of the *Streptomyces* genus bind via adhesin receptors to the salivary glycoprotein (green vertical bars with round ends) from the pellicle, that coats the teeth and gums. Microbiota of this primary layer is bound by secondary early colonizers via specific adhesin-receptor pairs (same color with complementary shape, exemplified in the rectangle inset at the top). Late colonizers coaggregate by bridging colonizers already attached to early colonizers.

The coaggregations shown are experimental proven in pairwise tests. Different kinds of coaggregations are represented by different colors and shapes. Same colors show similar but not necessarily identical coaggregations. Identical shapes represent functional but not necessary structural similar interactions. Rectangle-shaped interactions are inhibited by lactose. Inhibitors of interactions other than rectangle shaped are unknown. Heat or protease sensitive receptors are depicted with a stem, insensitive receptors are connected to the corresponding microorganism with black lines. [23]

1.4 PG influence on the oral commensal-host-community

The classification of PG as the "*keystone pathogen*"^[2] for periodontal diseases is based on the following PG-specific features:

PG is the only bacteria producing virulence factors at disease-free sites whereas *T. denticola* and *T. forsythia* start expression of virulence factors once disease progresses.^[54] Low amounts of PG (0.8 % of cells within mature biofilms of human periodontitis) are sufficient to induce dysbiosis.^[55] PG suppresses the growth of certain commensals and changes the gene expression of pathobionts. Upon colonization, PG induces expression of transposases and DNA fragmentation in health-associated bacteria such as *S. mitis* and *Lactobacillus casei*. Furthermore, PG increases the concentration of reactive oxygen species by 25-fold within the biofilm, which leads to a 10-fold increase of *S. mitis* cell death and thus reorganization of the biofilm composition by increasing the amounts of pathobionts.^{[56],[46]} Meanwhile, these activate transcription factors responsible for ABC-transport systems, cell division, chaperone or growth factor expression and release of virulence factors.^{[2],[56]}

PG exploits the immune system of the host. PG virulence factors induce a permanent immune-inflammatory response in which the migration and bactericidal activity of leukocytes, who in a healthy oral environment suppress inflammation, are inhibited. Suppression of the host's immune response together with increased inflammation enriches the nutritional environment of the biofilm by tissue break down, leading to enhanced growth of pathobionts which further increases the inflammation, thereby enhance periodontitis and eventually lead to inflammatory bone resorption.^{[4],[6],[24],[46],[57]}

1.5 PG virulence factors and their effects on the host immune system

Evasion of the immune system and remodeling of the host's immune response by PG are based on PG's virulence factors (Table TABLE 1.1). These are membrane bound or released to the environment either as soluble factors or within outer membrane vesicles (OMVs). The most important virulence factors of PG are fimbriae, lipopolysaccharides (LPS) and small cysteine proteases called gingipains, which are lysine-specific gingipains (Kgp) or arginine-specific gingipains (RgpA and RgpB).^{[4],[58],[59]}

PG Virulence Factor	Direct function	Effects	Signaling	Cell types	Reference
LPS-lipopoly-saccharides	- TLR2 activation	- proinflammatory	- NFκB activation by the TLR2-MyD88-pathway - NFκB sensitization by the P38-MAPK pathway	- leukocytes, endothelial cells, epithelial cells, fibroblasts	[60],[61], [62]
		- cell entry and intracellularly survival - antiinflammatory TLR2-MyD88 inhibition - proinflammatory TNF release - anti-bacterial activity	- enhancement of TLR2-CR5a cross talk (TLR2-PI3K pathway), cAMP-PKA dependent MyD88 degradation and TNF release	- Monocytes, NK cells, neutrophils, macrophages	[4]
	- TLR4 activation	- proinflammatory	- NFκB activation by the MD-2/TLR4-pathway	- leukocytes, epithelial cells, endothelial cells, fibroblasts	[4]
		- upregulation of TLR4 and TLR2 expression	- NFκB activation by the MD-2/TLR4-pathway	- Fibroblasts	[63]
	- Lipid A (hemin concentration dependent phosphorylation and acetylation)	- proinflammatory (low hemin) weak TLR4 activation	- penta-acyl-mono-phosphorylated LipidA-NFκB activation by the MD-2/TLR4-pathway	- leukocytes	[57],[63], [64],[4]
		- antiinflammatory (high hemin) TLR4 deactivation - decreased β-defensin release and NFκB sensitization - inhibition of MAPK pathway	- tetra-acyl-mono-phosphorylated LipidA-TLR4 inactivation	- epithelial cells	
	- CD180 activation	- antiinflammatory	- MD2/TLR4-NFκB pathway inhibition	- keratinocytes - dendritic cells	[65]
		- immune response remodeling	- TLR2 and TLR4 downregulation (not transcriptional) - enhanced LPS binding protein (LBP) expression - CD180 upregulation	- macrophages	[63]
Fimbriae	- TLR2 activation	- proinflammatory - enhanced bone resorption by osteoclasts, osteoclast differentiation from precursor cells	- NFκB activation by the TLR2-MyD88-pathway - NFκB sensitization P38-MAPK pathway	- Epithelial cells - Osteoclasts - immune cells	[66],[67], [61]

PG Virulence Factor	Direct function	Effects	Signaling	Cell types	Reference
Fimbriae	TLR2 +CXCR4	- antiinflammatory TLR2-MyD88 inhibition - preventing bactericidal activity - PG cell entry - phagosome maturation inhibition - PG survival - IL-10 expression (antiinflammatory)	- (C5aR-)TLR2-CXCR4 cross talk activating TLR2-PI3K pathway - cAMP-PKA dependent MyD88 degradation and NFκB inhibition - TLR2-PI3K dependent CR3 pathway activation	- Monocytes, NK cells, macrophages - neutrophils	[68],[69], [70],[71], [6],[72], [73],[74], [75],[76], [64]
	- adhesion	host cell attachment	via α5β1-Integrin	- Epithelial cells, fibroblasts	[77]
Gingipains	- adhesin	- adhesion to host cells		- Epithelial cells	[45]
		- coaggregation with biofilm		- <i>T. denticola</i> - <i>S. gordonii</i>	[45],[78]
		- Binding of complement inhibitor C4BP at hemagglutinin/adhesin domain - reduced leukocyte recruitment, protection from phagocytosis and cell lysis	-inhibition of immunoglobulin recognition (classical complement pathway) - inhibition of mannose- and n-acetylglucosamin recognition (lectin pathway)	PG	[79]
	Transpeptidase activity	- acquiring iron from host's hemoglobin - disturbance of immune tolerance - proinflammatory response- autoimmune reaction		- host cells	[80]
	Direct degradation of complement factors IL-1b, TNF, IgG1, IgG3	- immune system evasion for all oral bacteria, antiinflammatory			[64],[4]
	Successive IL-8 digestion	High concentrations: - inactivation of proinflammatory IL-8-72aa isoform Low concentrations: - cleavage from inactive IL-77aa isoform to hyperactive IL-8-69aa isoform	- reduced neutrophil recruitment and ROS release - high neutrophil recruitment and ROS release		[81],[82], [83],[64]

PG Virulence Factor	Direct function	Effects	Signaling	Cell types	Reference
	mCD14 (TLR coreceptor for LPS and fimbriae) down-regulation	- immune system evasion for all oral bacteria		- host cells	[84] [76]
	C5 degradation, C5a binding to C5aR	High concentrations: - cell entry and intracellularly survival - antiinflammatory TLR2-MyD88 inhibition - proinflammatory cytokine release - reduced anti-bacterial activity Low concentration/soluble gingipains: - proinflammatory signaling - leukocyte recruitment	- activation of TLR2-CR5a-CXCR4 cross talk (TLR2-PI3K pathway), see Fimbriae: TLR2+CXCR4) - neutrophil recruitment and activation	- Monocytes, NK cells, neutrophils, macrophages	[68],[69], [70],[71], [6],[72], [73],[74], [75],[76], [64]
	PAR (protease activated receptors)-activation	- proinflammatory	PAR-signaling	Epithelial cells, fibroblasts, neutrophils, T lymphocytes, osteoblasts, monocytes	[85]
	NOD-like receptors activation	- proinflammatory	- NLRP3 modified inflammatory corpuscle activation, IL-1b and TNF release - NOD2 activation		[86]
	haemolysin-, haemagglutination activities	- agglutination of erythrocytes - lysis of erythrocytes - degradation of hemoglobin	- hemin release-nutrition acquisition	- erythrocytes	[87],[88], [89],[90]
Gingipains and PPAD	- reduced transcription of IP-10, ITAC, and Mig chemokines (CXCR3 ligands)	- reduced adaptive immune system response - fewer antibacterial chemokines - T cell imbalance - chronic inflammation conditioning	- STAT1-IRF1	- Epithelial cells, monocytes, neutrophils - reduced CXCR3 response from T and B cells, NK cells, NKT cells, dendritic cells and monocytes	[91],[92], [93]

PG Virulence Factor	Direct function	Effects	Signaling	Cell types	Reference
	- cleavage and citrullination of host proteins	- autoimmune effects - increased inflammation	-	-	[94],[95]
SerB	- NFκB inactivation by phosphorylation	- antiinflammatory	- NFκB	- Epithelial cells	[96]
Rub-erythrin	- protection from ROS	- PG survival		- PG	[97],[64]
Superoxid dismutase	- protection from ROS	- PG survival		- PG	[4]
CPG70	metallocarboxypeptidase	- potent virulence factor secreted by the T9SS			[98]
HBP35	hemin binding protein	- potential virulence factor secreted by the T9SS - thioredoxin activity	- nutrition acquisition - essential during hemin-depleted conditions		[98]
OMVs (LPS and gingipain enriched vesicles) and soluble gingipains	- host cell targeting in proximity of PG - proinflammatory signaling - apoptosis inhibition	- proinflammatory signaling - NO release by macrophages - systemic inflammatory diseases	- TLR2, TLR4, NFκB activation, C5a depended leukocyte recruitment, IL-8-69aa digestion	- see LPS, gingipains	[81],[82], [83],[64], [58],[99], [98],[100]

TABLE 1.1: Overview on the most studied PG virulence factors.

Summary of the direct effects of virulence factors, the involved signaling pathways and known or likely involved cell types directly affected.

To remodel the immune response of the host, PG's virulence factors bind to the gingival cells and disturb intracellular signaling pathways that usually lead to local inflammation, leukocyte recruitment and anti-bactericidal activity against pathogens. In particular, also known as "*local chemokine paralysis*"^[82], PG disrupts the natural IL-8 gradient used to detect bacterial colonization.

In a healthy oral host-commensal relationship, these host cells recognize pathogen-associated molecular patterns (PAMPs) of bacterial cells by pattern recognition receptors (PRRs). Their intracellular signaling pathways lead to the expression of cytokines to attract leukocytes, initiate the immune response and prevent pathogenic invasion. Cytokines are small signaling proteins: Interleukins (ILs) for leukocyte induced inflammation, immune response activating interferons (INFs) and tumor necrosis factor (TNF), colony stimulating factors (CSFs) and chemokines that attract further immune cells. ^{[4],[101]}

The contact and recognition of the homeostatic bacterial biofilm by the host cells epithelium in a healthy state leads to the expression of cell adhesion molecules (e-selectin, ICAMS) and the release of proinflammatory chemokines, in

particular IL-8. The concentration gradient of IL-8 attracts leukocytes (including neutrophils, NK cells, T-lymphocytes and monocytes (macrophages and dendritic cells)) into the gingival crevice. There the attracted nucleophils build a barrier between the host's epithelial cells and the bacterial biofilm. They protect the host tissue from direct contact with the bacteria and prevent the pathobionts from overgrowth by bactericidal activity.

IL-8, among other cytokines are expressed under the control of the transcription regulator "nuclear factor kappa light chain enhancer of activated B-cells" (NF κ B, a heterodimer of RelA and p50 proteins). NF κ B activation is regulated by PAMP recognition by Toll-like-receptors (TLRs, complement PRRs) expressed in neutrophils, macrophages, fibroblasts and epithelial cells including keratinocytes. The signaling pathways of TLRs influence NF κ B activation (TLR2, 3 and 4, Figure 1.3). Their co-receptors with differential affinities (mCD14, sCD14, LBP, MD1, MD2, TLR1, TLR6), expression levels and cross-regulations (via receptors CD180, CXCR4, C5aR) vary between cell-types and are regulated by alternative signaling-pathways. [4],[6],[24],[57]

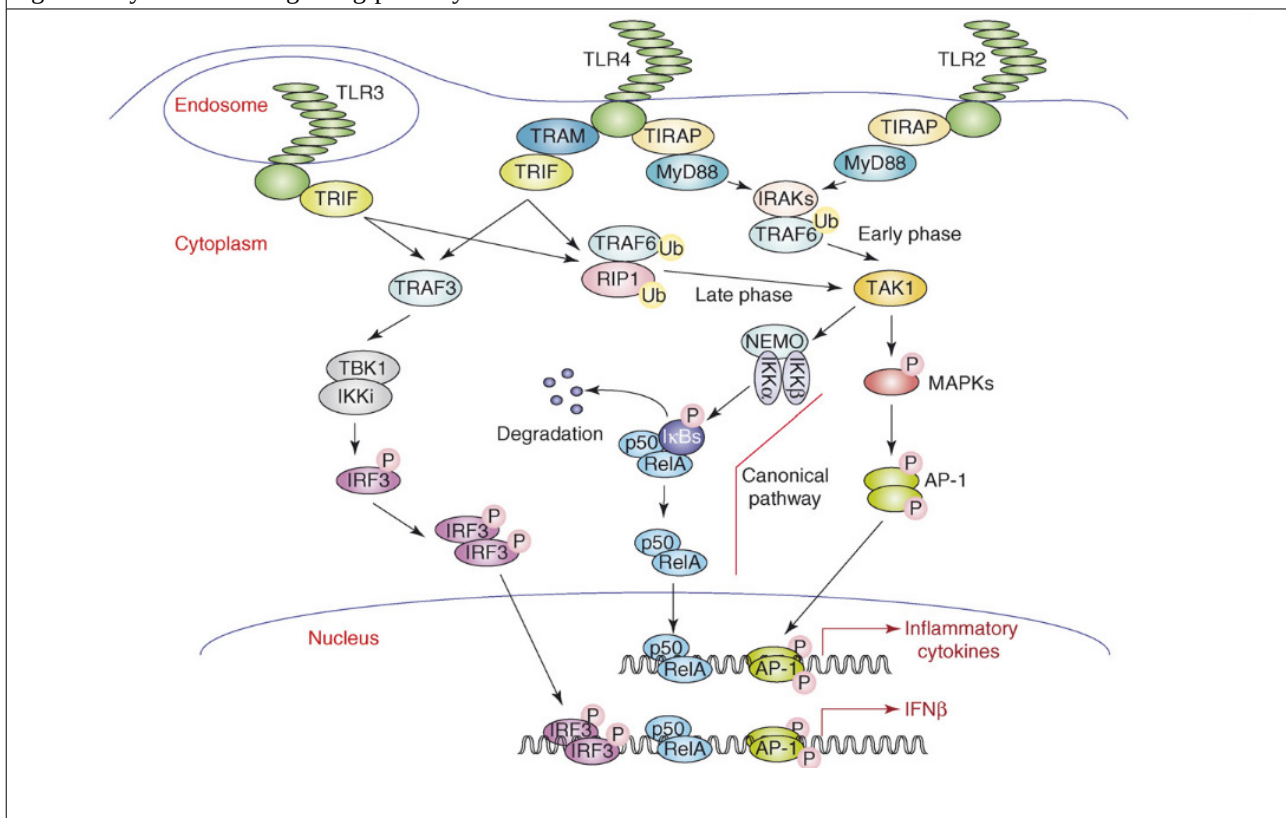


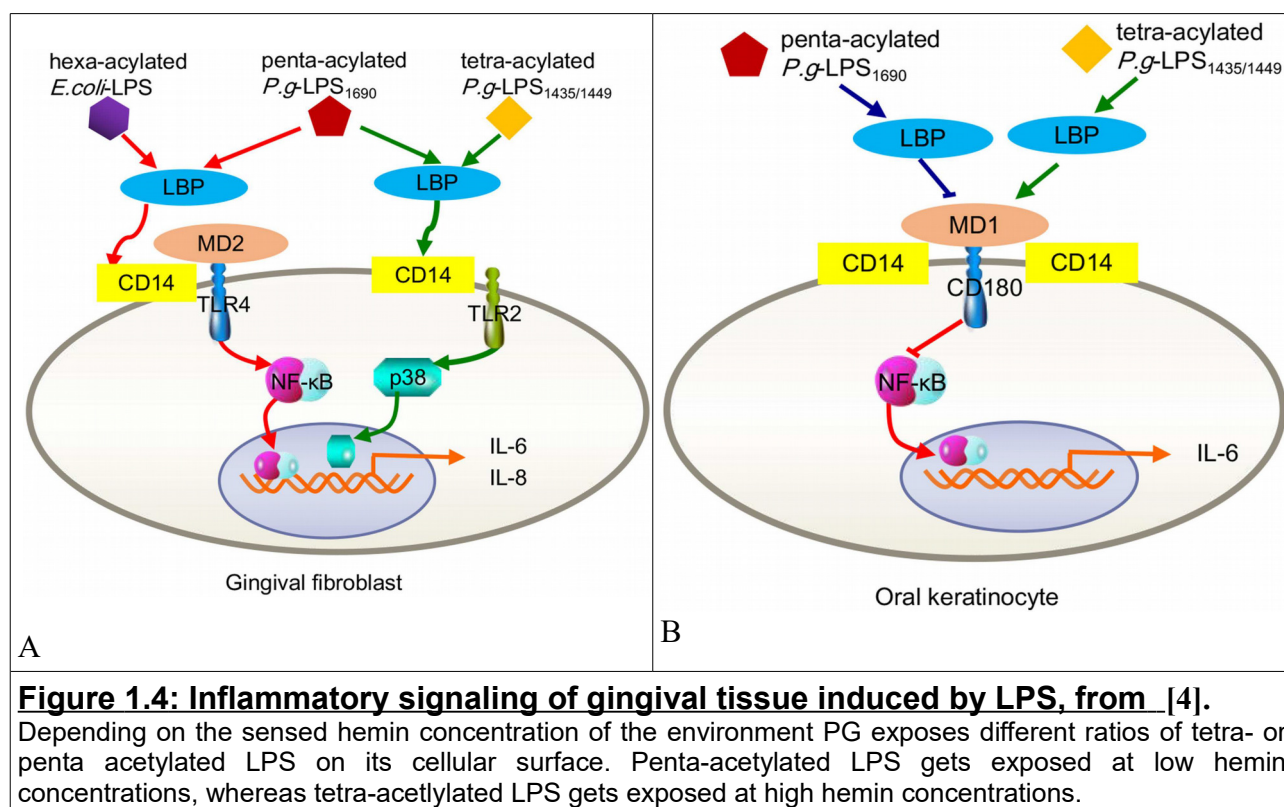
Figure 1.3: Schematic representation of TLR2, TLR3 and TLR4 signaling [102]

The transcription factor NF κ B, a p50/RelA-Dimer, is responsible for the expression of proinflammatory cytokines. TLR2 regulates NF κ B early phase activation dependent on the Myd88/TIRAP complex. TLR4 can activate NF κ B in a MyD88 dependent (Myd88/TIRAP) and a late phase activation Myd88 independent (TRAM/TRIF) signaling cascade. This results in TAK1 complex activation, which phosphorylates the inhibitor protein I κ B, which therefore is degraded and release active NF κ B into the nucleus. [102] The TAK complex can also activate the MAP kinase cascade (MAPKs). MAPKs activate the activator-proteins-1 (AP-1), a transcription factor family, modulating NF κ B response and vice versa [103] and with differential binding to cytokine promoters [40]. The MyD88 independent pathway via TRIF by TLR4 activation can also be activated via TLR3 and activates the interferon regulatory factor 3 (IRF3), responsible for expression and release of interferons. [102] In a healthy state of the oral environment TLR4 is activated by LPS, TLR2 by LPS, lipoproteins and fimbriae from commensals. [4] TLR3 gets activated by extracellular RNA released from necrotic cells/inflamed tissue. [104]

1.5.1 Lipopolysaccharides (LPS)

Change in expression, modification and exposure of LPS by PG can introduce a switch from pro- to an antiinflammatory response in the host cells. PG's LPS can activate NF κ B via the TLR2-MyD88 pathway^[60] and enhance TLR4 NF κ B activation by sensitization of the P38-MAPK pathway. Also Lipid A as part of PG LPS can activate NF κ B by TLR4. However, depended on sensed hemin concentration, PG can change the acetylation state of Lipid A (see Figure 1.4). Penta-acetylated Lipid A, expressed at low hemin concentration, is a strong TLR4 and TLR2 agonist that leads to an inflammatory response, while tetra-acetylated Lipid A, expressed at high hemin concentration, strongly prevents TLR4 activation and only weakly stimulates TLR2.^{[57],[63],[4]} This inhibits the immune response by leukocytes or human gingival epithelial cells against PG at already inflamed sites, where the hemin concentration is high. TLR4 activation is also responsible for a positive feedback loop inducing it's own expression, as well as the expression levels of TLR2.^[63] Thus, by inhibition of TLR4 there is a reduced response of TLR4 or TLR2 proinflammatory virulence factors from pathobionts.

Furthermore, PG's LPS binds to the TLR "CD180" (also known as "RP 105") on keratinocytes, dendritic cells and macrophages^{[65],[4]} whose signaling cascade decreases NF κ B activation as an TLR4 antagonist.^[4]



(A) In human gingival fibroblast penta-acetylated LPS weakly activates TLR 4 resulting in moderate proinflammatory IL-6 and IL-8 release in a similar manner than the LPS of canonical bacteria e.g. *E. coli*. However, tetra-acetylated LPS does not activate TLR4 resulting in NFκB activation and proinflammatory chemokine expression. Instead it activates the MAPK pathway via TLR2, resulting in NFκB sensitization but decreased proinflammatory effects compared to TLR4 activation. ^[4]

(B) In human oral keratinocytes both penta- and tetra-acetylated PG's LPS are able to bind to CD180, a TLR inhibiting the usual TLR4 pathways. However NFκB inhibition via CD180 is stronger mediated by tetra-acetylated than by penta-acetylated PG's LPS. ^[4]

1.5.2 Fimbriae

Fimbriae/pili of PG bind and activate TLR2 and CXCR4 receptors on the host cells. Fimbriae are elongated protein fibers at the bacterial surface, enhancing bacterial adhesion to other bacteria, host cells or inorganic surfaces and are involved in biofilm formation. ^{[105],[106],[4]}

Fimbria interactions with these cell receptors facilitate cell entry of PG, its intracellular survival, reduced phagocytosis and reduced bactericidal activity of leukocytes without inhibiting proinflammatory chemokine release. By binding and activation of TLR2, PG's fimbriae can activate NFκB by the MAPK pathway and induce a proinflammatory response in monocytes ^{[67],[4]}. However, in combination with PG's fimbriae activation of CXCR4 (expressed on leukocytes, endothelial and epithelial cells and acting as a TLR2 co-receptor), NFκB gets deactivated and the host cells express IL-10, an antiinflammatory chemokine. ^{[73],[4]}

1.5.3 TLR2-CXCR4-C5aR crosstalk

The reason for the NFκB deactivation by TLR2-CXCR4 crosstalk (in macrophages, neutrophils and NK cells) is the activation of the cAMP cyclase, protein-kinase-A (PKA) activation and subsequent deactivation of the TLR2-Myd88-NFκB proinflammatory pathway, responsible for nitroxid production in macrophages^[70] (Figure 1.5A) as well as the production of IL-6/-12, TNF and IFN- γ ^[107]. ^{[69],[4]}

The TLR2-Phosphoinositide-3-kinase (PI3K) pathway (also known as TLR2-MAL (-MyD88-adaptor-like) or TIRAP-PI3K- pathway) activated by TIRAP (a TLR adapter protein) further activates the TLR2-complement receptor CR3-(a hetero dimeric integrin) pathway which enhances the intracellular survival of PG. The PI3K/CR3 activation leads to a TLR2 dependent cellular entry of PG into macrophages ^{[68],[72]}, adhesion of leukocytes ^{[108],[109]} and a change in cytokine production (enhanced release of IL-1a, -1b, -6 and TNF from macrophages^[68], and reduced IL-12p35, IL-12/23p40 and IL-23p19 transcription, as well as reduced IL-12p70 and IL-23 translation and IFN- γ production)^[110]. ^{[4],[76]} Lower IL-12 and IL-23 concentrations lead to reduced tissue damage by T-helper cells^{[110],[72]}, whereas higher concentrations of IL-1b, -6 and TNF lead to differentiation and

activation of osteoclasts, which are responsible for bone resorption during periodontitis^[111]. In neutrophils PI3K/CR3 activation lead to MyD88 degradation after ubiquitination (thus no NFκB activation and bactericidal response) and PG's phagocytosis gets inhibited by blocking the Ras homolog family member A (RhoA) (Figure Figure 1.5 B). ^{[74],[42]}

The cAMP cyclase as well as the PI3K/CR3 pathways get further enhanced by C5a binding to the C5a receptor (C5aR) on macrophages and neutrophils. C5a is one of two subunits (C5a/C5b) of the host's complement factor C5 that usually gets cleaved as an immune response by the host's C5 convertase but, in case of PG infection, it will be cleaved by PG's cystein protease gingipains either as secreted proteases or bound on the bacterial surface of PG. ^{[6],[70],[71]}

The main influences on the host of the C5aR-CXCR4-TLR2 cross-talk are enhanced intracellular survival of PG and enhanced inflammation signaling of the host cells combined with a lower antibacterial response by leukocytes. ^{[24],[72]}

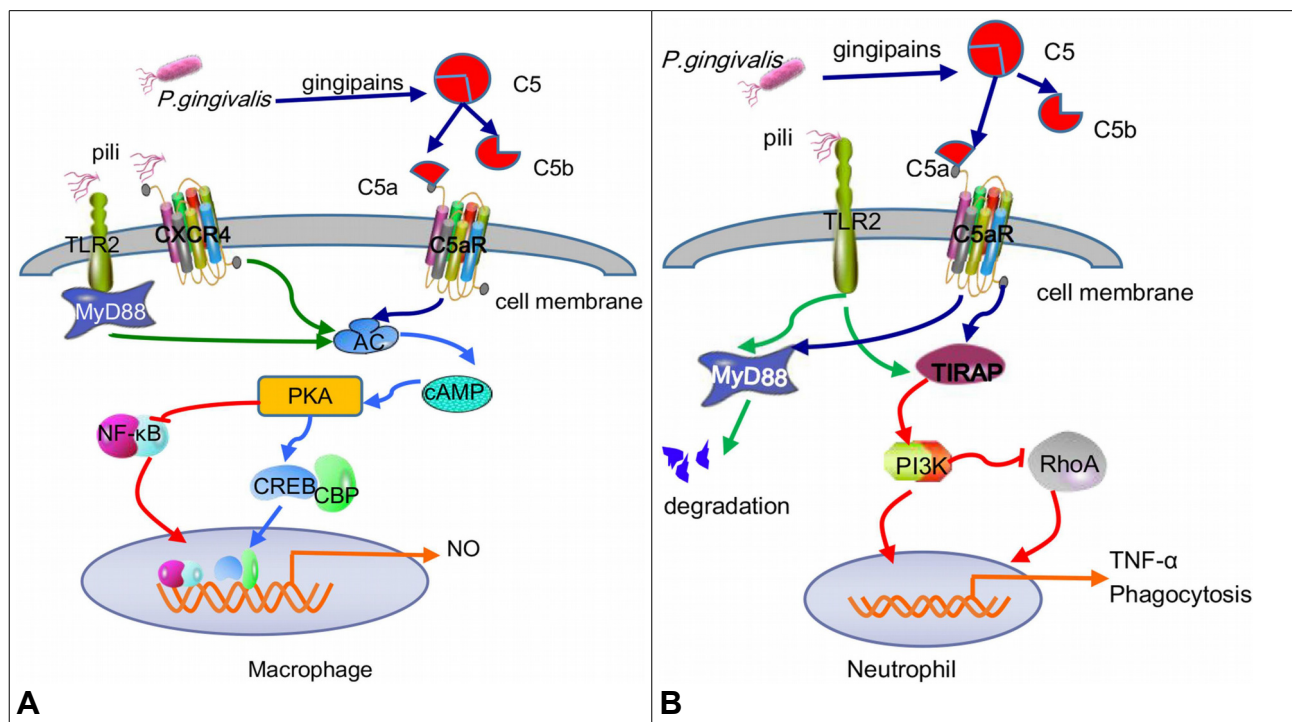


Figure 1.5: TLR2-CXCR4-C5aR crosstalk, from [4]

Different immune modulation of macrophages and neutrophils by the TLR2-CXCR4-C5aR crosstalk.

(A) In macrophages activation of TLR2 and CXCR4 (both by pili/fimbriae) and the indirect activation of C5aR by gingipains synergetically lead to increased cAMP levels by activating the adenylate cyclase (AC). cAMP activates the protein kinase A (PKA) that phosphorylates the cAMP response element binding protein (CREB) which in turn binds nuclear coactivator CREB-binding protein (CBP) which in its unbound form is needed by NFκB for transcriptional activation. ^[112]

(B) In neutrophils activation of the PI3K pathway by TLR2 and C5aR lead to MyD88 degradation, dependent on AC and PKA activation. This deactivates the proinflammatory signaling by the TLR2-MyD88 pathway which would lead to TNF secretion in neutrophils and furthermore PI3K deactivates RhoA resulting in impaired PG phagocytosis. ^[4]

1.5.4 PG suppression of the adaptive immune response

PG also suppresses the adaptive immune response of the host by inhibiting the release of T-cell chemokines CXCL9/10/11 from epithelial cells, monocytes and neutrophils. [24]

Interferon (INF)- γ induced protein 10 (CXCL10/IP-10), INF-induced T-cell chemo-attractant (CXCL11/ITAC) and monokine induced by INF- γ (CXCL9/MIG) are host cell released chemo-attractants that bind to the CXCR3 receptor and have anti-bactericidal activity on their own. CXCR3 is expressed on T-helper Th1CD4-cells and T effector CD8 cells as well as on monocytes, NK- and NKT-cells, dendritic-cells and B-cells. Release of and activation via CXCL9/10/11 leads to migration of these immune cells to inflammatory sites and antimicrobial and proapoptotic effects that are prevented by PG infection. [91]

PG suppresses the transcription of these T-cell chemokines by inhibiting the STAT1-IRF-1 pathways in epithelial cells and neutrophils by releasing of two or more soluble virulence factors. One of these factors are gingipains that have been shown to reduce the expression of CXCL10 in fibroblast cells. [92] Another one is peptidylarginine deiminase (PPAD) which, once released by PG inhibits IRF-1 expression. [93] PPAD is an enzyme unique to the *Porphyromonas* genus [94] that together with arginine specific gingipains citrullinates C-terminal arginine residues of proteins. Citrullinated proteins are bound by anticitrullinated protein antibodies (ACPA), so that an autoimmune reaction occurs including increased release of proinflammatory cytokines, which is a major driving factor for rheumatic arthritis. [94],[95]

Important host complement factors targeted and overexpressed upon PPAD are IL-1 immune modulators or leukocyte chemoattractants (IL-1 β , IL-8, IL-36 γ , CCL20), leading to increased inflammation. [94]

1.5.5 Concentration dependency of PG virulence factors

Intriguingly, different studies have shown an inflammatory or antiinflammatory response on the same virulence factors of PG. The above mentioned antiinflammatory responses are concentration dependent and take effect at high concentrations of virulence factors either when they are bound to the membrane or released in close proximity to PG cells. [64]

Furthermore, PG growing within epithelial cells express and release the serine phosphatase SerB, which phosphorylates the NF κ B transcription factor inhibiting proinflammatory cytokine expression and neutrophil recruitment. [96]

LPS and gingipains released as OMW [100] (which diffuse away from places of PG colonization)

induce a stronger proinflammatory signaling release (of TNF, IL-12p70, IL-6, IL-10, and IFN- β) in contrast to LPS and gingipains recognized in direct contact with PG cells. For example they promote severe tissue destruction by NO release by macrophages, whereas NO secretion in proximity to PG is inhibited. ^{[4],[58],[99]}

Also IL-8 degradation depends on gingipain concentration. IL-8 (e.g. induced by LPS recognition) is released by the host cells as active IL-8-72aa and weakly active IL-8-77aa isoforms. IL-8-72aa is cleaved and inactivated by gingipains, thus contributing to local antiinflammatory signaling, whereas the rather inactive IL-8-77aa diffuses into the extracellular medium where it is cleaved over time by soluble gingipains (at lower concentration) to a hyperactive IL-8-69aa variant leading to strong neutrophil recruitment and ROS release. ^{[64],[81],[82],[83]}

In addition the cleavage of the complement factor C5, which at high concentration of virulence factors leads to inhibition of the complement pathway and enhanced PG survival (via the C5R-TLR2 crosstalk), leads at low concentrations to neutrophil recruitment and activation. ^[64]

1.5.6 Gingipains as the main immune modulator of PG

As previously mentioned in more detail (Chapter 1.5.3), gingipains (in combination with TLR2 activation by fimbriae and LPS) are necessary for the C5aR-CXCR4-TLR2 cross-talk involving the PI3K and integrin CR3 pathways, which results in reduced proinflammatory signaling of host cells, impaired phagocytosis and improved intracellular survival of PG.

On their own, gingipains are the major virulence factors for PG immune evasion and biofilm adhesion. Gingipains can enhance binding to epithelial cells^[45] likely via adhesion domains, as has been shown to facilitate coaggregation with *T. denticola*^[78].

To evade the immune system, RgpA gingipain can recruit the host's complement inhibitor C4BP, which then covers the surface of PG. C4BP inhibits binding to immunoglobins (classical complement pathway) and the recognition of bacterial polysaccharids and N-acetylglucosamines (lectin complement pathway) by the host complement system. ^{[4],[79]}

Furthermore, PG gingipains degrade several complement factors i.a. IG1 and IG3, IL-1b and TNF ^{[4],[64]} and downregulate mCD14 in host cells, a TLR coreceptor for fimbriae, which lowers the pathogen recognition and immune response against all bacterial species, leading to bacterial overgrowth and chronic inflammation^{[76],[84]}.

Gingipains also have transpeptidase activity, which transfers amino acids between protein chains. This feature is used to acquire nutrition from human hemoglobin and disturbs the hosts immune

tolerance towards its own proteins, which leads to autoimmune reactions and enhanced inflammatory reactions. ^{[4],[80]}

Furthermore gingipains activate NOD-like receptors (responsible for cell stress recognition) further inducing proinflammatory signaling that includes IL-1b and TNF. ^[86]

Secreted gingipains also can have a negative influence on the PI3K pathway supporting cell survival (in contrast to its activation described earlier (Chapter 1.5.3). This activation and host cell survival is achieved by binding of PG fimbriae (TLR2 activation) in cell-cell contact to leukocytes. In contrast, an attenuation of the PI3K pathway has been shown for epithelial cells by secreted gingipains, presumably by cleaving extracellular PRRs, which in homeostasis activate the PI3K pathway in those cell types. This effect of soluble gingipains is stronger in epithelial cells, as they do not express mCD14 for LPS and fimbriae recognition. This might lead to enhanced apoptosis inside epithelial tissue and supports periodontal tissue-breakdown. ^{[4],[98]}

Gingipains also activate "protease activated receptors" (PAR)-signaling by cleaving the G protein coupled receptors PAR1, PAR2 and PAR3, resulting in release of proinflammatory cytokines, neutrophil and monocyte recruitment, and T cell activation leading to increased epithelial tissue destruction and bone resorption as well as a positive feedback loop increasing the expression of PAR2. ^[85]

1.6 Periodontal disease progression and polymicrobial synergy

The above mentioned PG virulence factors enable PG to evade the hosts immune system upon arrival and during persistence in the oral cavity. They enable PG to colonize the commensal biofilm as well as to colonize the host leukocytes and epithelial cells. PG then suppresses proinflammatory cytokine release ("local chemokine paralysis") which leads to impaired neutrophil recruitment and insufficient anti-bactericidal activity in areas of PG infection - towards PG itself and all other bacteria. This leads to overgrowth of pathobionts and the bacterial microflora in total, which leads to an uncontrolled inflammation further supported by PG soluble virulence factors.

This leads to enhanced host tissue breakdown, which releases nutrients from host cells to the biofilm, favoring inflammophilic microorganisms and further dysbiosis. These "*inflammophilic organisms*"^[113] contribute with further virulence factors, disable the immune response, change the biofilm environment, enhance the inflammation and favors the progress of periodontal disease that started on arrival of the keystone-pathogen PG. The complex interplay within the dysbiotic biofilm has been termed "polymicrobial synergy and dysbiosis (PSD) model"^[2] and (for the interested

reader) is reviewed in further detail in "*Polymicrobial synergy and dysbiosis in inflammatory disease*"^[113]. In addition, after periodontal colonization, soluble PG virulence factors can be found within the bloodstream^[20] and PG is able to colonize migrating leukocytes which can be used as carriers to reach and settle in other tissues, causing additional PG associated diseases.^{[4],[7],[8]}

The key factors for PG to start periodontal disease are: I) PG induced reduced neutrophil recruitment as shown by host immune deficiencies or chemical induction which lead to periodontitis even in the absence of PG;^[57] II) an intact complement system of the host with intact TLR4 and NF-kB signaling^[114], TLR2 and C5aR^[74], C3aR or C5aR^{[6],[71]} which are required for a proinflammatory host response, as shown by receptor inhibition or deficiencies in animal models that prevent developing of periodontitis; III) host cell proinflammatory cytokines and their receptors as shown by cytokine inhibition;^[57] IV) antiinflammatory host cytokine IL-10, required to prevent complement over-expression and resulting in higher proinflammatory response induced by commensals;^[115] V) the commensals themselves, as shown by the lack of inducing periodontitis by PG itself in germ-free mice;^[6] and VI) gingipains of PG as the main immune-modulatory virulence factor, as shown by immunization of a primate model^[116] or by a gingipain deficient isogenic mutant of PG which prevents periodontal disease induction^{[6], [4],[24],[57]}

1.7 Bacterial secretion systems

All bacteria, archaea and eukaryotes use secretion systems (SS) to transport biomolecules across their membranes^[117].

In bacteria nine different types, T1-T9, have been described so far.^[118] They either translocate biomolecules in a single step from the cytoplasm into the extracellular medium or in two steps, first through the inner membrane (IM) into the periplasm via the Sec-, tat- or holin(-like)-pathways, and subsequently across the outer membrane (OM) outside of the cell. The transferred biomolecules are then attached to the OM, released into the extracellular environment or injected into a target cell.^[59]
[119]

Substrate transports over the IM via Sec, tat or holin(-like) proteins are universal throughout all domains of life^[59]. The tat-pathway transports substrates in their active conformation, whereas the Sec-pathway transports substrates before they are completely posttranslationally folded and holin-like proteins act as homooligomeric pores for endolysins.^{[120],[121]}

Substrates of SS include adhesins, proteases, lipases, endolysins, amidases or motility proteins, which are responsible for environmental adaption, nutrition acquisition, intercellular/-bacterial

communication, surface attachment, virulence factor delivery, immune system evasion, attack of other microorganisms or host cells. ^{[59],[122],[123]}

1.8 Type IX secretion systems

The type 9 secretion system (T9SS) transports its substrates across the OM. It is the most recent characterized bacterial SS^[59] with the first component PorT described in 2005^[124]. It is also referred to as Por secretion system and its essential components as Por- proteins, named after the ability to form black pigmented colonies containing porphyrin (hence *Por*) due to the secretion of gingipains when PG bacteria are cultured on blood agar^[125]. ^[126]

The T9SS is, unique to the *Bacterioides* phylum^[127] and is present in 62 % of their currently sequenced genomes ^{[128],[129]}. ^[59]

T9SS is involved in biofilm formation of periodontal pathogens/pathobionts such as *Capnocytophaga ochracea*^[130] as well as several species of the *Fusobacterium*, *Porphyromonas* and *Prevotella* genera^{[130],[123]}. Virulence factors transported by T9SS lead to periodontitis by PG as well as several diseases in birds and fishes such as columnaris disease, rainbow trout fry syndrome and cold-water disease^{[131],[132],[133]}. There are also T9SS independent of virulence factors. In *Flavobacterium johnsoniae* the T9SS secretes its motility adhesins and enables the bacterium for gliding motility^{[127],[134]} and in the symbiotic bacterium *Candidatus Paraporphyromonas polyenzymogenes* the T9SS is responsible for the secretion of cellulases into the rumen of ruminant animals ^[135].

The T9SS of PG secretes at least 30 proteins^[118] including the virulence factors gingipains (Kgp, RgpA, RgpB) and PPAD together with proteins which are responsible for their transport and modification. Secreted proteins get either bound to the cellular surface, enriched in OMV or released to the extracellular environment. ^[136]

The proteins that constitute the T9SS include translocation components, protein modification machinery, energetic components and regulatory proteins^{[59],[127],[137]}. Essential proteins, their location, complexes and proposed function are listed in Table TABLE 1.2.

PG protein and complexes	Localization	Proposed function	Reference
PorA	OM	translocon component	[118]
PorE	OM lipoprotein	peptidoglycan, PorP and PG1035 binding	[131]
PorF	OM	Unknown	[131]

PG protein and complexes	Localization	Proposed function	Reference
PorG	OM	Binds to PorK and PorN	[131]
PorK	OM lipoprotein, periplasm	-Binds to PorN	[131]
PorL	IM	Motor stator	[127]
PorM	IM	Rotor and shaft/pinion inherits a histidine kinase-, DNA gyrase B-, and HSP90-like ATPase motif, putative ATPase	[127],[131]
PorN	periplasm anchored via PorK with the OM	Binds to PorK	[138],[131]
PorP	OM	Binds to PorE, PG1035 and type B-CTD proteins putative O-deacylase	[131],[118],[122]
PorQ	OM	anchors PorZ, required for CTD cleavage and attachment to A-LPS	[90],[131]
PorR	Cytoplasm	putative aminotransferase activity, biosynthesis of A-LPS	[122]
PorT	OM	putative O-deacylase	[139],[122]
PorU	OM	- sortase cleaves the CTD signal - conjugates the cargo to the A-LPS anchor	[140],[141]
PorV	OM	type-A cargo shuttle protein	[90],[131]
PorW	OM lipoprotein	links together with PGN_1783 the translocon to the PorK/N complex	[118],[122]
PorX	cytoplasm	transcriptional regulation, response regulator, PglZ output domain of unknown enzymatic function	[142],[143],[126]
PorY	IM	transcriptional regulation, sensor histidine kinase	[142],[143],[126]
PorZ	Cell surface	part of attachment complex required for CTD cleavage and attachment to A-LPS	[144],[90],[118]
SigP	cytoplasm	transcriptional regulation, ECF sigma factor, activator of transcription	[145],[142],[143],[126]
SprA (PG)/ Sov (F. johnsoniae)	OM	- translocon component - bind to either PorV or Plug	[133],[131],[118]
Plug	OM	- translocon component	[133],[118]
Omp17	OM	- may be involved in PorU translocation	[96],[137]
PG1035 (PG W83)	OM	forms a complex with PorP and PorE, only P. gingivalis protein with a type B CTD	[118]
PGN_1783 (PG ATCC 33277)	OM	Sov binding	[118]

PG protein and complexes	Localization	Proposed function	Reference
PorL/M	IM anchored, PorL extends into cytoplasm, PorM extends into the periplasm	molecular motor	[127],[119],[118]
PorG/K/N	Ring shaped structure in periplasm and OM	-	[138],[118]
PorK, PorE and PorW	periplasm, associated with the outer membrane	-	[127],[146],[138],[131]
SprA(Sov), PorA, PorV and Plug	OM	translocon complex	[133],[118]
PorF, PorG, PorP, Sov, PorV, PorQ and PorT	OM	OM beta-barrel	[144],[147],[131]
PorP/E/PG1035	OM	unknown function	[118]
PorU, PorV, PorQ, PorZ	cell surface anchored by PorV and PorQ [31]	attachment complex to the cell surface	[90],[131],[118]
PorW and PGN_1783	OM	connects Sov with PorK/N ring	[118]

TABLE 1.2: Essential proteins, complexes and their functions in the T9SS of PG.

1.8.1 T9SS transport

A schematic representation of PG T9SS protein transport and modification is shown at Figure 1.6. Substrates of the T9SS carry an N-terminal signal peptide that is recognized by the Sec translocon. During the transport, this signal peptide is cleaved and the protein released into the periplasm. ^[132]

For further transport the T9SS substrates require an additional C-terminal signaling domain, type A or type B CTD (each around 100 amino acids)^{[148],[149],[133]}. Type A and type B CTD differ in interactions with T9SS components during transportation. ^[149] Substrates with a CTD get translocated to the OM. At the periplasm they interact with the PorK/L/M/N complex that is anchored in the IM by PorL. ^{[147],[150],[132]} PorE is suggested to be a scaffold protein of this periplasmatic complex. ^{[146],[137]} It binds to peptidoglycan and might act as a periplasmatic anchor for complex assembly ^[131]. To the OM the periplasmatic complex is attached by PorN. PorN builds up a ring shaped structure with PorG and PorK, which interacts with PorW and PorE at the periplasm. ^{[127],[146],[138],[118]} PorM is a putative ATPase^[122] and it is assumed that it transduces energy through PorK/N to SprA(Sov). ^[131]

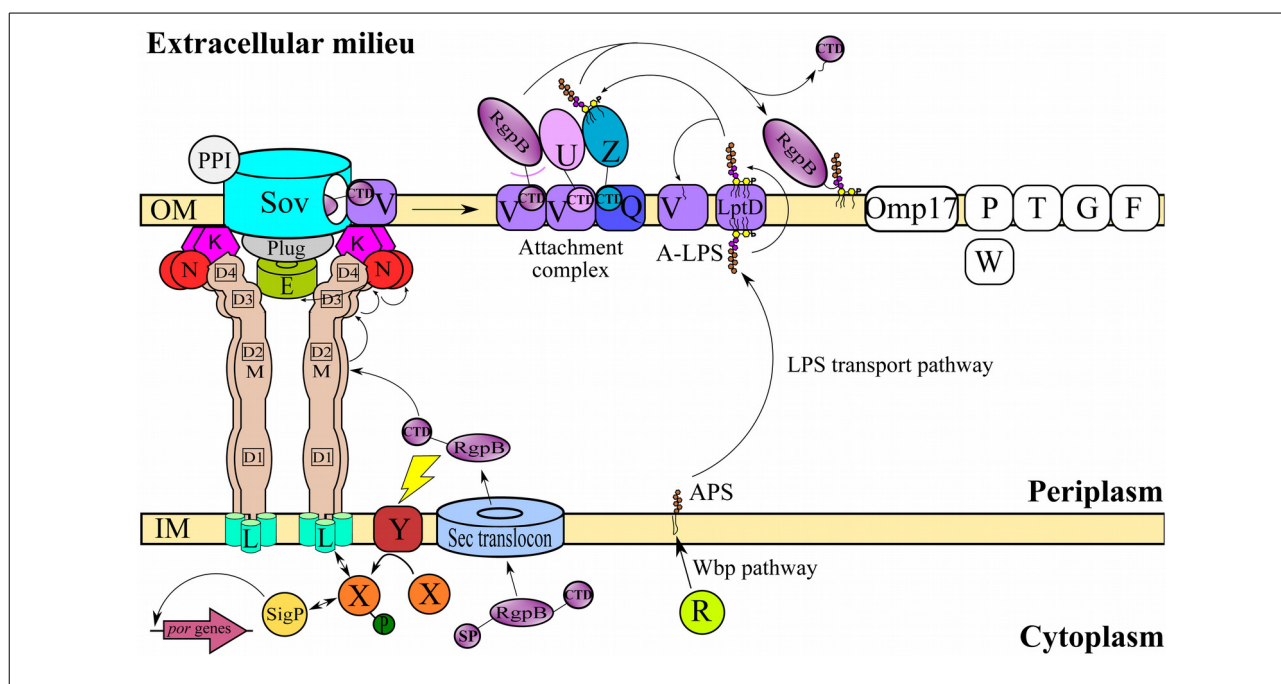


Figure 1.6: Transport, modification and regulatory components of the T9SS, from [132]

Substrate transport, taken RgpB gingipain as an example. Anionic-LPS (A-LPS) biosynthesis and transport (*wbp* pathway) have been simplified. Protein components of known function are colored. Type 9 secretion system (T9SS) Por-proteins are named by their acronym. [132] RgpB is transported over the inner membrane (IM) by the Sec translocon which recognizes its N-terminal signal peptide (SP). This SP is cleaved during IM translocation. Then RgpB is translocated through the periplasm where its C-terminal signaling domain (CTD) is recognized at the OM secretion pore including Sov and RgpB is transported through the outer membrane (OM). Outside the CTD gets bound by the attachment complex, where LPS are added and the C-terminal signaling domain is cleaved for secreted proteins. Regulation of the T9SS genes is initiated by a signal detected PorY at the IM, which autophosphorylates and subsequently phosphorylates PorX (X-P). X P then interacts with SigP, which activates the transcription of *por*-genes. [132]

SprA(Sov) is bound to the PorK/N ring by PorW and PGN_1783 and interacts directly with PorA, PorV and Plug [133],[118]. Spr(Sov), PorF, PorG, PorP, PorV, PorQ and PorT together build an OM beta-barrel secretion pore core complex^[131] whereas PorP builds a subcomplex with PorE and PG1035 of unknown function. It is proposed that, during the translocation, the SprA(Sov) alternates binding to PorA/PorV at the extra cellular side or by Plug on the periplasmatic side, guaranteeing unidirectional and specific transport. [118]

Once the substrates cross the OM through the secretion pore, they are bound by PorV, which shuttles them to the attachment complex, where they get further modified and anchored to the cell surface. [118]

The attachment complex consists of PorV/Q/U/Z, in which PorV/Q anchor the complex in the OM [90],[122]. PorU cleaves the CTD from the translocated substrates and anionic-LPS (APS), that attaches

substrates to the cell surface, is delivered to the new C-terminus by PorZ, which is part of the lipid A transport chain. ^{[59],[90],[122]}

PorQ and PorZ interact, and it is proposed that this PorQ/Z subcomplex binds to PorU/V when APS is bound by PorZ. ^[122]

APS is synthesized in the cytoplasm via the wbp pathway, which PorR as aminotransferase is part of ^{[125],[151],[132]}. PorP and PorT, which both carry a O-deacylase motive, may further modify substrates in the attachment complex. ^[122]

1.8.2 PG T9SS Regulation

T9SS gene expression is regulated by a two-component signal transduction system (TCS), which has evolved during adaption of virulence and metabolism onto sensing of external stimuli. ^[152] TCS are an ideal medical drug target because they are widespread in bacteria and archaea, are found in yeasts and plants, but are absent in animals ^[153]. Upon detection of an environmental signal the sensor histidine kinase (SHK) of a TCS autophosphorylates at a conserved histidine residue at its cytoplasmatic domain and subsequently phosphorylates its response regulator (RR) at a conserved aspartate of its receiver domain (RD). This leads either directly or indirectly controlled by the RR to a transcriptional activation and might depend on additional effector domains (ED) (83 % of RRs) that usually function as DNA binding domains (63 %), enzymatic domains (13 %), protein binding domains (3 %), RNA binding domains (1%) or fulfill other functions. ^{[153],[142],[154]}

The gene expression of the PG T9SS components is regulated by PorX, PorY and SigP. ^[126]

PorX and PorY constitute the two-component-system (TCS) to regulate the expression of at least 20 T9SS related genes indirectly via SigP, a extracytoplasmatic sigma factor. ^{[127],[147],[142],[132]}

The stimulus for PG T9SS is unknown but PorY, the sensor histidine kinase (SHK), is anchored in the IM by two transmembrane helices with a cytoplasmatic domain of 222 amino acids. ^{[127],[143],[59]}

The RR for PG T9SS is PorX (PG0928/PGN_1019) ^{[142],[126]}. It has been shown that PorX does not interact directly with DNA ^[143] but interacts with SigP which in turn binds the promoter region of T9SS related genes. ^{[142],[126]} However a recent project reported controversial results of PorX binding to *sigP* and *porT* promoters. ^{[160],[161]}

PorX has an N-terminal conserved receiver domain and in addition an C-terminal domain which belongs to the PglZ family (pfam: PF08665). ^[143] The PglZ domain is a putative phosphatase domain, ^{[154],[155],[156]} which in turn belongs to the alkaline phosphatase superfamily of enzymes. This enzymatic superfamily includes a wide spectrum of hydrolytic enzymes, not only phosphatases, but

also arylsulfatases, pyrophosphatases/phosphodiesterases and phosphomonoesterases. ^[157] Pgl stands for "phage growth limiting" (Pgl) as the Pgl-proteins were first reported being involved and essential for phage resistance in *Streptomyces coelicolor*. ^[158] PglZ domains are widespread upon prokaryotic life, found so far in 1357 species of bacteria and archaea. ^[159]

Apart from the PorX/PorY TCS, four other TCS have been identified in PG: FimR/FimS system which is related to fimbriation; GppX, a SHK-RR hybrid, involved in gingipain secretion; the RR RprY involved in stress response; and the PGN_0752/ PGN_0753 system involved in haemin assimilation. ^[142]

Apart from its RR function and interaction with SigP for gene expression, an other interaction of PorX has been found with PorL, ^[143] an IM protein and the cytoplasmatic part of the PorM/I molecular motor complex. ^{[127],[119],[118]} This contact might keep PorX in close proximity to PorY at the IM. In *Flavobacterium johnsoniae* it has been shown that phosphorylated PorX binds PorL which reverses the rotation of the T9SS which, in this species, acts as a rotary motor for gliding motility, and changes direction. Since PG is a non-motile bacterium, an other function needs to be assumed, as interaction of pPorX with PorL could influence the opening of the translocon complex by interaction via PorM. ^[122]

1.9 Aims of this Project

The objective of this thesis was the process of crystallization, optimization and structure solution of PorX. If crystallization trials with PorX alone would have failed, there were the options to cocrystallize PorX in complex with PorY or PG1660 proteins. Alternatively PorY or PG1660 could have been crystallized and characterized. Once structures of any PG protein were solved, the focus would have been their structure function relationship and its role in transcriptional activation of PG virulence factors and of cause publishing the results in an high impact factor journal.

In the case of successful crystallization and structure solution of PorX, the focus should have been, and in fact it was, the characterization of the PglZ domain of PorX, its structural features and functions of this previously uncharacterized protein domain and how it is related to the function of PorX as a response regulator. We therefore solved the PorX structure with and without substrates, analyzed its mechanism of dimerization, activation and enzymatic activities. We also highlighted similarities and differences to other proteins of the alkaline phosphatase superfamily and to its most similar relatives, the PglZ family of proteins which we have modeled by AlphaFold predictions and discussed how PorX is involved in transcriptional regulation.

Materials and Methods

NOTES.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

2. Materials and Methods

2.1 Chemicals and Reagents

Chemicals in our lab were used in "*pro analysis*" purity grade and purchased from ThermoFisher Scientific, ITW Reagents (PanReac AppliChem), MerckMillipore, Pronadisa, Roche or Sigma-Aldrich. Water was deionized by a Millipore Synergy UV water purification system, hereinafter referred to as Millipore water (MPW). Beryllium sulfate, sodium fluoride and acetyl phosphate were provided by our collaboration partner as 0.25 M BeSO₄, 1M NaF and 0.1 M acetyl phosphate stock solutions in MQW and were stored at -20 °C until use. Selenomethionine used by our collaboration partner for selenomethionine derivatized protein expression was purchased from Chem-Impex International, Inc..

2.1.1 Chromatography columns and materials

Methods	Material	Supplier
GST-Affinity Chromatography	Glutathione Sepharose 4 Fast Flow	Cytiva
Analytical Size-Exclusion-Chromatography (SEC)	Superdex 200 Increase 10/300 GL	Cytiva
SEC-Multi-Angle-Laser-Light-Scattering (SEC-MALLS)	Superdex 200 Increase 10/300 GL	Cytiva
Preparative Size-Exclusion-Chromatography (SEC)	HiLoad 16/60 Superdex 200 pg	Cytiva

2.1.2 Buffers and Solutions

Name	Method	Composition
PBS	Protein Purification-Affinity Chromatography	12 mM Na ₂ HPO ₄ , 2 mM KH ₂ PO ₄ , 137 mM NaCl, 3 mM KCl,
Dialysis-Buffer	Protein Purification-Pre-IEC	20 mM Bis-Tris pH 6.5 (RT)
IEC-low salt buffer	Protein Purification-IEC	20 mM Bis-Tris pH 6.5 (RT)
IEC-high salt buffer	Protein Purification-IEC	20 mM Bis-Tris pH 6.5 (RT), 1M NaCl
SEC-Buffer	Protein Purification-SEC	5 mM Tris pH 7.5, 150 mM NaCl
Crystallization-Buffer	Protein Purification-SEC pre-PorX crystallization	5 mM Tris pH 7.5, 50 mM NaCl
Phosphorylation-Buffer	PorX <i>in vitro</i> phosphorylation	50 mM Tris pH 8.0 (RT), 150 mM NaCl, 10 mM MgCl ₂ , 20 mM acetyl phosphate
BeF ₃ -Buffer	PorX BeF ₃ induced dimerization	10 mM Tris pH 8.0 (RT), 150 mM NaCl, 5 mM BeSO ₄ , 30 mM NaF, 15 mM MgCl ₂
MALLS-Buffer	SEC-MALLS (without zinc)	50 mM Tris pH 8.0, 150 mM NaCl and 10 mM MgCl ₂
MALLS-Zn-Buffer	SEC-MALLS (with zinc)	"MALLS-Buffer"+0.43 μM ZnCl ₂

2.1.3 Media

Name	Use	Composition
LB	Expression media	5 g/L yeast extract, 5 g/L sodium chloride, 10 g/L tryptone in MPW, autoclaved, 100 µg ampicilin added
LB - Agar	Bacterial (colony) separation	LB including 1.5 % agar and 100 µg ampicilin, autoclaved
SOC	Heat shock transformation	10 mM NaCl, 2.5 mM KCl, 10 mM MgCl ₂ , 20 g/L trypton, in MPW, 10 mM glucose, autoclaved
SeMet-Minimal Media	Protein expression media for selono methionine derivative PorX	Per liter: 8 g Na ₂ HPO ₄ , 4 g KH ₂ PO ₄ , 0,5 g NaCl, 0,5 g NH ₄ Cl, 5 µg EDTA, 8.3 µg FeCl ₃ ·6H ₂ O, 0.84 mg ZnCl ₂ , 0.13 mg CuCl ₂ ·2H ₂ O 0.10mg CoCl ₂ ·6H ₂ O, 0.10 mg H ₃ BO ₃ , 0.16 mg MnCl ₂ ·6H ₂ O, 4 g Glucose, 50 mg of each amino acid exept methionine, 1 mM MgSO ₄ , 0.3 mM CaCl ₂ , 10 mg Biotine, 10 mg Thiamine, 10 mg selenomethionine; 30 min before induction dilution with 100 ml MQW on 900 ml media, including additional 50 mg selenomethionine, 25 mg of each other amino acid, in addition 33 mg of Thr, Lys, Phe and 16 mg/ile, leu and val, 5 mg biotine, 5 mg Thiamine and 2g of glucose.

2.1.4 Plasmids

Name	Features	Fusion-tag	Supplier
pGEX-6-P1	pBR322 ori, tac promoter, lac operator, ribosome binding site,, PreScission protease site, MCS, AmpR	GST-tag	Cytiva

2.1.5 Primer

Name	Sequence 5'-3'	Use
PorX_pGEX_BamHI_F	acgggatccatggaaaaaacatg	wt PorX amplification from PG w83 genomic DNA extract
PorX_pGEX_Sall_R	caggtcgacttacttggttg	
PorX_D58A_F	atcgtattcctcgccgagaacatgcccgcatcgg	D58A mutation primers
PorX_D58A_R	ctcggcgaggaatacgaatgtcgaagtcgtattttgcacggc	
PorX_I134S_F	cagcagcacagtatcagcagcgaaaccacgaac	I134S mutation primers
PorX_I134S_R	cagggttttttgagcgacaggaggagctgattcgga	
PorX_T272A_F	ctgccgacagcgcccaatatgcacgcaatgccatctt	T272A mutation primers
PorX_T272A_R	ttgggccgctgtcggcaggatggacagggtacatatcttctcgaa	
PorX_M99K_D109A_F	aaaacacaagccatcgggggaagatagcggcctacatcaaa ccggtgaat	M99K plus D109A mutation primers
PorX_M99K_D109A_R	ggccgctatcttccccgatggctgtgtttgatatgctcctcttgcgtttt	
PorX_D361A_H365A_F	tggccatgatgtcggccgctcgtactgatagcaagatgatt	D361A plus H365A mutation primers
PorX_D361A_H365A_R	gcgccgacatcatggccacgaagttcaggactatcacat	
PorX_S385E_S398E_F	gaactgacgaaggaatggttcaagcattcgactacctac	S385E plus S398E mutation primers
PorX_S385E_S398E_R	ttccttcgtcagttcgcataggctgcttcgttg	

2.1.6 Bacterial strains

Name	Method	Genotype	Supplier
<i>E. coli</i> DH5 α	Vector amplification	F ⁻ ϕ 80/ <i>lacZ</i> ΔM15 Δ(<i>lacZYA-argF</i>)U169 <i>recA1 endA1 hsdR17</i> (rK ⁻ , mK ⁺) <i>phoA supE44 λ⁻thi-1 gyrA96 relA1</i>	ThermoFisher SCIENTIFIC
<i>E. coli</i> BL21(DE3)	Protein expression	F ⁻ <i>ompT hsdS_B</i> (rB ⁻ , mB ⁻) <i>gal dcm</i> (DE3)	ThermoFisher SCIENTIFIC

2.1.7 Informatics

Type	Name	Use	Reference
Crystallographic software packages	ccp4i	.blend for data analysis .PHASER-EP pipeline for experimental phasing .PHASER for molecular replacement .MOLREP for molecular replacement .PISA_qt for interfaces, surfaces and assembly analysis	[1], [2], [3], [4], [5], [6], [7], [8]
	ccp4i2	Buccaneer for initial model building	[9], [8]
	phenix	.autobuild for initial model building .refine for refinements .maps for map calculations	[10], [11], [12]
	buster	Electron density map calculations	[13]
	staraniso	aP_scale for final scaling, truncation	[14]
	SHELX C/D/E	Initial structure solution	[15]
Bioinformatics and biomolecule visualization software package	chimera	Graphical display of proteins, electron densities and molecular surfaces	[16], [17]
Molecular visualization software	PyMol	Graphical display of protein structures, and electron densities	[18]
Model-building tools	COOT	Model building and real space refinement	[19]
Data Collection software	MXCuBE	Beamline environment	[20], [21]
Diffraction image viewer	Adxv	Beamline environment	[22]
Data reduction software package	XDS	Data reduction	[23]
	XSCALE	Initial scaling and formatting	
SAXS software	Bioxtas RAW	SAXS data analysis Mw estimations	[24],[25], [26], [27], [28]
	ATSAS 3.0	Inverse Fourier Transformation	[29]
	GNOM		[30]
	DENSS	<i>Ab initio</i> electron density reconstruction	[31]
	SREFLEX	Refinements and model validation	[32]
	CRY SOL		[33]
English dictionaries and corrections	LEO: English-German online dictionary, LEO GmbH (2022); Google Translate, Google (2022); OpenAI, (2022)		

2.1.8 Commercial Kits and tools

Method	Kit/Tools	Supplier
Linear DNA purification	QIAQUICK PCR Purification Kit or QIAGEN Gel Extraction Kit	QIAGEN
Circular DNA purification	QIAprep Spin Miniprep Kit	QIAGEN
DNA Ligation	NEBuilder HiFi DNA Assembly Cloning Kit	New England Biolabs (NEB)
Crystal manipulations	CrystalProbe	Hampton
Crystal fishing/mounting	Magnetic cryo wand	Hampton
Cryo-cooling of crystals	Cryo pins and vials	Hampton, Molecular Dimensions
	Cryo loops	Hampton, MiTeGen

2.1.9 Equipment

Equipment	Use	Model	Supplier
Camera, objective lence, macro extensions	Crystal imaging	α5100, SEL-50F18, MK-S-AF3A 10mm+16mm	Sony, Sony, Meike
Microscope, objective lense, Eyepiece	Crystal manipulation, harvest, imaging	SMZ1000, P-Plan Apo 1x, C-W 10x (F.N. 22)	Nikon
FPLC+ autosampler.	MALLS	Prominence	SHIMADZU
On-line MALLS detector		DAWN-HELEOS-II	Wyatt Technology
On-line differential refractometer		OptiLab T-rEX	
Spectrophotometer	protein concentration measurements for MALLS experiments	GE NanoVUE spectrophotometer	Cytiva
	Measurement of DNA and protein concentrations (except MALLS)	Nanodrop 1000	ThermoFisher SCIENTIFIC

2.1.10 Protein Samples

Name	Description
Dim_PorX	In vitro dimerized PorX
Dim_PorX_SeMet	In vitro dimerized PorX with SeMet substitution
Mono_PorX	Monomeric PorX
Mono-PorX_SeMet	Monomeric PorX without SeMet substitution
BeF3_PorX	Dimerized by BeF ₃
BeF3_PorX_SeMet	Dimerized by BeF ₃ with SeMet substitution
D58A_PorX	Receiver domain (RD) mutant at the phosphorylation site
M94K/D104A/I129A_PorX	Triple mutant at RD dimerization interface
S385E/S389E_PorX	Double mutant at PglZ dimerization interface
D361A/H365A_PorX	Double mutant at PglZ HH conformation zinc ("Zn1") binding site
PorY185_PorX	N-terminal truncated (aa1-184) version of PorY
PG1660_PorX	PG extracytoplasmatic sigma factor PG1660

2.1.11 Crystallization Screens

Name/ Screen + extras	Date	Temp.	Protein	Conditions	Wells, Concen- tration, drop ratio [x100 nl]
PAC1 PAC2 PAC3 TOP96 PROPLEX	Sept. 2016 (set up before project takeover)	20 °C	Dim_PorX_ SeMet, Mono-PorX_SeMet, BeF3_PorX, BeF3_PorX_SeMet	Initial sparse matrix	4x 5x 96 wells 10 mg/ml
Grid_01 PEG_3350 (PAC1_C12-Grid)	25.01.17	20 °C	BeF3_PorX	Grid: Tris pH 7-9, PEG 3350 6-20%; 20 mM MgSO ₄	48 wells 10 mg/ml, 5:5
Grid01_ PEG_8000 (PAC1_C12-Grid)	25.01.17	20 °C	BeF3_PorX	Grid: Tris pH 7-9, PEG 8000 6-20%; 20 mM MgSO ₄	48 wells 10 mg/ml, 5:5
Grid01_ PEG_3350 (PAC1_C12-Grid)	26.01.17	20 °C	BeF3_PorX_SeMet	Grid: Tris pH 7-9, PEG 3350 6-20%; 20 mM MgSO ₄	48 wells 8.7 mg/ml, 7:7
Grid01_ PEG_8000 (PAC1_C12-Grid)	26.01.17	20 °C	BeF3_PorX_SeMet	Grid: Tris pH 7-9, PEG 8000 6-20%; 20 mM MgSO ₄	48 wells 8.7 mg/ml, 7:7
Opt_A	27.01.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 5.3 mg/ml 2:2; 96x 5.3 mg/ml 2:1
Opt_A	27.01.17	20 °C	BeF3_PorX_SeMet, buffer-exchanged to 50 mM NaCl buffer after concentration	See Chapter 7.10	96x 5.94 mg/ml
Opt_Grid0217	08.02.17	20 °C	BeF3_PorX_SeMet	Upscaling A7-D8 of Opt_A	8x 7.7 mg/ml 20:20; 8x 5.0 mg/ml 20:20
Opt_A +10% glycerol	08.02.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 7.8 mg/ml 5:5 96x 7.8 mg/ml 3:4
Opt_A +10% glycerol	08.02.17	15 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 7.8 mg/ml 3:3 96x 7.8 mg/ml 3:4
Opt_B	13.02.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 6 mg/ml 3:3 96x 6 mg/ml 3:4
Opt_B	13.02.17	4 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 6 mg/ml 3:3 96x 6 mg/ml 3:4
Opt_C	15.02.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 5 mg/ml 96x 2.8 mg/ml
Opt_C	15.02.17	4 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 5 mg/ml 96x 2.8 mg/ml

Name/ Screen + extras	Date	Temp.	Protein	Conditions	Wells, Concen- tration, drop ratio [x100 nl]
Opt_D	13.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 5.1 mg/ml
Opt_E	13.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 5.1 mg/ml
Opt_F	13.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 5.1 mg/ml
Opt_F	13.03.17	4 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 5.1 mg/ml
Opt_G	13.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 5.1 mg/ml
Opt_D	17.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 3.4 mg/ml 96x 2.5 mg/ml
Opt_E	17.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 3.4 mg/ml 96x 2.5 mg/ml
Opt_F	17.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 3.4 mg/ml 2:2; 96x 2.5 mg/ml 1:2
Opt_G	17.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 3.4 mg/ml 96x 2.5 mg/ml
BeF3_ 48opt_ 01	31.03.17	20 °C, later 37 °C	BeF3_PorX_SeMet	Glycine pH 8.0, 4 % PEG 3350 optimization grid: different drop ratios 1:5-5:1, reagents lab vs. platform, additives: MgCl ₂ , proline, Urea, KCl	48x 5.2 mg/ml
BeF3_ 48opt_ 02	15.02.17 13.03.17 15.03.17 17.03.17 31.03.17 13.04.17 20.4.17 31.04.17	20 °C	BeF3_PorX_SeMet	Crystallization hit upscaling, feeding trials, seeding trials	48x various concentrations and drop sizes up to 5 ul:5ul
Opt_H +50 mM MgCl ₂	29.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	24x 2.15 50:50
Opt_H +50 mM MnCl ₂	29.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	24x 2.15 50:50
Opt_H +50 mM MgCl ₂ +100 mM NaCl	29.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	24x 2.15 50:50
Opt_H +40 mM MgCl ₂ +20 mM proline +40 mM urea	30.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	24x 1.6 100:100
Opt_H +20 mM MgCl ₂	30.03.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	24x 1.6 160:20
Opt_I	13.04.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	48x 5.0 mg/ml
Opt_I	13.04.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	48x 2.5 mg/ml

Name/ Screen + extras	Date	Temp.	Protein	Conditions	Wells, Concentration, drop ratio [x100 nl]
Opt_J	20.04.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 4 mg/ml 96x 3 mg/ml
Opt_J	20.04.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 2.25 mg/ml 96x 1,6 mg/ml
Opt_J	20.04.17	37 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 4 mg/ml 96x 3 mg/ml
Opt_J	20.04.17	37 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 2.25 mg/ml 96x 1,6 mg/ml
Opt_J +0.1 M sodium formate	20.04.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 4 mg/ml 96x 3 mg/ml
Opt_J +0.1 M sodium formate	20.04.17	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 2.25 mg/ml 96x 1,6 mg/ml
BeF3_48opt_03	21.04.17	20 °C	BeF3_PorX_SeMet	Glycine pH 8.0, 4 % PEG3350	48x4 mg/ml, different drop ratios
BeF3_24opt_01	04.0617	20 °C	BeF3_PorX_SeMet	Opt_H_B6 reproduction screen	24xVarious drop sizes 250:50-50:250 conc. 2.3, 3.5 or 4.6 mg/ml
Opt_J-G5 1+2+3	20.06.17	20 °C	BeF3_PorX_SeMet	96x Opt_J_G5 control for "optisalt" screening	96x 5.1 mg/ml 96x 3.7 mg/ml 2x96x 2.8 mg/ml 2x96x 2.25mg/ml
Optisalt 1+2+3	20.06.17	20 °C	BeF3_PorX_SeMet	OptJ-G5 condition +0.1x PCB crystallization platform salt additive solutions ("Optisalt" See 7.10)	96x 5.1 mg/ml 96x 3.7 mg/ml 2x96x 2.8 mg/ml 2x96x 2.25 mg/ml
Opt_A	03.08.17	20 °C	Mono-PorX_SeMet	See Chapter 7.10	96x 4.9 mg/ml 96x 2.9 mg/ml
Opt_B	03.08.17	20 °C	Mono-PorX_SeMet	See Chapter 7.10	96x 4.9 mg/ml 96x 2.9 mg/ml
Opt_C	03.08.17	20 °C	Mono-PorX_SeMet	See Chapter 7.10	96x 4.9 mg/ml 96x 2.9 mg/ml
Mono_Grid_01	03.08.17	20 °C	Mono-PorX_SeMet	0-6 % PEG 3350 grid, in either glycine pH 8.0 or Hepes pH7.5; 100 mM sodium formate in all conditions	48x 3 or 5 mg/ml
Opt_D	10.08.17	15 °C	Mono-PorX_SeMet	See Chapter 7.10	96x 12.8 mg/ml 96x 8.0 mg/ml

Name/ Screen + extras	Date	Temp.	Protein	Conditions	Wells, Concentration, drop ratio [x100 nl]
Opt_E +20mM MgSO ₄	10.08.17	15 °C	Mono-PorX_SeMet	See Chapter 7.10	96x 12.8 mg/ml 96x 8.0 mg/ml
Opt_F	10.08.17	15 °C	Mono-PorX_SeMet	See Chapter 7.10	96x 12.8 mg/ml 96x 8.0 mg/ml
Opt_G +20mM MgSO ₄	10.08.17	15 °C	Mono-PorX_SeMet	See Chapter 7.10	96x 8.0 mg/ml 96x 2.8 mg/ml
Opt_J	05.09.17	20 °C	Mono-PorX_SeMet	See Chapter 7.10	96x 14.8 mg/ml
Opt_J +111 mM lithium citrate	05.09.17	20 °C	Mono-PorX_SeMet	See Chapter 7.10	96x 14.8 mg/ml
PAC1	13.10.17	20 °C	PorY185 +/- 1mM ADP	Sparse matrix screen	96x "20x" concentrate from collaboration partner
PAC2	13.10.17	20 °C	PorY185 +/- 1mM ADP	Sparse matrix screen	96x "20x" concentrate from collaboration partner
PAC3	13.10.17	20 °C	PorY185 +/- 1mM ADP	Sparse matrix screen	96x "20x" concentrate from collaboration partner
TOP96	13.10.17	20 °C	PorY185 +/- 1mM ADP	Sparse matrix screen	96x "20x" concentrate from collaboration partner
Proplex	13.10.17	20 °C	PorY185 +/- 1mM ADP	Sparse matrix screen	96x "20x" concentrate from collaboration partner
PAC1	17.10.17	20 °C	PG1660	Sparse matrix screen	96x 8.23 mg/ml
PAC2	17.10.17	20 °C	PG1660	Sparse matrix screen	96x 8.23 mg/ml
PAC3	17.10.17	20 °C	PG1660	Sparse matrix screen	96x 8.23 mg/ml
TOP96	17.10.17	20 °C	PG1660	Sparse matrix screen	96x 8.23 mg/ml
PROPLEX	17.10.17	20 °C	PG1660	Sparse matrix screen	96x 8.23 mg/ml
YOpt_A	21.01.18	20 °C	PorY185	Various	96x 2.25 mg/ml (50 mM NaCl) 96x 3.85 mg/ml (250 mM NaCl)
YOpt_A	24.01.18	20 °C	PorY185 + 1 mM ADP	Various	96x 2.25 mg/ml (50 mM NaCl) 96x 3.85 mg/ml (250 mM NaCl)
Opt_J	25.01.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 3.3 mg/ml 96x 2.25 mg/ml

Name/ Screen + extras	Date	Temp.	Protein	Conditions	Wells, Concen- tration, drop ratio [x100 nl]
Opt_J	25.01.18	20 °C	BeF3_PorX_SeMet	Equal upper screen, but protein solutions+ 1mM ZnCl ₂ and centrifuged	96x 0.77 mg/ml 96x 0.41 mg/ml
Opt_J	01.02.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 2.0 mg/ml 96x 1.3 mg/ml
Opt_J +0.1 mM ZnCl ₂	01.02.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 2.0 mg/ml 96x 1.3 mg/ml
YOpt_A	05.02.18	20 °C	PorY185	See Chapter 7.10	96x 4.9 mg/ml 96x 2.45 mg/ml
YOpt_A	05.02.18	20 °C	PorY185 + 1 mM AMPNP	See Chapter 7.10	96x 4.9 mg/ml 96x 2.45 mg/ml
Opt_J”+”(15 mM MgCl ₂ , 30 mM NaF, 5 mM BeSO ₄ , 0.1 mM ZnCl ₂)	08.02.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 12.8 mg/ml 96x 8.4 mg/ml
Opt_J+	08.02.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 4.5mg/ml 96x 2.3 mg/ml
Opt_K	28.03.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 5.0 mg/ml 96x 4.0 mg/ml
Opt_K	28.03.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 3.0 mg/ml 96x 1.5 mg/ml
Opt_K, +0.5 mM GMP	28.03.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 4.0 mg/ml 96x 3.0 mg/ml
Opt_K, +0.1 mM GTPγS	28.03.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 4.0 mg/ml 96x 3.0 mg/ml
BeF3_48opt_04(“K,+”)	12.04.18	20 °C	BeF3_PorX_SeMet	Optimization cacodylate pH7.0, 0.1 M sodium formate, 8 % HMW PEG, different concentrations, with and without GMP or GTPγS	48x various concentrations 2.2-4.4 mg/ml
BeF3_48opt_05(“K,++”)	12.04.18	20 °C	BeF3_PorX_SeMet	BeF3_48opt_03(“K,+”) with extra 0.2 mM ZnCl ₂	48x various concentrations 2.2-4.4 mg/ml
Opt_L+ (no Zn)	25.06.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 13 mg/ml 96x 10 mg/ml
Opt_L+ (no Zn)	25.06.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 7 mg/ml 96x 4 mg/ml
Opt_L+ (0.5mM Zn)	25.06.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 13 mg/ml 96x 10 mg/ml
Opt_L+ (0.5mM Zn)	25.06.18	20 °C	BeF3_PorX_SeMet	See Chapter 7.10	96x 7 mg/ml 96x 4 mg/ml
BeF3_24opt_02(“Opt_M”) hanging drop	various	20 °C	BeF3_PorX_SeMet	Upscaled crystal hits	various

Name/ Screen + extras	Date	Temp.	Protein	Conditions	Wells, Concen- tration, drop ratio [x100 nl]
BeF3_24opt_02("Opt_M") sitting drop	various	20 °C	BeF3_PorX_SeMet	Upscaled crystal hits	various
Opt_J+	26.11.18	20 °C	T272A PorX	See Chapter 7.10	96x 12 mg/ml 96x 8 mg/ml
Opt_J+ (pGpG)	26.11.18	20 °C	T272A PorX +0.5 mM pGpG	See Chapter 7.10	96x 12 mg/ml 96x 8 mg/ml
Opt_J+ (c-di-GMP)	26.11.18	20 °C	T272A PorX +0.5 mM c-di-GMP	See Chapter 7.10	96x 12 mg/ml 96x 8 mg/ml
Opt_L+ (0.5mM Zn)	26.11.18	20 °C	T272A PorX	See Chapter 7.10	96x 12 mg/ml 96x 8 mg/ml
Opt_L+ (0.5mM Zn)	26.11.18	20 °C	T272A PorX +0.5 mM pGpG	See Chapter 7.10	96x 12 mg/ml 96x 8 mg/ml
Opt_L+ (0.5mM Zn)	26.11.18	20 °C	T272A PorX +0.5 mM c-di-GMP	See Chapter 7.10	96x 12 mg/ml 96x 8 mg/ml
Opt_J+	19.12.18	20 °C	M94K/D104A/I129A_PorX	See Chapter 7.10	96x 7 mg/ml 96x 4 mg/ml
Opt_L+	19.12.18	20 °C	M94K/D104A/I129A_PorX	See Chapter 7.10	96x 7 mg/ml 96x 4 mg/ml
Opt_J+	19.12.18	20 °C	D58A_PorX	See Chapter 7.10	96x 10 mg/ml 96x 6 mg/ml
Opt_L+	19.12.18	20 °C	D58A_PorX	See Chapter 7.10	96x 10 mg/ml 96x 6 mg/ml

2.2 Molecular biology

All molecular biological methods were performed by our collaboration partners Zuza Nowakowska and Mariusz Madej from the group of Prof. Dr. Jan Potempa (Jagiellonian University in Kraków (Poland), Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology). Wt-PorX was amplified by PCR (method see 2.2.1) from PG W83 genomic DNA extract as template. Primers used (Chapter 2.1.5) were "PorX_pGEX_BamHI_F" (forward) and "PorX_pGEX_SalI_R" (reverse) which added BamHI and SalI restriction sites for subsequent cloning. PorX mutants were amplified by mutational primers from pGEX-6-P_wt-PorX.

