Charles University

Faculty of Science

Study programme: Cell Biology

Bc. Adéla Fejfarová

Design, preparation and structural studies of biologically relevant protein variants of cancer-related Carbonic Anhydrase IX

Návrh, příprava a strukturní studie biologicky relevantních variant enzymu karbonát lyáza IX

Type of thesis:

Diploma thesis

Supervisor: doc. RNDr. Pavlína Maloy Řezáčová, Ph.D.

Prague, 2024

Declaration

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 29. 4. 2024

Bc. Adéla Fejfarová

Acknowledgment

This work was elaborated in the Laboratory of Structural Biology at the Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences.

Diffraction data have been collected on BL14.2 at the BESSY II electron storage ring operated by the Helmholtz-Zentrum Berlin (Mueller et al., 2015). We would particularly like to acknowledge the help and support of Tatjana Barthel during the experiment.

This work was supported by the project National Institute for Cancer Research (Programme EXCELES, ID Project No. LX22NPO5102) - Funded by the European Union - Next Generation EU.

First, let me express my deepest gratitude to my supervisor Pavlína Maloy Řezáčová for her guidance through this project, for countless advice about experiments and their results, and most importantly for giving me all the opportunities to explore the world of science.

I would love to thank Irena Sieglová for introducing me to the protein biochemistry and for being there for me since the beginning, both personally and professionally. I wish to thank Tereza Vučková for teaching me to try find an answer before I go and ask her, and for all the kind responses to my questions because I ask her anyway.

My sincere thanks belong to Milan Fábry for the preparation of plasmid vectors and for the lovely poem recitations that always put smile on my face. I wish to thank Jiří Brynda and Petr Pachl for all the help with protein crystallography and for making me realize that the crystals must be mounted and measured anyway, and I do not have to be scared to do so. I thank Pavel Srb for the aid with NMR spectroscopy experiments and for taking a month-old master student on her first scientific expedition, which I genuinely enjoyed. I thank Václav Veverka for reassuring me that I still belong to cell biology and keeping me connected to this faculty.

My thanks also go to everyone from the Laboratory of Structural Biology at IOCB, Prague. You truly make me look forward to any other day full of chats about anything and everything. I am so grateful I have found friends within this group of amazing people.

My heartfelt thanks to Jenda, I know I make you go mad being mad myself.

Finally let me say few words in Czech. Babičko, dědečku, mami, děkuju vám za to, kým jsem. Tahle práce je pro vás, za to všechno, co jste pro mě kdy udělali.

Abstract

Carbonic anhydrase IX (CA IX) represents an attractive target for the development of anticancer drugs as it is overexpressed in various types of solid tumors. By its catalytic activity, CA IX assists the cancer cells to maintain the optimal intracellular pH and to acidify the extracellular milieu promoting tumor development. There are twelve enzymatically active carbonic anhydrases (CAs) present in human body, all sharing a high sequence identity and a typical β-sheet structural fold of the well-studied catalytic domain. Although, the activity of CAs is efficiently inhibited by sulfonamide-containing compounds, the design of inhibitor selective to the cancer-related CA IX has been hampered by the high sequence conservation. CA IX has several unique features compared to other members of the family, which investigation may help in the development of selective drug compounds. Namely, it is a type I transmembrane dimeric protein with unique N-terminal proteoglycan-like (PG) domain, extracellular catalytic domain, and short cytoplasmic C-terminal segment.

The above denoted traits make CA IX subject of structure-based drug design efforts. However, the expression and purification in high yield as well as crystallization experiments has been challenging. Therefore, protein variant bearing six amino acid substitutions for recombinant expression in *E. coli* has been published before. Nevertheless, some of the amino acid substitutions hinders the formation of the biologically relevant dimeric arrangement of CA IX. For that reason, three novel protein variants were designed and structurally characterized within this diploma thesis. The CA IX variants were expressed, purified, and characterized for their oligomeric state in solution, and confirmed to form dimers. These variants were crystallized and subjected for X-ray diffraction experiment which ultimately led to determination of three protein structures. Crystal structures revealed that when the wild-type amino acids are reintroduced to specific residues, the biological dimeric arrangement is restored.

Key words

Cancer, tumor, carbonic anhydrase, structural study, X-ray crystallography, NMR spectroscopy

Abstrakt

Lidská karbonát lyáza IX (CA IX) je enzym nadměrně produkovaný v různých typech nádorů, díky čemuž představuje atraktivní cíl pro vývoj protinádorových léčiv. Nádorovým buňkám tento enzym pomáhá udržovat optimální intracelulární pH a okyselováním vnějšího prostředí podporuje růst nádoru. V lidském těle existuje 12 enzymaticky aktivních izoforem karbonát lyáz, které sdílejí sekvenčně a strukturně konzervovanou katalytickou doménu tvořenou převážně β-listy. Přestože je aktivita těchto enzymů účinně inhibovaná látkami obsahující sulfonamidovou funkční skupinu, vysoká sekvenční konzervovanost katalytické domény znesnadňuje návrh inhibitorů selektivních pro CA IX. K vývoji selektivních léčiv by mohl pomoc výzkum specifických vlastností CA IX. CA IX je transmembránový protein typu I tvořící dimer a obsahující unikátní N-koncovou „proteoglycan-like" doménu, mimobuněčnou katalytickou doménu a krátký cytoplazmatický C-koncovým segment.

Z výše popsaných důvodů je karbonát lyáza IX předmětem pro strukturou-inspirovaný návrh léčiv. Avšak exprese a purifikace tohoto proteinu ve vysokých výtěžcích, stejně jako jeho krystalizace se ukázala být komplikovaná. Z tohoto důvodu byla připravena varianta CA IX obsahující šest aminokyselinových záměn pro rekombinantní expresi v *E. coli*. Některé z těchto šesti aminokyselinových záměn ale zamezují tvorbě biologického dimerního uspořádání. Proto byly v této práci navrženy a strukturně popsány tři nové varianty karbonát lyázy IX. Tyto proteinové varianty byly exprimovány, purifikovány a charakterizovány pro tvorbu dimeru v roztoku. Tyto varianty byly vykrystalizované a podrobené rentgenové strukturní analýze, díky které byly vyřešené celkem tři proteinové struktury. Krystalové struktury ukázaly, že obnovení biologického dimerního uspořádání lze dosáhnout zavedením původních aminokyselinových zbytků do specifických pozic.

Klíčová slova

Rakovina, nádor, karbonát lyáza, strukturní studie, RTG krystalografie, NMR spektroskopie

List of Abbreviations

AA	Acrylamide
AE	Anion Exchanger
AEC	Anion Exchange Chromatography
AR1	Affinity Regeneration solution 1
AR2	Affinity Regeneration solution 2
AU	Asymmetric Unit
BSA	Bovine Serum Albumin
CA	Carbonic Anhydrase
CA IX	Carbonic Anhydrase IX
CARP	Carbonic Anhydrase Related Proteins
CV	Column Volume
DMSO	Dimethyl Sulfoxide
DSF	Differential Scanning Fluorimetry
EB	Elution Buffer
ECM	Extracellular Matrix
ECR	Extracellular Region
EDTA	Ethylenediaminetetraacetic acid
EGF	Epithelial Growth Factor
EMT	Epithelial-Mesenchymal Transition
ETG	Ethyl-β-thiogalactopyranoside
FPLC	Fast Protein Liquid Chromatography
FT	Flow-through fraction
GPI	Glycosylphosphatidylinositol
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGF	Hepatocyte Growth Factor
HIF-1	Hypoxia-Inducible Factor-1
HRE	Hypoxia-Response Element
HSQC	Heteronuclear Single Quantum Coherence
HZB	Helmholtz-Zentrum Berlin
IC	Intracellular tail
IDR	Intrinsically Disordered Region
IN	Input
MCT	Monocarboxylate Transporter

MES	2-(N-morpholino)ethanesulfonic acid
MM	Minimal medium
MMP	Matrix Metalloproteases
MW	Molecular Weight
NBC	Sodium Bicarbonate Cotransporter
NMR	Nuclear Magnetic Resonance
PAGE	Polyacrylamide
PBS	Phosphate-Buffered Saline
PEG	Polyethylen glycol
PG domain	Proteoglycan-like domain
PHD	Prolyl Hydroxylase Domains
PKA	Protein Kinase A
PTM	Post-Translational Modification
SAC	Sulfonamide-Affinity Chromatography
SAXS	Small-Angle X-ray Scattering
SDS	Sodium Dodecyl Sulfate
SEC	Size Exclusion Chromatography
SP	Signaling Peptide
TCEP	Tris(2-carboxyethyl)phosphine
TEOA	Triethanolamine
TM	Transmembrane helix
TMB	3,3',5,5'-tetramethylbenzidine
Tris	Tris(hydrozymethyl)aminomethan
VHL	von-Hippel-Lindau-tumor suppressor
WB1	Washing Buffer 1
WB2	Washing Buffer 2

Table of Contents

1 Introduction and aims

During tumor development cancer cells express various proteins to help them overcome unfavorable conditions within the tumor environment. One of these proteins is enzyme carbonic anhydrase IX (CA IX) which is overexpressed on cellular surface in breast, bladder, colon, oropharyngeal and renal carcinomas. In cooperation with plasma membrane transporters and other proteins, CA IX maintain intracellular pH within physiological range and simultaneously acidify extracellular space which favors tumor dissemination within the body. This task is utilized mainly by CA IX enzymatic activity of carbon dioxide hydration to bicarbonate. The enzymatic activity is effectively inhibited by numerous molecules bearing sulfonamide functional group. However, usage of such inhibitors in antitumor therapy has been hindered by existence of other fourteen CA isoforms within human body. Therefore, CA IX has been subjected in structure-based drug design research to identify potential drug molecules that would be selective for only this cancer-related carbonic anhydrase.

During the structural studies recombinant proteins are typically produced in various expression systems. In case of CA IX, expression and purification protocols in eukaryotic or bacterial systems have been developed for number of protein variants bearing different amino acid substitutions. Introduction of amino acid substitutions may decrease the biological relevance as the protein differs from the wild type. This diploma project aimed to prepare and in detail characterize number of CA IX protein variants that would remain their biological relevance but still be suitable for high-yield production necessary for structural studies. The specific aims of this diploma thesis are as follows:

- Design of CA IX protein variants with amino acid substitutions
- Production of recombinant CA IX protein variants
    - Heterologous expression in *Escherichia coli*
    - Purification by optimized protocol
- Characterization of oligomeric state of CA IX proteins in solution
- Crystallization of recombinant CA IX proteins
- Structure determination of CA IX protein variants by X-ray crystallography
- Structural analysis focusing on dimeric interface

2 Literature review

2.1 Family of human carbonic anhydrases

The carbonic anhydrases (CAs) represent ubiquitous enzymes which are found across all domains of life. There are eight evolutionary distinct families of genes that encode these proteins in archaeal, bacterial, and eukaryotic organisms. All CAs found in human, as in any other mammal, belong to the α-CA family. Each of the fifteen human isoforms is expressed from discrete gene and has different expression levels, molecular and enzymatic properties, subcellular localization, as well as tissue and organ distribution. Hence, they participate in various physiological processes such as respiration, pH regulation, electrolyte secretion and ion transport, biosynthetic reactions (gluconeogenesis, lipogenesis and ureagenesis), bone resorption and calcification, taste perception or gastric acid secretion (Aspatwar *et al*, 2022).

Not only that CAs are crucial for the physiological functions, almost all of them has been connected to some diseases (Table 1). Thus, they are recognized as therapeutic targets and modulation of their activity has been explored for treatment or management of numerous pathophysiological conditions. The research of CAs roles and their modulation in diseases has been reviewed in detail elsewhere (Mishra *et al*, 2020; Supuran, 2023).

2.1.1 Tissue distribution and cellular localization

Since most of the cell types express some isoform, CAs are distributed in diverse organs and tissues in the human body. However, each isoform gene has its specific expression pattern, which can be either very strict within only one or two tissue types (for example *CA5A*, *CA6*, *CA7*, *CA9*), or more widespread in several tissues (for example *CA2*, *CA5B*, *CA3*, *CA12*). All CA isoforms, together with their expression pattern in the human body, are specified in Table 1. Human CA proteins are located in various cellular compartments, from cytosol (CA I, CA II, CA III, CA VII, CA XIII, CA VIII), over mitochondria (CA VA, CA VB), associated with cell membrane (CA IV, CA IX, CA XII, CA XIV), up to secreted forms (CA VI, CA X, CA XI) (Aspatwar *et al*, 2022). Out of the fifteen isoforms, three have lost their enzymatic capability (CA VIII, CA X, CA XI), and are therefore referred to as carbonic anhydrase related proteins (CARP). The subcellular localization of enzymatically active CAs and CARPs is schematically illustrated in Figure 1.

The extracellular location of plasma membrane-associated CAs is mediated either by glycosylphosphatidylinositol (GPI) anchor in the case of CA IV (Zhu & Sly, 1990; Stams *et al*, 1996), or by single-pass helix of transmembrane isoforms CA IX (Pastoreková *et al*, 1992; Opavský *et al*, 1996; Alterio *et al*, 2009), CA XII (Türeci *et al*, 1998; Whittington *et al*, 2001) and CA XIV (Fujikawa-Adachi *et al*, 1999; Alterio *et al*, 2014). The transmembrane helix connects the enzymatic N-terminal region situated outside the cell with short C-terminal intracellular sequence.

Table 1: Carbonic anhydrase isoforms expressed in human body, their tissue distribution and list of diseases they are associated with. The summarized data are based on evidence reviewed in (Mishra et al*, 2020; Aspatwar* et al*, 2022).*

Isoform	Gene	Tissue expression	Disease
CA I	*CA1*	colon, erythrocytes, eye	retinal/cerebral edema, hemolytic anemia
CA II	*CA2*	stomach, colon, kidney, erythrocytes, bone osteoclasts, lung, testis, brain	renal tubular acidosis, osteopetrosis, cerebral calcification, glaucoma, edema, epilepsy
CA III	*CA3*	epididymis, skeletal muscle, adipocytes	oxidative stress
CA IV	*CA4*	colon, cerebellum, kidney, lung, pancreas, brain capillaries, heart muscle, eye	retinis pigmentosa, glaucoma, stroke
CA VA	*CA5A*	liver	obesity, hyperammonaemia, hyperlactatemia
CA VB	*CA5B*	testis, ovary, vagina, heart and skeletal muscle, pancreas, kidney, spinal cord	
CA VI	*CA6*	skin, salivary and mammary gland	dental caries
CA VII	*CA7*	colon, CNS	epilepsy
CA VIII	*CA8*	cerebellum	cerebellar ataxia, cerebellar atrophy, cognitive defect
CA IX	*CA9*	stomach, pancreas	cancer
CA X	*CA10*	cerebellum	
CA XI	*CA11*	cerebellum, ovary	
CA XII	*CA12*	kidney, intestine, reproductive epithelia, eye, CNS	cancer, glaucoma
CA XIII	*CA13*	colon, liver, kidney, brain, lung, reproductive system	sterility
CA XIV	*CA14*	spinal cord, kidney, brain, liver, eye	epilepsy, retinopathy

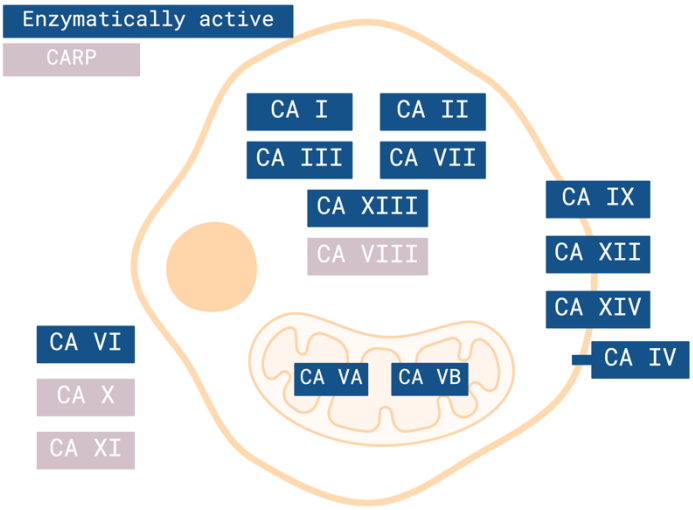


Figure 1: The subcellular location of all human CAs and carbonic anhydrase related proteins (CARPs). CA I, CA II, CA III, CA VII, CA XIII, CA VIII are confined in cytosol, CA VA and CA VB are situated in mitochondria, CA IV, CA IX, CA XII, CA XIV are associated with cell membrane, either by GPI-anchor (CA IV) or as single-pass transmembrane proteins (CA IX, CA XII, CA XIV), one catalytically active isoform (CA VI) is secreted similarly as two CARPs (CA X and CA XI). The figure was created in Biorender.com based on evidence reviewed in (Aspatwar et al*, 2022).*

2.1.2 Enzymatic function and active site

Carbonic anhydrases are metalloenzymes (EC 4.2.1.1.) and catalyze the reversible hydration of carbon dioxide to bicarbonate mediated by Zn^{2+} ion coordinated in the active site. The reaction is conducted by two-step ping-pong mechanism (Equation 1). While the first step, the bicarbonate is formed by nucleophilic attack of zinc-bound hydroxide [1.1]. Whereas during the second step the active site must be regenerated, which is the rate-limiting step of the whole catalysis [1.2] (Silverman & Lindskog, 1988).

$$E\mathrm{Zn}^{2+}\text{–}\mathrm{OH}^- + \mathrm{CO_2} \rightleftharpoons E\mathrm{Zn}^{2+}\text{–}\mathrm{HCO_3}^- \overset{\mathrm{H_2O}}{\rightleftharpoons} E\mathrm{Zn}^{2+}\text{–}\mathrm{H_2O} + \mathrm{HCO_3}^- \qquad [1.1]$$

$$E\mathrm{Zn}^{2+}\text{–}\mathrm{H_2O} + B \rightleftharpoons E\mathrm{Zn}^{2+}\text{–}\mathrm{OH}^- + B\mathrm{H}^+ \qquad [1.2]$$

Equation 1: Enzymatic reaction catalyzed by carbonic anhydrases, where E *stands for enzyme and* B *for buffer.*

Despite the rate-limiting regeneration, CAs are considered the fastest enzymes, with the catalysis efficiency close to the limits of diffusion. The high efficiency of the enzymatic reaction is given by well-organized architecture of the catalytic domain with the active site. The common catalytic domain is characterized as globular domain consisting of central 10-stranded antiparallel β-sheets core with additional surface loops and helices (Figure 2A). The active site positioned at bottom of a cavity constitutes of the zinc ion tetrahedrally coordinated by three histidine residues, which are conserved in the enzymatically active CAs (for example in CA II: H94, H96, H119, and CA IX: H226, H228, H251) (Figure 2B).

To further enhance the catalysis, the large cavity spanning from the protein surface comprises two clefts with opposing features. The hydrophobic site allows entry of CO_2 to the active site. On the other hand, the hydrophilic residues enable bicarbonate exit and are responsible for forming ordered water network for proton shuttling (Figure 2C). This is crucial for the active site regeneration when zinc-bound water must be deprotonated to form hydroxide. In the two most active isoforms, CA II and CA IX, the proton transfer from the bottom of the active site to the bulk solvent is facilitated by histidine residue H64 and H200, respectively (Alterio *et al*, 2009; Silverman & McKenna, 2007; Maupin *et al*, 2009) (Figure 2B). Overall, the enzymatically active isoforms share high sequence identity together with structural conservation of the catalytic domain with some residue variations especially in the loops on the surface (Pinard *et al*, 2015). An example of CAs structural similarity is given for CA II and CA IX in Figure 2D.

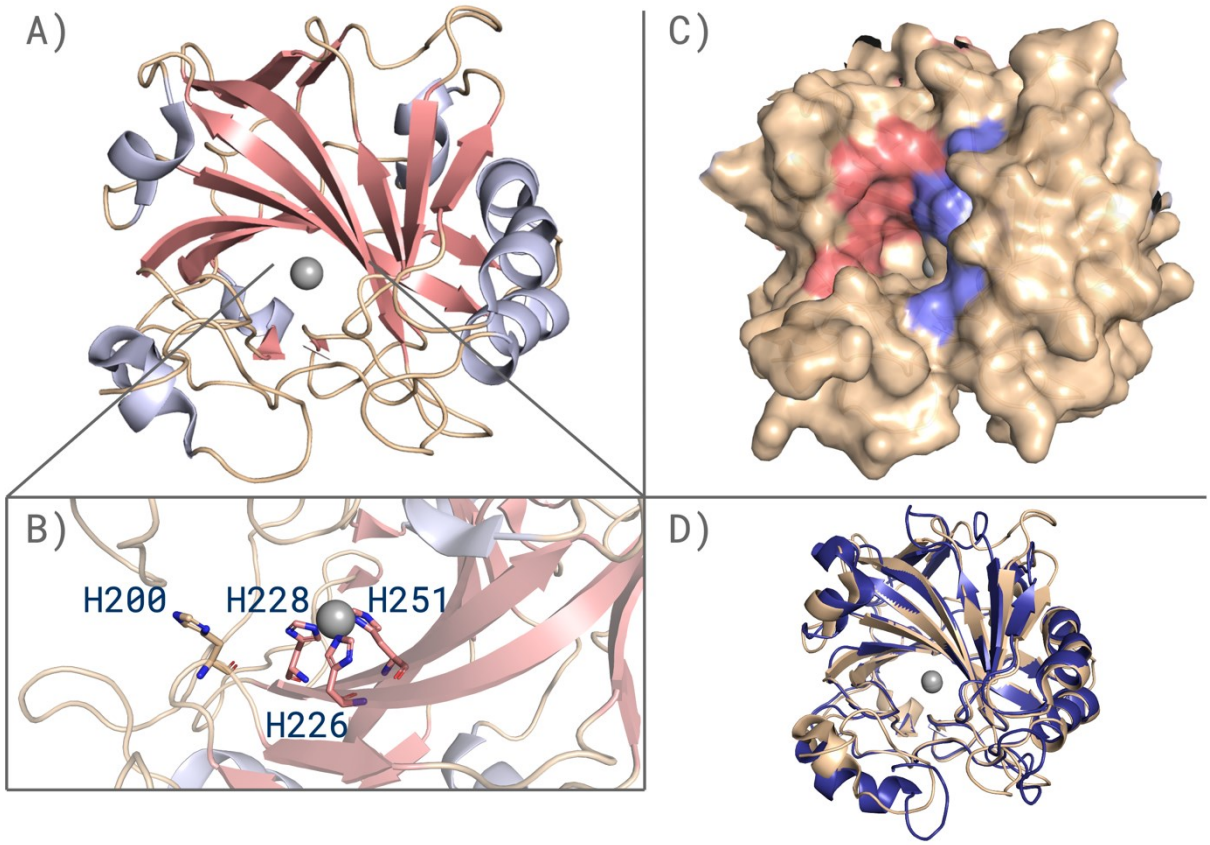


Figure 2: Some structural features common for human CAs. A) The core of the globular catalytic domain (shown as cartoon) is composed of 10 antiparallel β-strands (in pink) surrounded by several short helices (in light blue), all connected by numerous loops (in beige). B) Detail of the active site at the bottom of the cavity, three histidine residues which coordinate the Zn^{2+} ion are depicted as sticks and labelled according to the residue numbering in CA IX. The fourth histidine residue (in beige) is responsible for the proton transfer from the active site in the most active isoenzymes (H200 in CA IX as shown in here, and H64 in CA II). C) Surface visualization of CA IX demonstrating the hydrophobic (in pink) and hydrophilic (in violet) clefts. D) Structural comparison by superposition of CA II (in blue) and CA IX (in beige) catalytic domain (182 Cα atoms were aligned, RMSD 0.79 Å). Each figure was made in PyMOL software and assembled in Biorender.com. Figures A-D were made with PDB entry 3IAI of CA IX. PDB entry 3KS3 of CA II was used for figure D.

2.1.3 CAs related to cancer

Two isoforms of human carbonic anhydrases, CA IX and CA XII, have been associated with cancer, particularly with various types of solid tumors (renal cell, breast, colorectal, ovarian and other types of carcinomas, or brain tumors). Both are transmembrane isoenzymes, and their overexpression at the surface of cancer cells promotes tumor development by regulation of extracellular and consequently intracellular pH (Doyen *et al*, 2012). CA XII, unlike CA IX, is expressed in various healthy tissues (kidney, endometrium, gut, eye, sweat glands and others) and its correlation to poor prognosis is not as evident as it is for CA IX (Waheed & Sly, 2017). Nevertheless, silencing of both these genes led to reduced tumor size in mice model, and their combined targeting has been proposed as they play similar role, even though CA XII is less enzymatically active (Chiche *et al*, 2009). Their activity and function in tumors may be controlled by post-translational modifications (PTMs) such as phosphorylation, glycosylation, acetylation and ubiquitination (Di Fiore *et al*, 2022). The specific consequences of these PTMs have been experimentally elucidated only in some cases, and for CA IX these will be discussed in chapter 2.3. The mechanisms how CA IX contributes to tumor development will be described in detail in chapter 2.2.2. Since CA XII is not subject of this theses, it's role in tumor will not be further discussed and it can be found elsewhere (Waheed & Sly, 2017).

2.2 CA IX in cancer biology

CA IX has been validated as a tumor-associated enzyme which corelates with poor prognosis and patient survival, or increased metastasis when overexpressed in solid tumors of breast (Lou *et al*, 2011), oral/oropharyngeal squamous cell (Pérez-Sayáns *et al*, 2012; Sáenz-de-Santa-María *et al*, 2017), and clear renal cell (Stillebroer *et al*, 2010) carcinomas and others which are listed elsewhere (Ilardi *et al*, 2014). At physiological state, CA IX is expressed only in stomach and gall bladder, and at low levels in small intestine, duodenum and testes (from The Human Protein Atlas). However, when overexpressed at cancer parenchymal cells, it promotes the tumor development mainly by its pro-survival effect in hypoxic and acidic microenvironment of the solid tumors. Hence, CA IX is involved in many stages of the complex scenario of tumor growth, progression, and metastasis. This chapter will cover the cellular mechanisms underlying CA IX's regulation and function in cancer.

2.2.1 Regulation of CA IX overexpression

CA9 is recognized as a hypoxia-inducible gene, which is not expressed under normoxic conditions, but its expression is highly elevated in numerous cancer cell lines when exposed to low concentration of oxygen (0.1-2 % O_2), for example breast (MDA-MB-231, HBL-100, MCF-7) (Wykoff *et al*, 2000; Gieling *et al*, 2012; Jamali *et al*, 2015), cervical (HeLa) (Wykoff *et al*, 2000; Kazokaitė *et al*, 2018) and lung (A549) (Wykoff *et al*, 2000; Chiche *et al*, 2009). As a result, CA IX overexpression overlays with acidic regions within the tumor, which can either be in the hypoxic core, where it was initially described (Wykoff *et al*, 2000), or at the periphery with highly proliferative cells at the tumor-stroma interface (Rohani *et al*, 2019).

