Elucidating the Roles of ScORF3 and VAT-1 in Neocarzilin A Biosynthesis and Anticancer Activity: A Structural and Functional Study

Felix Maximilian Zäh

München 2024

Dissertation eingereicht am: 25.09.2024
Erstgutachterin: Dr. Sabine Schneider
Zweitgutachter: Prof. Dr. Thomas Carell
Tag der mündlichen Prüfung: 14.11.2024

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Dr. Sabine Schneider betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 19. Februar 2025

Felix Maximilian Zäh

Dissertation eingereicht am:	25.09.2024
1. Gutacher/in:	Dr. Sabine Schneider
2. Gutacher/in:	Prof. Dr. Thomas Carell
Mündliche Prüfung am:	14.11.2024

Abstract

Natural products have always been a source of inspiration for science. While understanding their biosynthesis can give rise to novel options for synthetic chemists, their interaction with cellular targets may serve as a starting point for pharmaceutical development.

In the first half of the thesis, the putative $FADH_2$-dependent halogenase ScORF3, a protein involved in the final chlorination step during the biosynthesis of neocarzilin A (NCA), a natural product with antiproliferative properties, was investigated. The goal of this project was to establish the structure activity relationship for the largely unknown protein. To this end x-ray crystallography on protein expressed by *E.coli* and *K. lactis* yeast cells, as well as *in-silico* structure prediction by AlphaFold were utilized to gain insight into the function of ScORF3. While the predicted structure showed similarities to known tryptophane halogenases, such as ctcP, clear differences are apparent as the protein is involved in halogenation of aliphatic, rather than aromatic carbons. Additionally, several FAD-binding domains were identified, strengthening the role of ScORF3 as a FAD-dependent halogenase.

The second half focuses on the human VAT-1 protein which is related to cancer and recently has been identified as a potential cellular target for neocarzilin A. The aim of this part of the thesis was to gain a deepened structural and mechanistic understanding of their interaction, by employing biochemical methods such as x-ray crystallography and activity based inhibition assays. The structure of the apo-protein was solved at 2.5 Å in a previously unknown space group I 4. Despite extensive cocrystallization and soaking experiments being setup, a complex crystal structure of VAT-1 and NCA unfortunately remained elusive. Thus it was not possible to directly investigate their binding modes. However, by utilizing VAT-1s' ability to reduce phenanthrenequinone in the presence of NADPH, an inhibition of its enzymatic activity through NCA could be verified. Additionally, several derivatives of NCA were tested as well, with a trifluoromethyl variant even outperforming the inhibitory properties of NCA, thereby supporting the initial hypothesis.

Table of Contents

Introduction

The membrane-anchored vesicle amine transport-1 protein (VAT-1) is overexpressed in glioblastomas, during prostate cancer, and in benign prostatic hyperplasia. Gleißner *et al.* recently identified VAT-1 as the potential cellular target of neocarzilin A (NCA), a natural product originating from *Streptomyces carzinostaticus*.[1] NCA irreversibly binds to VAT-1, thereby reducing cancer cell mobility. This interaction could serve as a possible starting point for the development of novel anti-cancer drugs. The biosynthesis of NCA is governed by the *S. carzinostaticus* biosynthetic gene cluster and was first proposed over 2 decades ago.[2] While the key player of the final chlorination step, the putative halogenase ScORF3, is known, the exact mechanism of this reaction has yet to be unraveled.

1.1 Chemoproteomics

To identify VAT-1, the key cellular target of NCA, Gleißner *et at.* employed chemoproteomics, specifically activity-based protein profiling (ABPP). This powerful technique enabled the precise identification of active protein within complex biological systems. In drug discovery two opposed approaches can be distinguished: phenotypic drug discovery (PDD) and target based drug discovery (TDD). In PDD, which is also called classical, or forward pharmacology, a phenotypic screening of natural product (NP) or small molecule libraries is carried out on animal or cellular models. The goal is to identify compounds with a potential therapeutic effect with the help of chemoproteomic methods. TDD on the other hand, is often called reverse pharmacology and is based on a general hypothesis: that a disease can be altered through modification of a specific biological target. To identify the target for a specific disease, said target is purified and then screened against NP or small molecule libraries. Positive hits cause a change in activity of the biological target and can then be tested in animals or cells.

1.1.1 Activity Based Protein Profiling

In 1999 Benjamin Cravatt developed ABPP which is popular not only for drug discovery, but also in the fields of small molecule screening, biomarker discovery, *in vivo* imaging, and for identification of cellular interaction partners.[3, 4] Activity based probes (ABPs) are used in biological samples to covalently label proteins, thereby modifying their active sites. ABPs typically consist of a warhead, a linker and a reporter group (figure 1.1 A).

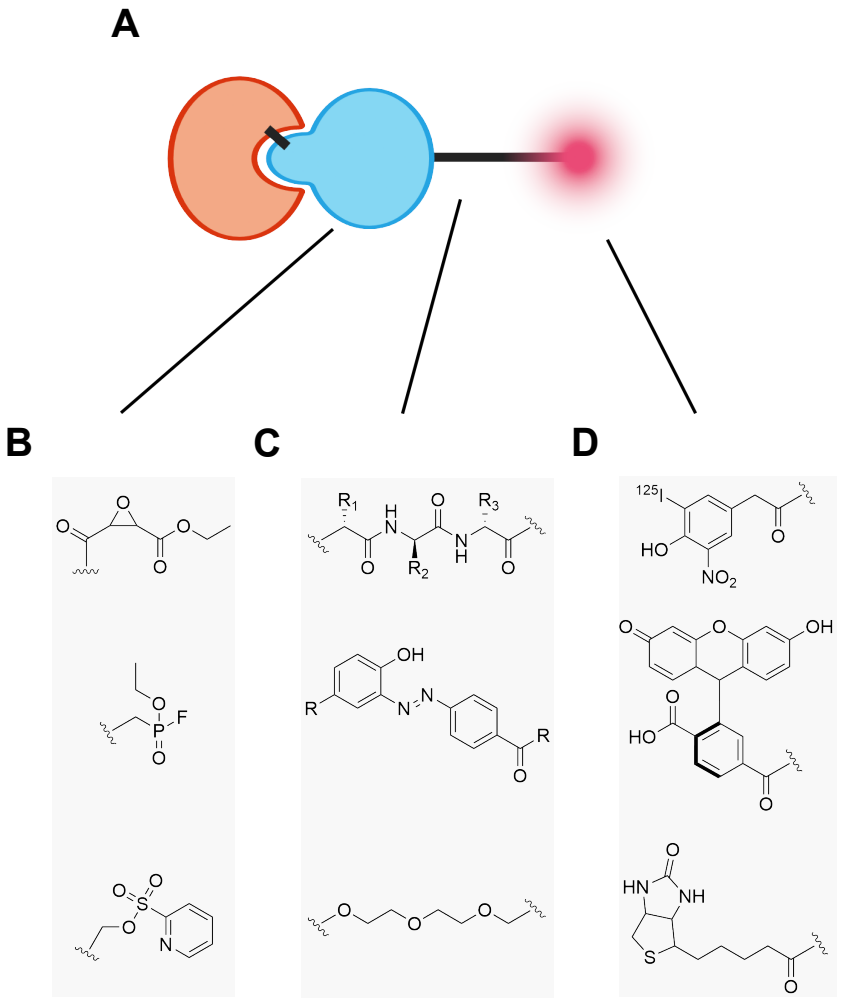


Figure 1.1: Structure of ABPs.
A The three main compontents of ABPs include a warhead (blue), a linker unit (black) and a reporter (magenta), with the target shown in orange. **B** Possible reactive warheads include epoxides, fluorophosphates and sulfonate esters.[3, 5–8] **C** Examples for linkers include peptides, cleavable diazo, or polyethylene glycol (PEG) chains.[9]. **D** Three types of probes are used either for visualization, such as ^{125}I (radioisotopes), or fluorescein (fluorescent tags), or for purification, such as biotin tag.[9]

The warhead is usually an electrophilic reactive group, covalently linking the probe to the target.[10, 11] The list of possible reactive groups is long, with examples, such as epoxides, fluorophosphates or sulfonate esters, as shown in figure 1.1 B. The design of these warheads is challenging, as they need to be reactive enough to modify the respective nucleophile. On the other hand, high reactivity leading to interactions with non-target proteins should be avoided. The linker region is a spacer, which not only connects the warhead to the probe, but also influences selectivity of the probe binding. Some linkers containing diazo or disulfide groups also allow cleavage and thereby separation of the warhead from the probe. Finally, the probe can fulfill either of two goals, enabling visualization of the biological target, or allowing purification thereof. This is achieved by addition of either of three different types of probe seen in figure 1.1 D. Radioisotopes such as ^{125}I allow for easy visualization by audiography, however short storage times and special handling conditions make it laborious to work with. Fluorophores on the other hand are usually safe to handle, albeit being susceptible to photobleaching. For the purpose of target purification and enrichment, affinity tags such as biotin, are far superior compared to fluorophores, with poor cell permeability being the limiting factor.[12] While addition of the probe *in*

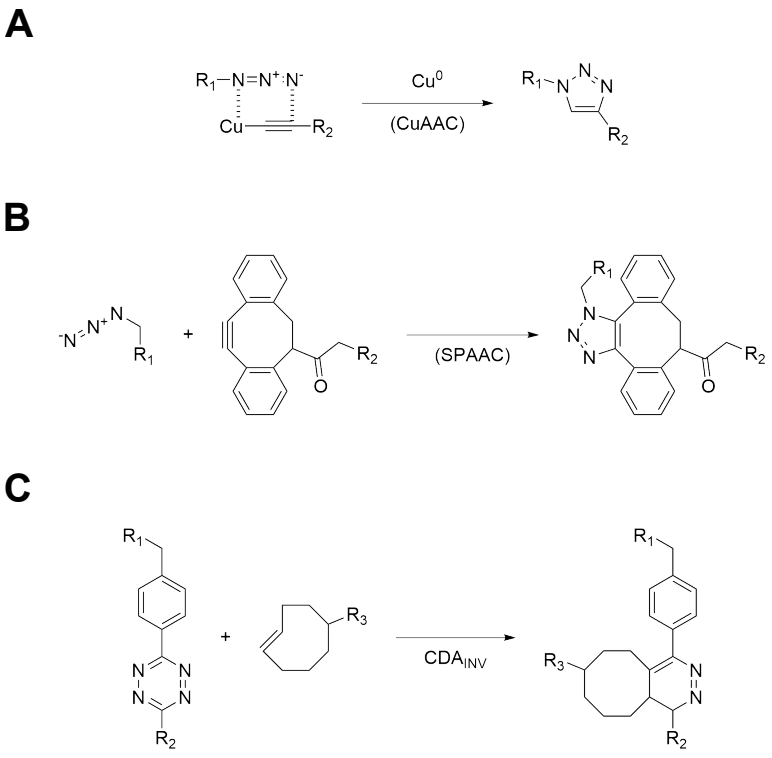


Figure 1.2: Bioorthogonal reactions used for installation of ABPs.
A CuAAC developed by Sharpless. **B** SPAAC by Bertozzi. **C**DA$_{\mathrm{INV}}$ according to Wiessler.[13–15]

vitro is rather simple, challenges arise when moving to *in vivo* conditions. Modifying living systems with minimal interference requires the usage of bioorthogonal reactions, such as the Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) developed by Barry Sharpless (figure 1.2 A).[13] Copper-free methods include strain-promoted alkyne-azide cycloadditions (SPAAC, figure) or an inverse electron-demand Diels-Alder reaction (DA_{INV}) seen in figure 1.2 B and C.[14, 15] If no covalently bound inhibitor for the biological target is possible, a probe can still be installed through photoaffinity based labelling. Upon photoirradiation a photocrosslinker forms a highly reactive state which then covalently binds to the protein of interest (POI). While benzophenones boast a high affinity towards methionine; aryl azides, or diazirines are commonly used for modification of neighboring C-H and heteroatom-H bonds.[16] Schultz *et al.* described a way to introduce genetically modified amino acids, suitable for site-specific photocrosslinking, into protein sequences by using the amber codon, TAG, in combination with a pair of aminoacyl-tRNA synthetase and tRNA.[17, 18] Depending on the mode of action, a photocrosslinker might bind to the active site, altering protein activity or ligand binding. This makes photoaffinity based labelling the method of choice, when activity based probes are not feasible.

1.1.2 Quantitative Methods

ABPP allows for qualitative analysis of protein-ligand interactions, nevertheless absolute protein levels do not equal active protein levels. The regulation of protein function within a cell requires approaches with minimal intervention, such as gel-free, or even label free methods. Isobaric tags for relative and absolute quantitation (iTRAQ), developed by Ross *et al.* uses digestion of proteins, that are labelled N-terminally and on side chains with tags of varying mass. The pooled samples are then fractioned and analyzed by tandem mass spectrometry (MS/MS), to allow identification of labelled peptides and assembly of the respective proteins. Quantification is possible due to fragmentation of the tags which keeps relative ratio constant.[19] Another option is tandem mass tag (TMT), which relies on the same principle as other isotope labelling techniques by generating pairs of chemically identical, tagged peptides. However, TMTs have a similar overall mass and comigrate in chromatographic separations. Upon fragmentation different reporter ions are obtained which can then be quantified as their relative ratio stays the same.[20] Depending on the choice of tag, one can differentiate between TMTzero, with no isotopic labelling present on the tag, TMTduplex, where isobaric pairing takes place after isotope labelling and TMTsixplex or TMT10-plex which use a larger number of isobaric mass tags (six and ten respectively).[21, 22]Additionally, dimethyl labelling (DiME) can also be used, which

labels the N-terminus as well as ϵ-amino groups of Lys through reductive amination with formaldehyde.[23]

All the methods described so far are based on chemical labelling of peptides. Stable isotope labelling by amino acids in cell culture (SILAC) presents an alternative, as it metabolically labels proteins rather than chemically. At least two cultures are grown on differently labelled media, resulting in uniquely labelled proteins upon their biosynthesis. This allows pooling of the samples early on, reducing errors due to different handling. Proteins are then analyzed by MS, where the ratio of peak intensities equals the peptide ratio in cultures.[24]

1.1.3 Derivatization-Free Methods

When no modification of the POI is possible, or desired, derivatization-free methods utilize an increase in stability upon protein binding. Thermal protein profiling (TPP) monitors the melting profile of proteins through multiplexed quantitative MS and can be used *in vivo*, *in vitro* as well as *in situ*.[25] Drug affinity responsive target stability (DARTS) utilizes reduced protease susceptibility upon drug binding.[26] In stability of proteins from rates of oxidation (SPROX), hydrogen peroxide and chemical denaturants are used to oxidize proteins, which in turn possess increased oxidation stability upon ligand binding.[27] Target identification by chromatographic co-elution (TICC) on the other hand is base upon proteins forming highly stable ligand-protein complexes, which allow chromatographic co-purification.

1.1.4 Computational Methods

In silico techniques, such as molecular docking, developed in the 1980s, are available for predicting interactions between proteins and ligands.[28] The aims of molecular docking include the accurate structural modelling and the correct prediction of activity, by predicting ligand conformation and orientation within a target binding site.[29] Potential ligands that can be identified include other proteins, nucleotides, such as DNA or RNA, metals, or other NP.[30]

Taken together, chemoproteomics can serve as the starting point for drug discovery and development through identification of protein targets for bioactive small-molecules, improvement of probe selectivity for protein screening, for example against a native proteome, or by providing platforms for screening of small-molecule libraries against poorly characterized proteins.[31, 32]

1.2 Natural Products in Drug Development

In the earlier stages of drug discovery, compound libraries play a crucial role in the detection of potential drug candidates. An initial selection is often made from the synthetic libraries by computational methods.[33] These libraries often contain a wide variety of compounds, such as already known drugs, or natural products (NP).[34] Historically, NPs have been a source for pharmaceutically relevant compounds, either as drugs themselves, or as an inspiration for novel drugs.[35] Therefore they fulfill a key role in drug discovery, especially in cancer treatment, but also in cardiovascular and metabolic diseases.[36–38]

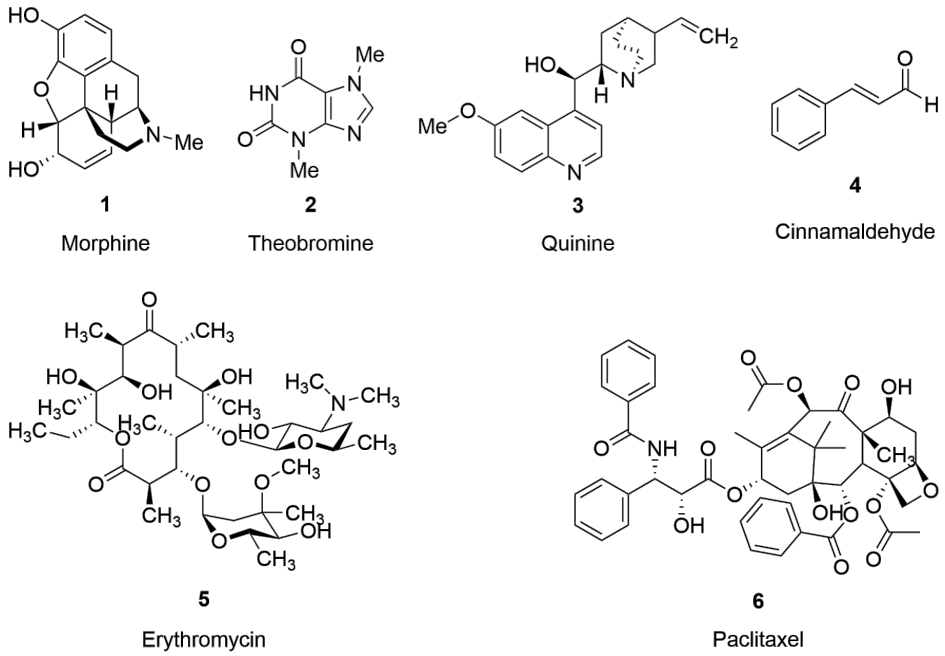


Figure 1.3: Example for classes of NPs.
Lewis structures of alkaloids morphine (**1**), theobromine (**2**), quinine (**3**), phenylpropanoid cinnamaldehyde (**4**), polyketide erythromycin (**5**) and terpenoid paclitaxel (**6**).

In nature, NPs are the result of a constant struggle for resources and defense against predators among plants, microbes and fungi.[39–41] These organisms, found both in soil and in marine environments, are under perpetual pressure to produce novel and more sophisticated chemical compounds, to ensure their own survival. Many of which can be isolated and utilized by humans for similar purposes, such as treating bacterial, or fungal infection, or combating proliferating cancer cells.[42–44]

NPs encompass a wide range of organic compounds originating from living organisms. Classifications may vary depending on the context in which they are discussed. Roberts and Caseiro suggested four different categories for classifying NPs: recurring structural features, the genus of the plant source, the physiological effect of the NP, or its biosynthesis.[45] For example, classification according to biosynthesis divides NPs into four groups (figure 1.3): the highly diverse alkaloids (figure 1.3 **1**, **2** and **3**), which contain at least one nitrogen atom but have no uniform classification otherwise. Phenylpropanoids like cinnamaldehyde (**4**) are derived from phenylalanine and tyrosine via the shikimic acid pathway and play a major role in plant defense mechanisms.[46] Polyketides such as erythromycin (**5**) are synthesized from acetyl-CoA in the acetate pathway and contain alternating carbonyl and methyl groups.[47, 48] Terpenoids like paclitaxel (**6**) are based on isoprene and its derivatives and are produced in the melvalonate methylerythritol phosphate pathway. On the other hand, classification based on underlying structural features gives rise to a vast array of groups such as aminoacid based structures (peptides, proteins), aromatic compounds (e.g. flavonoids, polyketides), biogenic amines, carbohydrates, glycosides, heterocycles such as nucleosides or alkaloids, isoprenoids (terpenoids, steroids, polyprenylhydroquinones), lipids or peptidoglycanes. However, issues arise with more complex compounds like glycoproteins which fit into multiple classes.

Over the last four decades, NPs have been promising candidates, either directly as drugs, or as a starting point for drug development. They originate from various sources, including bacteria, animals, fungi and plants, with the latter accounting up for 50% of the potential drugs identified in 2008. Approximately 40% of these drugs were intended for cancer treatment.[49] However, novel ways for isolating, screening and characterizing NPs have allowed them to maintain their relevance as a starting point for drug development, despite an initial decline in the 1990s.[50, 51]

1.3 Neocarzilins

Neocarzilins are a class of NPs first reported in 1992 by Nozoe *et al.*. Initially, only two members, neocarzilin A (NCA) and neocarzilin B (NCB), were identified by NMR and MS from *Streptomyces carzinostaticus var.* F41 (figure 1.4 A), which is commonly found in soil.[53] It was not until years later, that Otsuka *et al.* discovered a third member, neocarzilin C (NCC), during their efforts to characterize the genecluster responsible for the biosynthesis of neocarzilins (NCZs). They also discovered a novel type I polyketide syn-

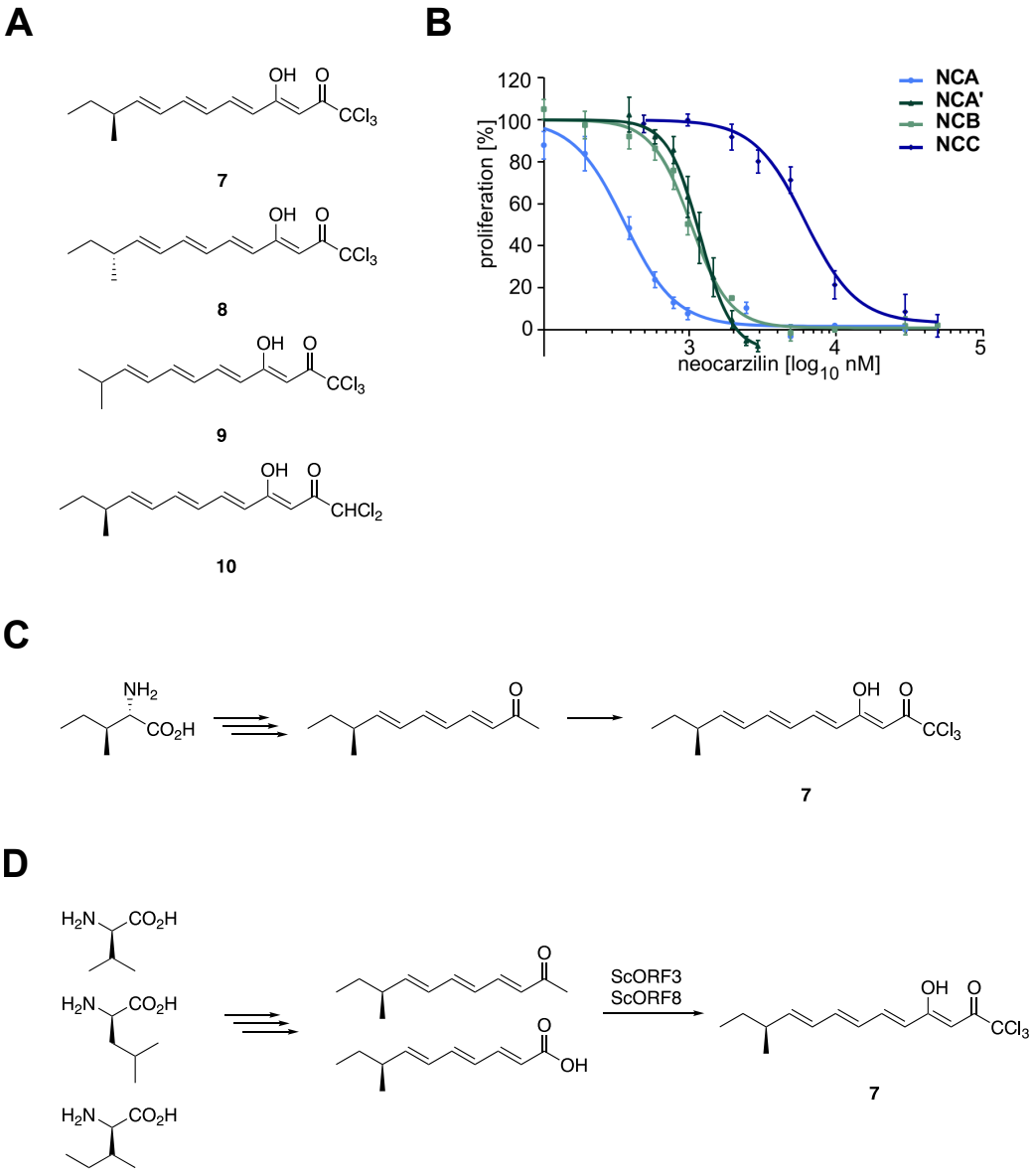


Figure 1.4: Types of neocarzilins.
A Neocarzilins A (**7**), A' (**8**), B (**9**) and C (**10**). **B** Antiproliferative activites of NCZs in MDA-MB-231 cells as reported by Gleißner *et al.* **C** Total synthesis of NCA by Nozoe **D** involved in the biosynthesis of NCA starting from L-valine, L-leucine, or L-isoleucine.[1, 2, 52]

thase (PKS, figure 1.4 A) in this process.[2] All three NCZs are long polyenones featuring a terminal trichloromethyl group, which is introduced in the final step of their biosynthesis through a halogenation reaction, catalyzed by a pair of enzymes (ScORF3/8). In 1992, Nozoe *et al.* demonstrated cytotoxic activity first against K562 chronic mylogenous leukemia cells with an IC_{50} of 0.06 µg/mL, proving to be nearly as effective as neocarzinostatin, which has an IC_{50} of 0.09 µg/mL. This cytotoxicity can partly be attributed to the presence of an enolic hydroxyl group making it slightly acidic, as the corresponding methylether was shown to be less efficient against cancer cells (IC_{50} = 2 µg/mL).[53] While the total synthesis of NCA was achieved in 1992, starting from L-isoleucine (figure 1.4 B), there is still no data on the catalytic mechanism of the biosynthesis to date.[52] The structure-activity relationship (SAR), crucial for drug development, remains unclear. However, recent studies have shown that NCZs inhibit VAT-1, exhibiting antiproliferative and antimigratory abilities, as seen in figure 1.4 C.[1]

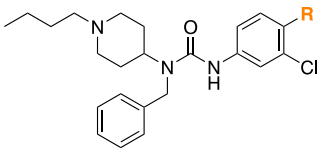
Biosynthesis

2.1 Introduction

2.1.1 Chlorination and Halogenation in Pharmaceuticals

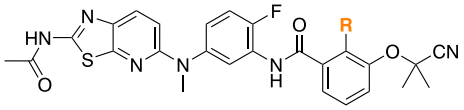
The presence of chlorine in neocarzilins highlights a broader trend observed in pharmaceutical chemistry, where chlorination is pivotal in enhancing the efficacy and stability of various compounds. These compounds often demonstrate formidable adsorption, distribution, metabolism, excretion and toxicity (ADMET) properties, crucial in the later phases of preclinical development.[54] Chlorination, as well as halogenation of C-H bonds in general, introduces a dipole along the bond axis. The strength of the dipoles formed follows the same trend as the respective halogen's electronegativity, with the C-F bond being strongly polar up to the barely polarized C-I bond.[55] This change in polarity may increase water solubility, or lipophilicity of the halogenated molecule.[56, 57] In a pharmacokinetic context, an increased water solubility causes better bioavailability and might also improve excretion of a drug. Better lipophilicity, on the other hand, increases membrane permeability. Thirdly halogenation might also increase a drugs stability, as well as blood-barrier permeability.[58] In the case of chlorination, Hammill *et al.* showed improved potency in the form of low half maximal inhibitory concentrations (IC_{50}), as well as $t_{1/2}$ and intrinsic clearance (CL_{int}), with the latter two being key pharmacokinetic parameters used to determine a patients drug dose (figure 2.1 A).[59–61] The area under the curve (AUC) gives the total exposure to a drug over time, a value important for comparison whether different means of administration of the same drug dose result in varying degrees of exposure. In the case of the pyridine derivatives in figure 2.1 B, a hydrogen-chlorine exchange lead to a 46 fold increase in AUC without decreasing other desirable pharmacological parameters.[62] Candidates containing chlorine possess increased apparent permeability (P_{app}) and reduced efflux ratio compared to non-halogenated compounds, or those with cyanide groups (figure 2.1 C).[63] [64] In orally active compounds chlorination was shown to reduce clearance, thereby increasing oral exposure in rats (figure 2.1 D). As a substituent, chlorine is often compared to fluorine or the methyl group, since it brings characteristics of both substituents

A



R	IC_{50} (nM)	$t_{\frac{1}{2}}$ (hr)	CL_{int} $(\frac{mL}{min \times kg})$
H	358	≤ 0.1	≥ 500
Cl	45	2.11 ± 0.336	25.6 ± 4.08

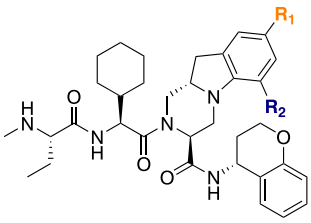
B



R	IC_{50} (nM)		AUC PO[1] $(\frac{\mu g \times h}{mL})$
	BRAF	VEGFR2	
H	23	10	0.311
Cl	45	14	14.308

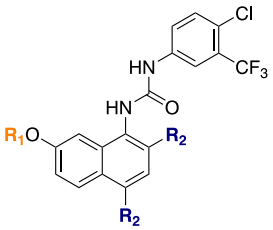
[1] PO: *per os* (by mouth)

C



R_1	R_2	P_{app} $(\frac{nm}{s})$	ER
H	H	25	8.2
CN	H	14	13
Cl	H	64	1.4
Cl	Cl	75	0.69

D



R_1	R_2	IC_{50} (nM)	Clearance $(\frac{l}{h \times kg})$
CH_3	H	19	2.0
H	Cl	2.1	0.16

Figure 2.1: Pharmacokinetic parameters upon chlorine introduction.
A *n*-Butylpiperidines suggested for inhibition and regulation of defective in cullin neddlyation 1-mediated cullin neddylation.[59] **B** 5-Amino-linked [1,3]thiazolo[5,4-*b*]pyridine derivatives designed as RAF/vascular endothelial growth factor receptor 2 inhibitors.[62] **C** Potential inhibitors of apoptosis protein antagonits, (*S*,10a*S*)-hexahydropyrazino[1,2-*a*]indole derivatives.[63] **D** Naphthol-based antagonists of vanilloid receptor TRPV1, an ion channel gating physical and chemical stimuli.[64]

as seen in table 2.1. With the electronegativity of chlorine being close to fluorine, both can act as electron withdrawing groups which is seen in their Hammett σ_m parameter. While chlorine retains this ability when installed in a *para* position, fluorine boasts elec-

Table 2.1: Parameters for H, F, Cl and CH_3 functional groups.[68]

functional group	H	F	Cl	CH_3
electronegativity	2.20	3.98	3.16	2.55
van der Waals radius (Å)	1.10	1.47	1.75	2.0
molecular refraction	1.03	0.92	6.03	5.65
Hammett σ_m constant	0.00	0.34	0.37	-0.07
Hammett σ_p constant	0.00	0.06	0.23	-0.17
bond length in C-X bond (Å)	1.09	1.35	1.77	1.54
dipole moment in C-X bond $(\frac{\mu}{D})^*$	0.3-0.4 (CH_4)	1.85 (CH_3F)	1.90 (CH_3Cl)	0

* [69–71]

tron donating characteristics due to its π-donating character.[65] Together with its *van der Waals* radius being significantly larger than fluorine, similar to CH_3, it combines favorable characteristics of both groups often making it the preferred choice as a functional group. Comparing their respective molar refraction, a parameter which yields information about the electronic polarization of ions, chlorine with 6.06 is just above a methyl group (5.65), far higher than fluorine (0.92).[66, 67] The dipole moment of the C-Cl bond is larger than for C-F, due to the combination of a logner C-Cl bond while having similar dipole moments.