DNA constructs were analyzed by agarose gel electrophoresis and purified by commercial DNA purification kits using spin columns (methods see 2.2.2). Concentration measurements were performed photometrically at their absorbance at 260 nm. PCR products as well as the plasmid pGEX-6-P was digested with BamHI and SalI restriction enzymes, purified and enzymatically

ligated (methods see 2.2.3). PorX mutant vectors were produced by PCR from PorX-pGEX-6-P1 vector template using NEBuilder HiFi DNA Assembly Cloning Kit and appropriate PCR primer design (method see 2.2.4). Vectors then were transformed into DH5 α cells for vector amplification and send for sequencing. Inserts of produced pGEX-1p vectors for PorX and its mutants were sequenced by either by GENOMED or EUROFINs sequencing facilities using pGEX 5' and pGEX 3' sequencing primers, that bind at the GST-tag C-terminal sequence and terminator region flanking the PorX inserts on the plasmid (method see 2.2.5).

2.2.1 Polymerase Chain Reaction (PCR)

The polymerase chain reaction, introduced by Saiki in 1987,^[34] is an enzymatic method to amplify DNA with oligonucleotide primers and the use of a thermostable DNA polymerase. The primers need to be specific and complementary to the opposite DNA strand ends flanking the target sequence. They can add additional sequences to both ends e. g. restriction sites or additional codons for amino acids or stop codons, which can also be used to insert mutations into the final amplified constructs. After an initial heat denaturation of a DNA template, a target gene is amplified by repeating cycles of 1.) DNA denaturation, 2.) primer annealing with the template and 3.) primer elongation by the DNA polymerase under the use of dNTPs. Yields and specificity at different temperatures or reaction times can be analyzed by agarose gelelectrophoresis.^{[35], [34]}

2.2.2 DNA purification and analysis

Agarose gel electrophoresis uses a agarose gel matrix as firstly described by Aaij & Borst in 1972^[36] to separate nucleic acids by electrophoresis, which was described by Olivera in 1964^[37]. At non-acidic pH nucleic acids have a negative charge based on their phosphate backbone, which is approximately proportional to their number of base pairs and overall mass. Thus nucleic acids will run in an electric field towards the anode with smaller species running faster through a porous matrix than larger ones (given a similar shape/secondary structure). Separated nucleic acids can be visualized under UV light with a transilluminator using fluorescent dyes binding noncovalently to the nucleic acids. DNA molecules can be analyzed on their molecular weight and quantity using control samples/standards. Fractions of similar sizes also can be cut out from gels e.g. with a scalpel and extracted for further use e.g. with DNA purification spin columns.^{[36], [35]}

Spin column DNA purifications from aqueous solution or agarose gels are based on the DNA solubility and adhesion properties to a silica membrane in different buffered solutions. Different surface areas of silica membranes are more suitable for different DNA length and yields (thus

different membranes for PCR fragments or circular plasmid DNA). DNA adhesion occurs in both cases under high ionic strength of chaotropic salts, contaminations are washed from the membrane using ethanol containing buffers and DNA elution takes place with MQW or a low salt buffer solution. ^[38]

2.2.3 Restriction digestion and ligations

Restriction digestion uses restriction enzymes as discovered by and applied to molecular genetics by Werner Arber, Daniel Nathans and Hamilton O. Smith in the 1960-70s:^[39] Type II restriction enzymes used for molecular cloning have a specific palindromic DNA recognition sequence and cleave the DNA double strand specific within this sequence. DNA ligation utilizes DNA ligases as firstly reviewed by Lehman 1974^[40]: DNA ligases catalyse the phosphodiester bond formation between the 3' hydroxy and 5' phosphate ends of two DNA strands, by using either ATP or NAD⁺ to form a reactive ligase-adenylate complex. Higher efficiencies and reactivities can be accomplished by using "sticky end" ligation, which means designing DNA ends with a single stranded 3' overhang of one DNA double strand, complementary to the 5' overhang of the other. This can be done choosing type II restriction enzymes with this particular restriction behavior during the restriction digestion step. ^{[35], [40]}

2.2.4 Site-directed mutagenesis

Gene mutations can be introduced by primer overhangs with additional 18-20 bp complementary sequences to flanking insert or plasmid sequences. The complementary fragments will be ligated with restriction digested plasmid in an isothermal DNA assembly step at 50 °C by incubation with DNA-ligase and DNA-polymerase for one hour. ^{[41], [42]}

2.2.5 DNA sequencing

Modern sequencing methods for plasmid DNA inserts up to around 1 kbp are based on the chain-termination method (also referred to as Sanger method) first published by Frederick Sanger in 1977^[43]. It uses 4 PCR reactions in presence of a small amount (to reach incomplete incorporation) of one ddNTP each. Whenever one of these nucleotides gets incorporated into the nascent DNA chain, the polymerase chain reaction terminates because of the lack of the terminal 3'-hydroxy group. This leads to a mixture of DNA fragments with a length distribution equal to the distances of the labeled nucleotide (in its reaction) to the beginning of the primer sequence. These distances then can be determined by measuring the fragment length by electrophoresis. The original method was

followed by slab gel polyacrylamide gelelectrophoresis and had a resolution of up to 300 bp. The DNA detection was due to radiography by using radioactive labeled ddNTPs. Modern automated sequencing uses florescent labeled ddNTPs which can be run in a single PCR reaction, use a single primer to amplify single stranded DNA and uses small capillary electrophoresis with in-flow florescent detectors. This results in a chromatogram of four fluorescent signals representing the different bases of the DNA chain with a retention volume/time proportional to their distance to the primer/their fragment length and increases the resolution limit to minimum 1kb. [35], [43], [44]

2.3 Microbiological methods

Microbiological methods were performed by our collaboration partner Mariusz Madej. For plasmids and (after ligations) vector amplification, chemically competent DH5 α *E. coli* were transformed with target DNA by heat shock transformation (method see 2.3.1) and were spread on LB agar. All LB media contained 100 μ g/ml ampicillin for vector selectivity. Single colonies selected were grown in LB medium, as well as their extracted and purified DNA (methods see 2.2.2) and send for sequencing to assure correct open reading frames of PorX and its mutants. For protein production vectors were transformed into the protein expression strain *E. coli* BL21(DE3) and cultured overnight in LB medium. From these cultures one liter expression cultures were inoculated 1:100 in fresh LB medium, grown to OD₆₀₀ of 0.6, induced by 0.2 mM isopropyl-1-thio- β -D-galactopyranoside (IPTG) (method see Chapter 2.3.2) and grown for 16 h at 20 °C. For selenomethionine derivatized PorX, for crystalization, minimal medium with selenomethionine was used. After expression, cell pellets were collected 15 min at 5.000 g by centrifugation, washed with PBS and stored at -80 °C until further use.

2.3.1 Heat shock transformation

Heat shock transformation of competent cells is based on the protocol of D. Hanahan published in 1983^[45]. The method is based on the increased bacterial membrane permeability in presence of calcium ions combined with a heat shock of a short incubation (0.5-1.5 min) at 37-42 °C after incubation with DNA at 0 °C. After recovery of the cells in SOC medium they were spread on LB agar plates to select single colonies. [46], [45]

2.3.2 Recombinant protein expression

Recombinant proteins are preferably expressed in *E. coli*, which in relation to other expression systems, yield higher amounts of recombinant proteins in a shorter amount of time, in a cheaper

expression media and with minimal physiologic risk. *E. coli* expression strain BL21(DE) carries one chromosomal copy of the T7 RNA polymerase gene under control of the UV5 promoter (lac promoter mutant) and an improved lacI-repressor lacI^q. The T7 polymerase is highly specific for its T7 promoters, which can be used on expression plasmids and has a very high activity. Furthermore BL21(DE3) is deficient in Lon and OmpT protease, which lowers degradation of over-expressed proteins. ^{[47], [48]}

pGEX-6P-1 (Cytiva) is an expression vector for heterologous protein expression as GST fusion protein. It carries the gen for β -lactamase for plasmid integrity by providing ampicillin resistance and protein expression is under control of the tac promoter, a strong promoter inducible by IPTG. The GST-fusion tag is located N-terminal of the protein of interest in order to purify the fusion protein by affinity chromatography (Chapter 2.4.1) and is cleaved after purification by PreScission protease. ^[49]

2.4 Protein Purification

Protein purification was performed by Mariusz Madej from our collaboration partners of group of Prof. Dr. Jan Potempa from the Jagiellonian University in Kraków (Poland).

Cell lysis was performed by resuspending cells in PBS and passing them at 23.000 psi through a cell disruptor, causing lysis due to high pressure release to ambient atmosphere. Insoluble debris was removed by 45 min centrifugation at 200.000 g and GST-PorX fusion protein was harvested via affinity chromatography on Glutathione Sepharose 4 Fast Flow (Cytiva) (method see 2.4.1), washed and eluted by GST-tagged PreScission protease (Cytiva) and cleaved on column at 4 °C over 24 h. PorX was separated from the matrix by centrifugation and dialysed 1:250 v/v with 20 mM Bis-Tris pH 6.5 at protein concentrations between 1-2 mg/ml and further purified by anion-exchange chromatography on a Resource-Q column (Cytiva) (method see 2.4.2). With a linear gradient applied, PorX eluted around 200 mM NaCl. PorX containing fractions were pooled together and further purified by size exclusion chromatography (SEC) (method see 2.4.3) on a preparative HiLoad 16/60 Superdex 200 column (Cytiva) without previous concentration in SEC-Buffer [5 mM Tris pH 7.5, 50 mM NaCl]. PorX eluted as a single peak with a hydrodynamic radius of PorX monomer (Madej, M. personal communication).

2.4.1 Affinity chromatography

Affinity chromatography as a method introduced by L.S. Lerman 1953^[50] uses the affinity of protein (alone or as fusion-tags fused to the protein of interest) to a specific ligand. This affinity enables the

(fusion-) protein to bind to an affinity matrix with immobilized ligands, whereas contaminations have no or at least a significant lower affinity to that matrix. ^[50]

Glutathione sepharose is used as matrix to purify glutathione S-transferase (GST)-tagged proteins from *E. coli* soluble lysate. Contaminants are removed by washing the matrix with buffer and the target protein is eluted from the matrix by proteolytical digestion from its fusion tag or alternatively by eluting with soluble ligand glutathione. ^{[51], [52]}

2.4.2 Ion-exchange Chromatography (IEC)

Ion-exchange chromatography for proteins as introduced by Peterson and Sober in 1955^[53] uses chromatographic matrices with immobilized positively charged cations (anion-exchange matrices, usually bound with quaternary ammonium) or negatively charged anions (cation-exchange matrices, usually acetylated). Oppositely charged proteins bind those matrices at low salt concentrations. and elute, when the ambient salt concentration is high enough to compete with the binding sites at matrix and/or protein. Elutions are performed either at linear or stepwise gradients. ^{[35], [53]}

2.4.3 Size exclusion chromatography (SEC)

Size exclusion chromatography (SEC, also called gel filtration (hydrophilic mobile phase) or gel permeation chromatography (hydrophobic mobile phase)) was firstly described in detail by Lathe and Ruthven in 1956^[54]. It is a chromatographic method to separate biomolecules of different molecular sizes. The separation is based on the difference between the radius/shape of the molecules (to be separated) to the radius/shape of the pore sizes of the chromatographic matrix. Small molecules, which can fully penetrate the pores within the matrix have a longer path length through the matrix within a column, whereas molecules larger than the pores have the shortest path length and thus elute with the smallest elution volume. The relationship of the elution volume and the hydrodynamic radius (of non-interacting particles) can be approximately described as linear between the $\log(M_w)$ and the elution volume within the separation range of the matrix between the void volume (shortest path) and column volume (longest path) of the matrix. The hydrodynamic radius of molecules to be separated furthermore corresponds to their molecular mass and three-dimensional shape and can be interpolated from reference samples of similar shape and atomic composition (e.g. globular proteins) with known molecular masses. Therefore an estimation of the molecular mass of a protein sample can be done by calculating a calibration curve of the elution volumes of standard proteins of a similar shape and thus a similar hydrodynamic radius to mass ratio. Preparative size exclusion always include sample fractionation and usually larger columns in

order to separate proteins in larger sample volumes, whereas analytical SEC have smaller columns to determine molecular weights more precisely in a shorter time per run. Protein peaks are usually detected with an in-line UV absorbance detector. [35], [55], [54]

2.5 Protein sample preparations

Protein storage was initially performed in either “*Phosphorylation-Buffer*” (Dim_PorX, Dim_PorX_SeMet), “*BeF₃-Buffer*” (BeF3_PorX, BeF3_PorX_SeMet) or “*SEC-Buffer*” containing 150 mM NaCl (*Mono_PorX*, PorY, PG1660 and PorX mutants) and protein concentrations below 1 mg/ml (Protein abbreviations see Chapter 2.1.10). Upon crystallization optimizations (plates Opt_B and following) salt concentrations for PorX were lowered to 50 mM NaCl (at SEC-purification step onwards, “*Crystallization-Buffer*”). *In vitro* phosphorylation (Chapter 2.5.1) of PorX was performed by 1 h incubation at 37 °C in “*Phosphorylation Buffer*”. For beryllium fluoride induced dimerization (method see 2.5.2) of PorX protein samples were incubated with “*BeF₃-Buffer*”. Protein concentrations were performed by ultrafiltration (method see 2.5.3) using Amicon Ultra-0.5 Centrifugal Filter Units (Merck). For crystallization trials concentrated proteins were centrifuged 10-20 minutes at 20.000 g before dispensing to the crystallization plates.

2.5.1 *In vitro* phosphorylation

Apart from RR activation by their activated SHK, RRs are able to catalyse their own phosphorylation in the presence of small phospho-donors, as shown *in vitro*^[56], and *in vivo*^[57]. Acetyl phosphate can be used to activate TCS response regulators in the presence of Mg²⁺, which might act as a metabolic regulation of environmental responses, and can be used *in vitro* to produce phosphorylated RRs at high efficiencies. This phosphorylation is reversible because RRs also inherit autophosphatase activity. Thus high acetyl phosphate concentrations are necessary for high concentrations of phosphorylated protein. [56], [57]

2.5.2 Beryllium fluoride induced PorX dimerization

Due to autophosphatase activity of RRs the active site phospho-aspartate and thus the active/phosphorylated state of RRs is unstable. *In vitro*, the active site phosphorylation could only be kept constant, as there would be no depletion of acetyl-phosphate in the buffer, which is not the case with halve life periods of RR phospho-aspartate between seconds-hours.. Thus for structural studies non covalent bound phosphate mimics are used in order to stabilize active RRs states/conformations. As shown by Yan et al. 1999^[58] beryllio-fluorides (BeF_x, with BeF₃⁻ postulated as the active ion species) provide such structural mimics that keep RRs biochemical, as well as

structural in their active forms (with comparable activities), in contrast to chemically similar known phosphate mimics aluminum fluoride or vanadate. ^[58]

2.5.3 Ultrafiltration

Ultrafiltration is a method used for protein concentration, purification or buffer exchange/desalting. Particles between 1-1000 kDa in solution can be separated by filtration based on their size by through a porous filter membrane (10-400 Å diameter). This membranes usually consists of cellulose with a specific molecular weight cutoff (MWCO), where the molecules/proteins of interest retain within the column and smaller particles including the solvent are driven out through a filter by centrifugal forces. Without adding new buffer larger molecules are concentrated within the retentate, whereas by adding new buffer small molecules, salts or buffer components can get washed out stepwise. ^{[59], [60]}

2.6 Protein analytics

Protein integrity was validated before or after every new crystallization optimization by SDS-PAGE (method see 2.6.1) and by analytical SEC (method see 2.4.3) after arrival of new sample or before new experiments. Protein concentrations were measured photometrically at their absorbance at 280 nm. To avoid misunderstandings with our collaboration partners there was no protein specific extinction coefficient correction applied at any step (thereby 1 AU = 1 mg/ml, see 2.6.2). Taking the extinction coefficient of PorX into account, concentrations used in this work have to be corrected by additional 23.9 percent according to *expasytools*^[61].

The oligomeric state of PorX under different conditions was analyzed by SEC-MALLS (method see 2.6.3) at the “*Protein Purification Facility*” of PCB (Barcelona Science Park). 50 µl samples of Mono-PorX_SeMet, PorX mutants D58A and D361A/H365A, each with and without previous *in vitro* phosphorylation were passed through a 0.2 µm filter and injected at room temperature, after buffer preequilibration and calibration with a BSA standard. The two different running buffers were “*MALLS-Buffer*” and “*MALLS-Zn Buffer*” (Chapter 2.1.2). Proteins were concentrated between 4.6 and 6.9 mg/ml, diluted to 2.68 (samples with zinc) or 2.61 mg/ml (without zinc) and further diluted to 2 mg/ml by adding additional salts in order to reach the final buffer concentrations of 50 mM Tris pH 8.0, 150 mM NaCl and 10 mM MgCl₂ for unphosphorylated samples, with additional 20 mM acetyl phosphate for phosphorylated samples and with additional 0.7 mM ZnCl₂ for samples to be run in the presents of zinc. Phosphorylated samples were further incubated one hour at 37 °C. All samples were filtered with 0.22 µm cutoff filters and kept at 4 °C in the autosampler. For each

buffer, injections were done every 1.5 column volumes at a flowrate of 0.5 ml/min. The order of sample injections in both runs was phosphorylated D361A/H365A; phosphorylated D58A; phosphorylated WT; D361A/H365A; D58A; and WT to minimize effects of protein(-auto)-dephosphorylation during the time within the autosampler.

Further analytics of PorX and its structure based mutants were performed by our collaboration partners, including extensive analytical SECs with or without *in vitro* phosphorylation, zinc or magnesium addition, radiography of PorX phosphorylation, enzymatic activity assays (phosphatase, phosphodiesterase) and nucleotide cleavage assays.

2.6.1 SDS-PAGE

SDS polyacrylamide gel electrophoresis (PAGE) as introduced by U. K. Laemmli 1970^[62] is a method to separate and visualize proteins within an electric field according to their size. Proteins are denatured and negatively charged in a sample buffer containing 25 % SDS and disulfide bridges are disrupted with β -mercaptoethanol. Polymerized acrylamide, cross-linked by methylenebisacrylamide acts as molecular mesh with pore sizes chosen by acrylamide concentration for the proteins of interest to separate. As the proteins are denatured and charged proportional to their molecular weight (ca. one SDS anion binds per 2 amino acids) they run in the electric field faster as smaller they are. A mixture of standard proteins can be used to comparatively determine the Mw of to be separated proteins of a sample. After electrophoresis protein bands can be visualized by silver staining, coomassie staining (or any other protein binding dyes) or radiography and later on scanned for documentation. ^{[35], [62]}

2.6.2 Spectrophotometry

Protein concentrations in aqueous solutions can be measured photometrically by their absorbance at 280 nm, where their tyrosine, tryptophan and cystine residues absorb. The concentration correlates (Lambert-Beer's Law): $c = A_{\lambda} / (\epsilon_{\lambda} \cdot b)$; with $A = \log_{10}(I_0/I_t)$ being the absorbance from transmitted light I_t to incoming light intensity I_0 , ϵ the extinction coefficient and b the path-length. Modern spectrophotometers can calculate the protein concentrations directly setting 1 absorbance units equal to 1 mg/ml, using a protein extinction coefficient based on average standard-proteins. If the protein of interest has an unusual percentage of absorbing amino acids the real absorption-coefficient can be calculated by: $c = A_{280} * \epsilon_{280} / Mw_{protein}$ with:

$$\epsilon_{\lambda} = \text{number}(Tys) \cdot \epsilon_{\lambda}(Tyr) + \text{number}(Trp) \cdot \epsilon_{\lambda}(Trp) + \text{number}(Cystine) \cdot \epsilon_{\lambda}(Cystine)$$

This kind of calculations imply there is negligible scattering or diffraction from the sample and reflection and absorption of the buffer are calculated out by baseline correction. Minor scattering contribution, e.g. due to protein aggregation can be reduced by baseline correction at 340 nm. Also the above extinction coefficient calculation does not take into account contributions of protein folding and is based on unfolded protein absorbances in 6 M guanidinium hydrochloride. ^{[63], [61], [64]}

2.6.3 SEC-Multi-Angle-Laser-Light-Scattering (MALLS)

SEC-MALLS (multi-angle-laser-light-scattering) is an analytical size exclusion chromatography (Chapter 2.4.3) with continuous scattering and refractive index detection of eluting solvent/sample mixtures. It is a technique used to determine the absolute molecular mass of proteins (upon other molecules), independent of their molecular shape, calibration standards or molecule-matrix-interactions. Molecular mass calculations from SEC-MALLS are reliable, if proteins elute sufficiently monodisperse from the SEC and if intermolecular interactions (e.g. from protein aggregation) are negligible. ^{[65], [66]}

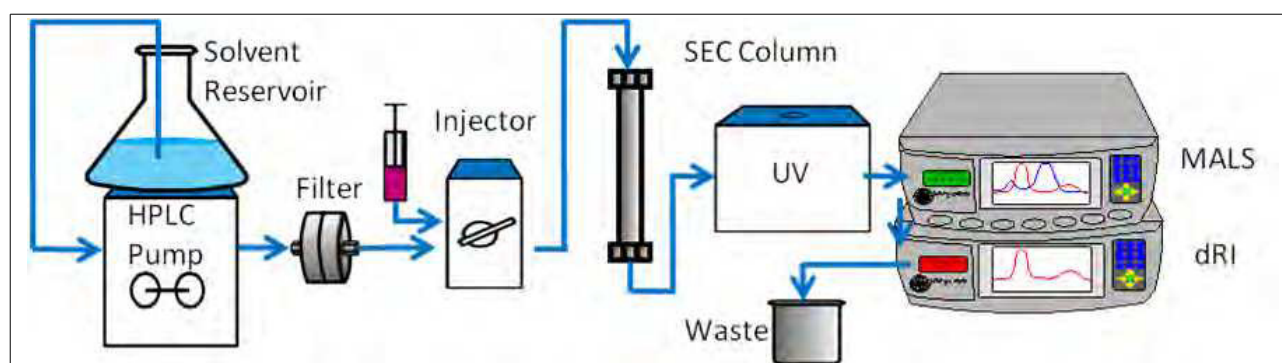


Figure 2.1: SEC-MALLS setup.^[66]

Experimental setup of SEC-MALLS: After a HPLC setup for analytical SEC with standard UV measurement, light scattering at multiple angles and change in refractive index are measured by two detectors (MAL(L)S and dRI) in-line with sample elution. ^[66]

Based on the work of Bruno Zimm 1948 on light scattering of polymer solutions^[67] and described by Philip Wyatt in 1993 ^[68] the absolute molecular weight M_w of eluting molecule species (at low concentrations including aqueous protein solutions) can be calculated by:

$$M_w = \frac{R_{(0)}}{K \cdot c \left(\frac{dn}{dc} \right)^2},$$

where $R_{(0)}$ is the Rayleigh ratio (scattered light intensity relative to laser light intensity) extrapolated to 0 angle, c the concentration and dn/dc the refractive increment of the analyte (refraction difference of analyte to buffer). The constant K depends on system variables like wavelength or

solvent. In SEC-MALLS $R_{(0)}$ is detected by the MALLS detector (multiple angles for more precise 0 angle extrapolation) and concentrations/refraction increments by a differential refractometer. The constant K and dn/dc (which is constant for all proteins with $\Delta_{\max}=2\%$) can get determined by system calibration. [65], [66]

2.7 Crystallization

Crystals "JG5"-, "HR"- and "pGpG"- from were obtained by sitting drop vapor diffusion by mixing 300-450 nl protein (PorX_SeMet) solution with precipitant solution in a 1:1 (v/v) ratio and incubating the resulting drops for 1-3 months at 20 °C. "JG5"-crystals (2.3 mg/ml in *BeF₃-Buffer*) crystallized in 100 mM glycine pH 8.0, 100 mM sodium formate and a PEG-mixture of 20 K, 10K, 8 K and 6 K at 1.5% each. "HR"-PorX (12.8 mg/ml) and "pGpG"-PorX (12.0 mg/ml), both in *Crystallization-Buffer*, crystallized in 100 mM glycine pH 8.0-8.5, 100 mM sodium formate, 3-4 % PEG 8 K, 15 mM MgCl₂, 30 mM NaF and 5 mM BeSO₄. "HR"-PorX and "pGpG"-PorX precipitant solutions contained additional 0.1 mM ZnCl₂ and 0.5 mM ZnCl₂, respectively. 0.5 mM pGpG and 0.5 mM GMP were added to "pGpG" and "HR" protein solutions respectively, before mixing. "HR"-PorX crystals grew as a cluster which was broken with CrystalProbe® (Hampton Research) into single crystals after adding cryo protectant. A second crystal from the "HR"- crystal cluster was soaked for 20 minutes in cryo-protectant containing additional 0.5 mM MnCl₂ (5 times excess Mn²⁺ over Zn²⁺).

2.7.1 Crystallization strategy and screenings

Crystallization screens were performed preferably (details and exceptions see 2.1.11) on sitting drop by using "MRC 2 Lens Crystallization Plates" (SWISSCI) nano-plates with 2 wells containing 96 conditions each. Volumes of 300-400 nl of reservoir solution were dispensed into the same volume of protein solution, at the "*Crystallization Platform Facility*" of PCB and stored at 20 °C. Crystal optimizations were also tried at 4, 15 and 37 °C, with drop sizes from 200 nl to 20 µl, 48 and 24 well sitting drop and 24 well hanging drop crystallization plates. Different protein drop to buffer ratios were screened as well as macro- and micro-seeding techniques. At the beginning of PorX crystallization trials, PorX_SeMet protein was concentrated and dispensed in *BeF₃-Buffer*. Upon optimization PorX_SeMet protein solutions were dispensed in *Purification buffer* (lower NaCl concentration) and with crystallization conditions containing BeF₃ for *in situ dimerization* in the crystallization drop to yield larger crystals. Details of crystallization screens performed can be seen in Chapter 2.1.11 (chronological order, page 5), the buffer compositions of PorX optimization

screenings Opt_A-M can be found in Chapter 7.4. The total amount of crystallization trials for PorX dimer, its derivatives and mutants contained over 10.000 conditions.

2.7.2 Crystal manipulation, imaging, soaking and cryo cooling

Crystal containing protein drops were diluted with their reservoir solution, after adding 27.5 % glycerol (to the reservoir), to a final optimized concentration within the cryo solution of around 18 %. During the crystallization trials, concentrations between 15 and 25 % glucose were tested, as well as concentrations of glucose, sucrose and trehalose, mineral oils, PEG200 and PEG 400, 1,3- and 1,4-butanediol, ethanol, organic salts (malate, formate) and ammonium sulfate. For soaking trials selected crystals (obtained from PorX or T272A mutant) were additionally incubated with 0.5-3 mM GTP, 0.5 mM pGpG or 0.5 mM c-diGMP in cryo condition buffers. Incubation times of 5, 10 and 20 min with the ligand were tested after addition to the protein drop before crystal harvest, as well as previous incubation of several weeks.

Crystals were harvested using cryo loops and flash frozen in liquid nitrogen. Crystal clusters were broken into single fragments with CrystalProbe[®] (Hampton) needles after cryo solution addition. Images were taken before and after cryo addition as well as before and after crystal manipulations using a digital camera in front of the microscope lens, often with additional polarization filter mounted onto the microscope lens.

2.7.3 Crystallization theory

Protein crystallization was first described in 1840 by Hünefeld^[69] who described obtaining hemoglobin crystals by slow evaporation of concentrated protein solutions and was used in the following century for purification reasons, as proof of purity and out of scientific curiosity. Methods for protein crystallization were greatly improved by J.B. Sumner, J.H. Northrop and their coworkers who Therefore received the Nobel-price in chemistry in 1946^[39]. It was applied as method for X-ray crystallography from the late 1930s onwards, when protein crystallization was used to obtain structural information of proteins by X-ray diffraction. ^{[70], [39]}

Crystals suitable for X-ray diffraction have to be of high quality which means have a high degree of internal order within the crystal that leads to high resolution diffraction. Thus they should preferably have a single crystal lattice, have low crystal defects, few to none alternative conformations, less solvent content and be large in absolute size. This requires highly pure protein samples and, due to low predictable nature of protein crystallization, a vast number of conditions to be screened and optimized in order to control protein nucleation, the phase transition from partially ordered (on

short distances) interacting protein intermediates to a first solid state, as well as growth of that nascent nuclei to yield a few highly ordered crystals of suitable size instead of many micro-crystals. Nucleation and crystal growth depend on the state of supersaturation in solution, whereas the concentration for nucleation usually has to be a multiple of that concentration required for crystal growth, but lower than the concentration, which leads to amorphous precipitation of the target molecule (Figure 2.2). Apart from crystallization or precipitation also phase separation, gels or partially ordered spherulites appear during supersaturation, depending on protein and buffer conditions. ^{[71], [70]}

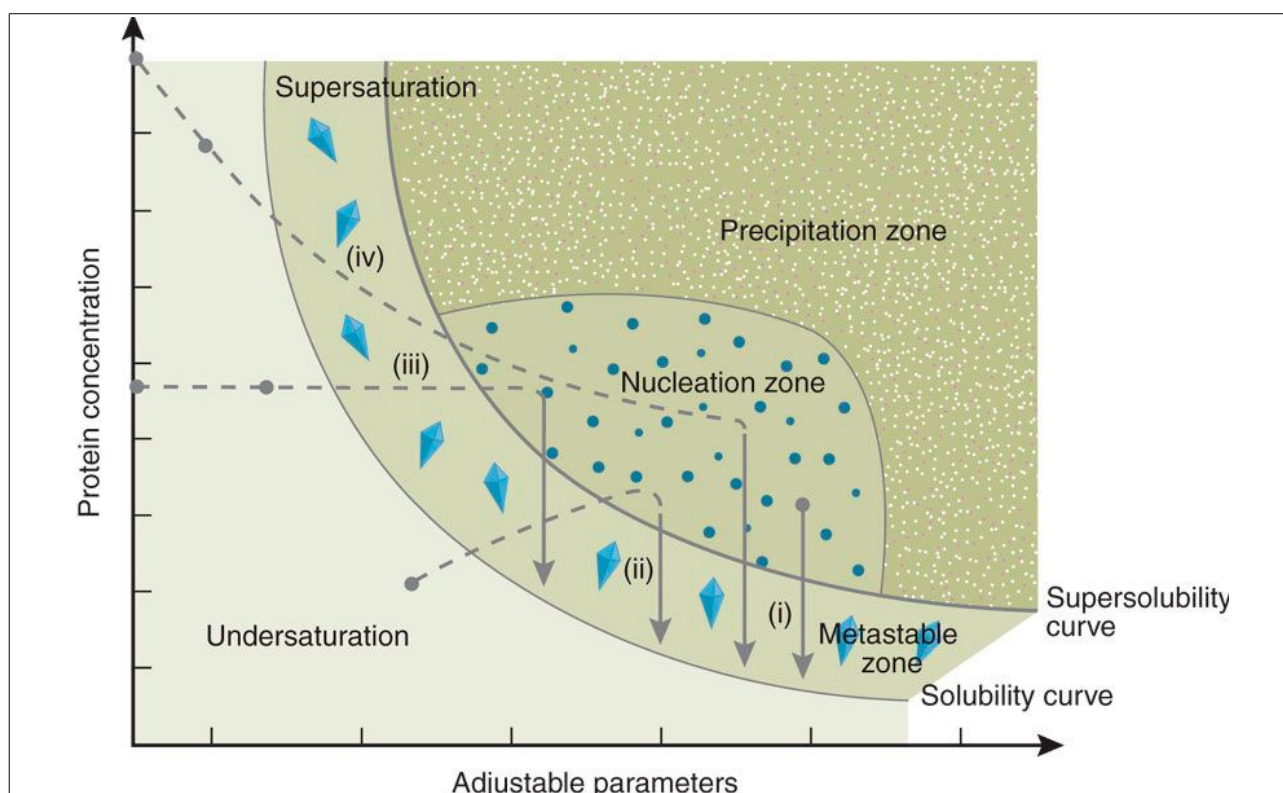


Figure 2.2: Macromolecular crystallization phase diagram.^[71]

Phase diagram of protein crystallization. Protein solubility depends on protein concentration (Y-axis) and adjustable physical and chemical parameters (X-axis) like pH, temperature, salt, organic solvent, polymer or ligand concentrations etc. in a non-linear manner, where multiple parameters need to fit in order to yield nucleation and/or crystal growth. In the undersaturated zone protein stays soluble, whereas the oversaturated zone is divided into a metastable zone, where crystals can further grow but no nucleation occurs and an oversaturated zone subdivided into nucleation zone (where nuclei can form) and precipitation zone. (i)-(iv) are suitable crystallization experiments (i: microbatch, ii):vapor diffusion, iii): dialysis, iv): diffusion) to obtain protein crystals. ^[71]

Supersaturation of protein in solution can be obtained through several different techniques involving batch techniques, evaporation, dialysis, diffusion and/or sequential extraction. Screened parameters usually include pH, temperature, salts or organic compounds, its variations and

combinations. Some added compounds act as precipitants which compete with the protein for solvent interactions thus strengthen protein-protein interactions. Other compounds can form interactions with the proteins itself and act as (or avoid) linking molecules between individual proteins units. One of the most common techniques used for screenings is the so called “sitting drop” vapor diffusion technique. In this technique the crystallization droplet, a mixture of protein solution and crystallization buffer “sits” on a platform above a large reservoir of crystallization buffer within an airtight chamber. Likewise in the so called “hanging drop” procedure the crystallization droplet does not sit in a separate well above the reservoir, but hangs above it, attached by adhesion to a cover glass plate. In both techniques, due to the usually higher osmotic strength of the crystallization buffer, water and/or volatile components of the solutions will equilibrate between droplet and reservoir due to vapor diffusion and the protein concentration in the shrinking crystallization droplet will gradually increase during evaporation. The hanging drop technique can be used highly automated in 96-well plates/buffer condition screenings and thus has a high reproducibility. ^[70]

The screening methods for crystallization conditions vary between “random” screening of adjustable, parameters in high throughput “*sparse matrix screenings*” and systematic approaches in “*grid screens*”. In sparse matrix screenings buffer components are screened in quality and quantity, including buffers, pH, salts, divalent cations, ligands, precipitants, additives etc. which are known to be able to aid protein crystallization, while protein concentrations, volumes and temperatures are kept constant.

The same sparse matrix screenings can also be duplicated at different temperatures, dispensed with different protein concentrations or different protein to reservoir volume ratio to observe effects on protein nucleation. These initial random conditions then can be scored and clustered into different qualities of precipitation (ordered or unordered precipitation likewise). Their score is empirical to be optimized to yield crystals and a higher score helps identifying stabilizing conditions like pH, additives etc.. The preferred conditions then can be fine-tuned and/or combined in grid screens, where single components (like pH, salt or precipitant concentration) are step-wise increased or the concentrations or pH are kept constant, but the type of buffer, salt, additive or precipitant are exchanged in order to yield better precipitation, nucleation or larger crystals. In final screenings addition of substrates or ligands can be screened in order to get crystals with higher symmetry, crystals in other protein conformations or obtain crystals of protein-ligand complexes. At each step of crystallization screening upscaling attempts can be performed in order to get larger crystals,

which could be combined with macro- or micro- seeding of crystals or nuclei which were grown under different conditions. However, these upscaling and seeding techniques are sometimes hard to reproduce due to a larger degree of manual performed steps than by automated procedures. [72], [70]

2.8 X-ray diffraction

Experimental data collection setups yielding our published pdbs are summarized:

Datasets	JG5	HR	pGpG
PDB entry	7PV7	7PVA	7PVK
Beamline	ESRF BEAMLINe ID29	ESRF BEAMLINe ID29	ALBA BEAMLINe XALOC
Detector	DECTRIS PILATUS 6M-F	DECTRIS PILATUS 6M-F	DECTRIS PILATUS 6M
Date of collection	2017-04-28	2018-04-18	2019-03-13
Temperature	100 K	100 K	100 K
Absorption edge	Selenium	Zinc	Selenium
Protein	BeF3_PorX_SeMet	BeF3_PorX_SeMet	BeF3_T272A_SeMet

TABLE 2.1: Data collection setups

Initial datasets acquired from BeF3_PorX did not contained selenomethionine and extensive trials to solved the structure by molecular replacement were done using PHASER^[2], MOLREP^[7], CHAINSAW^[6], and ARCIMBOLDO-SHREDDER^{[4], [5]}. Several homology models were built based on homologous receiver domain structures which were up to 30% identical to the first 120 residues of PorX, structural prediction models based on PorX aminoacid sequence build by Robetta or ITasser online servers, as well as their subdomains, truncated versions or poly-alanine models.

Upon crystallization optimization the structures from three crystals could be solved and modeled suitable for publication: crystals “JG5”, “HR” and “pGpG”. A single dataset of “JG5” without additional zinc addition in the buffers was collected at the Se absorption edge. “HR”-X-ray data was collected from one crystal at the Zn absorption edge, at 5 different positions with two kappa reorientations, and at another position at the Mn absorption edge. A manganese soaked fragment from the “HR”-PorX crystal cluster was diffracted at the Mn and Zn absorption edges as well. X-ray data from “pGpG” crystal was collected at the Se absorption edge at two different positions with different kappa orientations.

Data indexing and reduction (Chapters 2.8.4-2.8.7) was performed with XDS^[23]. Multiple datasets from the same crystal (“HR”- and “pGpG”- PorX data) were initially analyzed by ccp4i.BLEND^[1] for similarity, then merged into one file and correction factors applied with XSCALE^[23].

ap_SCALE^[14] was used for final scaling, merging, truncation and resolution cut-off based on $CC1/2 > 50$. The initial PorX structure was solved by SAD (single anomalous diffraction, see 2.9) using SHELX C/D/E^[15] using data from multiple different PorX-crystals. "JG5" and "pGpG" structures were solved by MR-SAD (Chapter 2.9.2) by PHASER-EP pipeline^{[2], [73], [3]} using previous structures as a searching model, the "HR"-structure by molecular replacement (Chapter 2.9.2) using PHASER^[2]. Model building (Chapter 2.9.3) started with phenix.autobuild^[12] and ccp4i2.buccaneer^{[9], [8]} and continued by iterative model refinement with phenix.refine and map calculations with phenix.maps^{[10], [11]} or buster^[13] and subsequent manual model building and real space refinement with COOT^[19].

2.8.1 Experimental setup

X-ray crystallography uses the diffraction of X-rays at crystals to determine the structure of its atomic contents. X-rays are required due to their low wavelength similar to atomic distances, which makes it possible to resolve neighboring atoms as separate diffracting entities and thus determine structures at atomic resolution. Crystals (of molecules of interest) are needed, because single molecule X-ray diffraction would not be possible to detect with modern methods. In a crystal symmetrical arranged identical molecules intensify their diffraction intensities due to their internal symmetry by constructive interference (at specific distances and angles) which then can be detected. In order to obtain intensified signals of diffraction the initial X-rays need to be coherent, which means they need to have a constant phase and/or amplitude relationship over time/after a certain distance (temporal coherence) and in the same direction of propagation/in phase over the whole beam cross section (spatial coherence). The length of coherence of light l_c depends on the monodispersity $\Delta\lambda$ of its source and is given by:

$$L_c = \frac{\lambda^2}{2\Delta\lambda}$$

Thus diffracted X-rays need to be ideally monochromatic (for temporal coherence) and the source of light is ideally a point source resulting in parallel waves (for spatial coherence).^{[63], [74], [75]}

The experimental setups, according to Rhodes et al. 2007,^[75] can be summarized as followed: There are three suitable X-Ray sources and synchrotron radiation is the most favorable X-ray source due to its high intensity and selectable wavelength. Synchrotron radiation is produced at particle accelerators where electrons or positrons are accelerated to high velocities and hold in a circular motion by strong magnetic fields. The change in direction induced by the magnetic fields lets those particles emit X-rays that can be used as light source for X-ray diffraction. These X-Rays passes

monochromators which restrict further the wavelength of the beam and focusing mirrors may enhance the flux at the sample. Alternative, X-ray sources with lower intensities and nonselectable wavelength are either X-ray tubes or rotating anodes, preferably used as "home-sources" at the local laboratory/facility. These X-Ray sources emit X-rays by radiating copper, molybdenum or chromium. The electromagnetic waves excite lower energy orbital electrons of these metals to higher energy level, which then emits monochromatic X-rays by falling back to its lower orbital. Due to multiple possible electron transitions and thus emitted wavelengths, undesired wavelengths are quantitatively absorbed by optical filters. Diffraction can be recorded by X-ray scintillation counters, sensitive films, imaging plates or charge coupled devices (CCD) detectors. CCDs are the most suitable for X-ray crystallography. These are solid state devices, a grid of light sensitive pixels, which accumulate electrons proportional to incoming visible light and accumulating charges are read out and digitally stored as diffraction patterns. In order to transform diffracted X-rays to visible light which is detected by the CCDs, optical fiber bundles are coated with a photoluminescent phosphor that converts X-rays to light in the visible spectrum.^[75]

The schematic combined experimental setup is illustrated in Figure 2.3:

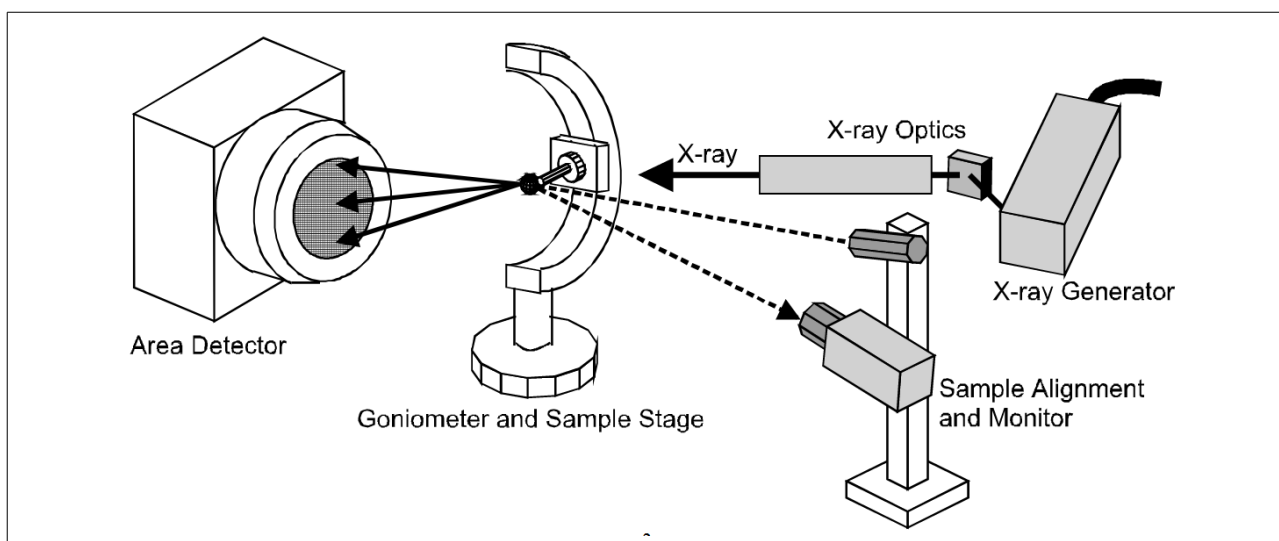


Figure 2.3: General X-Ray diffraction setup.^[76] Experimental setup of single crystal X-Ray diffraction experiments. X-Ray generators like synchrotron radiation, X-Ray tubes or rotating anode tubes act as light source. X-Ray optics as collimators and monochromators produce a coherent beam focused on the sample. A goniometer is used to rotate the sample to get diffraction data at multiple angles and an area detectors detect two-dimensional diffraction images. Behind the crystalline sample usually a so called beam stop (not shown in this graphic) is placed to protect the detector from the direct beam. Remote sample alignment and experimental monitoring by the user can be supported by additional visible light sources and cameras, in particular at synchrotron radiation facilities.^{[76], [75]}

2.8.2 Diffraction principles

X-ray crystal diffraction was firstly described and analyzed by Max von Laue^{[77], [78]} as well as William H. Bragg and William L. Bragg in 1912^[79].

X-Ray diffraction patterns measured from a crystal arise from the elastic scattering of the X-rays at the electrons of the sample. As X-rays are electromagnetic waves their magnetic field interacts with charged particles like electrons and protons. This brings them in oscillation, resulting in radiation emission of the same frequency/wavelength in all directions, thus scattering of the initial beam. Due to their high mass scattering of protons is negligible and measured diffraction patterns only results from photon scattering from electrons of a sample and that only at specific angles. These angles depend on the electron density distribution within the molecule, the wavelength of the light source and on the distances between the individual scattering centers that lead to constructive interference between the scattered waves. Constructive interference occurs, if an integer number n of the wavelength λ equals to twice the product of the distance d between equivalent parallel planes and the sinus of the angle θ between the incoming X-rays and these plains. This relationship is referred to as

$$\text{Bragg's Law : } n\lambda = 2d \sin \theta,^{[79]}$$

resulting in reflections of the order " n ". From a physical point of view and derivable from trigonometry (Figure 2.4) the term " $2 d \sin \theta$ " is the optical path difference δ the light has to travel when being scattered from the next plane plus the path difference to be in phase with the scattered wave from the previous plane. If the waves travels exact this difference (or an integer multiple n of it) so the scattered waves are in phase with each other and constructive interference occurs.^{[63], [80],[75]}

Following from Bragg's law (p. 27), as also visually seen in Figure 2.4, increasing of the distances d between planes, the diffraction angles will decrease. This causes a reciprocal relationship between the three-dimensional distances between equivalent distant planes of scatterers in the sample, referred to as real space, and the three-dimensional space of reflections between angles θ referred to as reciprocal space. Distances d in the real space between points in the lattice, with the point vectors a , b and c within the crystal are defined as

$$\vec{d} = u\vec{a} + v\vec{b} + w\vec{c} \quad \text{with } u, v, w \in \mathbb{Z}.^{[63], [75]}$$

In the reciprocal space the distances between reflections are defined as $\vec{d}^* = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*$

with $h, k, l \in \mathbb{Z}$ and $\vec{d}^* = \frac{1}{d}$, $\vec{a}^* = \frac{1}{a}$, $\vec{b}^* = \frac{1}{b}$ and $\vec{c}^* = \frac{1}{c}$.^{[63], [75]}

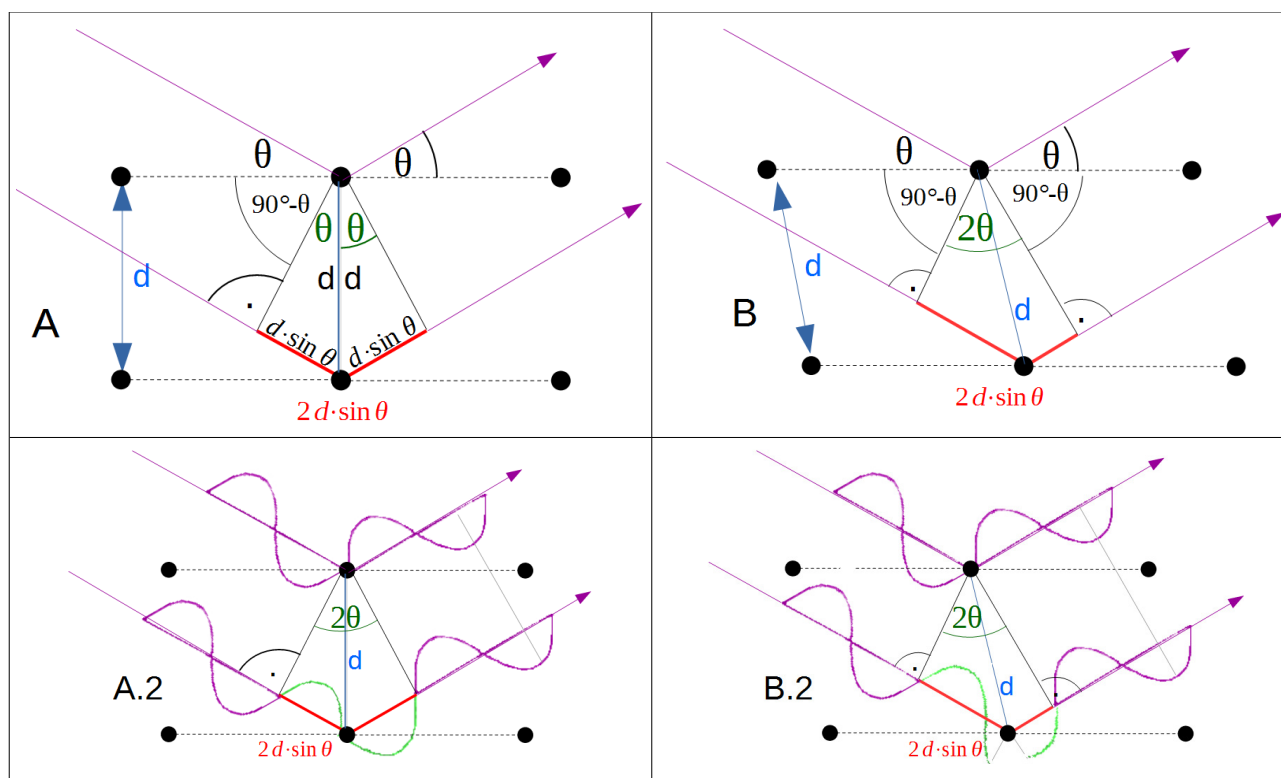
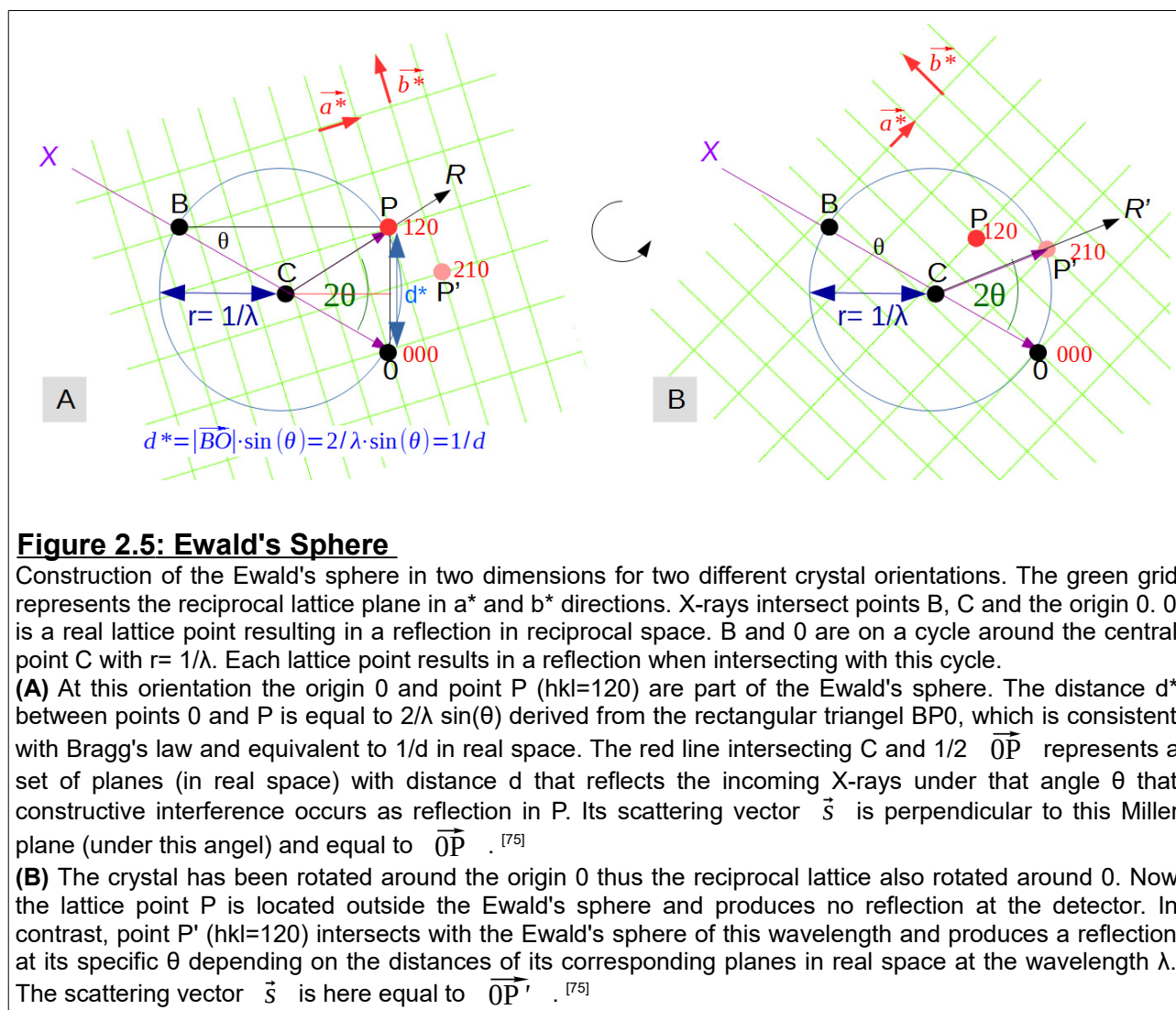


Figure 2.4: Geometrical explanation for Bragg's law.

At angles θ between incoming beam and parallel set of planes, where the path difference ($2d \sin \theta$) from incoming and scattered X-rays is n ($\in \mathbb{N}$) times the wavelength, then scattered waves are in phase, constructive interference occurs, diffraction can be observed and measured with a detector in the direction of the scattered waves. Two examples A and B are shown, where in (A) and (A.2) the vector of the distance d is perpendicular to two equidistant set of planes and (B) and (B.2) where it is not. Bragg's law can be derived from trigonometry inside the right-angled triangle in figures (A) and (A.2). This relationship is independent of the direction of the vector of d , as illustrated in (B)/(B.2). The distances d between two planes of scatterers are shown in blue, the angles θ in green, the distance " $2d \sin \theta$ " in red and the waves (with monochromatic wavelength λ) scattered from two planes are illustrated violet where they are in phase (in beam direction) or in green where they are out of phase. ^{[63], [75]}

h , k and l are the so called Miller-indices of the equidistant parallel planes in the real space and each of this planes produce one reflection in the reciprocal space, whose scattering vector $\vec{s}$ is perpendicular to its Miller plane. A commonly used visualization of the reciprocal space of reflections is the so called Ewald's sphere, described in the following (and shown in Figure 2.5) as a cycle in two dimensions for better visualization. Choosing an origin O upon the reciprocal lattice in direction of the incoming positively interfering/reflected X-rays as one point on the cycle, point C as center of the cycle in distance $1/\lambda$ on the incoming beam and point B as the first intersecting point of the cycle with the beam, all lattice points P fulfill Bragg's law and give reflections, if they

intersect with the cycle with radius $1/\lambda$ around C. Rotating the crystal (with fixed detector in distance to the reflected beam or vice versa) will rotate the lattices around origin 0 in real as in reciprocal space. In that way for every reflection (each set of equidistant set of planes in real space) there will be an angle θ (as larger the distances of the planes as smaller the angles, see Figure 2.5) that Bragg's law is fulfilled when the angle between PBO is θ and the distance d^* between 0 and P is $d^* = |\vec{BO}| \cdot \sin(\theta) = 2/\lambda \cdot \sin(\theta)$ which is $1/d$ by Bragg's law and the definition of d^* . [63], [75]



Crystalline samples intensify X-ray diffraction signals due to a multiplication of equidistant parallel planes of molecules in the direction of the incoming light source. Therefore linear multiplication of diffracted intensities will occur at angles that fulfill Bragg's law depending on the internal structure of the crystal, determined by its quality (single lattice, degree of crystal defects) and its "unit cell". [75]

2.8.3 Unit cells and asymmetric unit

The unit cell is defined as the smallest three-dimensional unit that can be translated into all directions to build up the entire lattice of a crystal. Without additional internal symmetry (rotation, inversion, rotoinversion) this unit is equal to the "asymmetric unit" (Figure 2.6) of the crystal which is defined by the equal three-dimensional units of the crystal which positions build up the grid points of the crystal lattice. If neighboring grid points of the crystal are graphically linearly connected, the unit cell of the crystal lattice will build up a "parallelepiped" - a three-dimensional space bordered by parallel planes with the side lengths a , b and c as well their included angles α , β and γ (Figure 2.7).

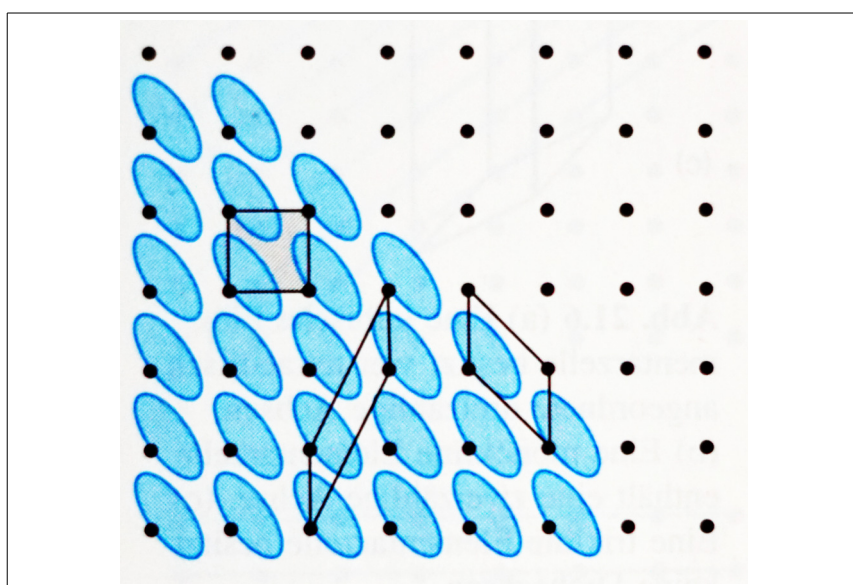


Figure 2.6: Unit cells and asymmetric unit within a crystal lattice^[63]

The grid points as black dots represent the crystal lattice, the asymmetric units are shown in blue and the possible unit cells as rectangles between neighboring grid points. Between the different possibilities of choosing one unit cell usually the one with the smallest side lengths is used.^[63]

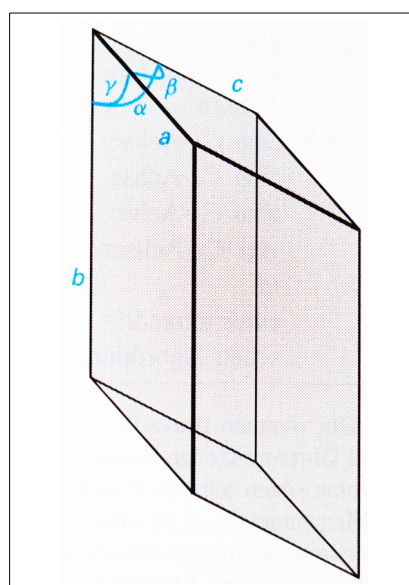


Figure 2.7: Unit cell attributes^[63]

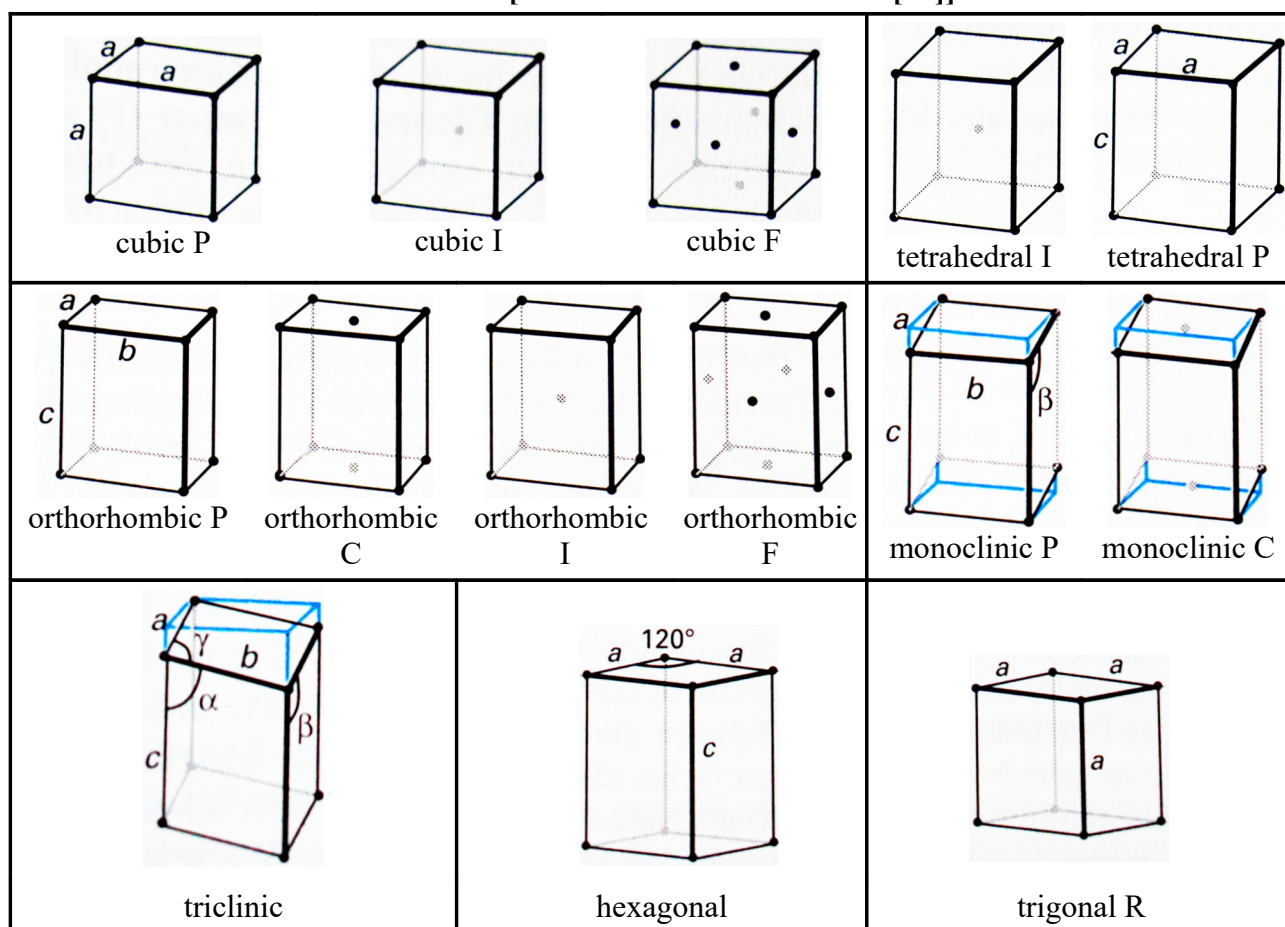
The parallelepiped is of a unit cell is defined by its side lengths a , b and c and their included angles α , β and γ ^[63]

Different unit cells are classified into seven different crystal systems (TABLE 2.2) according to their rotational symmetry and can build up 14 different types of crystal lattices referred to as bravais lattices (TABLE 2.3). Primitive unit cells which just have grid points on their edges are symbolized by P, room- or inner-centered unit cells with an additional grid point in the center as I, planar centered cells as F with additional grid points on the middle of its planes and the site-centered unit cells as A, B, and C with two additional grid points on two opposite planes.

TABLE 2.2: Crystal systems

System	essential symmetry ^[63]	Axis and angles ^[75]	System	essential symmetry ^[63]	Axis and angles ^[75]
triclinic	-	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma$	tetragonal	one C_4 axis	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$
monoclinic	one C_2 axis	$a \neq b \neq c$ $\alpha = \gamma, \beta > 90^\circ$	hexagonal	one C_6 axis	$a = b$ $\alpha = \beta = 90^\circ, \gamma = 120^\circ$
orthorhombic	three perpendicular C_2 axes	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	cubic	four C_3 axis in tetra-hedral packing	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$
rhombohedral	one C_3 axis	$a = b = c$ $\alpha = \beta = \gamma \neq 90^\circ$			

TABLE 2.3: The 14 Bravais lattices [translated and edited from [63]]



In real space the Miller indices (h, k, l) describe the equidistant parts in which these planes cut the unit cell vectors a, b and c (Figure 2.8), whereas in reciprocal space the Miller indices build the intersection points with the unit cell axes a^*, b^* and c^* (Figure 2.5). Thus the Miller indices are integral numbers and if negative (relative to the origin 0) written as $\bar{h}, \bar{k}$ and/or $\bar{l}$. An index of 0 is equivalent to parallelism of a plane to this axis and the origin 0 ($h, k, l = 0, 0, 0$) of this coordinate system will be on the edge of the unit cell. ^{[63], [75]}

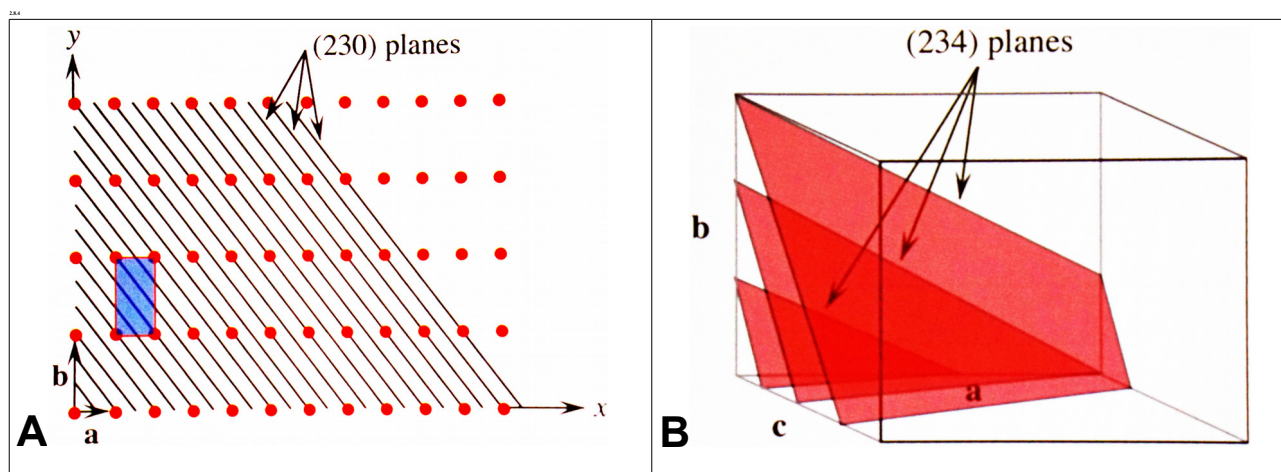


Figure 2.8: Miller plane intersection with the unit cell in 2 (A) and 3D (B)[75]

2.8.4 Indexing

Once the reflections of a crystal were measured and the Ewald's sphere reconstructed, from the position of each reflex it is possible to calculate the angles θ and thus calculate the distances of planes by Bragg's law (page 27). The spacings of the reciprocal lattice/the difference vectors of reflections are inverse proportional to the size of the unit cell, and with known wavelength and crystal to detector distance it is possible to calculate the unit cell's base vectors a , b and c and their length by trigonometry. a , b and c have to be chosen as smallest linear independent and their reciprocal a^* , b^* and c^* should describe all difference vectors between reflections as multiple triplets h , k , l . For simplicity, in an orthorhombic crystal, if P would be the reciprocal lattice point (100) in the Ewald's Sphere (Figure 2.5) and be detected as reflection P_D and 0 be detected as 0_D , then the length of the unit cell axis a equals to $1/(\text{the distance } OP)$. By geometry and distance $CP = 1/\lambda$:

$$\frac{OP}{CP} = \frac{P_D 0_D}{C 0_D} = OP \cdot \lambda \Rightarrow OP = \frac{P_D 0_D}{C 0_D \cdot \lambda}$$

$$\Rightarrow a = d_{100} = \frac{C 0_D \cdot \lambda}{P_D 0_D} \Rightarrow a = \frac{\text{detector distance} \cdot \lambda}{\text{distance}[(000)(100)] \text{ on detector}}$$

The same calculations can be done in order to get the length of b from $P(010)$ or c from $P(001)$. Likewise these calculations can be done for any other of the crystallographic systems. [63], [75]

With known angles and side lengths of the unit cell it is possible to assign the correct crystal systems (TABLE 2.2). Bravais types (TABLE 2.3) can be predicted based on internal rotational and inversional symmetry of the reflections. This leads to the assignment of potential space groups of the crystal, with potential alternatives if the crystal has additional screw axes. In these cases the real

space group may only be determined after solving the structure. By indexing also the relative intensities I_{hkl} is assigned to their reflex. The intensity of each reflection (h,k,l) is proportional to $|F_{hkl}|^2$ with F_{hkl} being the entire unit cell's structure factor, which is the sum of each structure factor F_i at positions $x_i a, y_i b, z_i c$. The single structure factors:

$$f = \frac{\omega^2}{\omega^2 - \omega_s^2 - i k \omega},$$

depend on the incoming photon frequency ω , the resonance frequency of the electron ω_s and damping factor k , which imaginary component only gets relevant if the photon frequency gets close to the atoms resonance frequency and absorption occurs. [63], [81], [75]

2.8.5 Space group determination

If there is internal symmetry inside the unit cell then specific sets of symmetry correlated reflections will be observed which can be used to determine the space group $X n_{(x)m_x} n_{(y)m_y} n_{(z)m_z}$ of the crystal. The space group nomenclature defines the lattice X (P, I, F, R, C see TABLE 2.3) followed by possible symmetry operations in x, y and z directions with n fold rotational axis and a translation of m/n. The indices n define the n-fold rotational axis with a translation of n/m in case of a screw axis. Certain unit cell symmetries can be determined by observing systematic absences in the diffraction pattern. These result from destructive interference of diffracted waves at a phase difference of 180° at certain planes of reflections. These are specific for each space group, depending on centering, glide planes or screw axis. [75] In case of multiple choices of the space group, the one with the highest symmetry is chosen. The 14 Bravais lattices can build up 32 different “point groups” by translational symmetry or 11 “Laue groups” which have an inversion center (TABLE 2.4). In combination with additional rotational symmetry these 32 point groups can build up to 230 different space groups. Due to the chirality of proteins, inversion centers or mirror planes are not possible within the unit cell thus only 65 of 230 space groups, so called Sohncke space groups, are possible in protein crystallography (TABLE 2.4). [82], [83], [75]

TABLE 2.4: The 65 Sohncke space groups in protein crystallography ^[82]

Laue group ^[75]	Bravais type	Possible space groups		Laue group ^[75]	Bravais type	Possible space groups	
-1	aP	P1	1	-3/m	hP	P321, P3(1)12, P3(2)12	149, 151, 153
2/m	mP	P2, P2(1)	3, 4	-3/m	hP	P321, P3121, P3(2)21	150, 152, 154
2/m	MC, ml	C2	5	-3/m	hR	R3(2)	155
mmm	oP	P222, P222(1), P2(1)2(1)2, P2(1)2(1)2(1)	16, 17, 18, 19	6/m	hP	P6, P6(1), P6(5), P6(2), P6(4), P6(3)	168, 169, 170, 171, 172, 173
mmm	oC	C222, C222(1)	21, 20	6/mmm	hP	P622, P6(1)22, P6(5)22, P6(2)22, P6(4)22, P6(3)22	177, 178, 179, 180, 181, 182
mmm	oF	F222	22	m-3	cP	P2(3), P2(1)3	195, 198
mmm	ol	I222, I222(1)	23, 24	m-3	cF	F23	196
4/m	tP	P4, P4(1), P4(2), P4(3)	75, 76, 77, 78	m-3	cl	I23, I2(1)3	197, 199
4/m	tl	I4, I4(1)	79, 80	m-3m	cP	P432, P4(2)32, P4(3)32, P4(1)32	207, 208, 212, 213
4/mmm	tP	P422, P42(1), P4(1)22, P4(1)2(1)2, P4(2)22, P4(2)2(1)2, P4(3)22, P4(3)2(1)2	89, 90, 91, 92, 93, 94, 95, 96	m-3m	cF	F432, F4(1)32	209, 210
4/mmm	tl	I422, I4(1)22	97, 98	m-3m	cl	I432, I4(1)32	211, 214
-3	hP	P3, P3(1), P3(2)	143, 144, 145				
-3	hR	R3	146				

2.8.6 Friedel pairs

The more symmetric a space group is, the more identical sets of unique reflections can be observed. Without internal symmetry (primitive cell, P1) there are just 2 sets of centrosymmetric unique reflections, one diffracted from each sides of the equidistant planes. These reflections are called Friedel pairs and in absence of X-ray absorption within the crystal have the same intensity and reversed signed phases:

$$I_{h,k,l} = I_{-h,-k,-l} ; \quad \phi_{h,k,l} = -\phi_{-h,-k,-l} , \quad \text{referred to as } \mathbf{Friedel's Law}.$$

With additional crystal internal symmetries the sets of unique reflections can reach up to 48 (24 sets of Friedel pairs, specific for P432 space group). After indexing and space group assignment these equivalent sets of unique reflections are merged, resulting in a higher multiplicity of unique reflections collected and a better signal to noise ratio according to position and intensity measured.