Transcription of *CA9* is activated by hypoxia-inducible factor (HIF-1), which recognizes specific sequence in its promoter (TGCACGTA), the so-called hypoxia-response element (HRE) (Wykoff *et al*, 2000). Binding to the promotor and initiation of expression is due to well-known stabilization of hypoxia-sensitive HIF-1α subunit upon lack of oxygen. In the presence of oxygen, HIF-1α (*HIF1A*) is post-translationally modified by enzymes with prolyl hydroxylase domains (PHD), which make the subunit a target for degradation in proteasome. The proteasomal destruction is mediated by E3 ubiquitin-ligase called von-Hippel-Lindau-tumor suppressor (VHL, *VHL*). Thus when there is low concentration of oxygen the HIF-1α is not modified, not degraded and it can heterodimerize with HIF-1β (*ARNT*) subunit and perform its function as a transcriptional factor (reviewed in (Schofield & Ratcliffe, 2004)). The regulation of *CA9* expression is schematically depicted in Figure 3. This regulation is disrupted in clear cell renal carcinoma, where the dominant negative regulator VHL is lost, therefore CA IX is highly expressed even in normoxia (Wykoff *et al*, 2000; Stillebroer *et al*, 2010).

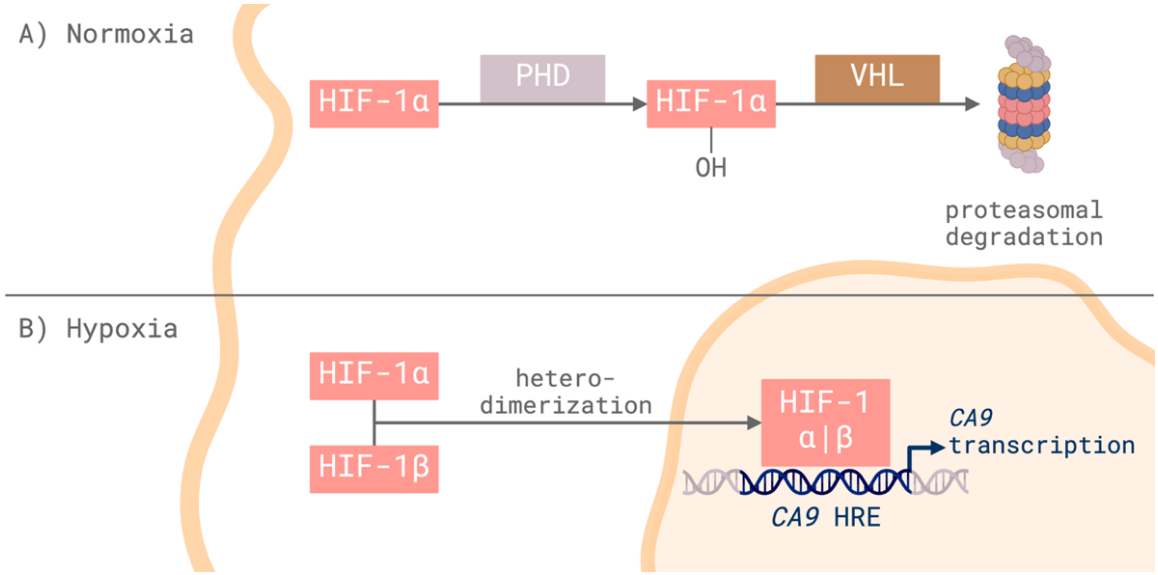


Figure 3: Regulation of CA9 *gene expression. A) Under normal concentration of oxygen, HIF-1α subunit is post-translationally modified with hydroxylation by PHD enzymes. This PTM makes it targeted for ubiquitination by E3 ligase VHL, which ultimately leads to degradation of HIF-1α in proteasome. Without the HIF-1α subunit the transcriptional factor HIF-1 is not assembled, therefore there is no expression of hypoxia-inducible genes, such as* CA9*. B) When the cells are exposed to hypoxic environment, PHD enzymes are inactivated so they do not modify the HIF-1α subunit, it is stabilized and it heterodimerizes with the HIF-1β subunit. The complex is translocated to nucleus where it binds to HRE sequence in promotor of* CA9 *gene, which is then transcribed and further expressed. The figure was created in Biorender.com*

Next to that, the positive regulation of *CA9* expression by HIF-1 has been connected to other typical oncogenic pathways like MAPK cascade with RAS, MEK and ERK activation stabilizing HIF-1α and leading to high levels of CA IX (McDonald *et al*, 2019). Increased level of CA IX and modulation of its localization on the plasmatic membrane is also regulated as a consequence to hepatocyte growth factor (HGF) signaling (Svastova *et al*, 2012).

2.2.2 Function of CA IX in tumor

The cancer development is characterized by many alterations in cellular processes, which are tightly interconnected, influence each other and altogether contribute to the cancer progression (Hanahan & Weinberg, 2011; Swietach *et al*, 2023). In the context of CA IX, the most important hallmarks are altered metabolism, hypoxia, pH regulation, cell proliferation, migration and metastasis. To start with the altered metabolism, all cancer cells are highly metabolically active, which leads to elevated concentration of acidic products. Either if it is an aerobically respirating cell producing CO_2, or a hypoxic cell extensively using glycolysis generating lactate, both types need to overcome the resulting decrease of intracellular pH (pH_i), since fine-tuned pH_i within physiological range (~7.2) is crucial for cell proliferation and survival (Swietach *et al*, 2008, 2009). By its enzymatic activity CA IX is responsible for a buffering system, by which the cancer cells are able to maintain high pH_i together with decrease of extracellular pH (pH_e) leading to an ability for invasion to surrounding tissues

(Swietach *et al*, 2023). The precise molecular and cellular mechanisms of CA IX function, together with identification of interacting proteins, have been studied mainly *in vitro* by gene depletion or transfection, pharmacologic inhibition, and biochemical or cell assays (proliferation, migration, invasion). Results of these studies are summarized in Table 2.

Table 2: Summary of CA IX function in tumor development.

Cancer progression	Cellular process	Mechanism	Reference
Tumor growth	cell survival proliferation metabolism	intracellular alkalization	(Swietach *et al*, 2008, 2009; Chiche *et al*, 2009; Jamali *et al*, 2015)
		survival in hypoxia	(Robertson *et al*, 2004; Swietach *et al*, 2008)
		survival in low pH	(Chiche *et al*, 2009)
		CO_2 removal	(Swietach *et al*, 2009)
		lactate efflux	(Jamali *et al*, 2015; Ames *et al*, 2018)
Metastasis	migration invasion ECM breakdown EMT transition	extracellular acidification	(Švastová *et al*, 2004; Chiche *et al*, 2009; Swietach *et al*, 2009; Radvak *et al*, 2013; Kazokaitė *et al*, 2018; Lee *et al*, 2018)
		lower cell-to-cell adhesion	(Švastová *et al*, 2003; Shin *et al*, 2011)
		focal contact/adhesion dynamics	(Švastová *et al*, 2003; Csaderova *et al*, 2013; Radvak *et al*, 2013)
		lamellipodia dynamics	(Csaderova *et al*, 2013)
		invadopodia dynamics	(Swayampakula *et al*, 2017; Debreova *et al*, 2019)
		MMP activation	(Swayampakula *et al*, 2017; Debreova *et al*, 2019)
		actin reorganization	(Shin *et al*, 2011)

Firstly, the CO_2, that is produced mostly in the mitochondrial tricarboxylic acid cycle (TCA), goes through the PM, where it is hydrated by CA IX forming HCO_3^-. The bicarbonate is then transported back to the cytosol by sodium bicarbonate cotransporters (NBC2, *SLC4A4* or NBC3, *SLC4A8*), or anion exchanger (AE, *SLC4A2*) (Svastova *et al*, 2012). The interplay between CA IX and these transporters ultimately leads to intracellular alkalization and spatial-temporal regulation of pH_i, which is essential for the cell survival mainly at multicellular levels as it has been shown in spheroids (Swietach *et al*, 2008, 2008). Additionally, the effect is even more profound under hypoxic conditions (Chiche *et al*, 2009; McDonald *et al*, 2019). Next to its catalytic activity CA IX contributes to the regulation of pHi by cooperation with monocarboxylate transporters (MCT1, *SLC16A1* or MCT4, *SLC16A3*). Their physical interaction increases the lactate and proton efflux from the cell, while CA IX act as so-called 'proton antenna' distributing H^+ (Jamali *et al*, 2015; Ames *et al*, 2018). The overview of how these proteins preserve the physiological pH_i in high-metabolic cell is illustrated in Figure 4A.

Not only that the described interplays alkalize the cell interior, they also acidify the extracellular environment (pH ~6.5) (Švastová *et al*, 2004; Swietach *et al*, 2008), The low pH_e is closely connected with the cancer cell ability to migrate, therefore further propagate within the body. It has been shown that CA IX together with NBC and MCT colocalize to cellular protrusions at the leading edge of migrating cells situated at the growing margin of the tumor (Svastova *et al*, 2012; Lee *et al*, 2018; Rohani *et al*, 2019). These proteins are redistributed to a lamellipodium, a general motility protrusion, where they regulate the ion transport and pH_e enabling cell invasion (Svastova *et al*, 2012; Csaderova *et al*, 2013). The invasion to surrounding tissue must be accompanied with extracellular matrix (ECM) breakdown by matrix metalloproteases (MMP) expressed in more aggressive protrusions called invadopodia (reviewed in (Linder *et al*, 2023)). CA IX has been recognized to interact with these enzymes, namely MMP14 (*MMP14*) and MMP9 (*MMP9*), and promote their function by supplying the protons necessary for MMP activity (Radvak *et al*, 2013; Swayampakula *et al*, 2017; Debreova *et al*, 2019; Rohani *et al*, 2019). CA IX involvement in cellular migration and invasion is depicted in Figure 4B.

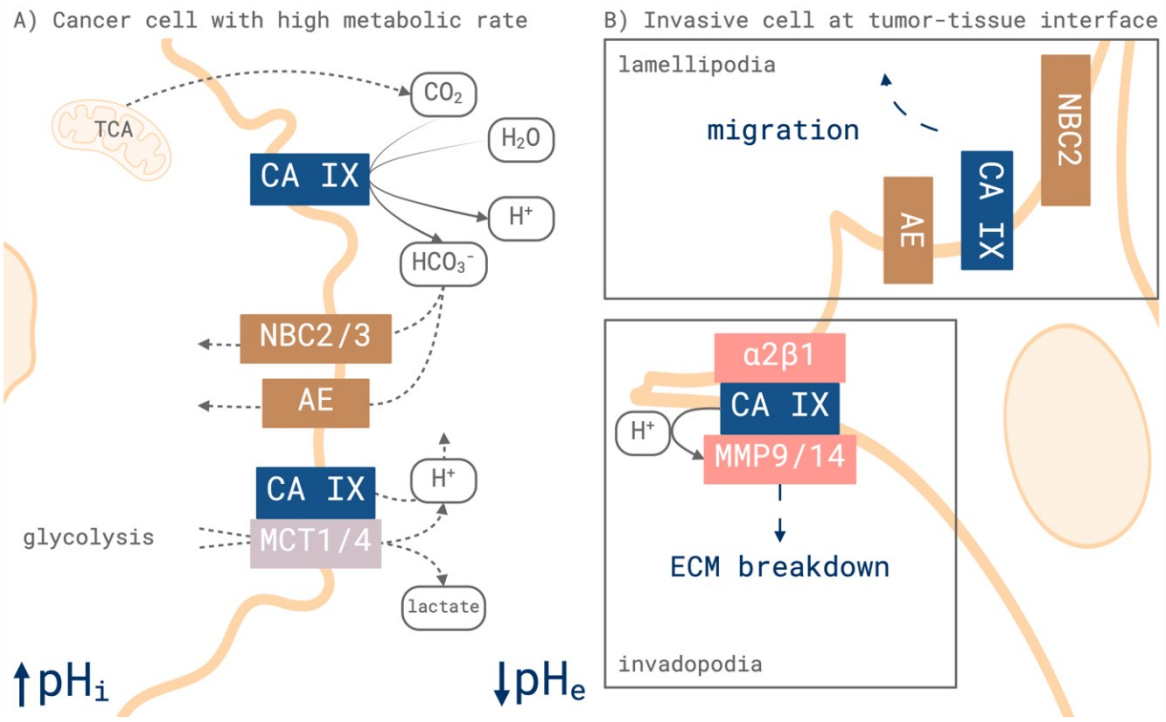


Figure 4: Schematic illustration of CA IX functions and interactions on a cancer cell. A) To moderate elevated levels of acidic products in highly metabolic cell, CA IX catalyzes hydration of CO_2 produced in mitochondrial TCA cycle (tricarboxylic acid cycle) to form bicarbonate, which is transported back to the cell by NBC (sodium bicarbonate cotransporter; together with sodium ion – not depicted) or AE (anion exchanger; antiport with chloride ion – not depicted). CA IX interacts and positively regulates activity of MCT (monocarboxylate transporter; proton-coupled symport) for export of lactate and hydrogen. Altogether this complex interplay leads to increase of intracellular pH (pH_i) to physiological levels essential for cancer cell survival and progression, as well as decrease of extracellular pH (pH_e) which helps cancer cell migration. B) In a migrating cancer cell invading to surrounding tissue, CA IX along with AE and NBC2 localize to cellular protrusions lamellipodia, key structures for cell motility, where their regulation of pH favors the tumor expansion. CA IX localizes also in the invadopodia, intrusive motility structures with accumulation of matrix metalloproteases (MMP) that play role in ECM breakdown and are positively regulated by interaction with CA IX. Created in Biorender.com

Besides the catalytic activity in the extracellular environment, CA IX enhances cell motility by comprising cell-to-cell contacts, either via binding to β-catenin (*CTNNB1*) thus interfering with E-cadherin-mediated adherence (Švastová *et al*, 2003), or by interaction with DKK1 (*DKK1*) and involvement in actin-regulating cascade of Rho/ROCK (Shin *et al*, 2011). CA IX expression level influences the dynamics between another important motility structures; nascent adhesion – focal contact – focal adhesion (Švastová *et al*, 2003; Csaderova *et al*, 2013; Radvak *et al*, 2013). Although this result has been questioned later on, CA IX does colocalize with αβ integrin (α2, *ITGA2*, β1, *ITGB1*) (Swayampakula *et al*, 2017), an important component of focal adhesion assembly (reviewed in (Kanchanawong & Calderwood, 2023)). Moreover, CA IX is associated with epithelial-mesenchymal transition (EMT) of cancer cells, since CA IX-expressing cells are more prone for mesenchymal morphology (Svastova *et al*, 2012).

2.3 Structural characteristics of CA IX

As mentioned earlier in chapter 2.1.1, human carbonic anhydrase IX is a transmembrane protein. Starting from the N-terminus, it is composed of signaling peptide (SP), which is cleaved during the translocation to plasma membrane, followed by large extracellular region (ECR), one transmembrane helix (TM) and short intracellular tail (IC). The ECR is divided into two domains connected by short linkers. The N-terminal domain called proteoglycan-like (PG) is a unique feature of CA IX as it is not present in any other human CA. The PG domain is followed by CA-common catalytic domain (Opavský *et al*, 1996). The domains with numbering of their amino acid residues are schematically depicted in Figure 5.

All the functions descripted in chapter 2.2.2 makes CA IX a potential drug target for cancer treatment. But the development of specific inhibitors is hindered by the fact that all human isoforms are highly conserved. Therefore, CA IX is a subject for the structure-based drug design which can provide important information for development of isoform-selective inhibitors. The most common method used in the structure-based drug design, or generally in structural studies, is X-ray crystallography. However, as it has been shown during the years of research, CA IX protein, namely its ECR, is quite challenging to express and successfully crystallize, and so far, only the common catalytic domain has been structurally determined (Alterio *et al*, 2009; Koruza *et al*, 2020). Nevertheless, other CA IX's features, in particular the PTMs, PG domain flexibility, and oligomerization state, has been studied by diverse biophysical methods and will be also covered in this chapter.

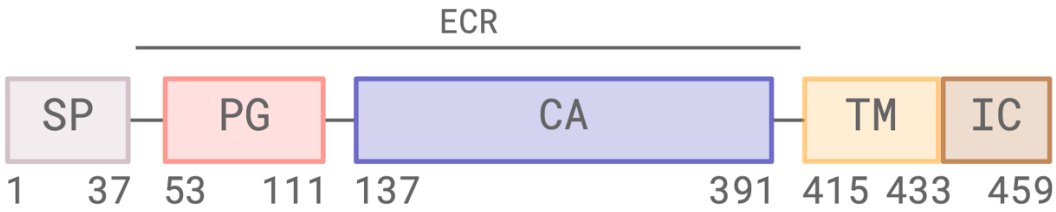


Figure 5: Schematic illustration of CA IX sequence. N-terminal part is recognized as signaling peptide (SP; 1-37) and it is cleaved during protein processing and transport to plasma membrane. Extracellular region (ECR) consists of unique proteoglycan-like domain (PG; 53-111) and common catalytic domain (CA; 137-391). C-terminus includes transmembrane helix (TM; 415-433) and short intracellular tail (IC; 434-459). Created in Biorender.com

2.3.1 PG domain

The proteoglycan-like domain was named based on its 38% sequential homology with a keratan sulfate attachment domain of aggrecan, a human cartilage protein which interact with ECM components (Opavský *et al*, 1996). Though sometimes it is delineated by residues 38-136 (from SP end to the start of the catalytic domain), the main region of PG domain is between residues 53 and 111. It is characterized by high number of acidic residues including 8 aspartates and 16 glutamates.

Majority of these 26 negatively charged residues is clustered within sixfold tandem repeat in region 61-96, where four repeats are completely identical (GEEDLP) and two contain two substitutions (SEEDSP and REEDPP) (Opavský *et al*, 1996). This feature makes PG domain a potential intrinsically disordered region (IDR) (Uversky, 2013).

Indeed, examination of PG by nuclear magnetic resonance (NMR) and circular dichroism revealed its inherent flexibility and unfolded state with only short regions at the N- and C-terminus with tendency to form secondary structures (Langella *et al*, 2018, 2021). Although, it should be noted that the experiments were carried out with isolated PG domain without the context of the CA domain. Investigation of CA IX protein variants consisting of the whole ECR (38-414) by small-angle X-ray scattering (SAXS) is in agreement with the flexible nature of PG, yet it showed that the PG remains close to the CA domain and that it is not completely extended in solution (Koruza *et al*, 2020). The interaction between PG and catalytic domain was proposed without any experimental evidence when the catalytic domain was firstly resolved, since there are three positively charged residues delimiting the active site (R194, R196, R268), which may interact with the negative residues of PG (Alterio *et al*, 2009). Overall, the nature of PG hinders the high-resolution structural characterization of the ECR by X-ray crystallography, due to the intrinsic flexibility which impairs crystal grow, or by its propensity for cleavage by proteases making the protein variants containing the ECR unstable over time (Alterio *et al*, 2009; Koruza *et al*, 2020).

Despite not having the catalytic activity, PG domain has been assigned with multiple functions in the cancer progression. The presence of PG domain enhances catalytic activity of CA domain, and also decreases the optimal working pH convenient for the CA IX role in tumors (Hilvo *et al*, 2008; Innocenti *et al*, 2009; Ames *et al*, 2018). PG domain works as the 'proton antenna', that buffers protons either from its own catalytic domain (Innocenti *et al*, 2009), or from MCT, thus enhancing lactate efflux (Jamali *et al*, 2015; Ames *et al*, 2018). Originally, CA IX was described as an adhesion molecule with the PG domain named after highly glycosylated protein which binds ECM (Opavský *et al*, 1996; Zavada *et al*, 1997). Only later to be found out that PG does covalently bind to ECM glycosaminoglycans (chondroitin or heparan sulfate) by O-linked glycosylation on residue S54, and therefore mediates cell-matrix adhesion (Christianson *et al*, 2017). Next to this modification with high-molecular weight glycans, the linker between PG and CA domain is glycosylated by di-, tri- or tetra-saccharide on T115 (Hilvo *et al*, 2008; Buanne *et al*, 2013).

2.3.2 Catalytic domain

The structure of CA IX catalytic domain has been explored extensively. There are 26 structures determined by X-ray crystallography and published in PDB database with resolution range 1.28-2.63 Å (Table 3), which clearly display the common well-folded globular domain of CA including the conserved catalytic site.

When comparing optimal working pH, catalytic domain alone is similar to other CAs with pK_a ~7, on the other hand for full-length CA IX it is ~6.5 (Innocenti *et al*, 2009). CA IX, with both CA and PG domain, performs the highest catalytic efficiency among CAs, up to k_{cat}/K_m = 1.5 x 10^9 [$M^{-1}s^{-1}$] (Hilvo *et al*, 2008). CA domain is similarly to the PG domain modified by glycosylation, here it is N-glycosylation by high mannose-type glycan on residue N346 (Hilvo *et al*, 2008; Alterio *et al*, 2009; Buanne *et al*, 2013).

Table 3: Summary of all structures of CA IX catalytic domain resolve by X-ray crystallography and published in PDB. Data acquired by Advanced search and Custom Tabular Report in RCSB as described in 4.4.1.

Entry ID	Resolution [Å]	Space Group	pH	Oligomeric State	Reference
3IAI	2.2	P 61	4	Homo 2-mer	(Alterio *et al*, 2009)
5FL4	1.82	H 3	4.5	Monomer	(Leitans *et al*, 2015)
5FL5	2.05	H 3	8.5	Monomer	(Leitans *et al*, 2015)
5FL6	1.95	H 3	4.5	Monomer	(Leitans *et al*, 2015)
5DVX	1.598	P 21 21 21	8.5	Monomer	(Mahon *et al*, 2016)
6FE0	1.91	H 3	4.5	Homo 2-mer	(Kazokaitė *et al*, 2018)
6FE1	1.95	H 3	4.5	Homo 2-mer	(Kazokaitė *et al*, 2018)
6FE2	1.87	H 3	4.5	Homo 2-mer	(Kazokaitė *et al*, 2018)
6G98	2.47	H 3	4.5	Homo 2-mer	(Kazokaitė *et al*, 2018)
6G9U	1.75	H 3	4.5	Homo 2-mer	(Kazokaitė *et al*, 2018)
6RQN	1.777	P 1 21 1	8.5	Monomer	(Koruza *et al*, 2019)
6RQQ	1.276	P 1 21 1	8.5	Monomer	(Koruza *et al*, 2019)
6RQW	1.488	P 1 21 1	8.5	Monomer	(Koruza *et al*, 2019)
6RQU	1.393	P 1 21 1	8.5	Monomer	(Koruza *et al*, 2019)
6QN2	1.95	H 3	4.5	Homo 2-mer	(Zakšauskas *et al*, 2020)
6QN5	1.96	H 3	4.5	Homo 2-mer	(Zakšauskas *et al*, 2020)
6QN6	2.25	H 3	4.5	Homo 2-mer	(Zakšauskas *et al*, 2020)
6QUT	1.96	H 3	4.5	Homo 2-mer	(Zakšauskas *et al*, 2020)
6Y74	1.53	C 1 2 1	4.3	Homo 2-mer	(Koruza *et al*, 2020)
6TL5	2.21	H 3	4.5	Homo 2-mer	(Dudutienė *et al*, 2020)
6TL6	2.15	H 3	4.5	Homo 2-mer	(Dudutienė *et al*, 2020)
7POM	1.98	H 3	4.5	Homo 2-mer	(Zakšauskas *et al*, 2021)
8Q18	2.13	H 3	4.5	Homo 2-mer	(Leitans *et al*, 2023)
8Q19	2.63	H 3	4.5	Homo 2-mer	(Leitans *et al*, 2023)
8Q1A	2.35	H 3	4.5	Homo 2-mer	(Leitans *et al*, 2023)
8CO0	2.3	H 3	4.5	Homo 4-mer	(Ivanova *et al*, 2023)

For the structural studies by X-ray crystallography, the protein of interest is expressed in recombinant form, purified, and crystallized to be subjected to the final X-ray experiment. Any variability during this process, from molecular cloning up to the data collection, impacts the final outcome of the protein structure (Müller, 2017). In the case of CA IX, the structurally characterized catalytic domain originates from a variety of protein variants and expression systems, summarized in Table 4. For the first structural experiments, the protein variant was expressed in insect-based system with *Spodoptera frugiperda* cells, with one mutation (C174S) that was introduced to avoid aggregation and precipitation of the protein (Alterio *et al*, 2009). A similar approach was followed later on for its high chances of glycosylated soluble protein, and resulted in a structure with no mutation (Koruza *et al*, 2020). On the contrary, the modification by glycans was eliminated when a protein variant with double mutations (C174S and N346Q) was produced in yeast-based system with *Komagataella pastoris* (Leitans *et al*, 2015, 2023; Zakšauskas *et al*, 2021; Ivanova *et al*, 2023). Regardless the N346Q mutation, all structures from yeast bearing only C174S mutation (Leitans *et al*, 2015; Kazokaitė *et al*, 2018; Dudutienė *et al*, 2020; Zakšauskas *et al*, 2020), are not glycosylated (from RCSB database Advanced Search).

Since the protein expression and purification from insect or yeast cells is quite laborious, a bacteria-based system has been established to produce CA IX catalytic domain in *Escherichia coli* (Mahon *et al*, 2016; Koruza *et al*, 2019). To ease the purification and crystallization process, six mutations were introduced on the surface: C174S to avoid the aggregation as in the previous designs, L180S and M360S to reduce hydrophobicity, and finally A210K, A258K and F259Y mimicking CA II which is easily purified and routinely crystallized (Domsic *et al*, 2008; Mahon *et al*, 2016).

Table 4: Summary of expression systems utilized to produce CA IX catalytic domain for X-ray crystallography experiments. Herein the references concern only publication where the expression and purification procedure is described, not all references to the corresponding PDB entries. Data acquired by Advanced search and Custom Tabular Report in RCSB.

Number of mutations	Mutations	Number of entries	Expression host	Entries	Reference
0	/	1	*S. frugiperda*	6Y74	(Alterio *et al*, 2009)
1	C174S	15	*S. frugiperda* (1), *K. pastoris* (14)	3IAI 5FL4, 5FL5, 5FL6, 6FE0, 6FE1, 6FE2, 6G98, 6G9U, 6QN2, 6QN5, 6QN6, 6QUT, 6TL5, 6TL6	(Alterio *et al*, 2009; Leitans *et al*, 2015)
2	C174S, N346Q	5	*K. pastoris*	7POM, 8CO0, 8Q18, 8Q19, 8Q1A	(Leitans *et al*, 2015)
6	C174S, L183S, A213K, A258K, F259Y, M350S	5	*E. coli*	5DVX, 6RQN, 6RQQ, 6RQU, 6RQW	(Mahon *et al*, 2016)

As mentioned before, the final structural outcome depends on numerous details during the experimental procedure, therefore it is important to disclose the variabilities published for CA IX crystallization. Out of the 26 structures, 19 constitute of enzyme with bound inhibitor, 4 of these were prepared by co-crystallization (Alterio *et al*, 2009; Dudutienė *et al*, 2020; Zakšauskas *et al*, 2021), when the CA IX was incubated with an inhibitor prior to the crystallization setup and 15 structures were determined from pre-formed crystals that were soaked with the inhibitor (Leitans *et al*, 2015, 2023; Kazokaitė *et al*, 2018; Zakšauskas *et al*, 2020; Ivanova *et al*, 2023). Remaining 7 structures are of the catalytic domain without any ligand (Mahon *et al*, 2016; Kazokaitė *et al*, 2018; Koruza *et al*, 2019, 2020).