2.1.2 Putative Halogenase ORF3 and Gap in Knowledge

The *Streptomyces carzinostaticus* gene cluster involved in the synthesis of NCZs contains 14 open reading frames (ORF). The function of several ORF can only be assumed from sequence similarities, such as ORF3, which is located upstream of the type I PKS gene. It encodes a putative $FADH_2$-dependent halogenase, responsible for chlorination of dechloroneocarzilins, as proposed by Otsuka *et al.* (figure 2.2).[2] A typical nucleotide binding motif, GXGXXG, is also found close to its N-terminus.[72] The mechanism underlying this reaction is still unknown, however, the enzyme possesses regulatory function over earlier steps of the biosynthesis. Disruption of ORF3 not only prevents NCZ production, it also eliminates the presence of the precursor dechloroneocarzilin.[2]

Unraveling the reaction mechanism could also provide more insight into eznymatic halogenations of aliphatic carbon. While halogenation of aromatic or olefinic centers are well understood, modification of aliphatic chains are far less studied.[73] Till date functionalization of inactivated alkyl chains remains challenging, meaning mechanisms used by

enzymes are of particularly great interest for synthetic chemists. Knowledge of the enzymatic mechanism and the structure-activity relationship could open up ways to halogenate other substrates as well.

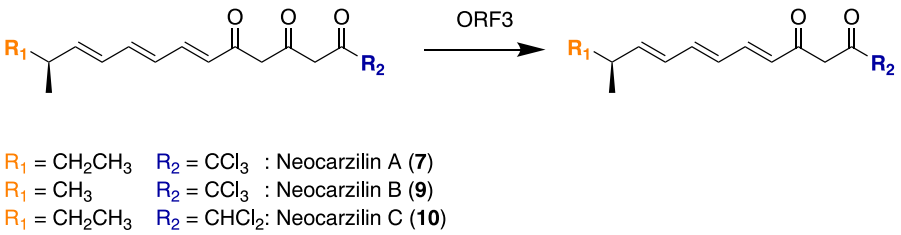


Figure 2.2: Final step in NCZ biosynthesis.
Chlorination of dechloroneocarzilins by putative halogenase ORF3.[2]

2.2 Aim & Objectives

The putative FAD-dependent *streptomyces carzinostaticus* ORF3 (ScORF3) halogenase is involved in the final step during the biosynthesis of NCZs. The 62 kDa protein ScORF3 is encoded by ORF3 of the *S. carzinostaticus* gene cluster, however, the mechanism of this chlorination reaction is still unknown. Its understanding could give rise to novel ways for halogenation of aliphatic compounds. The aim of this work is to shine light on the mechanism, by establishing the SAR of ScORF3. To this end the structure of the protein is to be solved and investigated through comparison with homologous FAD-dependent halogenases. In order to carry out functional and structural studies via x-ray crystallography, large amounts of pure protein are required. With *E. coli* and yeast cells, two different systems for heterologous protein expression are at hand and protein expression in these systems will be optimized.

2.3 Results & Discussion

For functional and structural studies, large quantities of homogeneous protein are required. Therefore the gene encoding for the putative halogenase ScORF3 was cloned and expressed in two different expression systems *E. coli* and the *Kluyveromyces lactis* yeast expression system.

2.3.1 Protein Expression

Heterologous Protein Expression in *E. coli*

The initial construct for expression of ScORF3 was a fusion protein consisting of a Strep tag II, a tobacco etch virus (TEV) site for tag removal and the protein sequence. While expression in autoinduction (AI) medium was successful at 25 and 37 °C, ScORF3 could only be observed in inclusion bodies (figure 2.3).

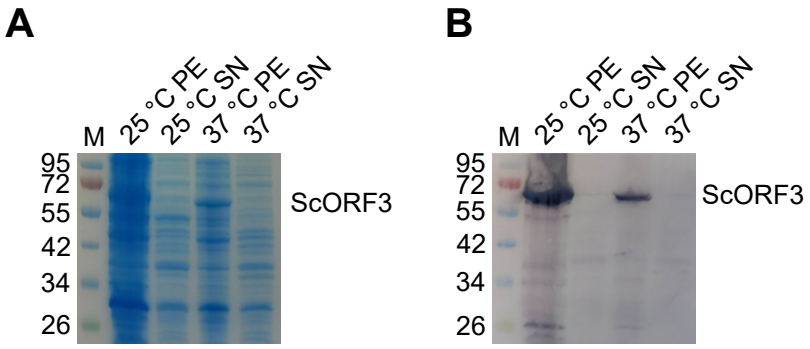


Figure 2.3: Expression of Strep-TEV-ScORF3.
A Coomassie-stained polyacrylamide(PAA)-gel of Strep-TEV-ScORF3 (64.8 kDa) expressed in BL21(DE3) cells. **B** Western blot showing Strep-tagged proteins. M: protein standard; 25 °C PE: pellet after lysis, expression at 25 °C; 25 °C SN: soluble fraction after lysis, expression at 25 °C, 37 °C PE: pellet after lysis, expression at 37 °C; 37 °C SN: soluble fraction after lysis, expression at 37 °C.

When expression of recombinant proteins overwhelms the bacterial quality control machinery, misfolded proteins aggregate inside the cells.[74] They contain the overexpressed protein as well as a multitude of other cellular proteins in the form of inclusion bodies.[75, 76] Since the protein of interest can make up to 50% of the aggregated proteins it opens up the possibility for extraction from inclusion bodies. Protein recovery is a two-step process, the first step being solubilisation from the aggregates, followed by refolding into its native

conformation. Solubilisation is usually carried out by addition of detergents or chaotropic agents. While the former interrupt hydrophobic interactions, the latter are also capable of breaking hydrogen bonds and destroying the secondary structure, thereby bringing proteins in solution. We achieved successful solubilisation of ScORF3 when treated with 2% sodium dodecyl sulfate (SDS) for 16 h. Protein stayed in aggregates when exposed to 1% (w/v) Tween-20, 2% (w/v) N-lauroylsarcosine, 5% (v/v) DMSO, 6 M guanidinium hydrochloride or 6 M urea. Extraction under milder conditions as low as 2 M urea was reported using the freeze-thawing method reported by Qi *et al.*. The resuspended inclusion bodies were frozen and thawed again, before separation of undissolved aggregates.[77] However, no solubilisation was observed until 7 M urea with this method, rendering it less ideal compared to usage of 2% (w/v) SDS.

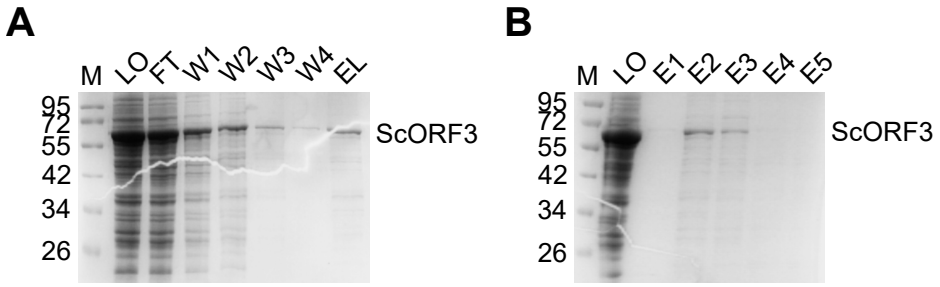


Figure 2.4: Purification of Strep-tagged ScORF3 after isolation from inclusion bodies. Coomassie-stained PAA-gels of Strep-TEV-ScORF3 (64.8 kDa) after **A** affinity purification. M: protein standard; LO: solubilised inclusion bodies; FT: unbound fraction; W1-4: washing fractions; EL: elution fraction; **B** gel filtration chromatography. M: protein standard; LO: loaded protein; E1-5: elution fractions.

Following solubilisation, the protein was refolded by dialysis against a detergent-free buffer, before Strep tag affinity chromatography purification (figure 2.4 A). Although solubilisation of ScORF3 was successful, purification yield was sparse, as most protein did not bind to the column matrix. In the case of ScORF3, low binding affinity was frequently observed, when solubilized from inclusion bodies. However, it did not correlate to the choice of solubilisation agent, or methods like freeze-thawing. This impacted subsequent size exclusion chromatography (SEC), which is shown in figure 2.4 B. Only negligible amounts of protein were obtained this way and the protein was still not being sufficiently pure for crystallization. While it was possible to isolate ScORF3 this way, the poor performance of affinity chromatography impeded following purification steps. Without sufficient amounts of protein, methods like SEC or ion-exchange chromatography would be futile, as high amounts of protein are necessary for subsequent screening for optimal crystallization conditions. We suspected refolding of the solubilised protein to be the issue. It was likely that only an

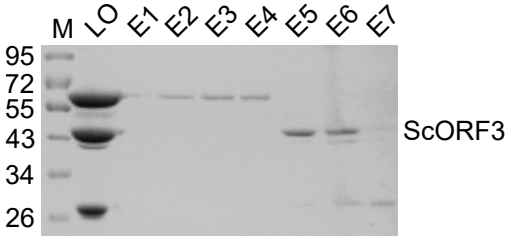


Figure 2.5: Final gel filtration step in the purification of ScORF3.
Coomassie-stained PAA gel of ScORF3 (61.7 kDa) after gel filtration for removal of MBP affinity tag and TEV protease. M: protein standard, LO: TEV-cleaved protein, E1-E7: elution fractions.

insignificant part of the proteins had accessible Strep tags, which were able to successfully interact with the stationary phase.

Although heterologous expression of ScORF3 in *E. coli* resulted in the formation of inclusion bodies, overall expression levels were satisfactory. To enhance solubility, we incorporated a maltose binding protein (MBP) tag onto ScORF3. MBP is a highly soluble, 42.5 kDa protein found in *E. coli* where it is part of the maltose catabolism and responsible for its transport.[78] Our construct consisted of the MBP tag, a TEV cleavage site for affinity tag removal and ScORF3. Following expression in two yeast extract tryptone (2YT) medium, affinity purification yielded high amounts of soluble protein. Cleavage by TEV protease was necessary, since the sterically demanding MBP tag would affect crystallization. Finally, pure protein could be obtained by gel filtration, as seen in figure 2.5. Solubility remained an issue throughout purification, especially after removal of the MBP tag, where ScORF3 was lost due to continued precipitation.

Heterologous Expression and Secretion of ScORF3 in *K. lactis*

As an alternative to protein expression in bacterial cells, we used the *K. lactis* yeast system from NEB. Compared to bacterial expression systems, yeast cells grow slower and are more laborious to handle, however, they give access to eukaryotic folding-machinery with their own set of *in-vivo* posttranslational modifications. The *K. lactis* strain presents several advantages over alternatives, such as *P. pastoris* or *S. cerevisiae*, rendering it the preferred choice for avoiding clustering of protein aggregates in inclusion bodies. Firstly, *K. lactis* contains the strong P_{LAC4} promoter, a variant of which (P_{LAC4}-PBI) is present in the pKLAC2 plasmid used for integration of ScORF3 in the yeast genome. P_{LAC4}-PBI features a mutated Pribnow box consensus sequence, a bacterial promoter, which is not active in *E. coli*, yet retains its full function in *K. lactis*.[79] Secondly, proteins expressed

Table 2.2: Domains identified for ScORF3 according two databases.

Domain Name	Residues	E-value[1]	Database
Flavin-dependent halogenase	121-350	-	Pfam/InterPro
FAD/NAD(P)-binding	4-197 316-379	-	Pfam/InterPro
Flavin-dependent halogenase in NP biosynthesis	3-430	-	Pfam/InterPro
FixC	14-380	$1.15e^{-36}$	NCBI
Trp_halogenase	7-247	$1.12e^{-12}$	NCBI
UbiH	3-193	$1.84e^{-12}$	NCBI
FAD_binding_3	6-186	$4.5e^{-7}$	NCBI

[1] scoring value for quality of homologue matches, lower E-values indicate better homology. E-values of 0.01 or less are usually considered good hits.

in *K. lactis* containing an α-mating factor (MF) secretion domain will be processed in the Golgi, leading to secretion from the yeast cells (figure 2.6 A).[80] This allows isolation from the nutrition medium, skipping lysis, and further preventing aggregation inside the cells. Finally various enzymes in *K. lactis* were identified, which improve protein folding.[81]

ScORF3 was inserted into the yeast genome by tranformation of a linear expression cassette, containing the cloned gene and a gene fragment coding for acetamidase (*amdS*), which serves as a selection marker (figure 2.6 B). Patches of properly integrated cells were used for cultivation. The protein was expressed in and secreted from yeast cells into YPGal medium over the course of 5 days. Secreted proteins were precipitated with $(NH_4)_2SO_4$, solubilised and purified by affinity chromatography using Strep-tactin as a stationary phase. Alternatively, the nutrition medium was directly applied to the resin. While the protein could be precipitated from the nutrition medium, overall affinity of ScORF3 for the resin was extremely low and it was not possible to isolate it by affinity chromatography (figure 2.6 C).

2.3.2 Crystallization

We attempted crystallization at protein concentrations of 2.5, 3.0, 7.0 and 8.0 mg/mL, however, only non-diffracting crystals were obtained under these conditions. Based on the

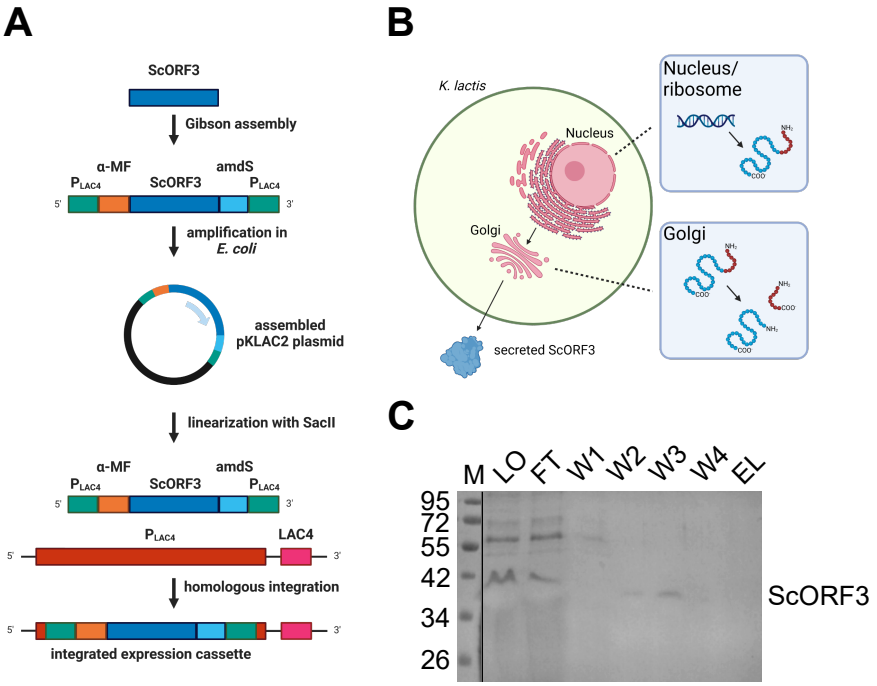


Figure 2.6: *K. lactis* yeast system for expression of ScORF3.
A Cloning strategy for insertion of ScORF3 gene containing expression cassette in the yeast genome. **B** Expression of ScORF3 containing the α-MF secretion domain, followed by its removal in Golgi and subsequent secretion of the protein. **C** Coomassie-stained PAA-gel of secreted proteins isolated from nutrition medium by precipitation and affinity chromatography. M: protein standard, LO: precipitated proteins, FT: unbound protein, W1-4: washing fractions, EL: elution fraction.

primary structure of ScORF3 we identified several conserved domains as shown in table 2.2. Both databases, InterPro and NCBI, identified a variety of FAD-binding, or FAD-dependent domains. FixC (figure 2.8, C, orange), which was the best hit according to the NCBI database, is a protein superfamily whose members are all flavin dependent hydrogenases, while UbiH (figure 2.8, D, orange) is based on 2-polyprenyl-6-methoxyphenol hydroxylase and contains related FAD-dependent oxidoreductases. The relatively low E-value for the FAD_binding_3 domain (figure 2.8, B, orange) may be due to the rather common motif present in a multitude of proteins. The Trp_halogenase domain (figure 2.8, E, orange) on the other hand is a set of conserved NADH-dependent halogenases, which catalyze the reaction from Trp to 7-Trp. This relationship is supported by sequence alignments of ScORF3 with the NCBI database, which gave 9 homologues with over 90% sequence identity and a bit score of over 1000. An identity of $>$ 30% and a bit score of at least 50 are commonly required for homology.[82] The majority of these homologues belong to Trp 7-halogenases, while one hit belonged to the D-amino acid oxidase, a FAD-

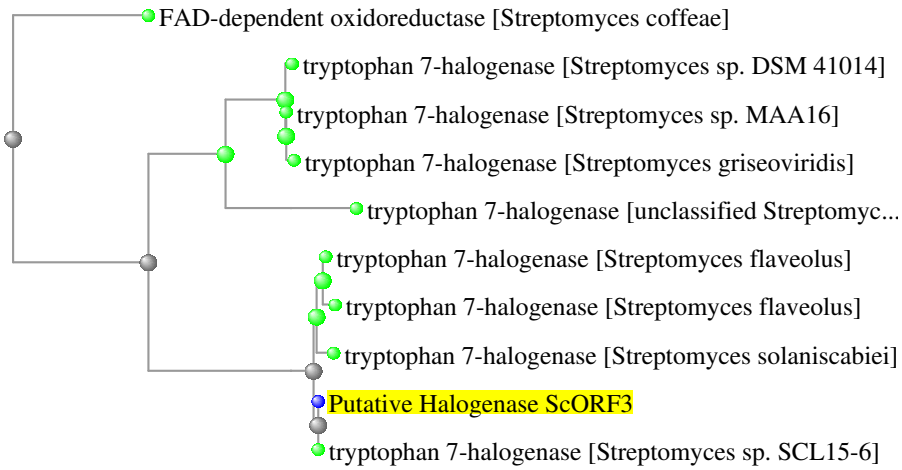


Figure 2.7: Family tree of homologues for ScORF3.
Generated according to sequence homologues identified by NCBI.

dependent oxidoreductase.[83] All homologue proteins found this way originate from various *Streptomyces* strains (figure 2.7). These findings are consistent with Otsuka *et al.* who initially found ScORF3 together with a putative flavin reductase (ORF8). These reductases are necessary to generate reduced flavin *in situ*, a requirement for the functionality of FAD-dependent proteins.[2] To improve crystallization we added 10 µM FAD as a cofactor to the lysis buffer. Binding of FAD might lead to more stable ScORF3 conformations, thereupon improving crystal quality.[84] Crystallizations were set up, at protein concentrations of 2.0 and 4.0 mg/mL, however, only non-diffracting crystals were obtained again. Based on the sequence of ScORF3 a folded structure was predicted using the AlphaFold model, shown in figure 2.8 A.[85] We compared the model to the published structures of three halogenases, PrnA (pdb code: 2ARD), RebH (pdb code: 2OA1) and CtcP (pdb code: 7XGB). PrnA and RebH are both FAD-dependent tryptophane 7-halogenases, from *Pseudomonas fluorescens* and *Lentzea aerocolonigenes*, respectively.[86–88] However they possess a rather low sequence similarity with 20% for PrnA and 19% for RebH. This correlates with poor root mean square deviation (RMSD) of 10.2 for PrnA and 9.5 for RebH when we attempted to align their structures to the predicted ScORF3. CtcP on the other hand is a halogenase from *Streptomyces aureofaciens* involved in the chlorination of tetracycline to chlorotetracycline.[89] While it stays behind previously mentioned tryptophane 7-halogenases from other *Streptomyces* strains with 45% sequence similarity, its folded structure was largely similar to the predicted ScORF3 structure with an RMSD of 0.965 (figure 2.8, F).

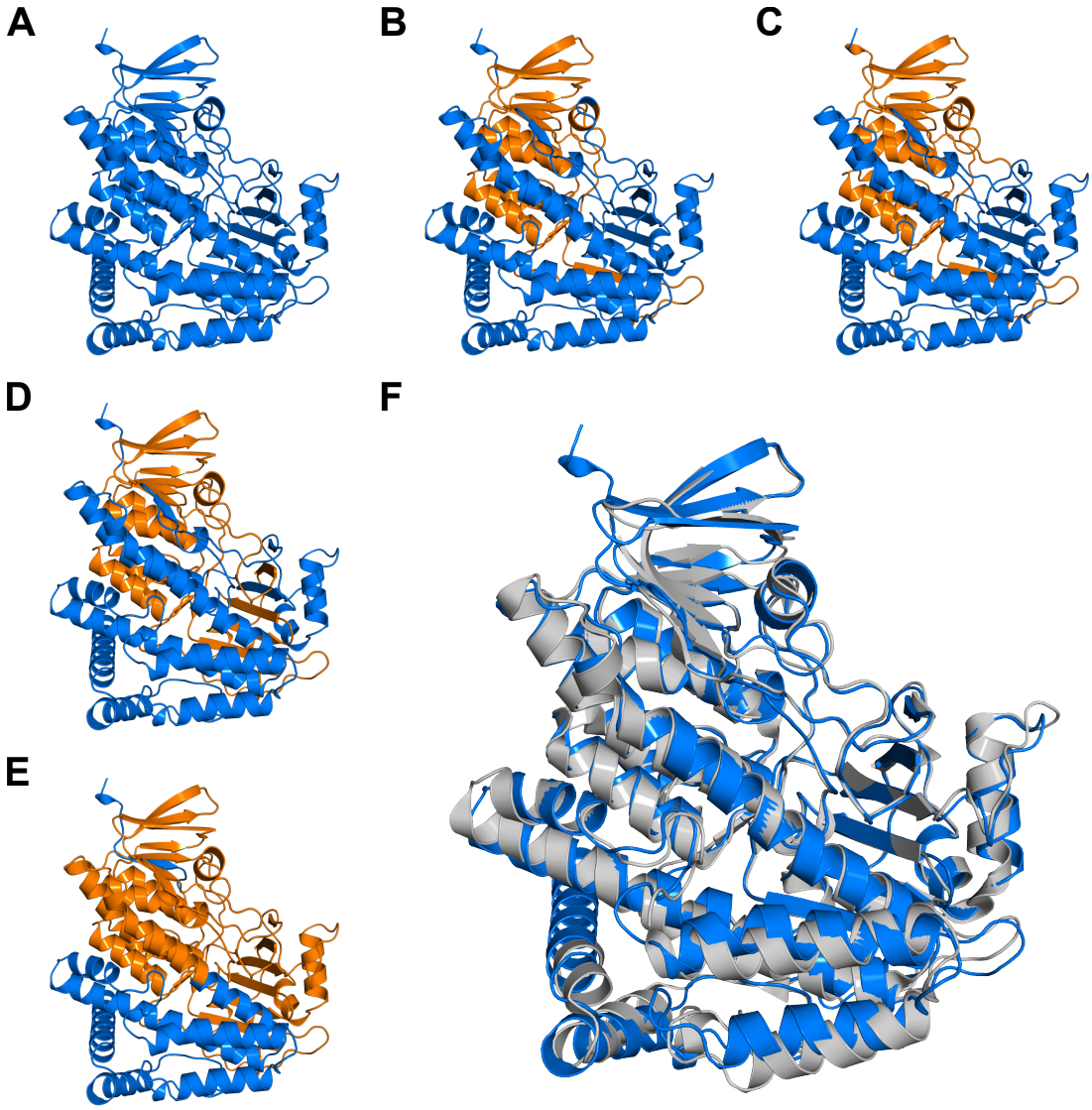


Figure 2.8: ScORF3 crystal structure and domains based on a AlphaFold prediction. **A** Folded ScORF3 as predicted by AlphaFold. Possible FAD-binding (**B**), FixC (**C**), UbiH (**D**) and Trp_halogenase (**E**) domains colored in orange. Domains are based on sequence similarities. **F** Superposition of ScORF3 (blue) and CtcP (pdb code: 7XGB, grey).