In case of measurable X-ray absorption there is one exception from the equality of intensity of Friedel pairs due to anomalous scattering/or anomalous dispersion, and the structure factors of Friedel pairs F^+/F^- should be indexed separately. X-Ray absorption occurs when electromagnetic waves are absorbed from electrons involved in scattering. This effect is strongest at the electron's element absorption edge, which for heavy atoms lies within the X-Ray spectrum. The absorption of X-rays of those "anomalous scatterers" will result in a phase shift of the scattered wave between planes, a change in the structure factor of its corresponding reflections and thus an intensity difference between Friedel pairs in non-centrosymmetric crystal systems. The reason for this phenomenon is a resonance effect when the absorbed photon promotes an electronic transition to an upper energy level, changing its oscillation frequency and breaking its centrosymmetric contribution, e.g. transition from an s to a d orbital. Thereby the scattering factor f of anomalous scatterer has to be corrected by a real and imaginary function: [63], [81], [75]

$$f = f_0 + f' + i f'' ,$$

f' phases are 180° shifted compared to the phases of f_0 (the scattering factor without anomalous correction) whereas the imaginary part f'' (which is responsible for the intensity difference of Friedel pairs) is shifted by 90° . The structure factor F_{hkl} of each reflection is proportional to the square root of its intensity I ($F^2=I$) and derives from the sum of each atom's structure factor f involved in diffraction: [63], [81], [75]

$$F_{hkl} = \sum_j f_j \exp(i\phi_j), \text{ with } \phi_j = 2\pi(h_j + k_j + l_j). \quad [63]$$

Thus different structure factor summations can obey or break Friedel's law,

a) between scatterers without X-ray absorption F_L where intensities and phases between F^+ and F^- are equal;

b) between anomalous scatterers F_H where Intensities are equal but phases differ;

and

c) between anomalous scatterers with nonanomalous scatterers F_{LH} , where intensities and phases can differ unless both structure factors are colinear (Figure 2.12). [63], [81], [75]

While normal scattering equivalent reflections follow Laue group symmetry, equivalent anomalous reflections follow point group symmetry (Chapter 2.8.5, TABLE 2.4). Pairs of symmetry related reflections that are not true Friedel pairs are called Bijvoet pairs. [81], [83], [75]

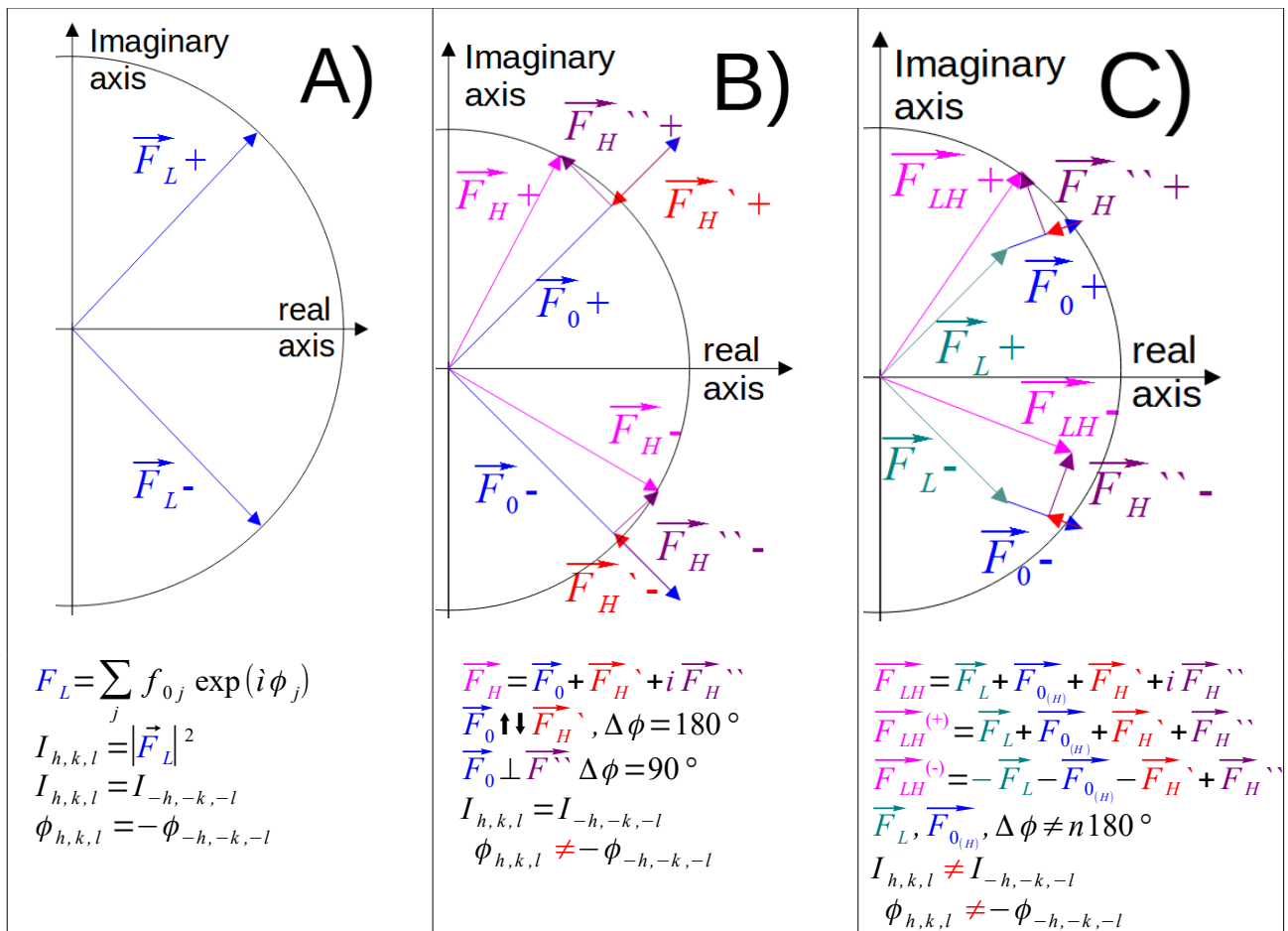


Figure 2.9: Vector “Harker-” diagram of structure factor summation for inverse reflections ((h, k, l) and (-h, -k, -l)) and corresponding structure factor equations

$|\vec{F}|^2$ is proportional to the measured intensities, while the angle of each $\vec{F}$ with the imaginary axis represents the phase angle. “L” stands for light, “H” for heavy atoms (with anomalous scattering).

A) In case of non-anomalous scatterers Friedel's law is obeyed because $F_L = \sum_j f_{0j} \exp(i\phi_j)$, the phases of each reflection at opposite directions change sign, $\phi_{h,k,l}$ is equal to $-\phi_{-h,-k,-l}$ and $I_{h,k,l}$ and $I_{-h,-k,-l}$ proportional to $|\vec{F}_L|^2$ are equal.

B) In case of scattering between anomalous scatterers, $\vec{F}_H = \vec{F}_0 + \vec{F}_H' + i\vec{F}_H''$ with $\vec{F}_0$ being the contribution of f_0 , $\vec{F}_H'$ of f' , $\vec{F}_H''$ of (f'') and if there is just one type of anomalous scatterer (f'/f'' are constant) $\vec{F}_0$ and $\vec{F}_H'$ are colinear ($\Delta\phi = 180^\circ$) and $\vec{F}_0$ and $\vec{F}_H''$ are perpendicular ($\Delta\phi = 90^\circ$). This changes the phase between $\vec{F}_H^{(+)(-)}$ but their intensities are equal.

C) In case of reflections between anomalous and non-anomalous scatterers $\vec{F}_{LH} = \vec{F}_L + \vec{F}_{0(h)} + \vec{F}_H' + i\vec{F}_H''$ the non-imaginary contributions $\vec{F}_L$ and $\vec{F}_{0(h)}$ are not colinear to each other, the imaginary contribution $i\vec{F}_H''$ isn't perpendicular to the real vector sum and absolute phases and intensities of inverse reflections will differ. These differences will be at maximum, when $\vec{F}_L$ and $\vec{F}_{0(h)}$ are close to being perpendicular. [81], [75]

2.8.7 Data reduction and quality control

Once Laue-group and point-group symmetry are chosen, measured data get reduced and analyzed. Intensities of reflections get integrated, unique reflections merged, all intensities put on a common scale and the correlation of equivalent reflections scored based on the chosen symmetries and data subsets. Several correction factors can be applied in order to reduce experimental variation, e.g. crystal size or quality in different directions (anisotropy), variation in beam intensity, time-dependent radiation damage etc.. Sub-sets with lower quality can be discarded, e.g. high and low resolution shells, shells compromised by ice rings, beam stop areas or shadows, late recorded diffraction images compromised by strong radiation damage, bad images resulting from the experimental setup or crystal defects in certain directions/time of measurement or single reflections belonging to another crystal lattice. For data quality assignment several calculations can be performed based on the number of reflections n_h of the reflection h :

R_{merge} indicates differences between single observations, which rises with dataset multiplicity and gets calculated as:

$$R_{\text{merge}} = \sum_h \sum_l |I_{hl} - \langle I_h \rangle| / \sum_h \sum_l \langle I_h \rangle .$$

R_{meas} is an enhanced multiplicity corrected version of R_{merge} reporting consistency of individual measurements with:

$$R_{\text{meas}} = \sum_h \left(\frac{n_h}{n_h - 1} \right) \sum_l |I_{hl} - \langle I_h \rangle| / \sum_h \sum_l \langle I_h \rangle .$$

and $R_{\text{p.i.m}}$ is an R factor that reports the precision of averaged reflections and is independent of higher multiplicities, with:

$$R_{\text{p.i.m.}} = \sum_h \left(\frac{1}{n_h - 1} \right) \sum_l |I_{hl} - \langle I_h \rangle| / \sum_h \sum_l \langle I_h \rangle .$$

Other coefficients to assign data quality are Pearson's correlation coefficients (CC) which compare independent sets of observations within the same resolution ranges and range from 0 to 1 with higher precision (after proper indexing of the data). CCs are a coefficient of statistical probability that the measured values are precise. $\text{CC}_{1/2\text{Anom}}$ and $\text{CC}_{1/2}$ are CCs between halve datasets. $\text{CC}_{1/2\text{Anom}}$ correlates anomalous differences between halve datasets, whereas $\text{CC}_{1/2}$ defines the correlation of intensities. Both decrease with lower signal to noise ratio (I/σ) at higher resolution and outliers in between shells need to be explained e.g. ice rings. Using the $\text{CC}_{1/2}$ as indicator for a high resolution cut-off instead of (I/σ) values has the advantage of higher reproducibility, as the (I/σ) output after

data reduction with different programs may differ significantly based on parameterization, algorithms, correction factors used etc.. $CC_{1/2}$ as used in this study can be calculated by:

$$CC_{1/2} = \frac{(\sigma_y^2 - 1/2 \sigma_\epsilon^2)}{(\sigma_y^2 + 1/2 \sigma_\epsilon^2)} ;$$

with σ_y^2 being the variance of merged intensities y of unique reflections and $1/2 \sigma_\epsilon^2$ being the variance of the random errors ϵ between observations. [84], [85], [86]

A series of studies have shown improvement in electron density maps and refined models with datasets utilizing high resolution shells with $CC_{1/2}$ as small as down to 0.1 even with a R_{meas} of 1000 % or I/σ values around 0.3. In recent years $CC_{1/2}$ increased as a parameter for high resolution truncation, instead of truncation at R factors (conservative >0.6) or I/σ (<1.0) especially when data with high multiplicity is merged and modern detectors with high sensitivity are used. In recent years $CC_{1/2}$ values above 0.5 were widely accepted as indicator for good quality data. [85]

2.9 Phasing

In order to reconstruct the electron density $\rho(x, y, z)$ from X-ray diffraction data, the so called “phase problem” has to be solved. The electron density can be described as the Fourier sum of the unit cell’s structure factors, as:

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} F_{hkl} \exp(-2\pi i(hx + ky + lz)) \quad [75]$$

where each structure factor F_{hkl} can be described by its frequency, phase and amplitude. The frequencies of each F_{hkl} depend on wavelength and distances within the crystal (Figure 2.5) and the frequency terms h , k and l are the indices of planes producing the reflection. Its amplitude is the square root of the measured intensity and the phases need to be determined indirectly. [75]

By knowing the wave function of $F = |F| \cdot \exp(i\alpha)$ and substituting α with $\alpha' = \alpha/2\pi$ we

get $\rho(x, y, z) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp(-2\pi i(hx + ky + lz - \alpha'_{hkl}))$ including terms for the known frequencies, intensities and the unknown phases α' . There are several methods in order to yield limited numbers of initial phases which, if they are correct enough, can be used as estimates for further phase extension to all F and further improvement. [75]

There are direct methods which use statistical relationships between structure factors and their sums, but direct phase calculations are computational intensive and usually not used for unit cells larger than one hundred atoms. [63]

An other method is the estimation of phases by “*Patterson synthesis*”, which uses a Fourier sum without phases. The resulting so called “*Patterson map*”, with density in the coordinate system u, v and w , displays vectors between scatterers. Thus a Patterson map has more peaks ($n(n-1)$, so called “*Patterson atoms*”) than atoms, due to multiple vectors between atomic positions. It becomes useful when introducing a limited number of heavy atoms (with higher electron density and diffraction contribution) to a protein crystal. These few heavy atoms can be distinguished from the light atoms in the Patterson map and their likely atomic coordinates (within the unit cell) can be calculated. [63], [75]

The Patterson function is given by

$$P_{u,v,w} = \frac{1}{V} \sum_{hkl} |F_{hkl}|^2 \exp(-2\pi i(hu + kv + lw)) ,$$

the difference Patterson amplitudes correspond to $(|F_{LH}| - |F_L|)^2$ and the map can get reduced from peaks of vectors between the light atoms if the native from derivative data get subtracted.

From the number of vectors of large scatterers the number of heavy atoms can be estimated and their three dimensional atomic coordinates been calculated. This method using heavy atoms without X-ray absorption is referred to as “*isomorphous replacement*” and usually uses a difference Patterson map between data from at least one heavy atom derivative and a native dataset of a protein with both isomorphous unit cells. Because Patterson atoms correspond to vectors between atoms, two inversed atomic coordinates and set of phases will yield the same solution. The correct set can be determined based electron density calculations with each phase set and the likelihood of their maps to be a correct solution (see Chapter 2.9.1). [81], [75],

Once the three dimensional coordinates of heavy atoms (or some of them) are known within the unit cell it is possible to calculate their structure factors F_H including the phases F_{LH} which then can be used to calculate the structure factors F_L since $F_L = F_{LH} - F_H$ using SIR “*single isomorphous replacement*” with native and one derivative datasets (Figure 2.10A). [81], [75]

If the protein crystal contains heavy atoms (e.g. by using selenomethionine incorporation during protein synthesis^[87]) that absorb X-rays and are measured at a wavelength corresponding to the heavy atoms absorption edge, the procedure is referred to as SAD “*single anomalous dispersion*” phasing, where only a single dataset is needed to solve the phase problem. As explained previously during anomalous dispersion Friedel's Law is broken and structure factors of anomalous scatterers have a real $\vec{F}_H'$ and imaginary $\vec{F}_H''$ correction factor. The magnitudes of these correction factors are known for every element and heavy atom positions can be calculated by Patterson

synthesis. With the positions of the heavy atoms their phases are known and will be used to estimate the phases F_{LH} and F_L using the amplitudes $|\vec{F}_{LH}|^{(+)}$ and $|\vec{F}_{LH}|^{(-)}$ (Figure 2.10 B).^{[81], [75], [88]}

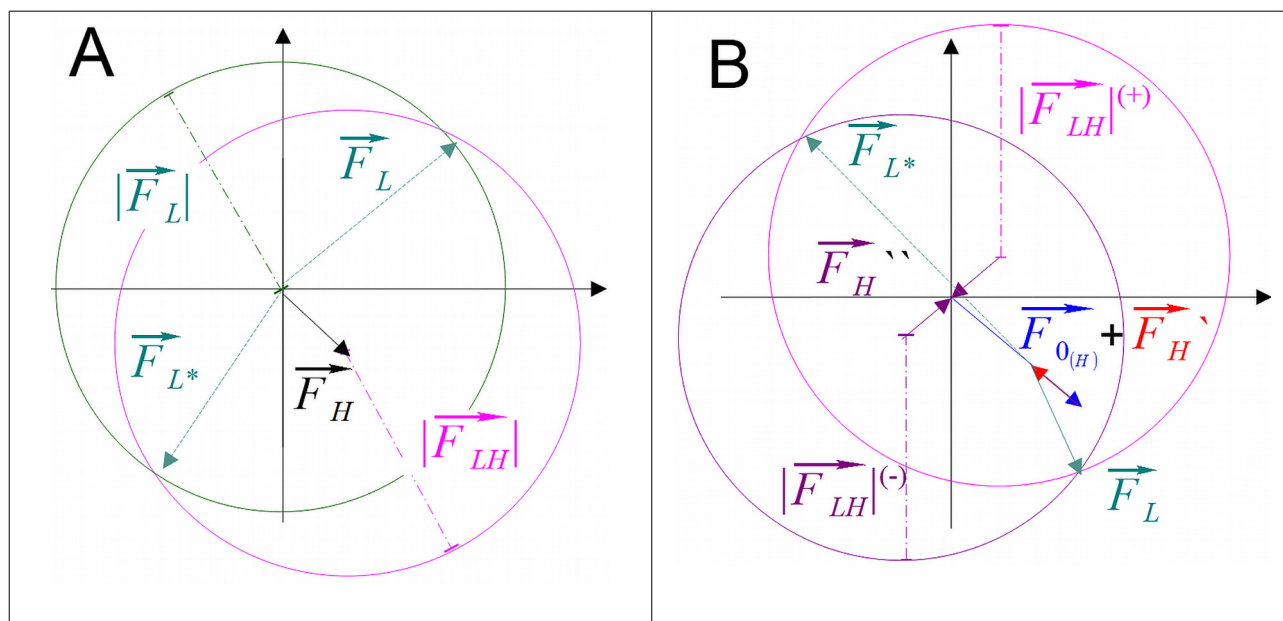


Figure 2.10: Harker diagram for phasing by isomorphous replacement (A) or anomalous dispersion (B)

A): Once the structure factor of F_H (H for heavy atoms) is calculated by Patterson synthesis, F_L (L for light atoms) can be calculated with a native and a heavy atom containing isomorphous dataset (SIR, single isomorphous replacement). Because $\vec{F}_L = \vec{F}_{LH} - \vec{F}_H$ and each $\vec{F} = |\vec{F}| \cdot \exp(i\alpha)$, the unknown phases can take any value around the shown cycles which represent the vector sums of all α with amplitudes $|\vec{F}_L|$ and $|\vec{F}_{LH}|$, respectively. There where the vector sums of potential $\vec{F}_L$ from both dataset overlap are two possible solutions $\vec{F}_L$ and $\vec{F}_{L*}$.

B): From a heavy atom containing dataset under X-ray absorption a single dataset can solve the phase problem (SAD, single anomalous dispersion). $\vec{F}_L = |\vec{F}_{LH}| \exp(i\alpha) - \vec{F}_H$ as in A) but there are two different measured amplitudes for $|\vec{F}_{LH}|$ due to $\vec{F}_H = \vec{F}_0 + \vec{F}_H' + i\vec{F}_H''$. Thus the vector sums of $\vec{F}_L = |\vec{F}_{LH}| \exp(i\alpha) - \vec{F}_{0(H)} - \vec{F}_H' - i\vec{F}_H''$ result in two possible solutions $\vec{F}_L$ and $\vec{F}_{L*}$ like in A).

^{[81], [75], [88]}

2.9.1 The hand problem and phase improvement

Both methods (SIR and SAD) result in two possible solutions with only one correct set of phases, referred to as "*the hand problem*" in phasing. There are two different approaches to solve the hand problem. First by try-and-error approach calculating the electron densities with both sets, modeling and phase extension and a calculation of likelihood of resulting maps as well as personal inspection for characteristic features like appearance of densities of protein helices in proteins. Second is the use of further datasets to add more informations with differing vector sums (additional circles in Figure 2.10A or B) of F_L . In MIR "*multiple isomorphous replacement*" multiple heavy metal derivatives with different F_H are used to yield a single solution. In SIRAS a "*single-*" or in

MIRAS “*multiple isomorphous replacement*” isomorphous replacement is combined with anomalous diffraction. In MAD “*multiple anomalous diffraction*”, SAD phasing is combined with datasets taken at different wavelength, such as a) without X-Ray diffraction (remote of X-Ray absorption edge), where $\vec{F}_H = \vec{F}_0$, b) at the absorption edge, where $i \vec{F}_H''$ is maximal and c) at the inflection point of the absorption curve, where the $\vec{F}_H'$ contribution is maximal. [75]

The reason for adding apparently redundant phase information, in particular combining of methods or additional datasets, when the previous ones already yield a single phase solution is to compensate on measurement errors and non-ideal data quality. Due to the uncertainty of the heavy atom positions, their structure factors, the exact amplitudes etc. the vector sum solutions do not result the “correct” single phase value but rather a likelihood region. By adding additional phase solutions random error effects can be reduced and the likelihood to reach correct phase solutions be increased. The likelihood is further increased by refining the atomic positions, occupancies, b-factors and correction factors of the heavy atoms, the amplitudes derived from scattering and iterative sub-structure completion and refinement, to narrow down the regions of correct atomic positions and phases. [73], [89]

2.9.2 Molecular Replacement (MR) and MR-SAD

If parts of a protein structure are already known, like single domains, homolog proteins, refined models from other crystal datasets of the same protein (e.g. with different ligand etc.) these models can be used as “phasing-” or “searching models” for initial phasing. In MR-SAD this phase information is used to improve the determination of heavy atom positions, if their anomalous data alone is not sufficient enough. Alternative an incomplete heavy atom substructure can already yield enough phase information that automatic building programs like Buccaneer^[9], ARP/wARP^[90] or Autobuild^[12] are able to identify regions of protein secondary structures (especially helices). This initial models can then be used to aid in finding additional heavy atom sites, improving subsequent iterative SAD phasing. If the provided initial model and the diffraction data are good enough, a protein structure can be solved from a single dataset by molecular replacement alone, without any additional SAD phasing. Therefore, if the model structure is isomorphous, the phases derived from the model are used directly for electron density reconstruction (by Fourier synthesis) instead of experimental phases:

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} |F_{TARGET\ hkl}| \exp(-2\pi i(hx + ky + lz - \alpha'_{(MODEL\ hkl)})) \quad (\text{Chapter 2.9})$$

If the unit cell, position and/or orientation of model and target structure differ, they first have to be superimposed with the target structure (by rotation and translation). The unit cell dimensions and symmetry of the target structure are already known from indexing (Chapter 2.8.4) and space group determination (Chapter 2.8.5) steps. The position and orientations of the model are computationally screened in order to give the best fit between calculated structure factors F from the model (F_c ; c: calculated) and observed F from the data (F_o ; o: observed). To calculate the best orientation of the model, the peaks in the Patterson map (Chapter 2.9) (which is independent of the atom positions in the unit cell, but not their orientation) are compared during rotational searches. Once the best orientation between model and target is found by Patterson peak comparison, the best positions/location in the unit cell are determined by translational search, comparing F_c and F_o . Comparison between F_c and F_o are usually quantified by the so called R-factor, given as:

$$R = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|} ,$$

where an R factor of 0 describes identical F_o and F_c , whereas R factors of 0.6 are observed between random protein models (0.8 for centric data).^{[73], [91], [89], [75]}.

2.9.3 Model building and refinement

If a structure could be phased by various methods (Chapters 2.9-2.9.2) to yield an interpretable electron density map, models can be further improved or *ab initio* build to yield better phases. These models then are improved iteratively and lead to better electron density maps and a final model. This process of iterative structure building, map fitting and refinement includes real-space and reciprocal space methods: In real space calculated electron density maps are used to manually extend and/or correct positions of residues and ligands in the model, using the known protein sequence, electron density features and known bond geometries that restrain positions of neighboring atoms. This map-fitting is commonly done against two kind of electron density maps,

$$\begin{aligned} \text{Fo-Fc: } \rho(x, y, z) &= \frac{1}{V} \sum_{hkl} (|F_o| - |F_c|) \exp(-2\pi i(hx + ky + lz - \alpha_{MODEL})) \\ &\text{and} \\ \text{2Fo-Fc: } \rho(x, y, z) &= \frac{1}{V} \sum_{hkl} (2|F_o| - |F_c|) \exp(-2\pi i(hx + ky + lz - \alpha_{MODEL})) . \end{aligned}$$

In the Fo-Fc map errors between model and data are more visible in the calculated difference density map. Thereby it is better suitable to correct wrongly build regions, whereas in the 2Fo-Fc Fourier transformation a more continuous surface density is calculated, better suitable for model

extension. In reciprocal space the model structure is computational refined to decrease the structure factor disagreement Φ between F_o and F_c by refining the atomic coordinates, atomic b-factors and/or occupancies. Usually, if least square methods are used (to find the local minima)

$$\Phi = \sum_{hkl} \omega_{hkl} (|F_o| - |F_c|)_{hkl}^2$$

where ω are specific weighting factors, depending on methods and programs used, and of all diffracting atoms j , for which:

$$F_c = G \cdot \sum_j n_j f_j \exp(2\pi(hx_j + kz_j + ly_j)) \cdot \exp(B_j[(\sin\theta)\lambda]^2) ,$$

with G being a scale factor, n_j being the atoms occupancy, f_j their scattering factor, B_j their B-factor and x, y, z their coordinates and $[(\sin\theta)\lambda]$ the angle of reflection. As it is done in real space refinement, restrains e.g. on bond geometries and energy minimization are applied as well. Alternating real and reciprocal space methods are usually used to yield the final pdb model of a protein structure, which may include its ordered hydration shell and bound ligands, until the calculated electron densities are explained by the model with reasonable chemical bond geometries, energies, distances etc.. Remaining F_c and F_o discrepancies can be explained by crystal defects and measurement uncertainties.

The convergences of molecular model improvement at each step of refinements are quantified by R-factors (Chapter 2.9.2), R_{work} and R_{free} . R_{work} are calculated from F_o of reflections, the structure was refined against, whereas R_{free} includes only a random small set of reflections that was excluded from refinement to avoid overfitting and represents a lesser biased validation reference. ^[75]

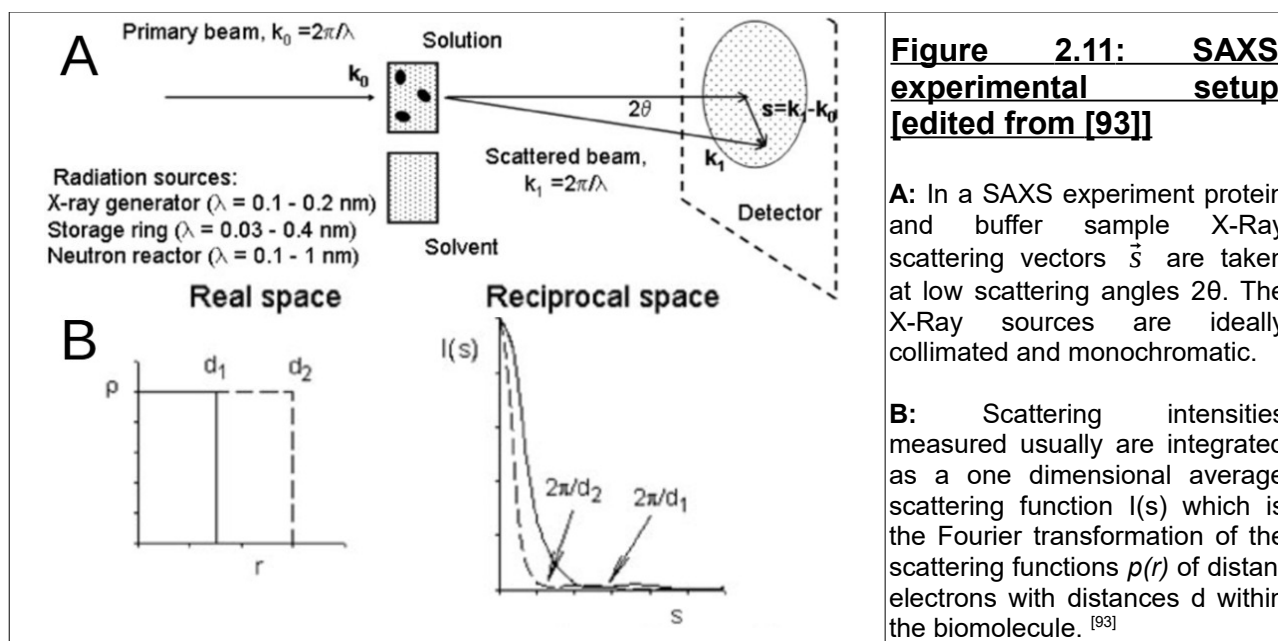
2.10 Small-angle X-Ray scattering (SAXS)

SAXS experiments were performed at BM29 BIOSAXS beamline at the European Synchrotron Radiation Facility (ESRF Grenoble, France) in batch mode. Samples were prepared by ultrafiltration of PorX monomer to final concentrations between 15.5 and 17 mg/ml (in *SEC-Buffer*) and were then diluted with appropriate buffers - *SEC-Buffer* for PorX-monomeric samples and *BeF₃-Buffer* for PorX dimeric samples (both without zinc addition). Final concentration of PorX monomeric samples were 0.25, 0.5, 1.0, 2.0, 5.0 and 10 mg/ml and final concentrations of PorX dimer were 0.5, 1.0 and 2.0 mg/ml. Zinc containing samples were concentrated from an initial concentration of 0.5 mg/ml PorX in *SEC-* (Monomer+Zn) or *BeF₃-Buffer* (Dimer+Zn) in the presence of 43 μ M $ZnCl_2$ to final concentrations between 2-3 mg/ml. Samples were then diluted to 0.5, 1.0 and 2.0 mg/ml each.

Bioxtas RAW software package^[24] was used to analyze and process data including data reduction, Guinier analysis and MW estimations^{[25], [26], [27], [28]}. Indirect Fourier Transform (IFT) was performed by GNOM^[30] from ATSAS 3.0 software package^[29]. Monomeric PorX datasets taken at 0.5, 2.0 and 10.0 mg/ml were superimposed, respective regions with best signal to noise ratio selected and merged into a single curve. Monomeric PorX samples at 0.5 and 2.0 mg/ml (which showed different behavior at the Guinier region) were processed separately and the data at 1.0 mg/ml was discarded due to polydispersity. For BeF₃-PorX samples, the best datasets at low, as well high q region has been used for further processing and analysis. These were 2.0 mg/ml for BeF₃-PorX and the 1.0 mg/ml for dimeric BeF₃-PorX with zinc. Low q regions showing intramolecular interactions were discarded from further analysis at each dataset. *Ab initio* electron density reconstruction was performed by DENSS^[31] using the Indirect Fourier Transform (IFT) calculated by GNOM. For the electron density maps of monomer and dimer without zinc 100 *ab initio* maps were calculated and averaged in "slow mode" with default parameters and refined against the P(r) function with an oversampling ratio of 5 to optimize map quality. Manual modeling of monomeric PorX was performed with CHIMERA v1.16^{[16], [45]} against the electron density calculated by DENSS. Refinements and model validation were performed with SREFLEX^[32] and CRY SOL^[33] with 1000 points of computed SAXS curve, 33 maximum order of harmonics (data based suggestion by CRY SOL) and constant background subtraction has been applied.

2.10.1 SAXS-Theoretical background

SAXS is a X-Ray diffraction method (diffraction theory see Chapter 2.8.2) to study the structure of proteins in solution, where diffraction contribution only derives from single particles without interparticle order. Compared to X-Ray crystallography, where the molecules have a long-range order resulting in high resolution, SAXS only yields data at low resolution (usually between 10-50 Å). As advantages SAXS does not need high quality protein crystals, the proteins can be measured at low concentration in solution, as they usually are functional *in vivo*. Thus SAXS acts as a complementary method to X-ray crystallography to avoid interpretation of crystallization artifacts (e.g. higher oligomers), potential analysis of flexible regions and conformational changes that might be invisible in a crystal structure or as aid in solving the phase problem (Chapter 2.9). SAXS requires protein samples that ideally stay soluble at rel. high concentrations (typically a few mg/ml for better signal to noise ratio) with less interparticle interactions/high monodispersity.^[92]



In Figure 2.11A a typical SAXS experimental setup is shown. Because the electron density differences and resulting scattering intensities I between aqueous buffers and protein solutions are rather small, SAXS sample data need a background subtraction using a buffer sample to determine the protein scattering curves. In contrast to X-Ray crystallography, during SAXS experiments biomolecules are arranged randomly in solution and only averaged scattering is measured over all orientations at the detector. Thus only information of the magnitude $q = (4\pi/\lambda) \cdot \sin(\theta)$ of the scattering vector $\vec{s}$ (Figure 2.11 A, cf. Figure 2.11) can be collected (hereinafter s and q are used as synonyms in diagrams) and the two dimensional images can be integrated about the beam center to plot all collected information in a one dimensional scattering curve I versus q . [94], [93]

The measured intensities $I(q)$ derive from the squared amplitude A of the detected wave, which derives from the sum of all interfering waves at their specific q at the detector:

$$I = |A|^2 = \left| \sum_n a_n \exp(i\phi_n) \right|^2$$

which in turn depend on the electron densities within each molecule in all directions, therefore:

$$I(q) = \left| \int \int \int \exp(iqy) \rho \, dx \, dy \, dz \right|^2$$

with y being the coordinate perpendicular to the plane oriented at the diffraction angle with constant phases $\phi_{(2\theta)}(x, y, z) = \phi_0 + qy$ and ρ being the electron density. Thus the intensity derives from the squared Fourier transformation of the electron density of a molecule. For smaller q all scattered waves from a particle are maximal in phase and thus the intensity is maximal, whereas for large q the scattered waves of the particle are out of phase and intensity decreases due to destructive

interference. The decrease in quality depends on the form factor of the particle, which in turn depends on its shape and volume. ^[95]

2.10.2 Kratky plot & Porod equation

The folded/unfolded state of a protein can be estimated by plotting $I(q)q^2$ versus q , the so called “Kratky plot” (Figure 2.12, ^{[96], [93]}), which can be used to determine protein sample quality, e.g. for subsequent X-Ray crystallization experiments. While globular single domain proteins show a distinct peak at low q (Figure 2.12, (1)), unfolded proteins show a continuous increase at higher angles (Figure 2.12, (2)-(4)) and the Kratky plot of multidomain proteins has a similar shape as partially unfolded proteins (Figure 2.12, (2)). ^{[92], [93]}

The hydrated particle Volume V_p “Porod volume” can be approximately calculated by the “Porod equation” (Porod, 1982) with Q being the concentration independent “Porod invariant” Q :

$$V_p = 2\pi^2 I_0 / Q \quad \text{with} \quad Q = \int_0^\infty q^2 I_q \cdot dq$$

Q is therefore the integration of the area of the Kratky curve over all angles and, because only limited datapoints for $I(q)$ can be measured, Q is approximately calculated by

$$Q = \int_0^{q^*} I_q q^2 dq + k_1 / q^*$$

with q^* being a sufficient large enough angle and k_1 the so called “end tail constant”. k_1 is determined by the plot $I(q)$ versus q^{-4} which oscillates around a straight line (for high q) with slope k_1 (Porod, 1951). $I(0)$, the so called “forward scattering intensity”, can be calculated by Guinier analysis (see below), it depends on the square number of electrons in the biomolecule and is proportional to its molecular mass. ^{[97], [98]}

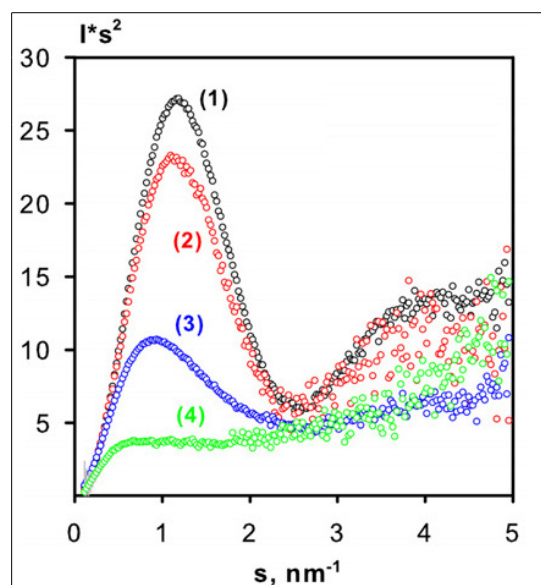


Figure 2.12: Kratky plot ^[92]

The graph shows the Kratky plot with experimental data from lysozyme at different folded states: 1 (black):folded, 2 (red) partially unfolded in 8 M urea, 3 (blue) partially unfolded at 90°C, 4 (green) unfolded lysozyme in 8M urea at 90°C. ^[92]

2.10.3 Guinier analysis

By plotting $\ln I$ against $(q)^2$, we obtain the “Guinier plot”. By extrapolating S to 0 we can calculate the intensity $I(0)$ and the radius of gyration R_g of a molecule, which is the root mean square of all distances of electron distances from their center of gravity^[98].

Following the “Guinier equation”^[99]

$$I(q) = I(0) \exp\left(-\frac{1}{3} R_g^2 s^2\right)$$

$$\text{or: } \ln(I(q)) = \ln(0) - \left(\frac{R_g^2}{3}\right) q^2 ,$$

R_g is proportional to the slope of the tangent with a scattering angle towards zero while $I(0)$ is the intercept with the y-axis. For biological macromolecules this approximation yields reasonable estimates of R_g for $q < 1.3/R_g$, while for higher q the linear dependency is increasingly broken as it does in the presence of

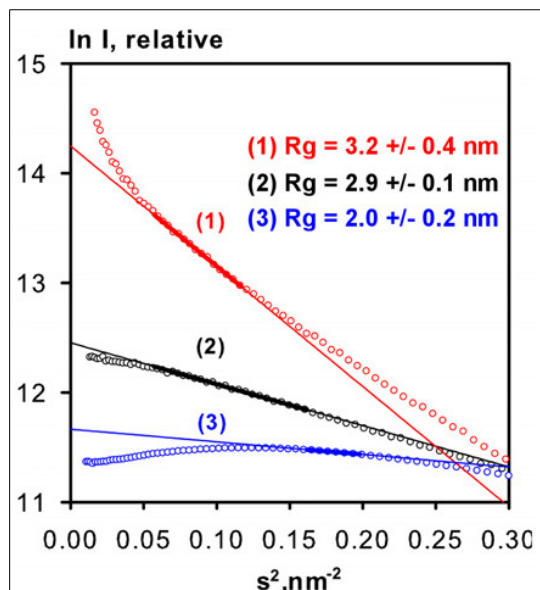


Figure 2.13: Guinier plot ^[92]

The graph shows the Guinier plot with experimental data of BSA samples. 1 (red) shows aggregated sample, 2 (black) shows a sample of good quality and 3 (blue) shows interparticle repulsion. ^[92]

unstable interparticle interactions or polydisperse samples (see Figure 2.13). ^{[92], [93]}

It is often necessary to measure a dilution series of samples to get the best Guinier fit at low concentrations for low q , to avoid effects of inter-particle interactions, while at high concentrations obtaining accurate high q , since the intensities at high q get increasingly weak. The relation of the R_g compared to the radius of a molecule depends on the shape, the radius of a sphere equals

$$\sqrt{5/3} R_g \quad \text{and that of a rod } 2\sqrt{3} R_g . \quad [98], [100]$$

2.10.4 Molecular weight estimations

With SAXS the molecular weight (M_w) of globular proteins can be calculated due to several methods, each having an average error of around 10 % depending on protein concentrations, calibration errors, size and shape of the protein as well as experimental errors and error propagation in the analysis. Thus often crossvalidation by independent methods is used for M_w estimation. ^[26],

^{[92], [101] [102], [27]}

One method is to calculate M_w from $I(0)$ of the Guinier analysis (Chapter 2.10.3) by using a standard (usually water or BSA) with known concentration (c) and M_w . The M_w of a measured protein (p) with known concentration compared to a standard protein (“st”) is:

$$Mw_p = \frac{I(0)_p}{c_p} \frac{Mw_{st}}{I(0)_{st}/c_{st}}$$

and it needs precise concentration measurements of both sample- and reference protein as well as similar protein-solvent contrasts.^[103]

When using water as a standard the Mw(p) can be determined “on absolute scale” by:

$$Mw_p = \frac{N_A}{c_p} \frac{I_0}{\Delta \rho_M^2} \quad [103]$$

with N(A) being the Avogadro number $6.023 \cdot 10^{23}$ and the scattering contrast per mass $\Delta \rho_M = [\Delta \rho(\text{protein}) - \Delta \rho(\text{solvent}) \cdot v] r_o$ with an electron density $\rho(\text{protein})$ of $3.22 \cdot 10^{23} \text{ e/g}$ (per mass) and $\rho(\text{solvent})$ $3.34 \cdot 10^{23} \text{ e/cm}^3$ (per volume), v the specific volume of the protein and $r_o = 2.8179 \cdot 10^{-13} \text{ cm}$ being an electrons scattering length. The Mw estimation of this method depends on exact concentration measurements, molecule contrast and its partial specific volume, as well as well performed calibration.^{[103], [104]}

An other method is by using the (q-) adjusted Porod Volume V_p where the molecular weight is equal to V_p multiplied by the proteins average mass density $\rho(m)$

$$Mw_p = V_p \rho_m = 2 \pi^2 \frac{I(0)}{Q} \rho_m, \text{ with } Q \text{ being the invariant over } q: \quad Q = \int_0^\infty I(q) q^2 dq. \quad [105]$$

q data up to a selectable q_{max} (usually q_{max} reliable for up to $8/R_g$) can be used for $q \rightarrow 0$ extrapolation in order to calculate the Mw of proteins, using an average mass density $\rho(m)$ of 1.37 g/cm^3 :

$$Mw_p = \frac{\rho_m V}{1.662 \cdot 10^{-21}}.$$

This method is concentration independent but works best for compact globular proteins without conformational flexibility.^[101]

For molecules undergoing conformational changes or partially unfolded proteins the Mw can be better estimated using the volume of correlation V_c , which is defined as the ratio of $I(0)$ to the total scattered intensity and is equal to the molecules volume divided by its self-correlation length:

$$V_c = \frac{I_0}{\int [q \cdot I(q)] dq} = V_p / (2 \pi l_c).$$

The molecular weight of proteins can be calculated from V_c by: $Mw(V_c) = \frac{V_c^2}{R_g} / 0.1231$.

This method is less reliable for noisy high q data, for protein nucleic acid complexes or proteins smaller 20 kDa.^{[106], [27]}

2.10.5 Distance distribution function $p(r)$

Alternatively to the Guinier analysis, R_g and $I(0)$ can be calculated by the distance distribution function $p(r)$, which is the inverse transformation of $I(q)$ (Figure 2.11B):

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty (q^2 I_q \sin qr / qr) dq$$

This function represents a molecules scattering in real space and the shape of the molecule can be estimated more intuitively as from $I(q)$ itself, by plotting $p(r)$ to their distances (Figure 2.14). While globular proteins generate a bell shaped curve with a maximum at their halve maximum electron distance, multi-domain proteins show several maxima due to internal and inter domain distances.^[93]

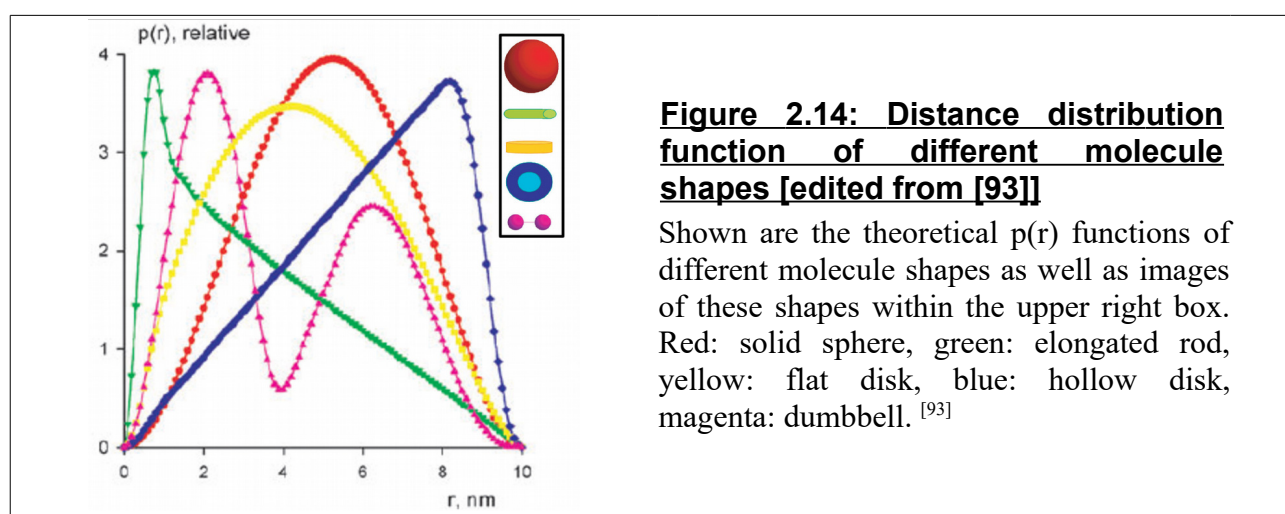
As data points q_i at q are limited ($q_{\min}$ to $q_{\max}$, rather than 0 to infinite) the Fourier transformation is being applied with the variable $D_{\max}$ ^[98]. The range from 0 to $D_{\max}$, $p(r)$ is defined as summation of K orthogonal functions $\phi_K(r)$:

$$p(r) = \sum_{k=1}^K c_K \phi_K(q_i),$$

with optimal c_K minimizing $\phi = \chi^2 + \alpha P(p)$ and χ^2 the sum of least square differences between experimental data and the direct Fourier transform of $p(r)$, α a weighting factor and $P(p)$ a term describing the $p(r)$ functions smoothness according to

$$I(0) = 4\pi \int_0^{D_{\max}} p(r) dr = (\Delta\rho)^2 V^2 \quad \text{and} \quad R_g = \left(\int_0^{D_{\max}} r^2 p(r) dr \right) / \left(2 \int_0^{D_{\max}} p(r) dr \right).$$

The values of $I(0)$ and R_g derived from the distance distribution function are usually more reliable due to data-cut off at low angles and can be used as cross-validation to the Guinier fit.^{[98], [92], [93]}



2.10.6 Structural reconstruction and validation

$p(r)$ functions contain information of the one dimensional amplitudes of the Fourier transform but phases are unknown. Thus $p(r)$ can only be used directly for validation of a crystal- (or homology-, EM- or NMR-) structure by comparing the theoretical scattering profile (via inverse Fourier transformation IFT) and the experimental data. In order to obtain three dimensional structural information, structure factors need to be iteratively retrieved by modeling algorithms. ^[31]

For subdomain modeling, theoretical scattering curves of domains, bead models or dummy residues, can be calculated by a spherical harmonics and refined with the amplitudes $A_{lm}(q)$ being the partial scattering amplitudes of the substructures, such that:

$$I(q) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l |A_{lm}(q)|^2$$

with minimal discrepancies χ or χ^2 between the theoretical SAXS curves and the experimental data. The algorithms usually assume a connected spherical structure (for proteins), solvent/hydration corrections, no overlaps, maximum dimensions of $D(\max)$, optional simulated annealing (depending on the method) and usually start with random distributions of position and orientations. Due to the ambiguity of compatible solutions (e.g. enantiomers, different conformations, flexible regions), these are often merged to yield better single models, ensembles or mixtures that render lower χ when compared to the experimental data. ^[92]

One program for structure evaluation is CRY SOL^[33] by Svergun, Barberato and Koch (1995). It calculates the hydration shell and the spherical averaged scattering intensities $I(q)$ from an atomic model it calculates in solution and compares it to the experimental SAXS data. ^[33]

SREFLEX^[32], a program by Panjkovich and Svergun (2016) uses a hybrid modeling approach using normal mode analysis. It iteratively divides a model into flexible “pseudo-domains” and refines their geometry (first with rigid body restrains, later on without) against the SAXS data to account for conformational changes and protein dynamics in solution. Pseudo-domains can be chosen manually or automatically and conformations that show clashes or breaks are filtered out. CRY SOL is used for validation and submodel selections by minimizing χ^2 calculated as:

$$\chi^2 = \frac{1}{N} \sum_{i=1}^{N_p} \left(\frac{I_e(q_i) - cI_t(q_i)}{\sigma(q_i)} \right)^2,$$

with $I(e)$ and $I(t)$ being experimental and theoretical calculated intensities, $N(p)$ number of experimental data points, σ experimental errors and c a scaling factor. ^[32]

2.10.7 DENSS Electron density reconstruction

A significant drawback of these modeling approaches, is that the resulting models assume uniform electron densities within the calculated model envelopes, which can not account for structural flexibility, dynamics or stability of subdomains (like refined B-factors used in crystal structures explaining high and low electron density regions). In order to account for these subregions alternatively (or in parallel to a structural modeling approach) a three dimensional electron density reconstruction can be applied. ^[100]

One method called “Density from Solution Scattering” (DENSS)^[31] developed by T.D. Grant uses an iterative 3D structure factor retrieval algorithm (Figure 2.15). In a first step a Grid of $N \times N \times N$ voxels with suitable dimensions (2-3 times of D_{max}) are used with random intensities to generate a starting random electron density map. By Fourier transformation (FT) this map generates a set of 3D structure factors (Figure 2.15A; with complex amplitude, phase and calculated intensities) at the reciprocal space. By spherical averaging (Figure 2.15B) then SAXS data is emulated. These averaged intensities $I(q)$ are scaled to the SAXS experimental data, which leads to a new set of structure factors, and transformed to yield a real space electron density (by inverse FT, Figure 2.15C). Starting from random electron densities this density might match the experimental data intensities (/restraints) but it has a high rate of obvious errors in real space that can be corrected by real space restrains:

The so called “shrink wrap algorithm” blurs the electron density and sets a support volume of

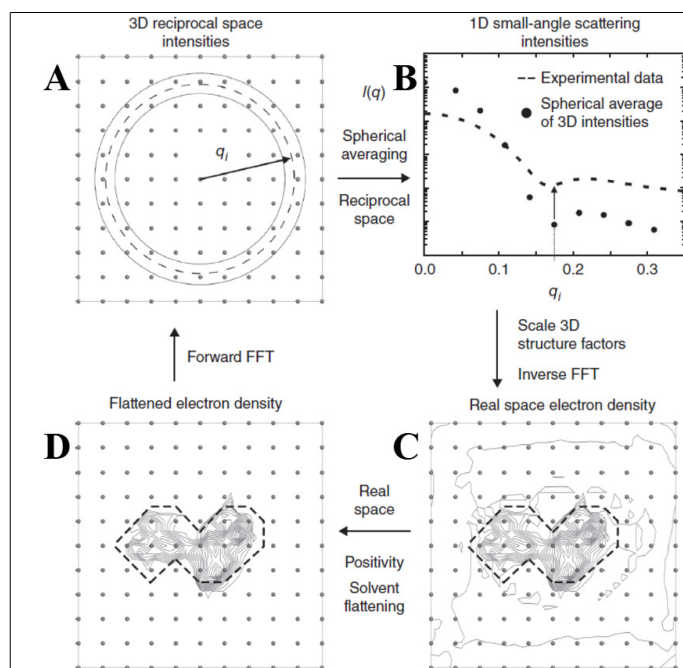


Figure 2.15: Schematic iterative electron density reconstruction steps used by DENSS [edited from [31]]

Ab initio electron density can be calculated by iterative phase improvement starting with a random density map Fourier transformed to 3D reciprocal space intensities (A). Spherical averaging by q-shells lead to simulated SAXS data (B), that are scaled to the experimental data in each q-shell and by IFT transformed to an improved electron density map (C). Applying real space restrains on this map like electron density positivity and solvent flattening (D) finishes the first cycle and lead to a further improved real space map, which will be forwarded into a new cycle of phase improvement until convergence and suitable fit to data and real space restrains.

positive density containing above 20 % of the experimental maximum, while solvent density outside is flattened and cleaned up at the end of the first cycle (Figure 2.15D).^[31] This initial cycle of structure factor retrieval is repeated thousands of times until convergence to optimize the calculated electron densities by repeating the FT, rescaling structure factors, recalculate new densities applying support restrains and solvent flattening to yield one (ambiguous) electron density model that matches the real space restrains as well as the intensity data. In order to reduce ambiguity of the final model between 20-100 models are usually produced and merged into one averaged map, where the comparison of averaged maps of each halve of the calculated models is used to estimate the trustworthy resolution at a Fourier shell correlation of 0.5. In order to finalize the model of electron density, the overall averaged map is refined against the experimental data.^[31]

2.10.8 PglZ ortholog comparisons with BREX-proteins

Apart from PorX there are six reported PglZ subfamilies from the BREX systems 1-6. In order to compare their ortholog sequences, we took the consensus sequences of those PglZ containing proteins from the NCBI database (National Center for Biotechnology Information). Then we predicted their three dimensional models using AlphaFold2^[107]. After monomeric predictions, or when already a monomer was longer than 1000 residues, we subdivided those proteins into subdomains and used their subsequences for dimer predictions of several overlapping regions, which we then joined together. This separated predictions were a necessity to keep the maximal model size below 1000 residues due to server limitations.

The resulting models we then used for structure based sequence alignments by domain regions and secondary structure predictions from JPRED^[108]. We used JALVIEW^[109] for manual comparisons and readjustments of those alignments, guided by automatic sequence alignments by regions using CLUSTALΩ, MAFFT, MUSCLE and Tcoffee.^[110] Phylogenetic trees were calculated by average distances (using Blosom62) within JALVIEW and visualized using Dendroscope^[111]. The identity matrix between PorX and BREX PglZs was calculated using Clustalv2.1^[112] in test mode (calculation without realignment).

Results

NOTES.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

3. Results

The essence of this project was recently published, 5th of December 2022, as an article in *Nucleic Acids Research*. Figures which have been reused here (or vice versa) are indicated accordingly with the following reference with index [0].

0: Claus Schmitz, Mariusz Madej, Zuzanna Nowakowska, Anna Cuppari, Anna Jacula, Mirosław Książek, Katarzyna Mikruta, Jerzy Wisniewski, Natalia Pudelko-Malik, Anshu Saran, Natalie Zeytuni, Piotr Mlynarz, Richard J. Lamont, Isabel Usón, Virginijus Siksnys, Jan Potempa, Maria Solà: "Response regulator *PorX* coordinates oligonucleotide signalling and gene expression to control the secretion of virulence factors.", *Nucleic acids research*, **2022**; DOI: 10.1093/nar/gkac1103.

This PorX project was as a collaboration between the group of Prof. Jan Potempa from the Jagiellonian University in Kraków (Poland) and our research group. Prof. Potempa's former PhD student Dr. Mariusz Madej (now a postdoctoral researcher in his group), as a side project, optimized the expression and purification protocol for PorX (and related proteins) during his PhD thesis and was able to supply us with sufficient amounts of protein for structural studies. Further he performed all enzymatic assays while we focused on structural studies. Whenever results from our collaborators are summarized, which was possible for understanding of this project, they have been refereed to accordingly. Results are ordered chronological as much as possible and written to best describe the progress of the project and decisions made within.

3.1 Preceding project status

Before I started on this project my former colleague Anna Cuppari, (who defended her thesis on an other project of hers) already performed first crystallization trials of monomeric PorX wild-type (*Mono_PorX*) and in vitro phosphorylated PorX (*Dim_PorX*) and she was able to obtain diffracting crystals of *Dim_PorX* from an ammonium sulfate grid screen. The best crystal tested already showed a nominal resolution around 3.5 Å. However, the diffraction data was not sufficient for a structural solution (personal communication). One reason for this insufficient diffraction of the initial crystals could have been the inhomogeneity of the protein samples, as the *in vitro* phosphorylation is reversible due to autodephosphorylation at the receiver domain (Figure 7.5B). Monomeric PorX was highly soluble, thus did not resulted in any crystallization.

We asked our collaboration partners for beryllium fluoride (BeF₃) dimerized PorX samples, to archive higher sample homogeneity and as well for selenomethionine (SeMet) substituted PorX derivative, to solve the PorX structure by SAD phasing later on. When we received this protein samples, for each new protein sample (*Mono_PorX*, *Mono_PorX_SeMet*, *BeF₃_PorX* and *BeF₃_PorX_SeMet*, see Chapter 2.1.10) five 96 well sparse matrix screens (*PAC1*, -2, -3, *Top96* and

Proplex from IBMB-CSIC, Automated Crystallography Platform (PAC), Barcelona, Spain) were set up for initial screening in our group. These screenings were set up by Dr. Arka Chakraborty, also a formed PhD student of our group, during his last months in in our group.

3.2 Initial screenings

3.2.1 Scoring analysis

When I started on the PorX project, the initial 2000 sparse matrix conditions (5 screens with 4 different protein samples) were screened after 4 months, including images taken from the crystallization drops in regular time intervals at our automated crystallization platform. The protein behavior was analyzed depending on salts, buffers, pH, precipitants and additives. For this purpose every condition was scored individually and screen components were analyzed in a statistical approach on their influence on PorX-aggregation (ordered and unordered likewise).

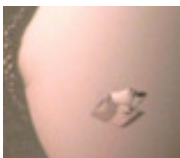
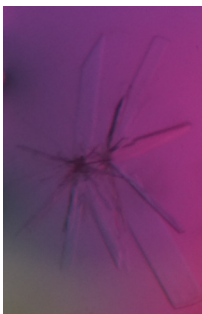
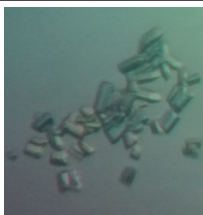
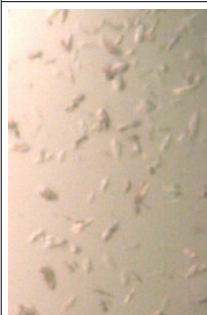
For all 4 protein samples 150 out of the 500 sparse matrix conditions scored in “clear drop” condition (after 4 month, proving at least protein stability in this buffer) or a better score (Scoring see Chapter 2.7.1). 74 of these 150 conditions yielded dense clusters of needles or plates. 36 of them were large enough to be manually broken in single needles or plates, which were measured at the ESRF at two subsequent occasions. One additional condition yielded single crystals but of poor diffraction quality. 2 out of 4 fragments broken out of a single crystal-cluster of BeF₃_PorX diffracted sufficiently to obtain complete datasets of around 2.7 Å (P21 space group) and 3.4 Å (C2) resolution (“*initial datasets*”). Furthermore it was observed that crystalline aggregations grew over a course of time out of initial precipitated protein. Images and description of the best conditions/highest scores from this screening are attached in TABLE 3.1 on the next page.

3.2.2 Structure determination by molecular replacement

Several homology models were used (Chapter 2.8) to solve the structure of initial datasets of selenomethionine unlabeled *BeF3_PorX* by molecular replacement. However these trials did not yield a structural solution and possible solutions found showed only dubious electron densities. Thus our focus moved to the optimization of SeMet labeled protein crystallization for experimental phasing.

TABLE 3.1: Best conditions of initial crystallization screening in 100 nl drop sizes.

The hereby shown kinds of crystallization clusters like sea urchins, needle and plate clusters twined bars or crystalline aggregates were, as it turned out during optimizations, characteristic for PorX crystallization and controlled growth of (or breaking into) single lattice crystals a challenge to overcome.

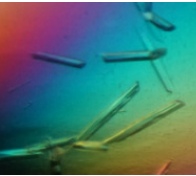
					
				before cryo addition	after cryo addition
A : <u>Crystal type:</u> BeF3_PorX, broken crystals cluster <u>Condition:</u> 0.1 M Tris pH 8.5, 8% PEG 8000; PAC 1, C12 <u>Diffraction:</u> Datasets of 2.7 Å (P21) and 3.4 Å (C1) (at CC½ 50)		B: <u>Crystal type:</u> BeF3_PorX, single crystals <u>Condition:</u> 0.1 M HEPES pH 7.5, 2 M Sodium formate; PAC 1, G9 <u>Diffraction:</u> up to 4.5 Å nominal		C: <u>Crystal type:</u> BeF3_PorX_SeMet, plate cluster <u>Condition:</u> 0.1 M HEPES pH 7.5, 0.2 M LiSO4, 25% PEG 3350, TOP 96, D6 <u>Diffraction:</u> none, bad cryo addition	
	D: <u>Crystal type:</u> BeF3_PorX_SeMet, needle cluster <u>Condition:</u> 0.1 M tri-sodium citrate pH 5.5, 20% PEG 3000, Top 96-D2, <u>Diffraction:</u> none		E: <u>Crystal type:</u> BeF3_PorX_SeMet, crystal cluster <u>Condition:</u> 0.1 M imidazol pH 8.0, 10% PEG 8000, PAC2-G10 <u>Diffraction:</u> 8-10 Å		F: <u>Crystal type:</u> BeF3_PorX_SeMet, small needles <u>Condition:</u> MES pH 6.5 12% PEG 20000 <u>Diffraction:</u> none (too small)
	G: <u>Crystal type:</u> BeF3_PorX, needle cluster <u>Condition:</u> 0.1 M Cacodylate pH.6.5, 1.26 M ammonium sulfate, PAC 2, B1; <u>Diffraction:</u> not tested		H: <u>Crystal type:</u> BeF3_PorX, needle cluster <u>Condition:</u> 0.1 M Na/KPO4 pH 6.2, 10% PEG 3000, PAC2 C8 <u>Diffraction:</u> not tested		I: <u>Crystal type:</u> BeF3_PorX_SeMet, needle cluster <u>Condition:</u> 0.1 M MES pH 6.5, 12% PEG 2000, PAC 1, F10 <u>Diffraction:</u> not tested
	J: <u>Crystal type:</u> BeF3_PorX, crystal cluster <u>Condition:</u> 0.1 M MES pH .6.5, 15% PEG 5500MME, Proplex A5 <u>Diffraction:</u> none		K: <u>Crystal type:</u> BeF3_PorX, crystal cluster <u>Condition:</u> 0.2 M Sodium acetate pH 5.5, 0.1 M tri-sodium citrate. 5% PEG 4000, Proplex A12 <u>Diffraction:</u> none		L: <u>Crystal type:</u> BeF3_PorX_SeMet, crystal bar cluster <u>Condition:</u> 0.1 M CHES pH 9.5, 0.2 M NaCl, 10% PEG 8000, PAC 2, C5 <u>Diffraction:</u> none

	M: <u>Crystal type:</u> BeF3_PorX, crystal cluster <u>Condition:</u> 0.1 M HEPES pH 7.0, 15% PEG 20.000, Proplex, F6 <u>Diffraction:</u> 5-10 Å multiple lattices		N: <u>Crystal type:</u> BeF3_PorX_Se Met, needle cluster <u>Condition:</u> 0.2 M Sodium acetate pH 5.5, 0.1 M tri-sodium citrate, 10 % PEG 4000, Proplex B3 <u>Diffraction:</u> not tested		O: <u>Crystal type:</u> BeF3_PorX_Se Met, needle cluster; <u>Condition:</u> 0.2 M Sodium acetate pH 5.5, 0.1 M tri-sodium citrate. 5% PEG 4000, Proplex A12; <u>Diffraction:</u> none
	P: <u>Crystal type:</u> BeF3_PorX, needle cluster <u>Condition:</u> 0.1 M MES pH 6.0, 0.2 M di-lithium sulfate, 20% PEG 4000, TOP 96, C4 <u>Diffraction:</u> not tested		Q: <u>Crystal type:</u> Dim_PorX, needle cluster <u>Condition:</u> 0.1 M HEPES pH 7.5, 0.2 M di-lithium sulfate, 25% PEG 3350, TOP 96, D6 <u>Diffraction:</u> no diffraction		R: <u>Crystal type:</u> BeF3_PorX_Se Met, needle cluster <u>Condition:</u> 0.2 M Na/K- tartrate, 20% PEG 3350, Top 96, C4 <u>Diffraction:</u> not tested

3.2.3 Reproduction trials of initial crystals

Reproduction trials of diffracting crystals from the C12 PAC1 condition were performed in four 48 well screens (screens see Chapter 2.1.11): Grid01_PEG_3350 and Grid01_PEG_8000 each with SeMet labeled and unlabeled protein. These upscaled trials (from original 100 nl drops after 4 months growing out of precipitate) were unsuccessful. They yielded twinned crystal bars (TABLE 3.2) with insufficient data quality due to multiple overlapping lattices. The nominal resolution of those bars was between 3.5 and 6.7 Å. SeMet derivative had lower protein solubility in the same conditions than unlabeled protein and crystallized by trend at a lower pH (ca. 1 pH unit). Most conditions (above pH 7.4) yielded sea urchins or highly clustered aggregates. Further observations were that crystals diffracted were highly sensitive to radiation damage despite their relative large size.

TABLE 3.2: Best Grid01_PEG_3350 crystals. BeF3_PorX_SeMet protein, sizes from around 10x10x30 to 15x15x500 µm. Most conditions yielded only precipitate or sea urchins.

	Tris pH 7.0 6% PEG 3350 grown in 4 days diffraction: 7 Å, twinned		Tris pH 7.4 6% PEG 3350 grown in 4 days diffraction: 7 Å, twinned		Tris pH 7.4 8% PEG 3350 grown in 4 days diffraction: 3.5 Å, twinned
---	---	---	---	---	---

3.3 Crystallization buffer optimizations

In order to optimize PorX crystallization buffers, the best (highest score) 37 conditions (which generated breakable clusters or single crystals out of 500 conditions) were compared and analyzed: Most conditions (78%) contained PEG with a molecular weight higher than 3000 as precipitant, and 86% had a pH between 6.5 and 8.5. In addition, a significant amount of these drops contained magnesium, sulfate or carboxylic acids (e.g. formate, tartrate, malonic acid) in the buffer composition. An other observation was that different PorX protein sample types reached their highest scores in the same crystallization conditions, but with different maximal scores, presumably due to their relative protein stability in those buffers. Only our two different dimeric protein samples (*Dim_PorX_SeMet* and *BeF3_PorX_SeMet*) yielded crystalline aggregates (sea urchins or higher score) whereas samples bound to BeF3 reached a higher score in the same conditions. Significant differences between native and SeMet substituted protein could not be observed except lower solubility of *BeF3_PorX_SeMet* derivative. This was selected for further crystallization optimization also in order to solve the structures by SAD.

The top scored 37 conditions mentioned above were screened and combined using seven consecutive 96 well grid screens (OptA-G, see Chapter 7.10), set up with various protein buffer conditions, protein concentrations and temperatures each, the progress in crystal quality can be seen in TABLE 3.3-TABLE 3.7. Early obtained single crystals were either not diffracting well (lower 7 Å) or too small. Some small crystals diffracted between 2.8 and 3.5 Å resolution but data could only get collected below the selenium absorption wavelength, due to high radiation damage close to Se absorption edge. Larger crystals, more resistant to radiation damage grew in crystal clusters and could not be broken into single fragments without resulting superposed diffraction patterns of multiple lattices (with a few exceptions of lower quality that could neither be reproduced nor optimized). Lowering protein concentrations (down to 2.5 mg/ml) prevented crystal clustering in exchange for smaller crystals with higher radiation damage and low anomalous signal. The same effect was observed for small amounts of glycerol added, which prevented crystallization at higher concentrations. Best pH regions for PorX crystal size and shape were around 5.5, 7.0 and 8.5 with best diffracting crystals at 8.5. Best precipitants were PEG with molecular weights of 3350, 8k or a medium or high molecular weight mixture (MMW(PEG:2K, 3350, 4k, 5k MME) or HMW (20k , 10k, 8k, 6k)). Additive screenings proved sodium formate to enhance protein diffraction quality and it promoted crystallization. High concentrations (>1 M) yielded to different crystal packing with low diffraction (lower 7 Å). At same pH glycine as buffer compound yielded the most and best

diffracting crystals compared to Tris or HEPES buffer but also enhanced crystallization as 20 mM additive to different buffers. From all tested cryo conditions, including cocrystallization (Opt_C), soaking in 18% glycerol added to the crystallization buffer yielded the best results during diffraction.

TABLE 3.3: Best Opt_A crystals.

Most conditions still yielded clustered aggregates but first separable crystals could be isolated and diffracted. Single large crystals in sodium formate precipitant conditions did not diffract well.

	Tris pH 7.5 8 % PEG 8000 grown in 4 days diffraction: 3.4 Å, ca. 10 µm		Tris pH 8.5 17 % PEG 8000 grown in 4 days diffraction: not tested, size 5 µm		HEPES pH 8.0 2.25 M Na formate, grown in 4 days, diffraction: 6-20 Å, size ca. 30 µm
	HEPES pH 8.5 2.0 M sodium formate, grown in 2 days		HEPES pH 8.0 200 mM LiSO ₄ 18 % PEG3350 10 % Glycerol, 6d, 3.9 Å, 30 µm		HEPES pH 8.0 200 mM LiSO ₄ 18 % PEG3350 without 10 % Glycerol, 1d

TABLE 3.4: Best Opt_B crystals.

Crystals yielded data better than 3 Å in PEG conditions. Sodium formate precipitant crystals grew larger but resolution was unsuitable for structure determination while low amounts of sodium formate in PEG conditions enhanced crystal quality in PEG conditions. Crystals were still small and prone to radiation damage.

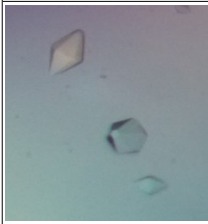
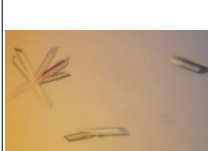
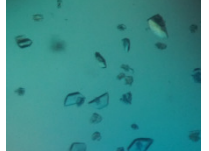
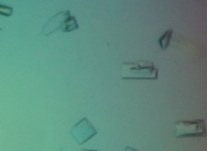
	Glycine pH 8.5 1.8 M sodium formate, grown in 2 days, diffraction: 4-7 Å, size ca. 50 µm		HEPES pH 8.25 2.0 M sodium formate, grown in 2 days, diffraction: 8 Å, size ca. 50 µm		Cacodylate pH 7.0, 8.5 % PEG 3350, grown in 1 days, degraded after 10 days, size ca 10x10x70 µm
	HEPES pH 7.0, 8.5 % PEG3350, grown in 3 days, diffraction: 3.4 Å size ca 10x10x40 µm,		Cacodylate pH 6.5, 6 % PEG 3350, grown in 1 day, diffraction: 3.7 Å size ca 40x30x20 µm,		Cacodylate pH 7.0, 6 % PEG 3350, 100 mM sodium formate grown in 3 days, diffraction 2.9 Å, ca 35x30x15 µm

TABLE 3.5: Best Opt_D crystals.

It could be observed that lower protein concentration was necessary to avoid crystal clustering, and PEG 3350 and PEG 2000 yielded larger single crystals but not as many as a MMW mixture and diffracting at lower resolution than crystals obtained with HMW PEGs.

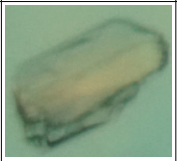
	Cacodylate pH 7.0, 7 % PEG 3350, weak diffraction, size ca. 40x10x10 µm		Glycine pH 8.5 7 % PEG 3350, weak diffraction, size ca. 40x15x15 µm		Cacodylate pH 6.5, 7 % MMW PEG, weak diffraction, ca. 100x20x20 µm
	Cacodylate pH 6.5, 7 % PEG 20000, diffraction up to 4 Å, plate size ca. 60x60x20 µm		Cacodylate pH 5.5, 7 % PEG 20000, diffraction: multiple lattices 2.9-8 Å, ca. 180x70x70 µm		Glycine pH 9.0 7 % HMW PEG, diffraction up to 2.4 Å, size ca. 130x60x60 µm

TABLE 3.6: Best Opt_E crystals. Additional glycerol increased PorX solubility. Only at glycine pH 8.5 conditions where relative unclustered crystals grew, improvements could be observed at 0.5 % glycerol concentration.