Another key factor is the concentration of the protein that is subjected for the crystallization experiment, and the catalytic domain of CA IX has crystallized in the concentration range from 5 to 17 mg/mL. Next to the protein concentration, the ratio between protein and precipitant volumes can vary, though this information is not published for all experiments with CA IX, some crystals were obtained by 1:1 ratio (Leitans *et al*, 2015, 2023; Koruza *et al*, 2019, 2020; Dudutienė *et al*, 2020; Zakšauskas *et al*, 2021). When the initial experimental arrangement was not successful to form high-quality crystals, optimization in form of micro-seeding method was employed (Koruza *et al*, 2019, 2020). Additionally, the age and storage conditions influence the ability of CA IX to crystallize, considering that fresh protein did not form any crystals, but protein sample stored in 4°C for six months resulted in a crystal diffracting up to 1.53 Å (Koruza *et al*, 2020).

Despite that the protein variants were produced in different expression systems, the structures are equal when compared by superposition (RMSD 0.28-0.35 Å) (Mahon *et al*, 2016). What distinguishes the proteins produced in *E. coli* from the others is the pH of crystallization condition, with pH 8.5 for bacterial recombinant proteins and pH 4-4.5 for any other structure (from RCSB search, denoted in Table 3:), with one exception from yeast which also crystallized at pH 8.5 (5FL5) (Leitans *et al*, 2015). Similarly, the proteins crystallized in numerous space groups (Table 3:), or oligomeric states which will be further discussed in chapter 2.3.4.

Finally, it should be noted that there were attempts to crystallize a protein variant bearing the whole ECR produced in insect cells, but they failed due to the proteolytic cleavage of PG domain (Alterio *et al*, 2009; Koruza *et al*, 2020).

2.3.3 Transmembrane helix and intracellular region

CA IX belongs to the type I single-pass transmembrane protein thanks to its sequence region 415-440 which corresponds to the so-called stop-transfer-membrane anchor. This sequence is together with the signaling peptide responsible for the proper protein targeting, firstly to the endoplasmic reticulum during translation, and ultimately to the plasma membrane with N-terminal region situated outside the cell. CA IX's transmembrane helix (415-433) complies with the rule of ~20 amino acid long

sequence with high content of hydrophobic residues (L, A, V, F) (Liu *et al*, 2002), followed by positively charged residues (R436, R437 and R440) (from Uniprot).

Apart from the SP and TM, the C-terminal intracellular tail is also needed for proper localization of CA IX on the plasma membrane (Hulikova *et al*, 2009). Moreover, the IC region contributes to the catalytic function as IC-deletion abolishes the enzymatic activity and extracellular acidification (Hulikova *et al*, 2009; Swayampakula *et al*, 2017). The underlying mechanism is not fully understood but protein-protein interactions and PTMs are considered to be involved. CA IX is phosphorylated by protein kinase A (PKA) on residue T443 inside the canonical PKA recognition motif RRxT, and this PTM is key for the enzymatic activity in the hypoxic cells (Ditte *et al*, 2011). Opposite to the activating phosphorylation of T443, the phosphorylation on residue S448 has a negative effect in CA IX regulation (Ditte *et al*, 2011). The last modified residue is Y449, which is phosphorylated upon epithelial growth factor (EGF) signaling (Dorai *et al*, 2005). However, it is not clear whether the phosphorylation is mediated directly by the receptor tyrosine kinase EGFR or by another kinase. Both phospho-T442 and phospho-Y449 are included in the CA IX association with proteins of nuclear transport (XPO, TNPO, CAND1) (Buanne *et al*, 2013; Buonanno *et al*, 2017).

2.3.4 Quaternary structure

Even before any structural studies, it was suggested that CA IX forms some oligomeric arrangement on the cellular plasma membrane (Závada *et al*, 1993, 2000; Hulikova *et al*, 2009; Ditte *et al*, 2011). The CA IX quaternary architecture on the cancer cell surface was narrowed down to a homodimeric arrangement (Švastová *et al*, 2004; Li *et al*, 2011). The following biochemical and structural investigation reported the physiological oligomeric structure to be a dimer stabilized by intermolecular disulfide bridge between two monomers of CA IX (Hilvo *et al*, 2008; Alterio *et al*, 2009). Although, this covalent bond between two C174 residues is not essential, as protein variant with C174S mutation forms structurally identical dimeric arrangement compared to the WT. The dimer is preserved by other interactions as H-bonds (R268, A258) and Van der Waals interactions within the contact area (A172-A176) (Alterio *et al*, 2009), pictured in Figure 6.

Each PDB entry for CA IX has its biological assembly which is defined by authors (as 'remark 350' in pdb file) and which corresponds to the oligomeric state of the catalytic domain structure in the crystal (Table 3). However, the oligomeric state of the recombinant CA IX protein used for structural characterization is often not reported in the corresponding literature. Except for the analysis by size exclusion chromatography (SEC) during purification, when either two peaks of monomeric and dimeric states (Koruza *et al*, 2018, 2019), or solely one monomeric (Mahon *et al*, 2016), were identified for the protein variants of catalytic domain produced in *E. coli* and bearing six mutations, for which all structures were assigned to be biological monomers. Two peaks corresponding

to monomeric and dimeric states of protein variant with both PG and catalytic domain were identified during purification from baculovirus system (Koruza *et al*, 2020).

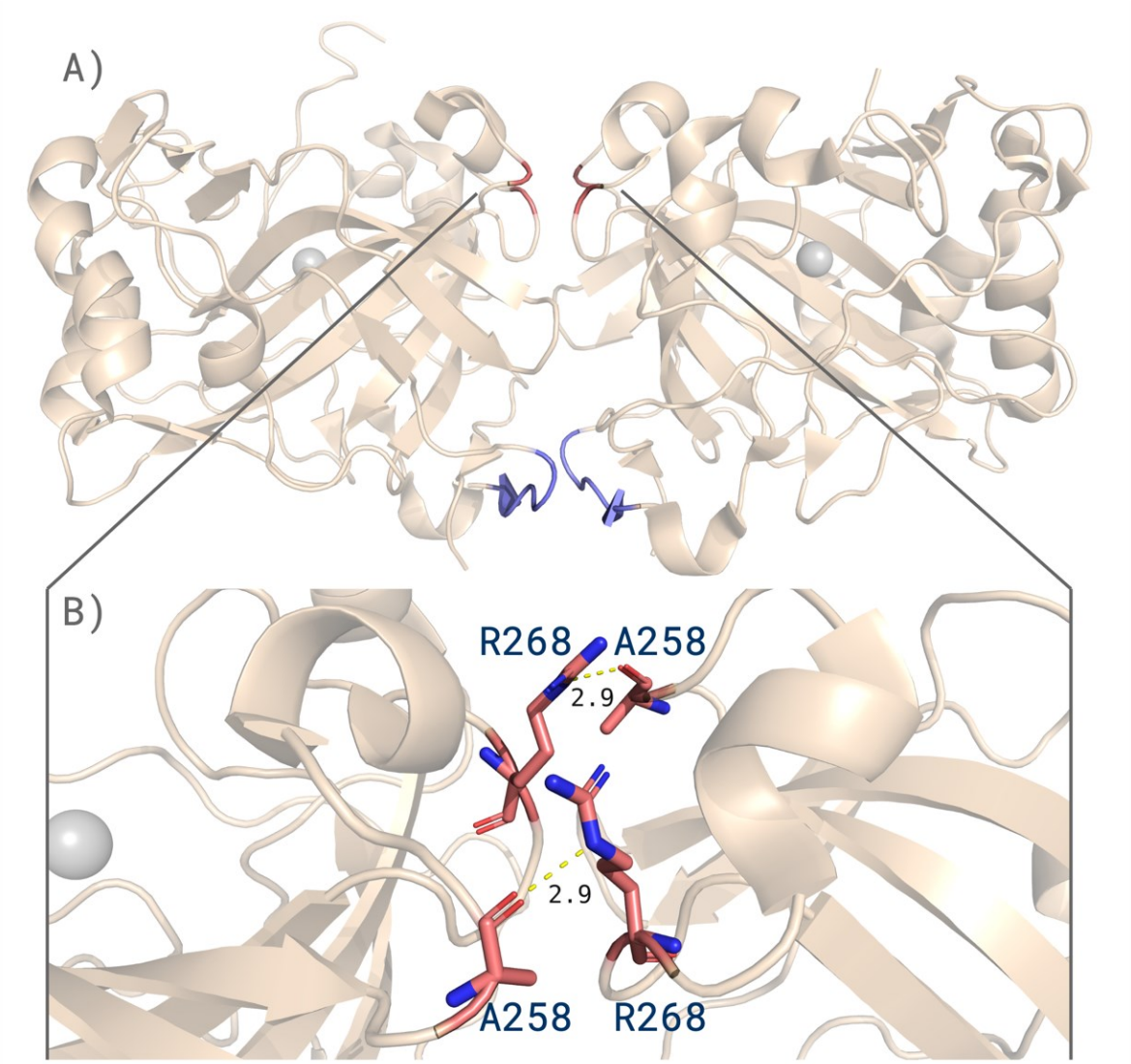


Figure 6: The benchmark for CA IX oligomeric state; PDB entry 3IAI. A) Dimeric interface of two CA IX monomers (in wheat), with highlighted contact area of residues A172-A176 (in violet), where the disulphide bridge is formed between two cysteines C174, if not mutated. B) Detail of hydrogen bonds formed between carbonyl of A258 and side chain of R268 of each monomer (shown as stick, in dark pink). The distances of H-bonds between the residues are depicted (2.9 Å). Each figure was made in PyMOL software and assembled in Biorender.com.

3 Materials

3.1 Bacterial strains and plasmids

Escherichia coli strains

Two competent bacterial strains (BL21 and NiCo21), both with λ DE3 bacteriophage construct bearing T7 RNA-polymerase under the transcription control of lacUV5 promoter, with the major difference of NiCo21 strain being resistant to T1 bacteriophage. Their genotypic description and sources are summarized in Table 5.

Table 5: Characteristics of bacterial strains used for CA IX recombinant expression.

E. coli strain	Genotype	Source
BL21 (DE3)	*F⁻ dcm ompT hsdS$_B$(r_B^- m_B^-) gal λ* (DE3)	Novagene, USA
NiCo21 (DE3)	*can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	NEB, USA

Plasmids

All expression vectors were prepared in the Laboratory of Structural Biology at IOCB by RNDr. Milan Fábry, CSc. The vectors were derived from pMCSG7 plasmid with the T7 promoter-driven expression of protein constructs of our interest cloned as NdeI-NotI insert. Each plasmid possessed sequence for the extracellular part of CA IX containing either solely the catalytic domain (781 bp, aa residues 137-391), or in attachment to the PG-domain (1078 bp, aa residues 38-391), as depicted in Figure 7. Various mutations were introduced resulting in eight protein variants of human CA IX bearing different amino acid substitutions, which are described in detail in Table 6.

Table 6: Protein variant features stating which domains were included, the given protein variant names and amino acid substitutions that were published before (ΔPG-CAIX/6 and CAIX/6 in (Mahon et al*, 2016)) or designed as part of this thesis (all the others).*

Domains	Protein variant name	Amino acid substitutions
catalytic domain	ΔPG-CAIX/6	C174S, L183S, A213K, A258K, F259Y, M350S
	ΔPG-CAIX/4	C174S, L183S, A213K, M350S
	ΔPG-CAIX/5S	C174S, L183S, A213K, A258S, M350S
	ΔPG-CAIX/5G	C174S, L183S, A213K, A260G, M350S
PG & catalytic domain	CAIX/6	C174S, L183S, A213K, A258K, F259Y, M350S
	CAIX/4	C174S, L183S, A213K, M350S
	CAIX/5S	C174S, L183S, A213K, A258S, M350S
	CAIX/5G	C174S, L183S, A213K, A260G, M350S

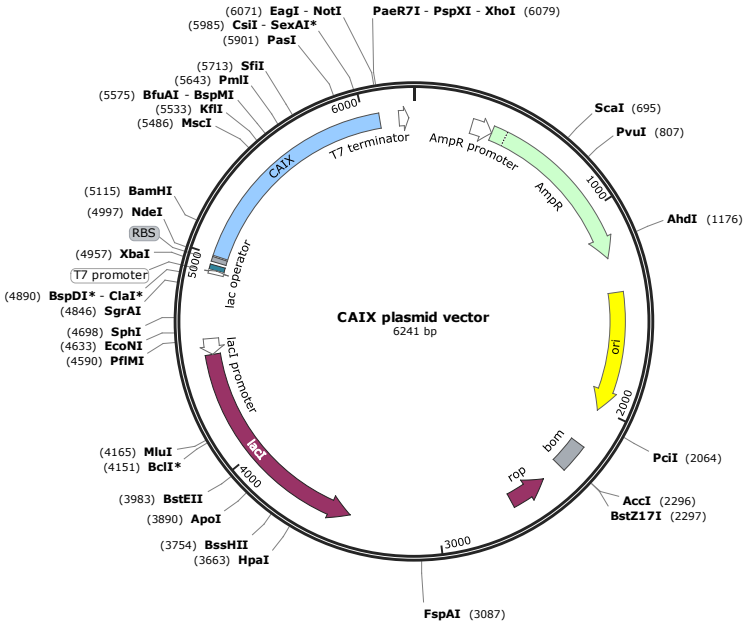


Figure 7: Plasmid map with the common features (AmpR = ampicillin resistance (+promoter), Ori = high-copy-number ColE1/pMB1/pBR322/pUC origin of replication, bom = basis of mobility region from pBR322, rop = Rop protein, which maintains plasmids at low copy number, lacI = lac repressor (+promoter), T7 promoter, lac operator, RBS= efficient ribosome binding site from bacteriophage T7 gene 10), and restriction enzyme sites such as NdeI and NotI which were used for expression of CA IX protein variants. The figure was made in SnapGene.

3.2 Chemicals

^{13}C glucose	Cambridge Isotope Laboratories, Inc., United Kingdom
^{15}N ammonium sulfate $(NH_4)_2SO_4$	CPI-Georgia LTD
2-(N-morpholino)ethanesulfonic acid (MES)	Sigma-Aldrich, USA
3,3',5,5'-tetramethylbenzidine (TMB)	Sigma-Aldrich, USA
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma-Aldrich, USA
acetic acid (CH_3COOH)	Lach-Ner, s.r.o., Czech Republic
acrylamide (AA)	Sigma-Aldrich, USA
ammonium acetate ($NH_4CH_2CO_2$)	Sigma-Aldrich, USA
ammonium persulfate (APS)	Sigma-Aldrich, USA
ampicillin	BB Pharma a.s., Czech Republic
boric acid (H_3BO_3)	Sigma-Aldrich, USA
bovine serum albumin (BSA)	Sigma-Aldrich, USA
bromfenol blue	Lachema, a.s., Czech Republic
calcium chloride dihydrate ($CaCl_2$ x 2 H_2O)	Merck, Germany
citric acid monohydrate	Sigma-Aldrich, USA
cobalt(II) chloride hexahydrate ($CoCl_2$ x 6 H_2O)	Sigma-Aldrich, USA
copper(II) chloride dihydrate ($CuCl_2$ x 2 H_2O)	Sigma-Aldrich, USA
d-Biotin	Sigma-Aldrich, USA
deuterium oxide (D_2O)	Cambridge Isotope Laboratories, Inc., United Kingdom
dimethyl sulfoxide (DMSO)	Sigma-Aldrich, USA
ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich, USA

ethyl-β-D-thiogalactopyranoside (ETG)	Carbosynth Limited, United Kingdom
formaldehyde (CH_2O)	Sigma-Aldrich, USA
glucose	Penta s.r.o., Czech Republic
glycerol	Penta s.r.o., Czech Republic
hydrochloric acid (HCl)	Lach-Ner, s.r.o., Czech Republic
iron(III) chloride hexahydrate ($FeCl_3$ x 6 H_2O)	Sigma-Aldrich, USA
LB Broth (Lennox)	Sigma-Aldrich, USA
LB Broth with agar (Lennox)	Sigma-Aldrich, USA
lithium chloride (LiCl)	Merck, Germany
lysozyme	Sigma-Aldrich, USA
magnesium sulfate ($MgSO_4$)	Sigma-Aldrich, USA
manganese(II) chloride ($MnCl_2$)	Sigma-Aldrich, USA
methanol (CH_3OH)	Lach-Ner, s.r.o., Czech Republic
N,N,N',N'-tetramethylethylendiamine (TEMED)	Sigma-Aldrich, USA
nonfat-dried milk	Nutristar, Aditiva CZ, s.r.o., Czech Republic
p-aminomethylbenzenesulfonamide–agarose	Sigma-Aldrich, USA
polyethylen glycol (PEG) 3500	Sigma-Aldrich, USA
PEG 4000	Merck, Germany
PEG 6000	Sigma-Aldrich, USA
PEG 8000	Sigma-Aldrich, USA
potassium dihydrogen phosphate (KH_2PO_4)	Sigma-Aldrich, USA
potassium hydrogen phosphate (K_2HPO_4)	Sigma-Aldrich, USA
potassium chloride (KCl)	Penta s.r.o., Czech Republic
silver nitrate ($AgNO_3$)	Sigma-Aldrich, USA
SimplyBlue™ SafeStain	Novex by Life technologies, Canada
sodium azide (NaN_3)	Sigma-Aldrich, USA
sodium carbonate (Na_2CO_3)	Sigma-Aldrich, USA
sodium dodecyl sulfate (SDS)	Sigma-Aldrich, USA
sodium chloride (NaCl)	Lach-Ner, s.r.o., Czech Republic
sodium phosphate dibasic dodecahydrate (Na_2HPO_4 x 12 H_2O)	Sigma-Aldrich, USA
sodium phosphate dibasic heptahydrate (Na_2HPO_4 x 7 H_2O)	Sigma-Aldrich, USA
sodium sulfate decahydrate (Na_2SO_4 x 10 H_2O)	Sigma-Aldrich, USA
sodium thiosulfate pentahydrate ($Na_2S_2O_3 \cdot 5H_2O$)	Lachema, a.s., Czech Republic
thiamine	Sigma-Aldrich, USA
triethanolamine (TEOA)	Sigma-Aldrich, USA
tris(2-carboxyethyl)phosphine (TCEP)	Merck, Germany
tris(hydrozymethyl)aminomethan (Tris)	Merck, Germany
Tween-20	Sigma-Aldrich, USA
zinc chloride ($ZnCl_2$)	Sigma-Aldrich, USA
zinc sulfate heptahydrate ($ZnSO_4$ x 7 H_2O)	Sigma-Aldrich, USA

3.3 Composition of solutions

3.3.1 Cultivating media

LB (Luria-Bertani) medium

20 g/L	LB Broth (Lennox) powder
0.1 g/L	ampicillin
0.43%	glycerol (v/v)
8.63 mg/L	zinc sulfate heptahydrate ($ZnSO_4$ x 7 H_2O)

LB Agar

35 g/L	LB Broth with agar (Lennox)
0.1 g/L	ampicillin
1% (w/v)	glucose

Minimal media (MM) for isotopically labelled samples (^{15}N and ^{13}C)

1 g/L	^{15}N ammonium sulfate $(NH_4)_2SO_4$
2 g/L	^{13}C glucose
0.1 g/L	ampicillin
8.6268 mg/L	zinc sulfate heptahydrate ($ZnSO_4$ x 7 H_2O)
17.66 g/L	sodium phosphate dibasic dodecahydrate (Na_2HPO_4 x 12 H_2O)
3 g/L	potassium dihydrogen phosphate (KH_2PO_4)
0.5 g/L	sodium chloride (NaCl)
96.5 mg/L	sodium sulfate decahydrate (Na_2SO_4 x 10 H_2O)
50 mg/L	ethylenediaminetetraacetic acid (EDTA)
16 mg/L	manganese(II) chloride ($MnCl_2$)
8 mg/L	iron(III) chloride hexahydrate ($FeCl_3$ x 6 H_2O)
0.5 mg/L	zinc chloride ($ZnCl_2$)
0.125 mg/L	copper(II) chloride dihydrate ($CuCl_2$ x 2 H_2O)
0.185 mg/L	cobalt(II) chloride hexahydrate ($CoCl_2$ x 6 H_2O)
0.1 mg/L	boric acid (H_3BO_3)
120.5 mg/L	magnesium sulfate ($MgSO_4$)
44 mg/L	calcium chloride dihydrate ($CaCl_2$ x 2 H_2O)
1 mg/L	d-Biotin
1 mg/L	thiamine

3.3.2 Solutions and buffers

Solutions for sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis

Stacking gel solution	5% (w/v)	AA
	125 mM	Tris, pH 6.8
	5% (v/v)	glycerol
	1 mM	EDTA
	0.1% (v/v)	TEMED
	0.1% (w/v)	SDS
	0.1% (w/v)	APS
Running gel solution	15% (w/v)	AA
	375 mM	Tris, pH 8.8
	5% (v/v)	glycerol
	1 mM	EDTA
	0.1% (v/v)	TEMED
	0.1% (w/v)	SDS
	0.1% (w/v)	APS
Sample buffer	125 mM	Tris, pH 6.8
	20% (v/v)	glycerol
	4% (w/v)	SDS
	0.2% (w/v)	bromfenol blue
Running (electrode) buffer	20 mM	Tris, pH 8.3
	200 mM	glycin
	0.1% (w/v)	SDS

Solutions for silver staining of SDS-PAGE gel

Fixative solution 1	50% (v/v)	methanol (CH_3OH)
	12% (v/v)	acetic acid (CH_3COOH)
	0.05% (v/v)	formaldehyde (CH_2O)
Fixative solution 2	50% (v/v)	methanol (CH_3OH)
Sensitizer solution	0.4 mM	sodium thiosulfate pentahydrate ($Na_2S_2O_3 \cdot 5H_2O$)
Stain solution	12 mM	silver nitrate ($AgNO_3$)
	0.075% (v/v)	formaldehyde (CH_2O)
Developer solution	570 mM	sodium carbonate (Na_2CO_3)
	0.05% (v/v)	formaldehyde (CH_2O)
	8 µM	sodium thiosulfate pentahydrate ($Na_2S_2O_3 \cdot 5H_2O$)
Stop solution	50% (v/v)	methanol (CH_3OH)
	12% (v/v)	acetic acid (CH_3COOH)

Frequently used buffers

Phosphate-buffered saline (PBS)	137 mM	natrium chloride (NaCl)
	2.7 mM	potassium chloride (KCl)
	13.4 mM	potassium hydrogen phosphate (K_2HPO_4)
	4.9 mM	sodium phosphate dibasic heptahydrate (Na_2HPO_4 x 7 H_2O)
Affinity chromatography Washing Buffer 1 (WB1)	100 mM	Tris-HCl, pH 9
	200 mM	sodium sulfate decahydrate (Na_2SO_4 x 10 H_2O)
Affinity chromatography Washing Buffer 2 (WB2)	100 mM	Tris-HCl, pH 7
	200 mM	sodium sulfate decahydrate (Na_2SO_4 x 10 H_2O)
Affinity chromatography Elution Buffer (EB)	50 mM	Tris-HCl, pH 7.8
	400 mM	sodium azide (NaN_3)
Affinity Regeneration solution 1 (AR1)	100 mM	acetic acid pH 5.6
	1 M	natrium chloride (NaCl)
Affinity Regeneration solution 2 (AR2)	25 mM	Tris-HCl, pH 8.7
	1 M	natrium chloride (NaCl)
AEC equilibration buffer	20 mM	triethanolamine (TEOA), pH 7.3
AEC elution buffer	20 mM	triethanolamine (TEOA), pH 7.3
	1 M	NaCl
Protein storage buffer A	20 mM	Tris-HCl, pH 7
	100 mM	NaCl
Protein storage buffer B	100 mM	MES, pH 6.5
Protein storage buffer C	100 mM	ammonium acetate, pH 7.3
NMR buffer	20 mM	Tris-d11, pH 7.3
	200 mM	NaCl
	1 mM	TCEP
	0.02%	azide

3.4 Instruments

incubators	Sanyo CO2 incubator MCO-ATAI, Japan
	New Brunswick Innova® 42/42R, Germany
	Epiphyte3 LEX-48 bioreactor, Canada
centrifuges	Benchmark Scientific MyFuge™ C1008-G, USA
	Eppendorf Mini Spin, Germany
	Hettich Zentrifugen Mikro 200R, Germany
	Beckman Coulter Allegra X-30R Centrifuge, SX4400 rotor, Germany
	Beckman Coulter Avanti JXN-30 Centrifuge, fixed angle rotors JA-20, J-LITE JLA 16.250, J-LITE JLA 9.1000, Germany
autoclave	Unisteri HP, BMT Medical Technology, Czech Republic
spectrophotometers	Thermo Scientific NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, USA
	Ultrospec 3100 Pro Amersham Biosciences, United Kingdom
western blot transfer system	Bio-Rad Trans-Blot Turbo Transfer System, USA
electrophoresis systems	Cytiva Mighty Small Multiple Gel Caster, Sweden
	Cytiva SE 250 Mini-Vertical Unit, Sweden
	Consort EV243, Belgium
thermostat	Labnet International D1100 AccuBlockTM Digital Dry Bath, USA
FPLC	ÄKTA pure™ 25, with Fraction collector F9-C, General Electric Healthcare Systems, USA
	Amersham Biosciences AKTA FPLC system, United Kingdom
optical microscopes	Olympus SZX10, Japan
	JANSi UVEX-m Imaging Systems, USA
photo camera	Olympus E-620, Japan
balances	Kern PRS 620-3, Germany
magnetic stirrer	ESP Ultraflat Magnetic Stirrer, Velp Scientifica, Italy
pH meter	GMH 3430 Griesinger Electronic, Germany
	Orion™ 9110DJWP Double Junction pH Electrode, Thermo Scientific, USA
conductometer	GMH 3430 Griesinger Electronic, Germany
cell disruptors	Soniprep 150 MSE, United Kingdom
	Artek Dynatech Fisher 300 Sonic Sonicator Dismembrator, USA
	Avestin Emulsiflex C3, Canada
crystallization robots	Crystal Gryphon Art Robbins Instruments, USA
	Oryx8 Douglas Instruments Ltd, United Kingdom
NMR spectrometer	Bruker Avance III™ HD 850 MHz spectrometer, Bruker Corporation, USA
imaging system	Azure 600, Biosystems Inc., USA
PCR machine	LightCycler® 480 Multiwell Plate 96

3.5 Other material

cultivation glassware	Petri dish plate, 90 mm, P-LAB, Czech Republic
	Glass petri dish plate, 200 mm, P-LAB, Czech Republic
	Erlenmeyer flasks, 2 L, P-LAB, Czech Republic
other glassware	Bottle with screw cap PP, clear, 0.1-1 L, DURAN®, P-LAB, Czech Republic
chromatographic columns	MonoQ™ 5/50 GL Tricorn MonoBeads™ Column, General Electric Healthcare Systems, USA
	SuperdexTM 200 Increase 10/300 GL, Cytiva, Sweden
dialysis utensils	Slide-A-Lyzer® Dialysis Cassette (Extra Strength), 10,000 MWCO capacity 0.5-3 mL, Thermo Scientific, USA Slide-A-Lyzer® G3 Dialysis Cassette, 10,000 MWCO capacity 15 mL, Thermo Scientific, USA
	Dialysis tubing, high retention seamless cellulose tubing, MWCO 12400, Merck, USA
concentrator	Amicon® Ultra Centrifugal Filter, 10 kDa MWCO, capacity 0.5-15 mL, Millipore, USA
MW standards for SDS-PAGE	Unstained Protein Standard, Broad Range (10-200 kDa), New England BioLabs, USA
	Spectra™ Multicolor Broad Range Protein Ladder, Thermo Scientific, USA
MW standard for SEC	Gel Filtration Standard #1511901 Bio-Rad, USA
western blot utensils	Bio-Rad Trans-Blot Turbo Transfer Midi-size nitrocellulose, USA
	Bio-Rad Trans-Blot Turbo Transfer Stacks, USA
	Bio-Rad Trans-Blot Turbo 5x Transfer Buffer, USA
DSF utensils	LightCycler ® 480 Multiwell Plate 96, white, Roche, Switzerland
	LightCycler ® 480 Sealing Foil, Roche, Switzerland
	SYPRO® Orange, Invitrogen, USA
crystallization utensils	MRC 3-well low crystallization plate, Jena Bioscience, Germany
	EasyXtal® 15-Well Tools, Qiagen, Germany
	ThermalSeal RTS™ Sealing Film, Sigma-Aldrich, USA
	X-Seal crystallization supports, Qiagen, Germany
	MicroMounts™, MiTeGen, USA
	Universal V1-Puck (Uni-Puck), MiTeGen, USA
crystallization screens	JSCG Plus, Hampton Research, USA
	Morpheus® Molecular Dimensions, United Kingdom
	The PEGs Suite, Qiagen, Germany
	SG1, MiTeGen, USA
	Additive Screen HT™ Hampton Research, USA
NMR tube	3 mm NMR tube, Bruker, USA

3.6 Software and databases

BioRender

Bruker NMR Software TopSpin™ 4.4.0

CCP4 v8.0 software package (MOLREP, REFMAC5, COOT)

Cytiva UNICORN™ 7 control software

Expasy ProtParam

MolProbity 4.5.2

NIH NLM NCBI Gene database

PDBePISA

PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC

RCSB PDB

SnapGene 7.2 by Dotmatics

The Human Protein Atlas

UniProt database

4 Experimental methods

4.1 Recombinant protein production

All recombinant protein variants of human CA IX were heterogeneously expressed in competent cells in various experimental conditions which are summarized in Table 7.