2.4 Summary

The aim of this chapter was to investigate the SAR of the putative halogenase ScORF3. The protein is involved in the chlorination of dechloroneocarzilins during the final step of neocarzilin biosynthesis, through a yet unknown mechanism. To conduct structural studies via x-ray crystallography, batches of highly pure protein were required. With *E. coli* and yeast cells, two different systems for heterologous protein expression were compared. ScORF3 contains several domains conserved among FAD-dependent halogenases and FAD-binding proteins indicating an FAD-dependant function of its own. Development of a successful expression and purification strategy for ScORF3 was largely dictated by its tendency to precipitate in the form of inclusion bodies. Isolation from the aggregates was achieved by solubilisation with 2% SDS, however, subsequent purification yielded only negligible amounts of protein, insufficient for crystallization. We avoided formation of inclusion bodies in *E. coli* by introduction of a solubility enhancing MBP tag. Using this fusion construct we obtained enough highly pure protein for crystal screening. Protein expression in the *K. lactis* yeast strain required an α-MF domain, for secretion of the protein into the nutrition medium. While isolation from the medium was successful, insufficient binding affinity for Strep-tactin was observed. This issue was previously encountered with inclusion bodies from *E. coli* cells, where it could be overcome by using solubility enhancing MBP-tags. Crystallization screening yielded several crystals of the protein alone and when isolated in the presence of cofactor FAD, nevertheless, all tested crystals were exclusively non-diffracting. A folded structure of the protein was predicted using AlphaFold. Comparison with homologous proteins revealed high similarity with the CtcP halogenase from *Streptomyces aureofaciens*, involved in chlorination of tetracycline, with an RMSD of only 0.965.

Cellular Target VAT-1

3.1 Introduction

Recently, the Sieber group identified the membrane-anchored protein VAT-1 as the cellular target of NCA.[1] They showed that presence of VAT-1 during treatment with NCA leads to a reduction in cancer cell migration. Furthermore, they reported Talin-1, a key protein in cellular migration, to be a direct interaction partner. This could pose as a possible starting point for the development of novel anti-cancer drugs.[1]

3.1.1 Background

Membrane Proteins

Membrane proteins (MPs) are an essential part of all cells and fulfill a variety of roles, making these vital hubs targets for development of novel drugs.[90] There are generally speaking two types of MPs, peripheral, or associated MPs, which are only temporary attached to a membrane, and integral MPs, which permeate the membrane. The latter can be broken down even further, in transmembrane proteins (e.g. ion channels) and integral monotopic proteins, like the phosphoglycosyl transferases.[91, 92]

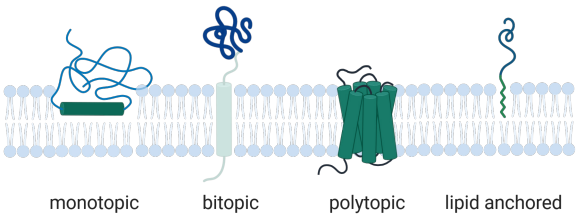


Figure 3.1: Membrane protein topology.
Four groups of structurally distinct membrane proteins, according to their topology.

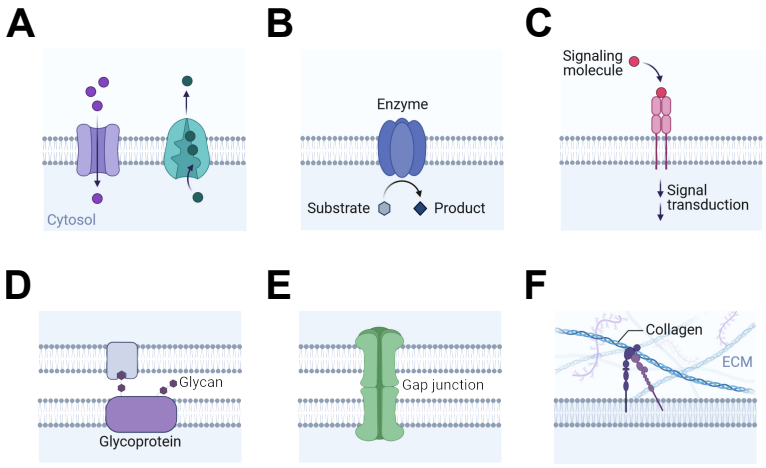


Figure 3.2: Functions of membrane proteins.
A Transport of molecules throughout the membrane. **B** Enzymatic transforming substrates into products. **C** Signal transduction across the membrane. **D** Cell-cell recognition based on glycan structures. **E** Intercellular joining of adjacent cells. **F** Anchorage and attachment of cells to the ECM.

According to the transporter classification database, there are currently around 100 superfamilies with over 1900 transporter families. Structurally, MPs can be divided into four groups, according to their topology: monotopic, bitopic, also referred to as single pass MPs, polytopic or lipid-anchored MPs as shown in figure 3.1.[93] MPs are involved in a variety of dynamic cellular processes, ranging from ionic and molecular transport, signal transduction and enzymatic activity to extracellular functions, such as cell-cell recognition, intercellular joinings and anchorage, or attachment of cells to the extra cellular matrix (ECM), or cytoskeleton (figure 3.2).[94] A well-known family of integral membrane proteins are G-protein coupled receptors (GPCRs), which consist of seven transmembrane domains. They are the largest family of receptors and membrane proteins in mammals and are involved in signal transduction across the membrane.[95] While diffusion across membranes exists, most molecules, ions and electrons are transported across membranes by transport proteins. This process can either be active, when energy is consumed (e.g. ABC transporters), or passive, in the form of facilitated diffusion, or osmosis. The solute carrier protein superfamily, for example, consists of over 400 proteins, a large portion of which are either passive, or secondary, active transporters.[96] Their important role in essential cellular processes makes MPs prime candidates for drug development with about 60% of of drug targets being located at the cell surface.[97] Yet characterization of these proteins remains challenging, mostly due to their problematic purification, compared to water-soluble proteins. Especially integral membrane proteins are highly hydrophobic and

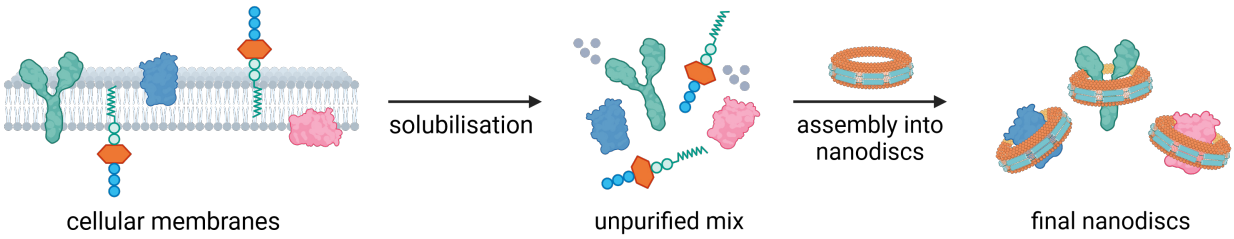


Figure 3.3: Assembly of Nanodiscs.
Solubilisation of membrane proteins followed by self-assembly into nanodiscs.

have the tendency to form aggregates.[98] Common ways to purify MPs include solubilization, or the application of nanodiscs. The former typically involves formation of a protein-lipid-detergent complex, which, under the right conditions, allows similar workup as for water soluble proteins. Therefore, nanodiscs on the other hand allow for purification of MPs under almost-native conditions. Membrane scaffolding proteins (MSPs) are used to capture artificial phospholipids in a disc-like structure, mimicing a cellular surface. Active MPs can then be trapped in the nanodisc in a stabilized environment. Alternatively, synthetic polymers, such as diisobutylene-maleic acids (DIBMA), styrene-maleic anhydride (SMA), or poly(acrylic acid co-styrene) (AASTY) are used to trap lipids from native cells, resulting in a more native-like environment.[99–101]

VAT-1 State of Knowledge

VAT-1 was first isolated from electromotor neurons of pacific electric ray, *Torpedo californica* in 1989 and could later be localized as a high-molecular-weight complex of about 176 kDa in synaptic vesicle membranes.[102, 103] Speculations about the complex consisting of four 41 kDa VAT-1 monomers were partially confirmed about 30 years later. Since then VAT-1 homologues were found in a variety of other organisms, such mammals, birds, turtles, frogs and fishes.[104] In zebrafish *Danio rerio*, it was expressed in the brain, pharynx and gut, with a possible relation to organ development.[105] However, so far no VAT-1 counterpart could be found in yeast cells.[106] In humans VAT-1, was detected in a variety of locations, often related to formation of cancer. In epithelial tissue, such as skin, mammary glands or ovaries, Koch *et al.* suggested an involvement in signalling cascades.[107] A suggested control by Ca^{2+} could be explained with its homology to calmodulin and annexin proteins, both of which can be found in or on cellular membranes.[108] Its expression was up-regulated in prostate tissue during progressive benign prostatic hyperplasia and prostate cancer, where its activity could be suppressed by a benzimidazole derivative.[109]

VAT-1 was further present in glioblastomes during glioma invasion and played a principal role in immunosuppression of diffuse gliome.[110, 111] To date, the exact function of VAT-1 remains unclear, however, several important findings towards its physiological role have been made. Comparisons were drawn with Zta1, ζ-crystsallin and quinone oxidoreductases. With the latter two it contains 38 conserved residues.[112] This comparison lead Linial *et al.* to assume a non-classical role of VAT-1 in *T. californica*, contrary to their previous interpretation.[113] Similarities with Zta1 and other members of the quinone oxidoreductase family include a large unoccupied space on the bound nicotinamide ring of nicotinamide adenine dinucleotide (NADP), or nicotinamide adenine dinucleotide phosphate (NADPH) respectively.[114] Even though Hakoshima *et al.* showed affinity for 9,10-phenanthrenequinone and, to a lower degree, 1,2-naphthoquinone, its biological substrates are yet unknown. Watanabe *et al.* sovled the crystal structure of human VAT-1 and found a total of four molecules in the asymmetric subunit, in the form of two homodimers.[115] In the crystal, VAT-1 consists of two major α/β domains, connected via a flexible switch segment of 25 amino acids as seen in figure 3.4.[114] Based on these results, they suggested its involvement in the transfer of phosphatidylserine form the endoplasmatic reticulum to mitochondria. However, they noted VAT-1 does not fit into any protein superfamily known so far, thereby contradicting previous publications, where it was introduced as a member of the medium-chain dehydrogenases/reductases family.[112, 115] Furthermore, VAT-1 was identified as a catechin binding protein, due to its interaction with (-)-epigallocatechin-3-gallate with a similar affinity as other binding proteins from neuroglioma cells.[116] Recently, Gleißner *et al.* identified VAT-1 as the potential target of NCA with ABPP. They demonstrated that the presence of VAT-1 during NCA treatment led to a reduction in cancer cell migration. Additionally, several interaction partners involved in the regulation of cell adhesion, integrin activation and lamellipodium organization were identified. Notably, among these was Talin-1, a key protein in cellular migration and direct interaction partner of VAT-1. This finding may serve as the basis for the development of novel anti-cancer drugs.[1]

3.2 Elucidation of the Structure-Activity Relationship: Protein Complex Structure Determination

To analyze drug binding in proteins and decipher their molecular mechanisms, crystallography is an indispensable tool. It allows determination of binding sites for characterization of novel protein-ligand interactions, while on the other other hand sheding light

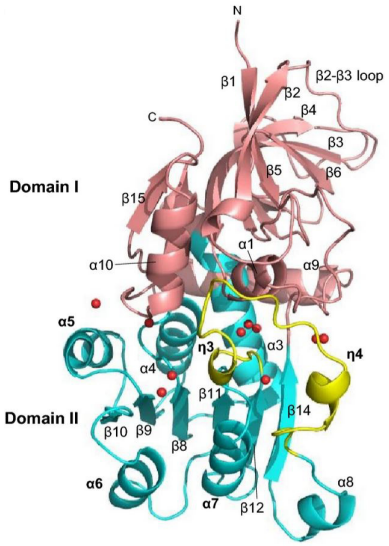


Figure 3.4: Crystal Structure of VAT-1.[114]
Two major α/β domains titled as Domain I and II (purple and blue), connected via a flexible switch segment (yellow). Red spheres indicate water molecules in the nucleotide binding site.

on binding modes and non-covalent interactions of potential drug candidates. To obtain protein-ligand crystals, two methods are applied today, co-crystallization, or soaking.[117] During co-crystallization, a ligand is added to the protein solution and crystals are grown from this mixture. Soaking on the other hand starts with the apo-crystal, to which the dissolved ligand is added. It can then enter the crystal through so called water channels and incorporate itself into the crystal structure. After the co-crystal has formed, it is treated identical to apo-crystals, diffraction data is collected with an x-ray source, phasing is performed, followed by model building. Phasing is necessary, due to the so called phase problem. During data collection only the position and intensity of the diffracted beam are recorded, the phase information is lost and has to be regained later. To this end, several methods are available, all of which provide initial phases for model building. In direct methods, or 'brute force', series of phases are tested, however this is only applicable for small molecules with a resolution of up to 1.2 Å.[118] When a related structure of the protein, or parts of it (like a specific domain) are known, these can be used as a model in molecular replacement to circumvent the phase problem. In molecular replacement an unknown crystal structure is solved by using a related, known molecular model. The known structure is fit to the measured data by trying all possible positions and orientations. Next, the phases of the known structure are taken and together with the measured intensities, a first electron density map is calculated. Since it is rather fast and cheap, com-

pared to experimental, or direct methods, molecular replacement is responsible for about two-thirds of solved structures annually.[119] With the emergence of artificial intelligence, such as AlphaFold or RoseTTAFold, 'known' structures can now be predicted and used for molecular replacement. This circumvents the necessity of needing a related structure, increasing the relevance of molecular replacement even further.[120] In case of an unknown structure single wavelength anomalous dispersion (SAD), or multi wavelength anomalous dispersion (MAD) can be applied. Most commonly, a methionine in the primary structure is replaced by seleno-methionine. During x-ray diffraction, three full datasets are collected, with wavelengths below, at and above the diffraction edge. Phase information can then be obtained from the wavelength-dependent quantitative differences in the anomalous scattering contribution of selenium atoms contained in the crystal.[121] In a similar manner, heavy atoms can be used for phasing. These are introduced through soaking or co-crystallization and phase information can be obtained with the Patterson function.[122] Finally, the model is build in an interactive process, where observed data are fit to a model. Each cycle results in a new set of phases, which can be used for further refinement, until maximum correlation is achieved. R_{work} and R_{free} indicate how well the simulated diffraction pattern matches the measured data, with values of 0 relating to a perfect fit, while a totally random set of atoms would result in an R value of 0.63. R_{work} values introduce bias due to the nature of their calculation, this bias is greatly reduced in R_{free}, where refinement is carried out with 90% of the experimental data. The result will then be compared to the remaining 10% not used in the process, to assess refinement quality.[123] Another set of parameters are the correlation coefficients $CC_{1/2}$ and CC^*, which give an estimation of the usefulness of statistical data, by correlating an observed dataset with the underlying true signal.[124]

3.3 Aim & Objectives

To this day, the role of VAT-1 in cells is not fully understood, yet its involvement in certain types of cancer is apparent. VAT-1 the cellular target of the natural product NCA, which upon binding, leads to a reduced mobility in cancer cells. The aim of this chapter is to provide insights into the recognition of VAT-1 by NCA and its potential dependence on the crucial halogenation of neocarzilins. To this end the structure of the apo protein and in complex with NCA will need to be solved by x-ray crystallography. Additionally, the kinetic nature of their interaction will be characterized through activity-based absorption assays. The required amount of highly pure protein to carry out structional and functional studies will be obtained through overexpression in *E. coli* cells.

3.4 Results & Discussion

For structural as well as activity studies, substantial amounts of highly pure protein are required. To achieve this, we utilized *E. coli* cells for heterologous expression of VAT-1. We then tested several constructs with varying affinity tags and optimized sequence to asses their impact on protein yield and quality.

3.4.1 Protein Expression in *E. coli* Cells

The initial construct for expression of truncated VAT-1 comprises 342 amino acids, a variant lacking the N-terminal 46 residues. It is a fusion protein, consisting of a hexahistidine tag and a cleavable TEV sequence, followed by the protein sequence. The protein DNA sequence was provided by our collaboration partner, Carolin Gleißner (Sieber group, TUM) and cloned into an empty pET28a vector according to chapter 6.2.5. The protein was expressed in Rosetta2(DE3) cells, purified by Ni^{2+}-immobilized metal affinity chromatography (IMAC) and analyzed with PAA gel electrophoresis (PAGE) and western blotting (figure 3.5). Though protein expression levels were satisfactory, most protein could be found in the cell pellet. Possible reasons could be improper folding during protein expression, or, since VAT-1 is a membrane protein, low protein solubility. The former is usually caused by non-ideal conditions like temperature or choice of medium which would need to be addressed.

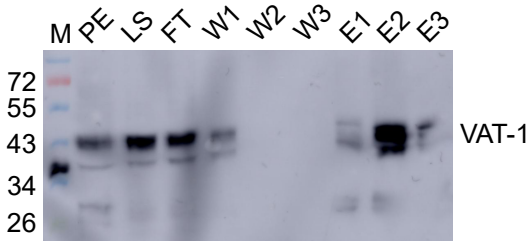


Figure 3.5: Expression of His-TEV-VAT-1.
Western blot showing His-tagged proteins. M: protein standard; PE: pellet after lysis; LS: lysate; FT: unbound protein; W1-3: washing fractions; E1-3: elution fractions.

3.4.2 Codon Optimization

During protein expression DNA triplets are transcribed into tRNA which is then translated into the amino acid sequence. For each amino acid, several tRNA isoacceptors are available, each belonging to their own DNA triplet. The frequency in which these tRNA appear

in cells is called codon usage and differs among organisms and even genes.[125] Codon usage serves as a powerful tool for cells for a multitude of cellular processes like transcription, translation efficiency or co-translational protein folding.[126–128] In the case of VAT-1 non optimized codon usage could lead to errors during protein folding. To overcome this issue and increase the yield of souluble protein, a codon-optimized DNA sequence was tested.

His tag

The codon-optimized, synthetic VAT-1 DNA sequence was obtained from GeneArt (Thermo Fischer Scientific) and cloned into an empty pET28a vector according to chapter 6.2.5, thereby creating a fusion construct comprising an N-terminal hexahistidine tag, followed by a six amino acid long spacer (SSGGRL), a TEV protease cleavage site and the protein sequence. The protein was expressed in BL21(DE3) cells, subsequently purified by Ni^{2+}-IMAC and analysed by PAA gel and western blotting. While most protein stayed insoluble in the pellet, it was possible to purify a small quantity, which was only detectable using western blot and did not show up in SDS-PAGE (figure 3.6). For further studies, we would need better yield by either increasing column binding of the soluble protein, or avoiding aggregation by improving protein solubility. We decided to focus on the latter, since changing extraction and purification conditions was expected to be less laborious compared to re-designing the expression system. Our approach was to improve interaction of the protein with the column matrix through modification of various parameters seen in table 3.1. Yet, regardless of the conditions, no increase in column binding could be seen with VAT-1 still ending up in inclusion bodies. A common observation throughout all attempts for protein purification was only partial column binding, which indicated poor accessibility of the affinity tag for resin binding.

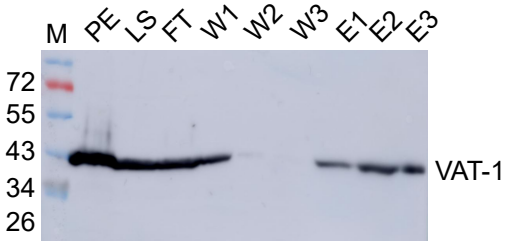


Figure 3.6: Expression of codon optimized His-TEV-VAT-1 in Rosetta2(DE3) cells.
Western blot showing His-tagged proteins. M: protein standard; PE: pellet after lysis; LS: lysate; FT: unbound protein; W1-3: washing fractions; E1-3: elution fractions.

Table 3.1: Parameters varied for improvement of VAT-1 column binding.

E. coli strain	column type	$c_{imidazol}$ (mM)	c_{NaCl} (mM)
Rosetta2(DE3)	His SpinTrap	10	150
BL21(DE3)	Ni-NTA Agarose (gravity flow)	20	500
BL21(DE3)pLysS	HisTrap FF	/	/

His tag Accessibility

A potential reason for the lackluster performance of VAT-1 during purification could be improper protein folding rendering the affinity tag inaccessible. Based on our initial design (figure 3.7 A), three constructs were designed, which should allow better access to the tag, while not impairing protein folding, by introducing a spacer, or moving the tag to the C-terminus. Our first spacer design (figure 3.7 B) comprised 17 glycine, serine, arginine and leucine, and was introduced between the hexahistidine tag and the TEV site, with site directed mutagenesis (SDM). Following expression in 2YT medium, affinity purification did not yield any soluble protein at all, as most of the protein stayed in the pellet (figure 3.8 A). The second design inserted 15 glycine and serine between the TEV site and the

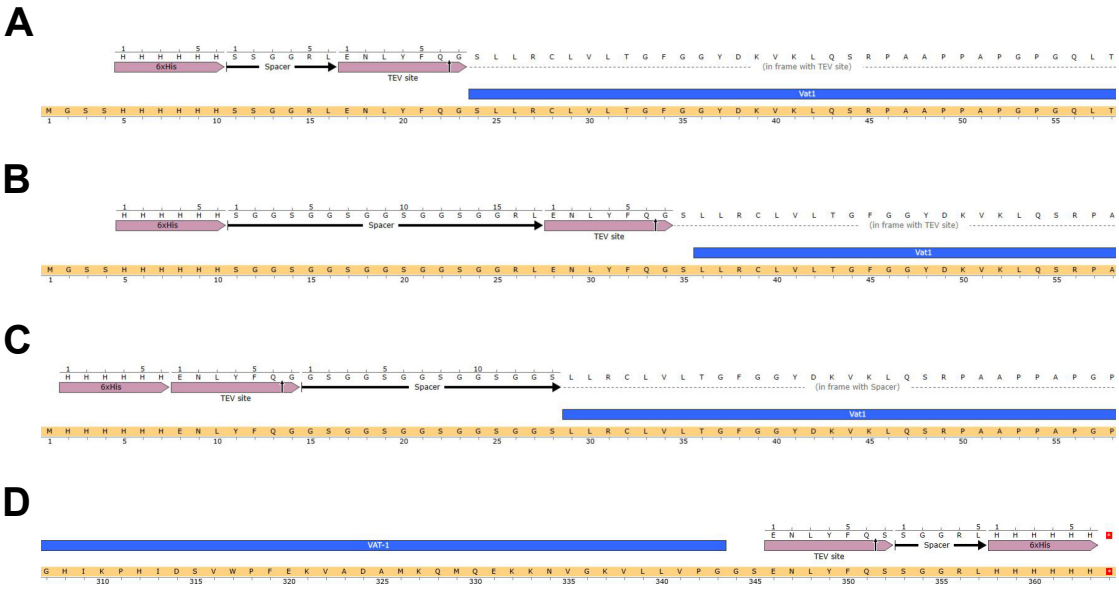


Figure 3.7: Constructs for expression of VAT-1.
A Initial construct consisting of a His tag, a TEV site and the protein sequence. Following designs are based upon this. **B** Spacer introduced between His tag and the TEV site. **C** Spacer introduced between TEV site and the protein sequence. **D** C-terminal TEV site, spacer and His tag.

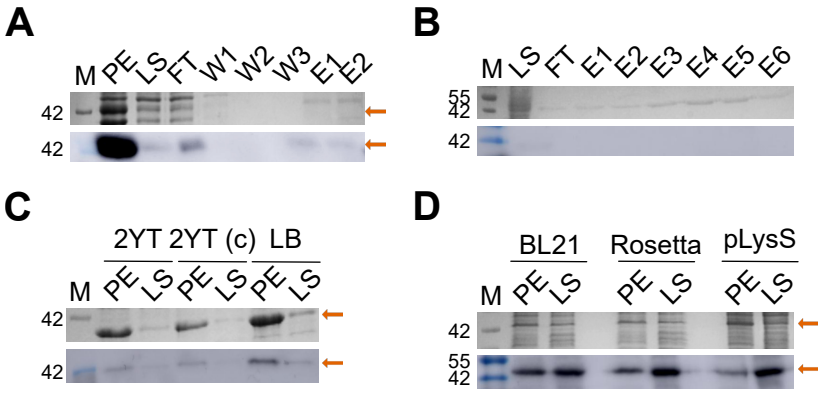


Figure 3.8: Expression of VAT-1 to assess His tag accessibility.
A His-Spcr-TEV-VAT-1 (orange arrow) expressed in Rosetta2(DE3) cells. **B** VAT-1 with a C-terminal His tag. **C** VAT-1 with a C-termianl His tag expressed in 2YT and LB medium. **D** His-Spcr-TEV-VAT-1 expressed by three *E. coli* strains. Top image shows coomassie-stained PAA-gel, bottom one shows corresponding western blot. M: protein standard; PE: pellet after lysis; LS: lysate; FT: unbound protein; W: washing fractions; E: elution fractions.

protein sequence (figure 3.7 C). Increasing the distance between the protein and the tags was could improve tag accessibility, while also allowing the TEV protease to remove a large portion of the tag. This time no protein could be found either in solution, nor in the cell pellet after expression in 2YT medium. Our third design focused on attaching the tag to the C-terminus of the protein, to evaluate whether tag accessibility would be better there without impairing protein folding. The TEV-cleavage site, a spacer comprising five amino acids (SGGRL), and a hexahistidine residue attached to the C-terminus of the protein (chapter 6.2.5). Expression was carried out in 2YT medium, yet again, no protein could be purified (figure 3.8 B) and was found in the pellet (figure 3.8 C. Introduction of spacers made did not improve protein solubility in any way, as for two designs the amount of soluble protein drastically decreased, while no protein could be detected at all for the third design. This led us to assume that low amounts of soluble protein, together with bad affinity tag accessibility could be caused by improper protein folding. Increasing the tag size would then sterically hinder correct protein folding causing misfoldings, which are partially or completely insoluble, while simultaneously limiring access to the tag. Our next approach focused on using molecular chaperones. Molecular chaperones assist in protein folding, either as foldases requiring ATP for actively folding protein, or as holdases, which bind and stabilize folding intermediates, to prevent aggregation. In literature, an increased level of chaperones was reported, after exposing cells to low temperatures as low as 4 °C.[129–133] However, adding a cold shock right before induction did not improve the soluble protein

fraction (figure 3.8 C; PE fraction of 2YT), neither did changing the expression medium to lysogeny broth (LB) Miller (figure 3.8 C). Using *E. coli* strains failed to improve the protein yield as well (figure 3.8 D). Altough our attempts to reduce aggregation were unsuccessful, several important insights were gained. First, codon-optimization of the VAT-1 DNA sequence did not increase the amount of soluble protein in *E. coli*. Second, throughout all expressions the major fraction of VAT-1 was found in the pellet, with varying amounts of soluble protein being present. Changing the expression medium, cold shocking the cells to induce chaperone production and improve protein folding, or protein expression in different *E. coli* strains like Rosetta2, or BL21(DE3)pLysS did not result in a significant increase in soluble protein. Third, purification of the soluble fraction was challenging due to large amounts of soluble VAT-1 not binding to the Ni^{2+} resin. The suspected bad accessibility was not the reason for this, as introducing a spacer between protein and affinity tag or moving the affinity tag from N- to C-terminus did not improve column binding.