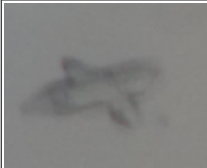
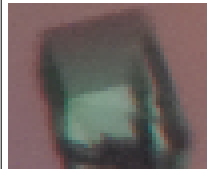
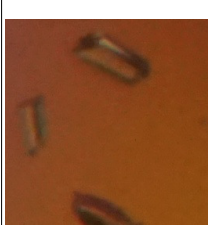
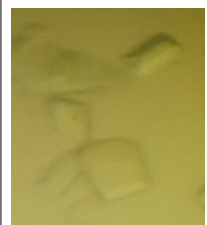
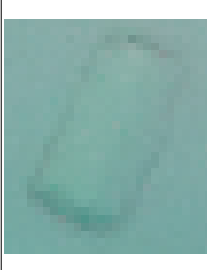
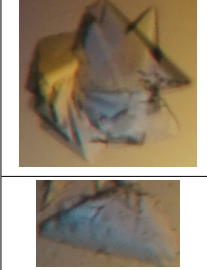
	Glycine pH 8.5 12 % PEG 3350, 0% Glycerol, size ca. 20x5x5 µm, diffraction nominal 3.3 Å		Glycine pH 8.5 12 % PEG 3350, 0.5% Glycerol, nominal 2.6 Å, 3.1 complete, 6.5 anomalous, ca. 40x15x15 µm		Glycine pH 8.5 12 % PEG 3350, 2.4% Glycerol, size ca. 20x15x5 µm
	Glycine pH 8.5 12 % PEG 3350, 4.3 % Glycerol, diffraction nominal 3 Å, size ca. 20x10x5 µm		Glycine pH 8.5 12 % PEG 8000, 0 % Glycerol, diffraction nominal 3 Å, size ca. 30x5x5 µm		Glycine pH 8.5 12 % PEG 8000, 0.5 % Glycerol, diffraction nomi- nal 2.9 Å, size ca. 40x40x40 µm

TABLE 3.7: Best Opt_F crystals. While large crystal clusters that diffract well had multiple lattices, sodium formate and glycine as additives supported the growth of many smaller single crystals instead of a large cluster within equivalent crystallization conditions. Screening of different protein concentrations proved 2.5 mg/ml as optimal concentration to avoid crystal clustering.

	Cacodylate pH 6.5, 6% PEG 8000, diffraction not tested due to size <15 µm		HEPES pH 7.0, 6% PEG 3350, bent plates size ca. 20x5x5 µm, no diffraction		HEPES pH 7.0, 6% PEG 8000, 100 mM sodium formate, size ca. 25,20x5 µm, also at pH7.5 and cacodylate pH 7.0
	HEPES pH 7.0, 6% PEG 3350, 50 mM sodium formate&20 mM glycine, 30x15x5 µm, also at pH 6.5-pH 9.0, diffraction 3.5Å nominal		Glycine pH 9.0 4 % PEG 8000, diffraction: broken fragments max. 3 Å, multiple lattices, size ca 75x75 µm		HEPES pH 7.5 4 % PEG 3350, fragment size ca. 100X50x25 µm, diffraction ca. 3.2 Å P21, 4.5 Å anomalous data

3.4 Crystallization strategy optimization

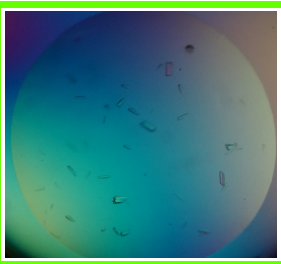
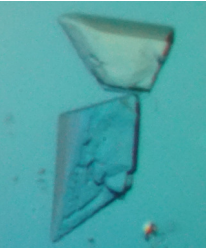
After screening crystallization buffer conditions that yielded either good diffracting larger clusters that were difficult to break or small crystals prone to radiation damage, two strategies were applied: First was to set up multiple identical plates (Opt_J, Chapter 7.10) with a series of PorX concentrations (4.0, 3.0, 2.25 and 1.6 mg/ml) at relative small grid screens in order to yield good diffracting crystals by chance (as observed before, "semi-high throughput" Chapter 3.4.1). These plates were also setup with and without sodium formate (100 mM) in order to confirm formate as an additive as observed before. The second approach was upscaling of crystallization conditions (Chapter 3.4.2) at low concentration (2.15 and 1.6 mg/ml) to have few nucleation events but longer crystal growth thus yielding larger single crystals (or at least large, better breakable clusters). In 24

or 48 well screens initial drop volumes up to 20 μ l (OptH, see Chapter 2.1.11) were used with and without additives and different salt concentrations. Further different protein to buffer drop ratios were screened, 1:5 to 5:1 or 4:0.5 (BeF3_48opt_1, -2, -3 and Opt_I). In addition also crystallization at 37 °C to slow down nucleation was screened, yielding diffracting crystals in a longer period of time but without improved crystal quality or amount.

3.4.1 Semi-high throughput, complete datasets better than 3 Å ("JG5" and "SCN" data)

The semi high throughput approach of Opt_J screens (ca. 1200 conditions) resulted in hundreds of crystals after 1-2 weeks. Ca. 150 were fished and diffracted and crystals from 4 conditions (TABLE 3.8) yielded data between 2.4 and 3 Å. Only crystals grown in the presence of sodium formate were resistant to radiation damage enough that full datasets could be collected. Therefore 100 mM sodium formate was added to all following crystallization screens (Opt_K, Opt_L). The best crystal grew in 100 mM glycine pH 8.0, 100 mM sodium formate and 6 % HMW PEG, which yielded a complete dataset at 2.5 Å and anomalous data of 3.3 Å at a multiplicity of 1.8. This condition Opt_J, G5 "JG5" (hereinafter used as name for dataset and crystal structure deriving from this data, TABLE 3.8, highlighted in green) was one of many crystals (over 30, 8 were diffracted) from this condition, all others diffracted only around 3-4 Å.

TABLE 3.8: Best Opt_J crystals. This screen between pH 7.0-9.5 and 2-8% PEG 3350, 8000, HMW or MMW as first setup rendered well diffracting crystals. Sodium formate improved radiation damage resistance enough to collect full datasets. However initial crystals were still small and radiation damage

	"JG5" condition, 2.25 mg/ml, glycine pH 8.0, 100 mM sodium formate, 6 % HMW PEG, size ca. 30x15x5 μ m, P1, 2.4 Å, 3.4 Å anomalous diffraction			4.0 mg/ml , glycine pH 8.5, 100 mM sodium formate, 8 % MMW PEG, ca. 200x100x60 μ m, P21, 2.55 Å, 4.1 Å anomalous
	4.0 mg/ml , glycine pH 8.5, 100 mM sodium formate, 8 % PEG 3350, ca. 100x80x50 μ m, P21, 2.60 Å, 4.3 Å anomalous		4.0 mg/ml , glycine pH 8.0, 100 mM sodium formate, 8 % PEG 3350, ca. 100x80x50 μ m, P21, 2.60 Å, 4.3 Å anomalous 37°C	2.25 mg/ml , cacodylate pH 7.0, 100 mM sodium formate, 8 % HMW PEG, ca. 100x20x20 μ m, P21, 2.70 Å, 4.2 Å anomalous

The multiplicity and resolution of JG5 was not high enough to solve its structure (P1, 360° data collection), neither were any single dataset of the collected 24 datasets so far. We tried either by SAD or MAD with or without merging. Merging between datasets of different crystals was difficult

due to differences of unit cell dimensions of up to 12 Å within the same space group and even 8 Å within a single crystal when it was diffracted at different positions. Thus in order to get as many similar datasets for merging as possible, 6x2x96 well plates with only this condition were set up at 5.1, 3.7, 2x2.8 and 2x2.25 mg/ml. In addition this 6x96 conditions were used as “*Optisalt*” screening, where 10% of additive solutions were added to the crystallization condition. This Optisalt-conditions yielded hundreds of crystals, the most promising 30 crystals were fished, diffracted and yielded data between 2.8-3.4 Å without a change in unit cell dimensions that would indicate different crystal packing. One exception was a crystal with additional 10% [2.4 M NaSCN] with 2.4 Å resolution in P1 space group “*SCN-crystal*” (TABLE 3.9, highlighted in green), which had six dimers within the asymmetric unit (ASU) and, as single crystal per well, appeared in both replicate conditions. From the purely reproduced conditions (double replicate of JG5 condition at exactly 2.25 mg/ml) only 5 % yielded similar single crystals as the original condition but none yielded higher resolution. Using anomalous data from this screenings (and merging all those similar datasets, as intended) for structure solution was (at this point of the project) not necessary anymore, as the parallel approach of upscaled conditions already yielded a crystal after 6 weeks that was suitable for structure solution.

TABLE 3.9: Best JG5-Optisalt-crystals. Addition of thiocyanide yielded to large single crystal of good diffraction quality but with very huge ASU difficult to model. Addition of citrate lead to single bars without the common PorX tendency to build clusters, but with slightly weaker diffraction quality.

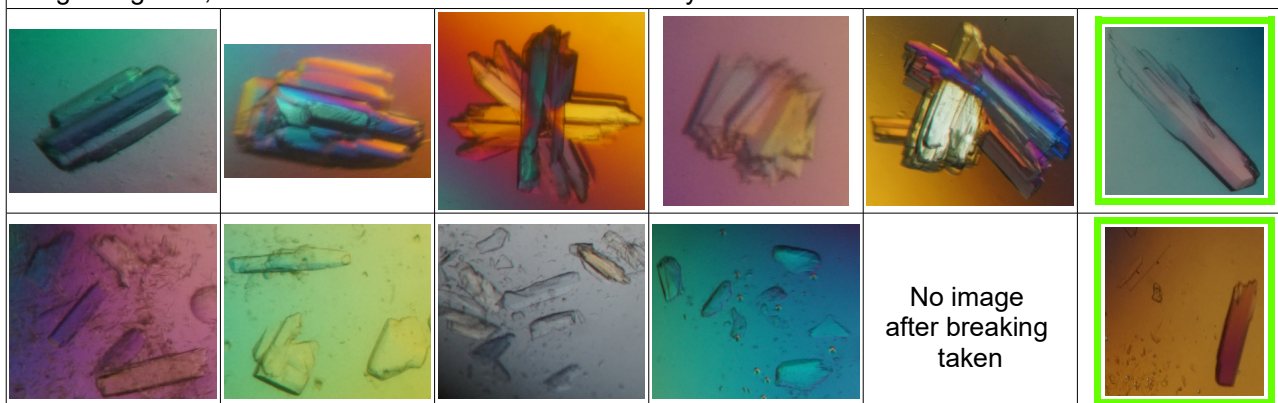
	JG5-condition+ 10% [0.1 M Tris pH 8.5 +2.4 M NaSCN] multiple lattices, broken plate cluster		"SCN-crystal" JG5-condition+ 10% [2.4 M NaSCN] 2.5 Å, 80x40x40 µm		JG5-condition+ 10% [0.6 M tri-sodium citrate+0.1 M sodium acetate 2.8-4.0 Å, many large bars
---	---	---	---	--	--

3.4.2 Upscaling, sufficient anomalous scattering ("HB6" data)

Seven upscaled screens in 24 or 48 well plates yielded several large breakable clusters (TABLE 3.10) but most fragments of those clusters still had superposed multiple lattices. When a single fragment could be obtained, there were still remaining additional fragments attached or within the cryo solution, which added noise to the collected data. One exception from this dilemma was a cluster grown in OptH_B6 (thus "*HB6*" condition, crystal or data) after 6 weeks at 10 µl drop size, 2.15 mg/ml, 100 mM HEPES pH7.5, with 50 mM additional MgCl₂, 100 mM sodium formate, 6 % PEG3350 (TABLE 3.10, highlighted in green). This cluster “HB6” of ca. 1 mm in length could be broken into a central bar of around 500x120x120 µm, which was covered on all sides by thin plates that could be separated and diffracted too. The central fragment yielded 3 different datasets

collected at multiple positions, including kappa reorientation, at at both the selenium absorption edge and inflection point (in addition to datasets not suitable for merging due to differences in unit cell dimensions). The highest resolution dataset was 2.6 Å, while an other dataset had the highest anomalous data of 3.3 Å. Separated plates from this cluster yielded data around 2.6 Å as well as 3.6 Å anomalous data as well, however with a difference in unit-cell dimensions larger than 8 Å. In terms of non-reproducibility this crystal was no exception to other 24 or 48 well derived crystals even at multiple replicate conditions, as well as screening of different concentrations and drop-to-drop ratios.

TABLE 3.10: Largest crystals (upper images) of PorX upscaling trials and their broken/diffracted fragments (lower panel). (BeF3_48Ops_01, -02, -03, Opt_H and Opt_I) Drop sizes of up to 20 µl yielded clusters up to 1 mm in size, less prone to radiation damage. Some were breakable into plates or bars but were difficult to separate and thus yielded superposed diffraction patterns. Only one cluster "HB6" (right side, highlighted in green) yielded in a bar like fragment with single crystal lattice, enveloped by multiple plates with different unit-cell dimensions. From this crystal 3 datasets could be collected that, merged together, were used to solve the PorX structure by SAD.



3.4.3 Soakings and co-crystallizations, "in situ dimerization", Opt_J, -K, -L, -M, K+/K++

As other PG proteins were unlikely to crystallize and the PorX JG5 structure model building progressed, we focused on obtaining additional PorX structures, in particular a high resolution structure, which could explain the dubious density at the catalytic site of PorX, the metal identities and correct ligand. Further our model of the JG5 data (Chapter 3.7.2) suggested that it was likely that PorX binds to nucleotides thus our collaborators performed *Prometheus* (Nanothempertech) high thermic stability characterization indicating interactions with GMP and GTP (on trend, Mariusz Madej, unpublished data). Looking at the PorX large ligand binding cavity and comparing it to structural homologs from the Protein Data Bank (PDB) (TABLE 7.4, Chapter 3.14) dinucleotides were likely to bind and were screened for co-crystallization too.

One additional optimization was to enhance resolution and solve the problem of crystal clustering by adding zinc to the crystallization conditions as our structure suggested that zinc is bound to PorX

active site. The idea was that a low occupancy of zinc might lead to mixed conformations hampering the resolution of crystals, or different batches of PorX molecules with different occupancies/conformations might successively add to a crystal and thus cause the clustering behavior we observed.

In a first attempt the screen with the highest resolution datasets, Opt_J was used at 3.30 and 2.25 mg/ml as control (and resulting crystals for soaking experiments) and in addition with 1 mM zinc chloride added to the protein containing buffer. Immediately after adding the zinc to the concentrated protein solutions (3.30 and 2.25 mg/ml) the samples got turbulent by precipitation and after precipitate removal by centrifugation PorX concentrations were at 0.77 and 0.41 mg/ml (ca. 33 % of initial concentration). Surprisingly these low concentrations still yielded a multitude in crystallization events, as single crystals but also small and prone to radiation damage.

To avoid immediate precipitation in the next trial the concentrations were lowered to 2.0 and 1.3 mg/ml and only 0.1 mM zinc chloride was added, this time to the crystallization buffer condition, which, while adding, dilutes the initial protein concentration even further. Here no visible precipitation was observed but the low amounts of zinc seemed to decrease clustering. Due to the low concentrations still this attempt did not yield better data.

We also tried to improve crystallographic data by macroseeding, transferring crystals from similar conditions and let them grow with fresh protein, with and without extra zinc in the cryo condition, but improvements in diffraction could not be observed.

Soaking with GMP or GTP γ S (a noncleaveable GTP analog) did not yield crystallographic data with ligands bound to PorX. CocrySTALLizations with GMP or GTP γ S ("Opt_K,+" or "Opt_K, ++") yielded for GMP three datasets up to 2.9 Å and 5 up to 2.8 Å for GTP γ S and also did not show stable binding of ligands but a different conformation and space group (Chapter 3.10.3&3.10.4). As these crystals were still strongly clustered together, two 48 well plates (BeF3_48opt_04 and -05) were used to optimize these conditions by a variation of drop size, higher protein or zinc concentrations and macro- or microseeding but without success.

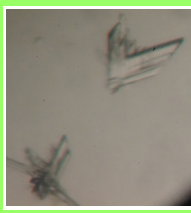
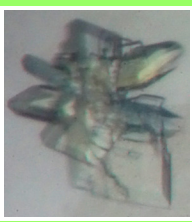
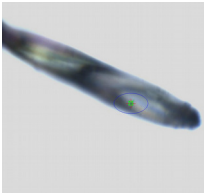
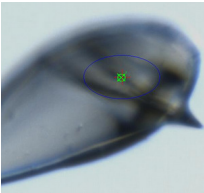
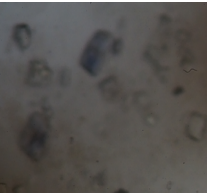
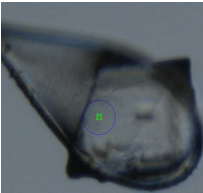
Crystals derived from screens Opt_J or Opt_K were also tried to be reproduced in a 24 well setup at various drop sizes, both in hanging and sitting drop setup but without success (screens Opt_M).

From concentration trials of PorX we knew that PorX monomer is highly soluble (at least up to 25 mg/ml without precipitation), in contrast to the dimer (precipitation above 10 mg/ml). We used this new information in order to add more protein into to the crystallization setup and induce dimerization inside the crystallization drop, a method we called "*in situ dimerization*". For this

purpose, dimerization additives (MgCl_2 , NaF, BeSO_4) were added into the crystallization buffers and crystallization was set up with high concentrated monomeric *PorX_SeMet*. Since we previously observed an improvement in crystal shape by addition of zinc, 0.1 mM zinc chloride was also added into the crystallization buffers. The in situ dimerization trials were the break-through in yielding large crystals of PorX at high concentrations (12-14 mg/ml). Instead of small crystals growing within days to about a week from clear conditions, in the *in situ* setup a substantial amount of PorX precipitated upon mixing with the crystallization condition. Out of this precipitate containing condition only a few large crystals or breakable clusters grew while feeding on the surrounding precipitate until a clear condition, however the process of crystal growths took between one and two month. These crystals diffracted up to 1.9 Å "HR"-crystal (structure Chapter 3.8) but took about 2 months to grow, while at pH 9.0 and 9.5 already large clusters of bars grown in two weeks that were easy to separate yet diffracted worse than 5 Å (space group P4).

The best diffracting crystal/fragment of a cluster "HR" significantly resisted radiation damage during 7 dataset collections (at different absorption edges (Zn, Se) and at different positions and crystal orientations) yielding data between 1.9-2.4 Å and, even one month later, during a second diffraction session, two datasets could be collected at 2.6 Å (Zn, Mn absorption edges).

TABLE 3.11: Crystals of dataset from co-crystallization and soaking experiments. Screens: Opt_J, -K, -L, -M, K+/K++, solved structures: "GMP", GTPyS, "HR". Images represent the crystallized clusters (upper left), broken clusters (lower left) and diffracted fragments (lower right).

GMP-cocrystallization crystal	GTPyS-cocrystallization crystal	HR-crystal (GMP soaked)
 Cacodylate pH 7.0, 9% HMW PEG, 100 mM sodium formate, 0.1 mM zinc chloride, 220x15x15 µm, 2.9 Å, P21	 Cacodylate pH 7.0, 8% HMW PEG, 100 mM sodium formate, 0.1 mM zinc chloride, 100x25x25 µm, 2.8 Å, P21	 Glycine pH 8.0, 4 %PEG 8000, 100 mM sodium formate, 0.1 mM ZnCl_2, 15 mM MgCl_2, 30 mM NaF, 5 mM BeSO_4, 100x75x40 µm, 1.9 Å, P1
 	 	 

From the condition of crystal "HR", with highest resolution data, a new screen "*Opt_L+*" was designed with multiple replicate conditions and only few condition variation. As a last optimization this screen was used without zinc addition and with 0.5 mM zinc chloride at protein concentrations 4, 7, 10 and 13 mg/ml. Upon those 400 conditions a clear improving trend in crystal shape when zinc was added was observed and the best crystals (in size and shape) were growing from the

highest protein concentrations. The resulting crystals were used for further soaking trials with different nucleotides but no ligand could be observed within the PorX active site in their corresponding solved crystal structures. Estimating the quality of crystals from this screen, when diffracted without previous soaking, they could have been used to collect data higher than 1.9 Å resolution.

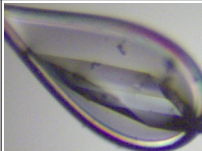
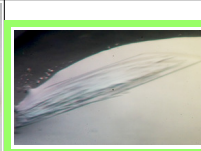
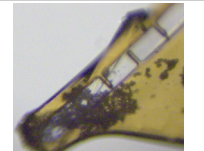
3.5 PorX mutants crystallization

In order to collect structural data for PorX mutants screens Opt_J+ as broad- grid and Opt_L+ as fine-grid screens were used for all mutants. Tested mutants included D58A in which the RD phosphorylation site of PorX was mutated to alanine, M94K/D104A/I129A in which the RD dimerization interface was mutated to prevent dimerization, and T272A. which had the PglZ's active site threonine mutated. This threonine supposedly binds the ligand's phosphate and should co-crystallize with native ligands without their cleavage. Thus T272A was set up with potential ligands pGpG and c-di-GMP at 0.5 mM concentration in both screens, while crystals of the plates of T272A grown without ligands were used for soaking with ATP or GTP.

Adding pGpG to the crystallization setups visually improved crystal size and shape that single crystal bars got more common than clustered plates, bars and needles. From these bars 24 were fished and almost all of them yielded structural data better then 3 Å with the best data of 2.4 Å (hereinafter “pGpG” for data or structure).

Cocrystallization with c-di-GMP did not improved crystallization behavior and from the best broken cluster-fragments data of 2.9 Å could be acquired from two twinned lattices, that needed to be processed separately and were merged afterwards.

TABLE 3.12: Best crystals of PorX mutants with or without ligands. Data aquired: “pGpG” with pGpG at the active side, “c-di-GMP” without bound ligand and “D58A” which was proteolytic digested and only yieled a structure of the PGLZ domain with parts of the HBD but without RD.

pGpG	c-di-GMP		D58A	
100 mM sodium formate, 0.5 mM ZnCl ₂ , 15 mM MgCl ₂ , 30 mM NaF, 5 mM BeSO ₄				
Glycine pH 8.5, 3 % PEG 8000 200x50x50 μm, 2.4 Å, P1 21 1	Glycine pH 9.0, 2 % PEG 3350, 200x50x10 μm, 2.9 Å, P21 21 21		Glycine pH 9.5, 6 % MMW PEG 60x20x10 μm, 2.4 Å, P31 2 1	
				

3.6 Crystallization trials of PorY, PG1660 and monomeric PorX

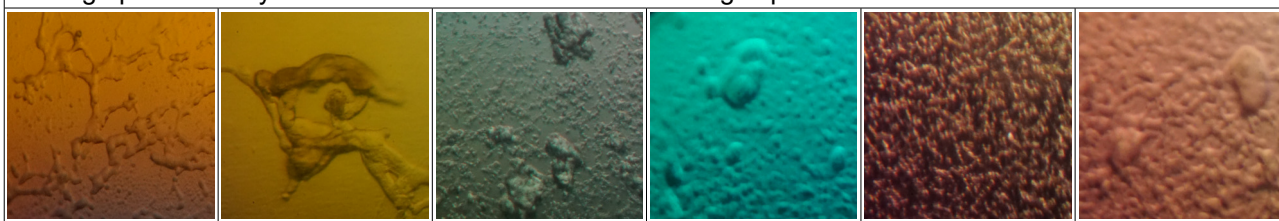
Although unsuccessful in yielding additional structures in the end, it is worth to mention our work included screenings for crystallization conditions of PorY, the sensor histidin kinase of PorX, PG1660, a sigma factor expressed at oxidative stress that might interact with PorX (Mariusz Madej, personal communication), and screenings for monomeric PorX.

Crystallization trials of the PorY cytosolic domain were performed in five 96-well sparse matrix screens, each with and without ADP addition, which seem to increase PorY solubility. Furthermore, additional 96 variations of promising conditions at 4 different protein concentrations were set up. In addition, two different protein concentrations each with ADP or AMPNP (a non cleaveable ATP analogue) as additive have been screened. However, these screens did not result in PorY crystallization.

Crystallization trails of PG1660 were set up in 500 sparse matrix conditions, without yielding promising conditions to optimize and mostly phase separation events.

Monomeric PorX was screened at ca. 1700 additional conditions (Chapter 2.1.11) but without crystallization hits. The screenings showed that monomeric PorX was highly soluble and tended to build gelatinous phase separations at non denaturing conditions rather than crystals (TABLE 3.13). Concentrations have been screened up to 14.8 mg/ml and could have been tried even further as concentration trials (performed later) showed that monomeric PorX-SeMet was even soluble at 25 mg/ml without precipitation (at least for a few minutes tested), while precipitation occurred for BeF3_PorX-SeMet already at 10 mg/ml.

TABLE 3.13: PorX monomer crystallization. Representative examples of gelatinous aggregates forming up in PorX crystallization due to oversaturation during vapor diffusion.



3.7 Structural results

3.7.1 Structural solution of "HB6" crystal

The first PorX structure could be solved by SAD with the help of Prof. Dr. Isabel Usón using SHELX_C/D/E and our 12 best datasets we collected so far. Seven of them deriving from JG5 condition (different crystals) and 5 from HB6. The program clustered the datasets and scaled and merged three similar datasets from HB (2.6, 3.1 and 3.1 Å) collected at two different positions in the crystal with

kappa reorientation, two at selenium absorption edge, one at the inflection point. Upon determination of the crystallographic phase, density modification and automatic model building followed. The best SHELX solution found 66 of 80 selenium positions in the ASU (containing two PorX dimers with 2,092 residues) and build 1187 alanines in 92 chains after 8 global autotracing cycles with 50 cycles of density modification. From then on, the structure was built by cycles of manual model building using COOT alternated with automatic refinement until the structure was mostly traced (except some missing loops and ligands) with an free R-factor (Rfree) of 26%. As further model building did not improved the Rfree and the merged data was of lesser resolution than that obtained from JG5, the incomplete HB6 model was used to solve and further build the structure of JG5.

3.7.2 Full length PorX "JG5" structure

When solving the phase problem for PorX from the HB6 data, which derived from a merged dataset at 2.6 Å resolution, the dataset JG5 was the highest resolution data we obtained so far (P1, SeMet-BeF₃, 2.4 Å, measured at the selenium absorption edge). It was solved with PHASER MR-SAD using individual PorX chains of the HB6-model as molecular replacement partial structure in order to determine the positions of the anomalous scatterers. This was followed by density modification, automatic chain tracing and manual model building alternated with automatic refinement cycles

until the Rfree dropped below 22%, all electron density was filled and geometry and bond distances were reasonable. The ASU of JG5 crystal contains two dimers of PorX (chains AB and CD, 2092 residues in total) each dimer in identical conformation but two different conformations within each dimer (RMSD $\Delta A/C=0.409$ Å, $\Delta B/D=0.455$ Å, $\Delta A/B=1.32$ Å, $\Delta C/D=1.30$ Å). Analysis of protein interface, surfaces and assemblies by ccp4_PISA_qt classified dimers AB and CD as biological units, each with around 17% interface and 87% surface residues and 3000 Å² interface area. The highest free Gibbs energy of dissociation ΔG_0 into monomers for molecule AB was reported 13.8 kcal/mol (while only 10.9 for the hypothetical

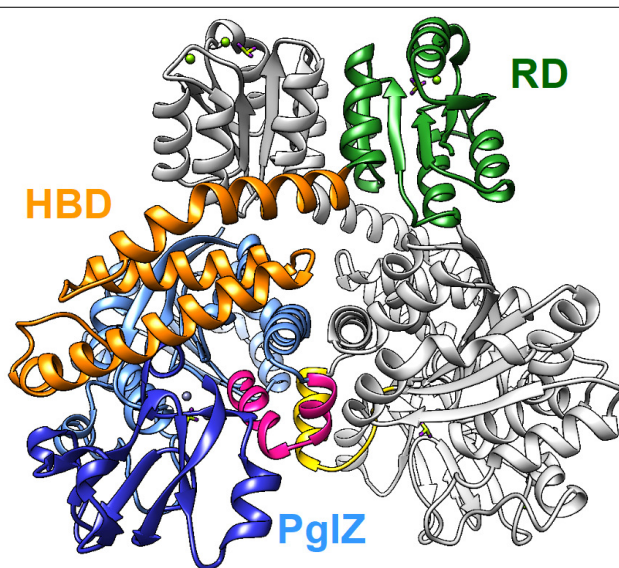


Figure 3.1: PorX dimer overall structure.

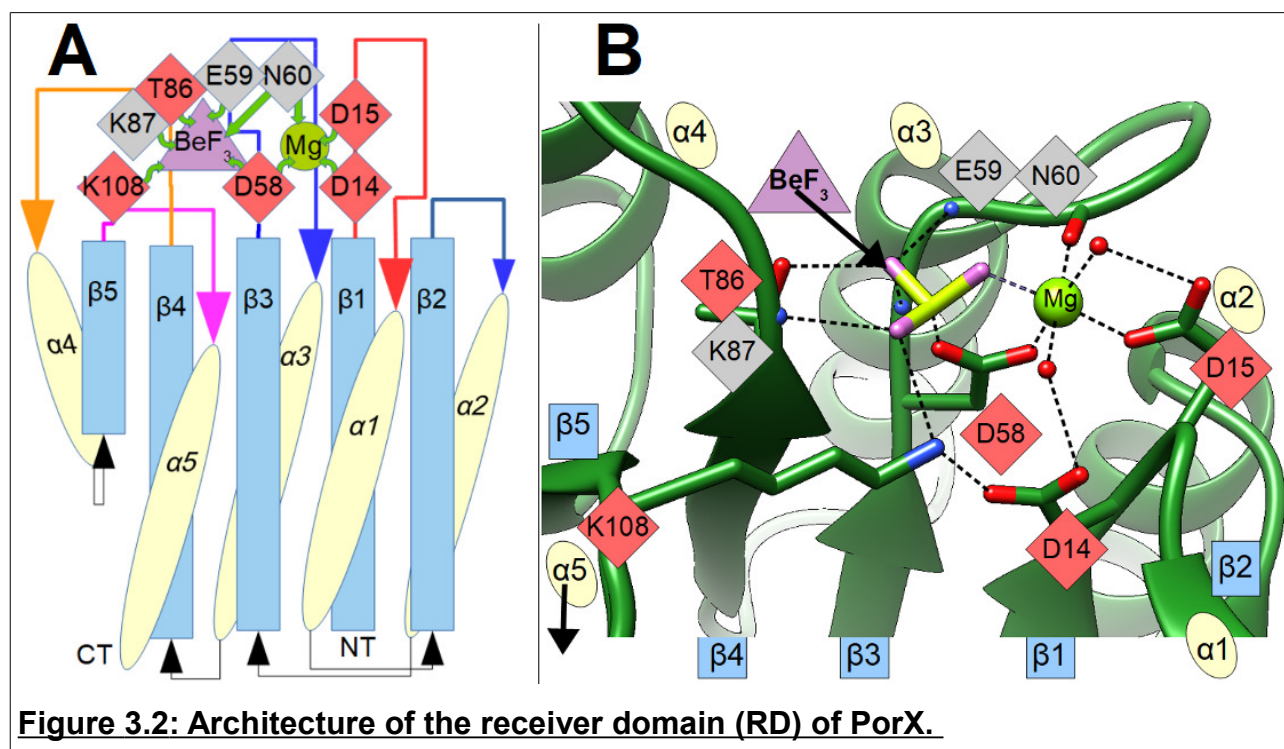
The N-terminal receiver domain (green) is connected over a three helix coiled-coil (orange) to the PglZ domain shown (blue). The PglZ's catalytic subdomain is shown in light, the CAP subdomain in dark blue. The other monomer within the dimer is shown in gray. The different conformations helix-helix (HH, magenta) and loop-helix (LH, yellow) are colored separately at the PglZ dimerization interface.

assembly of ABCD tetramer), its binding energy (AB) -43.4 kcal/mol and solvation energy gain of -32.4 kcal/mol. PorX (independent of chain) can be subdivided in an N-terminal receiver domain (RD) aa5-124, a helical bundle domain (HBD) aa125-213 and the PglZ domain (aa214-518)(Figure 3.1). The PglZ domain can be further subdivided into 1) a catalytic domain (hereinafter named "CAT") aa214-418 plus aa509-518 arranged around a central mixed β -sheet of 6 strands containing the active side metals and 2) an inserted subdomain aa419-508, which caps on side of the catalytic center (hereinafter named "CAP").

As JG5 is the most complete PorX model without substrate, structural features will be described exemplary for this structure in the following chapters. Statistics on data, model and refinement can be found in the supplementary at Chapter 7.4, domain borders and secondary structure subdivisions used hereinafter at Chapter 7.5 .

3.7.3 PorX receiver domain (RD)

The RD of PorX consist of a $\alpha_5\beta_5$ Rossmann-like fold with a central β -sheet surrounded by α -helices, helix 1 and 5 on one side, 2-4 on the other (Figure 3.2 A). The loops $L\alpha1\beta1$, $L\alpha3\beta3$ and $L\alpha4\beta4$ build an active site cavity on the C-terminal end of the sheet with D58 O $^{\delta1}$ being bound to phosphate mimic BeF₃ which in turn is stabilized by T86 O $^{\gamma1}$, K87 N, K108 N $^{\zeta}$ and E59 nitrogen and a Mg²⁺ further octahedral coordinated by D58 O $^{\delta2}$, N60 O, D15 O $^{\delta1}$ and two active site water molecules stabilized by D14 or D15, respectively (Figure 3.2 B).



α -helices are presented as pale yellow ellipses and β -strands as blue rectangles. Amino acids interacting with their side chains to the active site are indicated in red diamond shape, those with backbone interactions in gray, BeF_3 as magenta triangle. **(A)** Schematic representation of the RD's $\alpha 5\beta 5$ Rossmann-like fold of alternating β -strands and α -helices that build a central β -sheet surrounded by helices. The active site/phospho acceptor domain sits on top and is bound with beryllium fluoride which mimics phosphorylation and Mg^{2+} . Loops are represented as arrows. **(B)** Top view of PorX phosphorylation site: D58 is bound to BeF_3 which is coordinated by T86 O^{v1} and K108 N^{c} , as well as K87, E59 and N60's backbone nitrogens and one Mg^{2+} . This is further coordinated by N60's backbone oxygen and D15 and D58 O^{d1} , while two crystalline waters complete its octahedral coordination shell. Thereby one water molecule is stabilized by the O^{d2} of D15 side chain, while the second is stabilized by a water bridge to D14 O^{d1} , while O^{d2} stabilizes K108 N^{c} . Polar contacts are represented as dashed lines.

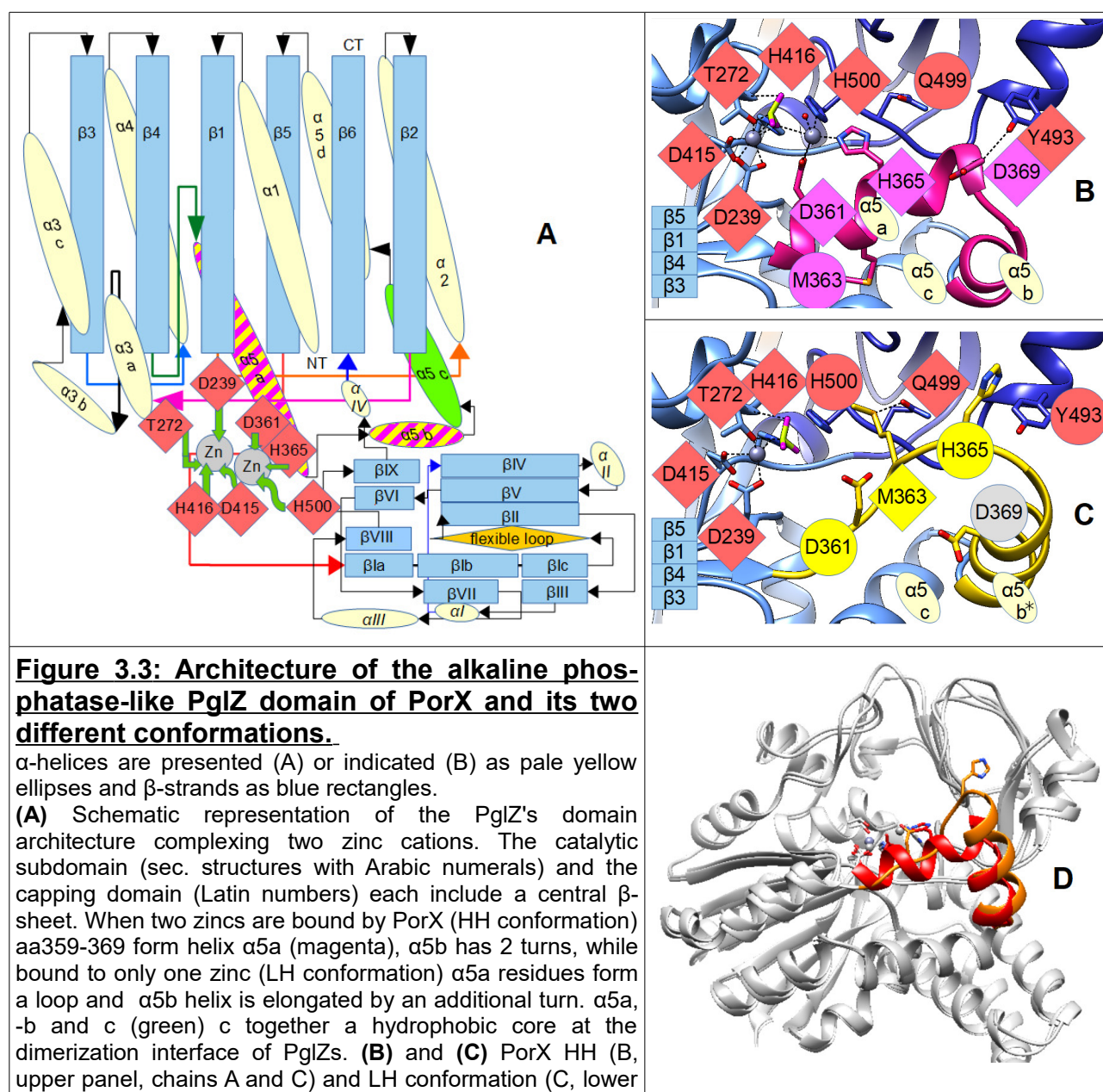
3.7.4 Helical bundle domain (HBD)

The C-terminal PglZ domain is attached to the RD via a helical bundle domain (HBD) of 81 residues (aa125-206). It starts at the end of RD $\alpha 5$, where the helical conformation is unwound, producing a bend of ca. 80° at aa125 and this bend encloses helix $\alpha 4$ of the other monomers RD (Figure 3.5). At this bend the HBD starts and includes three helices connected by short loops. The helices share a hydrophobic core and are linked to the PglZ domain by the elongated loop aa207-214. Helix 1 (aa125-148) contacts the RD of the other molecule via a network of water molecules and has a kink at F141 which makes contact into a hydrophobic core. This core is build up by helix 2 (aa154-172) and helix 3 (aa176-206) which sit on top of the PglZ domain of its own chain making multiple contacts.

3.7.5 PglZ domain

The adjacent PglZ domain consists of two globular subdomains, the larger catalytic site (CAT) and a smaller that caps the active site (CAP). The catalytic site is built around a central β -sheet of 6 strands surrounded by ten helices. The CAP domain consists of 9 β -strands interlinked in 5 β -sheets and one larger helix (αIII) with high tyrosine content (Figure 3.3A). The catalytic site can take on two alternative conformations either harboring one or two zinc ions (Figure 3.3B/C). This differences in conformation are found between both monomers within a dimer and at residues aa359-370. They include part of the zinc binding sites and of the dimerization site between PglZs and interact with the CAP domains. In chain A residues aa359-369 form an α -helix of three turns ($\alpha 5a$) coordinating ZN1 with D361 and H365. Further D369 at the end of the helix interacts with Y493 of the CAP domain, followed by an α -helix aa371-376 ($\alpha 5b$) of 2 turns. Since there are two consecutive helices, $\alpha 5a$ and $\alpha 5b$, this PorX conformation is termed “HH” conformation hereinafter. In contrast chain B, hereinafter “LH” conformation, residues aa359-366 form an elongated loop,

where D361 and H365 side chains points towards the solvent, whereas their position in HH conformation is taken by M363. Thus residues aa359-366 avoid Zn1 binding and instead interact with Q499 of the CAP domain (Figure 3.3B/C). Following this loop, aa367-376 form helix $\alpha 5b^*$ which, compared to $\alpha 5b$ (in HH conformation, chain A), is extended by an additional turn. In $\alpha 5b^*$, D369, the residue that in chain A binds to its own CAP domain at Y493, binds to S370 of the other monomer. Through this connection both $\alpha 5b$ and $\alpha 5b^*$ helices are connected and stabilize each other. Following $\alpha 5b$ both chains are followed by $\alpha 5c$ aa380-393 which is part of the PglZ dimerization interface and builds a hydrophobic core together with $\alpha 5a$ and $\alpha 5b$ (for further details see Figure 3.7).



panel, chains B and D). Residues forming the catalytic site are indicated in rectangles, or if they lost their interaction in circles. Red colored residues do not change their positions during conformational change, while magenta colored residues are the specific positions for HH and yellow colored for LH conformation. In HH conformation two zincs are coordinated and $\alpha 5a$ interacts with the CAP domain via D369/Y493. In LH conformation only one zinc (Zn2) is coordinated, whereas at the other zinc (Zn1) position, M363 is stabilized by Q499. In both conformations BeF_3 is bound by the active site to T272, mimicking a transition state intermediate during catalytic turnover with a phospho mono- or diester. **(D)** Superposition of Chains A and B PglZ domains, HH conformation of chain A is colored red, LH of chain B in orange. The area of alternative conformation is located at the center of the PglZ dimerization interface.

3.7.6 Substrate binding pocket

The catalytic site is buried within a substrate binding pocket shared and bordered by five regions (Figure 3.4): The CAP helix aa487-494 with four tyrosines facing the active site, the CAP domain flexible loop aa430-438, a flexible loop of the catalytic subdomain aa292-307, the solvent exposed Y332 and the area of PglZ that can take two different conformations ($\alpha 5a$ and $\alpha 5b$). Within the dimer the two active sites of each monomer are relative distant with distances between zincs of 32-38 Å (Chapter 7.6, Figure 7.2).

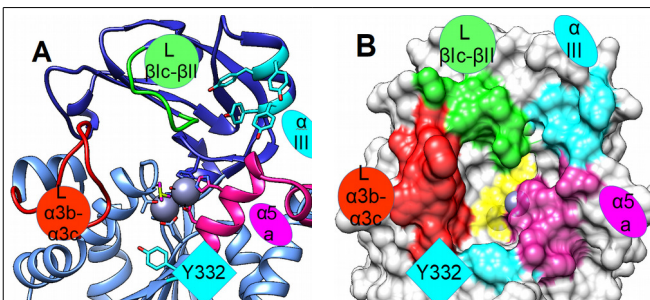


Figure 3.4: PorX-binding pocket (HH conformation). Secondary structure (A) and surface representation (B) of PorX substrate binding pocket. Flanking regions are represented in different colors, yellow was used for zinc coordinating residues in (B). The catalytic site of PorX is flanked by two flexible loops L β lc- β II and L α 3b- α 3c (in green and red respectively), Y332 and helix α III of the CAP domain (both in cyan), and helix α 5a (in magenta).

3.7.7 Dimerization interfaces

The interaction area of PorX dimer can be subdivided into two major areas. First: the RDs face each other and each RD interlaces with the HBD of the other monomer; and second: the face-to-face dimerization interface (DI) of the PglZs. Within the dimer, one RD is rotated against the other by 180° in y- and 120° in x-direction, the HBDs an additional 10° against each other around z (Figure 3.5). The RDs and HBDs from different subunits contact each other by four surface patches: At the center of interaction $\beta 5$ -D104 of both chains are connected by water bridges, surrounded by 3 larger interacting areas (Figure 3.6) a) HBD- $\alpha 1+2$ (I128, I129, T132, T136, R163 and E170) interacts with RD- $\alpha 3$ (Q71, E75), $\alpha 4$ (I98, G99) and K101; b) RD- $\alpha 4$ (E91 & T95) and $\beta 5$ -Y105 interact with $\alpha 1$ (K30) and $\alpha 5$ (L117 & K120); and c) $\beta 5$ -I107 and E89 interact with $\alpha 5$ (N111, N113, Q114) of the other chain's RD. In addition RD- $\alpha 4$ M94 connects the hydrophobic interactions of patches b) and c). The RDs have no strong interactions with the PglZ of the other monomer, with the exception of a small hydrophobic patch between RD:Y79 and PglZ:S254, F259 and K250 and the hydrogen bond

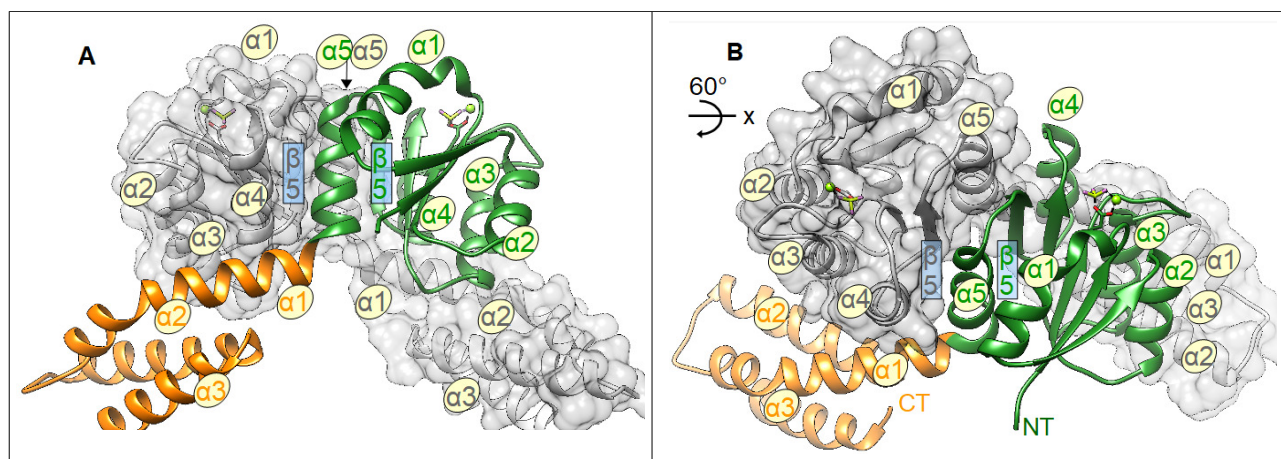


Figure 3.5: Relative position of PorX receiver domains (RD) within the PorX dimer. Similar as published [0]

RDs are related by a two fold axis with $\alpha 5$ and $\beta 5$ at the center. For relative superposition of both domains a rotation of 180° around y- and 120° around the x-axis is required. The HBDs are shifted an additional 10° around the Z-axis.

between A:E51 and B:T258 side chains between B:Y79 and A:259. The HBDs interact exclusively with their own PglZ. The mentioned face to face interaction of the PglZ domains is build up by $\alpha 5a$ (HH) or its corresponding loop L $\beta 5'$ - $\alpha 5b'$ in the LH chain and helices $\alpha 5b$ (elongated as " $\alpha 5b^*$ " in LH) and $\alpha 5c$ (Figure 3.7). The N-termini of the LH loop and HH $\alpha 5a$ are hold in place by hydrophobic contact between F359 with W390 of their own chain. These two residues are part of the central hydrophobic core together with residues from $\alpha 5b/\alpha 5b^*$ and $\alpha 5c$ (I373, M372, L376 and L386). In LH conformation the specific loop further contributes with V360, while in HH $\alpha 5a$ adds M363 and A366 to the hydrophobic core. The N-terminus of $\alpha 5bs$ is stabilized by hydrogen bonds between

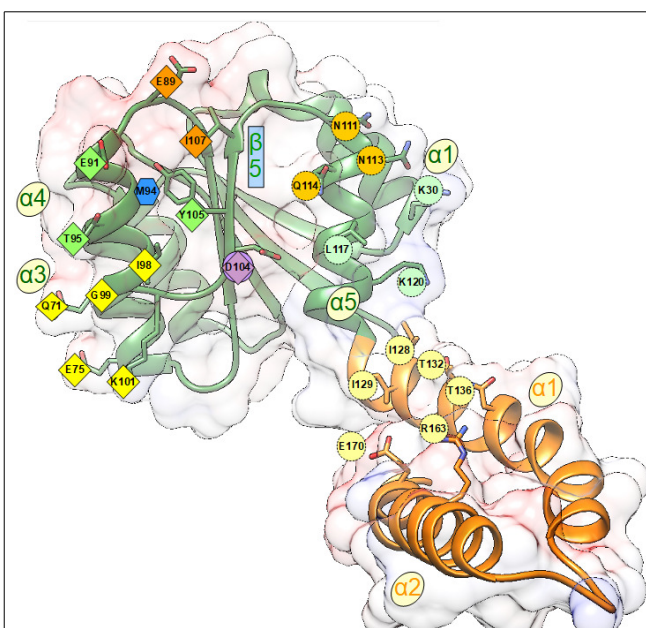


Figure 3.6: PorX interaction areas of receiver domains (RD) and helical bundle domains (HBD).

Shown are the RDs (green) and HBDs (orange) of one monomer within the dimer of PorX. Interacting residues are colored one color each (yellow, green and orange) in bright color for one monomer with the residues in light-color on the other. D104 (purple) sits at the center of the dimerization surface connecting both RD by water bridges and M94 (blue) sits at the center of a hydrophobic core linking the patches colored orange and green.

LH:HH T368:E369 and E369:S370, while their C-terminus is linked by E375 with T334. Helices $\alpha 3c$ participate in the hydrophobic core with L368 each but build an independent interaction area formed by A381, S395, S398 and H393 between both helices. While two HH/HH homodimers are sterically possible (as also shown in later structures) a double LH conformation of a PorX dimer would result in clashes between the first turns of $\alpha 5b^*$ helices at the residues of T368 with the opposites chain D369 and vice versa (Figure 3.7D).

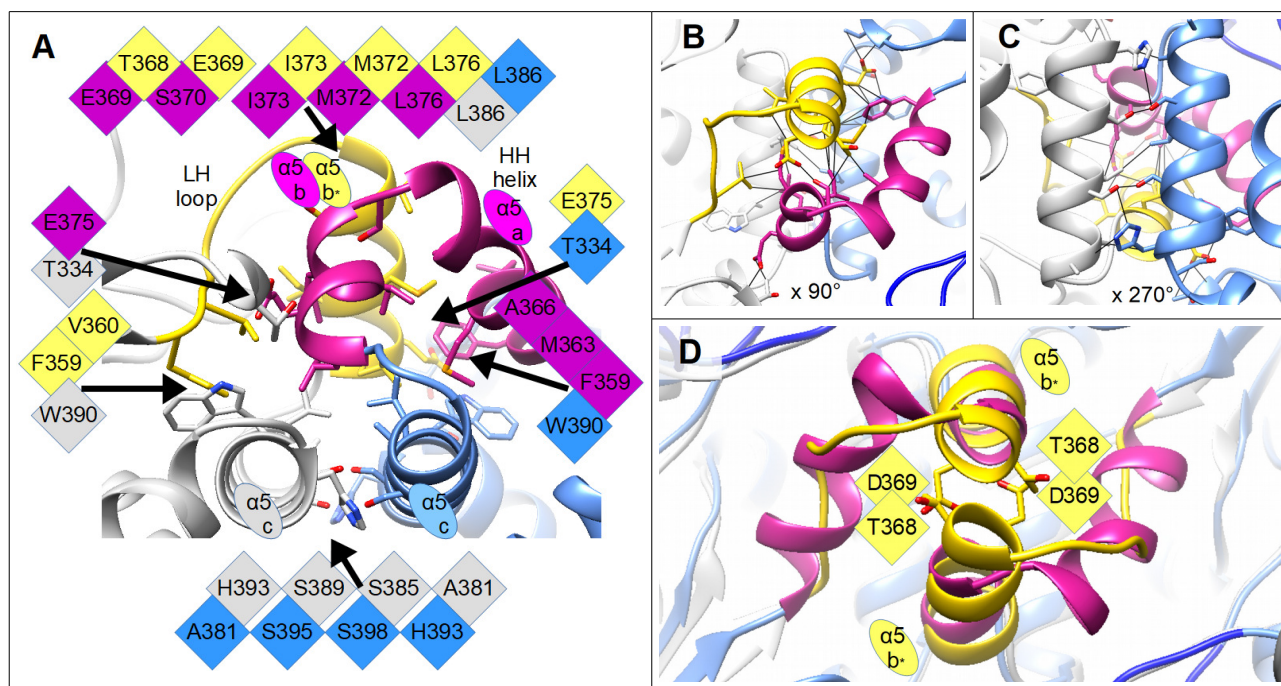


Figure 3.7: PorX interaction areas of PglZ domains in the PorX LH:HH dimer.

Chain A PglZ is colored blue, chain B gray, while B:LH is colored yellow and A:HH purple. The same colors are used for interacting residues and indication of helices. **(A)** Residues interacting with each other and stabilizing LH loop or HH $\alpha 5a$ are shown. The hydrophilic stabilization between $\alpha 5b$ (left) and the residues participating in the hydrophobic core (right) are shown on the top. On the bottom, residues of $\alpha 5c$ interaction surface are shown. **(B)** Representation of (A) turned 90° around the x-axis, interacting residues are indicated by black lines between them. **(C)** Representation of (A) turned 270° around the x-axis. **(D)** Superposition of molecules A and B to each others. There would be no clashes between two PorX dimers in double HH conformations, while a double LH conformation would result in clashes between helices $\alpha 5b^*$ s at the indicated residues E369 on one chain with T368 on the other.

3.8 High resolution structure "HR"

The HR structure (for "high resolution", 1.9 Å data, model Rfree=19.2) was obtained during progressive crystallization improvements (Chapter 3.4.3). In contrast to JG5 (crystallized without metal ion addition) the HR crystallization condition included 0.1 mM $ZnCl_2$ with the intention to reach higher occupancies of Zn and thus higher resolution. The crystal was soaked with 0.5 mM GMP but no nucleotide was found in the crystal structure. GMP was used to stabilize and presumably get a higher resolution structure due to preliminary thermoflour results (data not shown)

that indicated PorX stabilization by either GMP or GTP. Consequently, GMP, GTP and GTP γ S soakings or co-crystallization were performed in parallel. The overall structure, space group, unit cell dimensions and interactions of HR are similar to that of structure JG5. In addition due to higher resolution, this structure shows more details which were necessary to identify e.g. the unexpected BeF₃ at the PglZ site and clarified the dubious electron density at L β 1c- β 2 resulting from two alternative conformations. However alternative conformations at loop L α 3b- α 3c that were observed to a much lesser extend in JG5 data made this region difficult to trace in all chains (esp. LH conformations) resulting in a more incomplete model compared to JG5 (missing B:aa298-304 and D:aa298-307). Statistics on data, model and refinement can be found at Chapter 7.4 , domain borders and secondary structure are equal to JG5 structure (Chapter 7.5).

The crystallographic data used for refinement consists of a merged dataset of five data collections from a single crystal fragment, including different positions and two kappa reorientations (0°, 30°, 77°) out of seven datasets taken at Zn absorption edge. Additional datasets from this crystal were taken at Se and Mn absorption edges. Furthermore, data was taken from a second fragment of its cluster which we soaked with additional 0.5 mM MnCl₂ and diffracted Zn and Mn absorption edges to evaluate relative metal binding affinity between Zn and Mn.

The electron density of the refined model was good enough to confirm BeF₃ binding not only at RDs, as expected, but also at PglZ HH and LH conformations (Figure 3.8).

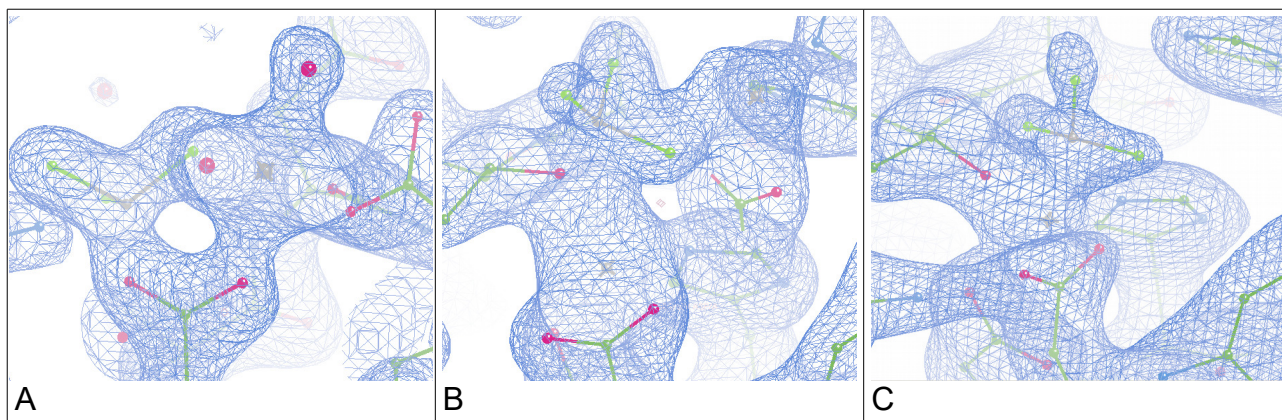
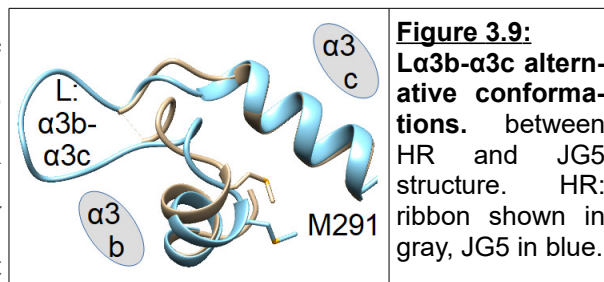


Figure 3.8: PorX phosphorylation sites bound to BeF₃.

2Fc-Fo electron density maps are shown in blue, PorX in ball and stick mode. The sigma levels were adjusted for best ligand electron density visibility. **(A)** BeF₃ bound to D58 O δ^1 (1.41 Å) of the RD, while D58 O δ^2 binds octahedrally coordinated Mg²⁺. **(B)** BeF₃ is covalently to T272 (distances 1.41 Å between Be and T272 O γ^1) in HH conformation while F² is stabilized between both Zn²⁺ and D239 and D361 oxygens and F¹ is bound to T272 backbone nitrogen. **(C)** in contrast to HH conformation in LH BeF₃ only forms an ionic interaction with T272 (2.35 Å) while F² is stabilized just by D239 and F¹ by T272 backbone nitrogen as in HH conformation.

The electron density of overlapping alternative conformations in HR structure at LH-PglZs $\alpha 3b$ and $\alpha 3c$ (Figure 3.9) could only be validated precariously due to very weak density and by anomalous signals for M291 Se from the dataset



taken at Se absorption edge (2.2 Å resolution). Both alternative positions of M291-selenium at $\alpha 3b$ at chains B and D (LH conformations) showed comparable low signals. Due to the already high b-factors and thus weak electron density in that region the alternative conformations (identical to that of JG5) were not built in the HR structure, even for helix $\alpha 3b$, where some side chains were visible (F292), because they did not increase the R_{free} after refinement. Anomalous data taken at the Zn absorption edge confirmed Zn at the PglZ active sites, both in LH (Figure 3.10A, left) or HH (Figure 3.10B, left) conformation, but not at the RD (Figure 3.10C). However this crystal was grown at 0.1 mM $ZnCl_2$, thus we could not exclude that Mn was replaced due to higher concentration of Zn. Data taken from another fragment of the HR crystal cluster, soaked with 5 times excess of Mn over Zn showed anomalous signal at Zn absorption edge (PglZ active site) but no anomalous signal at any metal positions at Mn absorption edge (Figure 3.10A, B). The RD active site did not show any anomalous signal, neither at Zn or Mn absorption edges (Figure 3.10C).

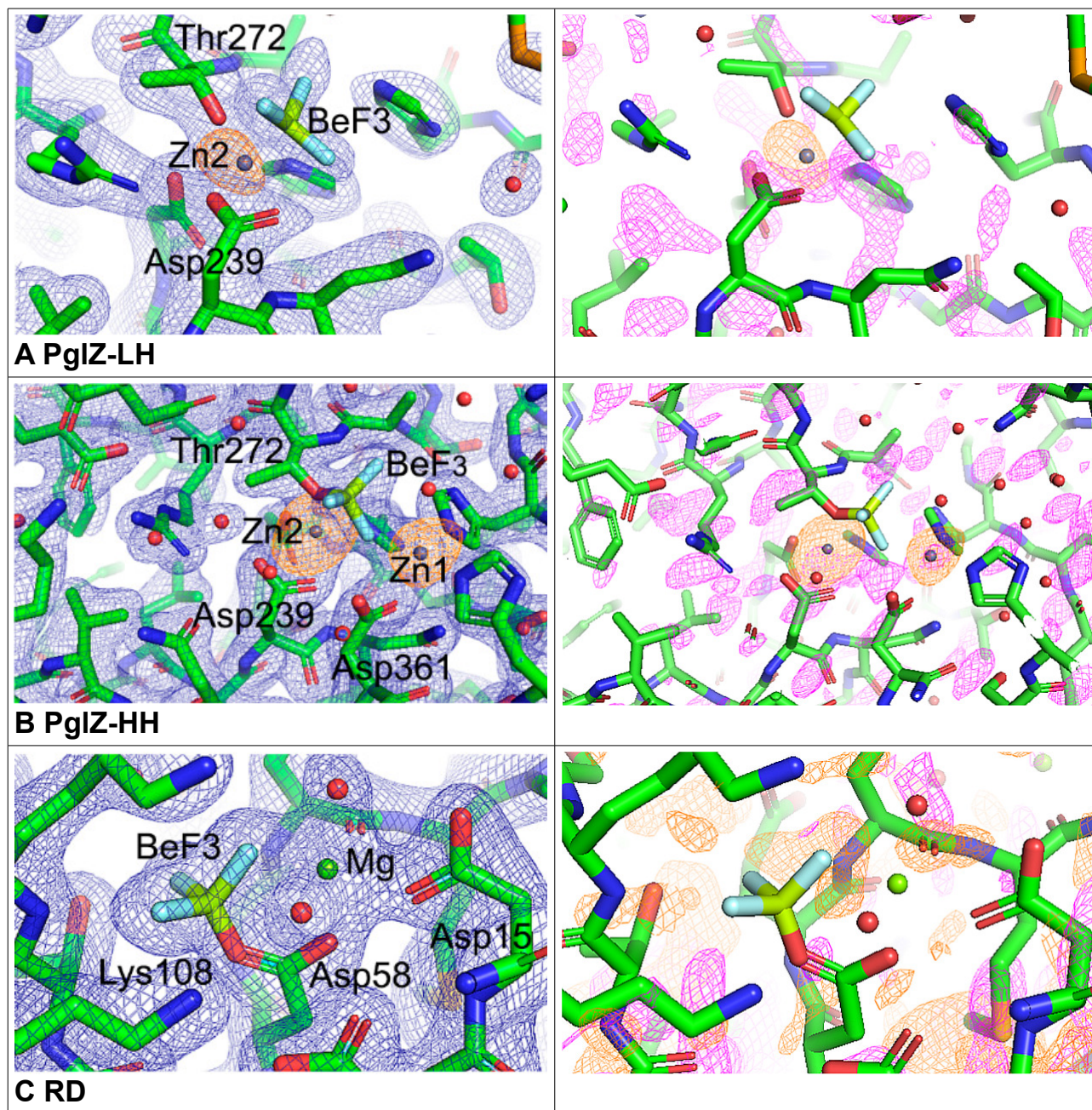
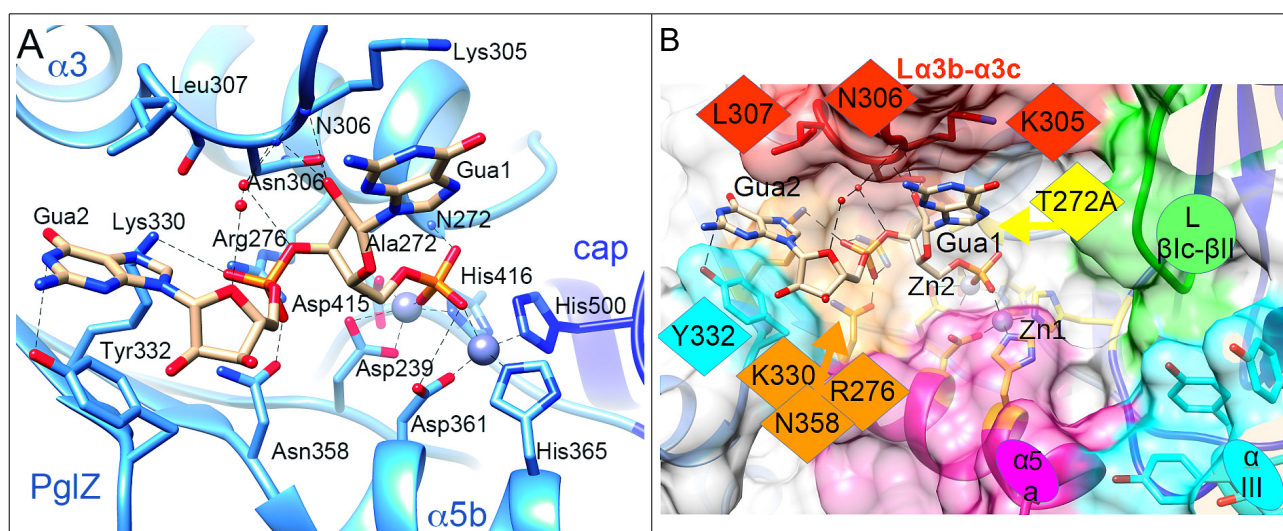


Figure 3.10: Anomalous diffraction electron density maps for HR structure. A-C left panels as published^[9]. **Left panels:** 2Fo-Fc (blue) electron density maps and Zn anomalous map (orange) of HR data. 2Fo-Fc pymol contour levels are 2.0 (A), 3.0 (B) and 1.0 (C). **right panels:** Mn anomalous map (magenta) and Zn anomalous map (orange). X-ray data of Mn anomalous signal was taken from a fragment of the same cluster as the HR data, soaked with a five times excess of Mn/Zn. Anomalous map pymol contour level for Zn is 15 for (A)&(B) or 2.0 for (C), for Mn they are 2.0 in each right panel. **(A)** PglZ active site in LH conformation. **(B)** PglZ active site in HH conformation. **(C)** Receiver domain.

3.9 "pGpG" structure, T272A mutant in complex with pGpG

pGpG is an intermediate hydrolysis product of c-di-GMP, which is a dinucleotide involved in environmental adaption, motility and biofilm formation (see introduction), thus we did co-crystallization trials with the supposed inactive mutant T272A. We successfully cocrystallized pGpG ("pGpG" structure, 2.4 Å, P21, 0.5 mM ZnCl₂ added) solved the structure by molecular replacement and built the corresponding model. Despite similar resolution to JG5 with a final R_{free} of the model of 21.6, the pGpG model could only be refined to an R_{free} of 22.3 due to pseudotranslational symmetry between the RDs of dimer AB with the RDs of the crystallographic symmetry related dimers CD* in P21 space group. Statistics on data, model and refinement can be found at Chapter 7.4. The ASU contained, as before, a double dimer. However, the contacts between both dimers changed and a less close crystal packing avoided the clashes that lead to alternative conformations at chains A/C in previous structures at the CAP flexible loop LβIc-βII. This region has the same conformation as seen for in previous chains B/D. Further PorX crystallized as observed before at the structures resulting from GMP or GTPγS co-crystallization trials. These crystallized as double dimers in the asymmetric unit, with each dimer AB and CD in double HH conformation and each bound to two zincs with comparable anomalous signals but without ligands bound. In contrast in the pGpG structure each chain carried one pGpG at the active site (Figure 3.11A, B). The 5'-phosphate of pGpG takes a similar position as BeF₃ in the JG5/HR structures (Figure 3.11C, D), and the phosphates O2 (without equivalent F in BeF₃) takes a similar



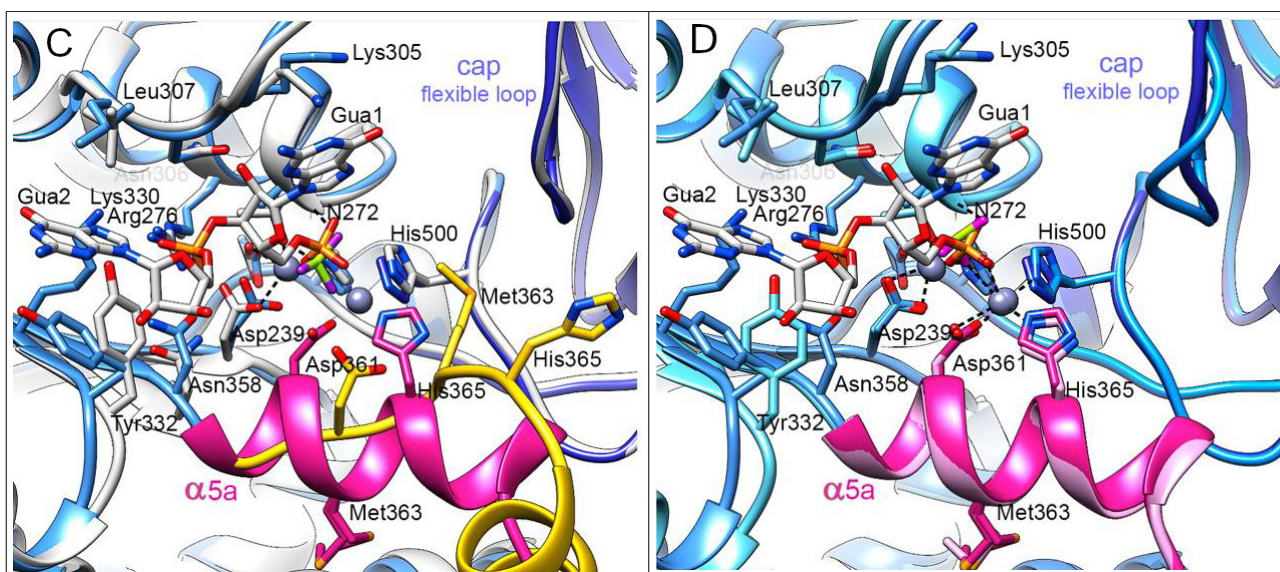


Figure 3.11: PorX T272A-pGpG active site. (A/C/D) as published [10] (A) shows the ligand binding in detail, binding residues and both guanosins are indicated (B) Substrate binding pocket. Loops and residues determining the binding surface are indicated. (C) Superposition between JG5/HR LH (gray and loop-helix in yellow) and pGpG bound HH conformation (blue and α5a in magenta). The cap flexible loop is indicated (D) Superposition between HH with (dark blue, dark magenta) and without (light colors) pGpG bound. (A/C/D) as published [10].

position as the active site threonine O^{γ1} close (3.05-3.41Å) to T272A alanine C^β (Figure 3.12A). The 5'-guanine "Gua1" is stabilized by loop Lα3b-α3c, while the 3' "Gua2" is stabilized by a hydrophobic clamp of L307 and Y332 which reorients upon ligand binding (Figure 3.11C, D). Helix α5a is not involved in substrate binding and pGpG is too small as a ligand to make contact with LβIc-βII or the tyrosine rich helix CAPαIII. Therefore, the PorX substrate binding cavity is only partially filled (Figure 3.11B). Lα3b-α3c stabilizes Gua1 by hydrophobic interaction with K305, as well as hydrophilic interaction of Gua1 O6' with K305 N^ζ and N2' with E302 backbone oxygen (Figure 3.12C, the later with chain-variations in bond-length between Gua1 and K305/E302 by chain: A:3.90/3.84; B:4.80/3.02; C:3.16/3.24; D:4.87/3.23; averg.: 4.18/3.33). The 5'-ribose is stabilized by hydrogen bonds between N306 O^{δ1} and backbone nitrogen with O2' as well as a water bridge stabilizing O3' contacting N306 and K305 backbone nitrogen (Lα3b-α3c). This water is further bridged through a second water with O5' of the 3'-ribose of pGpG (Figure 3.12B). Compared to previously described structures this substrate binding loop is dragged slightly closer into the active site and stabilized. This results in a much better defined electron density, especially compared to PorX crystallized in LH conformation, where this loop could not even be traced in the HR structure. In this new ligand bound conformation, Lα3b-α3c also binds and is stabilized by LαII-βV, a flexible loop at the CAP domain, not taking part in the active site. Specifically N466 of LαII-βV interacts with D298 and E300 of Lα3b-α3c. While the other regions of the PorX substrate

binding pocket (L β Ic- β II, α III and α 5a) do not undergo conformational changes during ligand binding, the side-chain rotation of Y332 by 60° opens a phosphate binding subpocket with residues K330, N358 and R276. In our structure those side chains coordinate the 3'-phosphate of pGpG, while Gua2 is stabilized by stacking into the hydrophobic clamp of Y332 in its new conformation and L307 (Figure 3.12D). While the electron density around the 5'-guanosine is well defined, at the 3'-end the electron density for Gua2 is rather weak in some chains, maybe due to alternative conformations. The flexible loop of the CAP region is in the same conformation as in the LH conformations due to a less tight crystal packing compared to JG5 or HR HH crystal packing between these loops of chains A and C.