Table 7: Summary of the experimental variables during recombinant protein expression for given CA IX protein variants.

Protein variant	*E.coli* strain[a]	Culture medium[b]	Plate diameter [Ø cm][c]	Incubating system[d]	Exp. time[d]
ΔPG-CAIX/6	BL21 (DE3)	LB	9	42/42R	4 h
ΔPG-CAIX/4	NiCo21 (DE3)	LB	9/20	42/42R	4 h
ΔPG-CAIX/5S	NiCo21 (DE3)	LB	9	42/42R	4 h
ΔPG-CAIX /5G	NiCo21 (DE3)	LB	9	42/42R	4 h
CAIX/6	BL21 (DE3)	LB	9/20	42/42R	4 h
		MM	9	42/42R	4 h
CAIX/4	NiCo21 (DE3)	LB	9/20	42/42R	4 h
		MM	20	LEX-48	O/N
CAIX/5S	NiCo21 (DE3)	LB	9	42/42R	4 h
CAIX/5G	NiCo21 (DE3)	LB	9	42/42R	4 h

[a] Two *E. coli* competent strains were used for the expression; BL21 (DE3) and NiCo21 (DE3), see Table 5.

[b] The cells were cultivated either in LB Broth (LB) or in minimal medium (MM) based on the experiment purpose (see chapter 3.3.1 for media composition).

[c] Initial incubation after transformation was either on smaller Petri dish plates with 9 cm diameter, or at bigger 20 cm diameter plates.

[d] Majority of the experiments was carried out in 42/42R incubator with the protein expression for 4 hours, one exception was performed in LEX-48 bioreactor with expression overnight (O/N).

4.1.1 Transformation and cultivation

The host bacteria were *E. coli* strains, either BL21 (DE3) for expression of protein variants ΔPG-CAIX/6 and CAIX/6, or NiCo21 (DE3) for all the others (see Table 7). For each protein expression experiment the competent cells were freshly transformed by adding 1 µL (~50-100 ng) of the relevant plasmid into 100 µL of bacteria suspension, followed by incubation and heat shock. For BL21 (DE3) transformation, the initial incubation on ice was 45 min with consequent heat shock by 90 s in water bath (42°C) and 2 min on ice. NiCo21 (DE3) transformation was performed according to the 'High Efficiency Transformation Protocol' (C2529, NEB, USA) by 30 min ice incubation, 10 s water bath heat shock and 5 min on ice. Both types of the transformed bacteria suspension were then inoculated into 900 µL of LB media without any antibiotics and incubated at 37°C and 200 rpm for 1 h in Sanyo CO2 incubator. Next, 300 µL or 100 µL of the suspension was spread on LB agar plate (35 g/L of LB agar with 0.1 g/L ampicillin and 1% w/v glucose) with 20 cm or 9 cm diameter respectively, and incubated overnight in 37°C inside Sanyo CO2 incubator.

4.1.1.1 Cultivation in LB medium

On the next day, the bacterial colonies were washed from the plate by 12 mL / 4 mL of LB broth medium (20 g/L) with ampicillin (0.1 g/L). The bacterial suspension optical density was measured as OD_{550nm} on the Ultrospec 3100 Pro Amersham Biosciences. Varying volume of the bacterial suspension, according to the measured $OD_{550nm,}$ was inoculated into two-liters Erlenmeyer flask with 0.5 L of LB broth with ampicillin (0.1 g/L), glycerol (0.43%, v/v) and zinc sulfate heptahydrate (8.6 mg/L), so the final OD_{550nm} of the culture would be ~0.1 at the starting point. The prepared cultures were incubated at 37°C and 200 rpm in New Brunswick Innova® 42/42R incubator and their OD_{550nm} was measured until it reached ~0.6-0.7, the following procedure is described in chapter 4.1.2.

4.1.1.2 Cultivation in minimal medium

Two protein variants were produced by bacteria cultivation in minimal medium (MM composition in chapter 3.3.1), during which they were isotopically labelled for more advanced NMR experiments:

- CAIX/6 as so-called double-labelled sample with both ^{15}N and ^{13}C isotopes, for which the MM was supplemented with ^{15}N-ammonium sulfate ($^{15}N_2$, 99%) and ^{13}C-glucose ($^{13}C_6$, 99%) = $^{15}N/^{13}C$-CAIX/6,
- CAIX/4 as ^{15}N-labelled protein from MM supplemented only with ^{15}N ammonium sulfate ($^{15}N_2$, 99%) = ^{15}N-CAIX/4.

The overnight colonies were washed from the plate by 12 mL / 4 mL of minimal medium (composition in chapter 3.3.1). The bacterial suspension optical density was measured as OD_{550nm} on the Ultrospec 3100 Pro Amersham Biosciences. For both protein variants the bacterial suspension was inoculated so the final OD_{550nm} of the culture would be ~0.05 at the starting point. CAIX/6 bacterial suspension was inoculated into two-liters Erlenmeyer flask with 0.5 L minimal medium and cultivated at 37°C and 200 rpm in New Brunswick Innova® 42/42R incubator until the OD_{550nm} reached ~0.6-0.7. CAIX/4 was inoculated into 1.5 L minimal medium in flask and incubated in Epiphyte3 LEX-48 bioreactor until the OD_{550nm} ~0.6-0.7 was achieved.

4.1.2 Expression and isolation

When the OD_{550nm} reached the values ~0.6-0.7, the bacterial culture was transferred into New Brunswick Innova® 42/42R incubator or Epiphyte3 LEX-48 bioreactor set to 30°C. Subsequently, the protein expression was induced by adding ETG to final concentration of 0.4 mM, and $ZnSO_4$ x 7 H2O to final concentration of 1 mM. The bacterial culture was then left to grow for 4 h, at 30°C and 200 rpm in New Brunswick Innova® 42/42R incubator, or overnight at 18°C in LEX-48 bioreactor system (specified for each protein variant in Table 7).

The bacterial biomass was isolated by centrifugation (4000 rpm, 20 min, 4°C, J-LITE JLA 9.1000 rotor), then resuspended, centrifuged again (4000 rpm, 20 min, 4°C, J-LITE JLA 16.250

rotor) and stored in -20°C until it was further processed. The frozen pellet was resuspended in 10 mL per 1 L of bacterial culture of WB1 (100 mM Tris-HCl, pH 9, 200 mM Na_2SO_4 x 10 H_2O) with lysosome (40 mg / 1 L of bacterial culture) on magnetic stirrer overnight at 4°C. The next day, the buffer volume was adjusted up to 50 mL per 1 L of bacterial culture with WB1 and stirred for 1 h. The cells were then disrupted either by sonication (Soniprep 150 MSE), or by cell homogenizer (Avestin Emulsiflex C3) and centrifugated (12 000 rpm, 90 min, 4°C, J-LITE JLA 16.250 rotor). The final supernatant solution served as the sample for the first purification step described in 4.1.3.3, the pellet was stored at -20°C. The protein expression process was analyzed by SDS-PAGE electrophoresis described in chapter 4.1.3.1.

4.1.3 Purification and biochemical analysis

Each protein variant underwent different purification steps all of which were analyzed by SDS-PAGE electrophoresis and are designated in Table 8.

Table 8: Specification of different purification steps for each protein variant.

Protein variant	Initial capture[a]	Intermediate purification[b]	Final step or analysis[c]
ΔPG-CAIX/6	SAC	-	SEC
ΔPG-CAIX/4	SAC	-	SEC
ΔPG-CAIX/5S	N/A	N/A	N/A
ΔPG-CAIX/5G[d]	SAC	AEC	SEC
CAIX/6	SAC	AEC	SEC
CAIX/4	SAC	-	SEC
CAIX/5S	SAC	-	SEC
CAIX/5G	SAC	-	SEC

[a]The initial capture of the CA IX protein was performed by sulfonamide-affinity chromatography (SAC).

[b]Two variants were intermediately purified by anion exchange chromatography (AEC).

[c]All protein samples were ultimately purified and analyzed by size exclusion chromatography (SEC).

[d]Purification of protein variant ΔPG-CAIX/5S was not performed (N/A).

4.1.3.1 SDS-PAGE electrophoresis

Throughout the process of recombinant protein production, the samples were analyzed by sodium dodecyl sulfate–polyacrylamide gel discontinuous electrophoresis in Cytiva SE 250 Mini-Vertical Unit in the presence of running buffer (20 mM Tris pH 8.3, 200 mM glycine, 0.1% (w/v) SDS) and supplied by Consort EV243 voltage source, typically for ~15 min at 100 V and ~45 min at 250 V. The gels prepared in Cytiva Mighty Small Multiple Gel Caster were of size 80 x 70 x 0,75 mm and consisted of stacking (5% AA) and running gel (15% AA) which composition are reported in chapter 3.3.2.

Sample preparation

All samples were combined with non-reducing sample buffer (125 mM Tris pH 6.8, 20% (v/v) glycerol, 4% (w/v) SDS, 0.2% (w/v) bromfenol blue). The samples for protein expression analysis were prepared from 0.5 mL of bacterial culture retrieved from the cultivation just before the induction, and after 4 h at the end of the protein expression. Both were centrifugated (12 000xG, 5 min, RT) and the medium supernatant was discarded. The cellular sample before induction was then prepared by addition of 40 µL sample buffer to the bacterial pellet. The bacterial pellet retrieved at the end of the protein expression was resuspended in 0.5 mL of PBS, sonicated (Soniprep 150 MSE) and centrifugated (12 000xG, 10 min, 4°C). The resulting supernatant was recovered, and 40 µL of it was mixed with 20 µL of the sample buffer, the bacterial pellet was resuspended in 40 µL of the same buffer. All the remaining samples for the purification analysis were prepared by mixing the protein solution with the sample buffer in ratio 2:1, typically 20 µL with 10 µL.

The SDS-PAGE samples were then heated for 5 min at 98°C in Labnet International D1100 AccuBlockTM Digital Dry Bath and centrifugated in Eppendorf MiniSpin (13 000xG, 30 s). Finally, they were applied to the gel together with the Unstained Protein Standard, Broad Range (10-200 kDa), or SpectraTM Multicolor Broad Range Protein Ladder when the electrophoresis gel was meant for further analysis by western blot (described in chapter 4.1.3.2). After the run, the gel was stained either by silver staining method or by SimplyBlue™ SafeStain, based on the aim of analysis or protein concentration and visualized by Azure 600 imaging system.

Silver staining

The silver staining of the SDS-PAGE gel was done by several steps and the precise composition of the used solutions is in chapter 3.3.2. First, the proteins were fixed for 30 min in the fixative solution 1, and then three times washed for 5 min with the fixative solution 2. Next the gel was rinsed in the sensitizer solution for 30 s and shortly washed three times with distilled water before it was incubated for 20 min in the stain solution, again shortly washed three times with distilled water, to be finally visualized by the developer solution. Once the protein bands were clearly visible the reaction was stopped by change into the stop solution.

Coomassie G-250 staining

The visualization by Coomassie G-250 was done by rinsing the gel in the SimplyBlue™ SafeStain solution for 30 min. The background shade was then destained by washing the gel in 25% (v/v) methanol and 10% (v/v) acetic acid.

4.1.3.2 Western blot with immunodetection

The recombinant protein expression of human CA IX was verified by western blot analysis together with immunodetection by a specific antibody. The samples were prepared by SDS-PAGE

analysis as described in chapter 4.1.3.1 and then transferred onto nitrocellulose membrane by Bio-rad Trans-blot Turbo system in Trans-blot Turbo Transfer Buffer with 20% (v/v) methanol.

Once the proteins were transferred, the membrane was blocked in 5% milk in PBS for at least 45 min, and then washed in PBS for 3x10 min. The primary antibody was mouse IgG M75 Fab fragment specific against PGEEDLP sequence in the PG domain, it was diluted 35 000x in a staining solution (PBS with 1% (w/v) BSA, 0.05% (v/v) Tween), in which the membrane was incubated for 2 h. The membrane was then again washed for 3x10 min in PBS and incubated for 1 h in the presence of 4 000x diluted secondary antibody, which was rabbit-anti-mouse conjugated with horseradish peroxidase. Before the final visualization the membrane was yet again washed 3x10 min in PBS, then splashed by 0.75 mL of chromogenic TMB substrate which is converted by the horseradish peroxidase into a colorful product. The result was visualized by Azure 600 imaging system.

4.1.3.3 Affinity chromatography

The first purification step is based on the capture of carbonic anhydrase via the affinity of the enzyme active site to a deprotonated sulfonamide group bound to a resin. A gravity-flow column was prepared with *p*-aminomethylbenzenesulfonamide–agarose (commercially bought by Sigma-Aldrich or kindly given by doc. RNDr. Mgr. Jozef Hritz, Ph.D. from Masaryk University, Brno), equilibrated by distilled water of 10times the column volume (10 CV) and 10 CV of WB1. The sample of soluble proteins delivered from the final step in selection 4.1.2 was loaded to the sulfonamide-column while the flow-through fraction (FT) was collected. The column was then washed by 10 CV of WB1 and WB2 (specified in chapter 3.3.2), both fractions were collected (W1, W2, respectively). The elution was performed by addition of 1 CV EB (specified in chapter 3.3.2) at least 15 times, producing 15 fraction samples. Based on the analysis by SDS-PAGE the fractions with CA IX protein construct were combined and dialyzed for further purification or storage in -80ºC.

The columns were used repeatedly and regenerated after each usage by extensive washing with affinity regeneration solutions defined in section 3.3.2. The resin inside the column was washed with 10 CV of distilled water, 10 CV of AR1 solution, 10 CV of distilled water, 10 CV of AR2, and stored in 4ºC.

4.1.3.4 Dialysis, protein concentration measurement and adjustment

The protein samples after the affinity chromatography were dialyzed to either an appropriate buffer for anion exchange chromatography (20 mM triethanolamine pH 7.3), or to protein buffer A (20 mM Tris-HCl pH 7, 100 mM NaCl) for storage or SEC analysis. At all times, the dialysis was performed in 4ºC during spinning on a magnetic stirrer, the buffer was changed three times with each dialysis for at least 5 h. The specific dialysis utensils differed based on the sample volume. For smaller volumes it was either Slide-A-Lyzer® G3 Dialysis Cassette with 15 mL capacity or Slide-A-Lyzer®

Dialysis Cassette (Extra Strength) with 0.5-3 mL, both with 10 kDa cut-off. Samples of higher volume were dialyzed in Dialysis tubing, high retention seamless cellulose tubing with 12 400 Da pore size.

During the purification process the protein concentration was frequently measured on NanoDrop™ One/OneC and calculated from the given values of absorbance at 280nm and absorbance coefficient for each protein variant. The protein concentration [mg/mL] was then recalculated to molar concentration [μM] based on the protein molecular weight (MW). Both absorbance coefficients and protein molecular weight were theoretically calculated by ExPASy – ProtParam tool and are summarized in Table 9 for each protein variant.

Table 9: Values of absorbance coefficient [g/L] and molecular weight [Da] for each protein variants theoretically calculated by ExPASy – ProtParam server.

Protein variant	Ab. coefficient [g/L]	MW [Da]
ΔPG-CAIX/6	1.297	28181.8
ΔPG-CAIX/4	1.248	28107.7
ΔPG-CAIX/5S	1.247	28124.7
ΔPG-CAIX/5G	1.248	28094.7
CAIX/6	0.939	38931.98
CAIX/4	0.903	38858.89
CAIX/5S	0.902	38874.89
CAIX/5G	0.903	38844.86

When the protein concentration needed to be adjusted, the solution was subjected for ultrafiltration on centrifuge filter Amicon® Ultra Centrifugal Filter with 10 kDa cut-off. Based on the sample volume the centrifugation time and speed varied from 4 000xG in the case of 4 mL capacity amicon, to 13 800xG for 0.5 mL capacity.

In specific cases of ΔPG-CAIX/6 and CAIX/6 protein variants, when change into a different buffer was required, they were dialyzed into protein buffer B or protein buffer C (see 3.3.2) for crystallization optimization by the same dialysis process as noted above. The sample $^{15}N/^{13}C$-CAIX/6 was changed into NMR buffer (see 3.3.2) by 6-times dilution and concentration on 4 mL centrifuge filter amicon.

4.1.3.5 Anion exchange chromatography

Two protein variants (ΔPG-CAIX/5G and CAIX/6) were intermediately purified by anion exchange chromatography method (AEC) on MonoQ™ 5/50 GL Tricorn MonoBeads™ Column and Amersham Biosciences AKTA FPLC system in 20 mM TEOA buffer system (defined in 3.3.2). Before the purification experiment, the conductivity was measured for the dialysis buffer, AEC equilibration buffer with 10% of AEC elution buffer (20 mM TEOA, 1 M NaCl), and protein sample. If the protein sample conductivity was higher than the AEC buffer, it was diluted until it was lower.

The column was equilibrated in the AEC equilibration buffer together with 10% gradient of AEC elution buffer (20 mM TEOA, 0.1 M NaCl), the sample was manually loaded into the instrument and injected onto the column, then washed with the same buffer system. The elution was performed by gradient of AEC elution buffer (thus NaCl increasement) in four segments with different buffer volume [CV] as defined in Table 10. During the experiment the flow rate and the maximal pressure followed the manufacture instruction with values of 1 mL/min and 4 MPa, respectively. The elution profile was spectrophotometrically tracked by absorbance of 280 nm [mAU] and plotted together with retention volume, elution buffer percentage or concentration, and conductivity. The plot was analyzed in Cytiva Unicorn Evaluation software.

Table 10: The method employed for protein elution during AEC with definition of each segment of growing percentage of AEC elution buffer and consequent increase of NaCl concentration and lapse of volume that passed through the system.

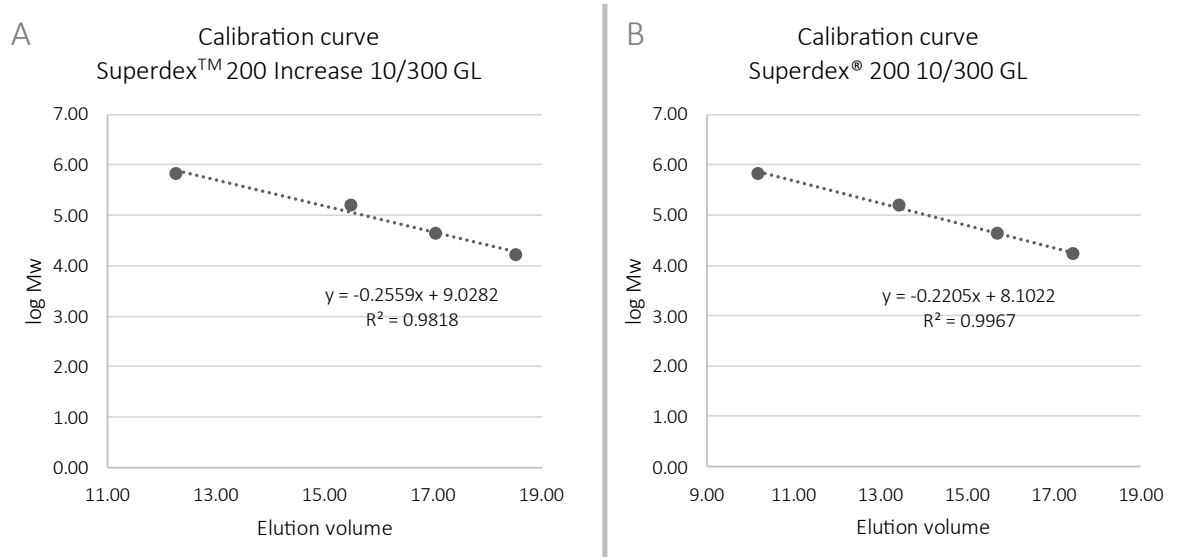
Segment	AEC elution buffer %	NaCl concentration	Volume of buffer
1	10-50%	0.1-0.5 M	20 CV
2	50-100%	0.5-1 M	10 CV
3	100%	1 M	5 CV
4	100-10%	1-0.1 M	5 CV

4.1.3.6 Size exclusion chromatography

The final polishing step of purification served also as an analytical experiment for analysis of quaternary protein structure as it was done on size exclusion chromatography column Superdex^{TM} 200 Increase 10/300 GL combined with ÄKTA pure™ 25 system with the ability for separation of protein of sizes within 10 000 to 600 000 Da range. The separation of proteins based on their size was performed in the presence of protein storage buffer A (20 mM Tris-HCl pH 7, 100 mM NaCl). The procedure followed manufacture instruction with flow rate 0.75 mL/min and maximum delta-column pressure 2.6 MPa. The column was washed with distilled water and the protein storage buffer A, both with volume of 1 CV, the sample was manually loaded into the system, then the column was equilibrated for 0.1 CV in the same buffer, the sample was injected onto the column. The fractions were eluted by isocratic elution of 1.1 CV and collected at fixed fraction volume 1.00 mL into deep-well block.

For protein variants ΔPG-CAIX/6 and ΔPG-CAIX/5G the size exclusion chromatography was performed mainly as an analytical step of already purified protein samples on Superdex® 200 10/300 GL connected to Amersham Biosciences AKTA FPLC system. The procedure was performed comparably to the described above, only it was at flow rate of 0.5 mL/min, within PBS buffer and fractions collection into microtubes.

The Gel Filtration Standard #1511901 Bio-Rad was subjected to both analysis and the resulting calibration curves (Figure 8A,B) were used for approximate calculation of MW of the CA IX protein variants. The elution profile was spectrophotometrically tracked by absorbance of 280 nm [mAU] and plotted with retention volume [mL], the plots were analyzed in Unicorn Cytiva Evaluation software.



Biorad #1511901	Elution V [ml]	MW [Da]	logMw
Thyroglobulin (V_0)	12.27	670000	5.83
γ-globulin	15.50	158000	5.20
Ovalbumin	17.05	44000	4.64
Myoglobulin	18.53	17000	4.23
Vitamin B12	20.92	1350	3.13

Biorad #1511901	Elution V [ml]	MW [Da]	logMw
Thyroglobulin (V_0)	10.18	670000	5.83
γ-globulin	13.42	158000	5.20
Ovalbumin	15.69	44000	4.64
Myoglobulin	17.44	17000	4.23
Vitamin B12	20.39	1350	3.13

Figure 8: Calibration curves based on size exclusion chromatography with MW Standard #1511901 Bio-Rad. The given equations were used for calculation of apparent molecular weight of CA IX protein variants. The values of Vitamin B12 were omitted from the calibration because such small molecules are not properly separated on the columns used herein.

4.2 Biophysical characterization

4.2.1 NMR spectroscopy

All protein variants were measured for one-dimensional hydrogen (1D ^{1}H) spectrum which gives general information about the protein folding state. The 2D $^{15}N/^{1}H$ correlation experiments (set of HSQC; Heteronuclear Single Quantum Coherence) of the isotopically labelled samples $^{15}N/^{13}C$-CAIX/6 and ^{15}N-PG-CAIX/4 were used for further analysis. The cross-correlated relaxation experiment TRACT (Lee *et al*, 2006; Robson *et al*, 2021) results in calculation of rotational correlation time, which is proportional to the MW of a macromolecule, therefore can be used for estimation of the multimeric state of ^{15}N-labelled samples.

The NMR protein samples contained 5% D_2O and were measured in a 3 mm tube (non-labelled samples) or in 5 mm Shigemi type tube (isotopically labelled samples). All data were collected at 25ºC (298 K), on 850 MHz Bruker Avance spectrometer equipped with a triple-resonance (15N/13C/1H)

cryoprobe together with Bruker's TopSpin™ software. The NMR spectroscopy experiments were done in collaboration with Mgr. Pavel Srb, Ph.D.

4.2.2 Mass photometry

Mass photometry method was used for accurate mass measurement to resolve the quaternary protein structure. In principle, the macromolecules approach a glass surface and are illuminated by a laser causing light scattering which interferes with the light reflected at the interface, the resulting interface contrast scales linearly with the mass of the biomolecule. The ratiometric contrast is converted into the molecular mass based on a calibrant (β-amylase protein of known MW). The result is displayed as histogram of the macromolecule mass [kDa] and counts (number of particles of given MW that were measured) (Soltermann *et al*, 2020; Wu & Piszczek, 2020). The measurable mass range relies between 30 kDa and 5 MDa. The experiment was measured by TwoMP-207 with AcquireMP 2023 R2 and analyzed on DiscoverMP 2023 R2 during Refeyn workshop at IOCB, Prague (October 2023).

Two protein variants CAIX/6 and CAIX/4 were analyzed by mass photometry method, with the experimental specification defined in Table 11.

Table 11: Experimental conditions of mass photometry experiments for analyzed protein variants.

Protein variant	Buffer	Aprox. concentration
CAIX/6	25 mM ammonium acetate pH 7.3	30 nM
CAIX/4	20 mM Tris pH 7, 100 mM NaCl	30 nM

4.2.3 Differential scanning fluorimetry

The differential scanning fluorimetry (DSF) method was used to identify the stabilizing buffer during crystallization optimization of protein variants ΔPG-CAIX/6 and CAIX/6 and partially followed protocol described before (Písačková *et al*, 2013). The experiment was measured by real-time PCR Roche LightCycler® 480 II machine in 96-well plate covered by LightCycler® 480 Sealing Foil. Each well contained 25 µL of buffer solution together with protein and fluorescent dye SYPRO® Orange, described in detail in Table 12. Number of buffers at various pH values with or without salt were analyzed, as the controls the original protein buffer (20 mM Tris pH 7, 100 mM NaCl for ΔPG-CAIX/6, 20 mM TEOA pH 7.3, 100 mM NaCl for CAIX/6) and distilled H_2O were measured as well. The instrument was set up as 'melting curves' mode, with fluorescence measurement of SYPRO® Orange (excitation at 465 nm, emission at 580 nm) during gradient increase of temperature from 20ºC to 95ºC with ramp rate of 0.01 ºC/s and 40 acquisitions per ºC. The experiment was analyzed in Roche LightCycler® 480 SW 1.5 real-time PCR software and Microsoft Excel.