3.4.3 Solubility Enhancing tags

Expression of His-tagged VAT-1 resulted in either low levels of soluble protein, or the protein showed only moderate affinity for Ni^{2+} resin. We decided to improve protein solubility, by fusion to MBP and glutathione-S-transferase (GST), which are both established affinity tags. Having a good solubility by themselves, these proteins should act as floaties, preventing VAT-1 from aggregating and keeping it in solution.[5, 134, 135] A construct comprising the MBP tag, a TEV cleavage site and the protein DNA sequence was generated. In case of the GST tag, a human rhinovirus 3C (HRV 3C) site was used for later tag removal, as is common.[136] We decided to use the non-codon optimized sequence, since it yielded more

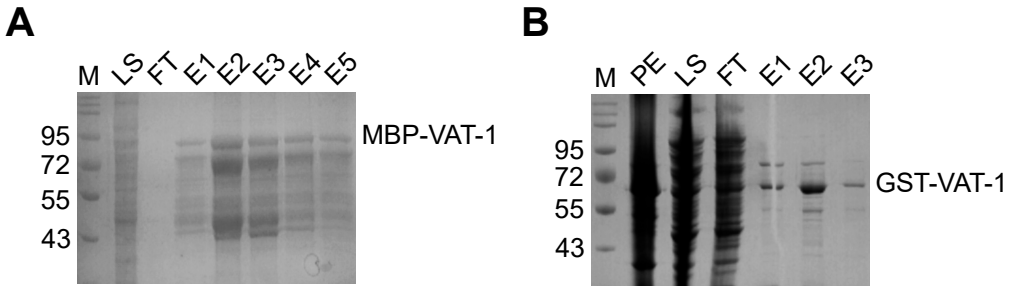


Figure 3.9: Expression of VAT-1 fused to solubility enhancing tags.
Coomassie-stained PAA gels of **A** MBP-VAT-1 and **B** GST-VAT-1 after affinity purification. M: protein standard; PE: pellet after lysis; LS: lysate; FT: unbound protein; E1-5: elution fractions. **B** Coomassie-stained PAA gel of GST-VAT-1 after affinity purification.

soluble protein in *E. coli*. Following expression of both fusion proteins in Rosetta2 cells, we obtained adequate amounts of soluble protein (figure 3.9 A, B).

3.4.4 Heterologous Expression of VAT-1

While usage of GST and MBP tag improved amounts of soluble protein, large quantities of VAT-1 were required for further studies. We then proceeded to optimize protein expression. A variety of factors influence protein expression levels in *E. coli* cells, such as vector choice, culture conditions, codon preference or bacterial strain.[137, 138] Four variables (bacterial strains, expression media, temperature, point of protein expression induction) were chosen, which were deemed to be most impactful. For all four parameters, GST-tagged protein was expressed and purified by affinity chromatography prior to analysis. Expression levels of protein were defined as the amount of protein in the 62 kDa bands in a PAA gel after affinity purification. The amount of protein was obtained by integrating over the intensities of each relevant band. The 55 kDa band of the marker was taken as a reference to compare samples on different gels.

Impact of Expression Strain Used

For the first parameter to optimize, the choice of *E. coli* cells, we tested 6 different strains: BL21(DE3), C41(DE3), NiCo21, Rosetta2(DE3) and T7 Express *lysY/I*^{q}. Expression levels of VAT-1 vary drastically, with the best results being obtained, when using the C43(DE3) strain followed by T7 Express *lysY/I*^{q} (figure 3.10 A). The other four strains used showed between 38% (C41(DE3)) and as little as 9% (Rosetta2(DE3)) of the expression level of the C43(DE3) cells. This was surprising as C43(DE3) is a variant of the C41(DE3) strain optimized for expression of toxic proteins.

Impact of Culture Media

Another factor impacting on protein expression levels, was the choice of nutrition media. Their composition determines POI yield not only, by increasing the cell density but also stabilizing the culture for longer times.[139] This is can be achieved, by adding buffering chemicals as is done in TB medium. Apart from LB (Miller), media with higher nutritional value like 2YT, four yeast extract tryptone (4YT), and TB were used for protein expression, as well as dynamite broth and AI medium. The latter was shown to perform

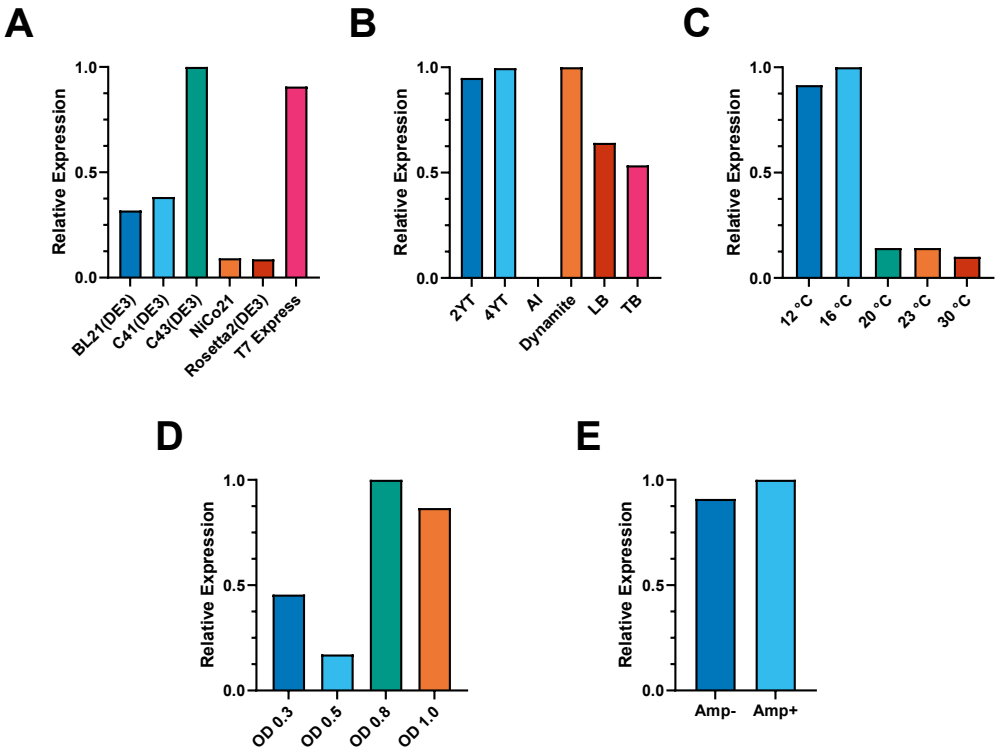


Figure 3.10: Expression levels of GST-tagged VAT-1 under varying conditions.
Several conditions tested to optimize yield of soluble protein after affinity purification. **A** *E. coli* strains: BL21(DE3), C41(DE3), NiCo21, Rosetta2(DE3) and T7 Express *lysY/I*^{q}. **B** Culture media: 2YT, 4YT, AI, dynamite, LB, TB. **C** Expression temperatures: 12 °C, 16 °C, 20 °C, 23 °C, 30 °C. **D** Cell density: OD_{600} = 0.3, 0.5, 0.8, 1.0. **E** Second addition of ampicillin after 4 h.

well for a wide range of proteins including membrane proteins.[140] Several carbon sources which are consumed in subsequent succession, resulting in induction of protein expression at defined cell densities. Additionally, this way of induction is much milder, compared to the sudden addition of Isopropyl-β-D-thiogalactopyranoside (IPTG). Finally, the temperature is usually not reduced during expression in AI medium, which could turn out to be a problem since lower temperatures usually improve correct folding of proteins.[141, 142] As a sixth option, we tested dynamite broth, developed by Taylor *et al.*, who reported superior performance compared to other, common media.[143] Following protein expression and purification, we observed similarly high yields for dynamite medium, 2YT, and 4YT with the former one slightly outperforming the other two (figure 3.10). Way less protein was obtained when using LB and TB, both with about 50% yield. No protein at all was obtained from AI medium.

Impact of Expression Temperature

One of the many reasons why protein aggregates form during expression are incorrectly folded amino acid sequences. It was shown that lower cultivation temperatures reduce the amount of aggregates, due to misfolding, thereby improving protein solubility.[144–148] Based thereupon, expression temperatures were tested with a focus on the below-room-temperature range, as seen in figure 3.10. VAT-1 was expressed at five temperatures ranging from 12 °C to 30 °C. Except for the highest temperature, all expressions were carried out over night. At 30 °C protein expression was cut down to 4 h, since prolonged exposure to higher temperatures would lead to aggregation, in spite of the presence of a solubility enhancing GST tag. By far the highest amount of soluble VAT-1 was obtained at 16 °C, direcly followed by 12 °C. Increasing the expression temperature to 20 °C and above led to a significant drop in soluble protein, with yields of below 15% compared to expressions at 16 °C.

Impact of Cell Density

Bacterial growth is not a linear process until all nutrients in a culture are depleted. It can be split in four different phases starting right after inoculation. In the initial lag phase cells sense their surroundings and prepare the organism for the following exponential phase.[149] Herein, the cells are exposed to optimal growth conditions in regards to temperature, nutrients and other cells. However, with increasing cell numbers the culture enters the stationary phase, which is defined by similar amounts of cells dying as are duplicating, due to reducing levels of nutrients. Finally, in the death phase, the majority of bacteria die, as waste products accumulate and nutrients are depleted. Protein expression is usually induced in the exponential phase, however, the optimal cell concentration might vary depending on the cells, plasmids used, and also environmental conditions. To find the right time point for induction of expression with IPTG, four cell densities (0.3, 0.5, 0.8, 1.0) were chosen (figure 3.10 D). Optical densities (ODs) were measured by absorption at 600 nm compared to empty medium. Except for an outlier at an OD_{600} increasing cell densities lead to more protein being expressed until a maximum at an OD_{600} of 0.8. Higher cell densities upon induction lead to decreased protein yields. *E. coli* probably entered the plateau phase shortly after, where suboptimal conditions prevail.

Impact of Repeated Addition of Ampicillin

To avoid contaminations and acting as a selection marker, antibiotics are added to bacterial cultures. The respective resistance genes are usually found on the plasmid containing the genetic information of the desired proteins. Examples for antibiotics used in protein expression are kanamycin, ampicillin, or chloramphenicol, which all affect different metabolical pathways. Apmicillin for example affects *E. coli* cells by binding to penicillin-binding proteins (PBPs), which affects their cell cycle. PBPs are membrane bound receptors, which play an important role in formation of the peptidoglycan structure of the cell wall. Their inhibition by ampicillin hinders peptidoglycan synthesis, reducing cell wall integrity, which results in lysis and finally death of the cells.[150, 151] Resistant cells produce β-lactamase, encoded for on the plasmid, which hydrolyzes ampicillin thereby rendering it harmless for the cell.[152] However, excessive β-lactamase is secreted into the culture medium where it retains its function and continues degrading the antibiotic. This reduces concentration of the selection marker and makes cultures prone to contamination. To counteract this reduction of selection for plasmid presence, additional ampicillin was added to the culture after 4 h of protein expression. Expression levels increased by about 10% compared to the control culture without additional ampicillin (figure 3.10 E).

3.4.5 Purification of VAT-1

Following the optimized expression protocol, pure protein could be obtained after a two step purification. The first step involved a hybrid affinity purification, coupled with proteolytic cleavage of resin bound protein (figure 3.11 A). The HRV 3C protease used for removal of the affinity tag was designed to cleave protein bound to affinity resin. After cleavage the protease would bind to the resin, due to the presence of an inherent GST tag and leave only VAT-1 in solution, which could then be washed out without usage of a potentially harmful eluent (figure 3.11 B). Subsequent cation exchange purification (CIEX) step using cation exchange resin resulted in the pure protein (figure 3.11 C). This step was preferred over SEC, due to similar sizes of VAT-1 (38 kDa) and GST-HRV 3C (42 kDa) making separation by size challenging.

Even though initial results were promising the purification of VAT-1 remained challenging, due to low reproducibility. A variety of factors influenced both purification steps and lead to plenty of changes throughout the years. Therefore, the following sections will focus on the search for these factors and how to consistently get pure VAT-1.

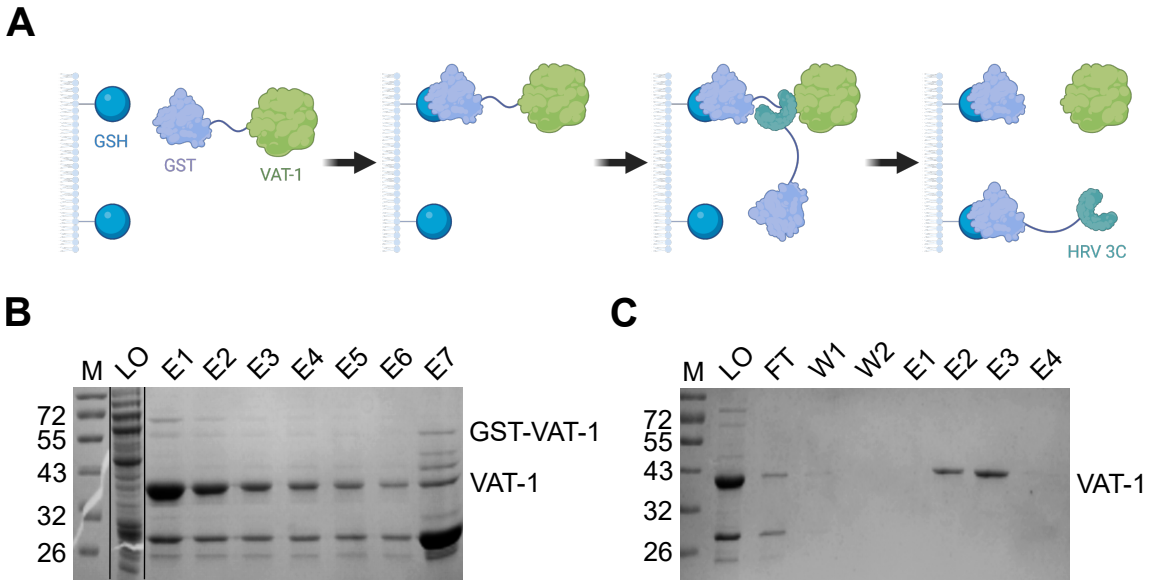


Figure 3.11: Purification of VAT-1 by on-column cleavage and CIEX.
A On-column cleavage of GST-tagged VAT-1 with HRV 3C protease. After binding the protein to the resin and washing out unbound sample the protease is injected onto the column. Following proteolytic removal of the GST tag from VAT-1 the protease then binds to the column itself. Finally, the pure protein can be washed out. **B** PAA gel of on-column cleavage of VAT-1. M: protein standard; LO: column load; E1-7: elution fractions. **C** Second step in VAT-1 purification by CIEX visualized on a PAA gel. M: protein standard; LO: column load; FT: unbound protein; W: washing fractions; E1-4: elution fractions.

Affinity Purification

The above mentioned on-column cleavage as an alternative to affinity purification, was initially carried out using the ÄKTA system (figure 3.12 A). However, protease cleavage efficiency on the column was not consistent over the course of several purifications, but declined without any apparent change in protocol (figure 3.12 B). Neither thorough cleaning of the resin, a new column, nor freshly purified protease could restore its performance. Changing from the prepacked columns to modified sepharose beads suited for benchtop purification turned out to be only a temporary solution (figure 3.12 C), as on-column cleavage performance went down as well after some time (figure 3.12 D). A viable alternative was found in the form of classic affinity purification (figure 3.12 E), followed by subsequent separate proteolytic cleavage of the eluted protein in solution (figure 3.12 F).

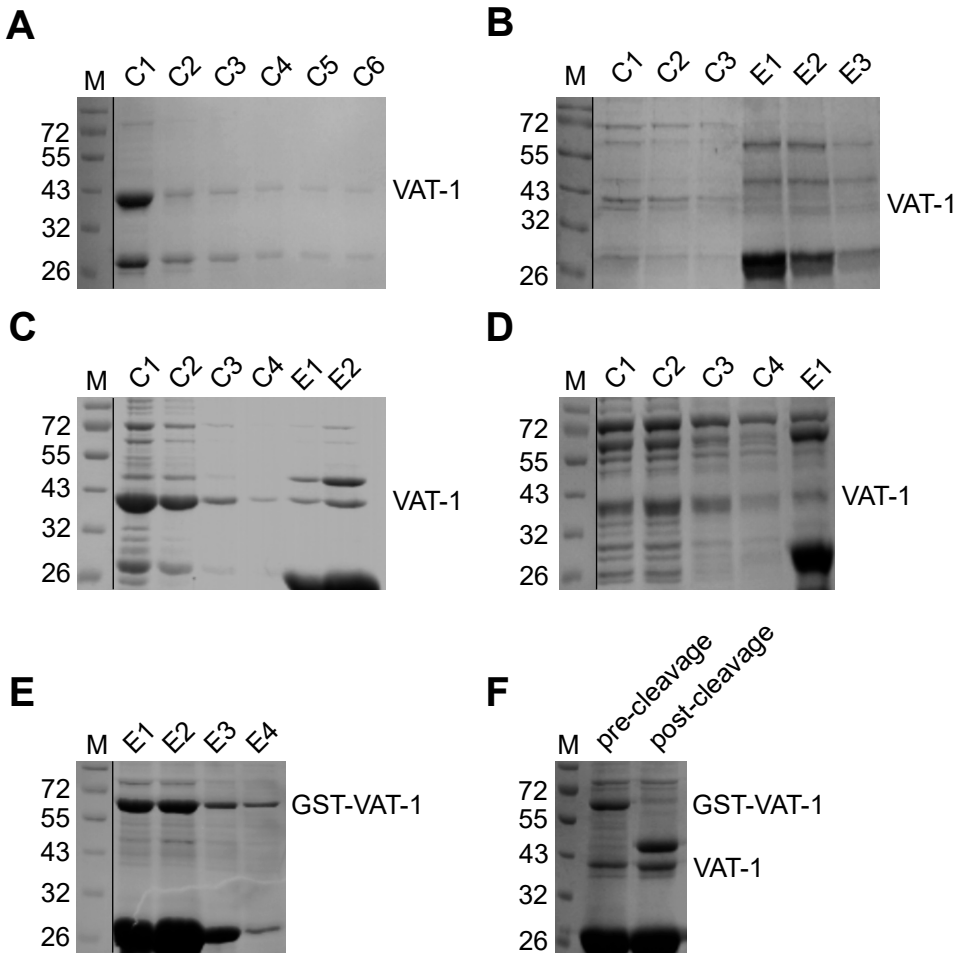


Figure 3.12: Various affinity purification approaches due to unreliable results.
Coomassie-stained PAA gels of: successful (**A**) and partially successful (**B**) on-column cleavage purification of VAT-1 using the ÄKTA system. Successful (**C**) and partially successful (**D**) on-column cleavage purification using Sepharose™ 4B resin. Classic affinity purification (**E**) using Sepharose™ 4B resin and subsequent proteolytic cleavage (**F**). M: protein standard; LS: lysate; C: on-column cleaved fractions; E: elution fractions.

Ion Exchange Purification

Ion exchange chromatography using a Resource S column (Cytiva) proved highly effective in removing the protease, the GST tag and residual impurities following affinity purification and proteolytic cleavage. (figure 3.13 A). Unfortunately, its reliability in binding and separation efficiency was inconsistent (figure 3.13 B). While the principle behind ion exchange chromatography, separation by charge, is a rather straightforward concept, its effectiveness can be influenced by several parameters that need to be addressed for optimal results. The ionic interaction between protein and matrix is mainly influenced by the ionic

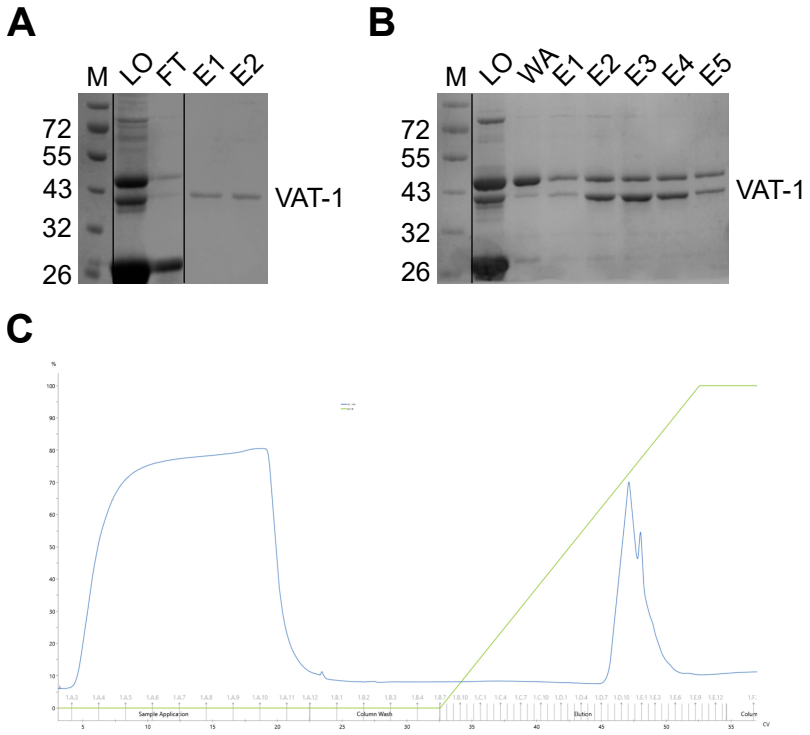


Figure 3.13: CIEX purification of VAT-1 after proteolytic cleavage.
Coomassie-stained PAA gel after ion exchange purification of VAT-1 under identical conditions with **A** good and **B** poor separation of impurities. M: protein marker; LO: column load; FT: unbound proteins; WA: washing fraction; E1-5: elution fractions. **C** chromatogram for ion exchange purification of VAT-1. Protein elutes at 500 mM NaCl.

strength of the buffer and the choice of pH. The former is defined by salt concentration in the buffer which competes with the protein in binding the charged surface of the resin. Low to moderate salt concentrations allow protein binding to the column and, if selected carefully, prevent weakly coordinating impurities to bind the column. Additionally, under high salt conditions the protein is eluted from the column. To ensure proper binding of VAT-1 to the column salt concentrations prior to loading were reduced to 10-20 mM NaCl by dilution or desalting column against a low salt buffer. Furthermore, it was shown that for VAT-1 specifically, column binding was possible at salt concentrations as low as 20 mM NaCl and elution did not start until 500 mM NaCl was reached (figure 3.13 C). Apart from salt concentration pH is a major player in IEX by deciding over the proteins charge, a parameter which is defined by the buffer pH and protein pI, a value specific for each protein. Amino acids with positively and negatively charged side chains contribute to the overall charge of a protein. Their charge in turn depends on the respective pKa value of each group, mainly -NH_2, -COOH, -SH and -OH. Calculating the pI according to the

Table 3.2: Varying pI depending on the pI calculator.

pI calculator	pI
protpi.ch	8.14
novoprolabs.com	8.14
protocalc.sourceforge.net	8.38
bioinformatics.org	8.57
expasy.org	8.57
geneinfinity.org	8.57

Henderson-Hasselbach equation is rather easy, assuming its approximations are valid under the present conditions.[153] However, challenges arise when determining which pKa values apply for ionizable groups, as these vary, depending on charge-charge interactions, charge-dipole interactions and the Born effect.[154] Taken together, pKa values vary depending on the underlying database for calculation tools.[153] In the case of cleaved VAT-1 varied as strong as 8.14-8.57 table 3.2). This range of ΔpI = 0.43 meant a pH of 7.0 or lower was necessary to ensure positively charged VAT-1 was present. Apart from the ionic strength of the buffer several other parameters, were addressed to ensure optimal conditions for IEX purification, such as choice of column resin, or addition of detergents. The matrix of resource S columns is modified with sulfonate groups, which classifies as a strong cation exchange resin. While this allows separation of weakly positively charged proteins, issues will arise when using strongly positively charged proteins, as elution might not be possible. Therefore, weak cation exchange resins, or anion exchange columns might pose an alternative solution. First, we purified VAT-1 using a CaptoQ column, which contains a strong anion exchange resin in the form of quatenary amines. Unfortunately we did not obtain any significant amount of protein. Then we switched to a weak cation binding resin in the form of a CMFF16/10 column, which is loaded with carboxyethyl groups, however, no protein binding could be observed this time. In conclusion, the weakly positively charged VAT-1 protein requires a strong cation exchange resin not only for column binding but also for achieving protein separation. Another parameter which was reported to improve efficiency of ion exchange purifications is the addition of detergents.[155, 156] To cover a wide variety of conditions, a nonionic detergent, Tween 20, and an anionic detergent, N-lauroylsarcosine, were used to test this hypothesis. A cationic detergent was not tested, as it would irreversibly bind to the cation exchange column resin. N-lauroylsarcosine was already proven to be suitable for VAT-1, as we used it previously in the lysis buffers, thereby increasing

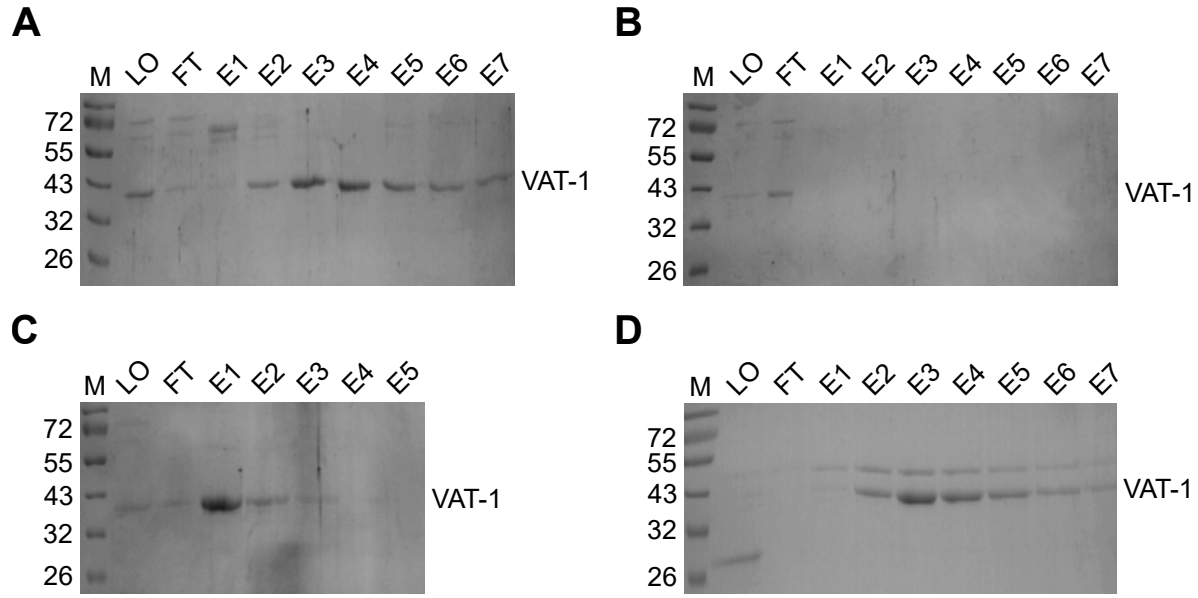


Figure 3.14: CIEX purification of VAT-1 under various conditions.
CIEX purification at pH 6.0 (**A**), of 0.2% Tween-20 (**B**), of 0.1% NLS (**C**) and at pH 7.0 (**D**). M: protein marker; LO: column load; FT: unbound proteins; E1-7: elution fractions.

the fraction of soluble VAT-1. Surprisingly, it performed rather poorly during purification, with all VAT-1 being found in the fraction for unbound proteins (figure 3.14 B). Tween 20 on the other hand, yielded almost as much protein (figure 3.14 C) as the purification without any additives (figure 3.14 A). In an attempt to reduce protein exposure to acidic conditions and potentially prevent protein degradatoin, the pH was increased from 6.0 to 7.0. However, even though VAT-1 did bind to the column, separation of protease was not possible anymore under these conditions (figure 3.14 D). This further highlights the careful selection of buffer pH for successful purification. Size exclusion chromatography was considered as an alternative to CIEX for separation of VAT-1 from HRV 3C. However similar sizes of VAT-1 (38 kDa) and HRV 3C (45 kDa) made separation difficult even with high resolution columns like superdex™ 75 10/300 GL (figure 3.15 A, B).