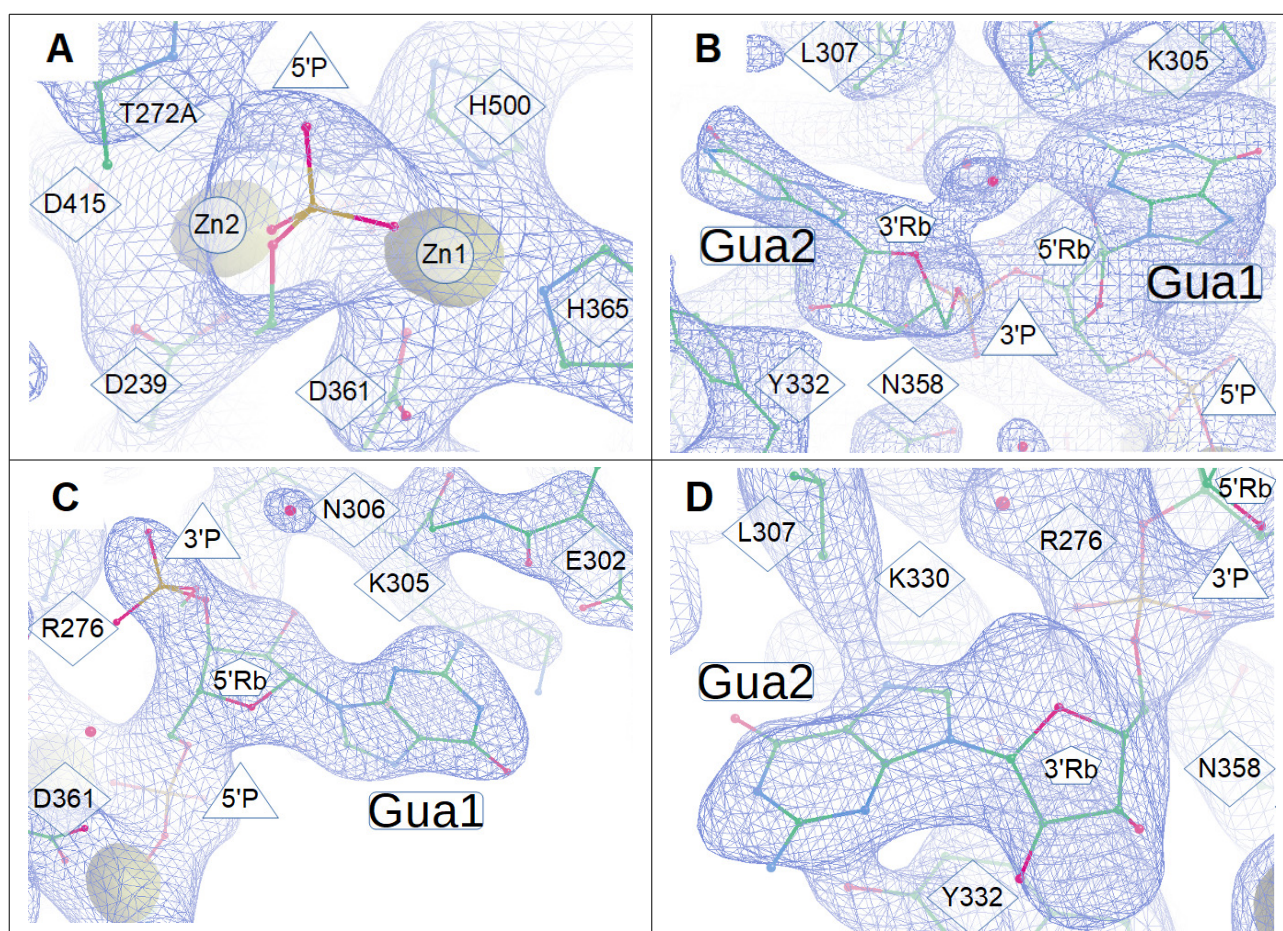


Figure 3.12: PorX T272A-pGpG active site electron density. PorX as ball and stick model, the 2F_c-F_o map represented as blue mesh, the anomalous map in yellow cut-glass representation. The σ -levels were selected for best ligand density visibility. **(A)** shows the coordination of the 5' Phosphate (5'P), which is located between both zincs and stabilized by a hydrogen bond with T272A backbone nitrogen. The O5' is close to A272 and in the wild type could form a covalent bond with native T272 under dehydration. **(B)** shows the relative position of both guanines as well the water bridge between them, **(C)** shows the coordination of guanosine 1 (5', "Gua1") and the 5'-ribose (5'Rb) which are stabilized by N306 and K305 (as well as E302 in some chains) by hydrogen bonds and hydrophobic interactions and **(D)** shows the coordination of guanosine 2 (3', "Gua2") including the 3'-ribose (3'Rb), phosphate is coordinated by R276, N358 and K330, whereas the guanine is stabilized by a hydrophobic clamp between L307 and Y332.

3.10 Other crystallization results and structures

During the development of this project hundreds of datasets have been collected and analyzed. Hereinafter only structures will be mentioned that yielded models that gave additional insights into PorX structure or substrate specificity. Further the results of crystallization trials with different protein types and mutants (mentioned at Chapter 2.1.11) or cocrystallization trials (mentioned at Chapter 3.4) are summarized.

3.10.1 Native structures

After solving the first PorX structures by SAD, it was possible to solve the phase problem of non-selenomethionine derivative PorX datasets by molecular replacement. This native structures were obtained at the beginning of the project and their data was mentioned as "initial datasets". The structures to these datasets in P21 (2.7Å, CC^{1/2}:50) and C2 (3.4 Å) were built and refined until Rfree of 27.0 and 34.0% respectively and did not show any structural differences. Thus and given they showed lower resolution, they haven't been build further.

3.10.2 "SCN" structure

In order to potentially have PorX in different crystal forms and conformations, we prepared the Optisalt-screening (Chapter 3.4.1), which with the additional 240 mM sodium thiocyanate yielded 2.4 Å data in P1 space group. The ASU contained 6 mixed LH/HH conformation dimers, two of which with translational non-crystallographic symmetry and two others with an additional rotation. The structure was built and refined to a Rfree of 27% mostly by superposition of the JG5 models into the respective chains, because the computational requirements for this data were too high to build. Superpositions of other PorX models resulted also in good phases and a useful electron density could be calculated to trace loops that were disordered or less defined for the JG5-PorX model. This was particular due to the fact that the data had useful reflections until 2.1 Å, although only at 46% completeness at the last resolution shell. Further it contained regions with different crystal-contacts and electron density maps of around 12 monomers could be merged to reduce noise levels at specific regions, especially less ordered loops.

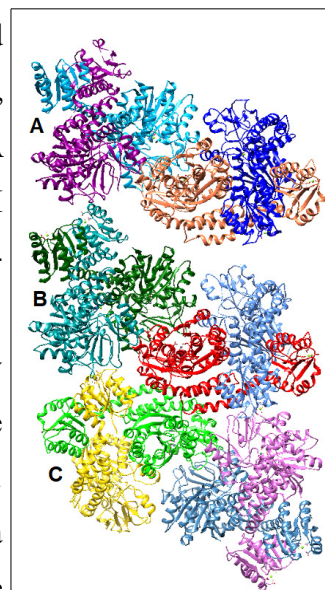


Figure 3.13: PorX SCN asymmetric unit (ASU)

The ASU of SCN data contains 6 PorX dimers in mixed LH/HH conformations and was used to trace regions in other models disturbed by different crystal contacts.

3.10.3 GMP cocrystallization

PorX crystallized in the presence of 0.5 mM GMP and 0.1 mM ZnCl₂ ("GMP" structure, 2.9 Å resolution, P21 space group) but the structure did not show any ligand bound to the active site. Surprisingly and in contrast to the HR structure which was soaked with GMP, both monomers within a dimer contained 2 zinc metals in respective PglZ active sites, both in HH conformation.

3.10.4 GTPγS cocrystallization

Co-crystallization attempts with a non-cleaveable GTP analogue yielded crystals and datasets up to 2.8 Å resolution ("GTPγS" structure, 0.1 mM ZnCl₂ added, P21). The protein crystallized in the same space group as the GMP structure (Chapter 3.10.3) with both monomers in HH conformation. The structure was built until an R_{free} of 25% but it was not possible to trace the position of any bound GTPγS in the active site. Also no rearrangements of the substrate binding pocket was observed and the anomalous signal for Zn1 was lower than that of Zn2, in contrast to pGpG-PorX but similar to the unliganded structures JG5 and HR. Statistics on data, model and refinement can be found at Chapter 7.4 .

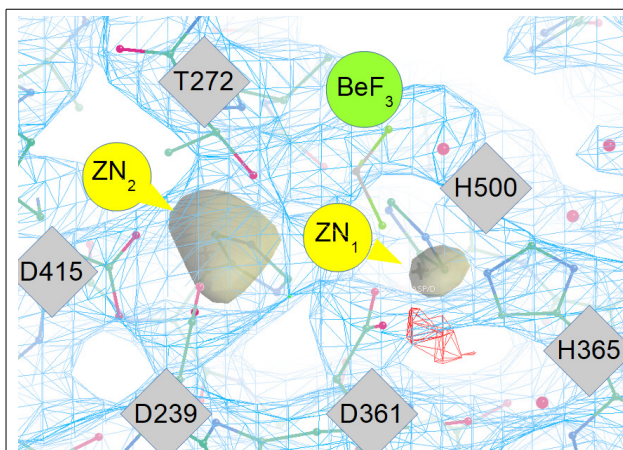


Figure 3.14:Active site GTPγS cocrystallization (without ligand bound).

Shown is one representative active site of the GTPγS structure. 2Fo-FC electron density map is shown in blue mesh representation and Fo-Fc difference map in green (not visible) and red (positive density). The anomalous map (Zn absorption edge data) is shown in yellow "cut-glass" representation validating the Zn positions, whereas one Zn signal is weaker as the other, maybe due lower occupancy.

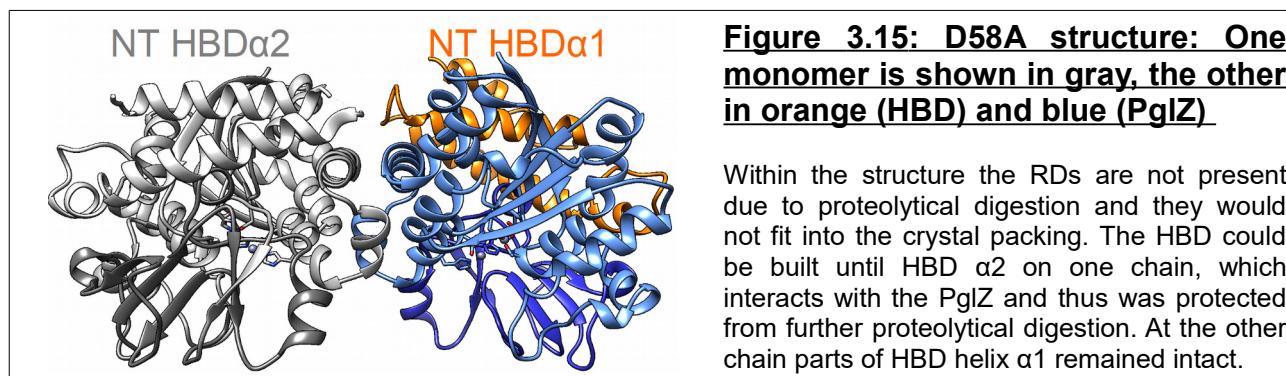
3.10.5 PorX M94K/D104A/I129A mutant

M94K/D104A/I129A samples were used for crystallization screening in 386 conditions, however no crystals were obtained due to high solubility, similar to the monomer.

3.10.6 PorX D58A mutant

The D58A mutant was crystallized and the corresponding structure solved and build to an R_{free} of 29% ("D58A" structure, 0.5 mM ZnCl₂, P31 2 1, 2.4 Å resolution). Unfortunately, building of the model and SDS-PAGE analysis of the crystallization drop showed proteolytical digestion within the coiled coil region linking the RD and PglZ domain. The fragment crystallized contained the PglZ region with part of the coiled-coil as a PglZ dimer in double HH conformation and two zincs each as confirmed by anomalous signals at the metal binding sites. Interestingly the proteolytical

digestion occurred at the N-terminal end of the HBD, specifically at the kink of helix $\alpha 1$ in one chain and at HBD helix $\alpha 2$ in the other. This confirmed the stability of HBD helices $\alpha 2$ and $\alpha 3$ in solution without contacts to $\alpha 1$. Furthermore it suggest mobility of $\alpha 1$ and confirms destabilization of $\alpha 1$ binding to $\alpha 2$ and $\alpha 3$ without contacts to the RD, which leads to $\alpha 1$ proteolytical digestion. Statistics on data, model and refinement can be found at Chapter 7.4 .



3.10.7 PorX T272A mutant

The T272A mutant was used for crystallization as negative controls of co-crystallization screenings to look inside the plates for different crystal shapes, which were not observed. Because we already obtained crystallographic data better than 2 Å, all T272A crystals were used for soaking experiments with pGpG, c-di-GMP, GTP and ATP. However, those crystals reproducibly "melted" after nucleotide addition, which was not visible under the microscope, but noticeable as smeared diffraction patterns during data collection. These crystals obtained from Opt_L optimization screening were large and unclustered and likely could have yield better data than 1.9 Å if they were not soaked but our focus at this point of the project already focused on substrate bound structures.

3.10.8 T272A+c-di-GMP cocrystallization

In order to crystallize potential substrates the supposed to be enzymatic inactive mutant T272A mutant was tried to be co-crystallized with c-di-GMP, which was likely to fit into the PorX substrate binding pocket. The structure of crystals obtained were solved (P21/21/21, 2.9 Å) and refined until an Rfree of 30%. No ligand was visible at the active site, nor changes in structure were observed. The conformations crystallized were LH/HH conformation dimers, despite the high concentration 0.5 mM ZnCl₂ in the crystallization buffer. Statistics on data, model and refinement can be found in Chapter 7.4 .

3.11 Mutational studies

Based on our JG5 structure we proposed a series of PorX mutants (some of them already mentioned before) to test them on their dimerization behavior (TABLE 3.14)

TABLE 3.14: PorX-Mutants	
D58A	RD phosphorylation site D58 to alanine mutation
T272A	PglZ temporal phosphorylation site T272 to alanine
M94K/D104A/I129S	Mutation of the dimerization surface of RDs
D361A/H365A	Zn1 binding residues (only in HH conformation)
S385E/S389E	PglZ dimerization surface mutation
D361A/H365A/S385E/S389E	combination of HH conformation destabilization from Zn1 combined with sterical hindrance of PglZ dimerization

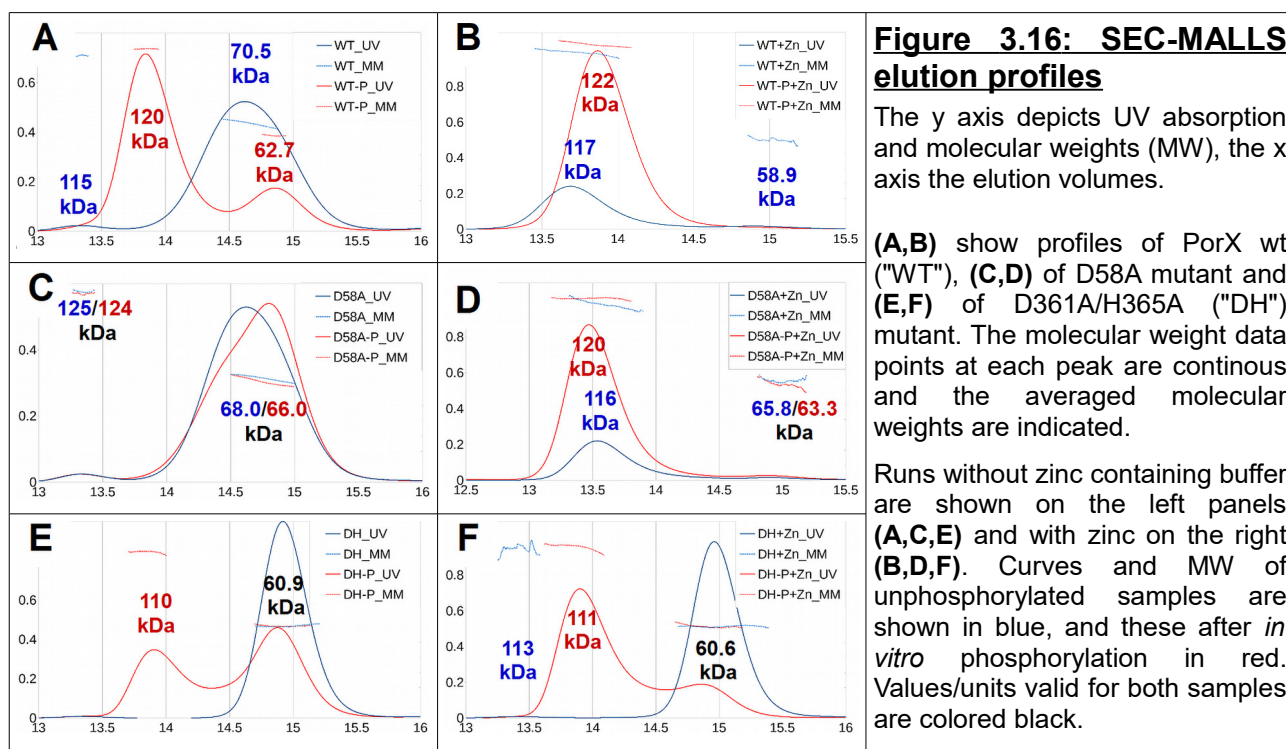
3.12 PorX dimerization analysis by SEC and SEC-MALLS

Dimerization behavior of PorX and its mutants was investigated via SEC by our collaborator Mariusz Madej from Jan Potempa's group, who tested the proposed PorX mutants with or without *in vitro* phosphorylation previous to the run (Figure 7.3). PorX was tested with and without *in vitro* phosphorylation and, with and without zinc addition (Figure 7.4), which might stabilize PorX PglZ HH conformation and thus an alternative zinc induced dimerization. The elution profiles of PorX and its mutants indicated that stable dimerization of PorX needed phosphorylation at the RD and intact RD dimerization interface, and a functional two zinc binding PglZ domain with intact PglZ dimerization interface.

In order to confirm these results and exclude delayed elution by matrix interactions we determined the absolute molecular weight of potential monomers/dimers (theoretical Mw= 61.1/122.2 kDa) by MALLS of wildtype ("wt") and representative mutants D58A (RD phosphorylation site) to disturb dimerization at RDs and D361A/H365A ("DH", mutated second zinc binding site) to disturb dimerization at the PglZs. We performed each run of the mutants and wt with or without previous phosphorylation and each in buffers with and without zinc addition. Samples running in zinc containing buffers were further preincubated with zinc for at least one hour. Statistics of each run can be seen in TABLE 7.3.

The elution profiles (Figure 3.16) showed that wt-PorX dimerized upon phosphorylation as well as by addition of zinc. D58A, the RD phosphorylation site mutant, showed no phosphorylation dependent dimerization, while the zinc induced dimerization was unaffected. DH-mutant, with disturbed Zn1 binding site, showed inhibited dimerization upon phosphorylation, as well as by zinc addition:

While wt-PorX is mostly present as monomer (98% of total mass) without zinc or phosphorylation (calculated Mw=70.5 kDa, higher than the theoretical value due to overlap with the broad dimeric peak at 115 kDa), phosphorylation shifted the oligomeric state into two separated peaks with mostly (81%) dimers (120 kDa versus 62.7 kDa monomers). The addition of zinc induced mostly dimers (93%, 117 kDa peak, with a small fraction of monomer at 58.9 kDa), while phosphorylation and zinc together let PorX to form only dimers (>99%, 122 kDa peak) (Figure 3.16, A, B).



The D58A mutant (Figure 3.16, C, D) eluted as monomer (97%) without zinc (68.0/66.0 kDa peak) independently of previous incubation with AcP due to the mutation of the phosphorylation site D58. Interestingly, like the wt, a small dimeric fraction (2%) was present at 125/124 kDa, which has to be independent of phosphorylation of D58. Likewise the unphosphorylated wt protein, in the presence of zinc D58A gets shifted to a dimeric state (120 and 116 kDa peaks, respectively) with still a remaining peak of 3-6% monomers at 56.8 or 63.3 kDa, independent of previous incubation with AcP.

The elution profiles of the DH-mutant (Figure 3.16, E, F) showed that this mutation at the PglZ domain had no effect on the unphosphorylated protein, which eluted as monomer (>99%) like the wt. However in contrast to phosphorylated wt, the phosphorylated DH-mutant had impaired ability to dimerize by RD phosphorylation and instead of a quantitative dimeric peak, an equilibrium of 42% dimer (110 kDa peak) and 58% monomer (60.9 kDa peak) was observed. The run in zinc

containing buffer showed stabilization of the dimeric state for the phosphorylated sample (111 kDa peak), but to a lesser extent (79%) than observed for wt PorX (>99%). This suggests that the additional intact zinc binding site of the wt stabilizes HH conformation and thus the dimeric state of PorX. On the other hand the unphosphorylated DH-sample could not be shifted quantitatively from monomers to dimers by zinc addition which lead to a dimeric mass fraction of under 1%.

3.13 Small Angle X-ray scattering

Because the crystallization trials of monomeric PorX were unsuccessful (ca. 1700 conditions screened, at concentrations up to 14.8 mg/ml, (see Chapter 2.1.11), we analyzed the overall structure of the monomer in solution by Small-Angle X-ray Scattering (SAXS) as well as the dimeric samples for comparison with the X-ray structure. Because we observed dimer formation during MALLS in the presence of zinc or by phosphorylation (Chapter 7.7), we tested samples in the presence of zinc as well. Further, from SEC experiments we knew of PorX RD autodephosphorylation (Figure 7.4) and used BeF_3 to mimic phosphorylation in order to avoid heterogeneity of the samples, as we did for the crystallization trials.

3.13.1 Oligomeric state and data quality analysis

The PorX samples without zinc or BeF_3 were mono-disperse and curves collected at different concentrations were merged. By discarding low q data we collected at high concentrations we avoided interparticle interaction scattering and by discarding low signals at high q from low concentration samples we avoided adding noise to our data. For the BeF_3 PorX-dimers, merging did not improved data quality and thus we used the data of the samples with the best signal to noise ratio (2.0 mg/ml for BeF_3 -PorX-dimer without zinc and 1.0 mg/ml for BeF_3 -PorX-Dimer with zinc). For "monomeric PorX" (without the addition of BeF_3) in the presence of zinc we observed different oligomeric states that depended on protein concentrations. Therefore we processed the datasets of 0.5 (monomeric) and 2.0 mg/ml (zinc induced dimer) separately and discarded the data of the polydisperse mixture we observed at 1.0 mg/ml.

Sample measurements and results including the calculated MWs are summarized in TABLE 3.15.

TABLE 3.15: PorX-SAXS data analysis

	PorX-monomer	PorX-monomer + zinc	PorX-monomer + zinc	PorX-Dimer	PorX-Dimer + zinc
Concentrations [mg/ml]	0.5, 2.0, 10.0 merged	0.5	2.0	2.0	1.0
Buffers	5 mM Tris pH 8.0, 150 mM NaCl				
	-	43 μ M ZnCl ₂		-	43 μ M ZnCl ₂
	-	-		15 mM MgCl ₂ , 30 mM NaF and 5 mM BeSO ₄	
q range [$\text{\AA}^{-1}$]	0.009-0.5152	0.0137-0.5152	0.0147-0.5152	0.0487-0.5183	0.0451-0.5065
I₀ (Guinier)	82.9	41.4	58.2	20.1	24.3
I₀ (P(r))	79.9	41.3	58.3	18.7	23.0
R_g (Guinier) [$\text{\AA}$]	30.9	35.6	39.1	34.5	36.3
R_g (P(r)) [$\text{\AA}$]	29.6	35.6	39.3	33.4	35.4
Dmax [$\text{\AA}$]	103	105	118	103	104
P(r) total estimate	0.854	0.859	0.855	0.942	0.955
P(r) χ^2 fit	1.09	1.05	1.06	1.07	1.07
MW [kDa]	61.1	61.1	122.2	122.2	122.2
MW (Vc, qmax=0.2) [kDa]	58.2	61.6	115.8	117.7	120.9
MW (Vp)	59.7	54.6	121.7	122.1	114.1

The MWs are consistent with either a monomer of 61.1 kDa (PorX without Zn or BeF₃ or PorX at low concentration (0.5 mg/ml) with zinc) or a dimer of 122.2 kDa (PorX with BeF₃ with or without zinc or PorX at high concentration (2.0 mg/ml with zinc but without BeF₃). The Kratky plots indicated compact protein folds with partial flexible regions. The monomeric proteins as well as the dimeric samples with BeF₃ were comparable in shape and size according to the R_g and Dmax, while the data of dimeric sample with zinc but without BeF₃ at higher concentration indicates a significant elongated shape. The corresponding curves for data analysis and quality control are shown in Figures 3.17-3.21.

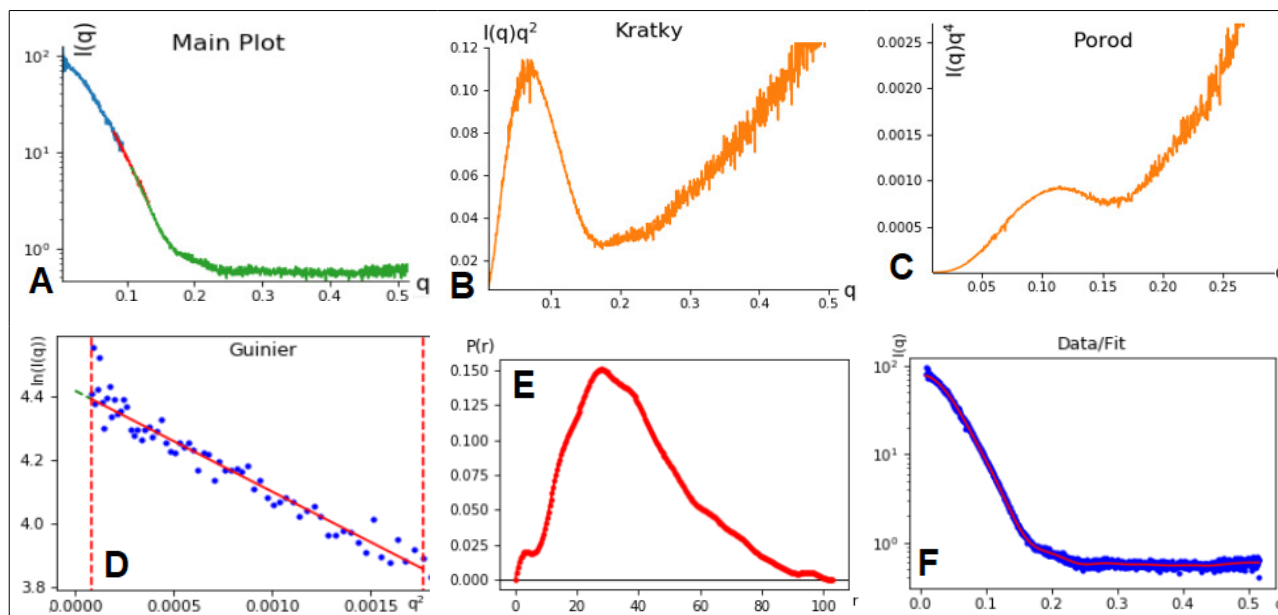


Figure 3.17: SAXS data/fit for monomeric PorX

(A) shows the raw data $I(q)$ which was merged from concentration of 0.5 mg/ml (blue), 2.0 mg/ml (red) and 10 mg/ml (green), (B) and (C) show Kratky and Porod plots, respectively, indicating folded protein with a few flexible regions. (D) shows the used Guinier fit at ideal low q region and (E) shows the distance distribution function $P(r)$ with its fit to the data shown in (F). Data quality is good and the $P(r)$ fit matches the data well. According to GNOM quality estimation the $P(r)$ fit is a "good" solution.

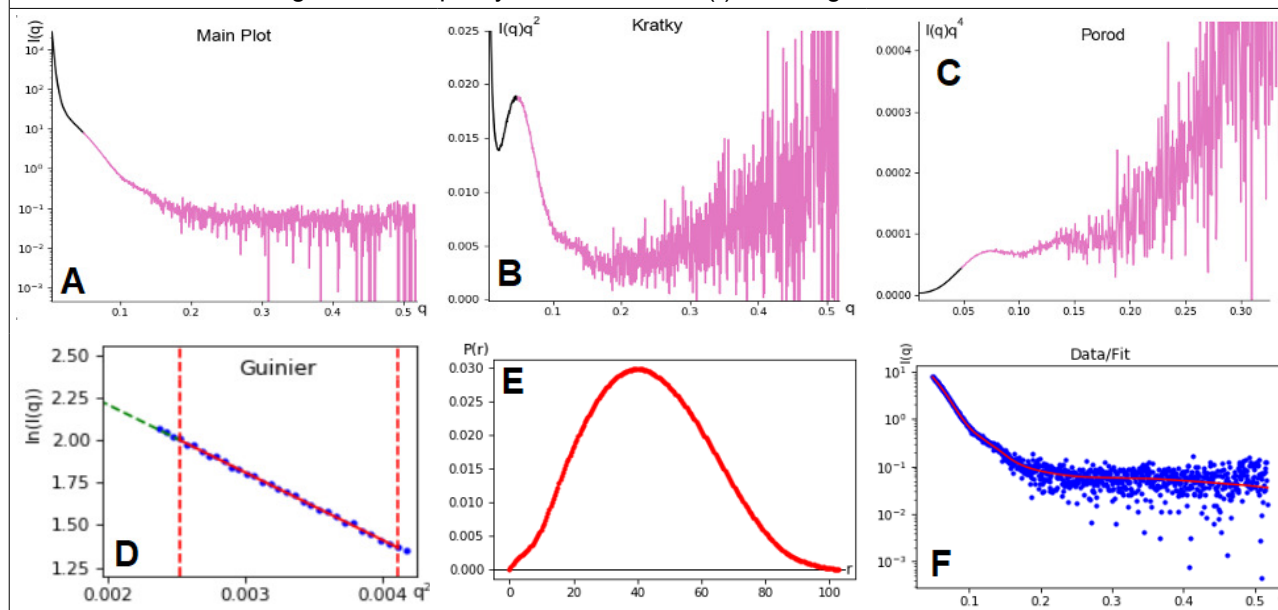


Figure 3.18: SAXS data/fit for dimeric PorX with BeF3 without zinc

(A) shows the raw data $I(q)$, data at low q (shown in black) shows strong inter-particle aggregation and was taken out for further analysis. (B) and (C) show Kratky and Porod plots, respectively, indicating folded protein with a few flexible regions. (D) shows the used Guinier fit, which could only be done at relatively high q due to interparticle interactions at low q but with a good fit in agreement with our MALLS data. (E) shows the distance distribution function $P(r)$ with its fit to the data shown in (F). Data quality is sufficient (bad low resolution data but good high resolution data) and fits match the data well. According to GNOM quality estimation the $P(r)$ fit is an "excellent" solution.

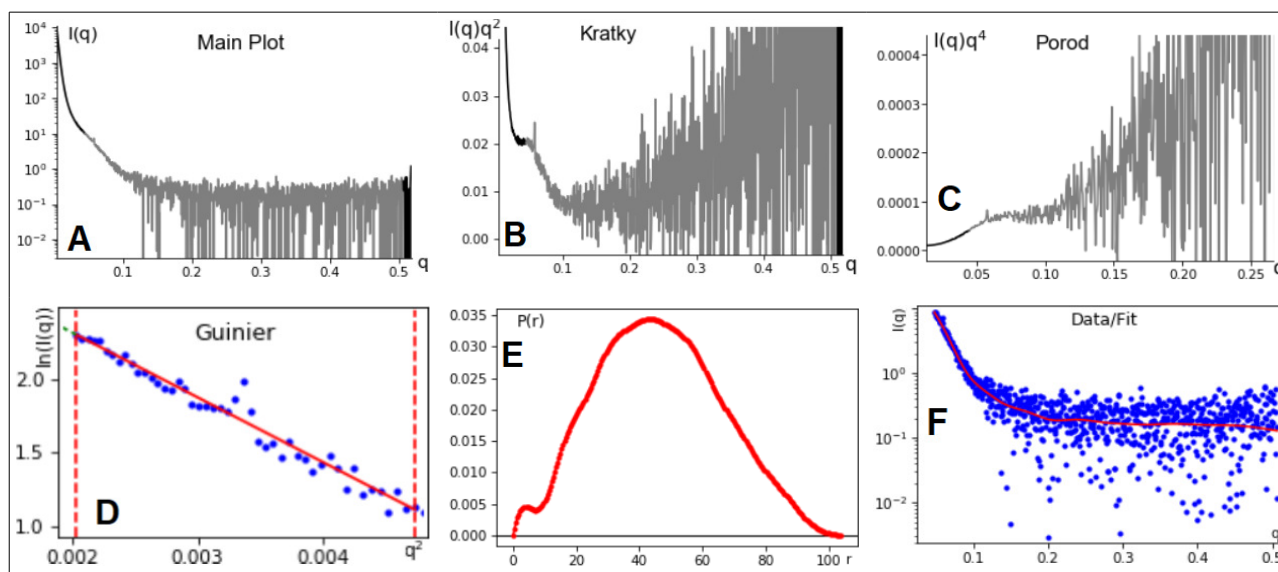


Figure 3.19: SAXS data/fit for dimeric PorX with BeF3 and zinc

(A) shows the raw data $I(q)$ in gray, data at low q (inter-particle aggregation) and high q (strong noise) was taken out from analysis and is shown in black. (B) and (C) show Kratky and Porod plots, respectively, indicating folded protein with a few flexible regions. (D) shows the used Guinier fit, which could only be done at relatively high q due to strong inter-particle interactions but with a good fit in agreement with our MALLS data. (E) shows the distance distribution function $P(r)$ with its fit to the data shown in (F). Data quality is sufficient and fits match the data well. According to GNOM quality estimation the $P(r)$ fit is an "excellent" solution.

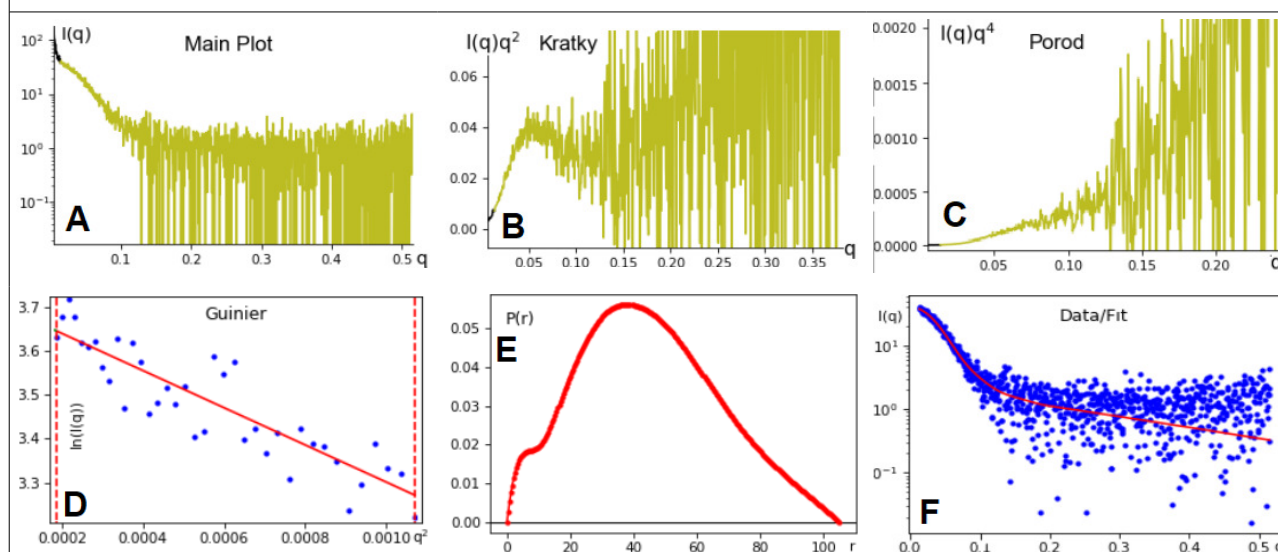


Figure 3.20: SAXS data/fit for monomeric PorX at low concentration without BeF3 but with zinc

(A) shows the raw data $I(q)$, data at low q (inter-particle aggregation) was taken out from analysis and is shown in black. (B) and (C) show Kratky and Porod plots, respectively, indicating folded protein. You see high noise levels at high q due to the low sample concentration. (D) shows the used Guinier fit, which is excellent due to low inter-particle interactions at low q . (E) shows the distance distribution function $P(r)$ with its fit to the data shown in (F). Data quality at low q is high, while high q /low has a lot of noise, thus the $P(r)$ fit at this region is suboptimal. However, according to GNOME quality estimation the $P(r)$ fit is an "good" solution to the data.

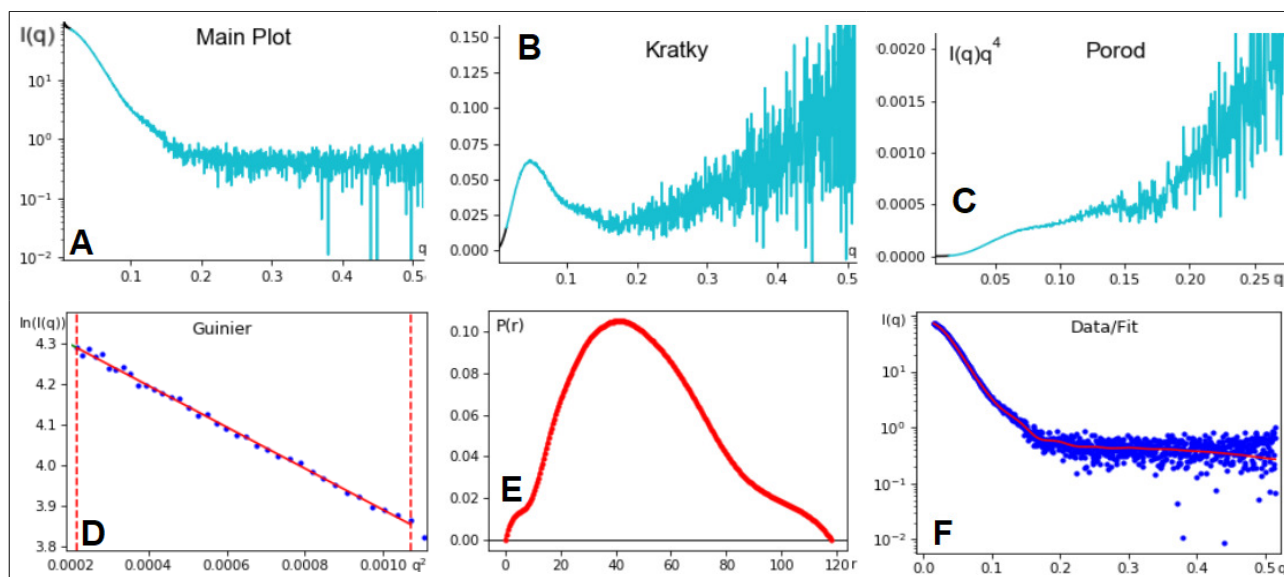


Figure 3.21: SAXS data/fit for dimeric PorX at high concentration without BeF₃ but with zinc

(A) shows the raw data $I(q)$, data at low q (inter-particle aggregation) was taken out from analysis and is shown in black. (B) and (C) show Kratky and Porod plots, respectively, indicating folded protein with a few flexible regions. (D) shows the used Guinier fit, which is excellent due to low interparticle interactions. (E) shows the distance distribution function $P(r)$ with its fit to the data shown in (F). Data quality is good both at low and high q and the $P(r)$ fit matches the data well. According to GNOME quality estimation the $P(r)$ fit is an "good" solution.

3.13.2 Electron density reconstructions by DENSS

Based on the SAXS data we used GNOM (software used see Chapter 2.1.7) to calculate the inverse Fourier transformation (IFT) which we used as input data for *ab initio* 3D electron density reconstructions using DENSS. Data quality was assessed by Ambimeter of the BioXtasRaw package. Lower Ambimeter scores indicate more reliable results than high scores. The calculated densities and refinement statistics are summarized in TABLE 3.16.

Regarding monomeric samples, the monomer (without BeF₃ or zinc) had the best data quality due to merging of different concentrations and yielded an electron density well suitable for modeling with a "potential unique" solution (Ambimeter score 1.18).

PorX monomeric data quality at low concentration (0.5 mg/ml) with zinc was considered insufficient and highly ambiguous for reliable density reconstruction (Ambimeter score 2.80). Despite the data at low q was good the signal over noise ratio was very low at middle and high q . Thus the calculated electron density has to be seen with caution. The map is similar to the map of the monomer without zinc and the relative position of RD and PglZ centers of mass are equal. Yet, the density of the PglZ is discontinuous with a lot of noise and at a much lower resolution which

can be explained by the low concentration of the sample measured and flexibility at the dimerization interface.

The data from PorX with zinc data at higher concentration (2.0 mg/ml) was better (Ambimeter score 2.08) and sufficient for a consistent electron density reconstruction but still with some ambiguity. The shape shows two smaller domains of the size of the RDs pointing at opposite sites of the dimer, while likely (in size and shape) the two PglZ domains dimerize at the center.

The BeF₃ dimer yielded the second best data, which by Ambimeter should lead to a potentially unique 3D reconstruction (Ambimeter score 1.43). The electron density fits our crystallographic dimer very well in size and shape, considering the differences in resolution.

The BeF₃ dimer in the presence of zinc yielded sufficient but ambiguous (Ambimeter score 1.57) data, however in size and shape it was equal to the dimer without zinc addition, thus fitting our crystallographic data. In contrast to the other datasets there was no improvement by increasing the number of voxels (slow mode) or the number of reconstructions, as this only increased the noise levels outside the continuous density due to large low q cut-off during data processing.

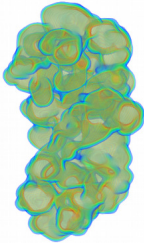
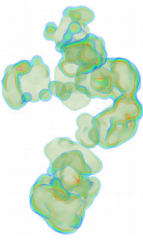
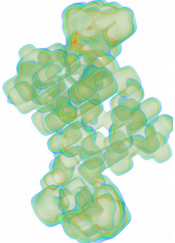
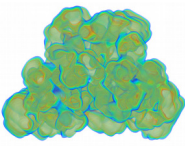
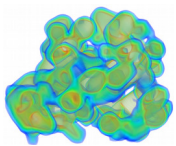
TABLE 3.16: PorX-DENSS electron density reconstructions

Visualization of electron densities were performed by Pymol

	PorX- (monomer)	PorX + zinc low concentration (monomer)	PorX + zinc high concentration (dimer)	PorX + BeF ₃ (Dimer)	PorX-Dimer + BeF ₃ & zinc (Dimer)
Ambimeter data quality assessment: <1.5 = good 1.5 < 2.5 = ambiguous <2.5 insufficient	Score : 1.176 Good 3D reconstruction is potentially unique	Score: 2.802 Insufficient 3D reconstruction is highly ambiguous	Score: 2.083 Ambiguous 3D reconstruction might be ambiguous	Score: 1.431 Good 3D reconstruction is potentially unique	Score: 1.568 Ambiguous 3D reconstruction might be ambiguous
Concentrations [mg/ml]	0.5, 2.0, 10.0	0.5	2.0	2.0	1.0
Number of averaged ab initio maps	100	20	100	100	20
Mode	slow, defaults	slow, defaults	slow, defaults	slow, defaults	fast, dmax=93 (noise reduction)

TABLE 3.16: PorX-DENSS electron density reconstructions

Visualization of electron densities were performed by Pymol

	PorX- (monomer)	PorX + zinc low concentration (monomer)	PorX + zinc high concentration (dimer)	PorX + BeF ₃ (Dimer)	PorX-Dimer + BeF ₃ & zinc (Dimer)
Ab initio electron density reconstruction s by DENSS. (not scaled to each other)					
Resolution [Å] map average	34.0	54.7	61.5	49.0	47.0
Refinement steps	4184	10824	10999	10999	10259
oversampling ratio	5	3	5	5	3
Final Chi²	1.112e-01	7.612e-05	4.252e-03	5.540e-01	1.727e-03
Final Rg [Å]	29.94	38.57	42.22	34.98	39.77
Support Volume [Å³]	93527	147490	317808	186665	430138

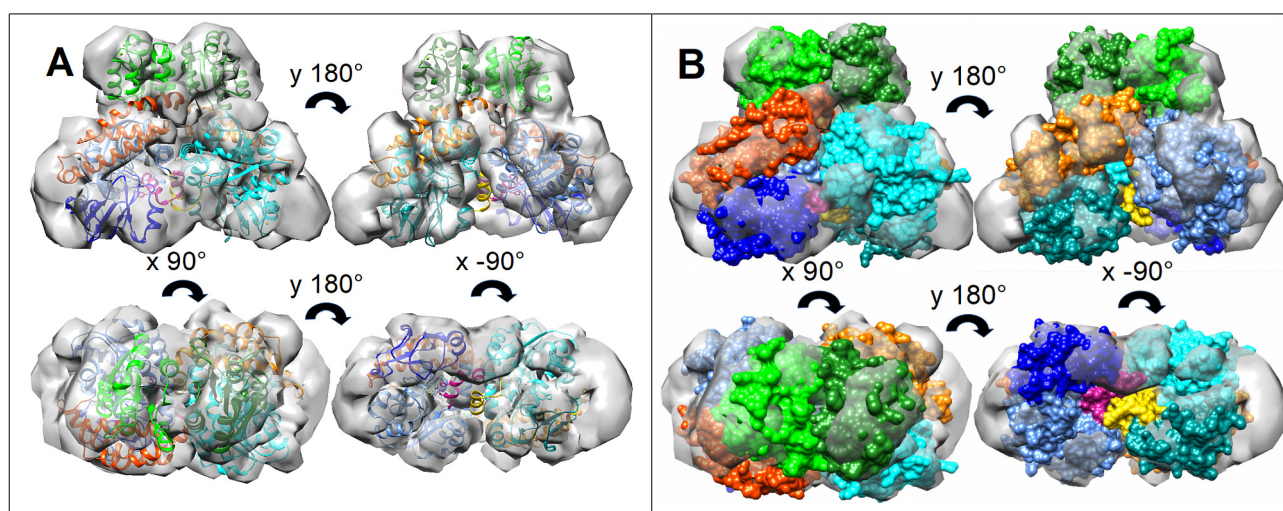
3.13.3 SAXS modeling

Based on data quality only the potential unique solutions were investigated further. The SAXS PorX+BeF₃ dimer without zinc was compared to our crystal structure and analyzed for conformational differences in solution. The data of PorX+BeF₃ and with zinc was identical in shape but at lower resolution and would not add additional informations (see TABLE 3.16).

The PorX monomeric data without BeF₃ or zinc-addition was analyzed to investigate conformational changes that could explain enzyme activation by dimerization. Also here, the other monomeric data (0.5 mg/ml with zinc) had a worst resolution at similar shape and was not modeled for the same reason as the BeF₃ dimer with zinc. The dimeric data of PorX with zinc but not BeF₃ was not modeled, as it is likely a non-physiological dimer only occurring at high - non-physiological concentrations.

3.13.4 SAXS model of PorX dimer

We used dimeric crystal structures of PorX as a model for the SAXS data and got a very good fit for the DENSS (*ab initio* reconstructed) electron density as well as between theoretical calculated SAXS curves and experimental measured data (Figure 3.22). The almost identical structures HH/LH JG5 and HR have a χ^2 goodness of fit of 1.85 and 1.81 (Figure 3.22A, B, F), respectively. Comparing the data to the HH/HH pGpG structure even resulted in a χ^2 of 1.58 (Figure 3.22E). However, for the latter the overall fit does not result from a better fit at the active site, but rather from a less compact crystal packing of the pGpG PorX dimer which might more reflect the protein in solution. The unideal $\chi^2 \neq 1.0$ results from inter-particle interactions at low q , visible as the divergence of the normalized curves (Figure 3.22E/F right panels) at low q . Taking a closer look to the active site superposition (Figure 3.22C), we can see a high solvent accessibility at the active site. Surprisingly at the dimerization interface there is less defined electron density at helices $\alpha 5b$, suggesting a higher degree of flexibility in this region and a potential equilibrium between HH/LH and LH/HH conformations, in the absence of substrates. The flexible regions of PorX (Figure 3.4) L β Ic- β II and L α 3b- α 3c are out of density and thus likely disordered in solution, as their high b -factors in the crystal structures suggested, while α III of the CAP domain is very well defined close to the active site. While, despite the low density map resolution, even smaller helices and loops can be located in this map, one major difference between the PorX crystallographic dimer and the density exists (Figure 3.22D). This is an extra electron density area in the same shape and dimensions as the CAP domain, in an position that would allow an even more opened active site pocket. However the position of the CAP subdomain in the model is necessary to explain the density there, which is consistent placed inside the density as well.



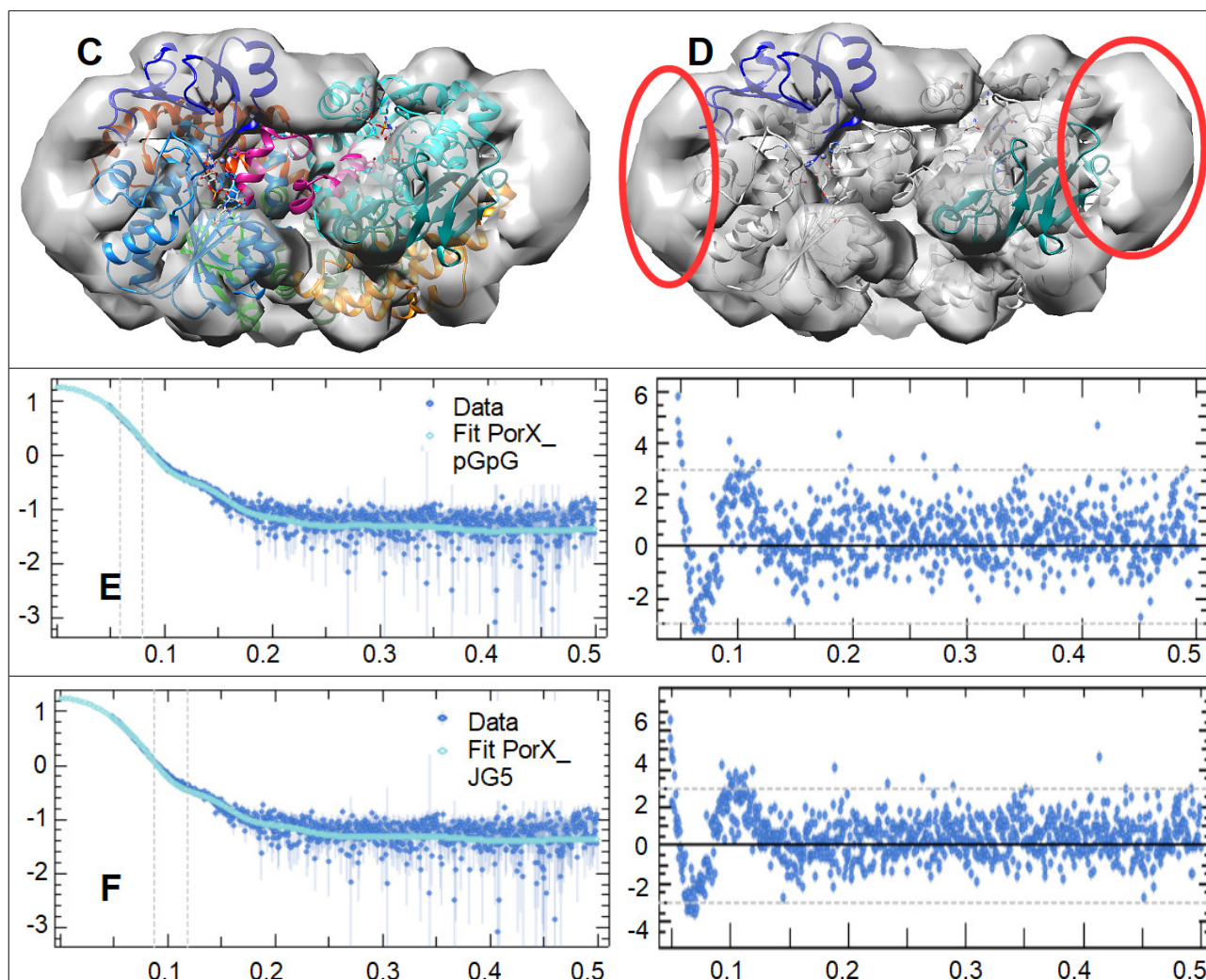


Figure 3.22: PorX SAXS-Dimer model. Superposition of crystal structure and DENSS electron density reconstruction.

(A) The crystallographic dimer (pGpG chains A&B) was used as SAXS model for the PorX-BeF₃ data and placed into the calculated DENSS ab initio electron density map. Chain A is represented in darker colors than B, RDs are colored green, HBDs orange, PglZs catalytic subdomains in cornflower blue for chain A/cyan for chain B and CAP subdomains in dark blue (chain A)/dark cyan (chain B). Further the two alternative conformation regions $\alpha 5a$ /loop and $\alpha 5b$ at the active site centers are colored yellow (LH) and magenta (HH). (B) same as A but PorX in surface representation. (C) Enlarged bottom view onto electron density of the dimer, superposed to pGpG-AB dimer structure. (D) same orientation as (C), only the two CAP domains are colored for better visibility of their fit. Red circles indicate empty electron density of similar size and shape than that of the CAP domains. (E) Data and PorX pGpG-AB model fit, left: representing rel. scattering intensity (y axis) vs. q range (x axis), right as normalized curves. Final $\chi^2=1.58$. (F) same as (E) but for JG5-AB as model, $\chi^2=1.81$.

3.13.5 SAXS model of PorX monomer

The electron density map of the monomeric sample fitted a single PorX molecule with two globular domains in the size of the PglZ and the RD, separated by a linker region similar as in a single PorX chain of the crystallographic dimer. The central HBD and the core of the catalytic subdomain of PglZ fit the density very well.

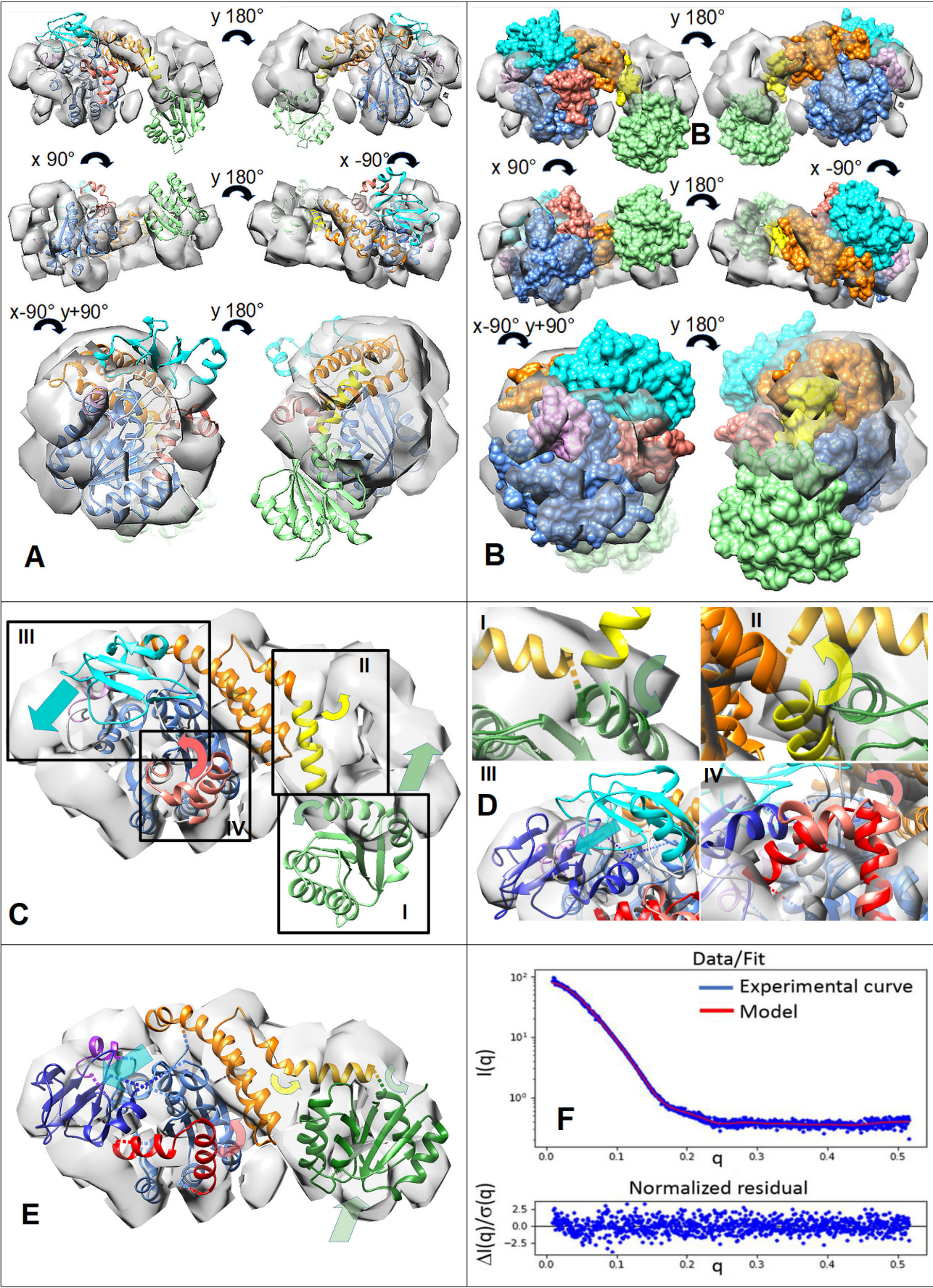


Figure 3.23: PorX SAXS-Monomer model, crystal structure superposition and DENSS electron density reconstruction.

(A, B) Superposition of a single PorX chain of the crystallographic dimer with the DENSS electron density map in ribbon (A) and surface (B) representation. Different subdomains that need to be individually re-positioned to optimally fill the density map, are shown in different colors: RD: green, HBD helix $\alpha 1a$: yellow, PglZ-CAP: cyan, dimerization interface: red, and already aligned domains are colored as in previous figures: HBD helices $\alpha 1b$ - $\alpha 3$: orange and the PglZ catalytic subdomain: cornflower blue. (C) highlights the specific regions that need to be re-positioned, the vectors are indicated by arrows. PglZ helix $\alpha 3$ (in violet) minor repositioning can be neglected. (D) zooms into the regions I-IV of (C) together with superpositions between crystallographic PorX chain of a dimer and the final SAXS monomeric model with re-positioned subdomains in darker colors. (E) shows the final SAXS monomeric SREFLEX model with all subdomains re-positioned and superposed into the DENSS electron density map. (F) calculated fit between SAXS data and model (as in (E)) as output by CRY SOL, final $\chi^2=1.15$.

The rotation of the PglZ catalytic subdomain could be recognized by the density of its surrounding helices (Figure 3.23A+B), which had the same relative position to the HBD as in the dimer. Apart from this regions, the CAP domain and the RD needed to be repositioned to fit the density. Furthermore, the RD repositioning needed a change in conformation at both ends of the kinked helix HBD_ $\alpha 1a$ (Figure 3.23C+D), including the connection to the remaining HBD and the RD as well. At the dimer interface of PorX PglZ helices $\alpha 1a$ -c was no clear continuous electron density, suggesting disorder in the monomer in solution.

Based on visual inspection and fits into the DENSS electron density initially we subdivided PorX into subdomains. These subdomains were then automatically repositioned by SREFLEX based on the raw data. The quality of fit was calculated by CRY SOL resulted in a final χ^2 of 1.15. A perfect fit ($\chi^2 = 1.00$) might need to include the missing loops that we have not modeled. The alternative conformation of the CAP subdomain (see Figure 3.23C+D, III) would close the active site of PorX and thus is not identical with the potential alternative conformation of the CAP domain we observed in the DENSS electron density of the dimer (Figure 3.22D).

3.14 Comparative homology

3.14.1 DALI analysis & homology search in the PDB

In order to determine the function of PorX, extensive structural homology searches were performed. Preliminary results of these comparisons lead to cocrystallization and soaking experiments with dinucleotides which yielded our PorX_pGpG structure. We tested dinucleotides because PorX substrate binding pocket seemed to be much larger than those of homologs binding only mononucleotides. The binding mode of pGpG in PorX, while other substrates including mononucleotides did not bound, lead us to assume potential substrates even larger than dinucleotides. Therefore a full screening by ultrahigh-performance liquid chromatography/mass spectrometry (UHPLC-MS) was performed by our collaboration partners.

As search model for similar proteins the PglZ domain (aa214-518 of chain A of the structure "PorX_pGpG") was used and compared to all structures in the PDB using DALI server (other search models of PorX e.g. in PglZ in LH conformation did not yield significant other results). The closest 115 homolog structures in the PDB were then summarized to the 20 closest structural homolog proteins (including all mutants, orthologs and substrate bound structures). They had a minimal Zscore of 11.5 (Chapter 7.8, TABLE 7.4) and were used to compare the overall structure, enzymatic metal coordination sphere and substrate binding pockets. In addition, PorX was superposed and compared manually to all APS members/"alkaline phosphatase clan" enzymes in the PDB to find similar bound ligands and/or a comparable dimerization interface. Also evolutionary more distant enzymes like adenylate or guanylate cyclases were compared for similarities of their substrate binding pockets. As a result PorX shows the specific qualities of phosphodiesterases (PDEs) of the APS with its conserved two metal coordination shell and active site threonine (T272 in PorX), while alkaline phosphatases (AP) usually have an additional third metal binding site (Mg^{2+}) and an active site serine. Sulfatases have a single metal (Ca^{2+}) binding site different from PDEs and APs (see Figure 3.24).

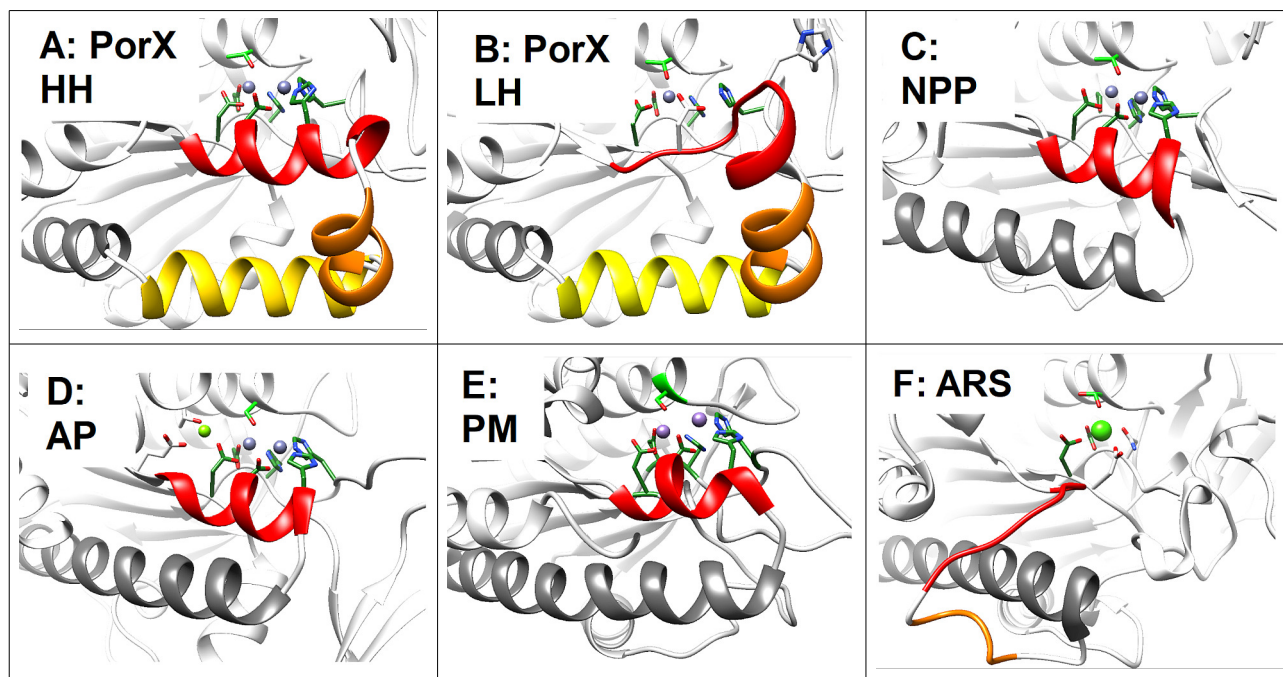


Figure 3.24: PorX active site and dimerization interface conservation throughout the alkaline phosphatase superfamily (APS).

A-F similar as published^[9]. Conserved residues between PorX and APS members are colored dark green, while the catalysis specific active site substrate acceptor is colored light green. An active site Thr is characteristic for nucleotide pyrophosphatases/phosphodiesterases (NPP), Ser for alkaline phosphatases (AP), Ser or Thr for phosphotransferases/-mutases (PM) and formyl glycine for arylsulfatases (ARS). Regions comparable in sequence, secondary structure or positions to the PorX dimerization interface are shown between APS in the same colors. These are red for PorX-HH $\alpha 5a$ or the corresponding loop in PorX-

LH, orange for $\alpha 5b$, yellow for $\alpha 5c$ and dark gray for $\alpha 5d$. All metal coordinating residues, except for sulfatases, are conserved within the superfamily. There are no regions comparable to PorX dimerization interface at helices $\alpha 5b/c$. **(A/B)** PorX reference structures in HH and LH conformation respectively. PorX HH possesses the canonical nucleotide pyrophosphatase (NPP) active site, which is broken in LH conformation. Respective rearranged residues A361 and H365 positions are colored white in (B). **(C)** Structure of Xac nucleotide pyro-phosphatase/phosphodiesterase (NPP, pdb: 2gsn). **(D)** E. coli alkaline phosphatase (AP, pdb: 1B8J), the active site contains an additional magnesium binding site. **(E)** Structure of phosphoglycerate mutase (phosphatase domain) as an example of active Mn containing enzyme common for phosphotransferases/mutases. **(F)** Structure of human arylsulfatase-A (ARS-A, pdb: 1n2K). In ARS helix $\alpha 5a$ is missing and the active site harbors a single catalytic calcium ion. Therefore the metal coordinating residues are exchanged in favor of calcium and Zn1 binding site is absent.^[9]

The PorX two metal binding site in HH conformation is identical to the canonical coordination shell and reaction pathway of bi-metallo APS as published by Sunden et al. (2017).^[1] Our PorX-PglZ structures, containing either BeF_3 or pGpG, represent the substrate hydrolysis transition state intermediate (similar as vanadate in the Xac NPP, PDB: 2gso)^[2] which is conserved throughout the superfamily, except those with single metal active sites. Therefore it can be assumed that PorX inherits the same reaction pathway as common for phosphatemono- and -diesterases as well as phosphotransferases (Figure 3.25).

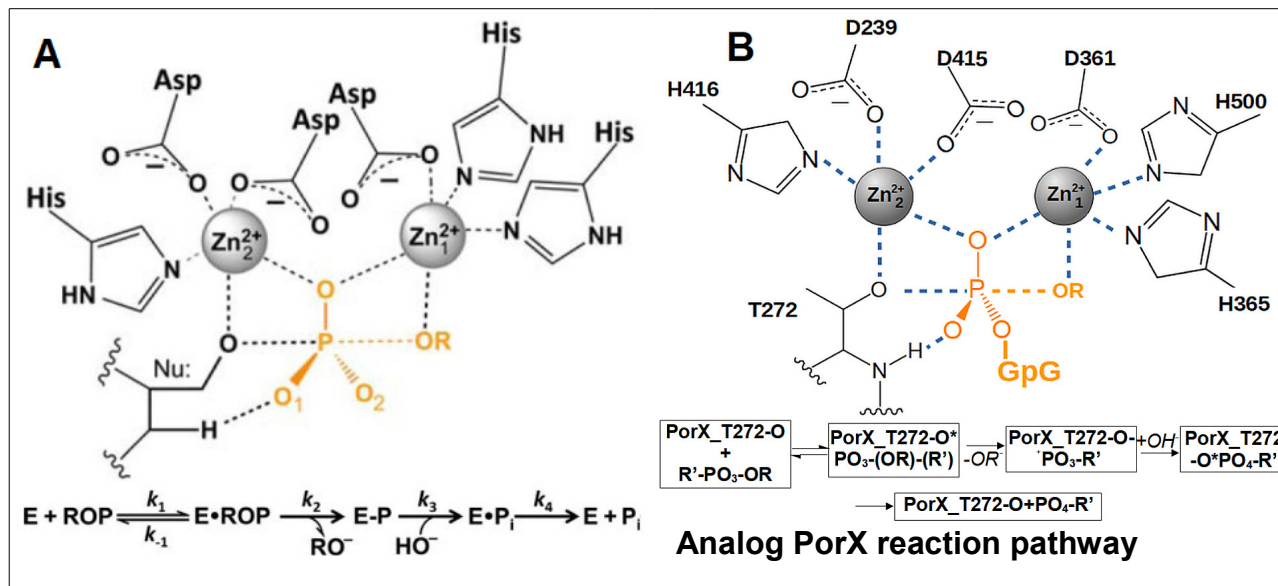


Figure 3.25: PorX shares the APS two metal active site.

(A) "AP superfamily bimetallo active site and reaction pathway" as published by Sunden et al. (2017)^[1].

(B) PorX-PglZ active site with bound pGpG-R ("R" for unspecified rest moiety), the metal coordination shell is identical to the one in (A), the reaction scheme was added analog to (A) in four steps a)-d). In a first reaction step a) the enzyme E binds covalently to the outgoing phosphate group OP in a transition state intermediate $E \cdot \text{ROP}$. This phosphoester bond is mimicked in our crystallographic structures of wild type PorX by binding to BeF_3 or in our pGpG_T272 PorX structure by binding to pGpG with R only being a hydrogen atom. In the next step b) the negative charged (now dephosphorylated group) RO^- leaves (in case of PorX_pGpG structure a hypothetical longer nucleotide fragment OR, while pGpG stays covalently bound to PorX). Then follows in c) a nucleophilic attack of a close hydroxy-ion to the phospho-Ser/Thr (T272 in PorX); At the last step d), inorganic phosphate P_i leaves, in case of monoesterases. R_2OP is the leaving group in the case of diesterases (pGpG in PorX), where O2 in the scheme in A) is substituted with a second organic rest OR_2 . In phosphotransfer reactions R_3^*O does the nucleophilic attack in step c) instead of HO^- .^[1]

Dendrograms generated by DALI [data not shown] did not show any specific evolutionary relationship of PorX with one of the homolog enzymes in the PDB. This suggests the PglZ domain diverted an early stage of alkaline phosphatase superfamily evolution and the degree of conserved residues outside the active center is too low to trace back any coevolution of PglZ with one of its homologs within the PDB database. The top 12 closest structural homologs (DALI Z-scores higher than 14.9) are phosphodiesterases (PDE), phosphonoacetate hydrolases or alkaline phosphatases (AP) with two conserved zinc metal centers (Figure 3.24). The phosphonoacetate hydrolases have two zinc metal center as PDEs but in contrast a very small substrate binding pocket. The APs PafA as well as its close homolog PhoK are exceptional two metal center APs. In PafA and PhoK structures the common third metal (Mg^{2+}) binding site of APs is absent and the active site serine is replaced by threonine as in PDEs (probably due to enzyme co-evolution as suggested by Sunden et al. (2017)^[1]. This might have been possible for PorX as well and this was one reason we also tested PorX for alkaline phosphatase activity (Chapter 3.15).

Further homologs containing two metal centers were the phosphoglycerate mutase GpmI and the phosphatidylethanolamine transferase MCR1. GpmI contains an active site serine as typical for the common AP active site, the other metal coordinating residues are identical to PorX but in complex with two manganese ions. This was one reason we also tested anomalous diffraction of PorX crystals at the manganese absorption edge (Figure 3.10). MCR1 is very similar to the phosphatidylethanolamine transferase EptC which contains a single zinc ion in the active site, whereas MCR1 could be crystallized with two zincs in the active site but the binding site of this second zinc differs from that of PorX. The remaining, more distant, homologs are either sulfatases or sulfamidases, which contain a single calcium in the active site whereas the active site residues binding the second metal in PorX are not present (Figure 3.24).

Importantly, and to the best of our knowledge none of the homolog structures as well as no APS protein structure published in the PDB database has a dimerization site similar to PorX. This includes the characteristic alternative conformations of helices $\alpha 5a/b$ and the separation between helices $\alpha 5c/d$ by a kink or a small loop, which seems to be unique for PorX so far (Figure 3.24).

These results and the large degree of substrate cross-promiscuity of the PDE with two zinc metal center (Mohamed&Hollfelder, 2012)^[3] lead us to comparisons of the substrate binding cavities surrounding the active site of PorX with that of its homologs found by DALI. These, as well as their enzymatic activities and literature references can be found in TABLE 7.4.

3.14.2 Substrate binding pocket comparisons

Likewise we compared the metal coordination sphere of PorX to its closest homologs we also compared their substrate binding sites. All surfaces, corresponding structures, references, function and substrate specificity can be found at the Supplementary materials at Chapter 7.8.

The PorX substrate binding site in PorX_pGpG can be divided into three larger binding subpockets A, B and C. A is occupied by one guanosine and, when we compare pGpG bound with PorX unliganded structures, A opens by ligand binding due to rotation of Y332 side

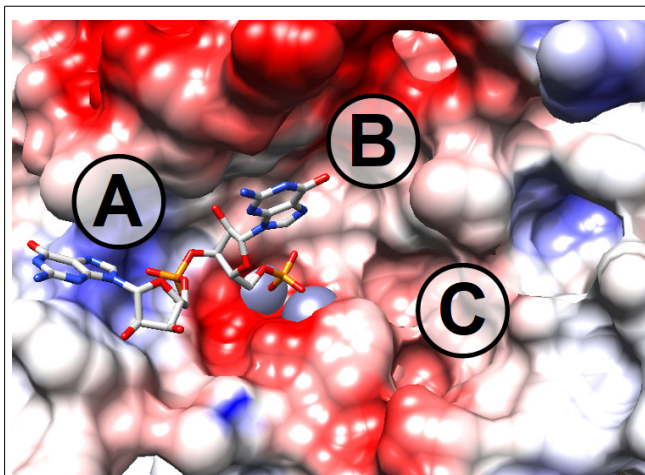


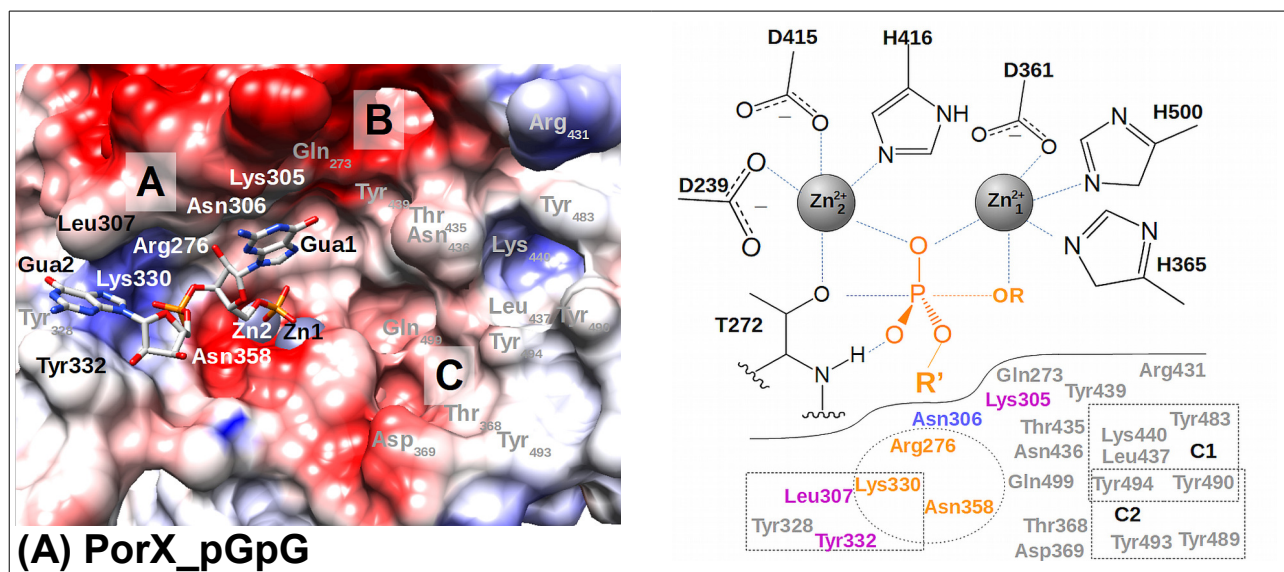
Figure 3.26: PorX_pGpG substrate binding pocket. Pocket A opens in PorX upon substrate binding, B is partially filled and C is a large subpocket able to fit in multiple nucleotides in size and could be further subdivided.

chain (see Figure 3.11D). B is partially occupied by the second guanosine in PorX_pGpG and could either harbor a larger substrate substructure or it could be used to bind or rotate a nucleosid moiety. C is a large empty cavity in our crystal structure that could even harbor multiple nucleotides. Also it allows rotational freedom of polynucleotide subunits (Figure 3.26), in a manner as it was proposed for the similar ENPP1 substrate binding pocket.^[23]

The large PorX substrate binding site with multiple subpocket differs significantly from those of homologs with small substrates as acetlyphosphate hydrolase PhoA (Figure 7.6_(1)), BT-159 sulfatase (Figure 7.6_(13)), phosphoglycerate mutase GpmI (Figure 7.6_(15)) or the proposed phosphoethanolamine transferase MRC1 (Figure 7.6_(20)). The alkaline phosphatase PafA, which dephosphorylates mononucleotides (NTPs to NDPs) (Figure 7.6_(11)) shows similarities only around the phosphate binding site and, in contrast to PorX, the substrate binding site is electropositively charged due to the presence of multiple lysines. All the remaining homologs are PDEs and show more similarities to the substrate cavity of PorX especially in pocket A. Upon the least similar enzymes the choline specific glycerophosphodiester phosphodiesterase ENPP6 (Figure 7.6_(7)) and the alkaline sphingomyelinase ENPP7 (Figure 7.6_(6)) show a hydrophobic pocket at position A that hosts several tyrosines, the so called "cation box" which is a choline binding site. In contrast a similar binding mechanism at position of PorX pocket A is seen at the homologs catalyzing mononucleotids, namely the snake venom phosphodiesterase SV-PDE (Figure 7.6_(8)) and XAC PDE (Figure 7.6_(4)). On the surface of these proteins the nucleotide is hold in position

either between a leucine and tyrosine as in PorX (XAC) or between a phenylalanine and a tyrosine (SV-PDE). The cavity at positions of the PorX pockets B and C on the other hand is occupied by electropositive side chains, unsuitable for further nucleotide bindings, but optimized to attract polyphosphates. ENPP5 homolog position B has a highly electronegative charge to be able to bind its substrate NAD^+ and further has a binding site for a third zinc for $\text{NAD}(+)$ hydrolysis, absent in PorX, and no subcavity at position C (Figure 7.6_(2)). The only homologs where all three binding subpockets A, B and C of PorX have a counterpart in their published structures are ENPP1-4, which each is proven to bind and catalyze the hydrolysis of dinucleotides. ENPP2/autaxin (Figure 7.6_(10)) has the least homology and prefers phospholipids over nucleotides due to a larger hydrophobic pocket at position A. However it also has activity for Ap3A hydrolysis but was not crystallized with a substrate analog yet to compare. The same lack of binding residue characterization exists for ENPP4 (Figure 7.6_(3)) which hydrolyses Ap3A and Ap4A but not ATP. However for ENPP3 (Figure 7.6_(5)) and ENPP1 (Figure 7.6_(9)) there are dinucleotides cocrystallized and the binding mechanism discussed in the literature.