List of buffers tested in DSF experiment

Buffer		pH	
Sodium Acetate	4.5	5.0	
Sodium Citrate	4.7		
Monopotassium phosphate	5.0	6.0	7.0
Monosodium phosphate	5.5	6.5	7.5
MES	5.8	6.2	6.5
HEPES	7.0	8.0	
Ammonium acetate	7.3		
Tris	7.5	8.0	8.5
Bicin	8.0	9.0	

Table 12: Content of one well within the 96-well plate for DSF experiment.

Volume [µL]	Concentration / dilution	Component	Final concentration / dilution
12.5	200 mM (400 mM)[a]	buffer system (NaCl)[a]	100 mM (200 mM)[a]
1	25x	SYPRO Orange	625x
2.5	~ 1 mg/mL	protein	0.1 mg/mL
9	-	distilled water	-

[a]All buffers were analyzed with and without natrium chloride (stated in brackets).

4.3 Structure determination by X-ray crystallography

4.3.1 Crystallization by vapor diffusion

Five protein variants were crystallized by vapor diffusion method in numerous setups which differed in protein concentration, drop arrangement (sitting or hanging drop) and ratio between protein and precipitant (1:1, 2:1, 1:2). The initial experiments were crystallization screens that were followed by various optimization steps in case of two protein variants. The summary of crystallization for each variant is in Table 13, the individual experimental procedures will be further described in detail.

All crystallization experiments were carried out in 18ºC, and positive crystallization events were evaluated by optical microscope Olympus SZX10 with photo camera Olympus E-620 and QuickPHOTO CAMERA 2.3, or by UV-fluorescent microscope JANSi UVEX-m imaging system with Crystal detect software. Before any crystallization experiment, the protein was shortly incubated with at least 2-times excess [µM] of an inhibitor and then centrifuged (13 000xG, 10 min, 4 min).

Table 13: Summary of crystallization experiments employed for each protein variant.

Protein variant	Protein conc. [mg/mL]	Method	Screening	Optimization
ΔPG-CAIX/6	7-10	sitting drop hanging drop	JSCG Plus, Hampton Research Morpheus® Mol Dimensions	buffer change precipitant composition
CAIX/6	3-18	sitting drop hanging drop	JSCG Plus, Hampton Research Morpheus® Mol Dimensions The PEGs Suite, Qiagen SG1, MiTeGen	microseeding streak seeding additive screen precipitant composition buffer change
CAIX/4	16.5	sitting drop	JSCG Plus, Hampton Research	-
CAIX/5S	16	sitting drop	JSCG Plus, Hampton Research	-
CAIX/5G	5.5	sitting drop	JSCG Plus, Hampton Research	-

4.3.1.1 Sitting drop

The first crystallization experiment performed with any protein variant was screening executed by crystallization robot Oryx8 Douglas Instruments into MRC 3-well low plate. Any well contained up to three drops, each of 0.3 µL volume with different protein to precipitant ratio, and a reservoir with 30 µL of a precipitant, as schematically depicted in Figure 9. The commercial crystallization screens with 96 precipitant conditions that were used are stated in Table 13. The addition of the precipitants into the reservoirs was done by robot Crystal Gryphon Art Robbins Instruments. Once the whole experiment was set up the crystallization plate with drops was covered by ThermalSeal RTS™.

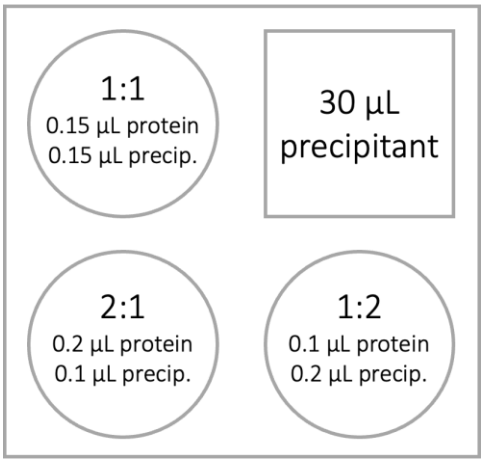


Figure 9: Layout of one well within the 96-well MRC 3-well low crystallization plate for initial screening experiment. The reservoir was filled with 30 µL of precipitant reagents of given commercial screen. Each 3 µL drop was of different protein to precipitant ratio composed of the designated volumes.

4.3.1.2 Hanging drop

The hanging drop method was used as optimization step for the ΔPG-CAIX/6 and CAIX/6 variants. The drops of 3 µL volume were prepared manually into the EasyXtal® 15-Well Tools plate with X-Seal crystallization supports. The drops were arranged of protein to precipitant ratio of 1:1, 2:1 or 2:1. The precipitant volume in the reservoir was 500 µL and the solution was mixed manually in all instances, the precise composition of the condition differed and will be described only for the positive crystallization results in chapter 5.4.2.1.

4.3.1.3 Crystallization optimization

Microseeding

The microseeding procedure for CAIX/6 followed the protocol 'Procedure for Making the Seed Stock' by Douglas Instruments. The seed stock was prepared from three different positive crystallization events obtained either from the sitting or the hanging drop experiment. The microseeding was performed as the sitting drop described in chapter 4.3.1.1 with the following adjustments:

- seeding experiment 1: all reservoirs were filled with 30 µL of chemical solution from tube #1-26 in JCSG Plus screen (1 M LiCl, 0.1 M citrate pH 4, 20% (w/v) PEG 6000)
- seeding experiment 2: reservoirs were filled with 96 conditions from JSCG Plus screen
- the different ratios in each well were setup as showed in Figure 10.

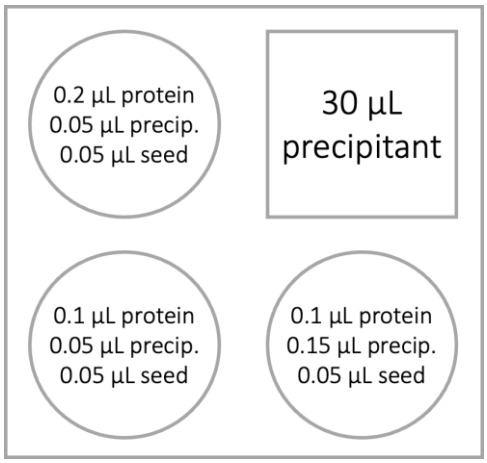


Figure 10: Layout of one well during the microseeding crystallization optimization experiment. The reservoirs were filled with either one precipitant condition throughout the whole 96-well plate, or with 96 different precipitant reagents from the JCSG Plus screen. To explore numerous crystallization conditions the volumes of the protein, precipitant and seed stock varied as stated.

Streak seeding

A different type of crystallization seeding was performed for the same CAIX/6 protein variant in the hanging drop setup. The drops were created similarly to the procedure described in chapter 4.3.1.2 with the reservoir filled with 500 µL of solution containing 1 M LiCl, 0.1 M citrate pH 4,

20% PEG 6000. The streak was utilized by a cat whisker which was first rubbed into a drop with positive crystallization events. The crystallization nuclei were then introduced by passing the cat whisker along a straight line through the newly established hanging drop. The streaking was done at two different timepoints after the drops were established – immediately or 8 hours after. As a control, one drop in the plate seal was left without any seeding.

Additive screen

The crystallization optimization for CAIX/6 with additive screen was performed in the sitting drop setup as described in chapter 4.3.1.1 with the following adjustments:

- all reservoirs were filled with 30 µL of chemical solution from tube #1-26 in JCSG Plus screen (1 M LiCl, 0.1 M citrate pH 4, 20% (w/v) PEG 6000)
- each drop contained:
    - 0.1 µL of CAIX/6 protein sample (3 mg/mL)
    - 0.17 µL of the precipitant from the reservoir
    - 0.03 µL of reagents from the commercial additive screen Additive Screen HT™ Hampton Research.

Precipitant condition

Since the initial hit from the crystallization screening of CAIX/6 was from precipitant condition 1 M LiCl, 0.1 M citrate pH 4, 20% PEG 6000 from the commercial JCSG Plus screen, such a solution was prepared manually along with various adjustment of pH values (4-6), salt concentration (0.5-1 M), PEG percentage (15-20%) and type (6000, 8000). These reagents were used as precipitant for crystallization of ΔPG-CAIX/6 and CAIX/6 in the hanging drop arrangement (see 4.3.1.2).

Buffer change

Based on the results from DSF experiment the protein variants ΔPG-CAIX/6 and CAIX/6 were transferred into two different buffers, either 100 mM MES pH 6.5, or 100 mM ammonium acetate pH 7.3 as described in chapter 4.1.3.4 and crystallized in the sitting drop with JCSG Plus screen as in chapter 4.3.1.1.

4.3.2 X-ray data collection and processing

Before the diffraction experiments, the crystals were mounted into MicroMounts™ loops with aperture size of 20-200 µm. Then flash-cooled and stored in liquid nitrogen inside Universal V1-Puck (Uni-Puck). Majority of the diffraction data were collected on MX14.1 beamline operated by the Helmholtz-Zentrum Berlin (HZB) at the BESSY II electron storage ring (Berlin-Adlershof, Germany) as standard data collection at 100 K and 0.9184 Å wavelength, with PILATUS3 S 6M detector

(Mueller *et al*, 2012, 2015; Gerlach *et al*, 2016). The datasets were automatically processed by XDSAPP (Sparta *et al*, 2016).

One dataset (CAIX/5G) was obtained by helical data collection, therefore it could not have been automatically processed during the data collection and it was processed by Ing. Petr Pachl, Ph.D. One dataset (CAIX/4 #1) was collected and processed by Ing. Petr Pachl, Ph.D. at Proxima-1 beamline of Synchrotron Soleil owned by the CNRS and the CEA Paris-Saclay (Saint-Aubin, France).

4.3.3 Structure determination and refinement

The structure determination was done in CCP4 software package within CCP4i2 suite (Agirre *et al*, 2023). The structures were solved by molecular replacement in MOLREP (Vagin & Teplyakov, 2010) using previously determined CA IX structure with PDB code 3IAI as the model. Structure refinement was performed by program REFMAC5 (Murshudov *et al*, 2011) together with manual adjustments in Coot (Emsley *et al*, 2010; Debreczeni & Emsley, 2012). The final structural model quality (sterics and geometry) was evaluated by MolProbity server (Williams *et al*, 2018). The data processing, refinement, and validation statistics are listed in Table 18 in results chapter 5.4.3.

4.4 *In silico* analysis

4.4.1 PDB database search

The Protein Data Bank (PDB) was searched on RCSB.org server (Berman *et al*, 2000; Burley *et al*, 2023) by Advanced Search Query Builder specified by the following queries:

Structural Attributes	Accession Code(s) - UniProt	is any of	Q16790
	Scientific Name of the Source Organism	is	Homo sapiens
	Scientific Name of the Source Organism	is NOT	Mus musculus
	Polymer Entity Mutation Count	=	0-6

The basic search by UniProt accession code Q16790 and source organism, without any definition of amino acid substitutions, was used to search for all published structures of human CA IX and to create the Table 3. For the investigation of amino acid substitutions within the structurally determined domain additional query was added (stated under the line), which value was different for each search, ranging from 0 to 6. These searches were used to create Table 4.

4.4.2 PDBePISA

The PDBePISA (Proteins, Interfaces, Structures and Assemblies) was utilized to investigate the CA IX probable multimeric state (assemblies) and interfaces between the subunits (Krissinel & Henrick, 2007; Krissinel, 2010). The PDB entries found as described in chapter 4.4.1 were used as inputs. Any PISA Query was performed as Structure Analysis with Auto Processing mode, when any

ligand, except for Zn ion, were excluded from the process. The result files were downloaded and analyzed in Microsoft Excel by conditional formatting to highlight the amino acid residues which were identified as solvent-inaccessible or interfacing or forming hydrogen bonds and salt bridges. The summary of this *in silico* analysis was used to propose a hypothesis about the dimer interface and for the design of amino acid substitutions in CA IX protein variants.

4.4.3 PyMOL

The protein structures determined either in this diploma thesis or published before were visualized and analyzed in PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC. A typical processing included these steps and commands:

- remove water molecules: `remove solvent`
- deletion of any other molecules besides protein chains and Zn ions
- select specific amino acid residues: `sele PDBentry/molecule/chain/resi+resi+resi`
- color coding according to the visualization purpose:
  - coloring secondary structures:
    - α-helix: `color colorvalue, ss h`
    - β-sheet: `color colorvalue, ss s`
    - loops: `color colorvalue, ss l+''`
  - selection colored for its characteristics: `color colorvalue, selection`
  - color according to atom type except for carbons: `util.cnc`
  - color according to AlphaFold pLDDT: `coloraf objectname`
- transparency adjustments:
  - surface display: `set transparency, 0.0-1.0, selection`
  - cartoon display: `set cartoon_transparency, 0.0-1.0, selection`
  - stick display:
    - `set stick_transparency, 0.0-1.0, objectname`
    - `set_bond stick_transparency, 0.0-1.0, selection`
- figures were rayed with transparent background at 150 DPI

To display the electron density maps corresponding map files were uploaded with consequent visualization by `isomesh name, map, level [,(selection) [,carve ]]` with level value = 1 (2*Fo-Fc* map countered at 1 σ), and carve = 70 (maps shown in radius of 70 Å around selection). The structural alignment was performed by `align` action within the internal GUI. The polar contacts between chains were identified by `find` action within the internal GUI.

5 Results

5.1 Design of CA IX protein variants

Based on the literature review, RCSB search for oligomeric state (Table 3) and some preliminary experiments, it was proposed that the six amino acid substitutions present on the surface of CA IX variant (herein named ΔPG-CAIX/6) expressed in *E. coli* system (Mahon *et al*, 2016; Koruza *et al*, 2019) hinder the biological dimeric arrangement. To design novel protein variants that may preserve the dimeric state, the 26 PDB entries of CA IX catalytic domain (Table 3 and chapter 4.4.1) were subjected for analysis by PDBePISA software for exploration of macromolecular interfaces as described in chapter 4.4.2. First two results for each entry from the interface list were downloaded and analyzed to identify the amino acid residues located in the interacting surface in most entries. The outcome sheets with the interfacing residues for all entries are not included in this thesis since they can be generated at any time at the PDBePISA server by subjecting the PDB entry code. The main result from the analysis, as shown in Table 14, was the definition of residues within the dimeric interface.

Table 14: Residues identified to be involved in formation of dimeric assembly.

Interfacing areas	Solvent-inaccessible residues	H-bonds residues		Salt bridge residues	
F160	S162	S174	S174	R221	E328
A172-L177	P163	R268	**A258**[a]	R159	E219
P216-Y220	N197			R268	E264
S256-G270	G236-H239				
N382-E387	H254				
	L279				
	S330				
	C336				
	L351				

[a]Residue A258 (bold) is substituted in CAIX/6 as A258K.

Some additional regions were identified to form the dimeric interface next to the well-defined H-bond between residues S174 within the A172-L177 which is a substitute for the disulfide bond in wild-type CA IX (Alterio *et al*, 2009). More importantly the interfacing area of residues S256-G270 overlays with two amino acid substitutions in ΔPG-CAIX/6 (C174S, L183S, A213K, A258K, A259Y, M350S). Namely the bulky lysine residue at position 258 and tyrosine at position 259 can obstruct the H-bond formation between arginine 268 with the main chain carbonyl of alanine 258 designated to form the biological assembly (Alterio *et al*, 2009). Therefore, amino acid substitutions were designed within this region, going back to wild-type amino acid at position 258 and 259 (CAIX/4), mutating A258 to small polar amino acid serin (CAIX/5S), and with wild-type A258 and F259 in addition to A260G substitution within the area (CAIX/5G). All variants were cloned into plasmid vector as gene variants expressing only the catalytic domain (ΔPG-CAIX/x, protein residues 137-391), or together with

the PG domain (CAIX/x, protein residues 38-391) as described in detail in chapter 3.1 and summarized in Table 6 with copy of the same table declared herein for clarity (not numbered).

Domains	Protein variant name	Amino acid substitutions
catalytic domain	ΔPG-CAIX/6	C174S, L183S, A213K, A258K, F259Y, M350S
	ΔPG-CAIX/4	C174S, L183S, A213K, M350S
	ΔPG-CAIX/5S	C174S, L183S, A213K, A258S, M350S
	ΔPG-CAIX/5G	C174S, L183S, A213K, A260G, M350S
PG & catalytic domain	CAIX/6	C174S, L183S, A213K, A258K, F259Y, M350S
	CAIX/4	C174S, L183S, A213K, M350S
	CAIX/5S	C174S, L183S, A213K, A258S, M350S
	CAIX/5G	C174S, L183S, A213K, A260G, M350S

5.2 Heterologous production of CA IX protein variants

5.2.1 Cultivation and expression

Eight variants of recombinant CA IX protein were expressed in *E. coli* competent strains BL21 (DE3) or NiCo21 (DE3) as stated in Table 7. The protocol is described in detail in chapter 4.1.1 and 4.1.2. Briefly, the transformed bacteria were cultivated at 37ºC until their OD_{550} reached value of ~0.6-0.7, when they were moved to 30ºC, and the protein expression was induced by addition of 0.4 M ETG and 1 mM $ZnSO_4$. ETG molecule (analog of lactose and IPTG) is an induction agent which leads to expression of T7 RNA-polymerase which then transcribes the gene of interest under the regulation of T7 promotor. Zinc sulfate was added to the medium as source for Zn^{2+} ion for the active site of CA IX catalytic domain.

Three examples of bacterial growth curves are depicted in Figure 11A, from which it is notable that the expression of CA IX protein variants is non-toxic as the bacterial growth was not ceased after the induction. Similar curves were observed for other protein variants including only the catalytic domain or both catalytic and PG domain. The cultivation of bacteria in minimal media for isotope-labelled samples led to increase time until the bacterial culture reached OD_{550} ~0.6-0.7 and the expression was induced, as shown for $^{15}N/^{13}C$-CAIX/6 in Figure 11A. For that reason the expression time for ^{15}N/-CAIX/4 protein was prolonged from four hours to over-night cultivation as claimed in chapter 4.1.1.2 and 4.1.2.

The protein expression was examined by SDS-PAGE electrophoresis of bacterial cells before and after the induction and of soluble fraction as shown for two examples in Figure 11B,C. Only in the case of ΔPG-CAIX/6 a band corresponding to the protein variant was visible in the soluble

fraction (Figure 11B). The heterologous expression of other protein variants in the soluble form was typically not confirmed by the SDS-PAGE experiment as shown on CAIX/6 example. However, the presence of protein variants bearing the PG-domain was detected by western-blot (WB) with immunodetection by M75 primary antibody, when protein band appeared in the soluble fraction as well as in the cellular pellet after induction (Figure 11C). A single band of CA IX was detected in the soluble fraction, but several other bands were evident in the cellular fraction after induction and for purified protein used as control due to extensive amount of protein load or non-specific interactions of the antibody. All of the protein variants were isolated and purified from the soluble fraction, except for ΔPG-CAIX/5S for which the purification procedure was not performed at all.

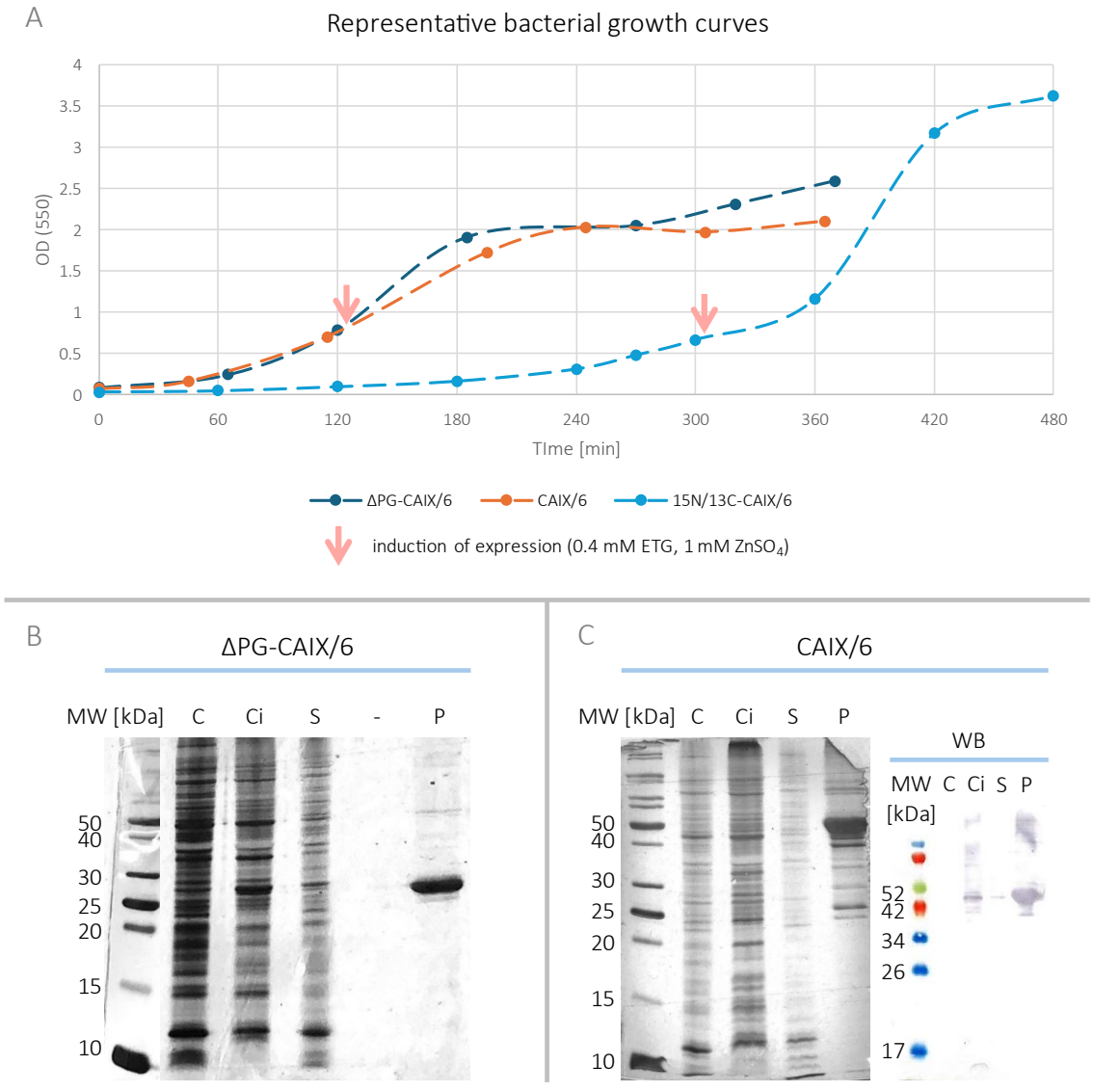


Figure 11: Recombinant expression of CAIX protein variants in E. coli *heterologous system. A) Representative bacterial growth curve given for ΔPG-CAIX/6 protein variant consisting only of the catalytic domain, CAIX/6 with both PG and catalytic domain, and sample $^{15}N/^{13}C$-CAIX/6 as example of cultivation in minimal media for preparation of isotope-labelled samples. B) SDS-PAGE gel analysis of ΔPG-CAIX/6 expression. C) SDS-PAGE gel and western-blot (WB) analysis of CAIX/6 expression. The abbreviations used in B) and C) stands for: C = bacterial cells right before induction, Ci = bacterial cells after induction at the end of protein expression, S = soluble fraction of bacterial cellular suspension, P = previously purified protein used as control.*

5.2.2 Isolation and purification

The isolation and purification procedure partially followed protocol published in (Mahon *et al*, 2016) and was optimized for each protein variant and the prospective experiments, the experimental procedures are described in detail in chapter 4.1.3. First, the soluble protein fraction was isolated by resuspending the bacterial pellet in WB1 followed by cell disruption and centrifugation. The soluble fraction was subjected for sulfonamide affinity column that selectively purify carbonic anhydrase, washed by buffers WB1 and WB2 and eluted by EB (solutions defined in chapter 3.3.2). The process was inspected by SDS-PAGE as shown for four examples in Figure 13.

Each depicted protein variant represents different purification strategies that were employed (schematic steps labelled A, B, C in Figure 12). Since the protein variants CAIX/6 and ΔPG-CAIX/5G were eluted from the SAC in fraction 2 and 3 together with some impurities (Figure 13A), these samples were further purified by anion exchange chromatography (as shown in chapter 5.2.2.1). On the other hand, ΔPG-CAIX/6 protein purity was satisfactory after SAC (Figure 13B) and thus only one-step purification protocol was employed. For any other protein variants (Figure 13C), the additional purification step was preparative size exclusion chromatography.

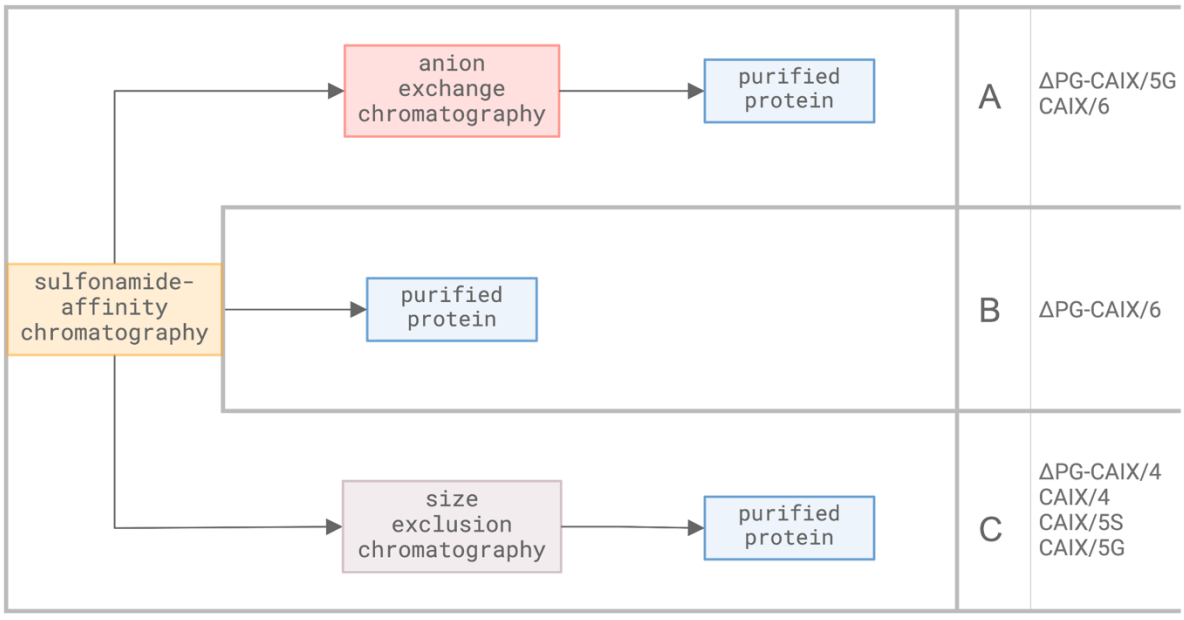


Figure 12: Schematic overview of various purification strategies used for preparation of protein variants within this thesis. The first purification step for all protein variants was affinity chromatography on sulfonamide column. A) Protein variants ΔPG-CAIX/5G and CAIX/6 were then purified by anion exchange chromatography. B) The first purification step was sufficient to prepare ΔPG-CAIX/6. C) Other protein variants were prepared by second step of size exclusion chromatography to obtain pure samples.