Purification Summary

Purification of VAT-1 turned out to be more challenging than expected. We improved the expression construct towards generation of highly soluble protein, by integration of the solubility enhancing GST-tag. Optimal expression conditions for VAT-1 were observed, when using C43(DE3) cells in Dynamite medium and inducing protein expression at 16 °C at an OD^{600} of 0.8. Protein yield could be slightly further increased upon a second addition

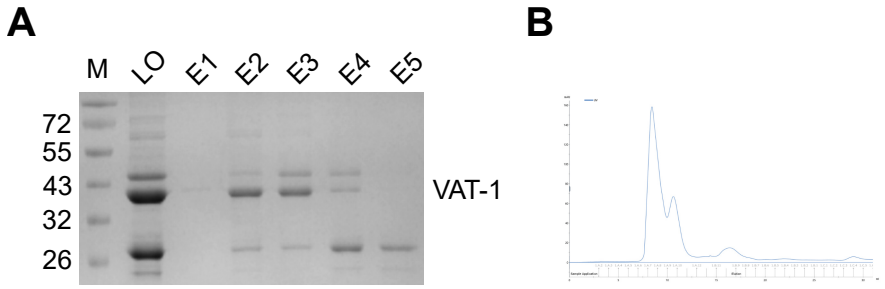


Figure 3.15: Purification result of VAT-1 by SEC.
A Coomassie-stained PAA gel of eluted protein fractions. M: protein marker; LO: column load; E1-5: eluted fractions. **B** Chromatogram of a Superdex™ 75 10/300 GL monitoring protein elution by UV absorbance at 280 nm.

of ampicillin, 4 h after induction. VAT-1 was purified in a two-step procedure, first, the majority of impurities was removed by on-column cleavage, which simultaneously cleaved-off the affinity tag. Second, we applied CIEX for tag removal and were able to obtain highly pure protein. Aggregation was a major hurdle, but this could partially be overcome with by incorporating solubility enhancing tags. Furthermore, it was essential to carry out all steps subsequently, as removal of the GST-tag significantly increased VAT-1 aggregation, making it impossible to freeze aliquots. Despite unreliable reproducibility, enough highly pure protein was obtained for crystallization.

3.4.6 Crystallization

To find optimal crystallization conditions for VAT-1, we set up high-throughput crystallization trials as described in chapter 6.7.1. Kim *et al.* reported protein crystals at 100 mM BisTris (pH 6.0), 200 mM $NaNO_3$ and 22% (v/v) PEG 3350.[114] Unfortunately, no crystal growth was observed under these conditions in our lab. We obtained VAT-1 crystals at a protein concentration of 3.7 mg/mL against a reservoir of 1.26 M $(NH_4)_2SO_4$, 100 mM TRIS pH 8.5 and 200 mM Li_2SO_4 (figure 3.16, A). Crystal growth was monitored

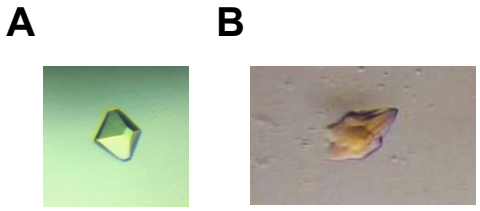


Figure 3.16: VAT-1 crystals.
VAT-1 apocrystal (**A**) and cocrystal with NCA (**B**) in brightfield.

by brightfield microscopy. A native dataset with a resolution of 2.5 Å was recorded at the Deutsches Elektronen Synchrotron (DESY) in Hamburg. The structure was determined through molecular replacement using MOLREP, employing a previously reported structure[115] (pdb code 6K9Y) as a template. Subsequent model building was performed using Coot. Refinement of the structure was carried out with Refmac5, part of the CCP4 software suite, achieving an R/R_{free} of 29.6% /31.5% (see figure 7, appendix). Due to disordering several loops on the protein surface could not be built. Structure solution statistics are listed in table 7 in the appendix.

Apo Structure

According to our crystallographic data, the asymmetric unit contained two VAT-1 molecules in the form of a homodimer, corresponding to the biomolecule (figure 3.17 A). Each monomer exhibits α/β topology, consisting of 8 α-helices and 15 β-strands. They are characterized by a central cluster of α-helices, flanked by two β-sheet, one in parallel and the other in antiparallel orientation. The antiparallel β-sheet is connected via flexible loops in a right handed β-α-β fold. The other, parallel β-sheet is connected vial α-helices in a typical Rossman fold. The similarity of both subunits, with an RMS of 0.400, can be seen in the superposition shown in figure 3.17 B and table 3.3, which highlights their very minor variations. Specifically chain A exhibits 35% helical structures (α-, or 3,10-helix), 26% β-strands and 39% coils and turns.[157, 158] Chain B on the other hand, possesses slightly fewer helices (31%) and β-strands (22%), with coils and turns (47%) being more prominent. Compared to previously solved structures, chain A retains a similar ratio of folds with only slight differences, resulting in an RMSD of 0.400. Fold-ratios for chain B vary strongly depending on the structure, except for the NAPD bound protein (pdb code: 6LHR) where both chains are almost identical (figure 3.17 C), possessing a RMSD of 0.594. Two of these structures, 6K9Y and 4A27 were assigned to the orthorhombic spacegroups P 2_1 2_1 2_1 and P 1 2_1 1, while 6LHR was found to be monoclinic space group C 1 2 1, making our solved structure the first VAT-1 in a tetragonal space group I4. To identify areas involved in catalytic activity of VAT-1, we performed a homologue search against the InterPro database. We found a GroES-like domain along residues 115-306 (figure 3.17 D, orange), containing an alcohol dehydrogenase (ADH)-like domain (residues 29-112, figure 3.17 E, orange). Residues 115-306, which are conserved among the NAD(P)-binding domain superfamily clearly adopt a Rossmann fold in the form of β-sheets connected through α-helices, often associated with nucleotide binding domains (figure 3.17 E, magenta). A conserved quinone oxidoreductase site also found ζ-crystallin can be seen in figure 3.17

D (magenta, residues 143-164). This supports reported findings of a NADP-dependent oxidoreductase functionality of VAT-1.

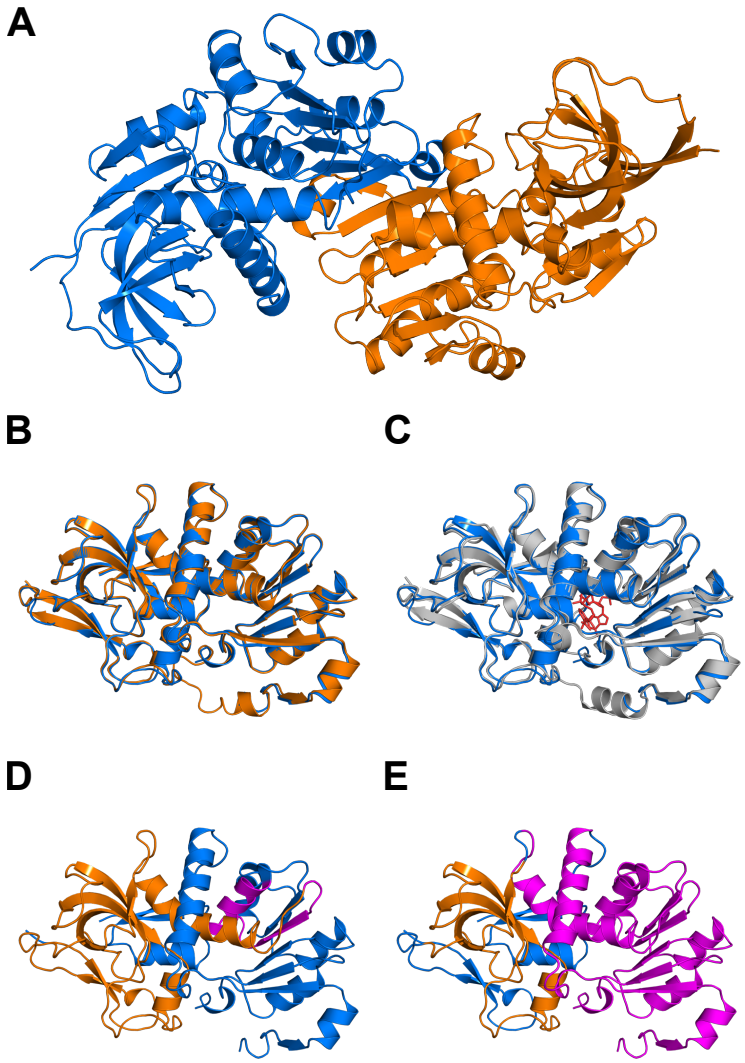


Figure 3.17: Crystal structure of VAT-1.
A Crystal structure of homodimeric VAT-1 with subunits colored in blue and orange. **B** Superposition of both VAT-1 subunits. **C** Superposition of Monomeric VAT-1 subunit (blue) with published structure (pdb code: 6LHR, grey). **D** VAT-1 (blue) with conserved sequences among GroES-like domain (orange) and quinone oxidoreductase site (magenta). **E** VAT-1 (blue) with ADH-like domain (orange) and the sequence conserved among NAD(P)-binding domain superfamily (magenta).

Table 3.3: Ratio of folds among different VAT-1 crystal structures.
Values were calculated according to Frishman *et al.*.[157, 158]

	Chain A				Chain B			
	VAT-1	6K9Y	6LHR	4A27	VAT-1	6K9Y	6LHR	4A27
helical	0.35	0.35	0.36	0.36	0.31	0.21	0.35	0.31
β-strands	0.26	0.27	0.27	0.28	0.22	0.29	0.28	0.31
coils/turns	0.39	0.38	0.36	0.36	0.47	0.38	0.37	0.37

Cocrystallization

Cocrystallizations of VAT-1 were set up with NCA (**7**), JW-51 (**11**), JW-21 (**12**), JW-73 (**13**) and WHY-309 (**14**) (figure 3.18) respectively at various protein concentrations (1.7, 2.0, 2.5, 3.0 mg/mL). At higher concentrations rapid precipitation of VAT-1 became imminent and prevented formation of crystal big enough for x-ray diffraction. Crystals were obtained with a 1:3 molar ratio of NCA at 2.0 mg/mL VAT-1, 1.6 M $(NH_4)_2SO_4$, 100 mM HEPES pH 7.5, 100 mM NaCl and 1 mM TEW (figure 3.16 B). While it was possible to collect a dataset to about 3.4 Å at the DESY in Hamburg, resolution was insufficient to properly analyse protein-inhibitor interaction. Seeding and improved screening conditions resulted in a multitude of crystals, none of which had a resolution better than 4.0 Å. A sharp drop at 4.0 Å beyond which no diffraction spots were present at all indicates an issue with lattice spacing in the crystals. To improve crystal quality, the cocrystals were dehydrated with sat. NaCl, 4 M and sat. $(NH_4)_2SO_4$ as described in chapter 6.7.1. Reduction of relative humidity in crystals by dehydration was shown to close solvent channels and trigger changes in the whole crystal structure, thereby improving the diffraction limit.[159]

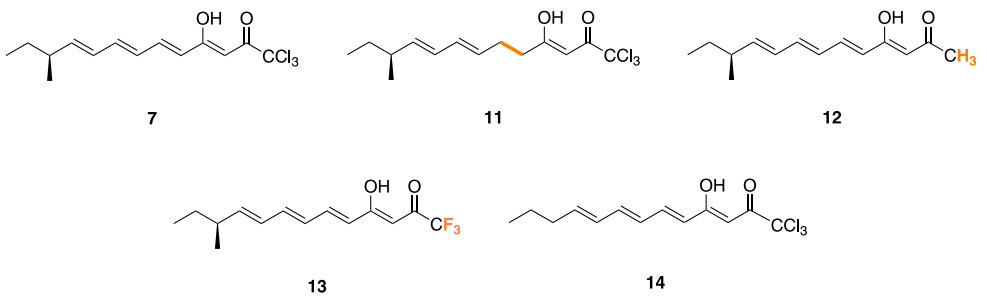


Figure 3.18: NCA and its derivatives JW-51 (**11**), JW-21 (**12**), JW-73 (**13**) and WHY-309 (**14**) used for cocrystallization of VAT-1.

Dehydration magnitude strongly depends on the relative humidity of the crystal and in some cases diffraction could not be improved.[160] While VAT-1 cocrystals were stable under dehydrating conditions for up to 20 h, dehydration did not lead to better diffraction.

3.4.7 Inhibition of VAT-1 Activity by Neocarzilin A and its Derivatives

Data published by Levius *et. al* indicated the presence of a domain conserved among oxidoreductases based on sequence similarities.[113] To assess the catalytic activity of VAT-1 an *in vitro* oxidoreductase assay was carried out. This assay monitored the conversion of NADPH (**15**) to NAPD$^+$ (**17**) in the presence of 9,10-phenanthrenequinone (**16**) over the course of 2 h, as shown in scheme 3.19. Our experiments clearly show conversion of NADPH (figure 3.21, blue curve). While the uncatalyzed reaction resulted in a 0.12 fold decrease in NADPH levels, under VAT-1 catalysis a 0.43 fold decrease was observed. This not only showed VAT-1 does indeed exhibit oxidoreductase functionality, it also allowed us, in cooperation with Josef Braun (Sieber group, TUM), to screen for potential inhibitors. Gleißner *et al.* already confirmed NCA to reduce tumor cell mobility, by interacting with VAT-1.[1] NCA, as well as three of its derivatives (figure 3.20 C) were tested, which would shed light on this interaction and allow identification of participating functional groups. As seen in figure 3.20 A, all four compounds showed a statistically significant inhibition of VAT-1s' catalytic activity (p values <0.0001). Higher concentrations naturally resulted in a stronger inhibition, with JW-73 (**13**) having the strongest effect of all tested compounds.

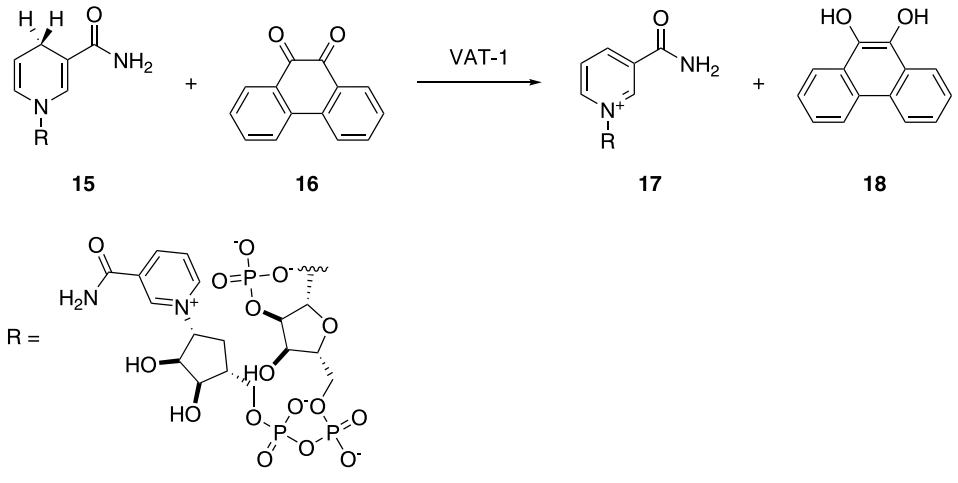


Figure 3.19: Redox reaction catalyzed by VAT-1.
Reaction used to determine protein activity based on oxidoreductase functionality of VAT-1.

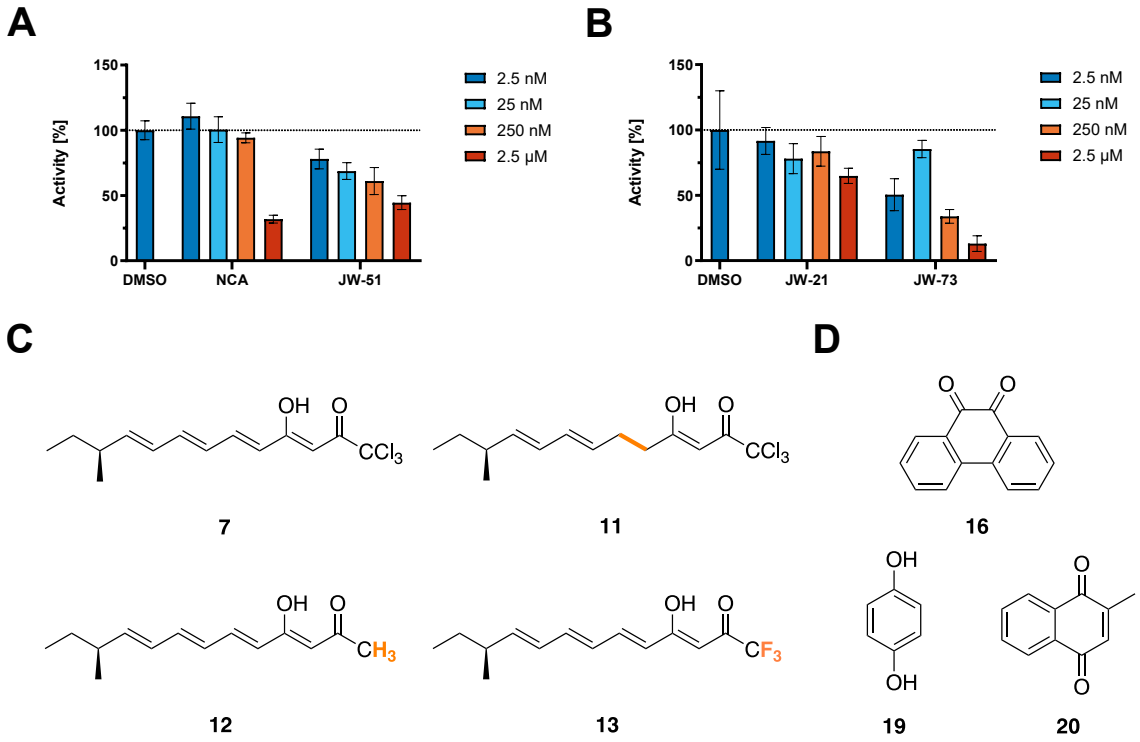


Figure 3.20: Redox reaction catalyzed by VAT-1.
A Inhibitor NCA (**7**) as well as its derivatives JW-51 (**11**), JW-21 (**12**) and JW-73 (**13**). Inhibition of VAT-1-catalyzed reaction by NCA, JW-51 (**B**), JW-21 and JW-73 (**C**). DMSO was used as negative control.

The former contains a trifluoromethyl group, boasting a stronger inductive effect compared to the trichloromethyl or methyl groups. JW-51 (**11**), whose conjugated π-system is much shorter than for the other compounds (6 vs 10 electrons), comes second in inhibition of VAT-1.Catalytic activity of VAT-1 varied from batch to batch. Factors such as salt and buffer concentration were taken into consideration, showing VAT-1 was less active in 20 mM NaCl (figure 3.21 brown curve), compared to 100 mM NaCl (figure 3.21 blue curve). Performing a SEC instead of CIEX in the final step of protein purification was more time consuming and greatly reduced the yield, due to protein aggregation, while also giving less active VAT-1 (figure 3.21, green curve). Moreover, alternative substitutes, such as hydroquinone (**19**, purple curve, figure 3.21) and menadione (**20**, red curve, figure 3.21) were also tested. However, neither of these were oxidized by VAT-1. The possibility of a batch effect cannot be excluded.

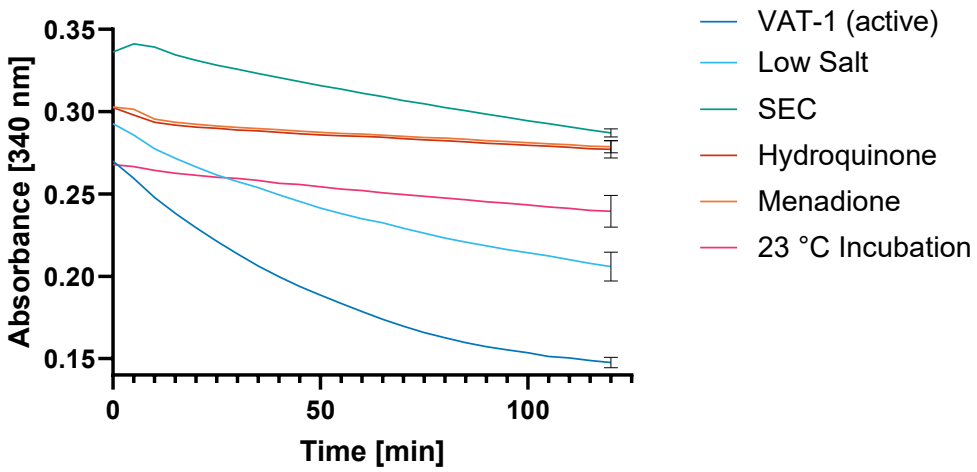


Figure 3.21: VAT-1 activity characterized by oxidoreductase assay.
A 9,10-Phenanthrenequinone (**16**) and alternative oxidizing agents hydroquinone (**19**) and menadione (**20**) used for oxidoreductase reaction. **B** NADPH concentration will decrease over the course of 2 h if VAT-1 is active due to conversion into $NADP^+$. Normalized absorbance of NADPH was plotted against time VAT-1 purified by CIEX (blue), SEC (green), assay at 20 mM NaCl (brown), as well as with menadione (red) and hydroquinone (purple) and after incubation with inhibitor at 23 °C (pink). Error bars are displayed in black.

3.5 Summary

The aim of this chapter was the characterization of VAT-1 and its recognition of NCA, an interaction which leads to a reduction in cancer cell mobility. To conduct structural and kinetic studies, large quantities of highly pure protein were required. At first we optimized expression parameters, such as expression vector, *E. coli* strain, growth medium and expression conditions, followed by the protein purification protocol. With this, we were able to obtain active and homogeneous protein, suitable for functional and structural studies. However, enzyme activity, quality, as well as quantity varied from batch to batch.

Together with our collaboration partner, J. Braun (Sieber group, TUM), we demonstrated specific inhibition of VAT-1's oxidoreductase activity by NCA and several derivatives. In addition, we were able to obtain crystals of VAT-1 under different crystallization conditions and in another spacegroup as previously reported. While we were able to crystallize VAT-1 together with NCA and derivatives, unfortunately none of the crystals reached a diffraction quality required for structural studies.

Conclusion & Outlook

The first part of this thesis aimed to investigate the putative halogenase ScORF3 involved in the final step of neocarzilin A biosynthesis. The protein was found alongside ORF8, a putative flavin halogenase, indicating an FAD dependant activity of ScORF3, which is supported by our findings of an FAD-binding domain. Our attempts to purify the protein were largely dictated by its poor solubility. Nevertheless, addition of FAD did not lead to formation of stable protein complexes in solution. Following protein purification from *E. coli*, we managed to obtain crystals, but none of sufficient quality for structure determination. Further research could focus on improving protein solubility, either by finding conditions better suited for purification and crystallization, or introducing solubility enhancing tags which do not hinder protein crystallization.
ScORF3 also contains a sequence conserved among Trp halogenases, therefore we compared a structure predicted by AlphaFold to structural and sequential homologues. We found CtcP, a halogenase responsible for chlorination of tetracycline, to be the best match for ScORF3 with a moderate sequence similarity and good RMSD. This reinforces ScORF3s role as a putative halogenase involved in chlorination. The two trp halogenases RebH and PrnA, on the other hand, which were previously compared to ScORF3, barely compared, due to low sequence similarities and poor RMSD of their folded structures. Their similar function might not be due to structural conformities. Further investigations could focus on substrate binding of ScORF3. To elucidate the chlorination mechanism, it might be necessary to obtain crystal structures of ScORF3 in complex with dechloroneocarzilin.

In the second part of this thesis, the interaction of NCA and VAT-1 should be characterized. NCA was previously shown to inhibit mobility of cancer cells, with VAT-1 being its cellular target. The nature of their interaction was investigated using activity based oxidoreductase assays, as well as x-ray crystallography. After purification and crystallization, the x-ray crystal structure of the apo-protein could be solved.
We observed dimeric VAT-1 in the previously unreported spacegroup I 4. The protein

clearly adopts a Rossman fold required for nucleotide binding as part of its NADP-binding site. On the other hand, the presence of an ADH-like domain supports the hypothesis of an NADP-driven oxidoreductase activity of the protein. We verified the catalytic activity of VAT-1, by its ability catalyze the reduction of 9,10-phenanthrenequinone, providing a photometric readout. Furthermore, addition of NCA to VAT-1 in the *in vitro* activity assay inhibited the catalytic conversion of 9,10-phenanthrenequinone by VAT-1, providing evidence of an interaction between VAT+1 and NCA. Several derivatives of NCA were also capable of interacting with and inhibiting VAT-1, with two of them, JW-73 and JW-51, possessing an even stronger effect as NCA itself. Despite extensive trials no diffraction quality complex crystals of VAT-1 and NCA could be obtained. Thus, the molecular basis of the inhibition of VAT-1 by NCA could not be determined. As a next step, it might thus be necessary to improve crystal diffraction through soaking of crystals, or finding better crystallization conditions.

Materials

5.1 Instruments

Table 5.1: Instruments and consumables in this thesis.

Purpose	**Product**	**Manufacturer**
Centrifuges	5804 R	Eppendorf
	MiniSpin	Eppendorf
	RC 6 Plus Superspeed Centrifuge	Sorvall
	Heraeus™ Multifuge™ X3R	Thermo Scientific
	Galaxy Mini	VWR
	Micro Star 12, 17, 17R	VWR
	Mini Star blueline, silverline	VWR
Rotors	SLA-3000 Super-Lite®	Sorvall
	SS-34	Sorvall
	Fiberlite™ F10-4x1000 LEX, F14-6x250 LE, F15-8x50cy	Thermo Scientific
	TX-1000	Thermo Scientific
Balances	IoT-Line Tabletop Balance FKB 6K0.02	Kern & Sohn
	BCE623I-1S Entris® II Prescision Balance	Sartorius
	BP 2100	Sartorius

Instruments and consumables in this thesis, continued.

Purpose	Product	Manufacturer
Gel electrophoresis and analysis	PowerPac Basic	Bio-Rad
	peqPOWER	PEQLAB
	EPS 600	Pharmacia Biotech
	PerfectBlue™ Power Supply Universal	VWR
	UVT-22 ME Transilluminator	Herolab
	Desatronic 500/500	DESAGA
	iBlot™ 2	Invitrogen
	Amersham ImageQuant™ 800	Cytiva
Incubators	MaxQ™ 6000	Thermo Scientific
	INCU-Line®	VWR
	MIR-553 Refrigerated Incubator	Sanyo
	SM-30 Universal Shaker Control	Edmund Bühler GmbH
	TH 30 Shaker Top	Edmund Bühler GmbH
FPLC	ÄKTA go™, pure™	Cytiva
	Superdex™ 200 Increase 10/300 GL	Cytiva
	HiLoad™ 16/60 Superdex™ 200 pg	Cytiva
	HiLoad 16/600 Superdex™ 75 pg	Cytiva
	Resource™ S	Cytiva
	HisTrap™ FF, HP	Cytiva
	StrepTrap™ HP	Cytiva
	HiTrap™ Capto Q	Cytiva
	HiPrep CM FF 16/10	Cytiva
Pipettes	accu-jet® pro	Eppendorf
	Research® plus 2.5 µL	Eppendorf
	Research® 10 µL	Eppendorf
	Research® plus 10 µL	Eppendorf
	Pipetman® 10 µL	Gilson
	Research® 20 µL	Eppendorf

Instruments and consumables in this thesis, continued.