ENPP1 is relatively unspecific for NTPs, which it hydrolyses to NMP and pyrophosphate while the single nucleotide is bound to the homolog of pocket A. It is also shown to hydrolyze Ap3A, but the binding mechanism of this substrate is not characterized. Yet it was shown to hydrolyze 2'-3'-cGAMP to AMP and GMP and was cocrystallized with its reaction intermediate pApG. ENPP3 hydrolyses NTPs, UDP-Glucose, NAD^+ as well as Ap3A and Ap4A and could be crystallized with Ap4A. Taking a closer look at these two dinucleotide binding pockets (Figure 3.27) and comparing them with the PorX cavity for each pocket A, B and C, several features of substrate binding of PorX can be comparatively characterized.



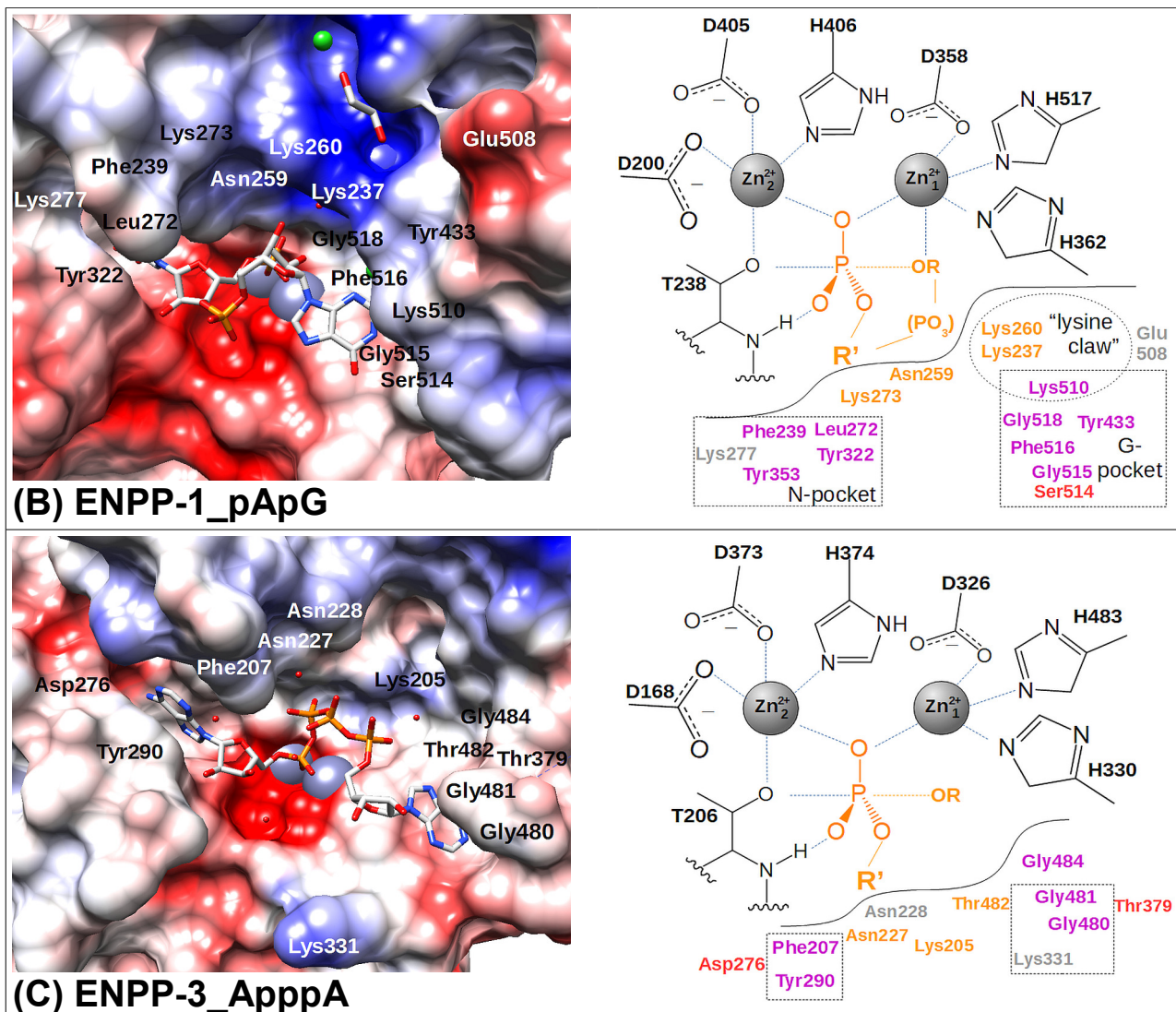


Figure 3.27: Substrate binding site of PorX (A) compared to available ligand bound structures of homologs, ENPP-1 (B) and ENPP-3 (C), as published^[9]. (Left column)

The substrate binding pockets are shown in coulombic colored Connolly surfaces. For better visibility the ionic radius of the zinc atoms was set to 1.9 Å. **(Right column):** Active site metal and substrate coordination are shown. Residues with hydrophobic contacts are colored magenta, residues stabilizing phosphates in orange, and hydrophilic residues in blue (electropositive) or red (electronegative). Residues in close proximity without direct interactions (from the structures themselves or as published in the literature) are shown in gray. Hydrophobic pockets are indicated as dotted rectangles, while additional phosphate binding sites outside the hydrolysis site are encircled likewise.

At pocket A, as for each nucleotide binding site of previous described homologs, the relative position of the hydrophobic clamp differs from ENPP-1 and ENPP-3. In these structures, their α -phosphate is positioned at the active site, whereas in PorX the nucleotide binding site is shifted further away from the catalytic center (see Figure 3.27A), creating space for an additional ribose subunit and a second phosphate. This suggests the leaving product of hydrolysis in this position would be a di-nucleotide (from a larger substrate at 5'). Also the 3'-hydroxy group of pGpG is pointing outside the substrate binding pocket and thereby there would not be clashes with an 3'

elongated nucleotide. The hydrophobic clamp between Tyr and Leu is shown to be unspecific for nucleotide types (as seen in the multiple structures in the PDB e.g. for ENPP1) and there are no hydrophilic side chains in the PorX hydrophobic pocket that could determine an obvious specificity for guanine over an adenine ring. Comparing pocket B between the homologs, there are no significant similarities between PorX and ENPP-1 or ENPP-3. Taking into consideration the shift in position of pocket A in PorX, this shift is functional plausible for ENPP-1 and ENPP-3. They cleave polyphosphates with NMP as leaving group so their α -phosphate needs to get located at the active site zincs and their polyphosphate chain needs to interact with an electropositive region present in those proteins. In PorX there is a highly electrophilic flexible loop region, which would prevent the binding of polyphosphates. Comparing the pocket C of PorX with ENPP-1 and ENPP-3 there is a dinucleotide binding site in both ENPP3 and ENPP1. In PorX two smaller subpockets within region C, C1 above H500 and C2 below F499, are found. The latter subpocket which is occupied by a diffuse electron density in the high resolution structure PorX-HR could be occupied by a loosely bound chloride anion that might fulfill a similar function as the chloride in the ENPP1 binding site, which stabilizes a nucleotide after binding at a similar position.

3.14.3 Structural comparisons to BREX PglZ orthologs

As our PorX structure is the first X-ray structure of a PglZ domain, which is ubiquitous among prokaryota,^[27] the question arose as to whether there are common structural features among PglZ domains. Apart from its role in type IX secretion systems as response regulators, PglZs are part of BREX (bacteriophage exclusion) genomic defense islands, including six families, BREX1 to -6, all with a unique but within its BREX system highly conserved PglZ protein. Therefore we compared PorX to their BREX consensus sequences for each system's PglZ (Figure 3.28). We predicted the three-dimensional structure of all six "BREX PglZ system phosphatases", as monomers and dimers (Figure 3.29A) by AlphaFold and used this structural information to adjust our primary sequence alignments. The final sequence alignment, homology quantified as percent identity matrix and phylogenetic dendrogram are shown in Figure 3.23A-D.

Each predicted model of BREX PglZ containing proteins contained an N-terminal globular domain, an elongated helical rich domain and a PglZ domain. The N-terminal domains have a similar size to PorX RD but lack noteworthy sequence or fold conservation to PorX. The elongated helical domain has no noteworthy sequence conservation, is much longer but in some BREX the last three helices of this domain are similar in fold to PorX HBD. Yet, the overall fold at the PglZ domains is highly conserved.

This includes the (structural) core residues, responsible for the overall fold of catalytic (CAT) and capping (CAP) subdomains. L269+P270+T271 (Ser in BREX1/5) is the center of a hydrophobic core that connects HBD H3 with $\alpha 3$ and the CAP subdomain at multiple positions (also at conserved G502), T414 (Ser in BREX2-4) interconnects $\beta 1$, -5, -6 and positions D415 and H416 close to Zn2, G417 and E507 connecting the CAP to the CAT subdomain, R438+G501+G502 position CAP on top of CAT and F481 (Tyr in BREX2) which builds the central residue in the hydrophobic core of CAP and P511 that links $\alpha 2$: $\beta 2$: $\beta 5$: $\beta 6$. Further all zinc binding residues are fully conserved: D239, T272, D361 D415, H416, H500 except H365 which is substituted by aspartate in BREX2 and glutamate in BREX5 and BREX6 (Figure 3.29D). Less conservation is at the residues that coordinate Gua2 in our pGpG structure: R276 (Lys in BREX1/4/5, Met in BREX6) and N358 (Asp in BREX-1/3/4, Glu in Brex6). $\alpha 1$ helix regions differs significantly due to connection to more divergent helical bundle domains and regions $\alpha 4$ and $\alpha 5b$ differ likely due to slightly different dimeric interfaces (see Figure 3.29A,C). Striking differences are found at the all regions determining the substrate binding pocket (PorX flexible loops $L\alpha 3b$ - $\alpha 3c$, αII + $L\alpha II\beta V$ (stabilizes $L\alpha 3b$ - $\alpha 3c$), $L\beta 1c$ - $\beta 2$ and αIII) and, outside the substrate binding pocket, there are two main differences in the CAP subdomains at the protein surface. First there is CAP(βIII + $\alpha 1$) which is solvent exposed and predicted as a well defined helix in BREX3 and second there is CAP αIII , unique for PorX, which extends outside the substrate binding pocket.

For BREX4/5 no dimers with a similar dimerization interface than PorX could be predicted, whereas BREX1-3 and BREX6 PglZ dimerization interfaces were predicted at the very same region as $\alpha 5a$ -d from the PorX PglZ dimerization interface but absent in other APS members, yet with large variations in contacts (see below). These BREX PglZ dimers were symmetric and similar to PorX in HH/HH conformation (Figure 3.29D).

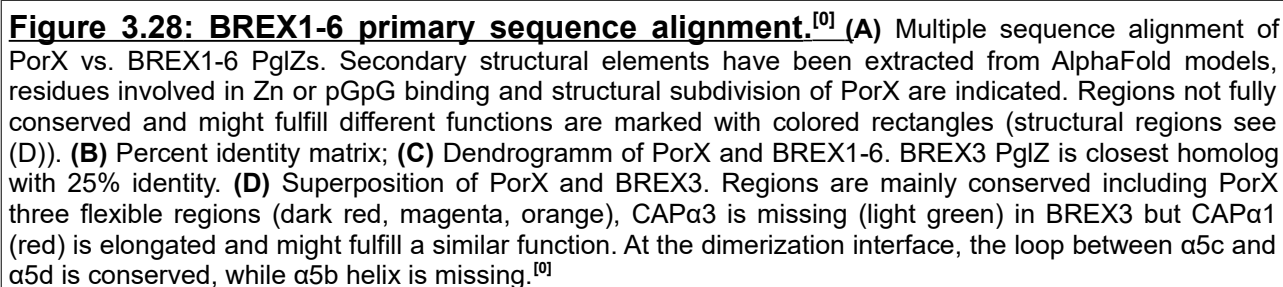
In PorX, helices $\alpha 5a$ -c build a hydrophobic core (F359, M363, I373, L376, L386, W390) and helices $\alpha 5a/b$ bind their counterpart on the other monomer $\alpha 5a'+b'$ (M362-K371', F350-E375'). $\alpha 5b/\alpha 5b'$ connect the hydrophobic cores of both chains in the center and orient $\alpha 5c/\alpha 5c'$ in an antiparallel direction perpendicular to the main axis of the domain. Helix $\alpha 5c$ interacts via polar S389-S385' and hydrophobic H393-A381' and L368-L368' contacts and bends towards $\alpha 5d$.

BREX1 dimerization interface is similar to PorX but there is no bend between $\alpha 5c/d$ ("helix $\alpha 5cd$ "). Also both $\alpha 5a$ contact each other via D552-H548', D552-D552' and build a hydrophobic core (between L549, L553, K556) with both $\alpha 5b/\alpha 5b'$ which bind to $\alpha 5cd$ via K556-E567' and R560-E563' hydrophilic interactions. $\alpha 5cd/\alpha 5cd'$ are too distant to contact each other directly.

BREX2, which has an Asp at H365 equivalent position of PorX (which coordinates Zn1), is missing helix $\alpha 5b$. Instead it harbors a long loop (D509-S519) connecting helices $\alpha 5a$ and $\alpha 5c$. This loop interconnects both chains with the opposite loop, as well as $\alpha 5a/c$ helices via D506-K510', D509-K510', E512-T524' and R512-D522'. In BREX2 the kink before $\alpha 5d$ is conserved but $\alpha 5c$ is predicted as only a single turn helix.

BREX3 dimeric model is the most similar to PorX but misses $\alpha 5b$ helix. Instead three residues I)-III) connect $\alpha 5a/c$: I) M363, which builds a hydrophobic core between helices $\alpha 5a/c$, including: K335, K355, I359, L376 and A378, II) N354, which connects to the opposite monomer's $\alpha 5a'$ K358' and I359' and III) L365, which connects to L376 and W377' from $\alpha 5c'$ that are part of the $\alpha 5a/c$ hydrophobic core. Helices $\alpha 5c$ and $\alpha 5c'$ which, interconnected by N372-N372' and L376-G367'/A368', are oriented antiparallel and bend into $\alpha 5d$ and $\alpha 5d'$ like in PorX, respectively.

BREX6, which has a glutamate in equivalent position of H365, has a strong terminal bent at helix $\alpha 5a$, similar as $\alpha 5b$ but no kink between -c/d. In the predicted model, BREX6 dimerizes with a strong hydrophobic core between both $\alpha 5 a/b/c$ (including A334, L339, A342, T343, I346 and F347 from both subunits), very similar to PorX and BREX1.



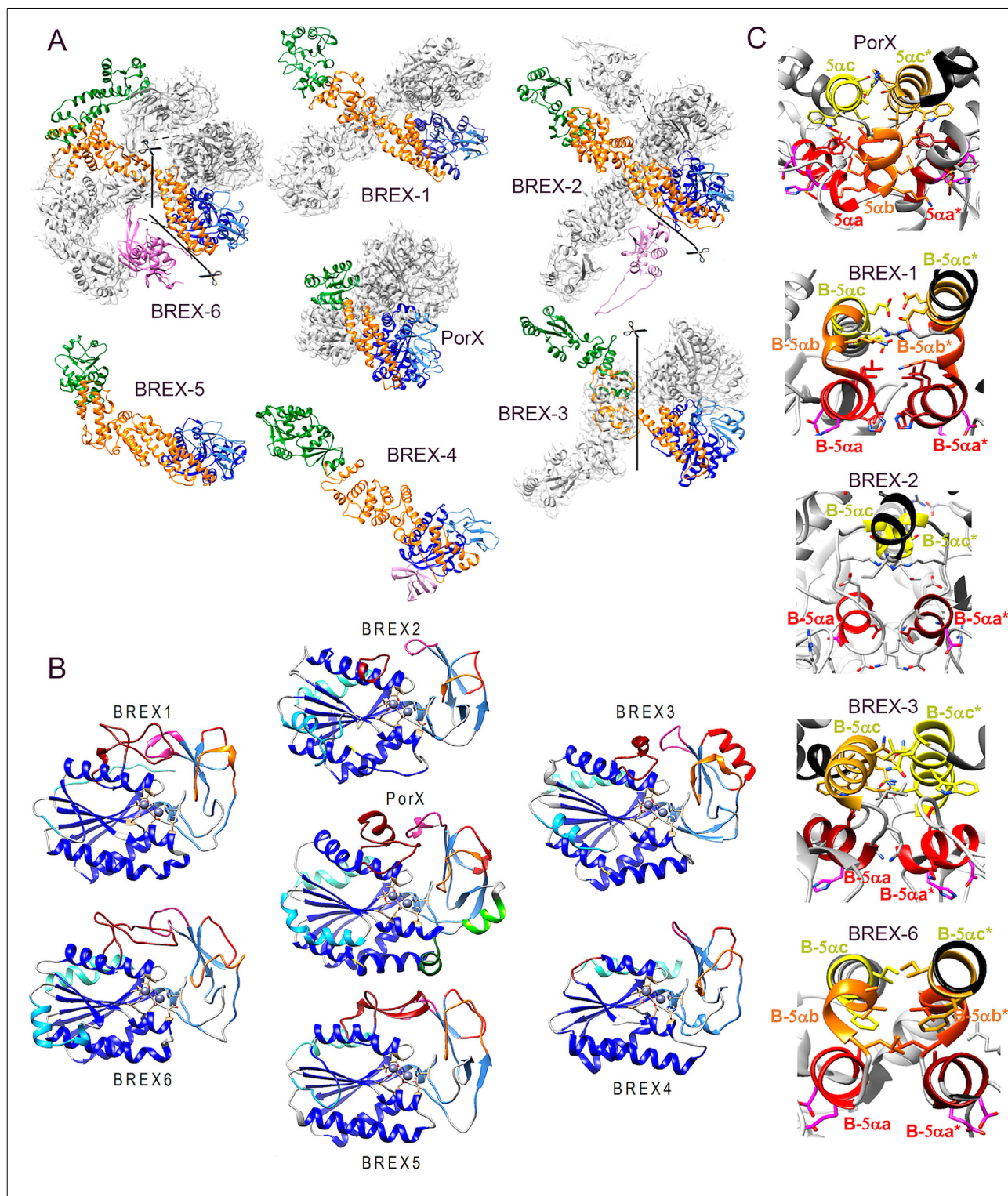


Figure 3.29: Structural alignments of BREX1-6 models.^[1] (A) Structural comparison of PorX and AlphaFold predicted full length dimeric models of BREX1-6 PglZs. PorX receiver domain or BREX N-terminal globular domains of similar size are colored green, the helical bundle domains, which are elongated in BREX PglZ proteins in orange, PglZ catalytic subdomains in blue, CAP subdomains in light blue and additional C-terminal domains in magenta. BREX4 and -5 were not predicted to build an interface at PorX homolog α5a-d region. Scissors indicate arbitrary domain positioning due to independent domain predictions. (B) Structural comparison between PorX and BREX PglZs, identical secondary structures are

colored blue (dark for catalytic subdomain, light for CAP) while divergent regions are colored differently, for primary sequence and PorX labels see (Figure 3.28A). Differences are at the PglZ N-terminal helix $\alpha 1$ where they are connected to different helical bundle domains (cyan), at helix $\alpha 4$ (sky blue) possible due to different dimeric contacts close to the dimeric interfaces; the flexible regions in PorX (L $\alpha 3b$ - $\alpha 3c$, dark red involved in pGpG bonding; CAP $\alpha 1$ +L $\alpha 2$ - βV , magenta; L $\beta 1c$ - βII , orange) which cap the active site or stabilize (CAP $\alpha 2$ +loop) the substrate binding loop (L $\alpha 3b$ - $\alpha 3c$), show high variations, presumably due to divergent substrate binding. Further differences are at helices CAP $\alpha 1$ (red) and CAP $\alpha 3$ (light green) which might be involved in protein-protein/DNA interactions in PorX and BREX3. **(C)** BREX1-6 dimeric Interfaces of AlphaFold models. Color scheme differs from panels (A/B): red depicts PorX $\alpha 5a$ helix and equivalent region in BREX proteins, orange is used for $\alpha 5b$, yellow for $\alpha 5c$ and dark gray for $\alpha 5d$ helices. For BREX4 and 5 no homolog dimerization interface to PorX was predicted.^[9]

3.15 Enzymatic assays

Testing enzymatic activities and assays of PorX and its mutants were performed by our collaborator Dr. Mariusz Madej from Jan Potempa's group (Małopolska Center of Biotechnology, Jagiellonian University, Krakow, Poland; HPLC experiments were performed at Virginijus Siksnys group at Institute of Biotechnology, Vilnius University, Lithuania). Because our comparisons, mutants, suggestions of substrates and enzymatic function analysis need to be put into context with the results of our collaborators, I will describe shortly their results here:

3.15.1 Phosphorylation assays

To study the effects of mutations (TABLE 3.14) on dimerization and activity, apart from further activity assays (Chapter 3.15.2), our collaborators performed ^{32}P -labeled phosphorylation assays of PorX with radioactive labeled acetyl phosphate (^{32}P -AcP, phosphorylation at RD) as well as labeled ^{32}P - γ -ATP (autophosphorylation at PglZ). The results from these experiments (Figure 7.7) proved phosphorylation at D58 (no phosphorylation at D58A) as well as the unspecific phosphorylation of T272 by ^{32}P - γ -ATP (no phosphorylation at T272A). The dimerization interface mutants between RDs (M94K/D104A/I134A mutant) or the PglZ (S385E/S389E or S385E/S389E/D361A/H365A) not only inhibited the phosphorylation of their own domain, but both domain phosphorylations were affected vice versa. Mutating the binding site of Zn²⁺ (T272A), which is not necessary for neither RD or PglZ dimerization, had a low effect (ca. -30%) on RD phosphorylation. Likewise, the D361A/H365A mutant, which destabilizes Zn²⁺ binding and thus HH conformation, inhibited (ca. -65%) but not prevented RD phosphorylation. Transient phosphorylation of T272 by ^{32}P - γ -ATP, which is an indicator of enzymatic activity, only occurred at the wt protein and was inhibited by all active site or interface mutations of PorX.

3.15.2 Substrate cleavage assays

Initial enzymatic assays for PorX were standard phosphatase assays with p-nitrophenyl phosphate as substrate, which were negative for PorX monomer as well as phosphorylated dimer (Mariusz Madej, unpublished data). On the opposite, the standard phosphodiesterase (PDE) assay with bis-4-nitrophenyl-phosphate, a synthetic substrate with a phosphodiester bond, showed phosphodiesterase activity of PorX, which led us to our cocrystallization and soaking trials with pGpG and c-diGMP (as putative natural substrates). Activity assays were performed with a series of metal ions, showing initially activity in the presence of 3 mM Zn or Mn (but not Mg, Ca, K Ni, unpublished data), another reason we performed the SAD experiments on PorX crystals for metal identification. After optimizations of buffers and conditions PorX showed activity only at 0.5 mM Zn (but not Cu, Mn, Mg, Ca; see Figure 3.27 A-D).

The activity of PorX dimers as well as monomers against bis-4-nitrophenyl-phosphate, was probably due to the relative small size and unspecific PDE binding of this substrate, which leads to very efficient cleavage by even small amounts of active enzyme. These small amounts of dimeric PorX in the presence of zinc, even without phosphorylation, are in agreement with the zinc induced PorX dimerization behavior we observed during SEC-MALLS (Figure 3.16) and SAXS (TABLE 3.15) experiments. In contrast, incubation of monomeric or dimeric PorX each with pGpG and manual analysis of cleavage products by HPLC (data not shown due to redundancy with Chapter 3.15.3, see below) showed cleavage of pGpG by the PorX dimer but not for the monomer, identifying PorX dimer as the active enzyme. For the same reason comparative assays (Figure 3.27 E-F) with phosphorylated and unphosphorylated PorX and its mutants (TABLE 3.14) only showed a slight increase (ca. 40%) of the turnover rate of PorX by phosphorylation. For the unphosphorylated PorX, activity was constant when residues at the RD not involved in zinc induced dimerization were mutated (D58A, M94K/D104A/I134A). In contrast activity was abolished by mutations involving the PglZ dimerization interface or the active site Thr272, stabilized by zinc (D361A/H365A, S385E/S389E, D361A/H365A/S385E/S389E, T272A). For phosphorylated PorX we saw a slight decrease involving residues necessary for RD phosphorylation (D58A, M94K/D104A/I134A), likely because unphysiological dimers induced by Zn (see SAXS results TABLE 3.16, 3.13.1,) remained active, while (like for the unphosphorylated protein) mutations involving the PglZ dimerization interface or the active site Thr272 abolished PorX activity.

3.15.3 Ultrahigh performance liquid chromatography mass spectrometry (UHPLC-MS)

Following the initial HPLC results for pGpG hydrolysis (performed by Mariusz Madej) it was decided to screen a full nucleoside library at Virginijus Siksnys group (Institute of Biotechnology, Vilnius University, Lithuania) and identification of products by UHPLC-MS ("ultrahigh-performance liquid chromatography-mass spectrometry", data evaluation by Mariusz Madej). The screening included non- and cyclic mono- or dinucleotides, cyclic oligonucleotide as well as polyphosphates (mono- or dinucleic) in particular AMP, ADP, ATP, GMP, GDP, GTP, c-di-GMP, c-di-AMP, 2',3'-cGAMP, 3',3'-cGAMP, c-AMP, c-GMP, cA3, cA4, cA6, pApA, pApG, pGpG, Ap3A, Ap4A and Ap4G.

PorX showed only hydrolytic activity at linear and cyclic oligonucleotides (pApA, pApG, pGpG, cA3, cA4 and cA6) as well as further hydrolysis of the intermediate hydrolysis products of oligonucleotides, resulting in NMPs as final PorX enzymatic products (Figure 7.9). There was no activity observed for nucleotide polyphosphates, cyclic dinucleotides or cyclic or linear mononucleotides tested (Figure 7.10). This shows high specificity for PorX regarding phosphodiester type bond, but no specificity against the nucleoside type, as we observed equivalent cleavage for adenosine and guanosine dinucleotides.

Discussion

NOTES.....

4. Discussion

4.1 Crystallization

4.1.1 Initial screenings

Result summary: In this study we used 480 sparse matrix conditions with four types of PorX samples. We observed the best crystallization behavior from selenomethionine derivative PorX with BeF₃. *In vitro* phosphorylated PorX dimer yielded more clustering and monomeric PorX samples stayed soluble. This initial screening identified buffer compounds and characteristics aiding PorX in crystallization and set our focus on PorX dimer structure. We chose selenomethionine derivative PorX because it showed lower solubility and thus less sample protein was necessary for screenings and second because initial MR trials failed and SAD phasing became aim of crystal optimization.

Contextualized discussion: Monomeric PorX remained soluble likely due to disordered regions at the dimerization interface, the flexible loops and multiple CAP subdomain positions, as we observed in our SAXS studies on PorX monomer (Chapter 3.13.5). BeF₃ proved to be a good compound as dimerization agent by mimicking phosphorylation at the receiver domain (RD). It improved crystal quality compared to *in vitro* phosphorylated PorX which decreased over time by autodephosphorylation. We showed that PorX autodephosphorylation has a half time period of ca. 2 h (Figure 7.5). Autodephosphorylation is typical for RDs of response regulators^{[3], [4]}, and thereby monodispersity of Dim_PorX samples (*in vitro* phosphorylated PorX by AcP) decreases and we used BeF₃ to form reversible ionic-interactions with PorX when monodispersity was necessary. Due to the high molar excess of BeF₃ during crystallization PorX was fully dimerized which led to a better crystallization. We validated this monodispersity of PorX with BeF₃ by our SAXS experiments (Chapter 3.13.1). The polydispersity of AcP *in vitro* phosphorylated PorX might have been the reason why previous crystallization attempts preceding this work did not yield high quality crystals.

Evaluation: Retrospective we could see why the MR trials failed, as the models we used were only homologs for the receiver domains (with up to 30% sequence identity). For the HBD or PglZ neither structural searches by sequence similarity (NCBI_blast, PDB) nor automatic homology searches (implemented in ITASSER or Robetta) could find or create suitable homology models e.g. based on alkaline phosphatase superfamily proteins. This might be due to the high variance within the APS family, multiple insertions and deletions and circular permutations. Considering the

extensive homology comparisons within this study, it is likely that we could have had more success using APS family proteins as MR models (especially phosphodiesterases). However we progressed fast in crystallization optimizations of selenomethionine derivative, BeF₃ dimerized, PorX. Solving the phase problem by SAD also had the benefit of not introducing model bias which would have delayed model building.

4.1.2 Buffer optimizations

Result summary: Around 5.400 conditions have been screened until the first single crystals were produced yielding a complete datasets (JG5 data) and a large crystal cluster (HB6 data) with sufficient anomalous signal we used for SAD phasing. Subsequent JG5 data could be solved by MR-SAD and we build the PorX dimeric model at 2.4 Å. Large breakable crystal clusters from larger drop sizes (0.5-10 µl) like HB6 were not reproducible and, in order to get single crystals that were not clustered, protein concentration had to be lowered to 1.3-2.5 mg/ml in nanodrops with multiple replicas and crystals fished and diffracted had to be small, before clustering started. After solving the first structure and finding zinc at the active site we added low amounts of zinc to all crystallization conditions and thereby decreased crystal clustering to an extent.

Contextualized discussion: One reason for the clustering behavior of crystals could have been the lack of sufficient zinc in the buffer to generate stable dimers. PorX molecules with different zinc occupancies might lead to low homogeneity which resulted in irregular patterns of crystal growth. This would also explain, how it was possible that PorX dimers could be crystallized with more than two zincs without adding zinc to PorX purification, or the crystallization buffers (in case of JG5 structure, LH/HH conformation, three zincs per dimer). Because our SAXS data suggest disorder at helices α5a-c it is likely PorX got only copurified with one zinc at Zn2 position per monomer. Thus it is likely that the third zinc of dimeric PorX (at Zn1 position, JG5) was taken from the crystallization solution. There zinc might be in equilibrium between free ions and bound to Zn2 position of PorX monomers. During crystal growth of dimeric PorX, which incorporates additional zinc, the solution and monomers of PorX would be extracted of zinc ions over time. This leads to decreasing zinc concentration during crystal growth and conformational variation over time of PorX. Such changes could be disorder at the PglZ dimerization interface of monomers, which incorporated into dimeric PorX crystals result in mixed conformations, crystal defects and disturbed crystal growth. This could be the reason for the increased clustering and explains well how we successfully decreased clustering by zinc addition.

Evaluation: We could solve the phase problem for PorX by SAD and build a PorX model for to 2.4 Å data. The compounds aiding PorX crystallization, as well as further zinc addition were an ideal starting point for further crystallization strategies. However only 5% of crystals grown in those conditions (of equivalent size and shape) yield a resolution of 2.5 Å or better.

4.1.3 Strategy optimizations

Result summary: By optimizing the PorX crystallization strategy we yielded high resolution data of PorX at 1.9 Å, as well as substrate bound PorX T272A mutant with pGpG bound to the active site. Other structures of PorX we obtained in the process of nucleotide screenings and crystallization trials of mutants. Further soaking experiments with nucleotides were performed but unsuccessful.

Contextualized discussion: We added the ions (BeF₃ and Mg) necessary for PorX dimerization into the crystallization buffers and by this we were able to concentrate PorX as the more stable monomer. Doing so we were able to disperse up to six times higher amounts of PorX per crystallization drop, resulting in larger crystals. The high amounts precipitated large amounts of PorX immediately but precipitation was reversible and thus non-denaturing. Because the initial protein concentration was lower, fewer nucleation events occurred and those nuclei grew feeding on precipitate, supplying a stable protein concentration over time. This stable conditions finally avoided crystal clustering and large single crystals were reproducibly produced. These were growing 5-6 weeks, up to one mm in size (instead of previously 30 µm max. within a few days), and diffraction was greatly improved. Soakings with nucleotides (except for GMP soaking in case of our HR structure without binding), lead to destabilization of the crystal packaging, as observed by smeared diffraction patterns. Those soakings with ATP, GTP and c-di-GMP and pGpG indicate some sort of interaction with the proteins even if they are not stably bound to the active site (as seen during cocrystallization with GTPγS or c-di-GMP), which also explains radioactive labeling with ATP (Chapter 3.15.1).

Evaluation: By these optimizations we provided a method and buffer conditions (Opt_J as wide grid- and Opt_L as fine grid screen) to reproducibly yield high quality PorX crystals (2.4 Å or better for most larger crystals) for future experiments, with or without substrates, substrate analogs/inhibitors and/or other mutations. If those conditions will yield to cocrystallization with PorY or sigma factors remains to be tested and might be the continuation of our work.

4.1.4 PorY, PG1660 and PorX monomer

Result summary: Crystallization trials of PorY (N-terminal truncated cytoplasmatic domain, two constructs, ca. 1300 conditions) and PG1660 (ca. 500 conditions) were only performed as side projects. PG1660 remained stable in solution and yielded phase separation events during crystallization trials. The PorY construct (C-terminal cytoplasmatic domain) was successfully optimized for higher solubility, based on domain borders we suggested, but still needed to be stabilized and tend to precipitate. Monomeric PorX was highly soluble (up to at least 25 mg/ml tested on short term, or over a year at concentrations below 1 mg/ml as storage condition). Crystallization trials of monomeric PorX (around 2200 conditions tested, up to 14 mg/ml) resulted in a significant number of phase separation events.

Contextualized discussion: Monomeric PorX, as we saw from our SAXS results, harbors flexible regions in solution (flexible loops, dimerization interface, see Chapter 3.13.5). These are likely the reason for the high solubility of PorX monomers that prevented crystallization. During PorY crystallization trials the PorY N-terminus was uncertain and with a the first optimization we heighten its solubility. Still PorY had a tendency to precipitate, needed stabilization by ADP and was not prone to crystallization. It might be possible that the C-terminus of PorY is unstable without interaction to PorX, monomers or dimers, or interaction with PorL which binds PorX at the plasma membrane where PorX and PorY interact^[5]. PG1660 was highly soluble and tended to yield phase separation during crystallization attempts. This indicates a stable soluble construct that might harbor disordered regions preventing crystal contacts. Those might be due to a lack of interaction partner, which might be the PorX dimer or DNA or both in order to stabilize PG1660 for crystallization.

Evaluation: Our work at PorX structure solution progressed rapidly, therefore these side projects got lesser attention over time and many strategies for structure solution of PorX monomer, PorY and PG1660 could still be applied.

4.2 PorX structural features

4.2.1 Receiver domain RD

Result summary: The PorX receiver domain (RD) bound to BeF₃ has the canonical^[4] $\alpha_5\beta_5$ Rossmann-like fold structure of two-component system response regulator (RR) RD. ^[4]At the C-terminal end of the central β -sheet, the canonical phosphorylation site is conserved at D58. Furthermore the so called TY pair, T86 and Y105 in PorX, is conserved, and is in the activated conformation. ^[4]

Contextualized discussion: The conserved D58 phosphorylation site is well known to be activated by either autophosphorylation with AcP^[6] or by ionic interaction with BeF₃^[7] as phosphate mimic, both strategies we were using during our research. We used BeF₃ in cases we needed a stable concentration of activated dimeric PorX. We didn't use BeF₃ in enzymatic assays because it also mimics phosphorylation at T272, which competes with substrate binding and thus inhibits competitively the protein's enzymatic activity. In those cases, as well as during liquid chromatography, we *in vitro* phosphorylated PorX by acetyl phosphate (AcP) that binds to D58 with a covalent bond, but not to T272 (Figure 7.7). Thus we did not inhibit the protein, and the (covalently bound) phosphate moiety could not be diluted by buffer exchanges. However we observed autodephosphorylation, which is characteristic for most RRs^{[3], [4]}. In case of PorX we determined a autodephosphorylation half life period of two hours (Figure 7.5). This leads to a steady decrease of AcP phosphorylated D58 until all AcP reacted, newly bound phosphorylations get hydrolyzed and phosphorylation/PorX dimerization and activation turns zero. We have proven that PorX phosphorylation indeed dimerizes and activates the enzyme, by ³²P-radiolabeling (phosphorylation Chapter 3.15.1, Figure 7.7), SEC/-MALLS (dimerization, Chapter 3.12 Figure 3.16), SAXS (dimerization, conformational changes, Chapter 3.13) and enzymatic assays (activity, Chapter 3.15.2, Figure 7.8).

The conserved phosphorylation site and TY pair classifies the RD of PorX belonging to the *OmpR* subfamily of RDs. In *OmpR* RDs the conformational change of this TY pair is responsible for the exposing the α_4 - β_5 - α_5 interface resulting in dimerization and activation, such it should be in PorX aswell.^[4] In the activated conformation, analog our PorX structures, the Y105 (in PorX) side chain points inward the active site and is stabilized by interaction with T86 (in PorX) and T86 is hold in position by binding to either phosphate or BeF₃. In the unphosphorylated/inactive form, the tyrosine would point outside and thereby blocks the dimerization interface. However, according to NMR studies, there is an equilibrium between both TY orientations in the *OmpR* family^[8]. In multidomain proteins the inactive state usually gets stabilized by interactions with the effector domain,^[9] while the active state gets fixed by phosphorylation^{[4], [10], [11]}. This equilibrium in PorX can be seen in our SEC-MALLS results (at Figure 3.16A), where there is a small dimeric fraction (2%, TABLE 7.3) of unphosphorylated wt-PorX dimer. As the active conformation without phosphorylation exist in equilibrium, *OmpR* RD phosphorylation for some RDs requires or is fasten up by an active conformation of its effector domain. This stabilizes the RD active state due to interactions with the RD, as has been shown e.g. for DNA binding by DBD effector domains.^[12] This effect we showed

when an inhibition of D58 phosphorylation was caused by mutations at the PglZ dimerization interface/active site (Chapter 3.15.1, Figure 7.7).

Evaluation: A two-component system response regulator with a PglZ effector domain has not been characterized before. Here we showed that PorX features a RD of the *OmpR* subfamily, including canonical structure and regulation of this class. Further we could show, by SEC/-MALL and phosphorylation assays, that indeed PorX RD in solution is in an equilibrium between active and inactive form. This equilibrium get fixed in active state by phosphorylation and stabilized by an active state of the effector domain, as it is typical for the *OmpR* family. It also is destabilized by the inactive state of the PglZ effector, which had been observed for other effector domains for the *OmpR* subfamily before.

4.2.2 Helical bundle domain (HBD)

Result summary: The HBD of PorX consists of three helices (H1-H3) and represents a flexible linking region between RD and PglZ domains. The hydrophobic core between H2 and H3 sits rigid on top of the PglZ and creates a platform for helix H1 that can act as an flexible arm switching between two alternative positions, one for monomeric PorX, one for the dimer. Therefore helix H1 has two flexible bends in it acting as hinges, which switch the relative position between RD and helices H2/H3/PglZ (see SAXS Figure 3.23C/D, I/), followed by a change in PorX activity (Chapter 3.15).

Contextualized discussion: Our data suggests that during transitioning from monomeric to dimeric state at either side of the protein (RD or PglZ) the helical bundle transfers a stabilizing effect on the other domain, so one monomeric or dimeric terminal domain stabilizes the monomeric or dimeric state of the other respectively. Thereby helix H1 acts as a flexible arm exchanging contacts between RD and helices H2 and H3 of the HBD which are linked to the PglZ.

Evaluation: We could not find any homolog structure among the PDB, therefore this functional domain is the first of its kind to be characterized. It might help to identify the function of homolog domains as one with dimerization transfer function, as in PorX.

4.2.3 PglZ domain

Result summary: The PglZ domain of PorX has the characteristic α/β -fold of the alkaline phosphatase superfamily (APS), a central β -sheet surrounded by almost parallel helices except from the commonly 8 β -strands,^[13] the sheet in PorX only consists out of six strands. Apart from the

metal binding residues there are no specific residues conserved within the APS superfamily and secondary structure elements aligning in the 3D structure might be found elsewhere in the primary sequence when comparing different APS homologs. The active center of the PglZ can take two alternative conformations helix-helix (HH, $\alpha 5a+b$) and loop-helix (LH, $L\beta 4'\alpha 5b'+\alpha 5b^*$) conformation, of which only HH conformation has the canonical metal binding site of APS phosphodiesterases^{[14], [1]} and is sterical necessary for dimerization. The active center of PorX in HH conformation shows two zinc ions coordinated by six residues that are found in identical arrangements in the active sites of alkaline phosphatases, phosphonoacetate hydrolases or nucleotide pyrophosphatases, whereas the seventh residue threonine (T272 in PorX) is characteristic for phosphodiesterases.^[1] This active sites have promiscuous activity throughout the superfamily^{[15], [16]} and substrate affinities are regulated by the substrate binding pockets. The active site is capped by a CAP^[17] subdomain, in its fold, a typical auxiliary domain of phosphodiesterases and evolutionary close alkaline phosphatases with bi-metallo core.^[18] The CAP subdomain consist of two small β -sheets surrounding a hydrophobic core, flanked by loops and a few helical regions. Of usually four antiparrallel β -strands forming the larger, more conserved sheet, PorX CAP subdomain only consists out of 3 β -strands. The sequence which builds the forth strand in homologs builds a flexible, partially disordered loop ($L\beta Ic-\beta II$) capping the active site. Further PorX has a hydrophobic clamp which opens upon ligand binding between Y332 and L307. A hydrophobic clamp can also be found in homologs, however in PorX it is displaced such that the 5' phosphate between a di-/oligonucleotide gets bound, rather than an α -phosphate of a polyphosphate in homologs like ENPP1 (see Figure 3.27). Aside from the hydrophobic clamp there are four additional regions determining the surface of the binding pocket. 1) First there is CAP $\alpha 3$, a helix of eight residues containing five tyrosines, without any homology upon the PDB or even within the BREX PglZ sequences and in our structures it is not involved in ligand binding. 2) Second there is $L\alpha 3b-\alpha 3c$ that binds to the active site nucleotide and positions its 3'phosphate between Zn1 and Zn2 and interacts with loop $L\alpha II-\beta V$. 3) Third the loop CAP $L\beta Ic-\beta II$ mentioned before capping the active site, which does not bind to pGpG in our structure; and 4) forth helix $\alpha 5a$ which binds the second zinc (Zn1) and is an extended loop $L\beta 5'-\alpha 5b^*$ in LH conformation.

Contextualized discussion: The flexible loops around the substrate binding pocket and helix αIII are not conserved between PorX and any homolog which indicates a different mode of substrate binding not previously reported. Features not described before are $\alpha 5a$ as part a dimerization interface, in which either $\alpha 5$ or $L\beta 5-\alpha 5b^*$ builds a hydrophobic core with $\alpha 5b/c$, $\alpha 5b$ binds $\alpha 5b'$ (or

$\alpha 5^*$ in LH) and $\alpha 5c$ and $\alpha 5c'$ bind each other. Also the folding of $\alpha 5a$ during activation, as this region seems disordered in the soluble monomer (Figure 3.23) has not been observed in other APS. Comparing monomeric and dimeric structure apart from the dimerization interface we also observed a remodeling of the CAP subdomain. In the SAXS dimer in solution the CAP takes two alternative positions, one identical to the crystal structure, where it contacts and stabilizes PorX dimerization interface at $\alpha 5b$, and apparently in a second position it leaves the active site more open (Figure 3.22). An open position likely supports easier substrate entry into the active site. In the monomer the CAP domain has a third position, which closes the active site (Figure 3.23) and thereby probably prevents substrate entry.

Aside from the overall fold and the conserved metal coordinating center of PorX, the PglZ domain is very diverse from other proteins of the APS but definitively belongs structural to the phosphodiesterase subfamily (Figure 3.24). The different conformations of PorX, HH and LH conformation, do not appear in any other protein of this family, however LH conformation might be an crystallization artifact and not be physiological relevant. Also if the metal binding is assisted by helix $\alpha 5a$ or helix formation by metal binding might be not two distinct consecutive steps but each stabilize or each.

The surface at the PorX active site is rather unique and larger than that of any known homolog structure, still most of PorX substrate binding pocket is unused in our PorX_pGpG structure. The functions of $\text{LaII-}\beta\text{V}$ and αIII remain elusive but it is tempting to speculate that $\text{LaII-}\beta\text{V}$ is involved in binding a nucleotide 3' the active site of an oligonucleotide and that it could take in that fourth β -strand conformation that we see in homolog phosphodiesterases. αIII helix could participate in nucleotide binding as well, as it forms two hydrophobic pockets in the substrate binding pocket area C, which is unused by pGpG in our structure (see Pockets C1&C2 Figure 3.27).

Evaluation: The PglZ domain is part of a superfamily of homolog domains (APS), but this superfamily divided very early in evolution and thus the PglZ has very low sequence identity with other enzymes. This explains why PglZ has a different amount of β -strands and lots of strands and helices are found elsewhere in the primary sequence due to circular permutations. Substrate binding motives are not conserved with any found homolog, the nucleotide binding site ("hydrophobic clamp) 5' the active site metals is shifted in such a manner that NTPs do not get hydrolysed (as e.g. in ENPP1) and the cavity is much larger than in known related structures. Further PorX is activated by dimerization at its $\alpha 5a$ -c dimerization interface, which is not found in any APS protein within the PDB. However as our BREX-PglZ models suggest, at least BREX1-3 and -6 dimerize at exact the

same surface, which makes it a PglZ specific feature and our PorX studies are the the first characterization of this feature of PglZ activated by dimerization.

4.3 Activation by interdomain crosstalk

Result summary: PorX is activated by phosphorylation which leads to subsequent RD dimerization, conformational change of PorX HBD H1 at two separate hinges/flexible kinks and dimerization of PglZ domains with repositioning of CAP subdomains. There is an interdomain crosstalk, which can be seen in stabilization of inactive state RD (that cannot be phosphorylated by AcP), if the active site/dimerization interface of PorX PglZ is mutated. Likewise preventing the active state of the RD prevents the formation of an enzymatic active PglZ (shown by phosphorylation assays (Figure 7.7), SEC-MALLS (Figure 3.16A/C/E) and SEC, (Figure 7.3) and enzymatic activities (Figure 7.8F)).

Contextualized discussion: PorX harbors a RD of the *OmpR* subfamily of response regulators, which in solution are in an equilibrium between active (dimeric) and inactive (monomeric) states^{[8], [11], [10]} Phosphorylation shifts the equilibrium to the active state conformation whereas inactive effector domains can stabilize the inactive state of the RD while activated effector domains the active state^{[9], [12]} We can see exactly this traits in PorX. SEC-MALLS of wt PorX showed a fraction of 2 % dimers, conclusive with an monomer-dimer equilibrium with a majority in inactive monomeric state without phosphorylation. Phosphorylation at the RD was also inhibited by mutants at PglZ active site and dimerization interface and transient radioactive labeling of the PglZ (a characteristic of the canonical bi-metallo core of APS, which indicates enzymatic activity)^{[17], [19], [20]}, was inhibited by RD mutations, which in turn inhibit PorX dimerization (Figure 3.16A/C/E, Figure 7.3) and decreases enzymatic activity (Chapter 3.15.2, Figure 7.8F).

This crosstalk between RD and PglZ domain is accomplished by the HBD and an exchange of interacting surfaces between monomeric and dimeric PorX. These either stabilize or destabilize one of those states and changes at one surface gets transmitted to the other. While changes between active and inactive state of RDs are well characterized,^[4] the role of the HBD and PglZ has to be discussed. Based on the surfaces of our dimeric crystal structure and the monomeric SAXS model, as well as comparing active and inactive state of the RD homolog PhoB (with BeF₃ (PDB: 1zes^[21]) and without BeF₃ (PDB:1B00^[22])), we propose the following conformational changes between PorX dimer and monomer, from the RD to the PglZ: In the active dimer T86 stabilizes Y105 and Y105 side chain points inwards the RD (conserved *OmpR* TY pair, stabilized by phosphorylation). There, Y105 takes part at the hydrophobic core of the RD dimerization interface with its own helix α 4 and

opposite RD helix $\alpha 5'$. In the inactive state Y105 would point outwards, as in homologs, and thereby blocks the dimerization interface, in PorX at M94 and I98 (on helix $\alpha 4$), which before contacted helix $\alpha 5'$. This new contact of helix $\alpha 4$ might lead to small rearrangements that disturb the contact with the opposites HBD helix H1 ($\alpha 4$:I98&G99 to HBD':I128'&I129'). This allows now the RD to rotate around the hinge between RD and HBD helix H1, around 100° . Thereby Y79, which before contacted the opposite PglZ' is now in proximity with (and likely contacts) the hydrophobic residues I128 and I129 of HBD helix H1, which before contacted the opposites RD'. Helix H1 loosing this contacts to RD' might destabilize the contacts to HBD helices H2 and H3 (between H2 R162 and helix H3 R191). This H2H3 region forms new contacts to its own RD $\alpha 3$ (aa66-76) and $\alpha 4$ (91-99), as we see in the SAXS model, while H1 bends at the second hinge around 60° . This region of H2-H3 is in direct contact with the PglZ dimerization interface (at PglZ $\alpha 5c'$, H2-PglZ' : E176-E385'; I174&E177-R389'; E177- A386') and the new contacts of H2H3 region to its own RD likely destabilize the interactions to PglZ' of the opposite chain. Further H3 as well as the PglZ dimerization interface contact and stabilize the CAP subdomain in its open position and changes in region H2H3 likely are responsible for CAP repositioning into the closed position we observe in the monomer. This, in combination with disorder of the dimerization interface, lets loose its contacts to H2H3 region and inactivates PorX. This surface exchange described above works analog in the other direction to activate PorX. The interactions of HBD H1 with H2H3 region upon dimerization of two molecules PorX at the RD brings the PglZ into close contact which facilitates PglZ dimerization, zinc binding at Zn1, opening of the of the substrate binding site by the CAP subdomain and leads to enzymatic activation.

Evaluation: Our structure confirms that PorX is a canonical response regulator that gets activated by RD phosphorylation with subsequent phosphorylation and activation of its output/effector domain. Between PorX RD and PglZ activation there is an interdomain crosstalk via HBD H1 reorientation that repositions RD and PglZ, which active/inactive states positively correlate each other, as observed in other response regulators. The here proposed mechanism of the crosstalk, we observed between domains, is reasonable, considering we can identify the surfaces in both crystal structure and SAXS model of the monomer. However we can only be certain of exact residues and contacts of our higher resolution crystal structures in the dimer, while specific residues mentioned above concerning the monomer remain speculative until further studies.

4.4 Reaction pathway

Result summary: The PorX two metal binding site in HH conformation is identical to the canonical coordination shell of bi-metallo APS as published by Sunden et al. (2017)^[1]. All APS enzymes with that reactive site have the same reaction pathway, only divergent in kinetics and with different substrate affinities (Figure 3.15A). Further PorX has threonine as active site nucleophile, the typical phosphodiesterase fold^[1] and we confirmed PorX as a specific phosphodiesterase by homology and enzymatic activity (Chapters 3.14&3.15).

Contextualized discussion: Therefore the reaction pathway for PorX catalyzed hydrolysis is analog to other APS phosphomono-/diester-/transferases reactions as shown in (Figure 3.25). In a first reaction step, PorX with its catalytic site alkoxide nucleophile T272 attacks and bind covalently to the phosphate, which phosphodiester bond will get cleaved, with its O^{γ1}, stabilized by its backbone nitrogen. The phosphorus, previously tetragonal coordinated (when bound to the substrate) takes on trigonal bipyramidal coordination in this transition state intermediate, where the substrate is covalently bound to the enzyme. Thereby one oxygen of the phosphate group positions in between Zn2 and Zn1 and stabilizes the transition state until autohydrolysis, where the leaving group gets cleaved and the phosphate stays, now again tetragonal coordinated, bound to PorX. The active site then gets restored by a nucleophilic attack by an activated hydroxy ion, which releases the second leaving group of the substrate and restores free T272. The structures we crystallized, both wt T272 bound to BeF₃ (similar as vanadate in the Xac NPP, PDB: 2gso)^[2] or T272A bound to pGpG (similar as pApG in ENPP1 T238A, PDB: 6aek)^[23] represent the transition state intermediate of substrate bound enzyme to the second leaving group before the enzyme is restored by a activated hydroxy ion close to the active site hydrolysing the T272-phosphate bond.

Evaluation: We haven't validated this reaction scheme for PorX but there is consensus about the canonical bi-metallo reaction pathway in APS for phosphatemono- and -diesterases as well as phosphotransferases and the reaction scheme described here is analog.

4.5 Substrate specificity

Result summary: During enzymatic assays PorX cleaved bis(4-nitrophenyl) phosphate (Chapter 3.15.2), a standard unspecific substrate used to measure phosphodiesterase activity. The standard assay for phosphatase activities by cleaving p-nitrophenyl phosphate (Mariusz Madej, personal communication) was negative. Further we determined PorX specificity by screening a nucleotide library (Chapter 3.15.3). PorX catalyzed the hydrolysis of linear and cyclic oligonucleotides (pGpG,

pApG, pApA, cA3, cA4 and cA6) as well as further hydrolysis of the intermediary hydrolysis products of oligonucleotides, resulting in NMPs as final PorX enzymatic products. PorX did not hydrolyse NMPs or NDPs AMP, GMP, ADP, GDP, cyclic mono- or dinucleotides (c-di-GMP, c-di-AMP, 2',3'-cGAMP, 3',3'-cGAMP, c-AMP, c-GMP) or mono or dinucleotide polyphosphates (ATP, GTP, AP3A, AP4A, AP4G). Conformation of cleavage and their products was conducted by UHPLC-MS. We did not observed any preference during our assays for either guanosine or adenosine nucleotides.

Contextualized discussion: The specificity for oligonucleotides we found in PorX is based on its large substrate binding pocket with multiple nucleotide binding sites (Chapter 3.14.2). In the binding site with the presumable highest affinity (based on our PorX_pGpG structure and pGpG substrate position), one nitrogenous base gets bound by a hydrophobic clamp between Y332 and L307 analog to other homolog proteins, its phosphate gets stabilized by R276/K330/N358. This subpocket opens during ligand binding as we observed comparing our ligand bound and free structures. In contrast to homolog phosphodiesterases this pocket is displaced further away from the active center and that is the reason why the homologs but not PorX can cleave NTPs between α - and β - phosphate (thus are pyrophosphatases), while this bond, as well as between β -/ γ -, are too distant in PorX. The reason, why PorX does not bind and cleaves AP4G or AP4A is likely due to the high electronegative charge from subpocket B as well residues from helix α 5a (Figure 3.26&Figure 7.2). The second nucleotide binding occurs at the active nucleoside which gets transient covalently bound to PorX at the active center. Its phosphate binds between the two zinc atoms as typical for bi-metallo APS transition state intermediates (Figure 3.25) while, in case of pGpG the nitrogen base gets stabilized by K305 and E302 which are part of PorX substrate binding loop L α 3b- α 3c. Further positioning of additional nucleotide moieties are speculative. Nucleotides 5' guanosine 2 (Gua2) will be likely unbound as the 5' oxygen of Gua2's ribose pointing to the solvent. 3' of Gua1 there are multiple hydrophobic pockets in PorX that could bind nucleotide analog to Gua 2 by a hydrophobic pocket (C1, C2 Figure 3.27A). If these act as nucleotide binding sites, it could explain the tyrosine-rich helix α III at this position, which is unique for PorX and not even found in other APS or PglZ domains of BREX proteins. Further there is the partial disordered loop L β Ic- β II, which also caps the active site. This loop builds a β -strand in homologs which is part central β -sheet of the CAP subdomain. The lack of stabilization by an additional nucleoside might be the reason why it does not follow the canonical fold of phosphodiesterases in our structures and might be an additional nucleoside binding site. Further the hydrophobic pocket C2 has some similarities to that of ENPP1

in which a chloride anion stabilizes the 5' nucleotide of pApG when bound (PDB: 6aek)^[23]. In PorX this chloride could be coordinated by Q499, T368 and H365. At this position a crystal water in tetrahedral coordination was build in our HR structure but an additional nucleotide would complete an octahedral coordination shell for a bound chloride analog as in ENPP1. Further pockets B and C might also be to reorientate ("flip")^[23] nucleoside moieties of oligonucleotides between pockets during cleavage as it is discussed for ENPP1^[23].

Evaluation: Synoptically we have shown that PorX is a specific phosphodiesterase cleaving linear and cyclic oligonucleotides to NMPs, with their two zinc APS catalytic center. The substrate specificity and cleavage we observed for PorX was not previously described for APS proteins and the binding mode as well as catalysis significantly differs from metal free catalysis of cyclic oligoadenylates by ring nucleases^[24] or Mn dependent oligoribonucleases^[25].^[0] Further we analyzed and explained the binding mode of two nucleoside moieties of pGpG and proposed further binding capacities and function of unused loops, surface subpockets and the PorX specific helix α III for larger substrates.

4.6 Enzymatic function

Result summary: Our studies identified PorX as a zinc dependent phosphodiesterase with hydrolytic activity against linear and circular oligonucleotides (cOAs) as well as linear dinucleotides, cleaving them successively into NMPs.

Contextualized discussion: Nothing was reported previously to our studies about any enzymatic activity of PglZ. In several publications PorX is described as "putative alkaline phosphatase",^[26],^[27] which we have proven to be wrong. However the enzymatic activity against cyclic oligonucleotids does not trivially explain the role of PorX in transcriptional activation. Enzymatic degradation of oligonucleotides could play a role in transcriptional activation by de-/activating transcription factors regulated by cyclic oligoadenylate (cOAs). As our collaborators reviewed the literature: In bacteria and archaea cOAs act as second messengers and are unregulated upon viral infection in anti-phage defence CBASS (cyclic oligonucleotide-based anti-phage signaling systems) and CRISPR-Cas type III. In CBASS cOAs bind and activate effector proteins that induce cell death by nucleic acid, metabolite and membrane degradation and thereby prevent the spread of phages within the prokaryotic population.^[0],^[28] In CRISPR-Cas cOAs lead to degradation of transcriptional-active nucleic acids by a Cas10 nuclease domain thereby protecting the cells from phage infection.^[0],^[29] Also the BREX (Bacteriophage-Exclusion) system, including PorX closest homolog PglZs, are

involved in phage defense and thereby gave the PglZ domain its name ("*phage growth limitation factor Z*"). BREX systems provide phage resistance by host DNA methylation and restriction of foreign DNA sequences^[26] and PglZ has been shown to be crucial for both DNA methylation as well as degradation^[30]. How any of these known anti-phage systems could relate or interact with the induction of virulence in *Porphyromonas gingivalis* (PG) remains elusive. Further PG does not encode any CBASS nucleotidyl-transferases synthesizing cOAs, nor do most of PG strains possess a CRISPR-Cas III system responsible for other known cOAs signal transductions. However there are no known phages that do infect PG.^{[0], [31], [32], [33], [34]}

Evaluation: As our collaborators suggest, it might be possible that PG possesses an previously undiscovered anti-phage system and PorX might degrade involved cOA signaling,^[0] or PorX is involved in DNA or RNA degradation. For example PorX might aid as oligoribonuclease in mRNA or nanoRNA decay (used for transcription initiation)^[35] and thus induces rapid environmental adaption upon an, yet unknown, environmental signal. We also cannot exclude that PorX enzymatic activity affects a signaling pathway outside of PG. When PG grows within the gingiva PorX might take part in cross-biofilm signaling. For example it could enhance c-di-GMP signal decay, a bacterial second messenger and a key signaling molecule in biofilm formation,^[36] by hydrolysis of pGpG, which inhibit c-di-GMP hydrolases by product inhibition.^{[37], [38]} An other potential function might be to disturb the host's innate immunity or viability (and therefore inducing inflammation), especially when PG lives intracellularly in host cells, which could be done by hydrolysis of host cells small messengers di- or oligonucleotides/RNAs. Validation of those theories would have needed further intensive investigations which went beyond the scope of this project but might be discovered based on our work.

4.7 Transcription regulation

Result summary: Our structural studies revealed the mechanism how phosphorylation of PorX leads to dimerization and enzymatic activation in order to cleave linear and circular oligonucleotides.

By analyzing and comparing different PglZ within their protein family (with BREX PglZs) we described potential surfaces suitable or unsuitable for protein-protein or protein-DNA/RNA interactions, in particular PorX specific tyrosine rich helix α III (YNYYVQYY), which is not found in other BREX PglZs and the region around CAP β III+ α 1, which builds an elongated helix in BREX3.

Contextualized discussion: In two component systems (TCS) the response regulators (RDs) get phosphorylated and thereby activated by a sensor histidin kinase (HK) upon external environmental stimulus. In *Porphyromonas gingivalis* (PG) type 9 secretion system (T9SS) PorX and PorY are assumed to be the RD/HK pair responsible for the expression of virulence factors.^[5] PorX gets phosphorylated by PorY,^{[5], [39]} which sits at the inner membrane where PorX interacts with PorL.^[5] Upon phosphorylation PorX dimerizes and activates the transcriptional regulation of *por* genes, which is thought to be through interaction with the sigma factor SigP.^[39] A recent study further claims PorX binding to sigP and porT promoters,^{[40], [41]} however it has been shown that PorX does not bind DNA, specifically not to those of the sigP and or sov genes/operons.^[5] There is also debate about PorX activation solely by PorY, as they are not encoded by nearby genes^[42] and our collaboration partners discussed alternative activation by alternative HKs or intracellular AcP in addition to the reported interaction with PorY,^[0] points that need to be further investigated. This leads to three different potential modes of regulation to discuss for PorX, a) the role of PorX enzymatic function in transcription regulation by cOAs hydrolysis, as discussed above (Chapter 4.6), b) potential protein-protein, as well as c) potential protein-nucleic acid interaction.

When comparing PorX with the BREX-PglZs, which are essential components of bacterial anti-phage/DNA restriction-modification systems,^{[27], [43]} some structural differences can be observed (Chapter 3.14.3). The differences in PorX flexible loop homolog regions are likely due to different substrate binding, while $\alpha 5b$ and $\alpha 4$ changes avoid clashes from slightly different relative positioning between monomers within a dimer (Figure 3.29B/C). CAP αIII and CAP $\beta III + \alpha I$ region on the other hand are potential candidates being involved in protein-protein or protein-DNA interaction. They are not part of the structural core or part of the immediate active site and they change their position as part of the CAP subdomain rearrangement upon dimerization. Likely they have a similar function in different BREX PglZs but with different interaction partners. For the closest PglZ homolog to PorX, BREX3, which is highly conserved even in the flexible regions (Figure 3.28), CAP $\beta III + \alpha I$ (suggested) binding site can be seen in exchange of helix CAP αIII in PorX which is missing in BREX3 and an elongated CAP αI at the opposing site of CAP is unique for BREX3 while other BREXs have flexible loops in both of those regions (Figure 3.29) supposedly they do not have this kind of interaction and those regions are not structural necessary. αIII flanks the two active sites of the dimer on both sites (Figure 7.2) and binds to the PglZ dimerization interface at $\alpha 5b$ helices and thereby keeps the CAP subdomains in an open position by hydrophilic interactions. These helices with their prominent aromatic side chain content might be

involved in interactions with PG sigma factors or with nucleic acids. Interactions at this helix could loosen the CAP interaction with $\alpha 5b$ and lead to an even more open position of the active sites, as we observed during SAXS of the dimer (Chapter 3.13.4, Figure 3.22). This would open a cavity starting from the C-terminus of the HBD via the active sites of both PglZs to the opposite monomer HBD, a large surface even for interaction with a nucleic acid double helix (DNA/DNA or DNA/RNA hybrid)(Figure 7.2).

The other region CAP $\beta III + \alpha I$ is pointing in the opposite direction as the active site (in contrast to capping loops) and is flanked by both flexible regions CAP $\alpha 2 + L\alpha 2 - \beta V$ and L $\beta Ic - \beta II$ (Figure 3.29B). In PorX this region only has a small helical region of one turn, while BREX3 PglZ has an elongated helix of the same size as PorX αIII . It is tempting to speculate that this region, which is not part of the much higher conserved central β -strand of the CAP subdomain, might unfold upon binding to other proteins and thereby explain reported binding of PorX to sigma factors.

An other interaction surface would be the dimerization surface of PorX in the monomer, which was shown to be disordered in solution (Chapter 3.13.5, Figure 3.23) but could bind to other proteins *in vivo*, e.g. PorL at the membrane or SigP (and thereby either activate or inhibit transcription).

Evaluation: Our discussion and suggestions about the potential protein-protein or protein nucleic acid binding regions are highly speculative and should be seen as what they are, an educated guess to explain reported and even contradictory results in the literature (e.g. DNA binding of PorX). From a structural point of view we did not find any DNA or other conserved interaction motives in PorX which would support any transcriptional regulation other than indirect regulation by enzymatic cOAs cleavage. Also binding motives or transcriptional regulation for homolog protein could not be found. However this was not the focus of our studies and the regions we suggested here might be a good starting point for future mutational studies on PorX interactions. Those should let PorX enzymatic activity intact, while potentially disturb those protein-protein/nucleic acid interactions reported, in contrast to e.g. T272A mutant with the opposite effect only affecting enzymatic activity but leaving the protein-protein/DNA/RNA interaction surfaces we suggested intact. This studies could include *in vivo* studies or biochemical interaction assays, or cocrystallization (PorY, SigP, reported promoter sequences etc.). Also we could not clarify how or if enzymatic activity and potential protein-protein (PorY, PorL) or protein-DNA (T9SS promoters) interactions of PorX are related. Likewise if they are independent functions of PorX, one involved in T9SS, the other in phage defense. Multiple functions of a TCS RR have not been described previously and combined with controversial discussion about protein/DNA binding, makes further

data necessary to draw more conclusions. Our data and characterization of potential surfaces provides the base for these studies.

4.8 Trouble shooting

As it happens in experimental sciences not all experiments and results are perfect. Here we would like to address some questions the careful reader might get and we have shortened out of the discussion.

4.8.1 Function of LH conformation in PorX

We see LH conformation as an crystallization/*in vitro* artifact, as it only appeared in structures without any nucleotide added before crystal formation (without physiological concentration of either nucleotides or inorganic phosphate). Phosphates *in vivo* or in buffers would bridge both zincs and thereby stabilize binding of Zn1 and $\alpha 5$ helix formation in both monomers inside a dimer. Thus we had LH/HH conformations at crystals HR (0.1 mM zinc added) but at the structure "GMP" or "GTP γ S" (GMP or GTP γ S cocrystallization setups without stable binding, 0.1 mM zinc) there we saw double HH. Likewise we observed double HH in our pGpG structure (0.5 mM zinc) but not in the LH/HH structure from the parallel prepared c-di-GMP cocrystallization setup (also 0.5 mM zinc, no binding of substrate) which had a cyclic bound phosphate group which did not bound to the PglZ (otherwise it would be cleaved as well during our enzymatic assays, Figure 7.10J). However, we observed zinc induced dimerization in our SAXS experiments, even if we do not know the exact conformation of this unphysiological dimer at higher resolution (TABLE 3.16), we know at least one HH conformation between two PglZ is sterical necessary. We also observed a double HH conformation in the proteolytical digested (missing RD and HBD H1) "D58A" structure (0.5 mM zinc) crystallized without any phosphate acting as a catalyst for HH formation. However this construct has more degrees of freedom then the full-length protein due to missing RD, which in monomeric HBD conformation would clash and destabilize PglZ dimerization.

4.8.2 Enzymatic activities of monomeric PorX samples in enzymatic assays

One inconsistency in the phosphodiesterase assays Chapter 3.15.2 has to be addressed. The still lower, but significant enzymatic activity of unphosphorylated PorX in those assays is due to the unphysiological high^[44] zinc concentrations. Excess concentrations of zinc lead to unphysiological PglZ dimers, as our SAXS data suggests (TABLE 3.15) and zinc-induced dimerization was observed in SEC with zinc containing buffer (Figure 7.4). In the SAXS electron density

reconstruction you can see two smaller domains in the size of the RDs pointing at opposite ends of the dimer, while likely (in size and shape) the two PglZ domains dimerize in the center, gaining residual enzymatic activity likely due to dimerization induced HH conformation. This explains the higher radius of gyration (R_g), maximum distances (D_{max}) and SEC elution behavior of this dimer in contrast to the BeF_3 induced dimer, which represents structural the native dimer. It also explains, why AcP did not increased the activity of D58A or M94K/D104A/I129A, which, as seen during SEC (Figure 7.3), do not dimerize PorX in the absence of zinc.

4.8.3 SAXS, concentrations and interparticle interactions

The strong interparticle interactions seen at low q at PorX experiments could lead to the assumption that PorX is unstable and tends to aggregate compared to other proteins. This is not the case rather we could not prepare the samples in the optimal ways suitable for SAXS experiments. Due to the COVID-19 pandemic restrictions we had to prepare our PorX samples in Barcelona and send them on ice to the ESRF in Grenoble to measure. Preparing samples and concentrating them just before the experiment would led to less interparticle interactions and would have also allowed higher concentrations yielding higher resolution maps. We suggest sample concentrations of at least 30 mg/ml for PorX monomer (instead of 10 mg/ml) and for dimeric PorX to 10 mg/ml instead of 2 mg/ml as highest concentrations of a concentration series. Also a multitude of samples and data merging would had improved the collected data as the noise levels were very high for some samples, therefore we had to take them out of our analysis.

5. Prospects

5.1 Crystallization

The crystallization protocols for dimeric PorX do not need to be further optimized and the two screens Opt_J and OptL are suitable for each kind of further substrate screening. As of those we should cocrystallize cyclic oligonucleotides with T272A inactive mutant. This would not only confirm their binding and yield the first structures of an alkaline phosphatase superfamily protein binding these substrates, it also would answer the open questions we have about functions and binding of PorX_pGpG unused binding pockets C1/C2 and the role of the flexible loop L β Ic- β II and helix CAP α III.

For PorX monomers we never tried higher concentrations than 14 mg/ml and we never added zinc into the reservoir solutions of screens. This, so we speculate, will induce a zinc induced dimer as we see during SAXS, which could crystallize and its structure solved. This dimer might not be physiological but the contacts between receiver domain (RD), helical bundle domain (HBD) and PglZ, as well as their respective conformations will shed light on the mechanism of PorX activation. Further it would be worth to test cocrystallization of PorX monomers and dimers with PorY cytoplasmatic domain, in the presence of ADP, ATP analogs and/or phosphate mimics. In addition this cocrystallizations could be tried with and without PorL, which binds to PorX at the plasma membrane where PorX and PorY interact.^[5] For those experiments also a proper screening of stable PorY constructs might be necessary, as those constructs we worked with were unstable in solution and difficult to concentrate. Also mutational studies to enhance stable binding between PorX and PorY could be a solution to solve the structure of PorY. Further cocrystallization of PorX with SigP or PG1660 (or other sigma factors, with and without DNA) could be tried, as well as cocrystallization with those promoter sequences published to interact directly with PorX (sigP and porT promoter)^{[40], [41]}. Preceding co-crystallographic experiments SAXS should be considered.

5.2 Comparative homology

We just started with to shed light on the functional roles of BREX-PglZ and our experience in producing and crystallizing PorX might help us in solving the structures of BREX-PglZ proteins. Also it would be fascinating if those proteins share the same activity as PorX and cleavage assays with those proteins could confirm that. On the other hand future discoveries of other researchers made for BREX-PglZ domains might help us understanding protein-protein or protein-DNA interactions discussed for PorX, signaling pathways involved and transcriptional regulation.

5.3 Enzymatic activity and transcriptional regulation

We have discovered the enzymatic activity of PorX against cyclic oligoadenylates (cOAs) but the role of cOAs in PG remains elusive. These might play a role in phage immunity or play a role in PGs live cycle degrading nanoRNAs. The cOAs could also be produced by other members of the biofilm (cross-biofilm signaling), or PorX enzymatic activity could play a role in digesting signal molecules of the host's immune system. These open questions need further research *in vivo* with mutational studies based on our structure. An important question is if the enzymatic activity of PorX is independent of the transcriptional activity and if it is possible to inactivate PorX (e.g. T272A mutant) and keep the potential transcription activation (as published) intact or vice versa by mutants of our suggested interaction interfaces CAP β III+ α I and CAP α III.