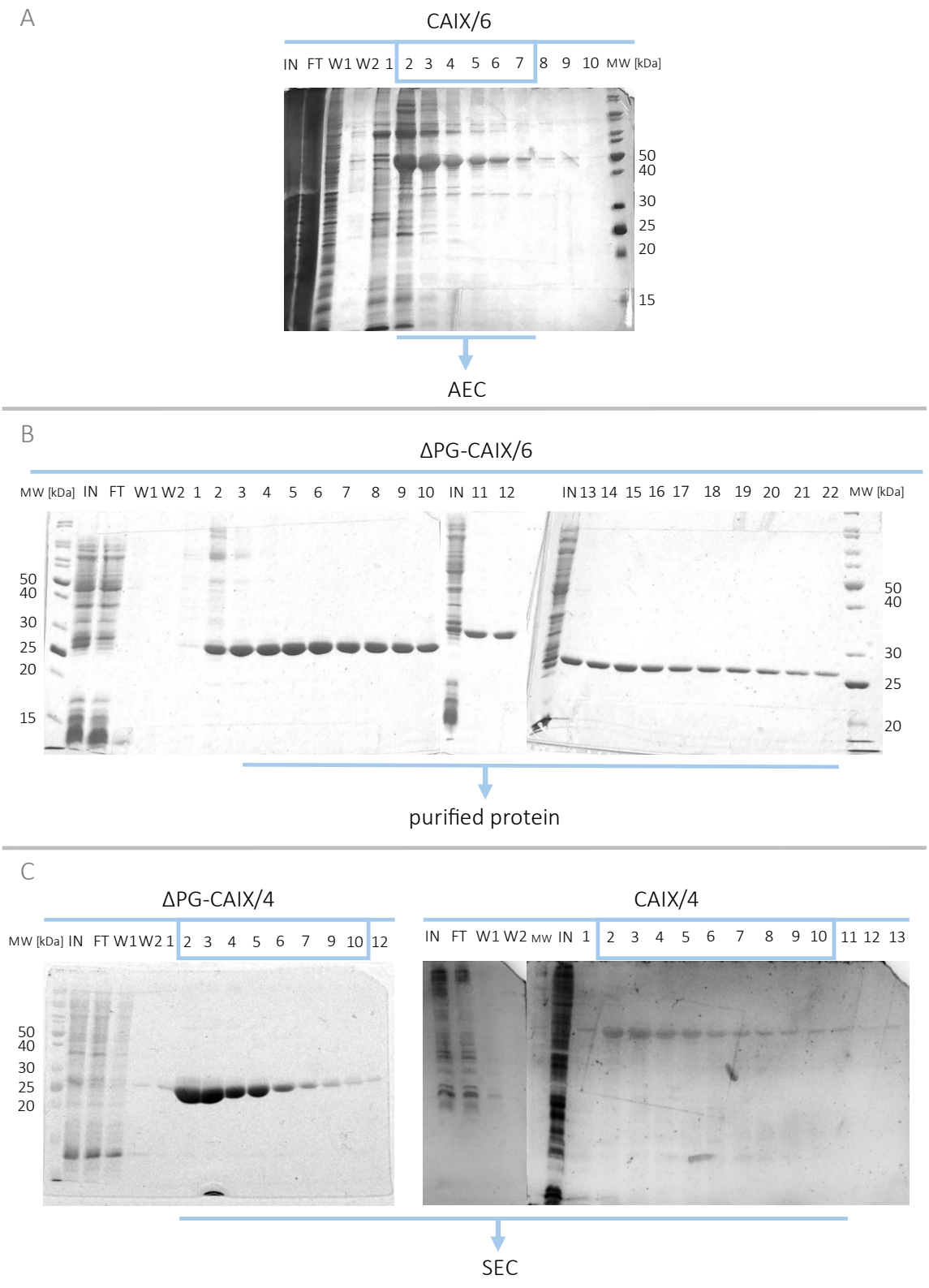


Figure 13: Representative SDS-PAGE gels analysis of first purification step by SAC. A) Fractions 2-7 clearly display protein variant CAIX/6 together with contaminants, therefore these were combined and further purified by AEC. B) Protein variant ΔPG-CAIX/6 is present in high purity within fractions 3-22 therefore these were combined and stored for further usage. C) Examples of protein variants that were extensively purified by the SAC but the combined fractions 2-10 were then subjected for SEC as the final purification step. The abbreviations used in A), B), C) stand for: IN = input to the column, FT = flow-through, W1 = wash by WB1, W2 = wash by W2. The fractions bordered in blue were combined and subjected for further steps. SDS-PAGE gels of CAIX/6 (A) and CAIX/4 (C) were visualized by silver staining, gels of ΔPG-CAIX/6 (B) and ΔPG-CAIX/4 (C) by Coomassie staining.

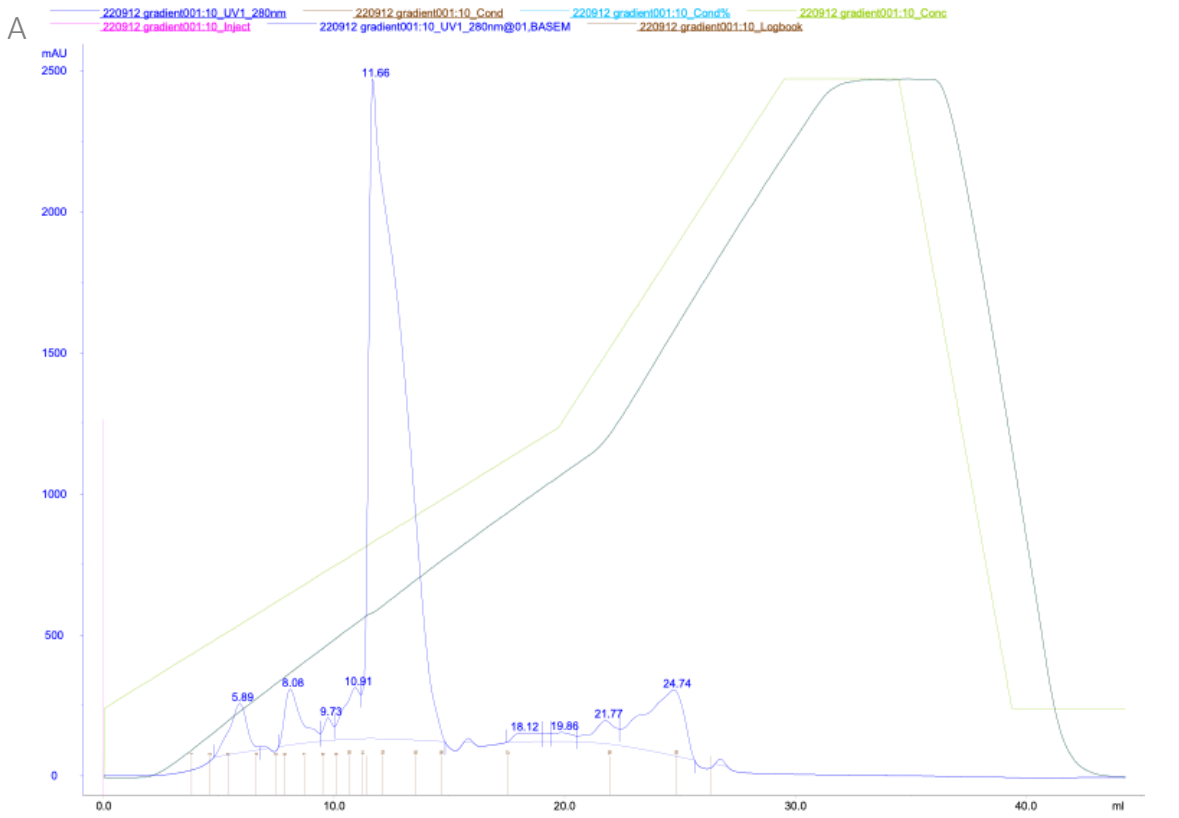
5.2.2.1 Anion exchange chromatography

Protein variants CAIX/6 and ΔPG-CAIX/5G were intermediately purified by anion exchange chromatography from the contaminants present in the sample after the first purification step (Figure 13A). Based on their isoelectric point (pI = 4.53 and 5.58, respectively) the buffer system for AEC was 20 mM triethanolamine pH 7.3 The chosen pH value is higher than the protein variant's pI therefore the protein molecules are negatively charged and bind to the positive amines in the MonoQ column resin. The sample was then eluted by natrium chloride gradient, when the bound proteins are exchanged for the ions as they compete for the binding sites. The optimized method is described in detail in chapter 4.1.3.5. Example of such a purification procedure is given for CAIX/6 as a respective chromatogram and subsequent SDS-PAGE analysis shown in Figure 14.

From the elution curve (Figure 14A), peak table (Figure 14B) and the gel (Figure 14C), it is evident that the CAIX/6 protein variant represents the majority of the sample as the highest peak (#5) with the retention volume of 11.66 mL is as high as 2339 mAU corresponding to the protein amount (fractions 13-15 on the gel). The contaminant proteins occurred mainly in the other fractions. Some moderate impurities can be seen in fraction 13 (Figure 14C), therefore this fraction was kept separate and fractions 14 and 15 were combined for further experiments in this batch. Analogous results were observed for other purification batches of CAIX/6, or ΔPG-CAIX/5G. In one instance of CAIX/6, when the fractions after AEC displayed major impurities, the sample was subjected to preparative SEC similarly to other protein variants, and for that the results will be discussed in chapter 5.3.1.

AEC: CAIX/6

A



B

No	Retention [ml]	Area [mAU*ml]	Height [mAU]
1	5.89	154.9226	175.38
2	8.08	164.5225	198.49
3	9.73	36.0192	80.685
4	10.91	145.8988	181.943
5	11.66	3843.629	2339.094
6	18.12	39.3218	31.031
7	19.32	10.4344	28.618
8	19.86	31.5826	31.886
9	21.77	93.1009	82.414
10	24.74	451.8702	234.739

C

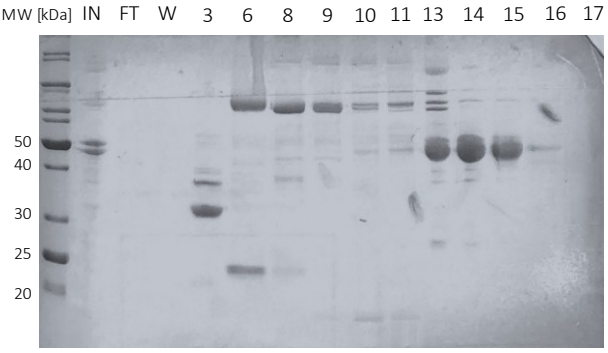


Figure 14: An example of anion exchange chromatography experiment given for CAIX/6. A) AEC elution curve with the major peak of 11.06 mL elution volume with the CAIX/6 protein variant. B) Table corresponding to peaks with values of retention volume, area under the peak and its height. C) SDS-PAGE analysis with silver staining for purification by AEC; the protein variant CAIX/4 is present in fractions 13-15. Fraction 13 was kept separately because it still contains some impurities, fractions 14 and 15 were combined and stored before further usage. The abbreviations used in C) stand for: IN = input to the column, FT = flow-through, W = column wash.

5.2.2.2 Purified protein variants and yield

Final protein samples that underwent any of the stated purification routes were dialyzed into relevant protein buffer and concentrated as denoted in chapter 4.1.3.4. Their purity was analyzed by SDS-PAGE electrophoresis. Initially the protein bands were detected by highly sensitive silver staining (Figure 15A), but for some samples that changed to visualization by Coomassie (Figure 15B). Nevertheless, all prepared protein samples demonstrated high purity as shown in Figure 15 for examples that were purified solely by SAC (ΔPG-CAIX/6), or in combination with AEC (CAIX/6 and ΔPG-CAIX/5G), or for protein samples after SEC (all in Figure 15B), which results will be shown in chapter 5.3.1. The seeming impurities in sample $^{15}N/^{13}C$-CAIX/6 are due to significant overload of protein amount on the gel, hence the sample is of sufficient quality for the NMR experiments.

Before the samples were stored in -80ºC or immediately used for other experiments, they were typically examined by NMR 1D experiment except for the isotopically labelled samples which were stored in 4ºC and subjected for the NMR experiments afterwards (discussed in chapter 5.4.1).

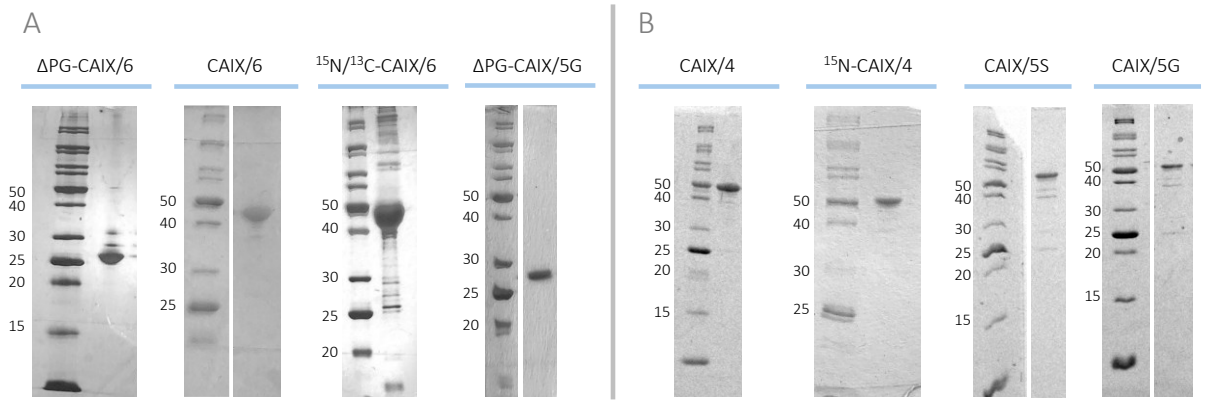


Figure 15: Representative SDS-PAGE electrophoretic analysis of the final purified protein samples. A) SDS-PAGE gel with detection by silver staining. B) SDS-PAGE gel with detection by Coomassie staining.

It is noteworthy that the given SDS-PAGE gels display irregular migration of the protein variants containing both PG and catalytic domain (CAIX/6, CAIX/4, CAIX/5S, CAIX/5G). Their theoretical molecular weight is just under ~39 kDa as stated in Table 9, however according to the MW standard the protein bands are around 50 kDa. The protein variants bearing only the catalytic domain migrate as their respective molecular weight of ~28 kDa.

The amount [mg] of protein in the final sample was calculated from the theoretical MW and spectrophotometrically measured concentration. The given value was then recalculated for 1 L of bacterial culture resulting in the yield for each protein variant as reported in Table 15.

Table 15: Calculated yield [mg/L] for each protein variant that was prepared.

Protein variant	Yield [mg/L]	
	LB medium	Minimal medium
ΔPG-CAIX/6	2.3	
ΔPG-CAIX/4	4.4	
ΔPG-CAIX/5G	1.8	
CAIX/6	1.6	4.6
CAIX/4	1	0.3
CAIX/5S	0.9	
CAIX/5G	0.3	

5.3 Characterization of the oligomeric state

The oligomeric state of the produced protein variants was studied by three methods (SEC, NMR spectroscopy and mass photometry), each of which is indicative for the molecular weight of biomolecules in solution and therefore can be used to clarify the quaternary structure of the protein samples.

5.3.1 Apparent molecular weight by SEC

Majority of the protein samples was analyzed for their apparent molecular weight by size exclusion chromatography on Superdex™ 200 Increase 10/300 GL column attached to ÄKTA pure™ 25 system. This experiment also served as the final purification step except for ΔPG-CAIX/6 and ΔPG-CAIX/5G which was purified only by SAC, or in combination with AEC, respectively (see chapter 5.2.2). These two protein variants were then subjected to SEC on Superdex® 200 10/300 GL with Amersham Biosciences AKTA FPLC system mainly as an analytical procedure. Both methods are delineated with the corresponding calibration curves in chapter 4.1.3.6. The calibration curves were used for calculation of apparent molecular weight for each peak of the elution curves as declared in comprehensive Table 16.

Table 16: Summary of the results from size exclusion chromatography experiments for analysis of the multimeric state of the CAIX protein variants.

Chromatogram	Protein variant	Peak	Ret V [mL]	Apparent MW [Da]	Protein variant band on gel	Multimeric state	Expected MW [Da]
Figure 16	CAIX/6	A #1	12.13	839717	no		
		A #2	14.51	206581	yes	Aggregates	
		A #3	16.12	80000	no		
Figure 16	CAIX/6	B #1	12.31	755216	no		
		B #2	14.48	210266	yes	Aggregates	
		B #3	16.29	72375	no		
Figure 17	CAIX/4	#1	15.62	107409	yes	Dimer	77718
		#2	17.05	46249	no		
Figure 17	CAIX/5S	#1	15.62	107409	yes	Dimer	77750
		#2	17.01	47352	no		
Figure 17	CAIX/5G	#1	15.51	114602	yes	Dimer	77690
		#2	17.07	45707	no		
Figure 19	ΔPG-CAIX/4	#1	17.92	27700	yes	Monomer	28108
Figure 20	ΔPG-CAIX/6	#1	17.11	21352	yes	Monomer	28182
Figure 20	ΔPG-CAIX/5G	#1	16.75	25635	yes	Monomer	28095

First result is given for CAIX/6 sample after AEC as described in chapter 5.2.2.1 which demonstrated several impurities as visible in the input lines (IN) on the gel (Figure 16). Two fractions after the AEC were kept separately and subjected for the SEC. Both results display comparable elution curves with the difference mainly in the amount of protein in each peak shown as mAU for UV_{280} at y axis (Figure 16A,B). Other than that, the peaks are of approximate same elution volumes with peak #1 corresponding to great MW aggregates, peak #2 which contains the CAIX/6 as demonstrated at the SDS-PAGE gel and peak #3 with smaller contaminants or degraded by-products. Though the protein from peak #2 appears as ~50 kDa band on the gel, a characteristic for the protein variants of PG and catalytic domain as denoted above, the apparent MW calculated from the elution volume of 14.51 or 14.48 mL is ~206 kDa and ~210 kDa, respectively. This indicates that the protein sample CAIX/6 forms high molecular weight aggregates in solution.

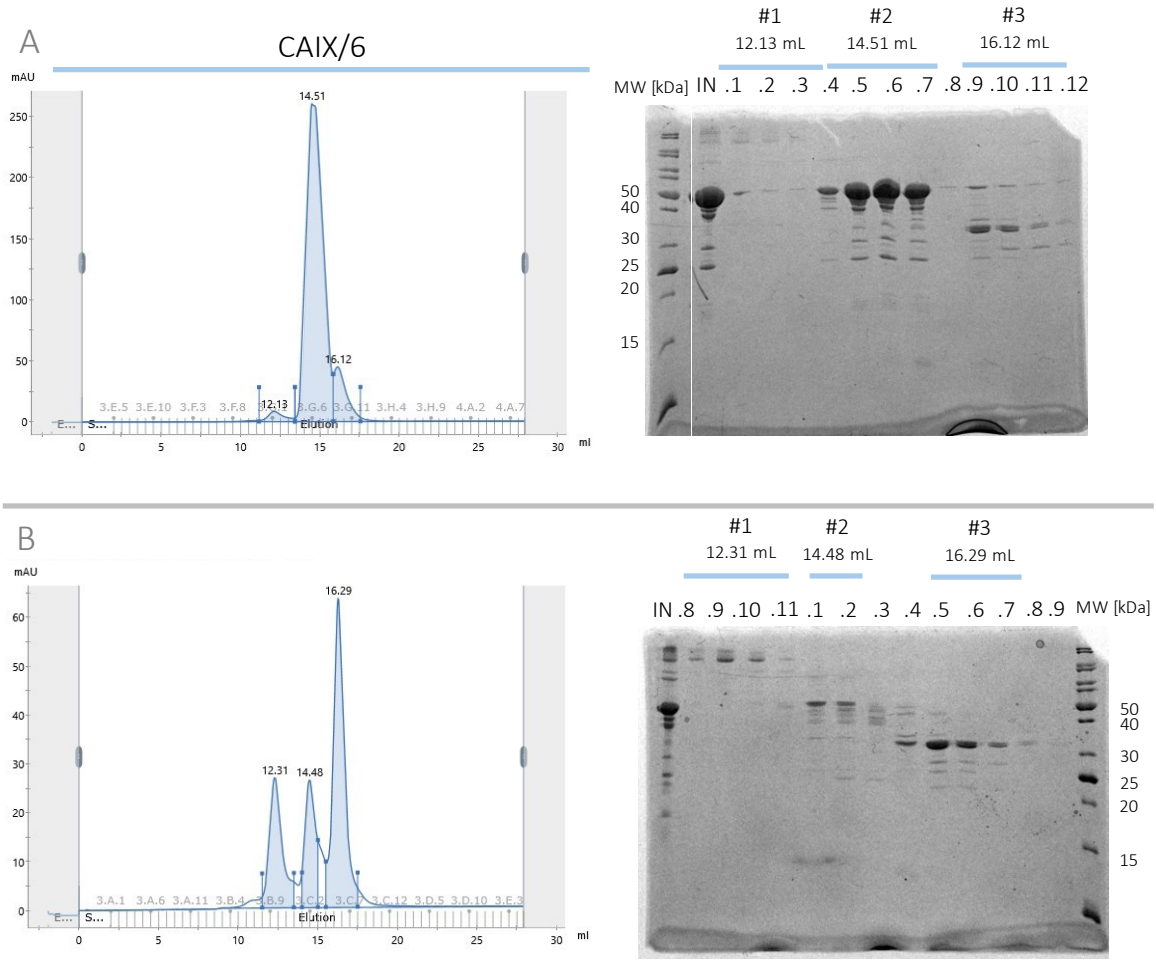


Figure 16: Size exclusion chromatography experiment of CAIX/6 protein variant. Two fractions after AEC were kept separately and subjected to SEC as individual experiments. Both display analogous elution curves with peaks of same elution volumes, where the peak #2 contains the CAIX/6 protein sample as evident on the SDS-PAGE gel visualized with Coomassie. IN = input to the column.

On the other hand, the novel protein variants of PG and catalytic domain (CAIX/4, CAIX/5S and CAIX/5G) displayed different elution profile to the CAIX/6 but similar to each other (Figure 17). The major peak #1 contains the desirable protein variant together with some minor impurities as shown at the SDS-PAGE gel stained by Coomassie. The elution volume 15.62 and 15.51 mL suggests apparent molecular weight of MW ~107 and ~114 kDa, respectively (Table 16). These MW values were interpreted as dimeric states even though when calculated solely theoretically the expected MW would be ~77.7 kDa. But since these CAIX variants contain PG domain which is considered to be unfolded (denoted in literature review chapter 2.3.1 and shown in results chapter 5.4.1), the protein biomolecules are not globular therefore they migrate through the column abnormally and appear as higher MW.

The peak #2 was initially assigned to monomeric state existing within the oligomeric equilibrium in the solution as its calculated MW is ~46 kDa which would resemble the monomer of non-globular CAIX (Table 16). But the SDS-PAGE analysis combined with Coomassie visualization was

not sufficient to clarify this peak's protein content, though some bands can be scarcely seen (Figure 17). When the more sensitive silver staining was used to analyze the same samples and overexposed during the detection, it is notable that the peak #2 has only minority of the CAIX protein variant (at ~50 kDa) and that it contains predominantly contaminant proteins detectable as bands at ~30 and ~25 kDa (Figure 18). This detection also displays the remaining impurities within peak #1 as bands around ~40 and ~25 kDa which were only barely visible by Coomassie staining. However, as the silver-stained gels were overexposed the detected bands cannot be interpreted quantitatively as protein amount within the sample. Hence the fractions of peak #1 were combined and considered to be purified samples of the desirable protein variants as demonstrated at Figure 15.

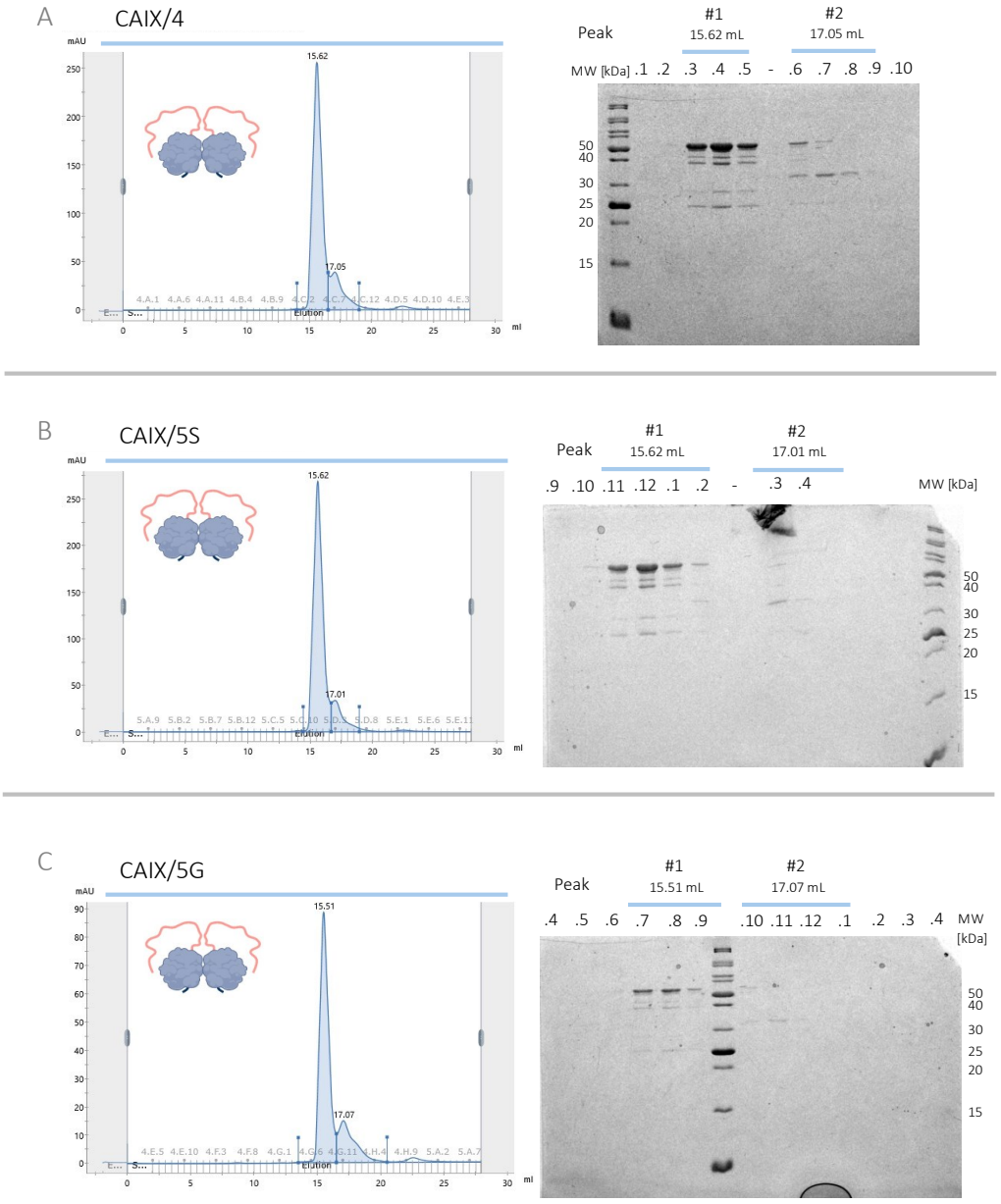


Figure 17: Size exclusion chromatography experiments of protein variants CAIX/4, CAIX/5S and CAIX/5G. All protein variants display similar elution curves with only difference in amount of protein given as UV_{280} [mAU] for CAIX/5G. The major peak #1 corresponds to dimeric state of the CAIX protein variant detectable in corresponding fractions on the SDS-PAGE gel. The composition of peak #2 was not resolved by the SDS-PAGE gel stained by Coomassie.