Purpose	Product	Manufacturer
Pipettes, continued	Research® plus 20 µL	Eppendorf
	Pipetman® 20 µL	Gilson
	Transferpette® S-12 200 µL	Brand
	Research® 200 µL	Eppendorf
	Research® plus µL	Eppendorf
	AXYPET™ 1000 µL	AXYGEN
	Research® 1000 µL	Eppendorf
	Research® plus 1000 µL	Eppendorf
	Research® 5000 µL	Eppendorf
Consumables	Intelli-Plate 96-3 LVR	ArtRobbins Instruments
	Transparent 96 Well Microplate	Greiner bio-one
	Amicon Ultra Ultracel-PL Membrane	Millipore
	Vivaspin® Centrifugal filter units	Sartorius
	MEMBRA-CEL® dialysis tubing	Serva
	ReadyLyzer 3	Serva
	Slide-A-Lyzer® Mini Dialysis Devices	Thermo Scientific
	Centrifugal Filter	VWR
Miscellaneous	EmulsiFlex-C5	AVESTIN
	Thermomixer comfort	Eppendorf
	BTD Dry Block Thermostat	Grant
	MR Hei-Standard heating stirrer	Heidolph
	MicroPulser™ Electroporator	Bio-Rad
	MJ Mini™ Thermal Cycler	Bio-Rad
	T100™Thermal Cycler	Bio-Rad
	DS-11 textsuperscript™	DeNovix
	Spark 10M	Tecan
	Ultrospec™ 10 Cell Density Meter	Harvard Biochrom
	DynaPro® NanoStar®	Wyatt Technology Corporation

Instruments and consumables in this thesis, continued.

Purpose	Product	Manufacturer
	S9i	Leica Microsystems
	Oryx8	Douglas Instruments
	LVSA 40/60	Zirbus
	Heratherm™ OGH60, OMH60 Dry Oven	Thermo Scientific
Miscellaneous, continued	inoLab® pH 720	WTW
	DRY-Line® 53 Dry Oven	VWR

5.2 Chemicals

All chemicals were purchased from common suppliers such as Carl Roth, Merck and VWR unless otherwise stated.

5.3 Cultivation Media

Media from Table 5.2 were autoclaved before use. Heat-labile components were sterile filtered.

Table 5.2: Composition of bacterial and yeast growth media used in this thesis.

Name	Ingredients
AI medium	0.5% (w/v) NaCl
	22 mM KH_2PO_4
	42 mM Na_2HPO_4
	2% (w/v) tryptone
	0.5% (w/v) yeast extract
	0.05% (w/v) glucose
	0.6% (v/v) glycerol
	0.2% (w/v) lactose

Composition of bacterial and yeast growth media used in this thesis, continued.

Name	**Ingredients**
2YT	0.5% (w/v) NaCl 1.6% (w/v) tryptone 1.0% (w/v) yeast extract
4YT	0.5% (w/v) NaCl 3.2% (w/v) tryptone 2.0% (w/v) yeast extract
Dynamite medium	17 mM KH_2PO_4 72 mM K_2HPO_4 1.6 mM $MgSO_4$ 1.2% (w/v) tryptone 2.4% (w/v) yeast extract 0.5% (w/v) glucose 0.5% (v/v) glycerol
LB Lennox	0.5% (w/v) NaCl 1% (w/v) tryptone 0.5% (w/v) yeast extract
Terrific broth (TB)	17 mM KH_2PO_4 72 mM K_2HPO_4 1.2% (w/v) tryptone 2.4% (w/v) yeast extract 0.4% (v/v) glycerol

Composition of bacterial and yeast growth media used in this thesis, continued.

Name	Ingredients
Super optimal broth with catabolic repression (SOC)	10 mM NaCl 1.5 mM KCl 10 mM $MgCl_2$ 10 mM $MgSO_4$ 2% (w/v) tryptone 0.5% (w/v) yeast extract 20 mM glucose
LB-agar	1% (w/v) NaCl 1% (w/v) tryptone 0.5% (w/v) yeast extract 1.5% (w/v) agar-agar
Yeast carbon base (YCB)	0.2-1% (v/v) acetamide 2% (w/v) agar powder 0.3% (v/v) Tris-HCl pH 7.0 0.12% (w/v) YCB medium powder
Yeast extract peptone dextrose (YPD)	2% (w/v) glucose 2% (w/v) tryptone 1% (w/v) yeast extract
Yeast extract peptone galactose (YPGal)	1% (w/v) galactose 2% (w/v) tryptone 1% (w/v) yeast extract

5.4 Enzymes, standards and kits

Table 5.3: DNA modifying enzymes.

Product	Source
Antarctic Phosphatase	NEB
Phusion® High-Fidelity DNA Polymerase	NEB
Q5 High-Fidelity DNA Polymerase	NEB
T4 DNA Ligase	NEB
T4 Polynucleotide Kinase	NEB
GoTaq® G2 DNA Polymerase	Promega
NEBuilder® HiFi DNA Assembly Master Mix	NEB
K. lactis Protein Expression Kit	NEB
Yeast Carbon Base Medium Powder	NEB
cOmplete™ ULTRA tablets, ethylene diamine tetraacetate (EDTA)-free	Roche

5.5 DNA/protein markers

Table 5.4: DNA and protein markers.

Ladder	Source
DNA-Marker short run *extended*	Carl Roth
1Kb DNA Ladder RTU	GeneDirex
1 kb Plus DNA Ladder	NEB
Quick-Load® Purple DNA ladder	NEB
TriDye™ Ultra Low Range DNA Ladder	NEB
TriDye™ 1kb Plus DNA Ladder	NEB
Unstained Protein Standard, Broad Range	NEB
peqGOLD DNA ladder	PEQLAB
Spectra™multicolor low range protein marker	Thermo Scientific

5.6 Kit systems

Table 5.5: Kits for DNA and protein purification.

Product	Supplier
Glutathione Sepharose™ 4B	Cytiva
His SpinTrap Columns	Cytiva
PD-10 desalting columns	Cytiva
StrepTactin®XT Spin Column Kit	IBA Lifesciences
Monarch® Plasmid Miniprep Kit	NEB
BugBuster® Protein Extraction Reagen	Novagen
Wizard® SV Gel and PCR Clean-Up System	Promega
Ni-NTA Agarose	Qiagen
Q5® Site-Directed Mutagenesis Kit	NEB

5.7 Primers

Table 5.6: Primers used for sequencing PCR reactions. Numbering according to database of the Schneider group.

Number	Name	Sequence
769	MBP-fw	gatgaagccctgaaagacgcgcag
1049	$FZ11\_pKLAC2\_fw1$	acttaaccggggatgaagtttc
1050	$T7\_fw\_sequence$	atgcgtccggcgtag
1051	$T7\_rev\_sequence$	ctagttattgctcagcgg
1052	$LAC4\_Terminator$	tatctcatagaaatatacctgtaagtacata cttatc
1103	$FZ18\_ScORF3\_GFP\_fwd1$	tttcgctgacaaggatgatcaaaa gaatgagtaaaggagaagaacttttcactgg
1132	Vat1s-stop-rev	taacccggaaccagcaga
1154	$FZ33\_Klactis\_Sequencing\_$	gcgataacaagctcaaca

Primers used for sequencing PCR reactions. Numbering according to database of the Schneider group, continued.

Number	Name	Sequence
	fwd	
1155	*FZ*34_*Klactis_Sequencing_rev*	ttatcgcacaagacaat
1159	*FZ*38_*pGEX_fwd*	gggctggcaagccacgtttggtg
1160	*FZ*39_*pGEX_rev*	ccgggagcttgcatgtgtcagagg
1172	*FZ*42_*Klactis_Sequencing_fwd*2	gtctgggtacccgtatcgttgac
1173	*FZ*43_*Klactis_Sequencing_rev*2	ccgtacctggatgctgggtac
1174	*FZ*44_*VAT* − 1	accgtgaaggttctccgca
1186	*FZ*47_*ScORF*3_*fwd*6	aatctttattttcagggcgccaaccgttctg accagtac
1187	*FZ*48_*ScORF*3_*rev*5	aagcttgtcgacggagctcgctatttacccg cacgcag
1214	*FZ*51_*ScORF*3_*Seq*_1	ctgacccgtctgatcg
1215	*FZ*52_*T*4_*Promoter_Seq*	cctttgcagggctggcaagc

Table 5.7: Primers used for cloning PCR reactions. Numbering according to database of the Schneider group.

Number	Name	Sequence
1004	*FZ*1_*Vat*1*fw*1	gatcaccatgggcagcagcc
1005	*FZ*1_*Vat*1*rev*1	cgatcactcgagctaacccgga
1037	*FZ*3_*ScORF*3*fw*1	gtcagtaccatggctagctg
1038	*FZ*4_*ScORF*3*rev*1	gcggccgttatacaagcttctatttacccg
1039	*FZ*5_*Vat*1*fw*2	agtggtggtggtggtggtgcccatgggcagc ag

Primers used for cloning PCR reactions. Numbering according to database of the Schneider group, continued.

Number	Name	Sequence
1040	$FZ6\_Vat1rev2$	actttaagaaggagatatacctcgagctaaccgg
1045	$FZ7\_ScORF3fw2$	ctcgagtaccatggctagctg
1046	$FZ8\_ScORF3rev2$	gcggccgcttatacaagcttctatttacccg
1047	$FZ9\_Vat1\_fw3$	actttaagaaggagatatacatgggcagcagccatcatc
1048	$FZ10\_Vat1\_rev3$	agtggtggtggtggtggtgcctaacccggaaccagcag
1049	$FZ11\_pKLAC2\_fw1$	acttaaccggggatgaagtttc
1070	MBP_p6_fwd6	tttcgctgacaaggatgatcgtaccatggctagctggag
1071	$FZ13\_ScORF3\_REV2$	ggatccgtcgacgatatcgctacaagcttctatttacccgc
1093	$FZ14\_Vat1\_fw4$	actttaagaaggagatatacatgctgctgcgttgcctg
1094	$FZ15\_Vat1\_rev4$	ggtgctcgagtgcggccgcattagtgatgatgatgatgatgcaggcggccgccgctgctctgaaaataaagattctcggatccacccggaaccagcagaactttacc
1095	$FZ16\_ScORF3\_fwd3$	gatcctcgagaaaagatggagccacccg
1096	$FZ17\_ScORF3\_rev3$	gatcgcggccgcctatttacccgcacgcag
1104	$FZ19\_ScORF3\_GFP\_rev1$	gatccgtcgacgatatcgcgaaaaaggggcagcccgc
1112	$FZ20\_Vat1\_Spacer\_Nterm\_fwd1$	tcgggaggtagcggctccctgctgcgttgcctg
1119	$FZ22\_Vat1\_Spacer\_Nterm\_fwd2$	gggaggtagcggcggtagcggcggccgcctggag
1120	$FZ23\_Vat1\_Spacer\_Nterm\_rev2$	gagccaccgctgccaccgctgtgatgatgatgatgatggctgctgccc

Primers used for cloning PCR reactions. Numbering according to database of the Schneider group, continued.

Number	Name	Sequence
1121	$FZ24\_Vat1\_Spcr1\_fwd3$	actttaagaaggagatatacatgcatcatca tcatcatcacgagaatctttattttcaggga ggcagcggtggctcgggaggtagcggcggta gcggcggctccctgctgcgttgcctg
1122	$FZ25\_Vat1\_Spcr1\_rev3$	gtcgacggagctcgaattcgtaacccggaac cagcagaac
1123	$FZ26\_Vat1\_Spcr2\_fwd4$	actttaagaaggagatatacatgcatcatca tcatcatcacagcggtggcagcggtggctcg ggaggtagcggcggtagcggcggcgagaatc tttattttcagggatcc
1130	$FZ28\_pBR322$	tatgtaagcagacagttttatt
1131	$Vat1s - stop - fw$	aattcgagctccgtcgac
1132	$Vat1s - stop - rev$	taacccggaaccagcaga
1150	$FZ29\_VAT - 1\_fwd5$	tccagggtaatgcaggatccctgctgcgctg cctagtg
1151	$FZ30\_VAT - 1\_rev5$	gctcgagtgcggccgcattaccctggaacca ggaggac
1152	$FZ31\_pET28a\_fwd1$	taatgcggccgcactcga
1153	$FZ32\_pET28a\_rev1$	ggatcctgcattaccctgg
1156	$FZ35\_ScORF3\_fwd4$	tttcgctgacaaggatgatctcgagaaaaga atggctagctggagccac
1157	$FZ36\_VAT - 1\_fwd6$	cctgggatccccgaattcccggctgctgcgc tgcctagtg
1158	$FZ37\_VAT - 1\_rev6$	tcgtcagtcagtcacgatgcttaccctggaa ccaggagg
1161	$FZ40\_VAT - 1\_fwd7$	aatctttattttcagggcgccctgctgcgct gcctagt
1162	$FZ41\_VAT - 1\_rev7$	tggtggtggtgctcgagtgcttaccctggaa ccaggagg

Primers used for cloning PCR reactions. Numbering according to database of the Schneider group, continued.

Number	Name	Sequence
1184	$FZ45\_ScORF3\_fwd5$	tctgttccaggggcccctgggtaaccgttct gaccagtac
1185	$FZ46\_ScORF3\_rev4$	tcgtcagtcagtcacgatgcctatttacccg cacgcag
1206	$FZ49\_ScORF3\_fwd7$	tttcgctgacaaggatgatctcgaaaaaaga tcccctatactaggttattgg
1207	$FZ50\_ScORF3\_rev6$	aattacctgcagggaattcgctatttacccg cacgcag
1330	$FZ53\_Integration\_Primer\_1$	acacacgtaaacgcgctcggt
1331	$FZ54\_Integration\_Primer\_2$	atcatccttgtcagcgaaagc

5.8 Plasmids

Table 5.8: Plasmids used in this thesis.

Number	Name
1	pRSETa-ScORF3
2	Vat1-pET28a-Cterm
3	ScORF3-pKLAC2
4	Vat1-pET28a
5	ScORF3-pKLAC2-Kex
5	Vat1-Spcr2-pET28a
6	VAT-1-Sprc2-pET28a
7	VAT-1-pET28a(β)
8	pKLAC2
9	pGEX-6P-3
10	VAT-1-pGEX-6P-3

Plasmids used in this thesis, continued.

Number	Name
11	MBP-VAT-1-pET28a
12	GST-ScORF3-pKLAC2

5.9 Bacterial Strains

Table 5.9: Bacterial strains used for plasmid amplification and protein expression.

Strain	Genotype	Usage
DH5α	F$^-$ φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 Δ$^-$ thi-1 *gyr*A96 *rel*A1	Cloning & plasmid amplification
NEBturbo	F' *proA*$^+$*B*$^+$ *lacI*^{q} Δ*lacZM15* / *fhuA2* Δ*(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)Tet*^{S} *endA1 thi-1* Δ*(hsdS-mcrB)5*	Cloning & plasmid amplification
BL21(DE3)	*fhuA2* [*lon*] *ompT gal (δ DE3)* [*dcm*] Δ*hsdS* λ *DE3* = λ *sBamHlo* Δ*EcoRI-B int::(lacI::PlacUV5::T7 gene1) is21* Δ*nin5*	Protein expression
BL21(DE3)pLysS	F$^-$ *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm* (DE3) plysS (Cam^R)	Protein expression
C41(DE3)	F$^-$ *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm* (DE3)	Protein expression
C43(DE3)plysS	F$^-$ *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm* (DE3) pLysS (Cm^R)	Protein expression
NiCo21(DE3)	*can::CBD fhuA2* [*lon*] *ompT gal* (λ *DE3)* [*dcm*] *amA::CBD slyD::CBD glmS6Ala* Δ*hsdS* λ *DE3* = λ *sBamHlo* Δ*EcoRI-B int::(lacI::PlacUV5:T7gene1) i21* Δ*nin5*	Protein expression
Rosetta2(DE3)	F$^-$ *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm* (DE3) pRARE2 (Cam^R)	Protein expression

Bacterial strains used for plasmid amplification and protein expression, continued.

Strain	Genotype	Usage
T7 Express*lysI/I^q*	MiniF *lysY lacI^q*(Cam^R) / *fhuA2 lacZ::T7 gene1* [*lon*] *ompT gal sulA11 R(mcr-73::miniTn10–Tet^S)2* [*dcm*] *R(zgb-210::Tn10–Tet^S) endA1Δ(mcrC-mrr) 114::Is10*	Protein expression

5.10 Buffers

Table 5.10: Formulation of other buffers used in this thesis.

Buffer	Components
Staining buffer	0.1% (w/v) coomassie brilliant blue G250 40% (v/v) EtOH 10% (v/v) acetic acid
SDS-Running buffer	25 mM tris pH 8.3 192 mM glycine 0.1% (w/v) SDS
1× TAE	40 mM Tris-HCl pH 8.0 40 mM acetic acid 1 mM (EDTA)
PBS	4 mM KH_2PO_4, pH 7.4 16 mM NA_2HPO_4 115 mM NaCl

Formulation of other buffers used in this thesis, continued.

Buffer	**Components**
PBS-blocking buffer	PBS buffer 3% (w/v) BSA 0.05% (v/v) Tween 20
PBS-T buffer	PBS buffer with 0.1% (v/v) Tween 20
5× SDS-sample buffer	200 mM Tris-HCl pH 6.8 12% (w/v) SDS 0.4% (w/v) romophenol blue 50% (w/v) glycerol 20% (v/v) β-mercaptoethanol
Buffer S1	100 mM Tris-HCl pH 8.0 150 mM NaCl 1 mM EDTA 1 pill cOmplete™ ULTRA tablets, EDTA-free per 50 mL 1 mg DNase per 50 mL
Buffer S2	100 mM Tris-HCl pH 8.0 150 mM NaCl 1 mM EDTA
Buffer S3	100 mM Tris-HCl pH 8.0 150 mM NaCl 1 mM EDTA 2 mM d-desthiobiotin
Buffer M1	24 mM N_2HPO_4 pH 8.0 274 mM NaCl 5.4 mM KCl 1 mM DTT 0.2% (v/v) Tween-20 10% (v/v) Glycerol

Formulation of other buffers used in this thesis, continued.

Buffer	Components
Buffer M1, continued	1 pill cOmplete™ ULTRA tablets, EDTA-free per 50 mL 1 mg DNase per 50 mL 10 mg lysozyme per 50 mL
Buffer M2	20 mM Tris-HCl pH 8.0 100 mM NaCl 1 mM DTT
Buffer M3	20 mM Tris-HCl pH 8.0 100 mM NaCl 1 mM DTT 10 mM maltose
Buffer G1	20 mM Tris-HCl pH 8.0 100 mM NaCl 1 mM DTT 10% (v/v) Glycerol
Buffer G2	20 mM Tris-HCl pH 8.0 100 mM NaCl 1 mM DTT 10% (v/v) Glycerol 15 mM reduced glutathione (GSH)
Buffer G3	20 mM Bis-Tris-Propane pH 6.0 1 mM DTT 10% Glycerol

Formulation of other buffers used in this thesis, continued.

Buffer	Components
Buffer G4	20 mM Bis-Tris-Propane pH 6.0 20 mM NaCl 1 mM DTT 10% Glycerol
Buffer G5	20 mM Bis-Tris-Propane pH 6.0 1 M NaCl 1 mM DTT 10% Glycerol

5.11 Crystallization Screens

Table 5.11: Commercial screens used for protein crystallization.

Product	Source
PEG/Ion	Hampton Research
PEG/Ionic Liquid	Hampton Research
Pre-Crytallization Test	Hampton Research
JBScreen Basic	Jena Bioscience
JBScreen Classic	Jena Bioscience
JBScreen JCSG ++	Jena Bioscience
JBScreen PACT ++	Jena Bioscience
JBScreen PEG/Salt	Jena Bioscience
JBScreen Pentaerythritol	Jena Bioscience
JBScreen Wizard	Jena Bioscience
XP Screen	Jena Bioscience
XP Up Screen	Jena Bioscience
$AmSO_4$ Suite	NeXtal

5.12 Software and tools

Table 5.12: Software and online tools used in this thesis.

Product	**Source**
Snap Gene Viewer	Snapgene
BioEdit	T. Hall
Unicorn 7	Cytiva
XDS	W. Kabsch, MPI for Medical Research, Heidelberg
ccp4	https://www.ccp4.ac.uk/
ChemDraw 22.0.0	Perkin Elmer
PyMOL	Schrödinger Inc.
Affinity Designer	Serif Europe Ltd.
Expasy ProtParam	SIB Swiss Institute of Bioinformatics
T_m Calculator	NEB

Methods

6.1 Working with Bacteria

Sterile personal protective equipment and gears were used for handling bacterial and yeast work. Before and after every experiment the work space was sterilized with disinfectants like Bacillol® or 70% EtOH.

6.2 Molecular Biology

6.2.1 Polymerase Chain Reaction

Amplification for incorporation of DNA fragments was performed by polymerase chain reaction (PCR). Annealing temperatures were calculated using the T_m calculator tool (NEB) based on the used primers' sequence or by gradient PCR.

Table 6.1: Standard programs used for various polymerases.

	Q5/Phusion		**GoTaq®**		
Step	**Temperature**	**Time**	**Temperature**	**Time**	
1	98 °C	30 s	98 °C	5 min	
2	98 °C	10 s	98 °C	15 s	} × 35
3	varying	30 s	varying	30 s	
4	72 °C	30 s per kb	72 °C	1 min per kb	
5	72 °C	2 min	72 °C	5 min	

The lid temperature was always 99 °C.

Standard PCR reactions were carried out using Q5 ® High-Fidelity DNA Polymerase (NEB) or Phusion® High-Fidelity DNA Polymerase (NEB) according to manufacturers' instructions with additional betaine and $MgSO_4$ (100 mM and 2.5 mM final concentration respectively) added. GoTaq® G2 DNA Polymerase was used for colony PCR according to manufacturers' instructions from single cell colonies. Colony PCR were performed using the primers listed in table 5.7.

Analysis was done by agarose gel (1% (w/v) agarose in 1× TAE buffer, table 5.10) electrophoresis and visualized using a UV-fluorescent dye (ROTI®GelStain from CarlRoth. 1:50 000). Amplified DNA fragments were purified using a commercial PCR purification kit form Promega.

6.2.2 Touchdown PCR

Touchdown PCR was used when nonspecific sequences were repeatedly obtained. Therefore gradually lowering T_m disadvantaged unwanted amplicons would be avoided.

Table 6.2: Program for Touchdown PCR with Q5 Polymerase.

Step	Temperature	Time	
1	98 °C	45 s	
2	98 °C	30 s	× 10
3	72 °C - 0.6 °C per cycle	30 s	
4	72 °C	50 s	
5	98 °C	30 s	× 22
6	varying	30 s	
7	72 °C	50 s	
8	72 °C	3 min	

The lid temperature was always 99 °C.

6.2.3 Colony PCR

Colony PCR were used to screen colony plates for positive clones containing the desired plasmid after transformation. They were carried out as mentioned in table 6.1 with single colonies as source of DNA template. Therefore the clones were picked from the plate and added in the PCR mixture with a total volume of 10 µL. A prolonged initial heating step should ensure proper lysis of the cells and acces to the plasmid DNA.

6.2.4 Cloning using Restriction Enzymes and Ligases

From purified vectors and inserts 10 µg each were digested with appropriate restriction enzymes for 1 h and purified from agarose gel. Subsequent ligation with T4 DNA ligase was carried out at a 3-fold molar excess of insert in 20 µL reaction volume either for 3 h at 16 °C or overnight at 4 °C. The obtained plasmids were then transformed into *E. coli*.

6.2.5 Gibson Assembly

Gibson assembly is an exonuclease-based method for isothermal DNA amplification of DNA double strands containing complementary ends with an overlap of 20-30 base pairs. In a one-pot reaction a T5 flap exonuclease generates sticky ends in the DNA fragments which can anneal. A polymerase then adds missing nucleotides and closes double strand gaps and a DNA ligase repairs any nicks. The so obtained plasmid can then directly be transformed into *E. coli* cells without further treatment. Any fragments not containing the overlap cannot form plasmids and are degraded by the cells.

Appropriate primers for PCR reaction were designed with the online NEBuilder® Assembly Tool. PCR was carried out according to table 6.1 using Q5 polymerase. The obtained fragments were purified by agarose gel (1% (w/v)) and assembled with the NEBuilder® HiFi DNA Assembly kit (NEB) according to manufacturers' instructions. A 2-fold molar excess of insert on a 50-100 ng scale was added to NEBuilder® HiFi DNA Assembly Master Mix (NEB) and incubated for 15 min at 50 °C. 2 µL of the ligated DNA was subsequently transformed in *E. coli* cells without further treatment.

6.2.6 Site-Directed Mutagenesis

SDM is a method to introduce single-point mutations in a double stranded plasmid DNA. A plasmid is amplified by PCR with back-to-back designed primers containing the mutation.

In the following step the nicked plasmid is repaired in a one-pot reaction containing T4 polynucleotide kinase, T4 DNA ligase and DpnI. The kinase phosphorylates the DNA 5' ends and the ligase repairs the nicks. Finally DpnI, which cleaves only methylated DNA, digests the unmodified, bacterial plasmid. SDM were carried out with the Q5® Site-Directed Mutagenesis Kit according to manufacturers' instructions. The obtained plasmids were then transformed into *E. coli*.

6.2.7 Transformation of Bacterial Cells

Transformation of plasmids into *E. coli* was carried out either by electroporation or heat-shock, depending on the target cells.

For electroporation 50 µL electrocompetent cells were thawed on ice. 10 µL of ligation reaction, or 50 ng of plasmid DNA were added, transferred into a cool cuvette and shocked with a pulse of 2.5 kV, taken up in 750 µL of prewarmed SOC medium at 37 °C for 45 min while shaking. The cells were spread on prewarmed agar plates containing the appropriate antibiotic and incubated at 37 °C over night.

For heat shock transformation 100 µL chemically competent cells were thawed on ice. 100 ng of plasmid were added, incubated on ice for 30 min and heat shocked at 42 °C for exactly 30 s. Follwing subsequent incubation on ice for 5 min the cells were taken up in 750 µL of prewarmed SOC medium at 37 °C for 45 min while shaking. The cells were spread on prewarmed agar plates containing the appropriate antibiotic and incubated at 37 °C over night.

6.2.8 Transformation in Yeast Cells

Prior to transformation the expression cassette had to be generated for incorporation into the yeast genome. Therefore the respective plasmid was linearized by digesting 2 µg of plasmid with 20 units of SacII at 37 °C for 2 h. The digested DNA was subsequently desalted and used for transformation.

For transformation into *K. lactis* were thawed on ice, taken up in 620 µL of Yeast Transformation Reagent (NEB). 1 µg of linearized DNA, or a maximum of 15 µL, were added, incubated at 30 °C for 30 min and subsequently heat shocked at 37 °C for 1 h. After subsequent centrifugation at 4 600 g for 2 min the supernatant was separated and the cells

taken up in 1 mL of YPD medium. Washing of the cells was repeated once before they were shaken at 30 °C for 3-4 h. The grown cells were centrifuged again at 4 600 g for 2 min, taken up in 1 mL phosphate buffered saline PBS and several dilutions of 10, 50 and 100 µL of cells in 500 µL of H_2O were prepared. Each was spread on YCB agar plates and incubated at 30 °C for 3-4 days for the cells to grow. About 10-20 colonies were transferred to fresh YCB agar plates for further usage.

Yeast colonies were tested for proper integration of the expression cassette by performing PCR with primers 1130 and 1131 (see 5.7 according to manufacturers' instructions).

6.2.9 Plasmid Purification

For the purifcation of plasmids 5 mL of LB medium were inoculated with cells containing the plasmid according to colony-PCR confirmation or Sanger Sequencing and grown under presence of apropriate antibiotics at 37 °C over night. Plasmids were obtained by using the Monarch® Plasmid MiniPrep kit (NEB). Elution was carried out with 30 µL of elution buffer or dd-H_2O.