5.4 PorX as a drug target

Because PorX is responsible for the pathogenicity of *Porphyromonas gingivalis*, our structures will aid in future drug design based on modified substrates known to inhibit phosphodiesterases of the APS. This approach is promising, as the active center of PorX is well conserved, so we can utilize functional groups from known PDE inhibitors. On the other hand the surrounding substrate binding pocket of PorX differs significantly from human homolog structures which are known so far (as ENPP1-6). Therefore, if substrates are specifically designed for PorX this would lower any cross-reactions with host proteins. A promising approach would be to elongate drugs targeting the hydrophobic clamp of PDEs. In PorX this clamp is shifted away from the active site allowing to fit an additional phosphate and ribose moiety and target this unique pocket. This research based on our PorX structure could then be tested *in vitro* for binding affinities, cocrystallizations etc. and then within time and after clinical studies might yield in medication to fight periodontal and other systemic inflammatory diseases PG is involved in, increasing human life span and quality.

6. Conclusions

- PorX is a multidomain protein with an N-terminal receiver domain (RD) linked by a helical bundle domain (HBD) to an C-terminal PglZ domain which was previously uncharacterized in structure and function.
- The RD belongs to the well conserved two component response regulator receiver domains of the *OmpR* family with a $\alpha_5\beta_5$ Rossmann-like fold, the conserved phosphorylation site at D58 and the so called TY pair T86/Y105.
- PorX gets activated upon phosphorylation which leads to conformational changes resulting in RD dimerization and exchange of interacting surfaces between RDs and HBDs and HBDs with PglZs.
- The HBD consists of 3 helices H1-3, two C-terminal helices H2&H3 connected with the PglZ domain and a N-terminal helix H1. Upon RD phosphorylation H1 changes surfaces from H2&H3 to the second monomers RD.
- The PglZ domain belongs to the alkaline phosphatase superfamily (APS) with a central six stranded β -sheet, surrounded by helices. It has the conserved active site residues of the phosphodiesterase subfamily coordinating 2 zinc ions, as well as a so called CAP domain, which partially caps the substrate binding pocket.
- Interactions of the HBDs with the other monomers RD's brings the PglZ domains together which thereby folding the active center and opening of the substrate binding pocket.
- Dimerization of PglZ domains occurs at disordered regions at the monomer that fold into three consecutive helices $\alpha 5a/b/c$ upon dimerization and $\alpha 5a$ coordinates a second zinc ion necessary for enzymatic activity. Further the CAP subdomain interacting with $\alpha 5b$ changes relative position to the catalytic center, which likely favors substrate entry.
- There is an interdomain crosstalk between RD and PglZ that each activated/dimerized state of either RD or PglZ stabilize each other via the HBD.
- PorX is a phosphodiesterase with the canonical reaction pathway of bi-metallo APS enzymes. Our structures mimic the typical transition state intermediates, where the leaving group phosphate (in our cases beryllium fluoride or 5' phosphate of pGpG) is covalently bound to the reactive T272.
- PorX hydrolyses cyclic and linear oligonucleotides as well as linear dinucleotides to NMPs, without activity against circular mono- or dinucleotides or mono or dinucleotidepolyphosphates.
- The signaling pathways of PorX involving oligoadenylates, mechanisms of transcription activation by PorX and ways to inhibit PorX based on our work will be subject of future studies.

Supplementary

[illegible]

7. Supplementary

7.1 Literature Introduction

- 1: Shah, HN & Collins MD: "*Proposal for Reclassification of Bacteroides asaccharolyticus* , *Bacteroides gingivalis*, and *Bacteroides endodontalis* in a New Genus, *Porphyromonas*", Int. Journal of Systematic Bacteriology , 38 (1), 128–131, 1988.
- 2: Hajishengallis, G.; Lamont, R. J.: "*Beyond the red complex and into more complexity*", Molecular oral microbiology 27, 6, 409–419, 2012.
- 3: DYE, B. A.: "*Global periodontal disease epidemiology*", Periodontology 2000 , 58, 10–25, 2012.
- 4: Jia, Lu; Han, Nannan; Du, Juan; Guo, Lijia; Luo, Zhenhua; Liu, Yi: "*Pathogenesis of Important Virulence Factors of Porphyromonas gingivalis via Toll-Like Receptors*", Frontiers in cellular and infection microbiology 9, 262, 2019.
- 5: Kassebaum, N. J.; Bernabé, E.; Dahiya, M.; Bhandari, B.; Murray, C. J. L.; Marcenes, W.: "*Global burden of severe periodontitis in 1990-2010*", Journal of dental research 93, 11, 1045–1053, 2014.
- 6: Hajishengallis G, Liang S, Payne MA, et al.: "*Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement*", Cell host & microbe 10, 5, 497–506, 2011.
- 7: Olsen, Ingar; Yilmaz, Özlem: "*Modulation of inflammasome activity by Porphyromonas gingivalis in periodontitis and associated systemic diseases*", Journal of oral microbiology 8, 30385, 2016.
- 8: Huck, Olivier; You, Jian; Han, Xianxian; Cai, Bin; Panek, James; Amar, Salomon: "*Reduction of Articular and Systemic Inflammation by Kava-241 in a Porphyromonas gingivalis-Induced Arthritis Murine Model*", Infection and immunity 86, 9, 2018.
- 9: Gholizadeh, Pourya; Eslami, Hosein; Yousefi, Mehdi; Asgharzadeh, Mohammad; Aghazadeh, Mohammad; Kafil, Hossein Samadi: "*Role of oral microbiome on oral cancers, a review*", Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 84, 552–558, 2016.
- 10: Detert, Jacqueline; Pischon, Nicole; Burmester, Gerd R.; Buttgereit, Frank: "*The association between rheumatoid arthritis and periodontal disease*", Arthritis research & therapy 12, 5, 218, 2010.
- 11: Reyes, Leticia; Herrera, David; Kozarov, Emil; Roldá, Silvia; Progulske-Fox, Ann: "*Periodontal bacterial invasion and infection*", Journal of periodontology , 84(4 Suppl), S30-50, 2013.
- 12: Huck, O; Saadi-Thiers, K; Tenenbaum, H; Davideau, JL; Romagna, C; Laurent, Y; Cottin, Y; Roul, JG.: "*Evaluating periodontal risk for patients at risk of or suffering from atherosclerosis*", Archives of cardiovascular diseases 104, 5, 352–358, 2011.
- 13: Chistiakov, Dimitry A.; Orekhov, Alexander N.; Bobryshev, Yuri V.: "*Links between atherosclerotic and periodontal disease*", Experimental and molecular pathology 100, 1, 220–235, 2016.
- 14: El-Shinnawi, Una; Soory, Mena: "*Associations between periodontitis and systemic inflammatory diseases-Response to treatment*", Recent patents on endocrine, metabolic & immune drug discovery 7, 3, 169–188, 2013.
- 15: Chapple, Iain L. C.; Genco, Robert: "*Diabetes and periodontal diseases*", Journal of periodontology 84, 4 Suppl, S106-12, 2013.
- 16: Wu, Zhou; Nakanishi, Hiroshi: "*Connection between periodontitis and Alzheimer's disease*", Journal of pharmacological sciences 126, 1, 8–13, 2014.
- 17: Michaud, Dominique S.: "*Role of bacterial infections in pancreatic cancer*", Carcinogenesis 34, 10, 2193–2197, 2013.
- 18: Imai, K. & Ochiai, K.: "*Role of histone modification on transcriptional regulation and HIV-1 gene expression: possible mechanisms of periodontal diseases in AIDS progression*", Journal of Oral Science , 53(1), 1–13, 2011.
- 19: Fiorillo, Luca; Cervino, Gabriele; Laino, Luigi; D'Amico, Cesare; Mauceri, Rodolfo; Tozum, Tolga Fikret; Gaeta, Michele; Cicciù, Marco: "*Porphyromonas gingivalis*, Periodontal and Systemic Implications-A Systematic Review", Dentistry journal 7, 4, 2019.
- 20: DeLeon-Pennell, Kristine Y.; Castro Brás, Lisandra E. de; Lindsey, Merry L.: "*Circulating Porphyromonas gingivalis lipopolysaccharide resets cardiac homeostasis in mice through a matrix metalloproteinase-9-dependent mechanism*", Physiological reports 1, 5, e00079, 2013.
- 21: Bao, Lirong; Zhang, Cheng; Dong, Jiajia; Zhao, Lei; Li, Yan; Sun, Jianxun: "*Oral Microbiome and SARS-CoV-2*", Frontiers in microbiology 11, 1840, 2020.
- 22: Dewhirst, FE.; Chen, T; Izard, J; Paster, BJ.; Tanner, ACR; Yu, WH; Lakshmanan, A; Wade, WG: "*The*

- human oral microbiome*", Journal of bacteriology 192, 19, 5002–5017, 2010.
- 23: Kolenbrander, Paul E.; Andersen, Roxanna N.; Blehert, David S.; Eglund, Paul G.; Foster, Jamie S.; Palmer, Robert J.: "*Communication among oral bacteria*", MMBR 66, 3, 486–505, 2002.
- 24: Lamont, Richard J.; Koo, Hyun; Hajishengallis, George: "*The oral microbiota*", Nature reviews. Microbiology 16, 12, 745–759, 2018.
- 25: Aas, Jørn A.; Paster, Bruce J.; Stokes, Lauren N.; Olsen, Ingar; Dewhirst, Floyd E.: "*Defining the normal bacterial flora of the oral cavity*", Journal of clinical microbiology 43, 11, 5721–5732, 2005.
- 26: Avia, M. and Yimaz, O.: "*The Oral Microbiota: Living with a Permanent Guest*", DNA AND CELL BIOLOGY, 28(8), 405–411, 2009.
- 27: Aruni, A. Wilson; Dou, Yuetan; Mishra, Arunima; Fletcher, Hansel M.: "*The Biofilm Community-Rebels with a Cause*", Current oral health reports 2, 1, 48–56, 2015.
- 28: Dethlefsen, Les; McFall-Ngai, Margaret; Relman, David A.: "*An ecological and evolutionary perspective on human-microbe mutualism and disease*", Nature 449, 7164, 811–818, 2007.
- 29: Said HS, Suda W, Nakagome S, et al. : "*Dysbiosis of salivary microbiota in inflammatory bowel disease and its association with oral immunological biomarkers*", DNA research : an international journal for rapid publication of reports on genes and genomes 21, 1, 15–25, 2014.
- 30: Jones, Mitchell L.; Martoni, Christopher J.; Ganopolsky, Jorge G.; Labbé, Alain; Prakash, Satya: "*The human microbiome and bile acid metabolism*", Expert opinion on biological therapy 14, 4, 467–482, 2014.
- 31: Bjarnsholt, Thomas; Jensen, Peter Østrup; Moser, Claus; Høiby, Niels: "*Biofilm Infections*", Springer Science+Business Media LLC, 2011.
- 32: Lewis, K.: "*Persister Cells and the Riddle of Biofilm Survival*", Biochemistry (Moscow), 70(2), 327–336, 2005.
- 33: Jamal, M; Ahmad, W; Andleeb, S; Jalil, F; Imran, M; Nawaz, MA; Hussain, Tahir; Ali, M; Rafiq, M; Kamil, MA: "*Bacterial biofilm and associated infections*", Journal of the Chinese Medical Association : JCMA 81, 1, 7–11, 2018.
- 34: Bos, R.; van der Mei, H. C.; Busscher, H. J.: "*Co-adhesion of oral microbial pairs under flow in the presence of saliva and lactose*", Journal of dental research 75, 2, 809–815, 1996.
- 35: Periasamy, Saravanan; Kolenbrander, Paul E.: "*Mutualistic biofilm communities develop with Porphyromonas gingivalis and initial, early, and late colonizers of enamel*", Journal of bacteriology 191, 22, 6804–6811, 2009.
- 36: Marcotte, H. and Lavoie, MC: "*Oral Microbial Ecology and the Role of Salivary Immunoglobulin A*", MMBR, 62(1), 71–109, 1998.
- 37: KOLENBRANDER, PE.; PALMER JR, RJ.; RICKARD, AH.; JAKUBOVICS, NS.; CHALMERS, NI & DIAZ, PI: "*Bacterial interactions and successions during plaque development*", Periodontology 2000, 42, 47–79, 2006.
- 38: Bassler, B. L.: "*How bacteria talk to each other: regulation of gene expression by quorum sensing*", Current Opinion in Microbiology, 2, 582–587, 1999.
- 39: Scheres, Nina; Lamont, Richard J.; Crielaard, Wim; Krom, Bastiaan P.: "*LuxS signaling in Porphyromonas gingivalis-host interactions*", Anaerobe, 35, 3–9, 2015.
- 40: Li, Y. H.; Lau, P. C.; Lee, J. H.; Ellen, R. P.; Cvitkovitch, D. G.: "*Natural genetic transformation of Streptococcus mutans growing in biofilms*", Journal of bacteriology 183, 3, 897–908, 2001.
- 41: Grenier, D.; Mayrand, D.: "*Nutritional relationships between oral bacteria*", Infection and immunity 53, 3, 616–620, 1986.
- 42: Hajishengallis, George: "*Periodontitis*", Nature reviews. Immunology 15, 1, 30–44, 2015.
- 43: Hajishengallis, George; Darveau, Richard P.; Curtis, Michael A.: "*The keystone-pathogen hypothesis*", Nature reviews. Microbiology 10, 10, 717–725, 2012.
- 44: Brogden, Kim A.; Guthmiller, Janet M.: "*Polymicrobial Diseases*", 2002.
- 45: Sakanaka, Akito; Takeuchi, Hiroki; Kuboniwa, Masae; Amano, Atsuo: "*Dual lifestyle of Porphyromonas gingivalis in biofilm and gingival cells*", Microbial pathogenesis 94, 42–47, 2016.
- 46: Duran-Pinedo, Ana E.; Baker, Vinesha D.; Frias-Lopez, Jorge: "*The periodontal pathogen Porphyromonas gingivalis induces expression of transposases and cell death of Streptococcus mitis in a biofilm model*", Infection and immunity 82, 8, 3374–3382, 2014.
- 47: Nyvad, B.; Kilian, M.: "*Microbiology of the early colonization of human enamel and root surfaces in vivo*", Scandinavian journal of dental research 95, 5, 369–380, 1987.
- 48: Lucas, V. S.; Beighton, D.; Roberts, G. J.: "*Composition of the oral streptococcal flora in healthy children*", Journal of Dentistry, 28, 45–50, 2000.
- 49: Bourgeois, Denis; Inquimbert, Camille; Ottolenghi, Livia; Carrouel, Florence: "*Periodontal Pathogens as*

- Risk Factors of Cardiovascular Diseases, Diabetes, Rheumatoid Arthritis, Cancer, and Chronic Obstructive Pulmonary Disease-Is There Cause for Consideration?*, Microorganisms 7, 10, 2019.
- 50: Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent Jr. RL: "Microbial complexes in subgingival plaque", J Clin Periodontol, 25, 134–144, 1998.
- 51: Diaz, P. I.; Zilm, P. S.; Rogers AH: "The response to oxidative stress of *Fusobacterium nucleatum* grown in continuous culture", FEMS Microbiology Letters, 187, 31–34, 2000.
- 52: Diaz, Patricia I.; Slakeski, Nada; Reynolds, Eric C.; Morona, Renato; Rogers, Anthony H.; Kolenbrander, Paul E.: "Role of *oxyR* in the oral anaerobe *Porphyromonas gingivalis*", Journal of bacteriology 188, 7, 2454–2462, 2006.
- 53: Kuboniwa, Masae; Hendrickson, Erik L.; Xia, Qiangwei; Wang, Tiansong; Xie, Hua; Hackett, Murray; Lamont, Richard J.: "Proteomics of *Porphyromonas gingivalis* within a model oral microbial community", BMC microbiology 9, 98, 2009.
- 54: Yost, Susan; Duran-Pinedo, Ana E.; Teles, Ricardo; Krishnan, Keerthana; Frias-Lopez, Jorge: "Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis", Genome medicine 7, 1, 27, 2015.
- 55: Kumar, Purnima S.; Leys, Eugene J.; Bryk, Jennifer M.; Martinez, Francisco J.; Moeschberger, Melvin L.; Griffen, Ann L.: "Changes in periodontal health status are associated with bacterial community shifts as assessed by quantitative 16S cloning and sequencing", Journal of clinical microbiology 44, 10, 3665–3673, 2006.
- 56: Frias-Lopez, Jorge; Duran-Pinedo, Ana: "Effect of periodontal pathogens on the metatranscriptome of a healthy multispecies biofilm model", Journal of bacteriology 194, 8, 2082–2095, 2012.
- 57: Darveau, R. P.: "Periodontitis-A polymicrobial disruption of host homeostasis", Nature reviews. Microbiology 8, 7, 481–490, 2010.
- 58: Fleetwood AJ, Lee MKS, Singleton W, et al. : "Metabolic Remodeling, Inflammasome Activation, and Pyroptosis in Macrophages Stimulated by *Porphyromonas gingivalis* and Its Outer Membrane Vesicles", Frontiers in cellular and infection microbiology 7, 351, 2017.
- 59: Lasica AM, Ksiazek M, Madej M, Potempa J: "The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function", Front Cell Infect Microbiol., 7, 215, 2017.
- 60: Darveau RP, Pham TT, Lemley K, et al.: "Porphyromonas gingivalis Lipopolysaccharide Contains Multiple Lipid A Species That Functionally Interact with Both Toll-Like Receptors 2 and 4", IAI (Infection and Immunity) 72, 9, 5041–5051, 2004.
- 61: Oliveira-Nascimento L, Massari P, Wetzler LM: "The Role of TLR2 in Infection and Immunity", Frontiers in immunology 3, 79, 2012.
- 62: Hashimoto, Masahito; Asai, Yasuyuki; Ogawa, Tomohiko: "Separation and structural analysis of lipoprotein in a lipopolysaccharide preparation from *Porphyromonas gingivalis*", International immunology 16, 10, 1431–1437, 2004.
- 63: Herath, Thanuja D. K.; Darveau, Richard P.; Seneviratne, Chaminda J.; Wang, Cun-Yu; Wang, Yu; Jin, Lijian; Ojcius, David M.: "Tetra- and Penta-Acylated Lipid A Structures of *Porphyromonas gingivalis* LPS Differentially Activate TLR4-Mediated NF- κ B Signal Transduction Cascade and Immuno-Inflammatory Response in Human Gingival Fibroblasts", PLoS ONE 8, 3, e58496, 2013.
- 64: Sochalska, Maja; Potempa, Jan: "Manipulation of Neutrophils by *Porphyromonas gingivalis* in the Development of Periodontitis", Frontiers in cellular and infection microbiology 7, 197, 2017.
- 65: Divanovic S, Trompette A, Atabani SF, et al. : "Negative regulation of Toll-like receptor 4 signaling by the Toll-like receptor homolog RP105.", Nat Immunol., 6, 571–578, 2005.
- 66: Kusumoto Y, Hirano H, Saitoh K et al.: "Human Gingival Epithelial Cells Produce Chemotactic Factors Interleukin and Monocyte Chemoattractant Protein After Stimulation With *Porphyromonas gingivalis* via TollLike Receptor 2", J Periodontol., 75(3), 370–379, 2004.
- 67: Pollreis, A.; Huang, Y.; Roth, G. A.; Cheng, B.; Kebschull, M.; Papapanou, P. N.; Schmidt, A. M.; Lalla, E.: "Enhanced monocyte migration and pro-inflammatory cytokine production by *Porphyromonas gingivalis* infection", Journal of Periodontal Research 45, 2, 239–245, 2010.
- 68: Wang, M; Shakhathreh, MAK.; James, D; Liang, S; Nishiyama, S; Yoshimura, F; Demuth, DR.; Hajishengallis, G: "Fimbrial Proteins of *Porphyromonas gingivalis* Mediate In Vivo Virulence and Exploit TLR2 and Complement Receptor 3 to Persist in Macrophages", J Immunol 179, 4, 2349–2358, 2007.
- 69: Hajishengallis G, Wang M, Liang S, Triantafilou M, Triantafilou K: "Pathogen induction of CXCR4/TLR2 cross-talk impairs host defense function.", Proc Natl Acad Sci U S A., 105(36), 13532–13537, 2008.
- 70: Wang, M.; Krauss, J. L.; Domon, H.; Hosur, K. B.; Liang, S.; Magotti, P.; Triantafilou, M.; Triantafilou, K.; Lambris, J. D.; Hajishengallis, G.: "Microbial Hijacking of Complement-Toll-Like Receptor Crosstalk", Science

Signaling 3, 109, ra11-ra11, 2010.

71: Liang, S; Krauss, JL.; Domon, H; McIntosh, ML.; Hosur, KB.; Qu, H; Li, F; Tzekou, A; Lambris, JD.; Hajishengallis, G: "*The C5a Receptor Impairs IL-12-Dependent Clearance of Porphyromonas gingivalis and Is Required for Induction of Periodontal Bone Loss*", J.I. 186, 2, 869–877, 2011.

72: Hajishengallis, G.; McIntosh, M. L.; Nishiyama, S.-I.; Yoshimura, F.: "*Mechanism and implications of CXCR4-mediated integrin activation by Porphyromonas gingivalis*", Mol oral Microbiol 28, 4, 239–249, 2013.

73: van Bergenhenegouwen, J; Plantinga, TS.; Joosten, LAB.; Netea, MG.; Folkerts, G; Kraneveld, AD.; Garssen, J; Vos, AP: "*TLR2 & Co*", Journal of Leukocyte Biology 94, 5, 885–902, 2013.

74: Maekawa T, Krauss JL, Abe T, et al. : "*Porphyromonas gingivalis Manipulates Complement and TLR Signaling to Uncouple Bacterial Clearance from Inflammation and Promote Dysbiosis*", Cell host & microbe 15, 6, 768–778, 2014.

75: Makkawi, Hasnaa; Hoch, Shifra; Burns, Elia; Hosur, Kavita; Hajishengallis, George; Kirschning, Carsten J.; Nussbaum, Gabriel: "*Porphyromonas gingivalis Stimulates TLR2-PI3K Signaling to Escape Immune Clearance and Induce Bone Resorption Independently of MyD88*", Frontiers in cellular and infection microbiology 7, 359, 2017.

76: Olsen, Ingar; Lambris, John D.; Hajishengallis, George: "*Porphyromonas gingivalis disturbs host-commensal homeostasis by changing complement function*", Journal of oral microbiology 9, 1, 1340085, 2017.

77: Amano, Atsuo: Frontiers in Bioscience , 12, 3965–3974, 2007/"*Diversity of structure and function of response regulator output domains*", Curr Opin Microbiol. 13(2),150-159, 2010.

78: Ito, Rieko; Ishihara, Kazuyuki; Shoji, Mikio; Nakayama, Koji; Okuda, Katsuji: "*Hemagglutinin/Adhesin domains of Porphyromonas gingivalis play key roles in coaggregation with Treponema denticola*", FEMS immunology and medical microbiology 60, 3, 251–260, 2010.

79: Potempa, M., Potempa, J., Okroj, M., Popadiak, K., Eick, S., Nguyen, K. A., Riesbeck, K., & Blom, A. M.: "*Binding of complement inhibitor C4b-binding protein contributes to serum resistance of Porphyromonas gingivalis*.", Journal of immunology , 181(8), 5537–5544, 2008.

80: Zhang, Lianyi; Veith, Paul D.; Huq, N. Laila; Chen, Yu-Yen; Seers, Christine A.; Cross, Keith J.; Gorasia, Dhana G.; Reynolds, Eric C.: "*Porphyromonas gingivalis Gingipains Display Transpeptidation Activity*", J. Proteome Res. 17, 8, 2803–2818, 2018.

81: Mikolajczyk-Pawlinska J, Travis J, Potempa J.: "*Modulation of interleukin-8 activity by gingipains from Porphyromonas gingivalis: implications for pathogenicity of periodontal disease*.", FEBS Lett. , 440(3), 282–286, 1998.

82: Darveau RP, Belton CM, Reife RA, Lamont RJ.: "*Local Chemokine Paralysis, a Novel Pathogenic Mechanism for Porphyromonas gingivalis*", Infect Immun. , 66(4), 1660–1665, 1998.

83: Dias, Irundika H. K.; Marshall, Lindsay; Lambert, Peter A.; Chapple, Iain L. C.; Matthews, John B.; Griffiths, Helen R.: "*Gingipains from Porphyromonas gingivalis increase the chemotactic and respiratory burst-priming properties of the 77-amino-acid interleukin-8 variant*", Infection and immunity 76, 1, 317–323, 2008.

84: Wilensky, Asaf; Tzach-Nahman, Rinat; Potempa, Jan; Shapira, Lior; Nussbaum, Gabriel: "*Porphyromonas gingivalis gingipains selectively reduce CD14 expression, leading to macrophage hyporesponsiveness to bacterial infection*", Journal of innate immunity 7, 2, 127–135, 2015.

85: Rovai, E. S.; Holzhausen, M.: "*The Role of Proteinase-Activated Receptors 1 and 2 in the Regulation of Periodontal Tissue Metabolism and Disease*", Journal of immunology research 2017, 5193572, 2017.

86: Jung, Young-Jung; Jun, Hye-Kyoung; Choi, Bong-Kyu: "*Contradictory roles of Porphyromonas gingivalis gingipains in caspase-1 activation*", Cellular microbiology 17, 9, 1304–1319, 2015.

87: Shah HN, Gharbia S. E.: "*Lysis of erythrocytes by the secreted cysteine proteinase of Porphyromonas gingivalis W83*", FEMS Microbiol Lett. , 52(1-2), 213–217, 1989.

88: Lewis JP, Dawson JA, Hannis JC, Muddiman D, Macrina FL: "*Hemoglobinase activity of the lysine gingipain protease (Kgp) of Porphyromonas gingivalis W83*.", J Bacteriol. , 181(16), 4905–4913, 1999.

89: Sroka, A.; Sztukowska, M.; Potempa, J.; Travis, J.; Genco, C. A.: "*Degradation of host heme proteins by lysine- and arginine-specific cysteine proteinases (gingipains) of Porphyromonas gingivalis*", Journal of bacteriology 183, 19, 5609–5616, 2001.

90: Glew, Michelle D.; Veith, Paul D.; Chen, Dina; Gorasia, Dhana G.; Peng, Ben; Reynolds, Eric C.: "*PorV is an Outer Membrane Shuttle Protein for the Type IX Secretion System*", Scientific reports 7, 1, 8790, 2017.

91: Jauregui, Catherine E.; Wang, Qian; Wright, Christopher J.; Takeuchi, Hiroki; Uriarte, Silvia M.; Lamont, Richard J.; Bliska, J. B.: "*Suppression of T-Cell Chemokines by Porphyromonas gingivalis*", Infect. Immun. 81, 7, 2288–2295, 2013.

- 92: Oido-Mori, M.; Rezzonico, R.; Wang, P. L.; Kowashi, Y.; Dayer, J. M.; Baehni, P. C.; Chizzolini, C.: "*Porphyromonas gingivalis* gingipain-R enhances interleukin-8 but decreases gamma interferon-inducible protein 10 production by human gingival fibroblasts in response to T-cell contact", *Infection and immunity* 9, 7, 4493–4501, 2001.
- 93: Pyrc K, Milewska A, Kantyka T, et al.: "*Inactivation of Epidermal Growth Factor by Porphyromonas gingivalis* as a Potential Mechanism for Periodontal Tissue Damage", *Infect. Immun.* 81, 1, 55–64, 2012.
- 94: Aliko A, Kamińska M, Bergum B, Gawron K, Benedyk M, Lamont RJ, Malicki S, Delaleu N, Potempa J, Mydel P: "*Impact of Porphyromonas gingivalis* Peptidylarginine Deiminase on Bacterial Biofilm Formation, Epithelial Cell Invasion, and Epithelial Cell Transcriptional Landscape", *Scientific reports* 8, 1, 14144, 2018.
- 95: Bereta, Grzegorz; Goulas, Theodoros; Madej, Mariusz; Bielecka, Ewa; Solà, Maria; Potempa, Jan; Xavier Gomis-Rüth, F.: "*Structure, function, and inhibition of a genomic/clinical variant of Porphyromonas gingivalis* peptidylarginine deiminase", *Protein Science* 28, 3, 478–486, 2019.
- 96: Taguchi Y, Sato K, Yukitake H, et al.: "*Involvement of an Skp-Like Protein, PGN_0300, in the Type IX Secretion System of Porphyromonas gingivalis*", *Infection and immunity* 84, 1, 230–240, 2016.
- 97: Mydel P, Takahashi Y, Yumoto H, et al. : "*Roles of the host oxidative immune response and bacterial antioxidant rubrerythrin during Porphyromonas gingivalis* infection", *PLoS pathogens* 2, 7, e76, 2006.
- 98: Nakayama, Masaaki; Inoue, Tetsuyoshi; Naito, Mariko; Nakayama, Koji; Ohara, Naoya: "*Attenuation of the phosphatidylinositol 3-kinase/Akt signaling pathway by Porphyromonas gingivalis* gingipains RgpA, RgpB, and Kgp", *The Journal of biological chemistry* 290, 8, 5190–5202, 2015.
- 99: Singhrao, Sim K.; Olsen, Ingar: "*Are Porphyromonas gingivalis* Outer Membrane Vesicles Microbullets for Sporadic Alzheimer's Disease Manifestation?", *ADR* 2, 1, 219–228, 2018.
- 100: Veith PD, Chen YY, Gorasia DG, Chen D, Glew MD, O'Brien-Simpson NM, Cecil JD, Holden JA, Reynolds EC: "*Porphyromonas gingivalis* Outer Membrane Vesicles Exclusively Contain Outer Membrane and Periplasmic Proteins and Carry a Cargo Enriched with Virulence Factors", *J. Proteome Res.* 13, 5, 2420–2432, 2014.
- 101: Murphy, Kenneth P.; Travers, Paul; Walport, Mark; Janeway, Charles; Seidler, Lothar; Ehrenstein, Michael; Haüßer-Stiller, Ingrid: "*Janeway Immunologie*", Spektrum Akademischer Verlag , 7. Auflage, 2009.
- 102: Kawai, Taro; Akira, Shizuo: "*Signaling to NF-kappaB by Toll-like receptors*", *Trends in molecular medicine* 13, 11, 460–469, 2007.
- 103: Fujioka, S; Niu, J; Schmidt, C; Sclabas, GM.; Peng, B; Uwagawa, T; Li, Z; Evans, DB.; Abbruzzese, JL.; Chiao, PJ.: "*NF-kappaB and AP-1 connection-Mechanism of NF-kappaB-dependent regulation of AP-1 activity*", *Molecular and cellular biology* 24, 17, 7806–7819, 2004.
- 104: Mori, K.; Yanagita, M.; Hasegawa, S.; Kubota, M.; Yamashita, M.; Yamada, S.; Kitamura, M.; Murakami, S.: "*Necrosis-induced TLR3 Activation Promotes TLR2 Expression in Gingival Cells*", *Journal of dental research* 94, 8, 1149–1157, 2015.
- 105: Sojar, Hakimuddin T.; Sharma, Ashu; Genco, Robert J.: "*Porphyromonas gingivalis* Fimbriae Bind to Cytokeratin of Epithelial Cells", *IAI* 70, 1, 96–101, 2002.
- 106: Nagano, K.; Abiko, Y.; Yoshida, Y.; Yoshimura, F.: "*Genetic and antigenic analyses of Porphyromonas gingivalis* FimA fimbriae", *Mol oral Microbio* 28, 5, 392–403, 2013.
- 107: Martin, Michael; Rehani, Kunal; Jope, Richard S.; Michalek, Suzanne M.: "*Toll-like receptor-mediated cytokine production is differentially regulated by glycogen synthase kinase 3*", *Nat Immunol (Nature Immunology)* 6, 8, 777–784, 2005.
- 108: Campbell, J. J.: "*Chemokines and the Arrest of Lymphocytes Rolling Under Flow Conditions*", *Science* 279, 5349, 381–384, 1998.
- 109: Constantin G, Majeed M, Giagulli C, et al.: "*Chemokines trigger immediate beta2 integrin affinity and mobility changes: differential regulation and roles in lymphocyte arrest under flow*", *Immunity* , 13(6), 759–769., 2000.
- 110: Weaver, Casey T.; Hatton, Robin D.; Mangan, Paul R.; Harrington, Laurie E.: "*IL-17 Family Cytokines and the Expanding Diversity of Effector T Cell Lineages*", *Annu. Rev. Immunol.* 25, 1, 821–852, 2007.
- 111: Hiramane, Hiroko; Watanabe, Kiyoko; Hamada, Nobushiro; Umemoto, Toshio: "*Porphyromonas gingivalis* 67-kDa fimbriae induced cytokine production and osteoclast differentiation utilizing TLR2", *FEMS Microbiology Letters* 229, 1, 49–55, 2003.
- 112: Dyson, H. Jane; Wright, Peter E.: "*Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300*", *J. Biol. Chem.* 291, 13, 6714–6722, 2016.
- 113: Lamont, Richard J.; Hajishengallis, George: "*Polymicrobial synergy and dysbiosis in inflammatory disease*", *Trends in molecular medicine* 21, 3, 172–183, 2015.

- 114: Li, Chenghua; Li, Bei; Dong, Zhiwei; Gao, Li; He, Xiaoning; Liao, Li; Hu, Chenghu; Wang, Qintao; Jin, Yan: "*Lipopolysaccharide differentially affects the osteogenic differentiation of periodontal ligament stem cells and bone marrow mesenchymal stem cells through Toll-like receptor 4 mediated nuclear factor κ B pathway*", Stem cell research & therapy 5, 3, 67, 2014.
- 115: Al-Rasheed, Abdulaziz; Scheerens, Heleen; Srivastava, Apurva K.; Rennick, Donna M.; Tatakis, Dimitris N.: "*Accelerated alveolar bone loss in mice lacking interleukin-10 - Late onset*", Journal of Periodontal Research 39, 3, 194–198, 2004.
- 116: Page RC, Lantz MS, Darveau R, Jeffcoat M, Mancl L, Houston L, Braham P, Persson GR: "*Immunization of Macaca fascicularis against experimental periodontitis using a vaccine containing cysteine proteases purified from Porphyromonas gingivalis*", Oral Microbiology Immunology , 22, 162–168, 2007.
- 117: Lomovatskaya, L.A., Romanenko, A.S.: "*secretion Systems of Bacterial Phytopathogens and Mutualists (Review)*", Appl Biochem Microbiol , 56, 115–129, 2020.
- 118: Gorasia DG, Chreifi G, Seers CA, Butler CA, Heath JE, Glew MD, McBride MJ, Subramanian P, Kjær A, Jensen GJ, Veith PD, Reynolds EC: "*In situ structure and organisation of the type IX secretion system*", 1–37, 2020_B.
- 119: Leone, P., Roche, J., Vincent, M.S. et al.: "*Type IX secretion system PorM and gliding machinery GldM form arches spanning the periplasmic space*", Nature communications 9, 1, 429, 2018.
- 120: Wu, Guichun; Su, Panpan; Wang, Bo; Zhang, Yuqiang; Qian, Guoliang; Liu, Fengquan: "*Novel Insights into Tat Pathway in Xanthomonas oryzae pv. oryzae Stress Adaption and Virulence*", Phytopathology 107, 9, 1011–1021, 2017.
- 121: Desvaux, Mickaël: "*Contribution of holins to protein trafficking-Secretion, leakage or lysis?*", Trends in microbiology 20, 6, 259–261, 2012.
- 122: Emrizal, Reeki; Nor Muhammad, Nor Azlan: "*Identification of Sequence Motifs for the Protein Components of Type IX Secretion System*", JSM 47, 12, 2941–2950, 2018.
- 123: Pena RT, Blasco L, Ambroa A, et al. : "*Relationship Between Quorum Sensing and Secretion Systems*", Frontiers in microbiology 10, 1100, 2019.
- 124: Sato, Keiko; Sakai, Eiko; Veith, Paul D.; Shoji, Mikio; Kikuchi, Yuichiro; Yukitake, Hideharu; Ohara, Naoya; Naito, Mariko; Okamoto, Kuniaki; Reynolds, Eric C.; Nakayama, Koji: "*Identification of a new membrane-associated protein that influences transport/maturation of gingipains and adhesins of Porphyromonas gingivalis*", The Journal of biological chemistry 280, 10, 8668–8677, 2005.
- 125: Shoji M, Ratnayake DB, Shi Y, Kadowaki T, Yamamoto K, Yoshimura F, Akamine A, Curtis MA, Nakayama K.: "*Construction and characterization of a nonpigmented mutant of Porphyromonas gingivalis: cell surface polysaccharide as an anchorage for gingipains.*", Microbiology , 148(4), 183–1191, 2002.
- 126: Veith, Paul D.; Glew, Michelle D.; Gorasia, Dhana G.; Reynolds, Eric C.: "*Type IX secretion*", Molecular microbiology 106, 1, 35–53, 2017.
- 127: Sato, Keiko; Naito, Mariko; Yukitake, Hideharu; Hirakawa, Hideki; Shoji, Mikio; McBride, Mark J.; Rhodes, Ryan G.; Nakayama, Koji: "*A protein secretion system linked to bacteroidete gliding motility and pathogenesis*", Proceedings of the National Academy of Sciences of the United States of America 107, 1, 276–281, 2010.
- 128: Abby, Sophie S.; Cury, Jean; Guglielmini, Julien; Néron, Bertrand; Touchon, Marie; Rocha, Eduardo P. C.: "*Identification of protein secretion systems in bacterial genomes*", Scientific reports 6, 23080, 2016.
- 129: McBride, Mark J.; Zhu, Yongtao: "*Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes*", Journal of bacteriology 195, 2, 270–278, 2013.
- 130: Kita, Daichi; Shibata, Satoshi; Kikuchi, Yuichiro; Kokubu, Eitoyo; Nakayama, Koji; Saito, Atsushi; Ishihara, Kazuyuki: "*Involvement of the Type IX Secretion System in Capnocytophaga ochracea Gliding Motility and Biofilm Formation*", Applied and environmental microbiology 82, 6, 1756–1766, 2016.
- 131: Gorasia DG; Veith PD and Reynolds EC: "*The Type IX Secretion System: Advances in Structure, Function and Organisation*", Microorganisms , 8, 1173, 2020.
- 132: Emrizal, Reeki; Nor Muhammad, Nor Azlan: "*Phylogenetic comparison between Type IX Secretion System (T9SS) protein components suggests evidence of horizontal gene transfer*", PeerJ 8, e9019, 2020.
- 133: Lauber, Frédéric; Deme, Justin C.; Lea, Susan M.; Berks, Ben C.: "*Type 9 secretion system structures reveal a new protein transport mechanism*", Nature 564, 7734, 77–82, 2018.
- 134: McBride, Mark J.; Nakane, Daisuke: "*Flavobacterium gliding motility and the type IX secretion system*", Current Opinion in Microbiology 28, 72–77, 2015.
- 135: Naas AE, Solden LM, Norbeck AD, et al. : "*"Candidatus Paraporphyromonas polyenzymogenes" encodes multi-modular cellulases linked to the type IX secretion system*", Microbiome 6, 1, 44, 2018.
- 136: Kim, Hey-Min; Davey, Mary E.: "*Synthesis of ppGpp impacts type IX secretion and biofilm matrix*

- formation in *Porphyromonas gingivalis*", NPJ biofilms and microbiomes 6, 1, 5, 2020.
- 137: Naito, Mariko; Tominaga, Takashi; Shoji, Mikio; Nakayama, Koji: "*PGN_0297 is an essential component of the type IX secretion system (T9SS) in Porphyromonas gingivalis*", Microbiology and immunology 63, 1, 11–20, 2019.
- 138: Gorasia, DG.; Veith, PD.; Hanssen, EG.; Glew, MD.; Sato, K; Yukitake, H; Nakayama, K; Reynolds, EC: "*Structural Insights into the PorK and PorN Components of the Porphyromonas gingivalis Type IX Secretion System*", PLoS pathogens 12, 8, e1005820, 2016.
- 139: Chen YY, Peng B, Yang Q, Glew MD, Veith PD, Cross KJ, Goldie KN, Chen D, O'Brien-Simpson N, Dashper SG, Reynolds EC: "*The outer membrane protein LptO is essential for the O-deacylation of LPS and the co-ordinated secretion and attachment of A-LPS and CTD proteins in Porphyromonas gingivalis*", Molecular microbiology 79, 5, 1380–1401, 2011.
- 140: Glew MD, Veith PD, Peng B, et al.: "*PG0026 is the C-terminal signal peptidase of a novel secretion system of Porphyromonas gingivalis*", The Journal of biological chemistry 287, 29, 24605–24617, 2012.
- 141: Gorasia DG, Veith PD, Chen D, Seers CA, Mitchell HA, Chen YY, Glew MD, Dashper SG, Reynolds EC: "*Porphyromonas gingivalis Type IX Secretion Substrates Are Cleaved and Modified by a Sortase-Like Mechanism*", PLoS pathogens 11, 9, e1005152, 2015.
- 142: Kadowaki, Tomoko; Yukitake, Hideharu; Naito, Mariko; Sato, Keiko; Kikuchi, Yuichiro; Kondo, Yoshio; Shoji, Mikio; Nakayama, Koji: "*A two-component system regulates gene expression of the type IX secretion component proteins via an ECF sigma factor*", Scientific reports 6, 23288, 2016.
- 143: Vincent, Maxence S.; Durand, Eric; Cascales, Eric: "*The PorX Response Regulator of the Porphyromonas gingivalis PorXY Two-Component System Does Not Directly Regulate the Type IX Secretion Genes but Binds the PorL Subunit*", Frontiers in cellular and infection microbiology 6, 96, 2016.
- 144: Lasica AM, Goulas T, Mizgalska D, et al. : "*Structural and functional probing of PorZ, an essential bacterial surface component of the type-IX secretion system of human oral-microbiomic Porphyromonas gingivalis*", Scientific reports 6, 37708, 2016.
- 145: Dou, Yuetan; Osbourne, Devon; McKenzie, Rachelle; Fletcher, Hansel M.: "*Involvement of extracytoplasmic function sigma factors in virulence regulation in Porphyromonas gingivalis W83*", FEMS Microbiology Letters 312, 1, 24–32, 2010.
- 146: Heath JE, Seers CA, Veith PD, et al. : "*PG1058 Is a Novel Multidomain Protein Component of the Bacterial Type IX Secretion System*", PLoS ONE 11, 10, e0164313, 2016.
- 147: Vincent MS, Canestrari MJ, Leone P, et al.: "*Characterization of the Porphyromonas gingivalis Type IX Secretion Trans-envelope PorKLMNP Core Complex*", The Journal of biological chemistry 292, 8, 3252–3261, 2017.
- 148: Shoji, Mikio; Sato, Keiko; Yukitake, Hideharu; Kondo, Yoshio; Narita, Yuka; Kadowaki, Tomoko; Naito, Mariko; Nakayama, Koji: "*Por secretion system-dependent secretion and glycosylation of Porphyromonas gingivalis hemin-binding protein 35*", PLoS ONE 6, 6, e21372, 2011.
- 149: Kulkarni SS, Zhu Y, Brendel CJ, McBride MJ.: "*Diverse C-Terminal Sequences Involved in Flavobacterium johnsoniae Protein Secretion*", Journal of bacteriology , 199(12), 2017.
- 150: Vincent, Maxence S.; Chabalier, Maïalène; Cascales, Eric: "*A conserved motif of Porphyromonas Type IX secretion effectors C-terminal secretion signal specifies interactions with the PorKLMN core complex*", bioRxiv , 483123, 2018.
- 151: Shoji, Mikio; Sato, Keiko; Yukitake, Hideharu; Naito, Mariko; Nakayama, Koji: "*Involvement of the Wbp pathway in the biosynthesis of Porphyromonas gingivalis lipopolysaccharide with anionic polysaccharide*", Scientific reports 4, 5056, 2014.
- 152: Mattos-Graner, Renata O.; Duncan, Margaret J.: "*Two-component signal transduction systems in oral bacteria*", Journal of oral microbiology 9, 1, 1400858, 2017.
- 153: Casino, Patricia; Rubio, Vicente; Marina, Alberto: "*The mechanism of signal transduction by two-component systems*", Current opinion in structural biology 20, 6, 763–771, 2010.
- 154: Galperin, MY: "*Structural Classification of Bacterial Response Regulators*", JB 188, 12, 4169–4182, 2006/ "*Diversity of structure and function of response regulator output domains*", Curr Opin Microbiol. 13(2),150-159, 2010.
- 155: Goldfarb T, Sberro H.; Weinstock E; Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G; Sorek R: "*BREX is a novel phage resistance system widespread in microbial genomes*", EMBO J 34, 2, 169–183, 2014.
- 156: Chaudhary K: "*BacteRiophage EXclusion (BREX)*", J Cell Physiol 233, 2, 771–773, 2017.
- 157: Mohamed, MF, Hollfelder, F: "*Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer*", Biochimica et Biophysica Acta 1834, 1, 417–424, 2013.

- 158: Chinenova TA, Mkrtumian NM, Lomovskaia ND : "*Geneticheskaia kharakteristika novogo priznaka fagoustoichivosti u Streptomyces coelicolor A3(2) [Genetic characteristics of a new phage resistance trait in Streptomyces coelicolor A3(2)]*", Genetika 18, 12, 1945-1952 , 1982.
- 159: El-Gebali S, Mistry J, Bateman A, et al. : "*The Pfam protein families database in 2019*", Nucleic Acids Res 47, D1, D427-D432, 2019.
- 160: Jiang, C; Yang, D; Hua, T; Hua, Z; Kong, W; Shi, Y, A: "*PorX/PorY and σ^P Feedforward Regulatory Loop Controls Gene Expression Essential for Porphyromonas gingivalis Virulence*", mSphere 6, 2021.
- 161: Yang, Dezhi; Jiang, Chizhou; Ning, Bo; Kong, Wei; Shi, Yixin, "*The PorX/PorY system is a virulence factor of Porphyromonas gingivalis and mediates the activation of the type IX secretion system*", The Journal of biological chemistry 296, p100574, 2021.

7.2 Literature Materials&Methods

- 1: Foadi, J.; Aller, P.; Alguel, Y.; Cameron, A.; Axford, D.; et al., "*Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography*", Acta Cryst D 69, 8, p1617–1632, 2013.
- 2: McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, .C.; Read, R.J., "*Phaser crystallographic software*", J Appl Crystallogr 40, Pt 4, p658–674, 2007.
- 3: McCoy, Airlie J.; Storoni, Laurent C.; Read, Randy J., "*Simple algorithm for a maximum-likelihood SAD function*", Acta Cryst D 60, 7, p1220–1228, 2004.
- 4: Millán, C; Sammito, MD; McCoy, AJ; Nascimento, AFZ; Petrillo, G et al., "*Exploiting distant homologues for phasing through the generation of compact fragments, local fold refinement and partial solution combination*", Acta Crystallogr D 74, , p290-304, 2018.
- 5: Sammito, M; Meindl, K; Ilarduya, IM de; Millán, C.; Artola-Recolons, C. et al. , "*Structure solution with ARCIMBOLDO using fragments derived from distant homology models*", FEBS Journal 281, , p4029–4045, 2014.
- 6: Stein, N., "*CHAINSaw: a program for mutating pdb files used as templates in molecular replacement.*", J. Appl. Cryst. 41, , p641 - 643, 2008.
- 7: Vagin, A.&Teplakov, A. , "*MOLREP: an Automated Program for Molecular Replacement*", J. Appl. Cryst. 30, , p1022-1025, 1997.
- 8: Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P. et al., "*Overview of the CCP 4 suite and current developments*", Acta Cryst D 67, 4, p235–242, 2011.
- 9: Cowtan, Kevin, "*The Buccaneer software for automated model building. 1. Tracing protein chains*", Acta Cryst D 62, 9, p1002–1011, 2006.
- 10: Afonine, P.V.; Grosse-Kunstleve, R.W.; Echols, N.; Headd, J.J.; Moriarty, N.W. et al. , "*Towards automated crystallographic structure refinement with phenix.refine*", Acta Cryst D 68, 4, p352–367, 2012.
- 11: Liebschner, D.; Afonine, P.V.; Baker, M.L.; Bunkóczi, G.; Chen, V.B. et al., "*Macromolecular structure determination using X-rays, neutrons and electrons, Recent developments in Phenix*", Acta Cryst D 75, 10, p861–877, 2019.
- 12: Terwilliger, T.C.; Grosse-Kunstleve, R.W.; Afonine, P.V.; Moriarty, N.W.; Zwart, P.H.; Hung, L.W.; Read, R.J.; Adams, P.D., "*Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard*", Acta Cryst D 64, Pt 1, p61–69, 2008.
- 13: Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P et al., "*BUSTER*", Global Phasing Ltd., 2017.
- 14: Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., et al. , "*STARANISO*", Global Phasing Ltd.: <https://www.globalphasing.com/>, 2018.
- 15: Sheldrick, George M., "*A short history of SHELX*", Acta Cryst A 64, 1, p112–122, 2008.
- 16: Pettersen, EF.; Goddard, TD.; Huang, CC.; Couch, GS.; Greenblatt, DM.; et al., "*UCSF Chimera-A visualization system for exploratory research and analysis*", J. Comput. Chem. (Journal of Computational Chemistry) 25, 13, p1605–1612, 2004.
- 17: Goddard, Thomas D.; Huang, Conrad C.; Ferrin, Thomas E., "*Visualizing density maps with UCSF Chimera*", Journal of structural biology 157, 1, p281–287, 2007.
- 18: Schrödinger, L., & DeLano, W., "*PyMOL*", The PyMOL Molecular Graphics System, 2020.
- 19: Emsley, P.; Cowtan, K., "*Coot, Model-building tools for molecular graphics*", Acta Cryst D 60, Pt 12 Pt 1, p2126–2132, 2004.
- 20: Gabadinho, J.; Beteva, A.; Guijarro, M.; Rey-Bakaikoa, V.; Spruce, D. et al., "*MxCuBE, A synchrotron beamline control environment customized for macromolecular crystallography experiments*", Journal of synchrotron radiation 17, 5, p700–707, 2010.
- 21: Oscarsson, M.; Beteva, A.; Flot, D.; Gordon, E.; Guijarro, M. et al., "*MXCuBE2, The dawn of MXCuBE*

- Collaboration*", Journal of synchrotron radiation 26, Pt 2, p393–405, 2019.
- 22: Arvai, A. , "Adxv – A Program to Display X-ray Diffraction Images", <https://www.scripps.edu/tainer/arvai/adxv.html>, 2013.
- 23: Kabsch, Wolfgang, "XDS/Integration, scaling, space-group assignment and post-refinement", Acta Cryst D 66, Pt 2, p125–132/133–144, 2010.
- 24: Hopkins, Jesse Bennett; Gillilan, Richard E.; Skou, Soren, "BioXTAS RAW", J Appl Crystallogr 50, Pt 5, p1545–1553, 2017.
- 25: Franke, Daniel; Jeffries, Cy M.; Svergun, Dmitri I., "Machine Learning Methods for X-Ray Scattering Data Analysis from Biomacromolecular Solutions", Biophysical journal 114, 11, p2485–2492, 2018.
- 26: Hajizadeh, Nelly R.; Franke, Daniel; Jeffries, Cy M.; Svergun, Dmitri I., "Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data", Scientific reports 8, 1, p7204, 2018.
- 27: Rambo, Robert P.; Tainer, John A., "Accurate assessment of mass, models and resolution by small-angle scattering", Nature 496, 7446, p477–481, 2013.
- 28: Piiadov, V.; Ares de Araújo, E.; Oliveira Neto, M.; Craievich, A.; Polikarpov, I., "SAXSMoW 2.0, Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale", Protein science 28, 2, p454–463, 2019.
- 29: Manalastas-Cantos, K.; Konarev, P.V.; Hajizadeh, N.R.; Kikhney, A.G.; Petoukhov, M.V. et al., "ATSAS 3.0, Expanded functionality and new tools for small-angle scattering data analysis", J Appl Crystallogr 54, Pt 1, p343–355, 2021.
- 30: Svergun, D. I., "Determination of the regularization parameter in indirect-transform methods using perceptual criteria", J Appl Crystallogr 25, 4, p495–503, 1992.
- 31: Grant, Thomas D., "Ab initio electron density determination directly from solution scattering data", Nature methods 15, 3, p191–193, 2018.
- 32: Panjkovich, Alejandro; Svergun, Dmitri I., "Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis", PCCP 18, 8, p5707–5719, 2016.
- 33: Svergun, D.; Barberato, C.; Koch, M. H. J., "CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates", J Appl Crystallogr 28, 6, p768–773, 1995.
- 34: Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R et al., "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.", Science , 239(4839), p487–491, 1988.
- 35: Berg, Jeremy M.; Tymoczko, John L.; Stryer, Lubert, "Biochemie, 5. Aufl.", Spektrum Akad. Verl., 2003.
- 36: Aaij, C.; Borst, P., "The gel electrophoresis of DNA", Biochimica et Biophysica Acta (BBA) 269, 2, p192–200, 1972.
- 37: Olivera, Baldomero M.; Baine, Peter; Davidson, Norman, "Electrophoresis of the nucleic acids", Biopolymers 2, 3, p245–257, 1964.
- 38: QIAGEN, "QIAquick® Spin Handbook", QIAGEN online: <https://www.qiagen.com/>, 2000.
- 39: Nobelprize in, "Physiology or Medicine", Nobelprize.org, Official web site of the Nobel Foundation , 1978.
- 40: Lehman, I. R., "DNA ligase, Structure, mechanism, and function", Science 186, 4166, p790–797, 1974.
- 41: Gibson, D.G.; Young, L.; Chuang, R.Y.; Venter, J.C.; Hutchison, C.A.; Smith, H.O., "Enzymatic assembly of DNA molecules up to several hundred kilobases", Nature methods 6, 5, p343–345, 2009.
- 42: NEB, "Improved methods for site directed mutagenesis using NEBuilder HiFi DNA Assembly Master Mix", NEB online: <https://international.neb.com/>, 2020.
- 43: Sanger F, Nicklen S, Coulson AR., "DNA sequencing with chain-terminating inhibitors.", Proc Natl Acad Sci U S A. , 74(12), p5463–5467, 1977.
- 44: Shendure, J.; Balasubramanian, S.; Church, G.M.; Gilbert, W.; Rogers, J. et al. , "DNA sequencing at 40, Past, present and future", Nature 550, 7676, p345–353, 2017.
- 45: Hanahan, D., "Studies on transformation of Escherichia coli with plasmids", Journal of Molecular Biology , 166,4, p557–580, 1983.
- 46: Green, Rachel; Rogers, Elizabeth J., "Transformation of chemically competent E. coli", Methods in enzymology 529, , p329–336, 2013.
- 47: Studier FW, Moffatt BA., "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.", J Mol Biol. , 189(1), p113–130, 1986.
- 48: Studier, F.W.; Daegelen, P.; Lenski, R.E.; Maslov, S.; Kim, J.F., "Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes", Journal of Molecular Biology 394, 4, p653–680, 2009.
- 49: VWR product informations, "pGEX Vectors, Cytiva | VWR", VWR online: <https://www.vwr.com/>, 2021.
- 50: LERMAN, L. S., "Antibody Chromatography on an Immunologically Specific Adsorbent", Nature 172,

4379, p635–636, 1953.

51: Schäfer, F.; Seip, N.; Maertens, B.; Block, H.; Kubicek, J., "*Purification of GST-Tagged Proteins*", Elsevier, 2015.

52: Simons, Peter C.; Vander Jagt, David L., "*Purification of glutathione S-transferases from human liver by glutathione-affinity chromatography*", Analytical Biochemistry 82, 2, p334–341, 1977.

53: Peterson, E.A.&Sober, H.A., "*Chromatography of Proteins. I. Cellulose Ion-exchange Adsorbents*", J. Am. Chem. Soc. 78, 4, p751–755, 1956.

54: Lathe, G. H.; Ruthven, C. R. J., "*The separation of substances and estimation of their relative molecular sizes by the use of columns of starch in water*", Biochemical Journal 62, 4, p665–674, 1956.

55: Cytiva, "*Size Exclusion Chromatography, Handbook*", Cytiva online: <https://www.cytivalifesciences.com/>, 2021.

56: Lukat GS, McCleary WR, Stock AM, Stock JB, "*Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors.*", Proc Natl Acad Sci U S A. , 89(2), p718–722, 1992.

57: McCleary, W. R.; Stock, J. B., "*Acetyl phosphate and the activation of two-component response regulators*", Journal of Biological Chemistry 269, 50, p31567–31572, 1994.

58: Yan, D.; Cho, H. S.; Hastings, C.A.; Igo, M. M.; Lee, S.-Y. Et al, "*Beryll fluoride mimics phosphorylation of NtrC and other bacterial response regulators*", Proceedings of the National Academy of Sciences 96, 26, p14789–14794, 1999.

59: Merck, "*Protein Sample Ultrafiltration, Amicon® Filter*", Merck online: <https://www.merckgroup.com/>, 2020.

60: van Reis, Robert; Zydney, Andrew L., "*Protein Ultrafiltration*", Encyclopedia of Industrial Biotechnology, 2010.

61: Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R. et al. , "*Protein Identification and Analysis Tools on the ExPASy Server*", Walker J.M. (eds) The Proteomics Protocols Handbook. Springer Protocols Handbooks., p571–607, 2005.

62: LAEMMLI, U. K., "*Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4*", Nature 227, 5259, p680–685, 1970.

63: Atkins, Peter W. , "*Physikalische Chemie, 3. korrigierte Aufl.*", Wiley-VCH-Verl., 2001.

64: ThermoScientific, "*NanoDrop 1000 Spectrophotometer, V3.8 User's Manual*", ThermoScientific online: <https://www.thermofisher.com/>, 2010.

65: Some, D.; Amartely, H.; Tsadok, A.; Lebendiker, M., "*Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS)*", Journal of visualized experiments : JoVE , 148, 2019.

66: Some, D.; Kenrick, S. , "*Characterization of Protein-Protein Interactions via Static and Dynamic Light Scattering*", Cai 2012 – Protein Interactions, p402-426, 2012.

67: Zimm, Bruno H., "*The Scattering of Light and the Radial Distribution Function of High Polymer Solutions*", The Journal of Chemical Physics 16, 12, p1093–1099, 1948.

68: Wyatt, Philip J., "*Light scattering and the absolute characterization of macromolecules*", Analytica Chimica Acta 272, 1, p1–40, 1993.

69: Hünefeld, Friedrich Ludwig, "*Der Chemismus in der thierischen Organisation*", Brockhaus, 1840.

70: McPherson, A.; Gavira, J. A., "*Introduction to protein crystallization*", Acta Cryst F 70, Pt 1, p2–20, 2014.

71: Chayen, Naomi E.; Saridakis, Emmanuel, "*Protein crystallization, Techniques, strategies, and tips ; a laboratory manual*", Nature methods 5, 2, p147–153, 2008.

72: Bergfors, Terese M., "*Protein crystallization, Techniques, strategies, and tips ; a laboratory manual*", International University Line 1, , 1999.

73: McCoy, Airlie J.; Read, Randy J., "*Experimental phasing, Best practice and pitfalls*", Acta Cryst D 66, Pt 4, p458–469, 2010.

74: Winter, R.; Noll, F., "*Methoden der biophysikalischen Chemie*", Teubner-Studienbücher : Chemie, 1998.

75: Rhodes, Gale, "*Crystallography made crystal clear, A guide for users of macromolecular models, 3. ed.*", Acad. Press , 2007.

76: He, B.B; Preckwinkel, U. and Smith K.L., "*FUNDAMENTALS OF TWO-DIMENSIONAL X-RAY DIFFRACTION (XRD2)*", Advances in X-Ray Analysis , 43, p273-280, 2000.

77: Friedrich, W.; Knipping, P.; Laue, M., "*Interferenzerscheinungen bei Röntgenstrahlen, Theoretischer Teil von M. Laue*", Ann. Phys. 346, 10, p971–988, 1913.

78: Laue, Max von, "*Eine quantitative Prüfung der Theorie für die Interferenz-Erscheinungen bei Röntgenstrahlen.*", Sitzungsberichte der mathematisch-physikalischen Klasse der K.B. Akademie der Wissenschaften zu München , II, p363–378, 1912.

- 79: BRAGG, W. H. & Bragg, W. L., "*The Reflection of X-Rays by Crystals*", Nature 91, 2280, p477, 1913.
- 80: DiVincenzo, D.P., "*Interaction of X-rays, Neutrons and Electrons with Matter*", Forschungszentrum Jülich Bd. 33, , 2012.
- 81: Blundell, T. L.; Johnson, L. N., "*Protein crystallography, 4th printing*", Academic Press, 1994.
- 82: Diederichs, K., "*Space group determination*", XDSwiki-Online resource for XDS users: <https://strucbio.biologie.uni-konstanz.de/xdswiki/>, 2022.
- 83: Evans, Philip, "*Scaling and assessment of data quality*", Acta Cryst D 62, Pt 1, p72–82, 2006.
- 84: Assmann, G.; Brehm, W.; Diederichs, K., "*Identification of rogue datasets in serial crystallography*", J. Appl. Cryst. 49, , p1021-1028, 2016.
- 85: Karplus, P. Andrew; Diederichs, Kay, "*Assessing and maximizing data quality in macromolecular crystallography*", Current opinion in structural biology 34, , p60–68, 2015.
- 86: Krojer, T.; Delft von, F., "*Assessment of radiation damage behaviour in a large collection of empirically optimized datasets highlights the importance of unmeasured complicating effects*", J Synchrotron Radiat. 18, 3, p387-97 , 2011.
- 87: Walden, Helen, "*Selenium incorporation using recombinant techniques*", Acta Cryst D 66, Pt 4, p352–357, 2010.
- 88: Wang, Jimin; Li, Yue; Modis, Yorgo, "*Exploiting subtle structural differences in heavy-atom derivatives for experimental phasing*", Acta Cryst D 70, Pt 7, p1873–1883, 2014.
- 89: Read, Randy J.; McCoy, Airlie J., "*Maximum-likelihood determination of anomalous substructures*", Acta Cryst D 74, Pt 2, p98–105, 2018.
- 90: Langer, G.; Cohen, S.X.; Lamzin, V.S.; Perrakis, A., "*Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7*", Nature protocols 3, 7, p1171–1179, 2008.
- 91: McRee, Duncan Everett; David, Peter R., "*Practical protein crystallography, 2nd ed.*", Academic Press, 1999.
- 92: Mertens, H.D.; Svergun, D.I., "*Structural characterization of proteins and complexes using small-angle X-ray solution scattering*", Journal of structural biology 172, 1, p128–141, 2010.
- 93: Svergun DI & Koch, MHJ, "*Small-angle scattering studies of biological macromolecules in solution*", Reports on Progress in Physics 66, 10, p1735, 2003.
- 94: Skou, Soren; Gillilan, Richard E.; Ando, Nozomi, "*Synchrotron-based small-angle X-ray scattering of proteins in solution*", Nature protocols 9, 7, p1727–1739, 2014.
- 95: Gommès, Cedric J.; Jaksch, Sebastian; Frielinghaus, Henrich, "*Small-angle scattering for beginners*", J Appl Crystallogr 54, Pt 6, p1832–1843, 2021.
- 96: Doniach, S., "*Changes in biomolecular conformation seen by small angle X-ray scattering*", Chemical reviews 101, 6, p1763–1778, 2001.
- 97: Choi, K.H. & Morais M., "*Use of Small-Angle X-ray Scattering to Investigate the Structure and Function of Dengue Virus NS3 and NS5*", Methods Mol. Biol. 1138, , p241-252, 2014.
- 98: Glatter, O.; Kratky, O., "*Small angle X-ray scattering*", Acad. Pr, 1982.
- 99: Guinier, André, "*La diffraction des rayons X aux très petits angles*", Ann. Phys. (Annalen der Physik) 11, 12, p161–237, 1939.
- 100: Gommès, Cedric J.; Jaksch, Sebastian; Frielinghaus, Henrich, "*Small-angle scattering for beginners*", Journal of applied crystallography 54, Pt 6, p1832–1843, 2021.
- 101: Piiadov, V.; Ares de Araújo, E.; Oliveira Neto, M.; Craievich, A.; Polikarpov, I., "*SAXSMoW 2.0, Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale*", Protein science : a publication of the Protein Society 28, 2, p454–463, 2019.
- 102: Rambo, Robert P.; Tainer, John A., "*Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law*", Biopolymers 95, 8, p559–571, 2011.
- 103: Mylonasa, E. & Svergun D.I., "*Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering*", J. Appl. Cryst. 40, , p245-249, 2007.
- 104: Orthaber, Doris; Bergmann, Alexander; Glatter, Otto, "*SAXS experiments on absolute scale with Kratky systems using water as a secondary standard*", J Appl Crystallogr (Journal of Applied Crystallography) 33, 2, p218–225, 2000.
- 105: Fischer, H.; Oliveira Neto, M. de; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F., "*Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale*", J Appl Crystallogr (Journal of Applied Crystallography) 43, 1, p101–109, 2010.
- 106: Korasick, David A.; Tanner, John J., "*Determination of protein oligomeric structure from small-angle X-ray scattering*", Protein science : a publication of the Protein Society 27, 4, p814–824, 2018.

- 107: Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M. et al., "*Highly accurate protein structure prediction with alphafold.*", *Nature* 596, , p583–589, 2021.
- 108: Drozdetskiy, A., Cole, C., Procter, J. and Barton, G.J., "*JPred4: a protein secondary structure prediction server.*", *Nucleic Acids Res.* 43, , p389–394, 2015.
- 109: Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J., "*Jalview version 2—a multiple sequence alignment editor and analysis workbench.*", *Bioinformatics* 25, , p1189–1191, 2009.
- 110: Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T. et al., "*The EMBL-EBI search and sequence analysis tools APIs in 2019.*", *Nucleic Acids Res.* 47, , p636–641, 2019.
- 111: Huson, D.H. and Scornavacca, C., "*Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks.*", *Syst. Biol.* 61, , p1061–1067, 2012.
- 112: Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A. et al., "*Clustal w and clustal x version 2.0.*", *Bioinformatics* 23, , p2947–2948, 2007.

7.3 Literature Results, Discussion, Prospects

0: Claus Schmitz, Mariusz Madej, Zuzanna Nowakowska, Anna Cuppari, Anna Jacula, Mirosław Ksiazek, Katarzyna Mikruta, Jerzy Wisniewski, Natalia Pudelko-Malik, Anshu Saran, Natalie Zeytuni, Piotr Mlynarz, Richard J. Lamont, Isabel Usón, Virginijus Siksnys, Jan Potempa, Maria Solà: "*Response regulator PorX coordinates oligonucleotide signalling and gene expression to control the secretion of virulence factors.*", *Nucleic acids research*, 2022;

- 1: Sunden F, AlSadhan I, Lyubimov A, Doukov T, Swan J, Herschlag D., "*Differential catalytic promiscuity of the alkaline phosphatase superfamily bimetallo core reveals mechanistic features underlying enzyme evolution.*", *J Biol Chem.* 292, 51, p20960–20974, 2017.
- 2: Zalatan, Jesse G.; Fenn, Timothy D.; Brunger, Axel T.; Herschlag, Daniel, "*Structural and Functional Comparisons of Nucleotide Pyrophosphatase/Phosphodiesterase and Alkaline Phosphatase*", *Biochemistry* 45, 32, p9788–9803, 2006.
- 3: Mohamed, MF.; Hollfelder, F., "*Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer*", *Biochimica et biophysica acta* 1834, 1, p417–424, 2013.
- 4: Barbieri, CM.; Mack, TR; Robinson, Vi.; Miller, MT.; Stock, AM, "*Regulation of response regulator autophosphorylation through interdomain contacts*", *The Journal of biological chemistry* 285, 42, p32325–32335, 2010.
- 5: Gao, Rong; Bouillet, Sophie; Stock, Ann M., *Structural Basis of Response Regulator Function*, 2019
- 6: Vincent, Maxence S.; Durand, Eric; Cascales, Eric, "*The PorX Response Regulator of the Porphyromonas gingivalis PorXY Two-Component System Does Not Directly Regulate the Type IX Secretion Genes but Binds the PorL Subunit*", *Frontiers in cellular and infection microbiology* 6, , p96, 2016.
- 7: Lukat, G. S.; McCleary, W. R.; Stock, A. M.; Stock, J. B., "*Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors*", *Proc. Natl. Acad. Sci. U.S.A.* (Proceedings of the National Academy of Sciences) 89, 2, p718–722, 1992.
- 8: Yan, D; Cho, HS; Hastings, CA; Igo, MM; Lee, Syet al. , "*Beryll fluoride mimics phosphorylation of NtrC and other bacterial response regulators*", *Proc. Natl. Acad. Sci. U.S.A.* (Proceedings of the National Academy of Sciences) 96, 26, p14789–14794, 1999.
- 9: Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D., "*Two-state allosteric behavior in a single-domain signaling protein*", *Science (New York, N.Y.)* 291, 5512, p2429–2433, 2001.
- 10: Corrêa, Fernando; Gardner, Kevin H., *Basis of Mutual Domain Inhibition in a Bacterial Response Regulator*, 2016
- 11: Zhu, X.; Rebello, J.; Matsumura, P.; Volz, K., "*Crystal structures of CheY mutants Y106W and T87I/Y106W. CheY activation correlates with movement of residue 106*", *Journal of Biological Chemistry* 272, 8, p5000–5006, 1997.
- 12: Cho, HS; Lee, SY; Yan, D; Pan, X; Parkinson, JS; Kustu, S; Wemmer, DE; Pelton, JG, *NMR structure of activated CheY*, 2000
- 13: Barbieri, CM; Wu, T; Stock, AM, *Comprehensive analysis of OmpR phosphorylation, dimerization, and DNA binding supports a canonical model for activation*, 2013
- 14: Galperin, Michael Y.; Koonin, Eugene V., *Divergence and convergence in enzyme evolution*, 2012
- 15: Kim, E. E.; Wyckoff, H. W., *Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis*, 1991
- 16: Pabis, Anna; Kamerlin, Shina Caroline Lynn, "*Promiscuity and electrostatic flexibility in the alkaline phosphatase superfamily*", *Current opinion in structural biology* 37, , p14–21, 2016.
- 17: Kim, A; Benning, MM; OkLee, S; Quinn, J Martin, BM; Holden, HM; Dunaway-Mariano, D, *Divergence of*

chemical function in the alkaline phosphatase superfamily, 2011.

- 18: Sundén, F.; AlSadhan, I.; Lyubimov, AY; Ressler, S.; Wiersma-Koch, H et al., "*Mechanistic and Evolutionary Insights from Comparative Enzymology of Phosphomonoesterases and Phosphodiesterases across the Alkaline Phosphatase Superfamily*", Journal of the American Chemical Society 138, 43, p14273–14287, 2016.
- 19: Clair, T.; Lee, H. Y.; Liotta, L. A.; Stracke, M. L., Autotaxin is an exoenzyme possessing 5'-nucleotide phosphodiesterase/ATP pyrophosphatase and ATPase activities, 1997
- 20: Oda, Y.; Kuo, M. D.; Huang, S. S.; Huang, J. S., "*The major acidic fibroblast growth factor (aFGF)-stimulated phosphoprotein from bovine liver plasma membranes has aFGF-stimulated kinase, autoadenylation, and alkaline nucleotide phosphodiesterase activities*", Journal of Biological Chemistry 268, 36, p27318–27326, 1993.
- 21: Bachhawat, P; Swapna, GVT; Montelione, GT.; Stock, AM, Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states, 2005
- 22: Solá, M.; Gomis-Rüth, F. X.; Serrano, L.; González, A.; Coll, M., "*Three-dimensional crystal structure of the transcription factor PhoB receiver domain*", Journal of molecular biology 285, 2, p675–687, 1999.
- 23: Kato, K; Nishimasu, H; Oikawa, D; Hirano, S; Hirano, H et al. , Structural insights into cGAMP degradation by Ecto-nucleotide pyrophosphatase phosphodiesterase 1, 2018
- 24: Athukoralage, JS; Rouillon, C; Graham, S; Grünschow, S; White, MF, Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate, 2018
- 25: Chu, Lee-Ya; Agrawal, Sashank; Chen, Yi-Ping; Yang, Wei-Zen; Yuan, Hanna S., "*Structural insights into nanoRNA degradation by human Rexo2*", RNA (New York, N.Y.) 25, 6, p737-746, 2019.
- 26: Chaudhary, Kulbhushan, Bacteriophage EXclusion (BREX), 2018
- 27: Goldfarb, T; Sberro, H; Weinstock, E; Cohen, O; Doron, Sy et al., BREX is a novel phage resistance system widespread in microbial genomes, 2015
- 28: Duncan-Lowey, Brianna; Kranzusch, Philip J., CBASS phage defense and evolution of antiviral nucleotide signaling, 2022
- 29: Kazlauskienė, M; Kostiuik, G; Venclovas, C; Tamulaitis, G; Siksnys, V, A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems, 2017
- 30: Zaworski, J; Dagva, O; Brandt, J; Baum, C; Ettwiller, L et al. , "*Reassembling a cannon in the DNA defense arsenal*", PLoS genetics 18, 4, pe1009943, 2022.
- 31: Picton, DM; Harling-Lee, JD; Duffner, SJ; Went, SC; Morgan, RD; Hinton, JCD; Blower, TR, "*A widespread family of WYL-domain transcriptional regulators co-localizes with diverse phage defence systems and islands*", Nucleic acids research 50, 9, p5191–5207, 2022.
- 32: Picton, DM.; Luyten, YA; Morgan, RD; Nelson, A; Smith, DL et al., "*The phage defence island of a multidrug resistant plasmid uses both BREX and type IV restriction for complementary protection from viruses*", Nucleic acids research 49, 19, p11257–11273, 2021.
- 33: Sandmeier, H.; Bär, K.; Meyer, J., "*Search for bacteriophages of black-pigmented gram-negative anaerobes from dental plaque*", FEMS immunology and medical microbiology 6, 2-3, p193–194, 1993.
- 34: Luyten, YA.; Hausman, DE; Young, JC; Doyle, LA; Higashi, KM et al. , "*Identification and characterization of the WYL BrxR protein and its gene as separable regulatory elements of a BREX phage restriction system*", Nucleic acids research 50, 9, p5171–5190, 2022.
- 35: Goldman, SR.; Sharp, JS; Vvedenskaya, IO.; Livny, J; Dove, L.; Nickels, BE, NanoRNAs prime transcription initiation in vivo, 2011
- 36: Valentini, Martina; Filloux, Alain, "*Biofilms and Cyclic di-GMP (c-di-GMP) Signaling*", The Journal of biological chemistry 291, 24, p12547–12555, 2016.
- 37: Cohen, D; Mechold, U; Nevenzal, H; Yarmiyahu, Y; Randall, TE et al., Oligoribonuclease is a central feature of cyclic diguanylate signaling in Pseudomonas aeruginosa, 2015
- 38: Orr, MW; Donaldson, GP; Severin, GB; Wang, J; Sintim, HO; Waters, CM; Lee, VT, "*Oligoribonuclease is the primary degradative enzyme for pGpG in Pseudomonas aeruginosa that is required for cyclic-di-GMP turnover*", Proceedings of the National Academy of Sciences of the United States of America 112, 36, pE5048-57, 2015.
- 39: Kadowaki, T; Yukitake, H; Naito, M; Sato, K; Kikuchi, Y; Kondo, Y; Shoji, M; Nakayama, K, A two-component system regulates gene expression of the type IX secretion component proteins via an ECF sigma factor, 2016
- 40: Jiang, C; Yang, D; Hua, T; Hua, Z; Kong, W; Shi, Y, "*A PorX/PorY and ?P Feedforward Regulatory Loop Controls Gene Expression Essential for Porphyromonas gingivalis Virulence*", 2021,
- 41: Yang, Dezhi; Jiang, Chizhou; Ning, Bo; Kong, Wei; Shi, Yixin, "*The PorX/PorY system is a virulence*

factor of Porphyromonas gingivalis and mediates the activation of the type IX secretion system", The Journal of biological chemistry 296, , p100574, 2021.