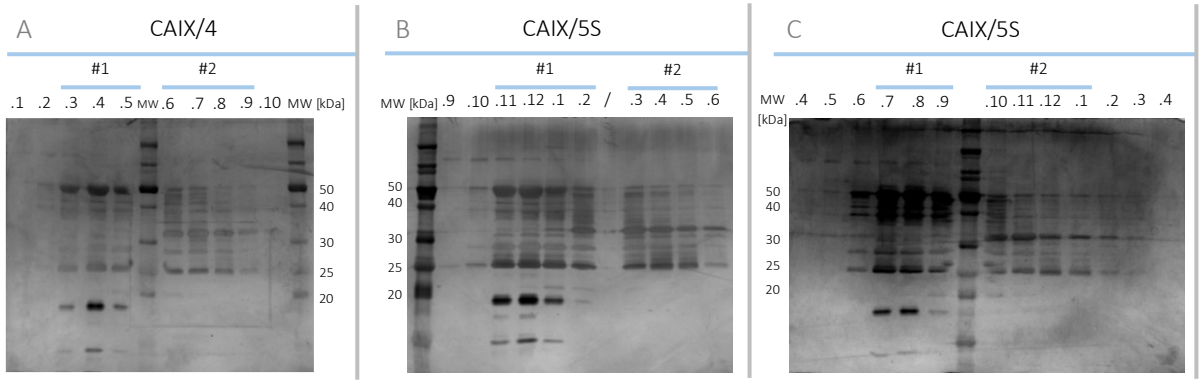


Figure 18: SDS-PAGE analysis of protein variants CAIX/4, CAIX/5S and CAIX/5G after SEC with detection by sensitive silver staining which display that peak #2 from SEC experiments shown at Figure 16 does not contain monomeric state of the protein samples.

The dimeric state was not confirmed for any of the protein variants bearing solely the catalytic domain as shown at Figure 19 for ΔPG-CAIX/4 and Figure 20 for ΔPG-CAIX/6 and ΔPG-CAIX/5G. All of which resulted in similar elution curve with single sharp peak of apparent MW corresponding monomeric state as calculated in Table 16.

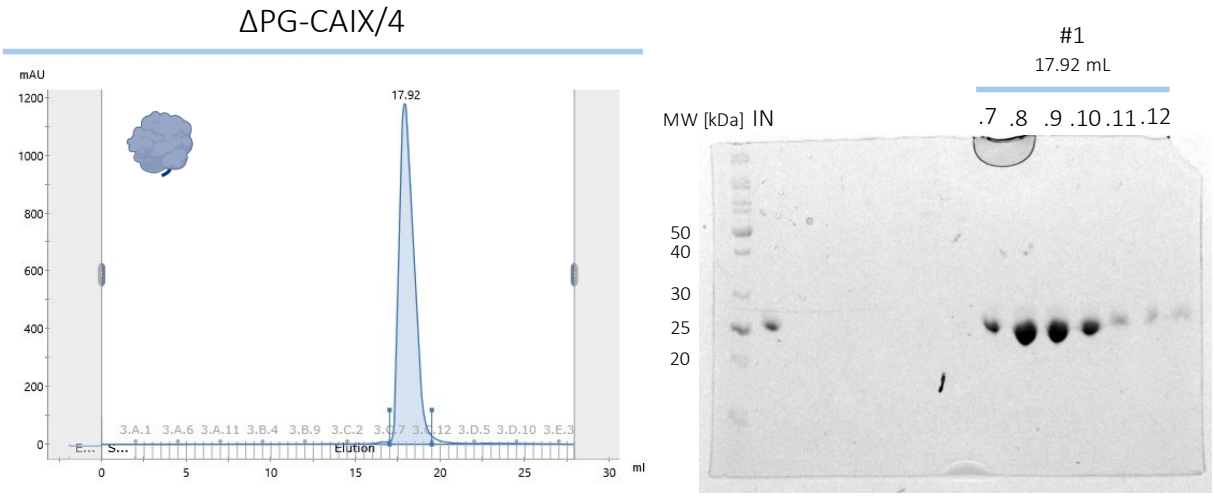


Figure 19: Size exclusion chromatography experiments of protein variant ΔPG-CAIX/4 with SDS-PAGE analysis followed by Coomassie staining. IN = input to the column. Single peak corresponding to monomeric form of ΔPG-CAIX/4 is evident from the elution curve. The protein was detected mainly in fractions .7-.10 as visible from the SDS-PAGE gel.

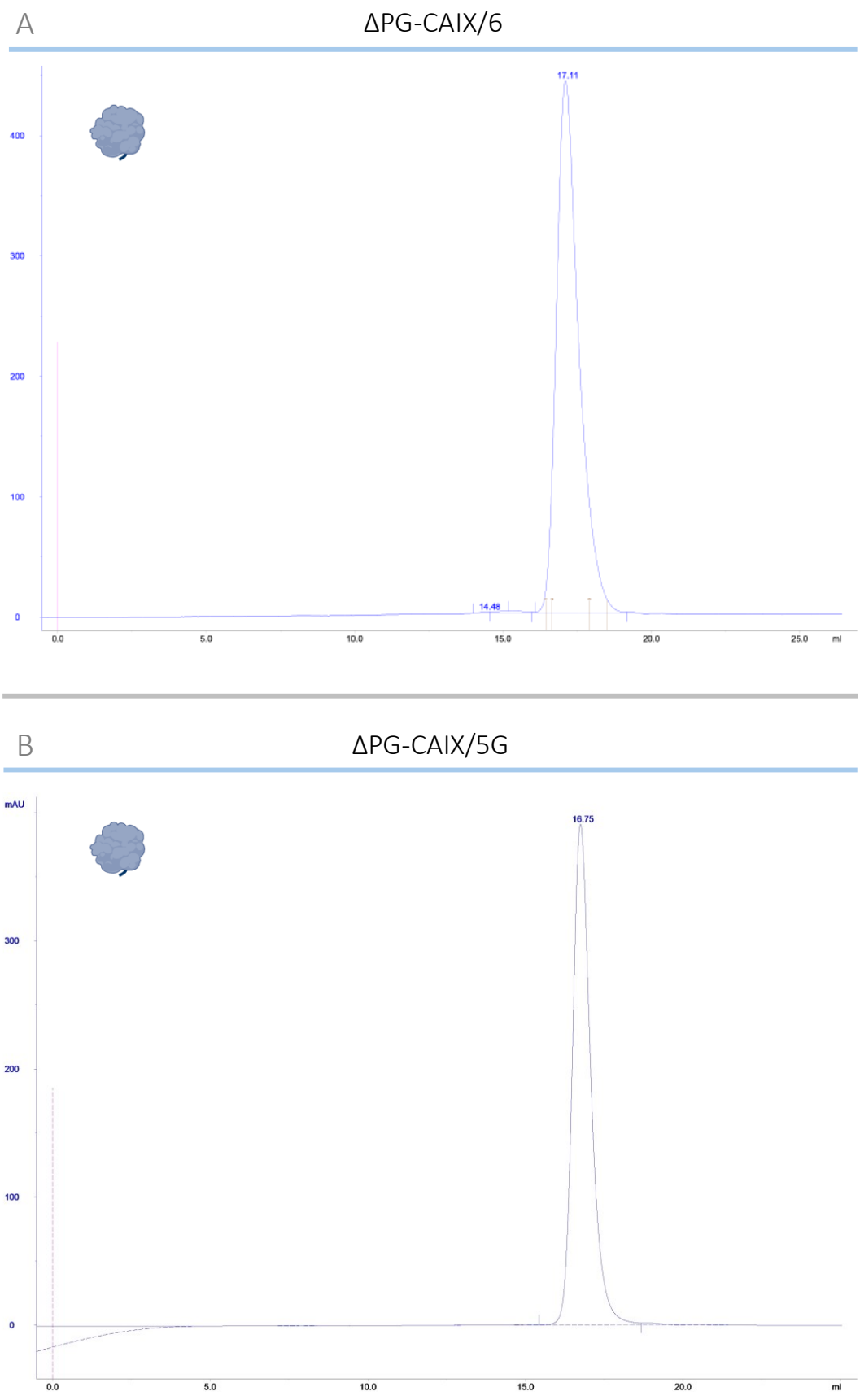


Figure 20: Size exclusion chromatography experiments of protein variants ΔPG-CAIX/6 (A) and ΔPG-CAIX/5G (B). In both instances single peak is evident from the elution curve with elution volume of 17.11 mL (A) and 16.76 mL (B) corresponding to monomers.

5.3.2 Rotational correlation time by NMR spectroscopy

To further examine the multimeric state two protein variants (CAIX/6 and CAIX/4) were chosen to be produced as isotopically labelled samples for NMR spectroscopy experiments. The samples were expressed as described in chapter 4.1.1.2, purified similarly to the standard non-labelled sample and their final SDS-PAGE analysis is in Figure 15. Since the CAIX/6 was intended for more advanced NMR experiments which are not part of this theses, the protein variant was produced as double labelled ^{15}N/^{13}C-CAIX/6. However, even only ^{15}N-labelled sample would have been sufficient for the TRACT experiment used herein (Robson *et al*, 2021; Lee *et al*, 2006), so the CAIX/4 was prepared as ^{15}N-CAIX/4.

To describe the principle of the experiment in brief, the rotational correlation time (Tauc) corresponding to the MW is calculated from the cross-correlated relaxation experiments. It can be theoretically calculated for globular proteins as Tauc [ns] = 0.6*MW [kDa], and this value is then compared with Tauc from the experimentally measured spectra. Since the calculation is only valid for the globular entities, the values presented in Figure 21 were calculated only from spectra in range of 8.6-9.2 ppm equivalent to the CAIX catalytic domain, so the result would not be influenced by the potential unfolded state of the PG domain.

Considering the catalytic domain is MW of 28.1 kDa, it's theoretical Tauc is 16.86 ns for monomer, and two times higher (33.72 ns) for the dimeric state. As apparent from the Figure 21, the calculated Tauc from experiment of ^{15}N/^{13}C-CAIX/6 is indicative of monomer (20.3 ns) and conversely as dimer for ^{15}N-CAIX/4 (37.5 ns).

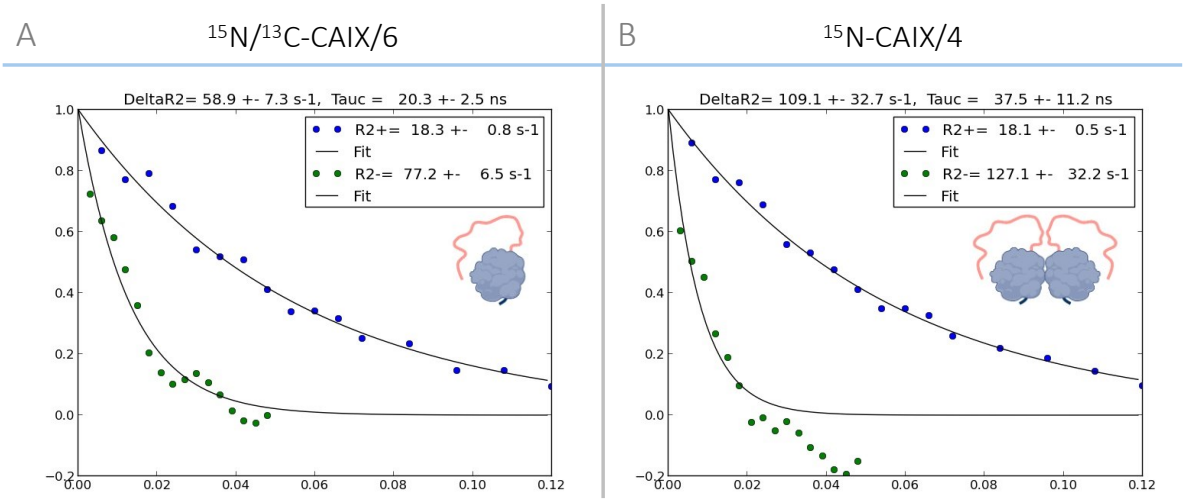


Figure 21: Results of NMR spectroscopy experiment for rotational correlation time of ^{15}N/^{13}C-CAIX/6 and ^{15}N-CAIX/4. The rotational correlation time (Tauc) corelates to the MW of a globular protein. Comparison of the theoretically calculated Tauc for CAIX monomer or dimer to the Tauc calculated based on the collected data implies monomeric state of ^{15}N/^{13}C-CAIX/6 and dimeric of ^{15}N-CAIX/4.

5.3.3 Population analysis by mass photometry

During mass photometry experiment the protein variants CAIX/4 and CAIX/6 were analyzed at low concentration of ~30 nM. The principle and the experiment are described in detail in chapter 4.2.2. As evident from Figure 22A, CAIX/4 sample consists of two different populations, one corresponding to MW of 46 kDa (potential monomer) and second of 77 kDa (potential dimer). On the other hand, the CAIX/6 sample comprises only one apparently monomeric population with particles of 47 kDa MW. Even though this method should not be influenced by the disordered characteristics of the PG domain, the apparent monomeric population exhibited ~8 kDa higher than the theoretically calculated MW (38.8 kDa).

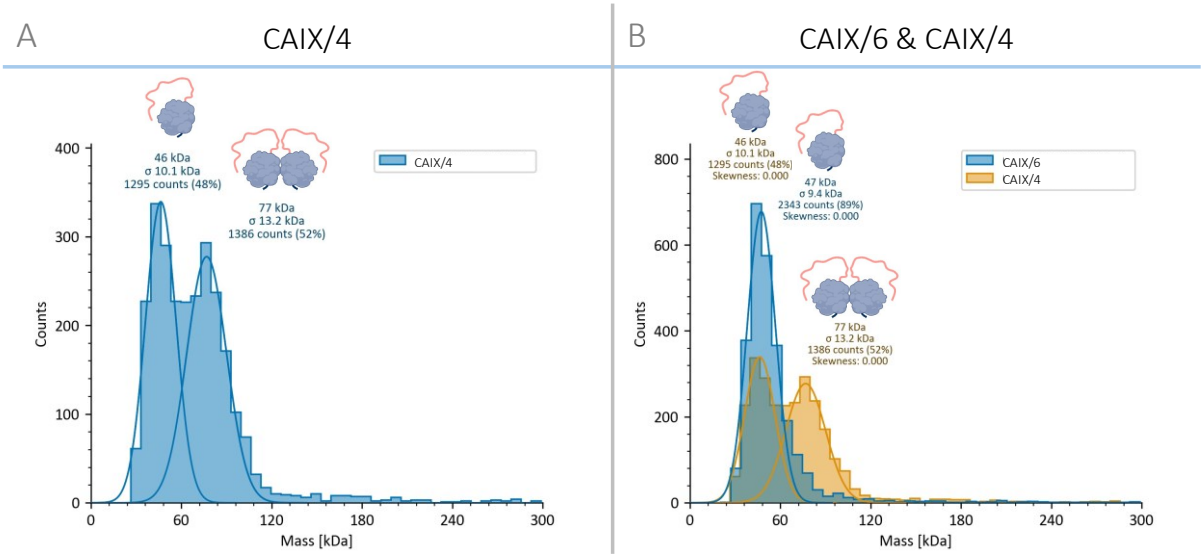


Figure 22: Mass photometry experiment results obtained for protein variants CAIX/4 and CAIX/6. A) Histogram with two populations present within the CAIX/4 protein sample (in blue) at concentration about 30 nM. Their MW signify equilibrium between monomers and dimers at such a low concentration. B) Overlay of CAIX/6 (in blue) and CAIX/4 (in orange) histograms showing that sample CAIX/6 exists only as one population which corresponds to the monomeric one in CAIX/4.

5.4 Protein structure characterization

Two methods of structural biology were used to investigate the structure of CA IX protein variants. Namely, NMR spectroscopy to generally describe the fold of given CA IX domains and macromolecular X-ray crystallography, which enabled determination of the protein structure at nearly atomic resolution.

5.4.1 Fold characterization by NMR spectroscopy

All prepared protein variant were subjected for general characterization by one-dimensional 1H NMR spectroscopy experiment to evaluate the folding state of its domains. An example of such an experiment is given for three selected protein variants containing both PG and catalytic domain, and

for one bearing only the catalytic domain (ΔPG-CAIX/5G) (Figure 23A). Analogous results were collected for other variants as well. The fold state was analyzed mainly from two regions within the 1D spectrum: (1) amide region with signals from hydrogen bound to nitrogen within the protein backbone (7 to 10 ppm), and (2) methyl region of hydrogens inside well-folded core (-1 to 1 ppm).

The presence of PG domain is signified by major broad peak in the amide region centered around 8.2 ppm a typical feature for disordered proteins, as this peak is absent in the ΔPG-CAIX/5G variant. The disordered state of the PG domain is profoundly visible at the 2D ^{1}H-^{15}N HSQC spectra, herein exemplified by spectra for of $^{15}N/^{13}C$-CAIX/6, with intense peaks clustered again around 8.2 ppm (at Figure 23B in red). Contrary the well-folded catalytic domain present in all protein variants is detectable as definite peaks in the methyl region (around 0 ppm) (Figure 23C).

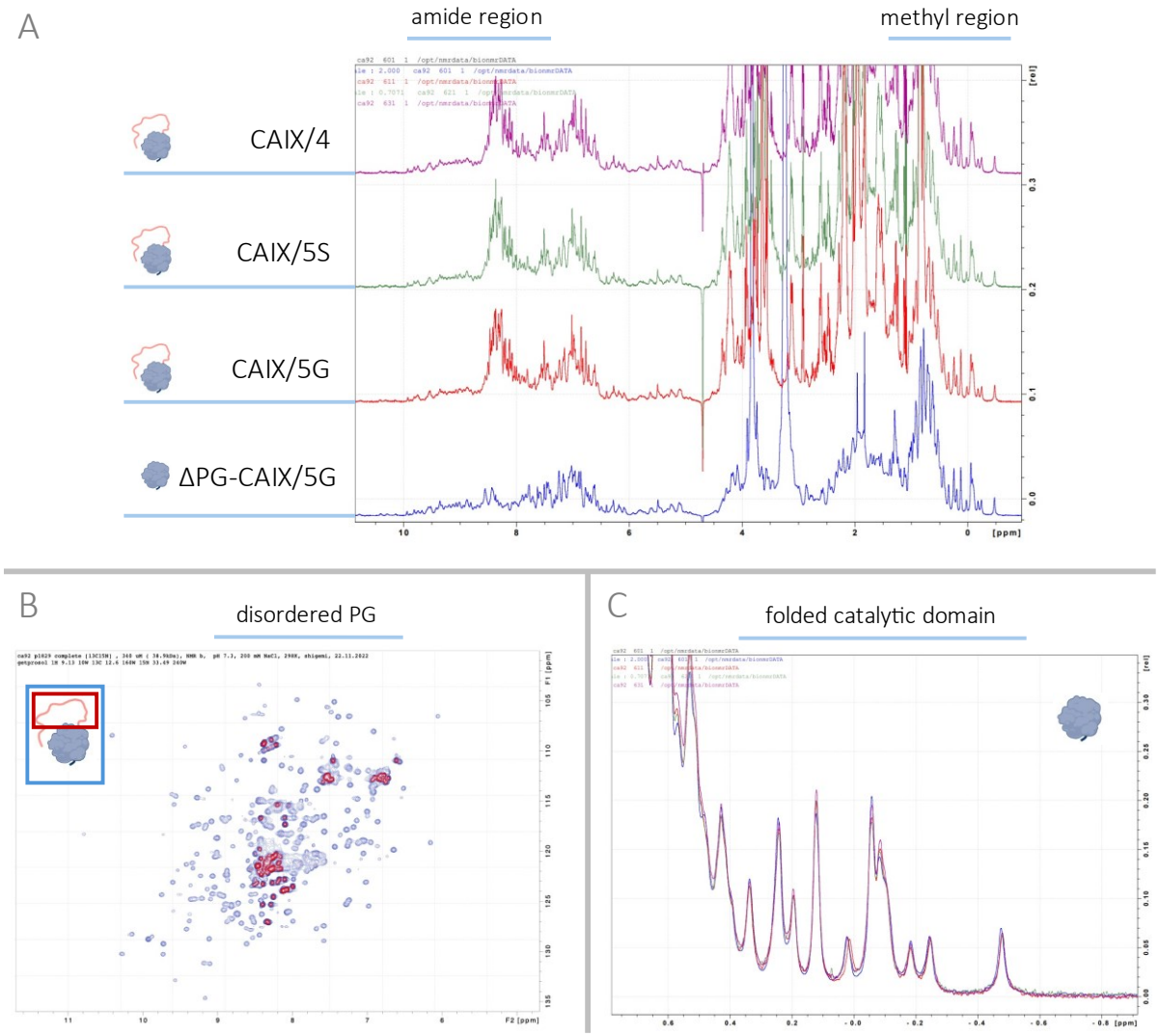


Figure 23: Fold characterization by NMR spectroscopy. A) Comparison of 1D 1H NMR spectra of representative protein samples. The presence of PG domain within protein variants CAIX/4, CAIX/5S, CAIX/5G is evident from the distinctive broad peak centered around 8.2 within the amide region. B) 2D 1H-15N HSQC spectra of $^{15}N/^{13}C$-CAIX/6 where the peaks for the PG domain are contoured in red and for the catalytic domain in blue. C) Overlay of 1D 1H NMR spectra shown in figure part A zoomed to methyl region with distinctive peak which designate the well-folded catalytic domain.

5.4.2 Crystallization procedure

Five of the prepared protein variants were used for various types of crystallization experiments as described in detail in chapter 4.3.1 and summarized in Table 13. As it is not possible to predict the crystallization conditions that would lead to crystal formation, the initial experiments were extensive screens with commercial kits in combination with robotic systems. The usage of robots enabled screening of many conditions with consumption of small volumes of protein sample.

Since it is impossible to include all results from these experiments, they were summarized into Table 17 and examples for selected experimental arrangements which resulted in crystals are discussed in this chapter. The positive crystallization events were typically analyzed by UV microscope as protein crystals can be identified thanks to the tryptophan fluorescence when excited with UV light of 280 nm (Dierks *et al*, 2010).

Even though significantly higher number of crystallization drops, which can be perceived as individual experiments, was set up for the known protein variants ΔPG-CAIX/6 and CAIX/6 they did not yield any crystals suitable for diffraction analysis and subsequent structure determination. The positive crystallization events that were identified during the initial screening or in the optimization were typically crystals of poor quality, so-called spherulites, clustered plates or needles, and quasi crystals. Therefore, only small number out of the 172 positive crystallization events was subjected for X-ray experiment, with only one poorly diffracting crystal with no collected dataset.

Protein variants CAIX/4, CAIX/5S, CAIX/5G were subjected to limited crystallization screening immediately after they were purified and examined by NMR spectroscopy. These experiments resulted in crystallization events found in 13, 4 and 3 crystallization drops, respectively. These crystals were mounted and analyzed in X-ray experiments with number of diffracting crystals, which will be described later in this chapter. From these experiments three protein structures were determined and will be discussed in chapter 5.4.3.

Table 17: Summary of crystallization experiments performed with five protein variants.

Protein variant	# setup drops	# crystal. events	# X-ray experiment	# diffracting crystals	# dataset collected	# resolved structures
ΔPG-CAIX/6	773	1	0	0	0	0
CAIX/6	1526	171	7	1	0	0
CAIX/4	192	13	14	11	7	2
CAIX/5S	192	4	5	2	0	0
CAIX/5G	192	3	3	1	1	1

5.4.2.1 Optimization of ΔPG-CAIX/6 and CAIX/6

The first hit identified for CAIX/6 was during crystallization in sitting drop setup within the commercial screening kit JCSG Plus. The precipitation solution was composed of 1 M LiCl, 0.1 M citrate pH 4, 20% (w/v) PEG 6000, the protein concentration was ~14 mg/mL and it was dissolved in 20 mM TEOA pH 7.3, 100 mM NaCl. Since the crystals appearance was of small clustered needles or plates (Figure 24A,B) these crystals were used for microseeding optimization technique as described in chapter 4.3.1.3. Clustered plates emerged in the majority of drops set up with the same precipitation solution, example is shown in Figure 24C,D. Even though these kinds of crystals were not exposed to X-ray experiment, it can be deduced that they are made of protein and not salt as they glow under the UV light (Figure 24B,D).

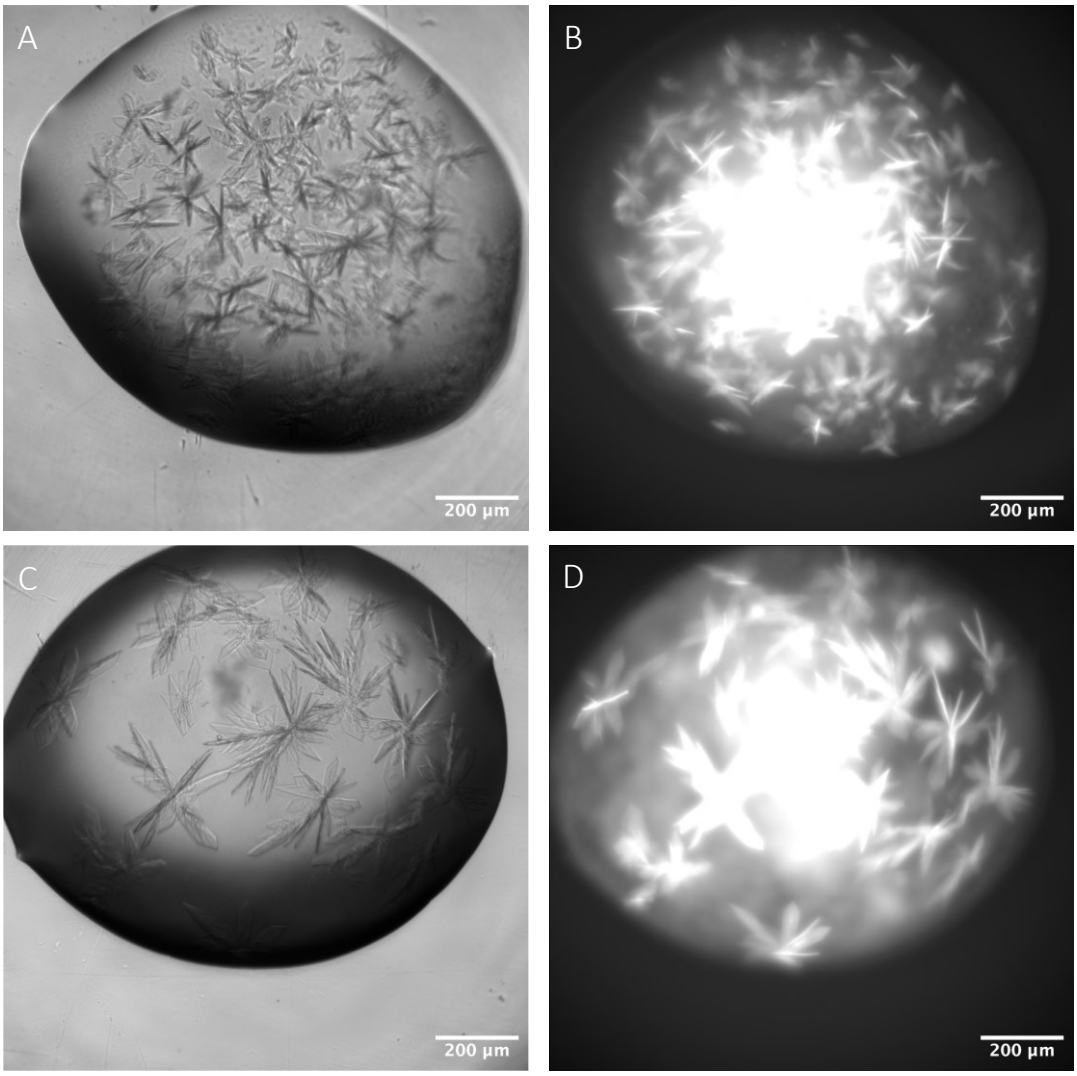


Figure 24: Positive crystallization events of CAIX/6 protein variant. Images were taken by JANSi UVEX-m Imaging Systems in bright field (A, C) or under UV light (B, D).