6.2.10 Control Digestion and Sequencing

Correct plasmids were chosen based on control digestion with restriction enzymes and subsequent Sanger Sequencing. For the control digestion 500 ng of the plasmid were cleaved with 0.5 U of one or more restriction enzymes at 37 °C for 30 min. Cleavage patterns were analyzed by agarose gel (1% (w/v)) electrophoresis. Plasmids were prepared for sequencing by adding 150-500 ng of DNA to 25 pmol of the respective sequencing primer with a total volume of 10 µL. Sequencing was carried out by GENEWIZ.

6.2.11 Determining Protein and Nucleic Acid Concentration

The concentration of protein and nucleic acid samples was determined by spectrophotometric analysis at 280 and 260 nm respectively. For this a DS 11 Spectrophotometer (DeNovix) was used with a sample volume 1.0 µL. Theoretical extinction coefficients and molecular masses to determine protein concentrations were calculated with ProtParam (https://web.expasy.org/protparam/).

6.3 Protein Expression

6.3.1 VAT-1

The VAT-1-containing pGEX-6P-3 plasmid was transformed into chemically competent C43(DE3)plysS cells according to section 6.2.7. Dynamite medium was inoculated 1:50 with overnight culture, cells were grown under ampicillin selection at 37 °C and 180 rpm until an OD_{600} of 0.8 was reached. Protein expression was induced by addition of IPTG to a final concentration of 200 µM at 16 °C for 18 h. Cells were harvested at 5 000 g and 4 °C for 30 min, washed once with (PBS), pelleted at 4 000 g and 4 °C for 20 min and the pellet was stored at -20 °C.

6.3.2 ScORF3

The ScORF3 containing plasmids (pRSETa and pET28a) were transformed into electro-competent BL21(DE3) cells according to section 6.2.7. Expressions in 2YT or LB medium were carried out by 1:1000 inoculation with overnight culture. The cells were grown under the respective antibiotics selection at 37 °C to an OD_{600} of 0.6 and induced with 500 µL at 16 °C for 18 h. For expressions in AI medium was inoculated 1:100 with overnight culture, cells were grown under ampicillin selection at 37 °C and 180 rpm for 24 h. All cells were harvested at 5 000 g and 4 °C for 30 min, washed once with PBS, pelleted at 4 000 g and 4 °C for 20 min and the pellet was stored at -20 °C.

6.3.3 Yeast Cells

The appropriate linearized pKLAC1-based expression cassette was transformed into *K. lactis* competent cells according to section 6.2.8. YPGal medium was inoculated with cells from a single colony and grown at 30 °C and 250 rpm for 5 d. The cells were pelleted at 4 000 g and 4 °C for 20 min and the supernatant was separated. The secreted protein was precipitated by addition of $(NH_4)_2SO_4$ to a final concentration of 3 M, shaken at 8 °C and 80 rpm for 1 h and the protein was pelleted at 15 000 g and 4 °C for 30 min. The protein pellet was stored at -20 °C.

6.4 Protein Purification

All purification steps were carried out on ice or at 4 °C. Centrifugation steps were carried out at 4 °C. All buffers were filtered and degased prior to usage.

6.4.1 *Strep*-tagged® ScORF3

The buffers used for the purification are described in table 5.10. Cells were resuspended in buffer S1, lysed in 1× Bug Buster (Novagen) according to manufacturers' instructions. Insoluble cell debris was separated by centrifugation at 9 000 g and 4 °C for 10 min. The supernatant was purified using a *Strep*-Tactin®XT spin column (iba lifesciences) according to manufacturers' instructions. The eluted protein was subsequently used for SDS-PAGE and western blotting (see sections 6.5.2).

6.4.2 MBP-tagged ScORF3

The buffers used for the purification are described in table 5.10. Cells were resuspended in buffer M1, lysed by running the sample 3× through a homogenizer. Insoluble cell debris was separated by centrifugation at 48 000 g and 4 °C for 30 min. The supernatant was filtered through a fold filter and bound to a dextrin sepharose packed MBPTrap HP column (cytiva), washed with buffer M2 and eluted with buffer M3. The affintiy tag was removed by incubation with TEV protease for at 4 °C for 16 h and concentrated with centrifugal filter units (30 000 kDa molecular weight cutoff (MWCO)). Subsequent purification by SEC using a Superdex™ 200 Increase 10/300 GL using buffer M2. Fractions containing the pure protein according to SDS-PAGE (see section 6.5.1) were pooled, concentrated with centrifugal filter units (30 000 kDa MWCO) to 3-4 $mg \cdot mL^{-1}$ and either subsequently used for crystallization or aliquoted, flash frozen in liq. N_2 and stored at -80 °C.

6.4.3 GST-tagged VAT-1

The buffers used for the purification are described in table 5.10. Cells were resuspended in buffer M1 and lysed by running the sample 4× through a homogenizer. Insoluble cell debris was separated by centrifugation at 48 000 g and 4 °C for 30 min. The supernatant was filtered through a residual filter and bound to a glutathione packed sepharose beads and unbound samples was washed out with buffer G1. The GST-tag was removed by addition of PreScission Protease. Following a 3 h incubation the cleaved protein was washed out with buffer G1 and bound proteins eluted with buffer G2. Fractions containing the cleaved

protein were pooled, diluted with buffer G3 and subsequently purified by CIEX using a Resource S 1 mL column. The protein was bound to the column, washed with buffer G4 and eluted using a linear gradient to 100% buffer G5. Samples containing the protein were determined by SDS-PAGE, pooled and concentrated with centrifugal filter units (3 000 kDa MWCO) to 2-3 mg·mL^{-1} and either subsequently used for crystallization or aliquoted, flash frozen in liq. N_2 and stored at -80 °C.

6.5 Protein Analysis

6.5.1 SDS-PAGE and Visualization

Quality assessment of protein samples was done using SDS-PAGE analysis. Therefore 20 µL of the samples were mixed with 5 µL SDS-sample buffer (table 5.10), heat denatured (5 min at 95 °C) and loaded on a PAA gel. SDS-PAGE was carried out at 300 V, 5 mA cm^{-1} with SDS-Running buffer (see table 5.10). Gels were stained with either Seragen QC Stain (avantor™) or by applying Staining buffer (see table 5.10 and heating once. Excessively stained gels were destained using 10% (v/v) AcOH. PAA gels were either commercially available (Bio-Rad Laboratories) 4-20% gradient gels or hand made using the Hoefer system (see table 6.3). Therefore 3.5 mL resolving gel solution were prepared, polymerization was started upon addition of ammonium persulafte (APS) and *N,N,N',N'*-tetramethylethylenediamine (TEMED), added between glass plates, covered with isopropanol (iPrOH) to smoothen the surface and remove bubbles and polymerized. After removal of iPrOH the freshly prepared stacking gel was added and a comb was inserted. The polymerized gels were wrapped in wet paper and stored at 4 °C until usage.

Table 6.3: PAA gels produced with the Hoefer system.
Formulation for hand made gels with a total volume of 5 mL.

Compound	Resolving Gel	Stacking Gel
acrylamide 29:1	12% (v/v)	4% (v/v)
tris-HCl	60 mM, pH 8.8	100 mM, pH 6.8
SDS	0.01% (w/v)	0.016% (w/v)
H_2O	46% (v/v)	70% (v/v)
10% (v/v) APS	0.1% (v/v)	0.2% (v/v)

Compound	Resolving Gel	Stacking Gel
TEMED	0.1% (v/v)	0.2% (v/v)

6.5.2 Semi-Dry Western Blotting

Western blotting and subsequent immunostaining were used for specific detection of tagged proteins depending on their respective affinity tag. For buffer composition see table 5.10.

Strep-tagged® proteins were detected by interaction with *Strep*-Tactin®-linked horse radish peroxidase (HRPip). After SDS-PAGE the protein was transferred to a polyvinylidene fluoride (PVDF) membrane with 2.4 mA cm^{-2}. To evaluate transfer efficiency the membrane was reversibly stained with ponceau S and destained with H_2O. Free binding sites were saturated with PBS-blocking buffer containing BSA at 4 °C over night, or at room temperature for 1 h. Excessive BSA was removed by washing twice with PBS-T buffer before application of a 1:100 dilution of *Strep*-Tactin®-linked HRP in PBS-T buffer at room temperature for 1 h under gentle shaking and washing the membrane thrice with PBS-T buffer and PBS buffer each. HRP was detected by activation of its enzymatic functionality. Therefore 5 mL of 5-bromo-4-chloro-3-indolyl phosphate (BCIP®)/nitro-blue tetrazolium chloride (NBT)-blue liquid substrate (Merck) was added to the membrane at room temperature and incubated for 1 h. If necessary the reaction was terminated by addition of H_2O.

His-tagged proteins were visualized similar to those with *Strep*-tags® except Ni^{2+}-activated HRP was used and chemoluminescent reaction was started by addition SuperSignal™ West Pico Plus substrate and subsequent incubation at 23 °C for 5 min.

Western blots were imaged using an Amersham ImageQuant™.

6.6 Oxidoredcutase Assay

VAT-1 activity was assessed by an oxidoreductase assay. Therefore pure VAT-1 was diluted in 100 mM Tris-HCl at pH 7.8 to 2.5 µM. 50 µL were added in a 96 well half-area microplate and the reaction was started by addition of 1 µL of a 2:1 mixture of NADPH and 9,10-phenanthrenequinone. Absorption was measured at 340 nm in 5 min intervalls

over the course of 2 h. For determination of the background a reaction without protein was set up (see table 6.4). All data points were measured as triplicates with the error bars showing the respective standard deviations.

Table 6.4: Composition of the oxidoreductase assay used to determine VAT-1 activity.

Compound	Reaction	Background
VAT-1	2.5 µM	0 µM
NADPH	100 µM	100 µM
9,10-phenanthrenequinone	50 µM	50 µM
tris-HCl, pH 8.0	100 mM	100 mM

6.7 Crystallography

6.7.1 Protein Crystallization

Prior to setting up crystallizations aggregates were removed by centrifugation at 17 000 g for 10 min and 4 °C. Protein concentration suitable for screening was determined by Pre-Crystallization Test (Hampton Research) according to manufacturers' instructions. Screening for hit conditions was done with commercially available screens from Hampton Research, Jena Bioscience and NeXtal in an Intelli-Plate 96-3 using an Oryx8 crystallization robot. The drops were prepared by adding together precipitant and protein solution in a 1:1 volumetric ratio with a total volume of 300 nL. Based on initial hits finescreens were designed and prepared around these hit conditions. Seeding experiments were set up by adding 100 nL seedstock to 300 nL of crystallization drop. For preparation of seedstocks the crystals in one well were crushed, washed excessively with mother liquor and flash frozen in liquid N_2. To dehydrate crystals a well was cut open, 300 nL of 4 M $(NH_4)_2SO_4$, saturated $(NH_4)_2SO_4$ or saturated NaCl were added and the plate was sealed again for 12-20 h.

Crystals of VAT-1 were obtained at a protein concentration of 3.7 mg/mL, a precipitant of 1.26 M ammonium sulfate, 200 mM Li_2SO_4 and 100 mM tris at pH 8.5 and were grown at room temperature. Crystals were picked with nylon loops (Hampton Research), covered in 4 M Li_2SO_4 cryoprotectant and flash frozen in liquid N_2.

6.7.2 X-Ray Data Collection and Structure Determination

X-ray diffraction data was collected with the P13 or P14 beamline of the European Molecular Biology Laboratory (EMBL) at the DESY in Hamburg and ID23-1 or ID23-2 beamline of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. During data collection samples were cooled with liquid N_2. Phasing was performed by molecular replacement with MOLREP with pdb entry 6K9Y as the template. Model building was carried out with Coot and the structure was refined with Refmac5 as part of the CCP4 software suite. Statistics for data collection and refinement are summarized in table 7.

Abbreviations

2YT two yeast extract tryptone

4YT four yeast extract tryptone

AASTY poly(acrylic acid co-styrene

ABP activity based probes

ABPP activity based protein profiling

ADH alcohol dehydrogenase

AI autoinduction medium

APS ammonium persulfate

AUC area under the curve

BCIP 5-bromo-4-chloro-3-indolyl phosphate

CIEX cation exchange chromatography

CL_{int} intrinsic clearance

CuAAC Cu(I)-catalyzed azide-alkyne cycloaddition

DARTS drug affinity responsive target stability

DESY Deutsches Elektronen Synchrotron

DIBMA diisobutylene-maleic acids

DiME dimethyl labelling

ECM extracellular matrix

EDTA ethylene diamine tetraacetate

EMBL European Molecular Biology Laboratory

ESRF European Synchrotron Research Facility

GPCR G-protein coupled receptor

GSH glutathione

GST glutathione S-transferase

HRP horseradish peroxidase

HRV 3C human rhinovirus 3C

IMAC immobilized metal affinity chromatography

iPrOH isopropanol

IPTG Isopropyl-β -D-thiogalactopyranoside

iTRAQ isobaric tags for relative and absolute quantitation

LB lysogeny broth

MAD multi wavelength anomalous dispersion

MBP maltose-binding protein

MF mating factor

MP membrane protein

MS mass spectrometry

MSP membrane scaffold protein

MWCO molecular weight cutoff

NADP nicotinamide adenine dinucleotide

NADPH nicotinamide adenine dinucleotide phosphate

NBT nitro-blue tetrazolium chloride

NCA neocarzilin A

NCB neocarzilin B

NCC neocarzilin C

NCZ neocarzilin

NP natural products

OD optical density

ORF open reading frame

P_{app} apparent permeability

PAA polyacrylamide

PBP penicillin binding protein

PBS phosphate buffered saline

PCR polymerase chain reaction

PDD phenotypic drug discovery

PEG polyethylene glycol

PKS polyketide synthase

POI protein of interest

PVDF polyvinylidene fluoride

RMSD root mean square deviation

SAD single wavelength anomalous dispersion

SAR structure acrivity relationship

ScORF3 *streptomyces carzinostaticus* ORF3

SDM site directed mutagenesis

SDS sodium dodecyl sulfate

SEC size exclusion chromatography

SILAC stable isotope labelling by amino acids in cell culture

SMA styrene-maleic anhydride

SOC super optimal broth with catabolic repressor

SPROX stability of proteins from rates of oxidation

TB terrific broth

TDD target based drug discovery

TEMED *N*,*N*,*N'*,*N'*-tetramethylethylenediamine

TEV tobacco etch virus

TICC target identification by chromatographic co-elution

TMT tandem mass tag

TPP thermal protein profiling

VAT-1 vesicle amine transport protein 1

YCB yeast carbon base

YPD yeast extract peptone dextrose

YPGal yeast extract peptone galactosel

Appendix

Sequences

Table 5: Sequences of ScORF3 constructs.

ScORF3*Strep*	MASWSHPQFEKGARLENLYFQGGSNRSDQYDVAILGSGMAGGMLGAVLARNGVKVLLLDA GTHPRFAVGESTIPYTSGLTRLIADRYRVPELRALSSFKGIREQVSRNCGQKQNFGFVYH REGSPQDWQEINQLVVPSVLRTETHLFRQDIDAYLFHVAVKYGAHPRLGTRIVDVETDPD TGAVLRTDGGEEFRAHYVVDGSGFRSPLAEKFALRETPTRARTHSRCLFTHMIGVEPFDK APAARRHDQPNPWHHGTLHHVFDGGWLWVIPFDNNEHSLNPLCSVGLTLDPRVHPKGDRT PQQEFDDFLARYPEIAHQFRGAKAVRPWVSTGRLQYSAKQVVGERFCLTSHAAGFIDALY SRGLTNTMELVNALGWRLIAASKDGDWSMERFGYLEDLQQGLFDFHDDIVYSSFVGFRDY ELWNAVNRTWMLGTMLGNVMLEDAYYRFERTGDDGVFRELEEFGHPGSPLPVSEGFTRMG PLTRELCEAVDQGTETSGEAARKILSYIRDADFIAPSFRFGERDTRCFAMSPAKMAMNAR WCRKDAPPEIGPRMINASKGLVRMRLRAGK
ScORF3*Lact*	MKFSTILAASTALISVVMAAPVSTETDIDDLPISVPEEALIGFIDLTGDEVSLLPVNNGT HTGILFLNTTIAEAAFADKDDLEKRWSHPQFEKGARLENLYFQGGSNRSDQYDVAILGSG MAGGMLGAVLARNGVKVLLLDAGTHPRFAVGESTIPYTSGLTRLIADRYRVPELRALSSF KGIREQVSRNCGQKQNFGFVYHREGSPQDWQEINQLVVPSVLRTETHLFRQDIDAYLFHV AVKYGAHPRLGTRIVDVETDPDTGAVLRTDGGEEFRAHYVVDGSGFRSPLAEKFALRETP TRARTHSRCLFTHMIGVEPFDKAPAARRHDQPNPWHHGTLHHVFDGGWLWVIPFDNNEHS LNPLCSVGLTLDPRVHPKGDRTPQQEFDDFLARYPEIAHQFRGAKAVRPWVSTGRLQYSA KQVVGERFCLTSHAAGFIDALYSRGLTNTMELVNALGWRLIAASKDGDWSMERFGYLEDL QQGLFDFHDDIVYSSFVGFRDYELWNAVNRTWMLGTMLGNVMLEDAYYRFERTGDDGVFR ELEEFGHPGSPLPVSEGFTRMGPLTRELCEAVDQGTETSGEAARKILSYIRDADFIAPSF RFGERDTRCFAMSPAKMAMNARWCRKDAPPEIGPRMINASKGLVRMRLRAGK

Sequences of ScORF3 constructs, continued

ScORF3*MBP*	MKHHHHHHPMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQV AATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAV EALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFK YENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPW AWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGL EAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA ASGRQTVDEALKDAQTNSGSGSGSENLYFQGANRSDQYDVAILGSGMAGGMLGAVLARNG VKVLLLDAGTHPRFAVGESTIPYTSGLTRLIADRYRVPELRALSSFKGIREQVSRNCGQK QNFGFVYHREGSPQDWQEINQLVVPSVLRTETHLFRQDIDAYLFHVAVKYGAHPRLGTRI VDVETDPDTGAVLRTDGGEEFRAHYVVDGSGFRSPLAEKFALRETPTRARTHSRCLFTHM IGVEPFDKAPAARRHDQPNPWHHGTLHHVFDGGWLWVIPFDNNEHSLNPLCSVGLTLDPR VHPKGDRTPQQEFDDFLARYPEIAHQFRGAKAVRPWVSTGRLQYSAKQVVGERFCLTSHA AGFIDALYSRGLTNTMELVNALGWRLIAASKDGDWSMERFGYLEDLQQGLFDFHDDIVYS SFVGFRDYELWNAVNRTWMLGTMLGNVMLEDAYYRFERTGDDGVFRELEEFGHPGSPLPV SEGFTRMGPLTRELCEAVDQGTETSGEAARKILSYIRDADFIAPSFRFGERDTRCFAMSP KMAMNARWCRKDAPPEIGPRMINASKGLVRMRLRAGK

Table 6: Sequences of VAT-1 constructs.

VAT-1*His*	MHHHHHHENLYFQGNAGSLLRCLVLTGFGGYDKVKLQSRPAAPPAPGPGQLTLRLRACGL NFADLMARQGLYDRLPPLPVTPGMEGAGVVIAVGEGVSDRKAGDRVMVLNRSGMWQEEVT VPSVQTFLIPEAMTFEEAAALLVNYITAYMVLFDFGNLQPGHSVLVHMAAGGVGMAAVQL CRTVENVTVFGTASASKHEALKENGVTHPIDYHTTDYVDEIKKISPKGVDIVMDPLGGSD TAKGYNLLKPMGKVVTYGMANLLTGPKRNLMALARTWWNQFSVTALQLLQANRAVCGFHL GYLDGEVELVSGVVARLLALYNQGHIKPHIDSVWPFEKVADAMKQMQEKKNVGKVLLVPG
VAT-1*Opt*	MGSSHHHHHHSSGGRLENLYFQGSLLRCLVLTGFGGYDKVKLQSRPAAPPAPGPGQLTLR LRACGLNFADLMARQGLYDRLPPLPVTPGMEGAGVVIAVGEGVSDRKAGDRVMVLNRSGM WQEEVTVPSVQTFLIPEAMTFEEAAALLVNYITAYMVLFDFGNLQPGHSVLVHMAAGGVG MAAVQLCRTVENVTVFGTASASKHEALKENGVTHPIDYHTTDYVDEIKKISPKGVDIVMD PLGGSDTAKGYNLLKPMGKVVTYGMANLLTGPKRNLMALARTWWNQFSVTALQLLQANRA VCGFHLGYLDGEVELVSGVVARLLALYNQGHIKPHIDSVWPFEKVADAMKQMQEKKNVGK VLLVPG
VAT-1*Spcr1*	MGSSHHHHHHSGGSGGSGGSGGSGGRLENLYFQGSLLRCLVLTGFGGYDKVKLQSRPAAP PAPGPGQLTLRLRACGLNFADLMARQGLYDRLPPLPVTPGMEGAGVVIAVGEGVSDRKAG DRVMVLNRSGMWQEEVTVPSVQTFLIPEAMTFEEAAALLVNYITAYMVLFDFGNLQPGHS VLVHMAAGGVGMAAVQLCRTVENVTVFGTASASKHEALKENGVTHPIDYHTTDYVDEIKK ISPKGVDIVMDPLGGSDTAKGYNLLKPMGKVVTYGMANLLTGPKRNLMALARTWWNQFSV TALQLLQANRAVCGFHLGYLDGEVELVSGVVARLLALYNQGHIKPHIDSVWPFEKVADAM KQMQEKKNVGKVLLVPG
VAT-1*Spcr2*	MGSSHHHHHHSSGGRLENLYFQGGGSGGSGGSGSLLRCLVLTGFGGYDKVKLQSRPAAPP APGPGQLTLRLRACGLNFADLMARQGLYDRLPPLPVTPGMEGAGVVIAVGEGVSDRKAGD RVMVLNRSGMWQEEVTVPSVQTFLIPEAMTFEEAAALLVNYITAYMVLFDFGNLQPGHSV LVHMAAGGVGMAAVQLCRTVENVTVFGTASASKHEALKENGVTHPIDYHTTDYVDEIKKI SPKGVDIVMDPLGGSDTAKGYNLLKPMGKVVTYGMANLLTGPKRNLMALARTWWNQFSVT ALQLLQANRAVCGFHLGYLDGEVELVSGVVARLLALYNQGHIKPHIDSVWPFEKVADAMK QMQEKKNVGKVLLVPG
VAT-1*Spcr3*	MLLRCLVLTGFGGYDKVKLQSRPAAPPAPGPGQLTLRLRACGLNFADLMARQGLYDRLPP LPVTPGMEGAGVVIAVGEGVSDRKAGDRVMVLNRSGMWQEEVTVPSVQTFLIPEAMTFEE AAALLVNYITAYMVLFDFGNLQPGHSVLVHMAAGGVGMAAVQLCRTVENVTVFGTASASK HEALKENGVTHPIDYHTTDYVDEIKKISPKGVDIVMDPLGGSDTAKGYNLLKPMGKVVTY GMANLLTGPKRNLMALARTWWNQFSVTALQLLQANRAVCGFHLGYLDGEVELVSGVVARL LALYNQGHIKPHIDSVWPFEKVADAMKQMQEKKNVGKVLLVPGGSENLYFQSSGGRLHHH HHH

Sequences of VAT-1 constructs, continued

VAT-1*MBP*	MKHHHHHHPMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQV AATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAV EALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFK YENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPW AWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGL EAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA ASGRQTVDEALKDAQTNSGSGSGSENLYFQGALLRCLVLTGFGGYDKVKLQSRPAAPPAP GPGQLTLRLRACGLNFADLMARQGLYDRLPPLPVTPGMEGAGVVIAVGEGVSDRKAGDRV MVLNRSGMWQEEVTVPSVQTFLIPEAMTFEEAAALLVNYITAYMVLFDFGNLQPGHSVLV HMAAGGVGMAAVQLCRTVENVTVFGTASASKHEALKENGVTHPIDYHTTDYVDEIKKISP KGVDIVMDPLGGSDTAKGYNLLKPMGKVVTYGMANLLTGPKRNLMALARTWWNQFSVTAL QLLQANRAVCGFHLGYLDGEVELVSGVVARLLALYNQGHIKPHIDSVWPFEKVADAMKQM QEKKNVGKVLLVPG
VAT-1*GST*	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID GDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKV DFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFK KRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSPNSRLLRCL VLTGFGGYDKVKLQSRPAAPPAPGPGQLTLRLRACGLNFADLMARQGLYDRLPPLPVTPG MEGAGVVIAVGEGVSDRKAGDRVMVLNRSGMWQEEVTVPSVQTFLIPEAMTFEEAAALLV NYITAYMVLFDFGNLQPGHSVLVHMAAGGVGMAAVQLCRTVENVTVFGTASASKHEALKE NGVTHPIDYHTTDYVDEIKKISPKGVDIVMDPLGGSDTAKGYNLLKPMGKVVTYGMANLL TGPKRNLMALARTWWNQFSVTALQLLQANRAVCGFHLGYLDGEVELVSGVVARLLALYNQ GHIKPHIDSVWPFEKVADAMKQMQEKKNVGKVLLVPG

Crystal Structure Solution Statistics

Table 7: Statistics for data collection and refinement of VAT-1 crystal structure.
Statistics for highest resolution shell in parentheses. Generated by phenix.

Statistic	Value
Wavelength (Å)	0.9763
Resolution range (Å)	39.62-2.5 (2.589-2.5)
Space group	I 4
Unit cell (Å)	125.3 125.3 148.0 90 90 90
Total reflections	528,547 (49,219)
Unique reflections	39,418 (3,899)
Multiplicity	13.4 (12.6)
Completeness (%)	99.5 (95.4)
Mean I/σ(I)	12.2 (0.49)
Wilson B-factor	79.2
R-merge	0.138 (3.835)
$CC_{1/2}$	0.999 (0.281)
CC^*	1 (0.663)
Reflections used in refinement	39,217 (3,719)
Reflections used for R-free	1,961 (185)
R-work	0.296 (0.431)
R-free	0.315 (0.479)
CC(work)	0.882 (0.545)
CC(free)	0.874 (0.557)
Number of non-H atoms	5,11
Number of macromolecules	5,11
Protein residues	676
RMS(bonds)	0.010
RMS(angles	1.90
Average B-factor	81.69

Literature

[1] C. M. Gleissner, C. L. Pyka, W. Heydenreuter, T. F. Gronauer, C. Atzberger, V. S. Korotkov, W. Cheng, S. M. Hacker, A. M. Vollmar, S. Braig, S. A. Sieber, *ACS Central Science* **2019**, *5*, 1170–1178.

[2] M. Otsuka, K. Ichinose, I. Fujii, Y. Ebizuka, *Antimicrobial Agents and Chemotherapy* **2004**, *48*, 3468–3476.

[3] Y. Liu, M. P. Patricelli, B. F. Cravatt, the Scripps Research Institute, **1999**.

[4] A. B. Berger, P. M. Vitorino, M. Bogyo, Activity-Based Protein Profiling Applications to Biomarker Discovery, In Vivo Imaging and Drug Discovery, **2004**.

[5] J. D. Fox, K. M. Routzahn, M. H. Bucher, D. S. Waugh, *FEBS Letters* **2003**, *537*, 53–57.

[6] J. P. Kahler, S. H. Verhelst, *RSC Chemical Biology* **2021**, *2*, 1285–1290.

[7] D. Greenbaum, K. F. Medzihradszky, A. Burlingame, M. Bogyo, Epoxide electrophiles as activity-dependent cysteine protease proćling and discovery tools, **2000**.

[8] G. C. Adam, E. J. Sorensen, B. F. Cravatt, *Nature biotechnology* **2002**, *20*, 805–9.

[9] M. Fonović, M. Bogyo, *Expert Review of Proteomics* **2008**, *5*, 721–730.