42: Sato, K; Naito, M; Yukitake, H; Hirakawa, H; Shoji, M et al., "A protein secretion system linked to bacteroidete gliding motility and pathogenesis", Proceedings of the National Academy of Sciences of the United States of America 107, 1, p276–281, 2010.

43: Gordeeva, J; Morozova, N; Sierro, N; Isaev, A; Sinkunas, T et al., "BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site.", 2019

44: Outten, C. E.; O'Halloran, T. V., "Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis", Science (New York, N.Y.) 292, 5526, p2488–2492, 2001.

7.4 Data, model and refinement statistics

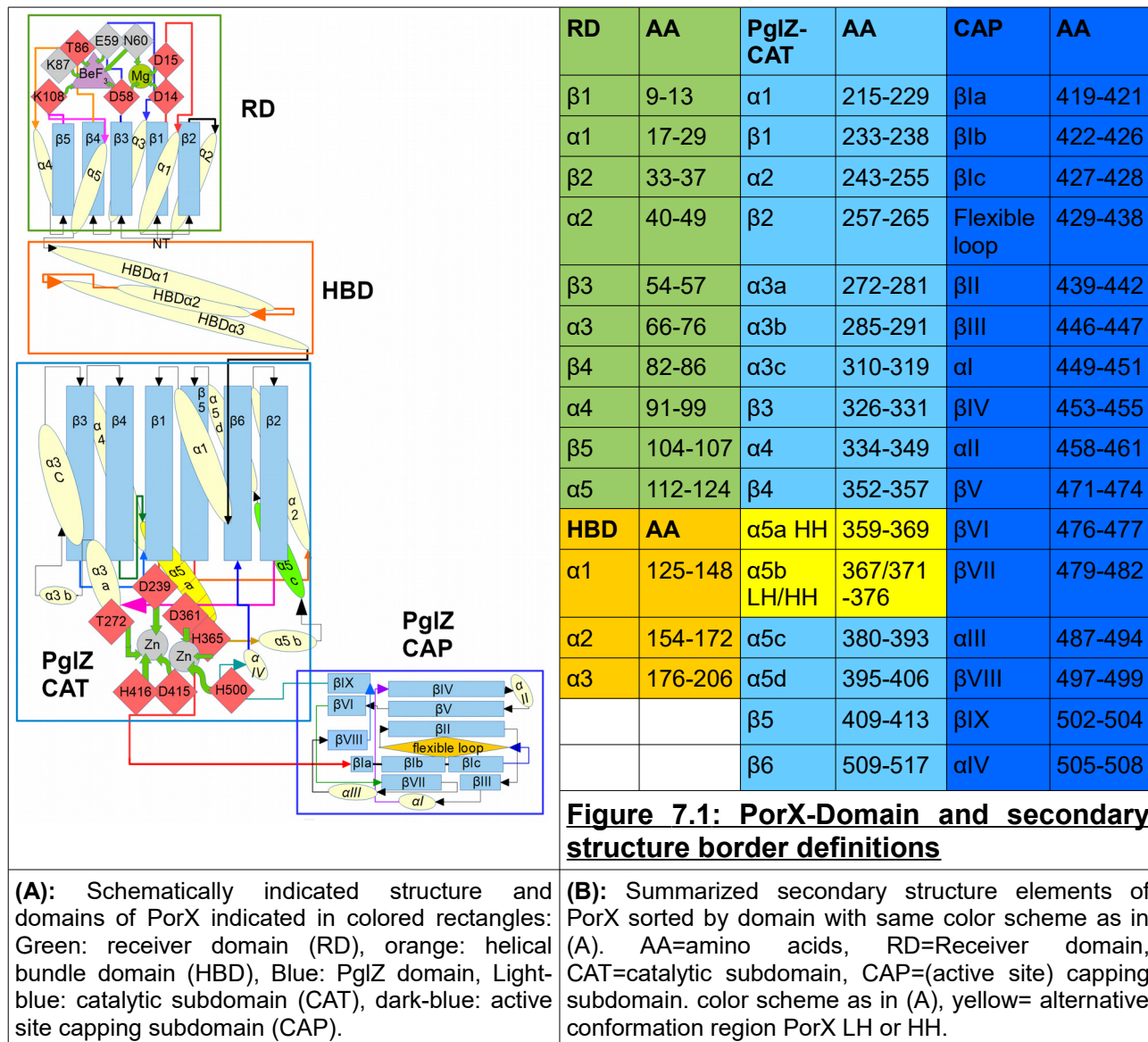
TABLE 7.1: published structures	JG5 - pdb: 7PV7	HR - pdb: 7PVA	pGpG-pdb: 7PVK
Wavelength [Å]	0.97893	1.28202	0.97923
Resolution range	68.4 - 2.41 (2.50 - 2.41)	68.59 - 1.91 (1.98 - 1.91)	51.9 - 2.40 (2.49 - 2.40)
Space group	P 1	P 1	P 1 21 1
Unit cell	61.8 77.1 138.7 82.3 81.4 76.4	61.5 77.3 138.8 82.4 81.6 76.5	94.6 103.8 132.2 90.0 98.5 90.0
Total reflections	335929 (31630)	3157066 (314354)	1419719 (139958)
Unique reflections	92407 (8754)	186912 (18467)	98575 (9713)
Multiplicity	3.6 (3.6)	16.9 (17.0)	14.4 (14.3)
Completeness (%)	97.91 (93.11)	98.30 (96.75)	99.39 (98.28)
Mean I/sigma(I)	6.45 (0.99)	12.29 (1.24)	15.93 (0.90)
Wilson B-factor	49.34	40.57	65.79
R-merge	0.1264 (0.8829)	0.1495 (2.903)	0.1222 (3.498)
R-meas	0.149 (1.039)	0.1542 (2.991)	0.1267 (3.626)
R-pim	0.07766 (0.5407)	0.03758 (0.7175)	0.03333 (0.9475)
CC1/2	0.992 (0.613)	0.998 (0.609)	0.999 (0.635)
CC*	0.998 (0.872)	1 (0.87)	1 (0.882)
Reflections used in refinement	92121 (8743)	186867 (18463)	98355 (9684)
Reflections used for R-free	971 (82)	1877 (183)	1041 (120)
R-work	0.1857 (0.3277)	0.1588 (0.2999)	0.1937 (0.3800)
R-free	0.2155 (0.3419)	0.1920 (0.3080)	0.2230 (0.4137)
CC(work)	0.963 (0.761)	0.971 (0.773)	0.966 (0.701)
CC(free)	0.955 (0.629)	0.963 (0.771)	0.954 (0.579)
Number of non-hydrogen atoms	18453	19606	17800
macromolecules	17144	17604	17188
ligands	143	171	156
solvent	1166	1831	456
Protein residues	2054	2033	2052
RMS(bonds)	0.004	0.016	0.005
RMS(angles)	1.06	1.50	0.98
Ramachandran favored (%)	96.68	97.92	96.38
Ramachandran allowed (%)	3.32	2.08	3.62
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	1.52	1.02	1.17
Clashscore	4.78	7.56	2.33
Average B-factor	61.04	58.76	89.35
macromolecules	61.20	57.75	89.37
ligands	75.33	77.36	112.28
solvent	56.84	66.73	80.85

Statistics for the highest-resolution shell are shown in parentheses.

TABLE 7.2: supporting structures	GTPγS	c-di-GMP	D58A
Wavelength	Zn absorption edge	Zn absorption edge	Zn absorption edge
Resolution range	62.91 - 2.82 (2.921 - 2.82)	82 - 2.91 (3.02 - 2.91)	61.36 - 2.40 (2.48 - 2.40)
Space group	P 1	P 21 21 21	P 31 2 1
Unit cell	62.0 82.2 127.5 89.6 80.8 80.4	61.6 149.2 294.5 90.0 90.0 90.0	88.8 88.8 203.8 90.0 90.0 120.0
Total reflections	207983 (19939)	896943 (61956)	1427740 (69018)
Unique reflections	58102 (5759)	53508 (4286)	37063 (3365)
Multiplicity	3.6 (3.5)	16.8 (14.0)	38.5 (20.5)
Completeness (%)	97.84 (93.55)	87.82 (72.29)	98.30 (83.80)
Mean I/sigma(I)	9.04 (0.65)	5.44 (1.14)	14.72 (0.69)
Wilson B-factor	81.55	82.24	68.04
R-merge	0.13 (2.479)	0.4094 (2.17)	0.1875 (4.118)
R-meas	0.1538 (2.944)	0.4225 (2.244)	0.19 (4.224)
R-pim	0.08086 (1.56)	0.0985 (0.5319)	0.03033 (0.8956)
CC1/2	0.978 (0.347)	0.962 (0.575)	0.969 (0.23)
CC*	0.994 (0.717)	0.99 (0.854)	0.992 (0.612)
Reflections used in refinement	57688 (5497)	53176 (4281)	36734 (3068)
Reflections used for R-free	1193 (114)	552 (44)	384 (34)
R-work	0.2259 (0.4189)	0.2695 (0.3920)	0.2389 (0.3721)
R-free	0.2528 (0.4587)	0.3198 (0.3765)	0.2922 (0.3648)
CC(work)	0.933 (0.602)	0.851 (0.681)	0.918 (0.440)
CC(free)	0.931 (0.434)	0.840 (0.550)	0.905 (0.245)
Number of non-hydrogen atoms	18013	19732	6466
macromolecules	16928	17742	6438
ligands	225	148	2
solvent	860	1842	26
Protein residues	2062	2066	778
RMS(bonds)	0.008	0.003	0.010
RMS(angles)	1.38	0.60	1.37
Ramachandran favored (%)	97.59	95.84	93.73
Ramachandran allowed (%)	2.12	3.96	5.22
Ramachandran outliers (%)	0.30	0.20	1.04
Rotamer outliers (%)	2.50	1.52	5.44
Clashscore	11.23	10.59	12.79
Average B-factor	93.53	85.96	82.54
macromolecules	93.11	85.68	82.63
ligands	138.17	90.70	72.51
solvent	90.03	88.29	61.84

Statistics for the highest-resolution shell are shown in parentheses.

7.5 PorX-structural subdivision



7.6 PorX-dimeric surface

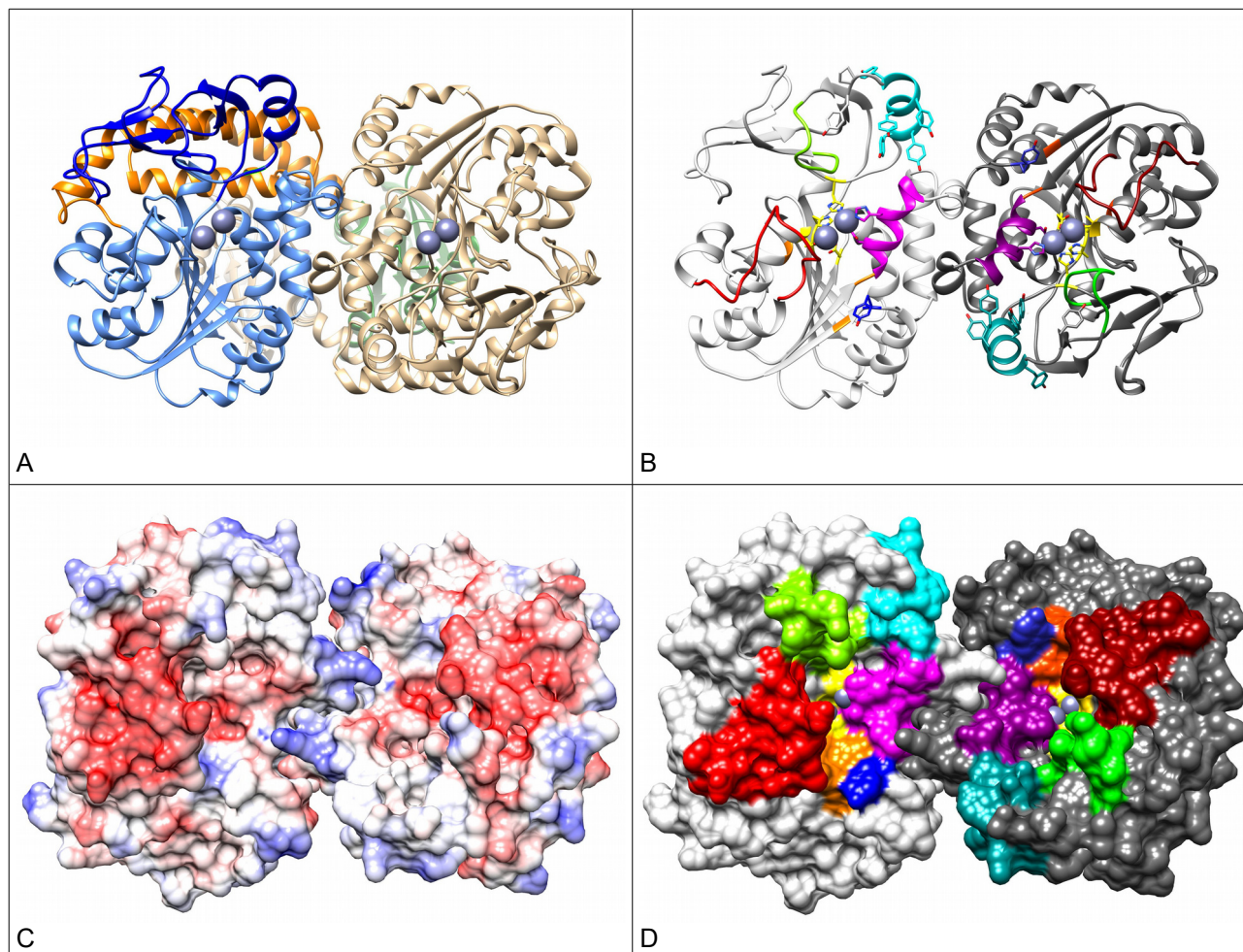
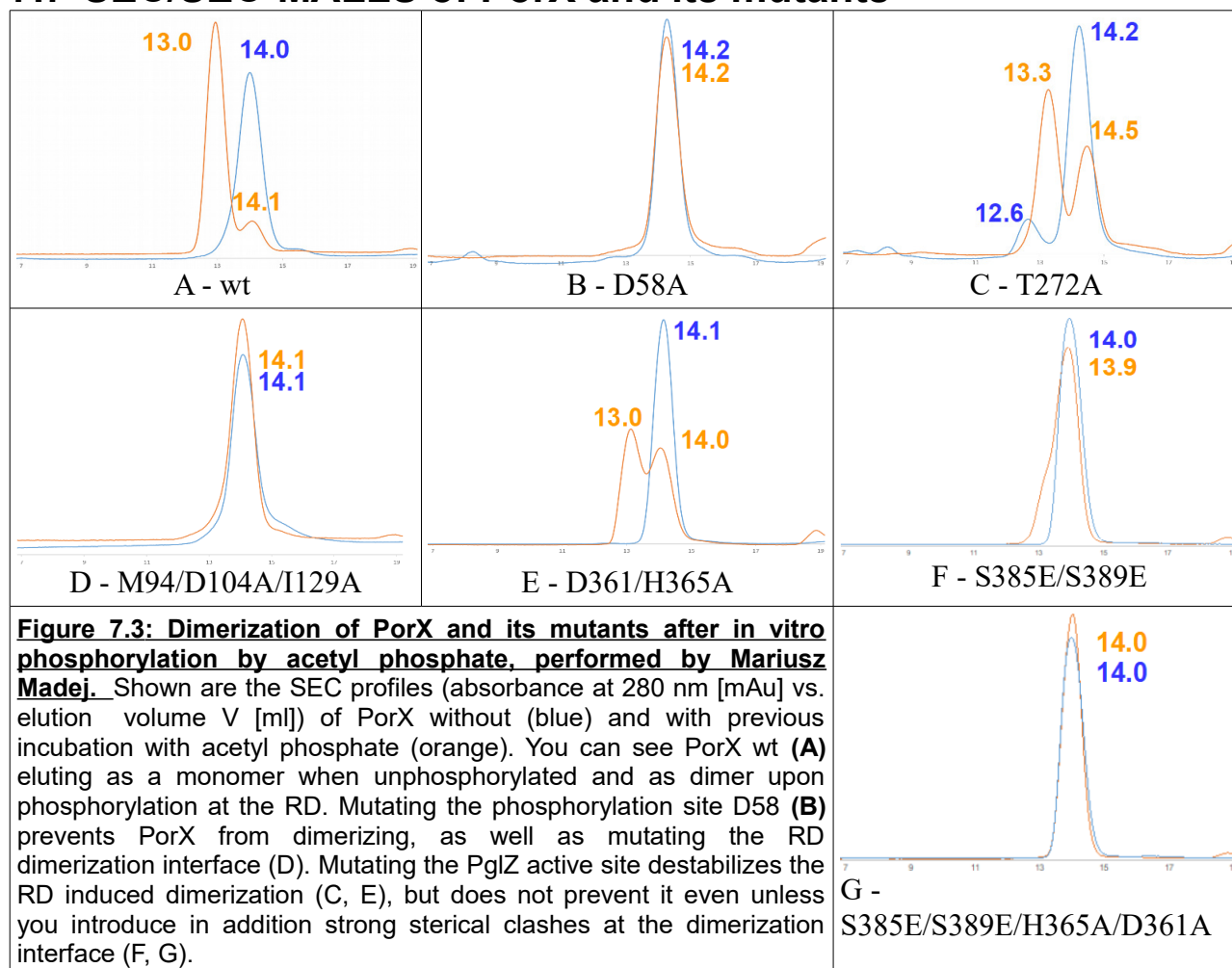


Figure 7.2: PorX-dimeric surfaces The active sites of PorX monomers within a dimer are at a large distance of 32-38 Å between zinc ions, therefore, assuming small di-/oligonucleotides as substrates, we can assume independent catalysis of each active site. **(A)** Ribbon representation of PorX (pGpG structure), Chain A: RD: green, HBD: orange, PglZ CAT:light blue, CAP:dark blue. Chain B in gray, Zn radius enlarged to 1.9 Å. **(B/D)** PorX surface regions in Ribbon (B) and surface (D) representation. Lα3b-α3c red, Lβ1c-β1I cyan, α5a magenta, metal coordination yellow, 5'phosphate coordination orange, Y332 blue. Chain B darker colors than A. **(C)** Coulombic surface coloring. RED: electronegative, BLUE: electropositive.

7.7 SEC/SEC-MALLS of PorX and its mutants



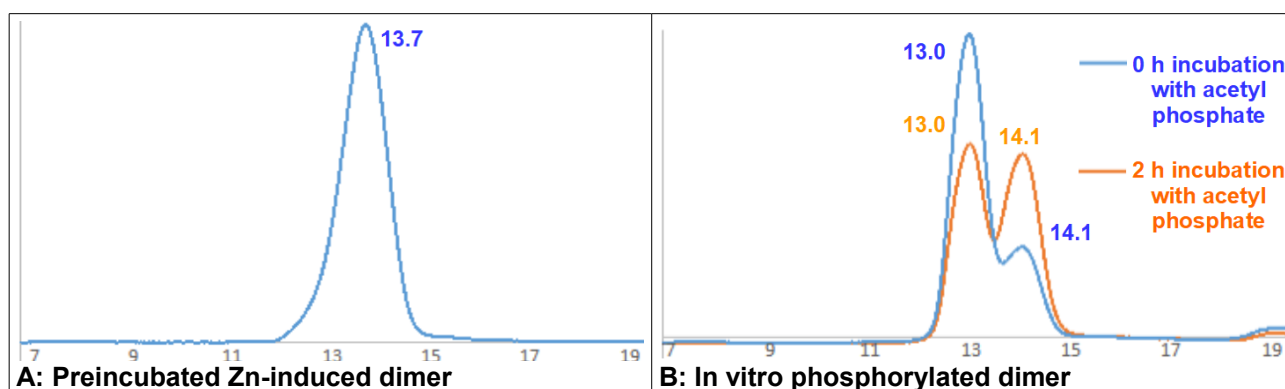
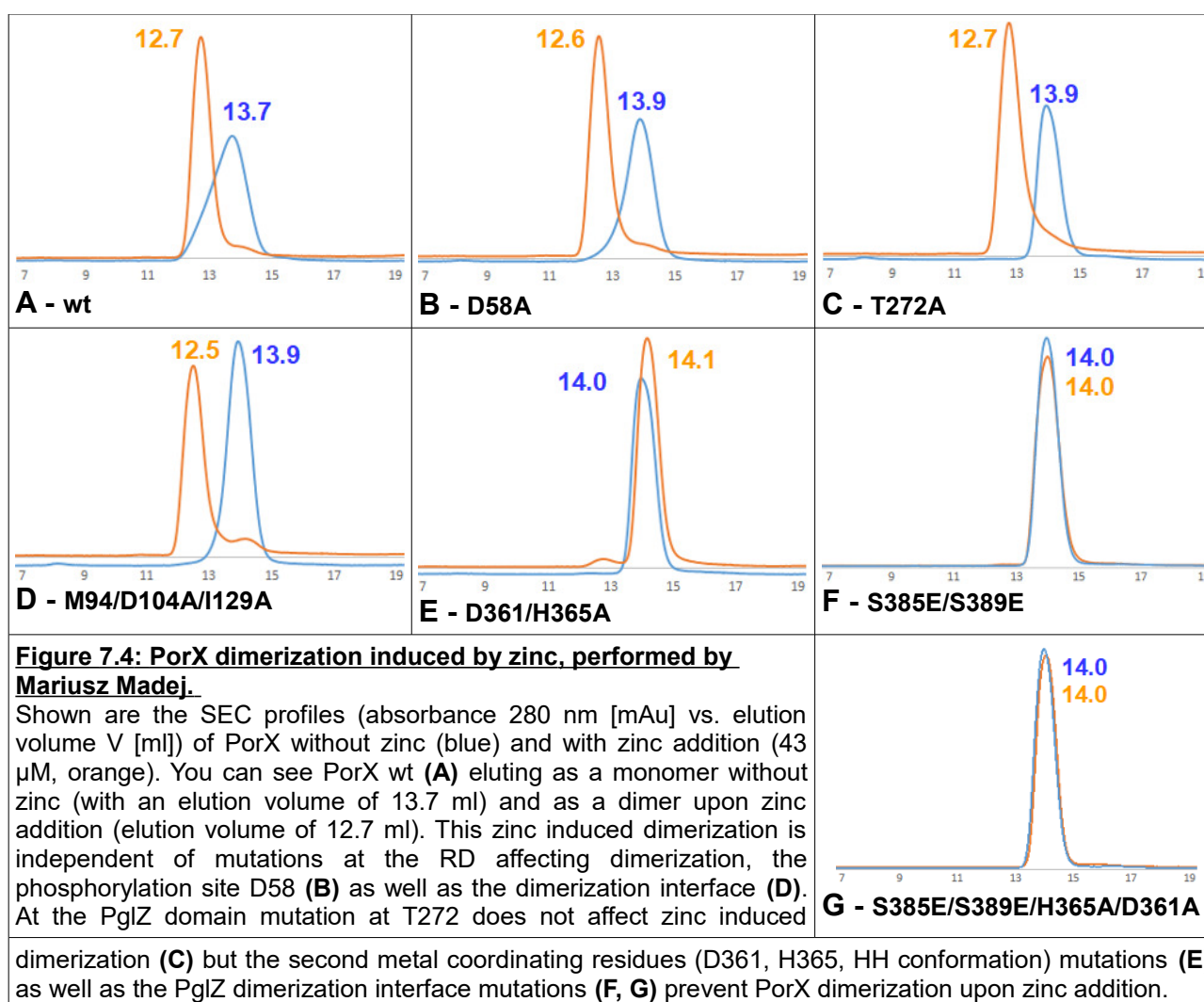


TABLE 7.3: SEC-MALLS inicial quality assesment. As published ^[0]				
MgCl ₂	Mn (kDa)	Mw (kDa)	Polydispersity (Mw/Mn)	Calculated mass (μg)
	Peak 1			
BSA	66.87 ± 0.05	66.87 ± 0.05	1.00 ± 0.00	36.45
wt	69.92 ± 0.10	70.15 ± 0.09	1.00 ± 0.00	62.23
wt-P	118.15 ± 0.17	118.21 ± 0.17	1.00 ± 0.00	48.74
D58A	68.58 ± 0.09	68.69 ± 0.09	1.00 ± 0.00	61.23
D58A-P	66.68 ± 0.10	66.88 ± 0.10	1.00 ± 0.00	61.37
D361A/H365A	60.83 ± 0.08	60.84 ± 0.08	1.00 ± 0.00	61.84
D361A/H365A-P	61.28 ± 0.08	61.29 ± 0.08	1.00 ± 0.00	32.60
	Peak 2			
BSA	133.75 ± 0.22	133.77 ± 0.22	1.00 ± 0.00	5.70
wt	114.54 ± 0.65	114.54 ± 0.65	1.00 ± 0.01	1.29
wt-P	63.44 ± 0.24	63.47 ± 0.24	1.00 ± 0.01	11.48
D58A	128.69 ± 0.73	128.89 ± 0.74	1.00 ± 0.01	1.22
D58A-P	124.18 ± 2.43	124.19 ± 2.43	1.00 ± 0.03	1.28
D361A/H365A	198.44 ± 2.72	207.20 ± 3.06	1.04 ± 0.02	0.28
D361A/H365A-P	105.88 ± 0.09	106.23 ± 0.09	1.00 ± 0.00	24.05

MgCl ₂ , ZnCl ₂	Peak 1			
BSA	65.79 ± 0.09	65.79 ± 0.09	1.00 ± 0.00	36.44
wt	116.73 ± 0.45	116.79 ± 0.44	1.00 ± 0.01	15.62
wt-P	121.63 ± 0.12	121.66 ± 0.12	1.00 ± 0.00	59.42
D58A	115.59 ± 0.24	115.71 ± 0.24	1.00 ± 0.00	14.59
D58A-P	119.54 ± 0.13	119.55 ± 0.13	1.00 ± 0.00	53.97
D361A/H365A	60.49 ± 0.10	60.49 ± 0.10	1.00 ± 0.00	56.24
D361A/H365A-P	109.34 ± 0.09	109.78 ± 0.09	1.00 ± 0.00	44.41
	Peak 2			
BSA	130.38 ± 0.35	130.42 ± 0.35	1.00 ± 0.00	6.25
wt	62.57 ± 3.76	63.00 ± 3.72	1.01 ± 0.08	1.25
wt-P	210.35 ± 2.19	210.47 ± 2.19	1.00 ± 0.01	0.47
D58A	67.74 ± 2.48	67.93 ± 2.47	1.00 ± 0.05	0.91
D58A-P	65.97 ± 1.69	66.29 ± 1.66	1.00 ± 0.04	1.58
D361A/H365A	110.96 ± 4.69	111.05 ± 4.65	1.00 ± 0.06	0.44
D361A/H365A-P	60.73 ± 0.18	60.75 ± 0.18	1.00 ± 0.00	11.54

7.8 Comparative homology

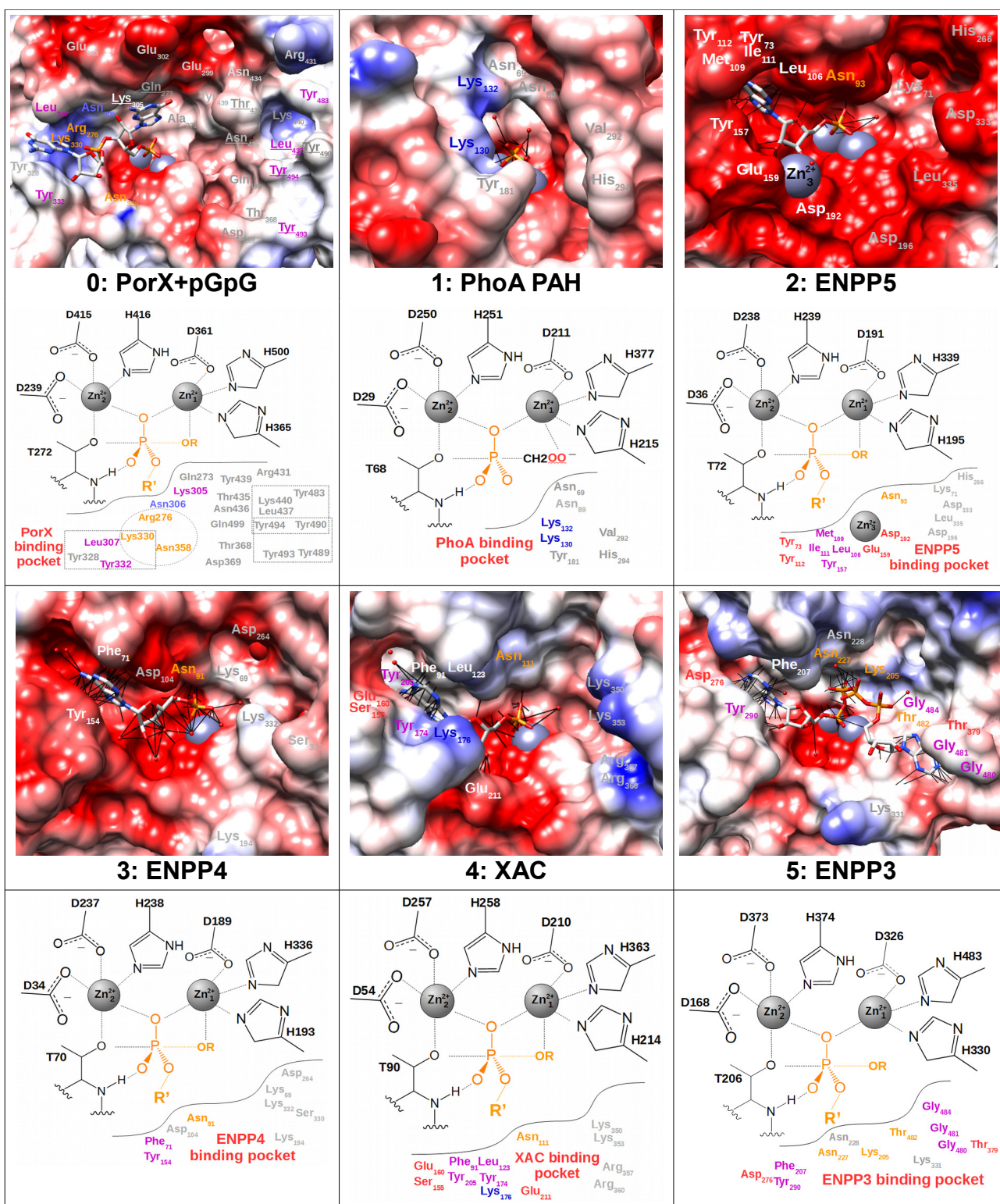
TABLE 7.4: TOP20 homolog proteins to PorX PglZ domain. Modified from publication^[0]

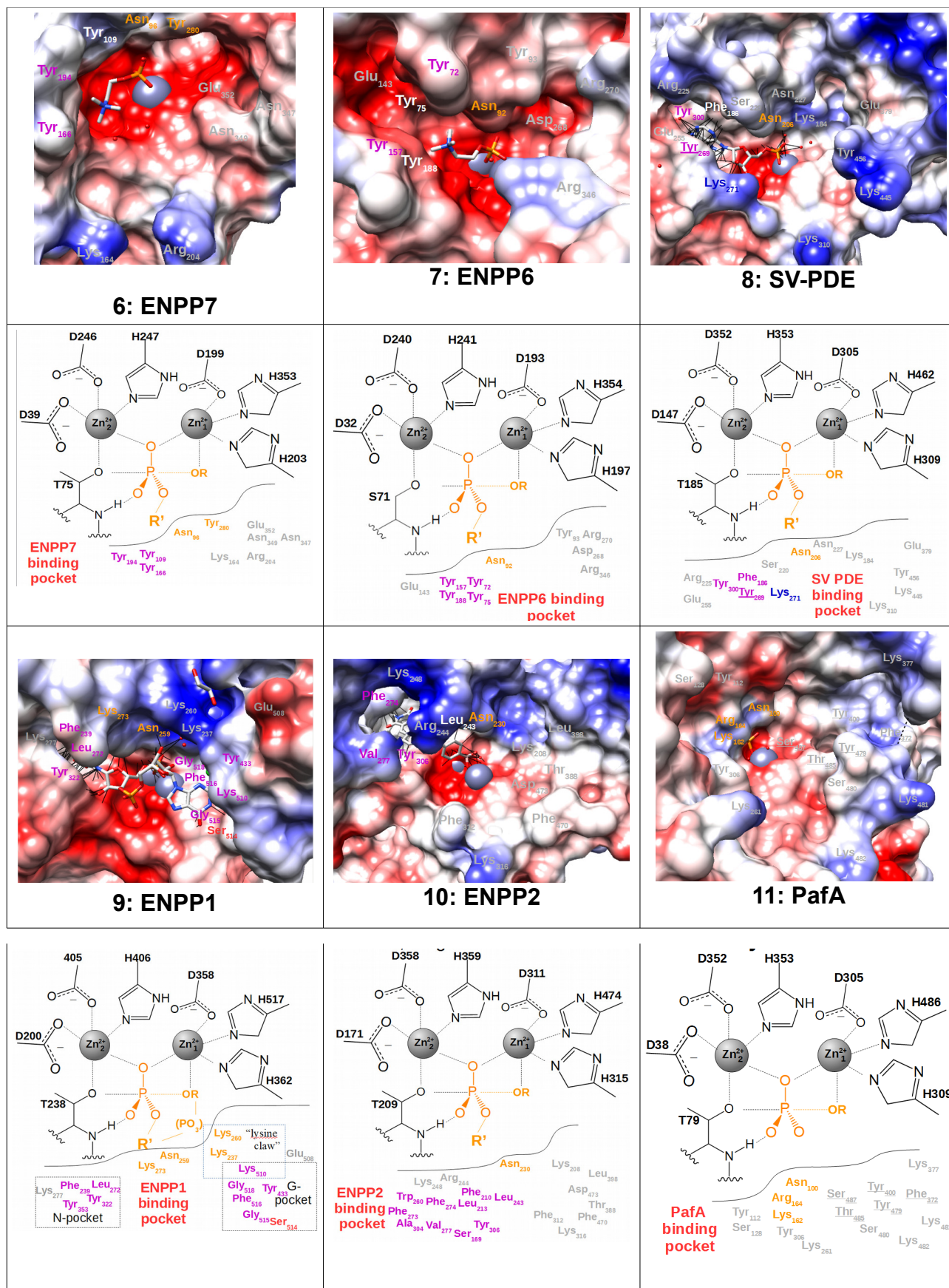
PDBs were summarized by protein, highest Z-scores, rmsd (root mean square deviation) and lali (length aligned) are shown for top pdb, different pdbs and species origins are ordered by homology as appeared in the TOP115 DALI output.^[0] Data has been mined from Uniprot data base: "UniProt: the universal protein knowledge-base, Nucleic Acids Res. 49:D1 (2021)." further references are written below each protein.

Nr.	Zscore/RM SD/LALI	Protein	PDB	Organism	Substrates	Function	Act. site metals
1	17.4/ 3.7/ 247	PhnA, Phosphono- acetate hydrolase	3szy, 3t01, 1ei6,3szz, 3t00, 3t02	S. meliloti, P. fluorescen s	- Phosphonoacetate	Phosphono- acetate metabolism	2 x Zn ²⁺
Agarwal, V; Borisova, SA; Metcalf, WW; vd Donk, WA; Nair, SK, <i>Chemistry & Biology</i> 18 (10), S. 1230–1240. (2011).							
2	17.3/ 3.4/ 247	ENPP5, Type I phospho- diesterase / nucleoti- de pyro- phospha- tase (PDE)	5vem, 5veo, 5ven	H. sapiens, M. musculus	- Hydrolyzes NAD /NAD+ - able to hydrolyse ADP- ribose (ADPR) and UDP glucose (UDPG) - no activity with nucleotide di- and triphosphates - no lysopholipase D activity	Assumed role in neuronal cell communication.	2 x Zn ²⁺
Gorelik, A; Randriamihaja, A; Illes, Kn; Nagar, B, <i>The FEBS journal</i> 284 (21), S. 3718–3726, (2017).							
3	17.1/ 3.7/ 248	ENPP4, Type I PDE	4lqy, 4lr2	H. sapiens,	- Hydrolyzes extracellular Ap3A into AMP and ADP and Ap4A into AMP and ATP. No ATPase activity	Vascular smooth muscle cell proliferation, blood coagulation.	2 x Zn ²⁺
Albright, RA; Ornstein, DL; Cao, W; Chang, WC; Robert, D; Tehan, M et al., <i>The Journal of biological chemistry</i> 289 (6), S. 3294–3306,(2014).							
4	16.9/ 4.0/ 251	Xac PDE, Type I PDE	2gsn, 2gso, 2rh6, 2gsu	X. citri	- NPP activity - ATP hydrolysis	- not classified	2 x Zn ²⁺
Zalatan, Jesse G.; Fenn, Timothy D.; Brunger, Axel T.; Herschlag, Daniel, <i>Biochemistry</i> 45 (32), S. 9788–9803, (2006).							
5	16.8/ 3.5/ 245	ENPP3, Type I PDE	6c01,6f2t, 6f30, 6c02, 6f33, 6f2y, 6c02, 6c01, 6f2v	H. sapiens, R. norvegicu s	- Hydrolyzes of extracellular nucleotides ATP, GTP, UTP and CTP - Hydrolyses Ap3A and Ap4A, UDPGlc, UDPGlcNAc and NAD+	- mast cell and basophil inhibition - prevents infl. cytokines release - glycosyltransfer regulatory activity	2 x Zn ²⁺
Döhler, Christoph; Zebisch, Matthias; Sträter, Norbert, <i>Scientific reports</i> 8 (1), S. 10874, (2018).							
6	16.8/ 3.5/ 247	ENPP7, Type I PDE	5tcd, 5udy	H. sapiens, R. norvegicu s	alkaline sphingomyelinase - extracellular - phospholipase C activity - no nucleotide -pyrophosphatase activity	- gastro-intestinal sphingomyelin digestion, cerami- de formation, fatty acid absorption	2 x Zn ²⁺
Gorelik, A; Liu, F; Illes, K; Nagar, B, <i>The Journal of biological chemistry</i> 292 (17), S. 7087–7094, (2017).							

7	16.8/ 3.9/ 246	ENPP6, Type I PDE	5egh, 5ege	H. sapien s, M. muscul us	- choline-specific glycerophosphodiester- phosphodiesterase - hydrolyzes glycerophosphocholine and lysophosphatidylcholine	- lipid metabolism	2 x Zn ²⁺
Morita, Junko; Kano, Kuniyuki; Kato, Kazuki; Takita, Hiroyuki; Sakagami, Hideki; Yamamoto, Yasuo et al., <i>Scientific reports</i> 6, S. 20995, (2016).							
8	16.7/ 3.5/ 246	Snake venom phosph odiester ase	5gz5, 5gz4	Naja atra atra	- Hydrolyzes ADP with high activity - weak or no activity on 5'- AMP, 5'-GMP, 3'-AMP, ATP, cAMP or cGMP	- platelet aggregation inhibition	2 x Zn ²⁺
9	16.3 / 3.7/ 247	ENPP1, Type I PDE	6aek, 4gtz, 4gtw, 4b56, 4gtx, 4gty, 4gtz, 4gty, 4b56, 6ael, 4gtw	H. sapien s, M. muscul us	- NTP hydrolysis to NMP and PPi, prefers ATP, hydrolyses diadenosine polyphosphates - cleaves 3',5'-cAMP to AMP - hydrolyzes 2'-3'-cGAMP (to AMP and GMP) but not 3'-3'- cGAMP - hydrolyses Ap3A	- involved in bone mineralization and calcification of soft tissue - regulation of nucleotide levels in ER and Golgi - regulation of purinergic signaling - immune suppression by inhibiting TMEM173/STING and type-I interferon production	2 x Zn ²⁺
Albright, RA; Ornstein, DL; Cao, W; Chang, WC; Robert, D; Tehan, M et al., <i>The Journal of biological chemistry</i> 289 (6), S. 3294–3306,(2014). Dennis, ML.; Newman, J; Dolezal, O; Hattarki, M; Surjadi, RN.; Nuttall et al. : Crystal structures of human ENPP1 in apo and bound forms. In: <i>Acta crystallographica. Section D, Structural biology</i> 76 (Pt 9), S. 889–898, (2020). Jansen, S; Perrakis, A; Ulens, C; Winkler, C; Andries, M; Joosten, R. et al., <i>Structure</i> 20 (11), S. 1948–1959, (2012). Kato, K; Nishimasu, H; Oikawa, D; Hirano, S; Hirano, H et al., <i>Nature communications</i> 9 (1), S. 4424, (2018). Kato, K; Nishimasu, H; Okudaira, S; Mihara, E; Ishitani, R et al., <i>Nat. Ac. Sci.</i> 109 (42), S. 16876–16881, (2012).							
10	16.1/ 3.6/ 244	ENPP2, Auto- taxin	3wav,6w35, 3nkr, 3nkn, 5lia, 3nko, 6leh,3nkq, 5m0e, 5dlw, 5l0b, 5dlv, 5m7m, 5ohi, 5ijq, 5olb, 3wax,2xr9, 5hrt, 5l0k,3way, 5m0m,4zg7,5kxa, 5mhp,4zg6,5m0d, 5l0b, 4zg6,5ijs, 5lqq, 5l0e, 3nkp, 4zg9	M. muscul us, R. norveg icus, H. sapien s	- lysophospholipase D hydrolyzes lysophospholipids to lysophosphatidic acid - hydrolyzes sphingosylphosphorylcholine to sphingosine-1-phosphate -Ap3A hydrolysis "selectivity for lipid over nucleotide substrates "	- chemotaxis, microtubule formation, neurite remodeling, proliferation and survival/apoptosis - tumor formation and metastasis	2 x Zn ²⁺
Albright, RA; Ornstein, DL; Cao, W; Chang, WC; Robert, D; Tehan, M et al., <i>The Journal of biological chemistry</i> 289 (6), S. 3294–3306, (2014). Hausmann, J; Kamtekar, S; Christodoulou, E; Day, JE; Wu, T et al., <i>Nature structural & molecular biology</i> 18 (2), S. 198–204, (2011).							

11	14.9/ 3.6/ 237	PafA alkaline phosphatase	5tj3, 5too	E. meningoseptica	- rel. unspecific alkaline phosphatase - hydrolyses NTPs to NDPs, dephosphorylates phosphate monoesters except 2',3'-cyclic AMP and myo-inositol hexakisphosphate	- peri-plasmatic phosphatase	2 x Zn ²⁺
Sunden, F; AlSadhan, I; Lyubimov, A; Doukov, T; Swan, J; Herschlag, D, <i>JBC</i> 292 (51), S. 20960–20974, (2017)							
12	14.9/ 3.6/ 246	SPAP/Alkaline phosphatase PhoK; [homolog to PafA]	3q3q, 5xwi, 5xwk	Sphingomonas sp.	- phosphomonoester hydrolases	- secreted enzyme	2 x Zn ²⁺
Bihani, SC; Das, A; Nilgiriwala, KS; Prashar, V; Pirocchi, M; Apte et al., <i>PloS one</i> 6 (7), e22767, (2011).							
13	12.9/ 3.1/ 184	BT_1596, Delta 4,5-hexuronate-2-O-sulfatase	5g2t, 5g2u	B. thetaiotaomicron	- hydrolysis of glycosaminoglycans, chondroitin sulfate, dermatan sulfate and heparan sulfate	- Exosulfatase - nutrition acquisition as human gut symbiont	none
14	12.2/ 3.3/ 181	Inner membrane protein YejM sulfatase domain 245-586	5i5h	E. coli	- hydrolysis of sulfuric esters	- unknown function	none
15	11.7/ 3.8/ 173	2,3-bisphosphoglycerate-independent phospho-glycerate mutase gpml	2ify	B. anthracis	- Isomerization of 2- and 3-phosphoglycerate	- Glucose metabolism	2 x Mn ²⁺
16	11.6/ 3.1/ 179	Choline sulfatase, betC	6g60	S. meliloti	- hydrolyses cholinesulfate into choline	- Choline metabolism	Ca ²⁺
17	11.5/ 3.3/ 170	Arylsulfatase A	1e2s	H. sapiens	- Hydrolyzes cerebroside sulfate	- Glycosphingolipid metabolism	Ca ²⁺
18	11.5/ 3.4/ 183	Phospho-ethanolamine Transferase EptC	4tn0	C. jejuni	Hydrolases sulfuric esters - Phosphotransferase activity	- cell surface modification	Zn ²⁺
19	11.5/ 3.3/ 179	Sulfamidase	4mhx	H. sapiens	- N-sulphoglucosamine sulphohydrolase	- involved in heparan sulfate degradation	Ca ²⁺
20	11.5/ 3.2/ 182	MRC1, probable phosphatidylethanolamine transferase	5lrm, 6li5, 5lrm, 5grr, 6li6, 5lrm	E. coli	- Phosphoethanolamine addition to lipid A	- provides polymyxin resistance	1 or 2xZn ²⁺





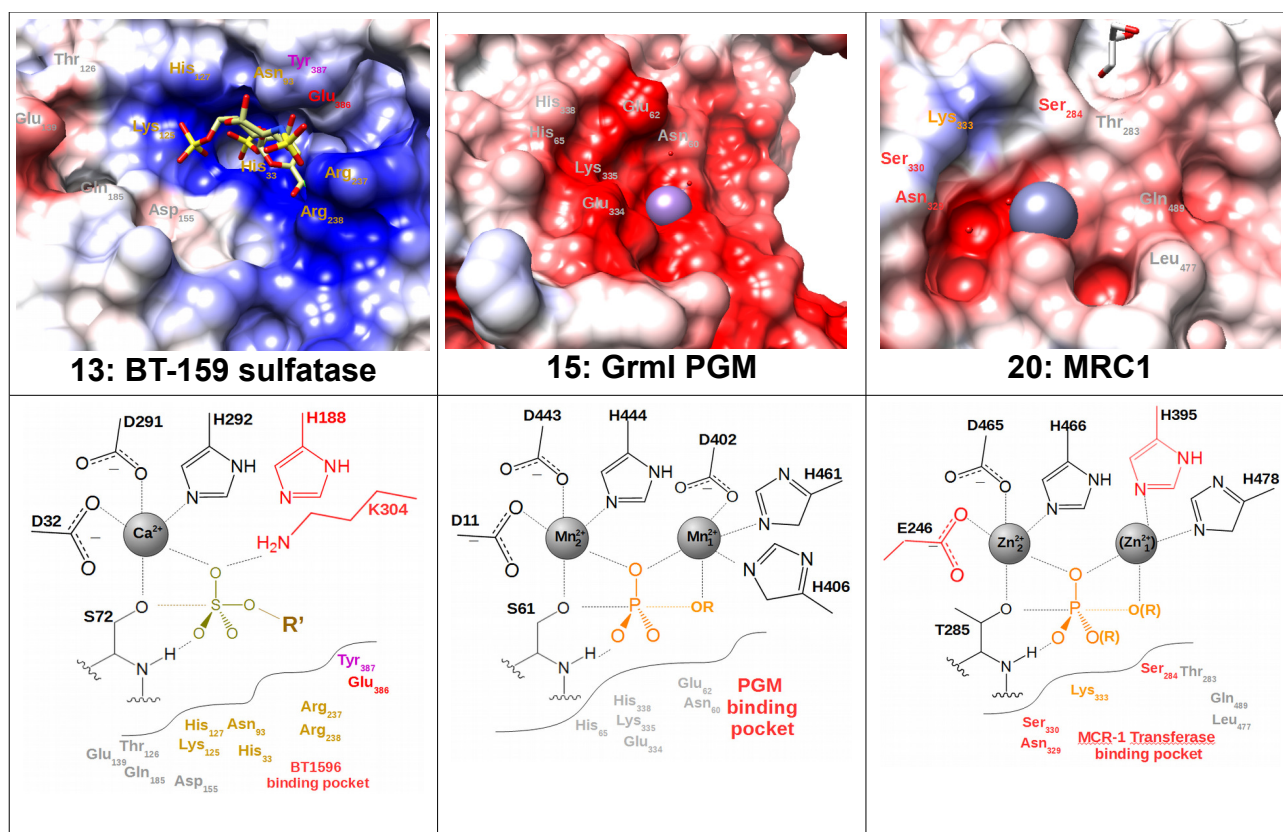


Figure 7.6: PorX homologs (DALI) active site coordination (lower panels) and substrate binding pockets (upper panels)

Active site conserved residues to PorX are colored black, whereas different residues are colored red. Phosphorylated substrates are colored orange, sulfurylated in yellow. The active site substrate binding pockets are colored in coulombic colored Connolly surface. Residues with hydrophobic contacts are colored magenta or white, residues stabilizing phosphates in orange, stabilizing sulfates in yellow and hydrophilic residues in blue (electropositive) or red electronegative). For better visibility the ionic radius of the zinc atoms has been set to 1.9 Å. Residues without direct interactions but in close proximity are shown in gray, those with assumed binding function are furthermore underlined. Redundant surfaces from similar proteins from the DALI results are not shown and the index number shows their rank of homology as described in TABLE 7.4.

7.9 Enzymatic assays

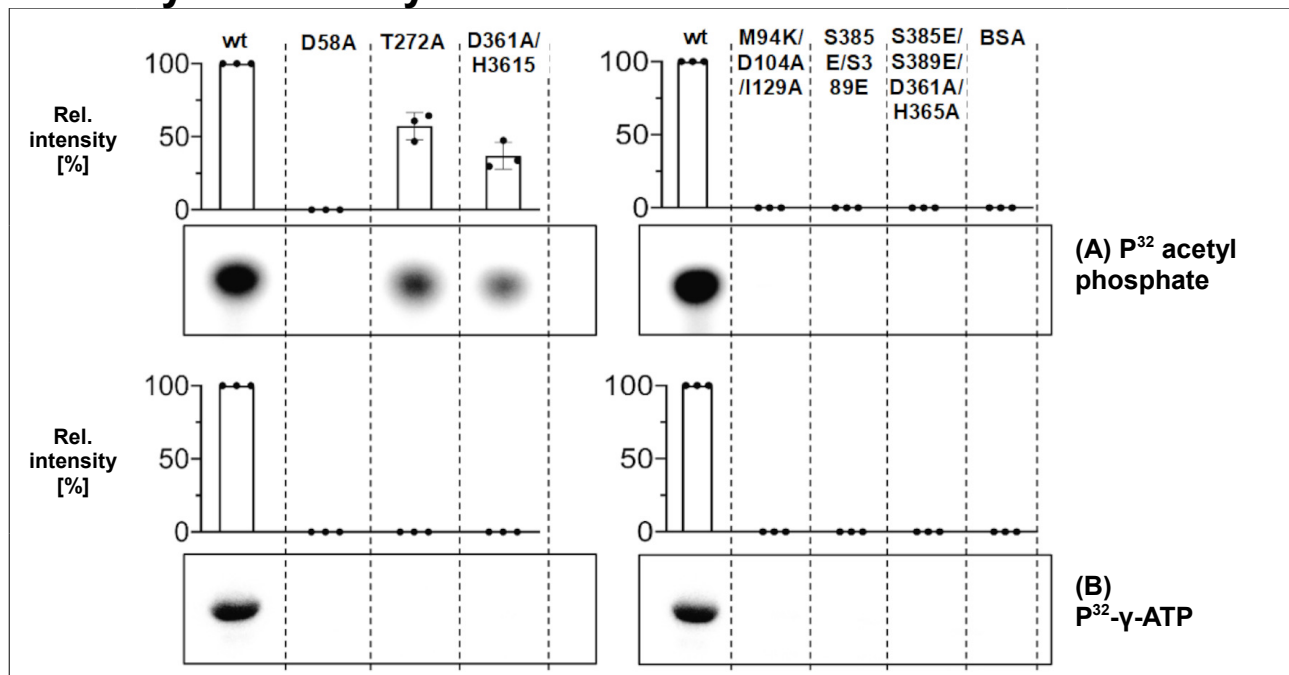


Figure 7.7: Phosphorylation assays of PorX and mutants, performed by Mariusz Madej, modified from publication^[0]

(A) Phosphorylation of PorX by acetyl phosphate at the RD gets inhibited by mutating the phosphorylation site D58 or any of the dimerization interfaces at RD (M94K/D104A/I134A) or PglZ (S385E/S389E or S385E/S389E/D361A/H365A). However inhibition occurs as well but to a lesser extent by inserting mutations at the active site of PorX coordinating Zn1 which stabilizes the HH conformation (D361A/H365A) and takes part in stabilizing the PglZ dimers, and by even to a lesser extent by T272A mutation that is not necessary for dimerization. **(B)** ATP induced phosphorylation, presumably at the PglZ active site T272 gets inhibited by mutating T272 to alanine, as well as any other mutation tested inhibiting dimer formation.^[0]

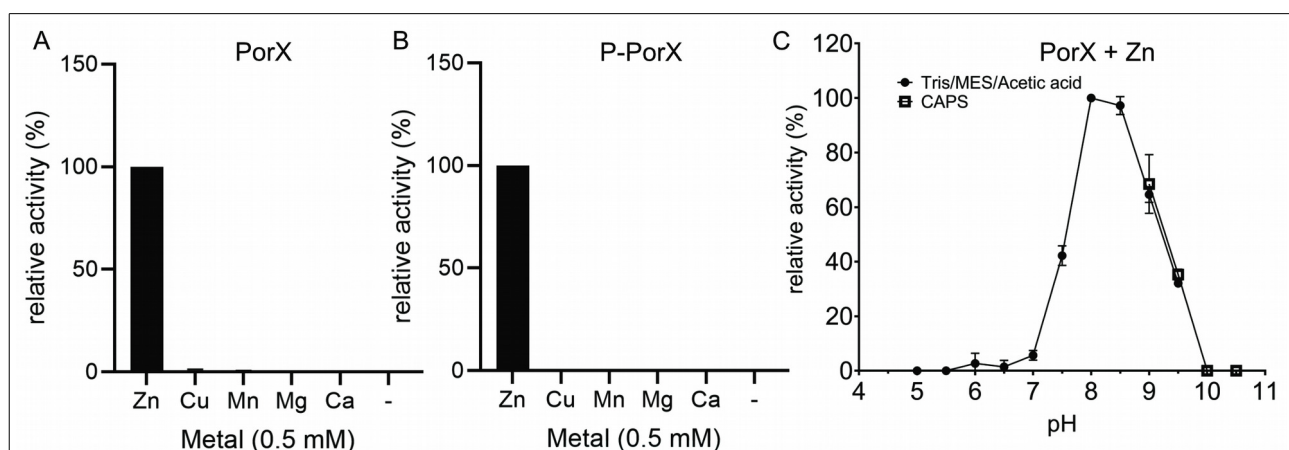


Figure 7.8: Phosphodiesterase assays, performed by Mariusz Madej, modified from publication^[0]

(A, B) PorX (A) and phosphorylated PorX (B) phosphodiesterase activities in the presence of different divalent cations by measuring the hydrolysis of bis(p-nitrophenyl) phosphate bis-(pNPP). The relative activity is normed to Zn at 100%. **(C, D)** PorX (C) and phosphorylated PorX (D) activity at different pH. pH range 4.0–9.5 is shown as dots, pH range 9–10.5 as squares. Buffer composition is indicated. **(D, E)** PorX (D) and phosphorylated PorX (E) bis-(pNPP) hydrolysis kinetics for wt and mutants.

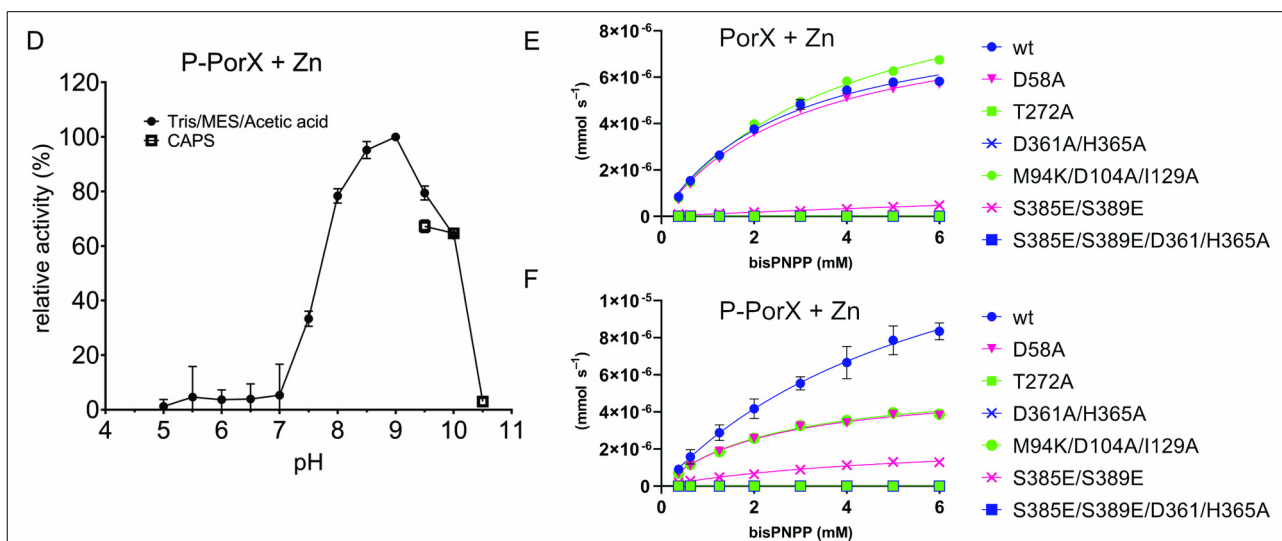


Figure 7.8: Phosphodiesterase assays, performed by Mariusz Madej, modified from publication^[1]

(A, B) PorX (A) and phosphorylated PorX (B) phosphodiesterase activities in the presence of different divalent cations by measuring the hydrolysis of bis(p-nitrophenyl) phosphate bis-(pNPP). The relative activity is normed to Zn at 100%. (C, D) PorX (C) and phosphorylated PorX (D) activity at different pH. pH range 4.0–9.5 is shown as dots, pH range 9–10.5 as squares. Buffer composition is indicated. (D, E) PorX (D) and phosphorylated PorX (E) bis-(pNPP) hydrolysis kinetics for wt and mutants.

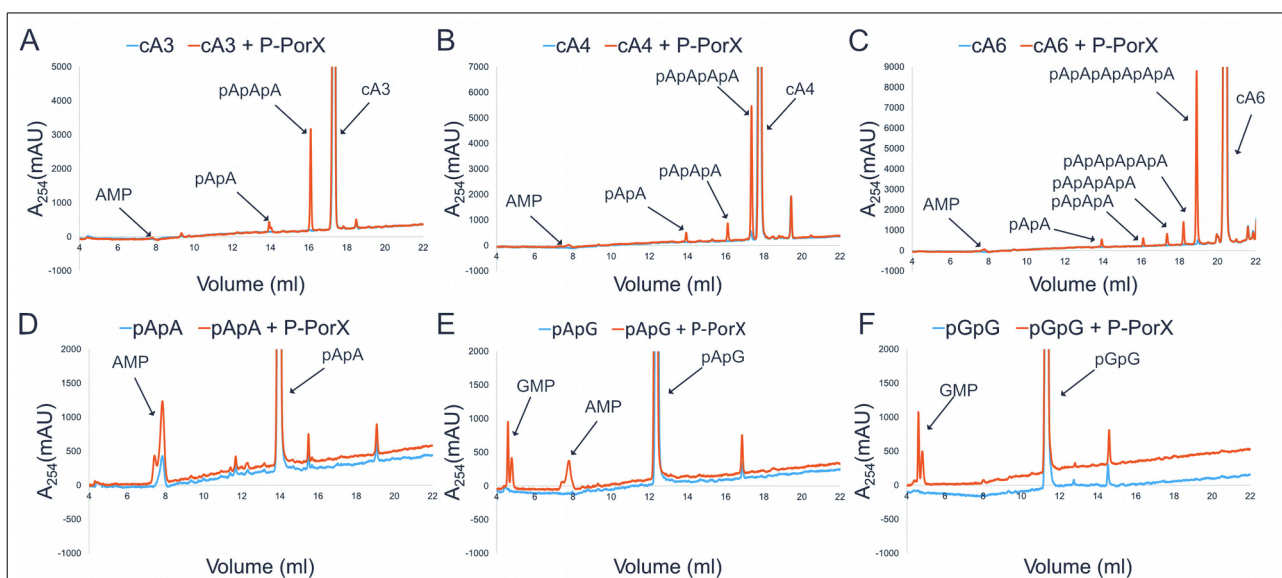


Figure 7.9: Positive substrate identifications of PorX by UHPLC-MS ("ultrahigh-performance liquid chromatography-mass spectrometry", performed by Mariusz Madej) as published^[1]

UHPLC elution profiles (Absorption (254 nm)[mAU]/elution volume [ml]) together with MS verified peak identifications for substrates with (orange curves) or without (blue curves) previous incubation with PorX. Shown are only positive results observed for: (A) cA3, (B) cA4, (C) cA6, (D) pApA, (E) pApG, (F) pGpG. All substrates are linear dinucleotides or linear or cyclic oligonucleotides.^[1]

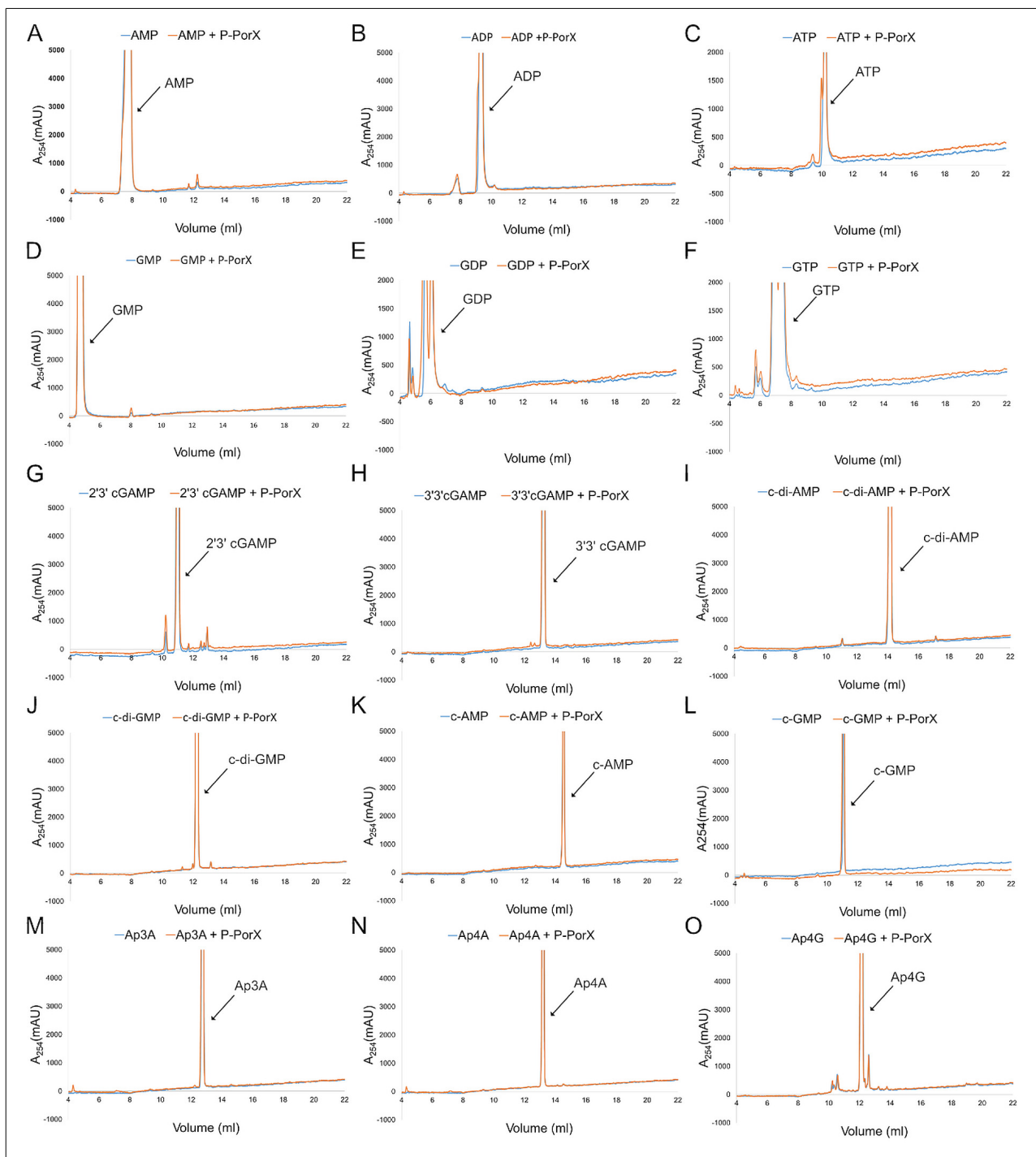


Figure 7.10: Negative substrate identifications of PorX by UHPLC-MS ("ultrahigh-performance liquid chromatography-mass spectrometry", performed by Mariusz Madej) as published ^[0]

UHPLC elution profiles (Absorption (254 nm)[mAU]/elution volume [ml]) together with MS verified peak identifications for substrates with (orange curves) or without (blue curves) previous incubation with PorX. Shown are only negative results A-O as indicated above each profile. These negative substrates are polyphosphates, cyclic dinucleotides or cyclic or lineal mononucleotides.^[0]

Opt_C (Cryo) Cryos: (A+B) equal (C+D) equal (E+F) equal G+H Buffers/Salts/Precipitants: A=B; C=D; E=F, G=H												
	1	2	3	4	5	6	7	8	9	10	11	12
A	35 % MPD	40 % MPD	40 % ethylene glycol	50 % ethylene glycol	30 % PEG 200	40 % PEG 200	50 % PEG200	20 % PEG 300	30 % PEG 300	40 % PEG 300	40 % Ethanol	35 % 2-Etoxy-ethanol
100 mM Glycine pH 8.5; 20 mM Sodium formate												
B	30 % PEG 400	40 % PEG 400	50 % PEG 400	30 % PEG 600	40 % PEG 600	35 % iso-propanol	35 % glycerol	45 % glycerol	25 % 1,2-propandiol	40 % 1,2-propandiol	30 % glucose	30 % sucrose
100 mM Glycine pH 8.5; 20 mM Sodium formate												
C	35 % MPD	40 % MPD	40 % ethylene glycol	50 % ethylene glycol	30 % PEG 200	40 % PEG 200	50 % PEG200	20 % PEG 300	30 % PEG 300	40 % PEG 300	40 % Ethanol	35 % 2-Etoxy-ethanol
100 mM HEPES pH 8.5; 20 mM Sodium formate												
D	30 % PEG 400	40 % PEG 400	50 % PEG 400	30 % PEG 600	40 % PEG 600	35 % iso-propanol	35 % glycerol	45 % glycerol	25 % 1,2-propandiol	40 % 1,2-propandiol	30 % glucose	30 % sucrose
100 mM HEPES pH 8.5; 20 mM Sodium formate												
E	35 % MPD	40 % MPD	40 % ethylene glycol	50 % ethylene glycol	30 % PEG 200	40 % PEG 200	50 % PEG200	20 % PEG 300	30 % PEG 300	40 % PEG 300	40 % Ethanol	35 % 2-Etoxy-ethanol
100 mM Cacodylate pH 7.0; 2 % PEG 3350												
F	30 % PEG 400	40 % PEG 400	50 % PEG 400	30 % PEG 600	40 % PEG 600	35 % iso-propanol	35 % glycerol	45 % glycerol	25 % 1,2-propandiol	40 % 1,2-propandiol	30 % glucose	30 % sucrose
100 mM Cacodylate pH 7.0; 2 % PEG 3350												
G	35 % MPD	40 % MPD	40 % ethylene glycol	50 % ethylene glycol	30 % PEG 200	40 % PEG 200	50 % PEG200	20 % PEG 300	30 % PEG 300	40 % PEG 300	40 % Ethanol	35 % 2-Etoxy-ethanol
100 mM Cacodylate pH 7.0; 2 % PEG 3350; 20 mM Glycine pH 8.0 (as additive)												
H	30 % PEG 400	40 % PEG 400	50 % PEG 400	30 % PEG 600	40 % PEG 600	35 % iso-propanol	35 % glycerol	45 % glycerol	25 % 1,2-propandiol	40 % 1,2-propandiol	30 % glucose	30 % sucrose
100 mM Cacodylate pH 7.0; 2 % PEG 3350; 20 mM Glycine pH 8.0 (as additive)												

Opt_D		7 % PEG 2000	7 % PEG 2000 MME	7 % PEG 3000	7 % PEG 3350	7 % PEG 4000	7 % PEG 5000 MME	7 % PEG 6000	7 % PEG 8000	7 % PEG 10000	7 % PEG 20000	7 % MM W PEG	7 % HM W PEG
pH versus PEG influence		1	2	3	4	5	6	7	8	9	10	11	12
Cacodylate pH 5.5	A												
Cacodylate pH 6.0	B												
Cacodylate pH 6.5	C												
Cacodylate pH 7.0	D												
HEPES pH 7.5	E												
Glycine pH 8.0	F												
Glycine pH 8.5	G												
Glycine pH 9.0	H												

Opt_E															
Glycerol screening		1	Gradient 0.5-10 % Glycerol							Gradient 15-35 % Glycerol					
			2	3	4	5	6	7	8	9	10	11	12		
Cacodylate pH 7.0	A													12 % PEG 3350	
HEPES pH 8.0	B														
HEPES pH 8.5	C													12 % PEG 8000	
Glycine pH 8.5	D														
Cacodylate pH 7.0	E													12 % PEG 8000	
HEPES pH 8.0	F														
HEPES pH 8.5	G														
Glycine pH 8.5	H														
Opt_F															
		1	2	3	4	5	6	7	8	9	10	11	12		
4 % PEG 3350	A														
6 % PEG 3350	B														
8 % PEG 3350	C														
4 % PEG 8000	D														
6 % PEG 8000	E														
6 % PEG 8000	F														
100 mM NaFormate															
6 % PEG 3350															
100 mM NaFormate	G														
6 % PEG 3350															
50 mM NaFormate + 20 mM Glycine pH 8	H														
fine-tuning		Cacodylate pH 6.0	Cacodylate pH 6.5	Cacodylate pH 6.7	Cacodylate pH 7.0	Cacodylate pH 7.0	HEPES pH 7.5	HEPES pH 7.5	HEPES pH 8.5	HEPES pH 8.5	TRIS pH 8.5	TRIS pH 9	Glycine pH 8.0	Glycine pH 8.5	Glycine pH 9.0

PorX Opt_G															
Deep-well 1 ml, 2 plates			5 mM MnCl	2 % Gly-cerol	20 % Gly-cerol	2 % trehal ose	20 % trehal ose	0.5 M TMA O pH 8.5	3 M TMA O pH 8.5	20 mM Gly-cine 8.5	50 mM NaFor mate	50 mM NaFor mate+ 20 mM Glycin e	10 mM CHES 8.5+ 20 mM Gly-cine 8.5		
			1	2	3	4	5	6	7	8	9	10	11	12	HEPES pH 8.5
1 M NaFormate	A														
2 M NaFormate	B														
6 % PEG 20.000	C														
6 % PEG 8000	D														
6 % PEG 3350	E														
PEG 8000 1-15%	F														
PEG 3350 1-15%	G														
PEG 3350 1-15%	H														
200 mM LiSO4															

PorX Opt_H		Cacodylate pH 6.0			Cacodylate pH 6.5					
			1	2	3	4	5	6		
	A		7 % MMW PEG		7 % PEG 2000	7 % PEG 3350	7 % PEG 20.000	7 % MMW PEG		
Cacodylate pH 7.0					HEPES pH 8.5					HEPES pH 7.5
	B		7 % PEG 3350	7 % MMW PEG	7 % PEG 3350	2 M NaFormate	7 % PEG 6000	6 % PEG 3350		
									100 mM NaFormate	
Glycin pH 8.0						Glycin pH 9.0				Glycin pH 10.0
	C		4 % PEG 3350	4 % PEG 8000	6 % PEG 8000	8 % PEG 3350	7 % HMW PEG	4 % PEG 3350		
Glycin pH 8.5										Glycin pH 9.5
	D		6 % PEG 3350	7 % PEG 4000	7 % PEG 5000 MME	12 % PEG 8000	6 % PEG 3350	2 % PEG 3350		
									Buffer	
						0.5 % Glycerol			Salt/Additive	
									Precipitant (/Salt)	

PorX Opt_I											
Drop ratios											
2.5 + 2.5											
4+0.5											
PorX Opt_J											
each pH the same, 2 columns/8 rows:											
Deep-well 1.0 ml, 6 plates:											
Diff cononcentrations, +/- 100 mM NaFormate, diff temperatures +/- ADDS											
HMW PEG: PEG 20 K, 10 K, 8 K, 6 K: ¼ of 50 % each											
MMW PEG: PEG 2 K, 3350, 4 K, 5000 MME: ¼ of 40% each											

PorX Opt_K			1	2	3	4	5	6	7	8	9	10	11	12
Cacodylate pH 7.0	A		5 - 10 % PEG MMW+HMW(5:4)						5 - 10 % PEG MMW Gradient					
	B		5 - 10 % PEG 3350 Gradient						5 - 10 % PEG HMW Gradient					
HEPES pH 7.5	C		5 - 10 % PEG MMW+HMW(5:4)						5 - 10 % PEG MMW Gradient					
	D		5 - 10 % PEG 3350 Gradient						5 - 10 % PEG HMW Gradient					
Glycine pH 8.0	E		5 - 10 % PEG MMW+HMW(5:4)						5 - 10 % PEG MMW Gradient					
	F		5 - 10 % PEG 3350 Gradient						5 - 10 % PEG HMW Gradient					
Glycine pH 8.5	G		5 - 10 % PEG MMW+HMW(5:4)						5 - 10 % PEG MMW Gradient					
	H		5 - 10 % PEG 3350 Gradient						5 - 10 % PEG HMW Gradient					
ALL PLATE: 100 mM NaFormate + 0.1 mM ZnCl2														
PorX_Opt_L			1	2	3	4	5	6	7	8	9	10	11	12
HEPES PH 7.5	A		1 - 6 % PEG 8000 Gradient											
	B		1 - 6 % PEG 8000 Gradient											
Glycine PH 8.0	C		1 - 6 % PEG 8000 Gradient											
	D		1 - 6 % PEG 8000 Gradient											
	E		1 - 6 % PEG 8000 Gradient											
	F		1 - 6 % PEG 8000 Gradient											
Glycine pH 8.5	G		1 - 6 % PEG 8000 Gradient											
	H		1 - 6 % PEG 8000 Gradient											
Deep-well 1 ml, 8 plates: ALL plus 1 X		ADDS												
: 100 mM NaFormate (pH 8.5), 15 mM MgCl, 5 mM BeSO4, 30 mM NaF (provided		10 X Stock)												