[10] D. Conole, M. Mondal, J. D. Majmudar, E. W. Tate, Recent Developments in Cell Permeable Deubiquitinating Enzyme Activity-Based Probes, **2019**.

[11] H. J. Benns, C. J. Wincott, E. W. Tate, M. A. Child, Activity- and reactivity-based proteomics: Recent technological advances and applications in drug discovery, **2021**.

[12] C. W. Cunningham, A. Mukhopadhyay, G. H. Lushington, B. S. Blagg, T. E. Prisinzano, J. P. Krise, *Molecular Pharmaceutics* **2010**, *7*, 1301–1310.

[13] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, H. K. A. C. Coolen, P. W. N. M. van Leeuwen, R. J. M. Nolte, Harrowfield in Calixarenes, **2002**.

[14] J. C. Jewett, E. M. Sletten, C. R. Bertozzi, *Journal of the American Chemical Society* **2010**, *132*, 3688–3690.

[15] M. Wiessler, W. Waldeck, C. Kliem, R. Pipkorn, K. Braun, The Diels-Alder-Reaction with inverse-Electron-Demand, a very efficient versatile Click-Reaction Concept for proper Ligation of variable molecular Partners, **2010**.

[16] D. P. Murale, S. C. Hong, M. M. Haque, J. S. Lee, Photo-affinity labeling (PAL) in chemical proteomics: A handy tool to investigate protein-protein interactions (PPIs), **2017**.

[17] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proceedings of the National Academy of Sciences* **2002**, *99*, 11020–11024.

[18] J. W. Chin, P. G. Schultz, *ChemBioChem* **2002**, *3*, 1135–1137.

[19] P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D. J. Pappin, *Molecular and Cellular Proteomics* **2004**, *3*, 1154–1169.

[20] A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, C. Hamon, *Analytical Chemistry* **2003**, *75*, 1895–1904.

[21] L. Dayon, A. Hainard, V. Licker, N. Turck, K. Kuhn, D. F. Hochstrasser, P. R. Burkhard, J. C. Sanchez, *Analytical Chemistry* **2008**, *80*, 2921–2931.

[22] T. Werner, G. Sweetman, M. F. Savitski, T. Mathieson, M. Bantscheff, M. M. Savitski, *Analytical Chemistry* **2014**, *86*, 3594–3601.

[23] J. L. Hsu, S. Y. Huang, N. H. Chow, S. H. Chen, *Analytical Chemistry* **2003**, *75*, 6843–6852.

[24] Y. Oda, K. Huang, F. R. Cross, D. Cowburn, B. T. Chait, *Proceedings of the National Academy of Sciences* **1999**, *96*, 6591–6596.

[25] A. Mateus, N. Kurzawa, I. Becher, S. Sridharan, D. Helm, F. Stein, A. Typas, M. M. Savitski, *Molecular Systems Biology* **2020**, *16*, DOI `10.15252/msb.20199232`.

[26] B. Lomenick, R. Hao, N. Jonai, R. M. Chin, M. Aghajan, S. Warburton, J. Wang, R. P. Wu, F. Gomez, J. A. Loo, J. A. Wohlschlegel, T. M. Vondriska, J. Pelletier, H. R. Herschman, J. Clardy, C. F. Clarke, J. Huang, *Proceedings of the National Academy of Sciences* **2009**, *106*, 21984–21989.

[27] G. M. West, L. Tang, M. C. Fitzgerald, *Analytical Chemistry* **2008**, *80*, 4175–4185.

[28] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, A Geometric Approach to Macromolecule-Ligand Interactions, **1982**.

[29] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: Methods and applications, **2004**.

[30] J. N. Chan, D. Vuckovic, L. Sleno, J. B. Olsen, O. Pogoutse, P. Havugimana, J. A. Hewel, N. Bajaj, Y. Wang, M. F. Musteata, C. Nislow, A. Emili, *Molecular and Cellular Proteomics* **2012**, *11*, DOI 10.1074/mcp.M111.016642.

[31] R. E. Moellering, B. F. Cravatt, How chemoproteomics can enable drug discovery and development, **2012**.

[32] E. O. Porta, P. G. Steel, Activity-based protein profiling: A graphical review, **2023**.

[33] C. M. Huwe, Synthetic library design, **2006**.

[34] G. L. Thomas, C. W. Johannes, Natural product-like synthetic libraries, **2011**.

[35] W. Sneader, *Drug Prototypes and Their Exploitation*, 1st, Wiley, **1996**.

[36] A. G. Atanasov, B. Waltenberger, E. M. Pferschy-Wenzig, T. Linder, C. Wawrosch, P. Uhrin, V. Temml, L. Wang, S. Schwaiger, E. H. Heiss, J. M. Rollinger, D. Schuster, J. M. Breuss, V. Bochkov, M. D. Mihovilovic, B. Kopp, R. Bauer, V. M. Dirsch, H. Stuppner, Discovery and resupply of pharmacologically active plant-derived natural products: A review, **2015**.

[37] A. L. Harvey, R. Edrada-Ebel, R. J. Quinn, *Nature Reviews Drug Discovery* **2015**, *14*, 111–129.

[38] B. Waltenberger, A. Mocan, K. Šmejkal, E. H. Heiss, A. G. Atanasov, Natural products to counteract the epidemic of cardiovascular and metabolic disorders, **2016**.

[39] T. Arif, J. D. Bhosale, N. Kumar, T. K. Mandal, R. S. Bendre, G. S. Lavekar, R. Dabur, Natural products - Antifungal agents derived from plants, **2009**.

[40] Y. Zhang, R. Loria, Y. Ding in Soil Science Society of America, **2015**, pp. 51–77.

[41] T. Hautbergue, E. L. Jamin, L. Debrauwer, O. Puel, I. P. Oswald, From genomics to metabolomics, moving toward an integrated strategy for the discovery of fungal secondary metabolites, **2018**.

[42] M. F. Traxler, R. Kolter, *Natural Product Reports* **2015**, *32*, 956–970.

[43] H. J. Shin, Natural products from marine fungi, **2020**.

[44] N. Chaachouay, L. Zidane, *Drugs and Drug Candidates* **2024**, *3*, 184–207.

[45] J. D. Roberts, M. C. Caserio, *Basic Principles of Organic Chemistry* **1977**, *1*, 1419–1459.

[46] R. A. Dixon, L. Achnine, P. Kota, C. J. Liu, M. S. Reddy, L. Wang, *Molecular Plant Pathology* **2002**, *3*, 371–390.

[47] D. O'Hagan, *Natural Product Reports* **1993**, *10*, 593.

[48] B. Shen, C. R. Hutchinson, *Science* **1993**, *262*, 1535–1540.

[49] A. L. Harvey, Natural products in drug discovery, **2008**.

[50] A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, I. E. Orhan, M. Banach, J. M. Rollinger, D. Barreca, W. Weckwerth, R. Bauer, E. A. Bayer, M. Majeed, A. Bishayee, V. Bochkov, G. K. Bonn, N. Braidy, F. Bucar, A. Cifuentes, G. D'Onofrio, M. Bodkin, M. Diederich, A. T. Dinkova-Kostova, T. Efferth, K. E. Bairi, N. Arkells, T. P. Fan, B. L. Fiebich, M. Freissmuth, M. I. Georgiev, S. Gibbons, K. M. Godfrey, C. W. Gruber, J. Heer, L. A. Huber, E. Ibanez, A. Kijjoa, A. K. Kiss, A. Lu, F. A. Macias, M. J. Miller, A. Mocan, R. Müller, F. Nicoletti, G. Perry, V. Pittalà, L. Rastrelli, M. Ristow, G. L. Russo, A. S. Silva, D. Schuster, H. Sheridan, K. Skalicka-Woźniak, L. Skaltsounis, E. Sobarzo-Sánchez, D. S. Bredt, H. Stuppner, A. Sureda, N. T. Tzvetkov, R. A. Vacca, B. B. Aggarwal, M. Battino, F. Giampieri, M. Wink, J. L. Wolfender, J. Xiao, A. W. K. Yeung, G. Lizard, M. A. Popp, M. Heinrich, I. Berindan-Neagoe, M. Stadler, M. Daglia, R. Verpoorte, C. T. Supuran, *Nature Reviews Drug Discovery* **2021**, *20*, 200–216.

[51] D. J. Newman, G. M. Cragg, *Journal of Natural Products* **2016**, *79*, 629–661.

[52] S. Nozoe, K. Kikuchi, N. Ishii, T. Ohta, *Tetrahedron Letters* **1992**, *33*, 7551–7552.

[53] S. Nozoe, N. Ishii, G. Kusano, K. Kikuchi, T. Ohta, *Tetrahedron Letters* **1992**, *33*, 7547–7550.

[54] L. Guan, H. Yang, Y. Cai, L. Sun, P. Di, W. Li, G. Liu, Y. Tang, *MedChemComm* **2019**, *10*, 148–157.

[55] S. Menant, F. Guégan, V. Tognetti, L. Merzoud, L. Joubert, H. Chermette, C. Morell, *Molecules* **2021**, *26*, DOI `10.3390/molecules26206218`.

[56] J. Zhang, K. V. Gulyás, J. Li, M. Ma, L. Zhou, L. Wu, R. Xiong, M. Erdelyi, W. Zhu, Z. Xu, *Computers in Biology and Medicine* **2024**, *172*, 108209.

[57] S. Roy, E. Colombo, R. Vinck, C. Mari, R. Rubbiani, M. Patra, G. Gasser, *ChemBioChem* **2020**, *21*, 2966–2973.

[58] C. Gentry, R. Egleton, T. Gillespie, T. Abbruscato, H. Bechowski, V. Hruby, T. Davis, *Peptides* **1999**, *20*, 1229–1238.

[59] J. T. Hammill, D. Bhasin, D. C. Scott, J. Min, Y. Chen, Y. Lu, L. Yang, H. S. Kim, M. C. Connelly, C. Hammill, G. Holbrook, C. Jeffries, B. Singh, B. A. Schulman, R. K. Guy, *Journal of Medicinal Chemistry* **2018**, *61*, 2694–2706.

[60] D. A. Smith, K. Beaumont, T. S. Maurer, L. Di, *Journal of Medicinal Chemistry* **2018**, *61*, 4273–4282.

[61] D. A. Smith, K. Beaumont, T. S. Maurer, L. Di, *Journal of Medicinal Chemistry* **2019**, *62*, 2245–2255.

[62] M. Hirose, M. Okaniwa, T. Miyazaki, T. Imada, T. Ohashi, Y. Tanaka, T. Arita, M. Yabuki, T. Kawamoto, S. Tsutsumi, A. Sumita, T. Takagi, B. C. Sang, J. Yano, K. Aertgeerts, S. Yoshida, T. Ishikawa, *Bioorganic and Medicinal Chemistry* **2012**, *20*, 5600–5615.

[63] Z. Shiokawa, K. Hashimoto, B. Saito, Y. Oguro, H. Sumi, M. Yabuki, M. Yoshimatsu, Y. Kosugi, Y. Debori, N. Morishita, D. R. Dougan, G. P. Snell, S. Yoshida, T. Ishikawa, *Bioorganic and Medicinal Chemistry* **2013**, *21*, 7938–7954.

[64] K. Urbahns, T. Yura, M. Mogi, M. Tajimi, H. Fujishima, T. Masuda, N. Yoshida, T. Moriwaki, T. B. Lowinger, H. Meier, F. Chan, D. Madge, J. B. Gupta, *Bioorganic and Medicinal Chemistry Letters* **2011**, *21*, 3354–3357.

[65] T. X. Carroll, T. D. Thomas, H. Bergersen, K. J. Børve, L. J. Sæthre, *Journal of Organic Chemistry* **2006**, *71*, 1961–1968.

[66] A. B. Dembinski, L. R. Johnson, G. A. Castro, *American Journal of Physiology - Regulatory Integrative and Comparative Physiology* **1979**, *6*, 1207–1216.

[67] P. Pacak, *Chemical Papers* **1989**, *43*, 489–500.

[68] D. Chiodi, Y. Ishihara, *Journal of Medicinal Chemistry* **2023**, *66*, 5305–5331.

[69] D. O'hagan, *Chemical Society Reviews* **2008**, *37*, 308–319.

[70] C. K. Miller, B. J. Orr, J. F. Ward, *The Journal of Chemical Physics* **1977**, *67*, 2109–2118.

[71] D. Haase, *Chemical Physics* **1992**, *161*, 403–418.

[72] A. G. Ryazanov, M. D. Ward, C. E. Mendola, K. S. Pavur, M. V. Dorovkov, M. Wiedmann, H. Erdjument-Bromage, P. Tempst, T. G. Parmer, C. R. Prostko, F. J. Germino, W. N. Hait, Identification of a new class of protein kinases represented by eukaryotic elongation factor-2 kinase, **1997**.

[73] P. Chankhamjon, Y. Tsunematsu, M. Ishida-Ito, Y. Sasa, F. Meyer, D. Boettger-Schmidt, B. Urbansky, K. D. Menzel, K. Scherlach, K. Watanabe, C. Hertweck, *Angewandte Chemie - International Edition* **2016**, *55*, 11955–11959.

[74] F. Baneyx, M. Mujacic, *Nature Biotechnology* **2004**, *22*, 1399–1407.

[75] U. Rinas, J. E. Bailey, *Applied Microbiology and Biotechnology* **1992**, *37*, 609–614.

[76] U. Rinas, J. E. Bailey, *Applied and Environmental Microbiology* **1993**, *59*, 561–566.

[77] X. Qi, Y. Sun, S. Xiong, *Microbial Cell Factories* **2015**, *14*, 1–12.

[78] J. C. Spurlino, G. Y. Lu, F. A. Quiocho, *Journal of Biological Chemistry* **1991**, *266*, 5202–5219.

[79] P. A. Colussi, C. H. Taron, *Applied and Environmental Microbiology* **2005**, *71*, 7092–7098.

[80] J. D. Platko, M. Deeg, V. Thompson, Z. Al-Hinai, H. Glick, K. Pontius, P. Colussi, C. Taron, D. L. Kaplan, *Protein Expression and Purification* **2008**, *57*, 57–62.

[81] W. G. Bao, H. Fukuhara, *Gene* **2001**, *272*, 103–110.

[82] W. R. Pearson, *Current Protocols in Bioinformatics* **2013**, DOI `10.1002/0471250953.bi0301s42`.

[83] F. Todone, M. A. Vanoni, A. Mozzarelli, M. Bolognesi, A. Coda, B. Curti, A. Mattevi, Active Site Plasticity in D-Amino Acid Oxidase: A Crystallographic Analysis †, ‡, **1997**.

[84] A. M. Hassell, G. An, R. K. Bledsoe, J. M. Bynum, H. L. Carter, S. J. J. Deng, R. T. Gampe, T. E. Grisard, K. P. Madauss, R. T. Nolte, W. J. Rocque, L. Wang, K. L. Weaver, S. P. Williams, G. B. Wisely, R. Xu, L. M. Shewchuk, *Acta Crystallographica Section D: Biological Crystallography* **2006**, *63*, 72–79.

[85] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C. C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, *Nature* **2024**, *630*, 493–500.

[86] C. Dong, S. Flecks, S. Unversucht, C. Haupt, K.-H. van Pée, J. H. Naismith, *Science* **2005**, *309*, 2216–2219.

[87] E. Bitto, Y. Huang, C. A. Bingman, S. Singh, J. S. Thorson, G. N. Phillips, *Proteins: Structure Function and Genetics* **2008**, *70*, 289–293.

[88] B. Nowak-Thompson, N. Chaney, J. S. Wing, S. J. Gould, J. E. Loper, *Journal of Bacteriology* **1999**, *181*, 2166–2174.

[89] L. Yin, *Acta Crystallographica Section F Structural Biology Communications* **2022**, *78*, 270–275.

[90] J. Gong, Y. Chen, F. Pu, P. Sun, F. He, L. Zhang, Y. Li, Z. Ma, H. Wang, *Current Drug Targets* **2019**, *20*, 551–564.

[91] J. W. Robertson, J. J. Kasianowicz, S. Banerjee, Analytical approaches for studying transporters, channels and porins, **2012**.

[92] L. C. Ray, D. Das, S. Entova, V. Lukose, A. J. Lynch, B. Imperiali, K. N. Allen, *Nature Chemical Biology* **2018**, *14*, 528–541.

[93] G. Blobel, Intracellular protein topogenesis (protein translocation across membranes/protein integration into membranes/posttranslocational sorting/topogenic sequences/ phylogeny of membranes and compartments), **1980**.

[94] M. Jelokhani-Niaraki, *International Journal of Molecular Sciences* **2022**, *24*, 468.

[95] J. Bockaert, *The EMBO Journal* **1999**, *18*, 1723–1729.

[96] M. D. Pizzagalli, A. Bensimon, G. Superti-Furga, A guide to plasma membrane solute carrier proteins, **2021**.

[97] J. P. Overington, B. Al-Lazikani, A. L. Hopkins, *Nature Reviews Drug Discovery* **2006**, *5*, 993–996.

[98] S. M. Smith in *Vol. 1485*, Humana Press Inc., **2017**, pp. 389–400.

[99] V. A. Bjørnestad, M. Orwick-Rydmark, R. Lund, *Langmuir* **2021**, *37*, 6178–6188.

[100] A. A. Smith, H. E. Autzen, B. Faust, J. L. Mann, B. W. Muir, S. Howard, A. Postma, A. J. Spakowitz, Y. Cheng, E. A. Appel, *Chem* **2020**, *6*, 2782–2795.

[101] L. E. Ball, L. J. Riley, W. Hadasha, R. Pfukwa, C. J. Smith, T. R. Dafforn, B. Klumperman, *Biomacromolecules* **2021**, *22*, 763–772.

[102] M. Linial, K. Miller, R. H. Scheller, *Neuron* **1989**, *2*, 1265–1273.

[103] M. LINIAL, *European Journal of Biochemistry* **1993**, *216*, 189–197.

[104] K. Hayess, R. Kraft, J. Sachsinger, J. Janke, G. Beckmann, K. Rohde, B. Jandrig, R. Benndorf, *Journal of cellular biochemistry* **1998**, *69*, 304–15.

[105] C. Loeb-Hennard, X. Cousin, I. Prengel, E. Kremmer, *Gene Expression Patterns* **2004**, *5*, 91–96.

[106] R. Kojima, T. Endo, Y. Tamura, *Scientific Reports* **2016**, *6*, DOI 10.1038/srep30777.

[107] J. Koch, J. Foekens, M. Timmermans, W. Fink, A. Wirzbach, M. D. Kramer, B. M. Schaefer, *Archives of Dermatological Research* **2003**, *295*, 203–210.

[108] O. Levius, M. Linial, VAT-1 from Torpedo Synaptic Vesicles Is a Calcium Binding Protein: A Study in Bacterial Expression Systems, **1993**.

[109] F. Mori, K. Tanigawa, K. Endo, K. Minamiguchi, M. Abe, S. Yamada, K. Miyoshi, *Prostate* **2011**, *71*, 1579–1586.

[110] S. Mertsch, M. Becker, A. Lichota, W. Paulus, V. Senner, *Neuropathology and Applied Neurobiology* **2009**, *35*, 342–352.

[111] P. Yang, K. Wang, C. Zhang, Z. Wang, Q. Liu, J. Wang, T. Jiang, X. Shan, *Cancer Immunology Immunotherapy* **2021**, *70*, 2589–2600.

[112] B. Persson, J. S. Zigler, H. Jornvall, *European Journal of Biochemistry* **1994**, *226*, 15–22.

[113] M. Linial, O. Levius, *FEBS Letters* **1993**, *315*, 91–94.

[114] S. Y. Kim, T. Mori, M. F. Chek, S. Furuya, K. Matsumoto, T. Yajima, T. Ogura, T. Hakoshima, *Scientific Reports* **2021**, *11*, DOI 10.1038/s41598-021-81409-y.

[115] Y. Watanabe, T. Endo, *Journal of Biological Chemistry* **2020**, *295*, 3257–3268.

[116] A. Ikemizu, D. Hatta, K. Fujimoto, M. Honda, K. Watanabe, K. Ohyama, N. Kuroda, T. Tanaka, K. Shirotani, N. Iwata, *Biological and Pharmaceutical Bulletin* **2024**, *47*, b23–00830.

[117] B. Wienen-Schmidt, M. Oebbeke, K. Ngo, A. Heine, G. Klebe, *ChemMedChem* **2021**, *16*, 292–300.

[118] I. Usón, G. M. Sheldrick, *Current Opinion in Structural Biology* **1999**, *9*, 643–648.

[119] P. Evans, A. McCoy in Acta Crystallographica Section D: Biological Crystallography, *Vol. 64*, **2007**, pp. 1–10.

[120] A. J. McCoy, M. D. Sammito, R. J. Read, Implications of AlphaFold2 for crystallographic phasing by molecular replacement, **2022**.

[121] R. J. Read, S. G. Urzhumtsev, V. Y. Lunin, *Advancing Methods for Biomolecular Crystallography*, (Eds.: R. Read, A. G. Urzhumtsev, V. Y. Lunin), Springer Netherlands, **2013**.

[122] A. L. Patterson, *Physical Review* **1934**, *46*, 372–376.

[123] G. J. Kleywegt, T. A. Jones, [1 1] Model Building and Refinement Practice, **1997**.

[124] P. A. Karplus, K. Diederichs, *Science* **2012**, *336*, 1030–1033.

[125] R. Percudani, A. Pavesi, S. Ottonello, *Journal of Molecular Biology* **1997**, *268*, 322–330.

[126] T. Tuller, Y. Y. Waldman, M. Kupiec, E. Ruppin, *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 3645–3650.

[127] M. Zhou, T. Wang, J. Fu, G. Xiao, Y. Liu, *Molecular Microbiology* **2015**, *97*, 974–987.

[128] J. F. Kane, *Current Opinion in Biotechnology* **1995**, *6*, 494–500.

[129] O. Kandror, A. L. Goldberg, *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 4978–4981.

[130] M. J. Lelivelt, T. H. Kawula, *Journal of Bacteriology* **1995**, *177*, 4900–4907.

[131] P. Graumann, M. A. Marahiel, *Archives of Microbiology* **1996**, *166*, 293–300.

[132] G. Blaha, D. N. Wilson, G. Stoller, G. Fischer, R. Willumeit, K. H. Nierhaus, *Journal of Molecular Biology* **2003**, *326*, 887–897.

[133] S. Phadtare, *Current Issues in Molecular Biology* **2004**, *6*, 125–136.

[134] R. B. Kapust, D. S. Waugh, *Protein Science* **1999**, *8*, 1668–1674.

[135] D. B. Smith, K. S. Johnson, *Gene* **1988**, *67*, 31–40.

[136] D. B. Smith, *Methods in Enzymology* **2000**, *326*, 254–270.

[137] M. Pouresmaeil, S. Azizi-Dargahlou, *Archives of Microbiology* **2023**, *205*, DOI 10.1007/s00203-023-03541-9.

[138] A. B. Alber, D. M. Suter, Dynamics of protein synthesis and degradation through the cell cycle, **2019**.

[139] M. Losen, B. Frölich, M. Pohl, J. Büchs, *Biotechnology Progress* **2004**, *20*, 1062–1068.

[140] F. W. Studier, *Protein expression and purification* **2005**, *41*, 207–234.

[141] C.-j. Huang, H.-L. Peng, A. K. Patel, R. R. Singhania, C.-D. Dong, C.-y. Cheng, *Catalysts* **2021**, *11*, 13.

[142] T. San-Miguel, P. Pérez-Bermúdez, I. Gavidia, *SpringerPlus* **2013**, *2*, 1–4.

[143] T. Taylor, J.-P. Denson, D. Esposito in **2017**, pp. 65–82.

[144] Y. Shirano, D. Shibata, *FEBS Letters* **1990**, *271*, 128–130.

[145] I. Kataeva, J. Chang, H. Xu, C. H. Luan, J. Zhou, V. N. Uversky, D. Lin, P. Horanyi, Z. J. Liu, L. G. Ljungdahl, J. Rose, M. Luo, B. C. Wang, *Journal of Proteome Research* **2005**, *4*, 1942–1951.

[146] F. Volontè, F. Marinelli, L. Gastaldo, S. Sacchi, M. S. Pilone, L. Pollegioni, G. Molla, *Protein Expression and Purification* **2008**, *61*, 131–137.

[147] A. Piserchio, R. Ghose, D. Cowburn, *Journal of Biomolecular NMR* **2009**, *44*, 87–93.

[148] A. Vera, N. González-Montalbán, A. Arís, A. Villaverde, *Biotechnology and Bioengineering* **2007**, *96*, 1101–1106.

[149] M. D. Rolfe, C. J. Rice, S. Lucchini, C. Pin, A. Thompson, A. D. Cameron, M. Alston, M. F. Stringer, R. P. Betts, J. Baranyi, M. W. Peck, J. C. Hinton, *Journal of Bacteriology* **2012**, *194*, 686–701.

[150] D. Horne, A. Tomasz, *Antimicrobial Agents and Chemotherapy* **1977**, *11*, 888–896.

[151] D. Horne, A. Tomasz, *Antimicrobial Agents and Chemotherapy* **1980**, *17*, 235–246.

[152] G. Minasov, X. Wang, B. K. Shoichet, *Journal of the American Chemical Society* **2002**, *124*, 5333–5340.

[153] E. Audain, Y. Ramos, H. Hermjakob, D. R. Flower, Y. Perez-Riverol, *Bioinformatics* **2016**, *32*, 821–827.

[154] G. R. Grimsley, J. M. Scholtz, C. N. Pace, *Protein Science* **2009**, *18*, 247–251.

[155] A. Poetsch, H. Seelert, J. M. Z. Tittingdorf, N. A. Dencher, *Biochemical and Biophysical Research Communications* **1999**, *265*, 520–524.

[156] G. W. Welling, Y. Hiemstra, M. Feijlbrief, C. Örvell, J. van Ede, S. Welling-Wester, *Journal of Chromatography A* **1992**, *599*, 157–162.

[157] D. Frishman, P. Argos, Knowledge-Based Protein Secondary Structure Assignment, **1995**.

[158] M. Heinig, D. Frishman, *Nucleic Acids Research* **2004**, *32*, W500–W502.

[159] N. J. Hu, S. Iwata, A. D. Cameron, D. Drew, *Nature* **2011**, *478*, 408–411.

[160] B. Heras, M. A. Edeling, K. A. Byriel, A. Jones, S. Raina, J. L. Martin, *Structure* **2003**, *11*, 139–145.

List of Figures

List of Tables

Acknowledgments

The journey toward this dissertation has been long and not always easy. I am deeply grateful to those who supported me along the way:

- Dr. Sabine Schneider, for giving me the opportunity to carry out my thesis in her lab, for her unwavering support throughout my research, and for the many insightful group meetings.
- Josef Braun and Tri Manh Phi, for the excellent collaborations, creating a positive atmosphere, and engaging in helpful discussions.
- The research groups on the ICE-M ground floor, for the cocktail evenings, lunch breaks, and coffee chats that brightened many of my days.
- My colleagues in the Schneider group over the years — Leo, Christopher, Philipp, Marie, Olga, Corinna, Özge, Dong, Ken, and Bennett — you have been a constant source of fun, support, and inspiration, making my time here truly memorable.
- My students, Johannes, Rebecca, and Edin, for their dedication and contributions to this thesis.
- Our HiWis, Laura, Alex, and Andi, for their invaluable assistance with everyday lab work.
- Aarthi and Tulika from the Huc group, for their support with crystal measurements.
- The FSI Chemie und Biochemie, for the fun times and support during my years at LMU.
- The members of my pen & paper groups—Joe, Maxi, Matte, Caro, Ksenija, Joni, and Philipp — you provided much-needed moments of escape from work and science.
- My parents, for their unconditional love and support, without which I would not have come this far.

- My best friend Laura, for her unwavering friendship, constant support, and always knowing how to lift my spirits.