A, B) CAIX/4 concentrated to ~14 mg/mL in 20 mM TEOA pH 7.3, 100 mM NaCl, precipitant solution: 1 M LiCl, 0.1 M citrate pH 4, 20% (w/v) PEG 6000, protein:precipitant ratio 1:1.

C, D) CAIX/4 concentrated to ~18 mg/mL in 20 mM TEOA pH 7.3, 100 mM NaCl, precipitant solution: 1 M LiCl, 0.1 M citrate pH 4, 20% (w/v) PEG 6000, protein:precipitant ratio 2:1, with microseeding.

Another type of crystallization optimization was performed based on the results from DSF experiment (Figure 25). During that experiment protein samples were dissolved in number of different buffers and analyzed for thermal stability by detection of fluorescence caused by interaction of SYPRO® Orange dye with hydrophobic amino acids while the protein is denaturated. The best buffers are shown in Figure 25, two of them were chosen to be optimal for the protein stability of both variants, therefore ΔPG-CAIX/6 and CAIX/6 were changed to protein storage buffer B and C (in blue in Figure 25 and defined in chapter 3.3.2).

These samples were then screened for crystallization condition by the JSCG Plus screening kit. This experiment's best results appeared as clustered plates of CAIX/6, similar to the ones in Figure 24C,D, with the difference of only one or few clusters present inside a drop. These events were found within the same precipitation solution of 1 M LiCl, 0.1 M citrate pH 4, 20% (w/v) PEG 6000, with protein concentration of ~4 mg/mL and in all protein:precipitant ratios.

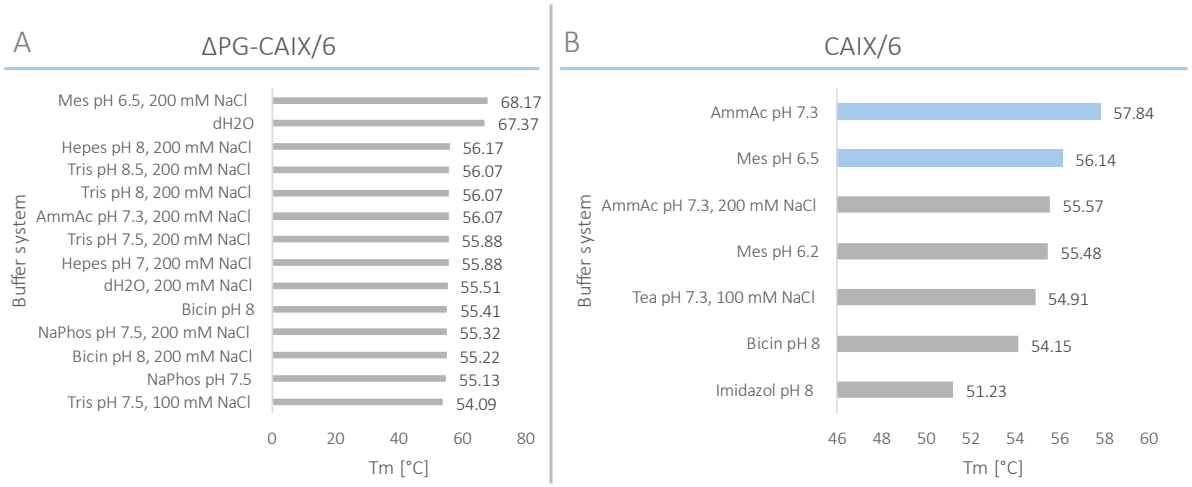


Figure 25: DSF results for ΔPG-CAIX/6 (A) and CAIX/6 (B), when number of buffer systems was tried out for the protein stabilization during increase of temperature. The depicted values are Tm [ºC] = melting temperature (protein denaturation) for each buffer system. The buffers highlighted in blue were chosen for the crystallization optimization process of both protein variants.

Only two arrangements of the hanging drop setup gave rise to some crystals as shown in Figure 26A for CAIX/6 in protein:precipitant ratio of 2:1 and in Figure 26B with streak seeding upon 8 hours equilibration of the drop. In both cases the precipitant solution was again 1 M LiCl, 0.1 M citrate pH 4, 20% (w/v) PEG 6000, the protein was concentrated to ~4 mg/mL in protein buffer B. However, similarly to other crystals of CAIX/6 these are not of sufficient quality for the X-ray experiment. Other optimization procedures performed (chapter 4.3.1.3), namely additive screen and precipitant condition adjustments did not lead to any crystallization results.

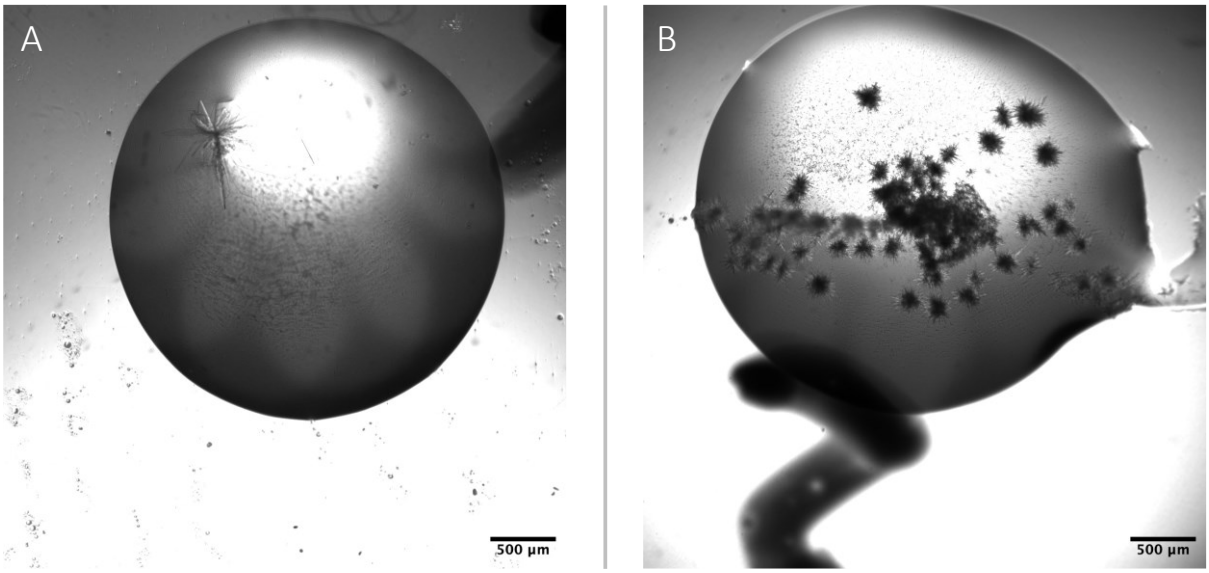


Figure 26: Positive crystallization events of CAIX/6 protein variant in hanging drop arrangement. Images were taken by JANSi UVEX-m Imaging Systems in bright field.

A) CAIX/4 concentrated to ~4 mg/mL in 100 mM MES pH 6.5, precipitant solution: 1 M LiCl, 0.1 M citrate pH 4, 20% (w/v) PEG 6000, protein:precipitant ratio 2:1.

B) CAIX/4 concentrated to ~4 mg/mL in 100 mM MES pH 6.5, precipitant solution: 1 M LiCl, 0.1 M citrate pH 4, 20% (w/v) PEG 6000, protein:precipitant ratio 1:1, with streak seeding 8 hours after the drop was established.

5.4.2.2 Crystals of the novel CAIX protein variants

The crystallization screen of protein variants CAIX/4, CAIX/5S and CAIX/5G with commercial JCSG Plus screening kit resulted in numerous crystals of variable size and quality. The most promising results and the composition of crystallization solution are described in the following text and figures. When compared to crystals obtained for CAIX/6 variant, these appeared to be more three dimensional, although still clustered. Thus, some of them could be separated to individual crystals for diffraction X-ray experiment (example of this set of crystals in Figure 27). The crystals illuminated when visualized in UV microscope which indicates they consist of protein.

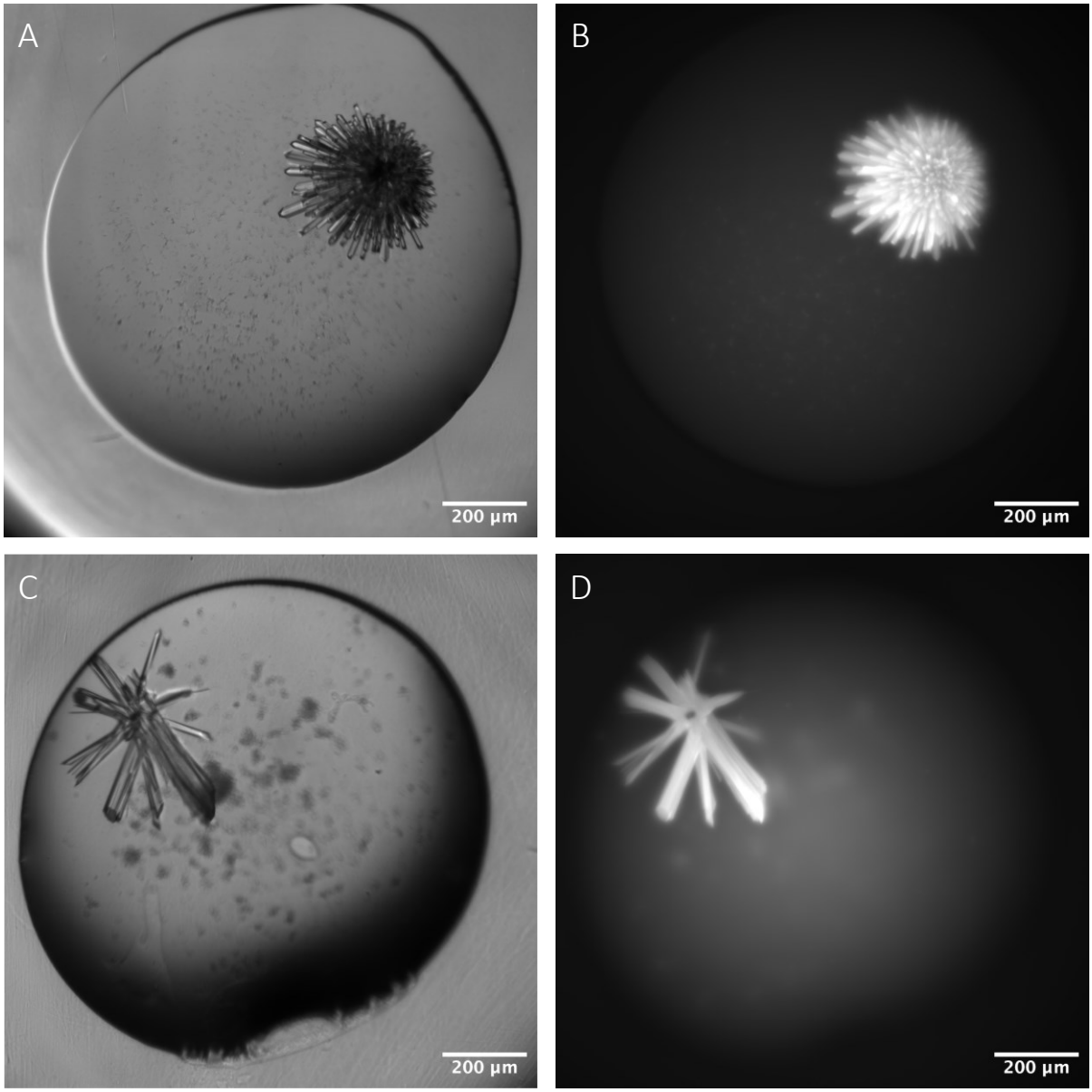


Figure 27: Crystals of CAIX/4 protein variant. Images were taken by JANSi UVEX-m Imaging Systems in bright field (A, C) or under UV light (B, D).

A, B) CAIX/4 concentrated to ~16.5 mg/mL in 20 mM Tris-HCl pH 7, 100 mM NaCl, precipitant solution: 0.18 M tri-ammonium citrate, 20% (w/v) PEG 3350, protein:precipitant ratio 2:1.

C, D) CAIX/4 concentrated to ~16.5 mg/mL in 20 mM Tris-HCl pH 7, 100 mM NaCl, precipitant solution: 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350, protein:precipitant ratio 2:1.

The second set of crystal images shows single crystals of CAIX/4 and CAIX/5S protein variants that were analyzed by X-ray and diffracted to low resolution of 2.4 Å and 4.5 Å (Figure 28) and therefore did not result in structure determination.

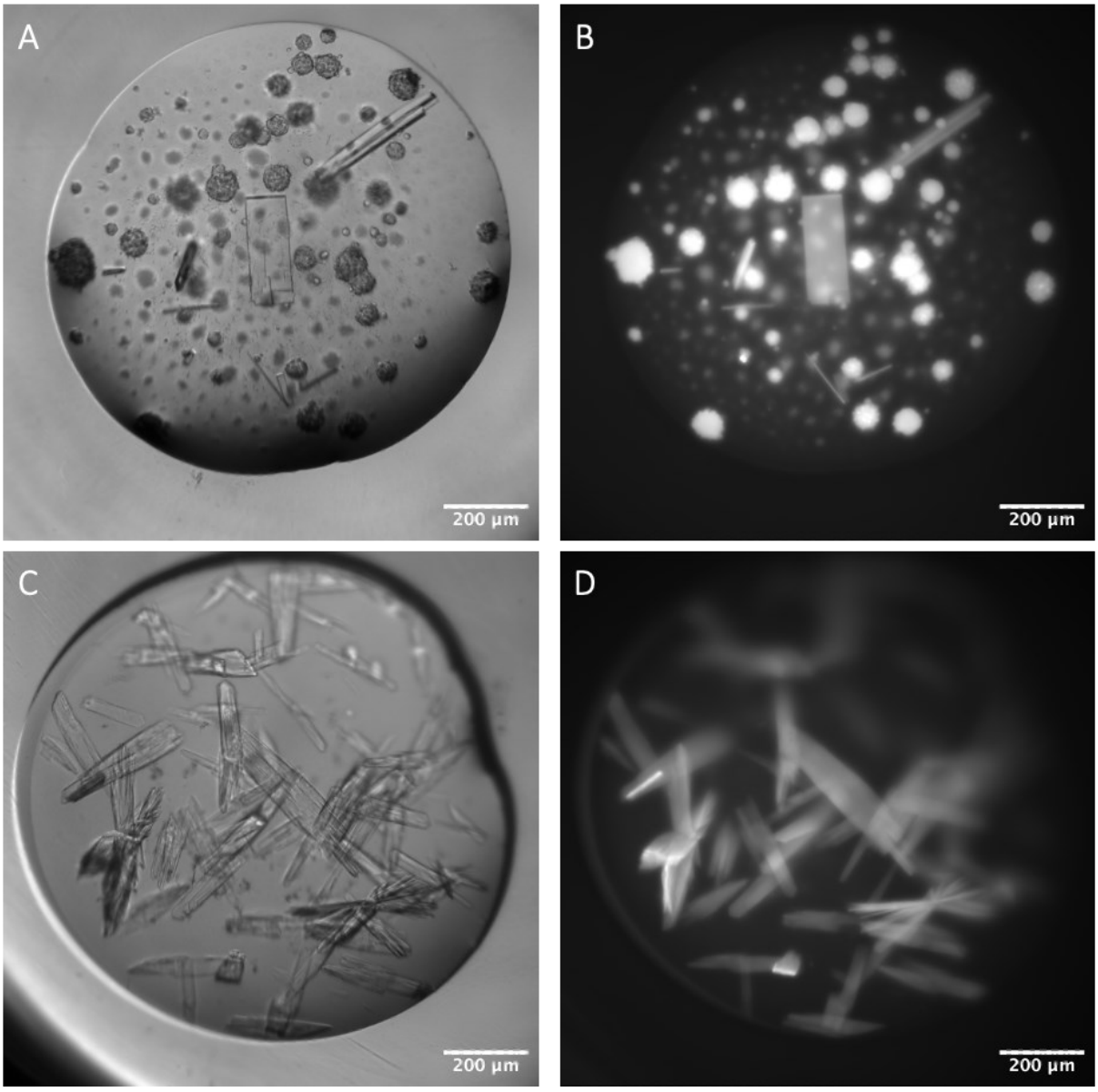


Figure 28: Crystals of CAIX/4 and CAIX/5S protein variants. Images were taken by JANSi UVEX-m Imaging Systems in bright field (A, C) or under UV light (B, D).

A, B) CAIX/4 concentrated to ~16.5 mg/mL in 20 mM Tris-HCl pH 7, 100 mM NaCl, precipitant solution: 0.1 M tri-sodium acetate pH 4.5, 0.1 M BIS-TRIS pH 5.5, 25% (w/v) PEG 3350, protein:precipitant ratio 2:1, diffraction up to 2.4 Å.

C, D) CAIX/5S concentrated to ~16 mg/mL in 20 mM Tris-HCl pH 7, 100 mM NaCl, precipitant solution: 0.2 M lithium sulfate, 0.1 M Tris pH 8.5, 1.26 M ammonium sulfate, protein:precipitant ratio 2:1, diffraction up to 4.5 Å.

Crystals for which complete diffraction datasets were collected, processed, and used for solution of the CAIX structure are shown in Figure 29 and Figure 30. The statistics of data collection, description of structure determination and refinement is in chapter 5.4.3. Even though the crystals of CAIX/4 in Figure 29 are highly clustered, small part of the cluster representing a monocrystal was separated during the mounting and enabled diffraction up to 1.75 Å. Similarly, small crystal was separated from the clustered rod of CAIX/4 in Figure 30A,B and diffracted up to 1.91 Å. The protein variant CAIX/5G generated two single crystals, out of which only one was successfully mounted into the loop (Figure 30C,D). As it is obvious from the Figure 30C,D, the crystal was elongated, therefore nonstandard, helical data collection was utilized during the synchrotron experiment and generated dataset with resolution of 1.35 Å.

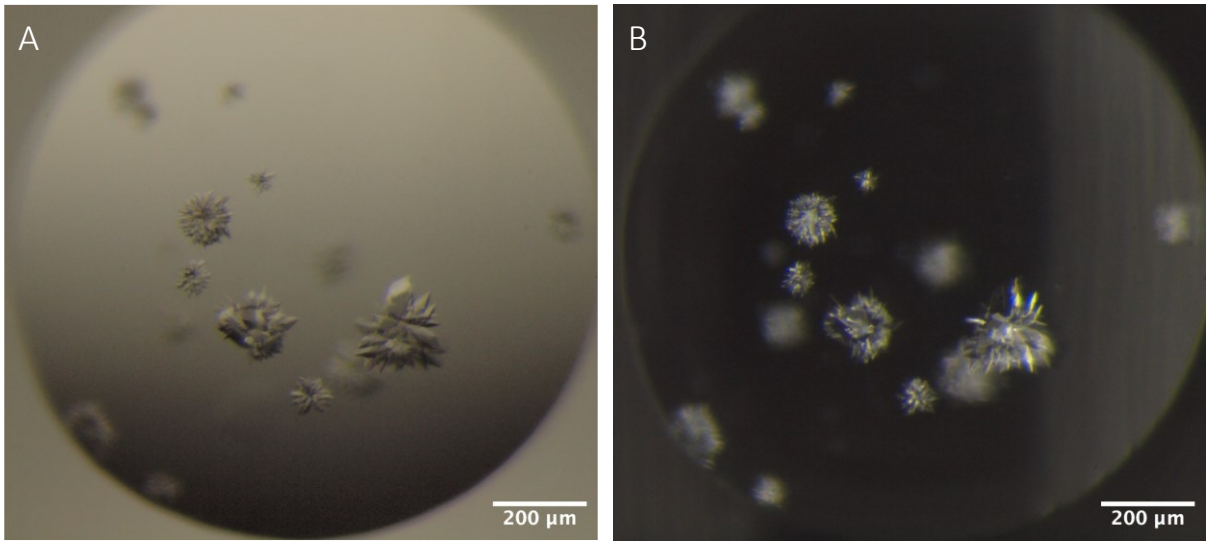


Figure 29: Crystals of CAIX/4 protein variant. Images were taken by optical microscope Olympus SZX10 with photo camera Olympus E-620 in bright field with grey (A) or black (B) background.

A, B) CAIX/4 concentrated to ~16.5 mg/mL in 20 mM Tris-HCl pH 7, 100 mM NaCl, precipitant solution: 0.1 M potassium thiocyanate, 30% (w/v) PEG MME 2000, protein:precipitant ratio 2:1, dataset: CAIX/4 #1.

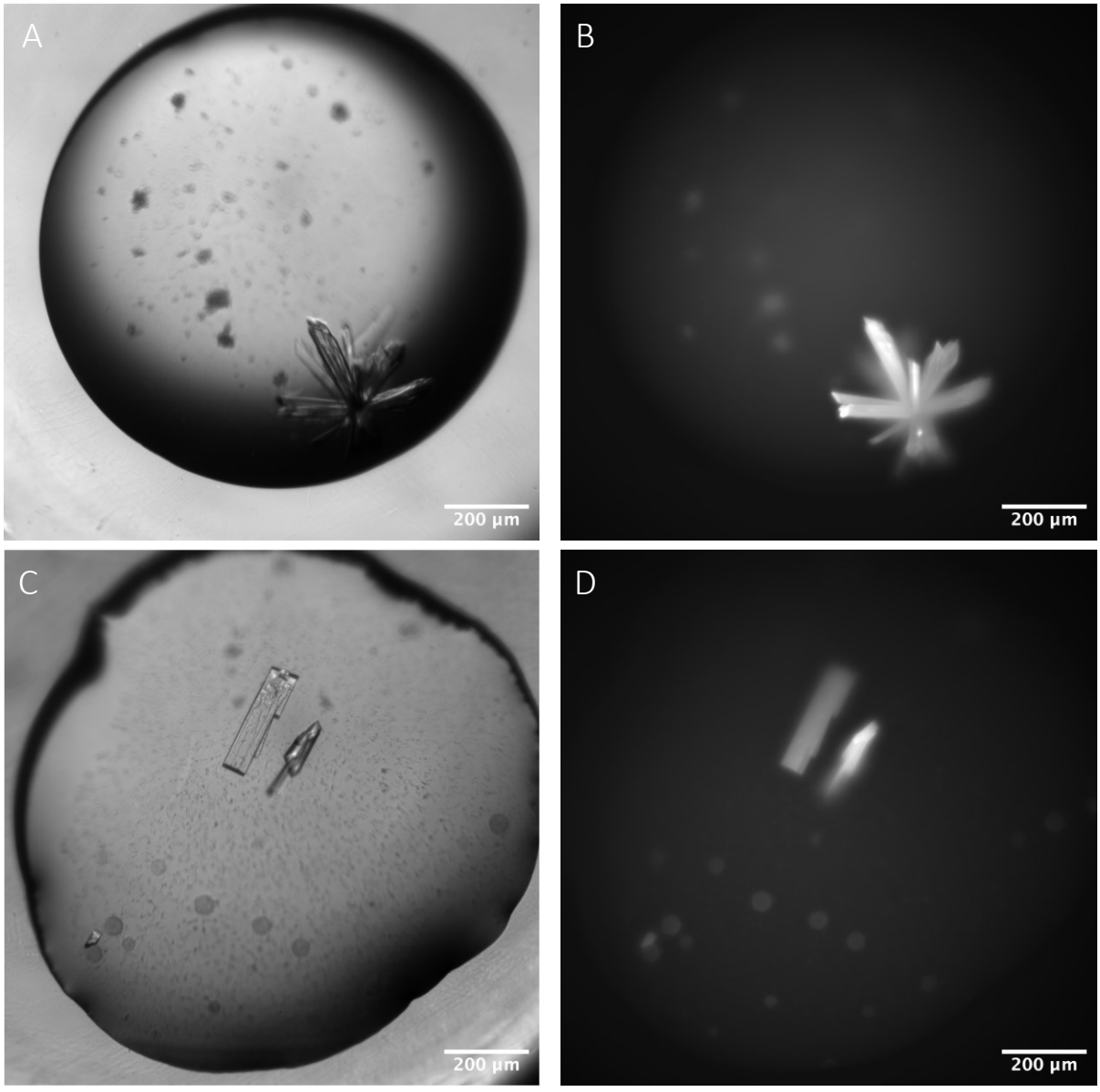


Figure 30: Crystals of CAIX/4 and CAIX/5G protein variants. Images were taken by JANSi UVEX-m Imaging Systems in bright field (A, C) or under UV light (B, D).

A, B) CAIX/4 concentrated to ~16.5 mg/mL in 20 mM Tris-HCl pH 7, 100 mM NaCl, precipitant solution: 0.2 M sodium thiocyanate, 20% PEG (w/v) 3350, protein:precipitant ratio 1:1, dataset: CAIX/4 #2.

C, D) CAIX/5G concentrated to ~5.5 mg/mL in 20 mM Tris-HCl pH 7, 100 mM NaCl, precipitant solution: 0.15 M potassium bromide, 30% PEG (w/v) MME 2000, protein:precipitant ratio 1:1, dataset: CAIX/5G.

5.4.2.3 Crystal content analysis

Based on the disordered feature of PG domain described in chapter 5.4.1, we wanted to analyze whether the domain is present in the crystal lattice as disordered protein regions are tolerated within crystals only to some extent and to investigate potential proteolysis (Oldfield *et al*, 2013). Therefore, the protein content of selected crystals was analyzed by SDS-PAGE gel electrophoresis. Crystals were dissolved in 20 µL of electrophoresis sample buffer either after or without any X-ray experiment and

loaded on the gel. Example of such analysis is given for CAIX/4 crystal previously depicted in Figure 29, when the major band migrates as MW of 28 kDa corresponding to the catalytic domain (Figure 31A) pointing to proteolytic cleavage, most probably by residual contaminant proteases.

To explore the nature of observed proteolysis further, the CAIX/4 protein sample was stored at 37ºC for 1 day and 1 week in the presence or absence of protease inhibitors (Figure 31B). The SDS-PAGE showed that there is some proteolysis which may be caused by residual protease impurities as the bands at 34 kDa are less prominent in samples with protease inhibitors after one day. Similar patterns are noticeable after one week, despite the fact the bands visualization is poor on this gel. The absence of PG domain within the crystal will be discussed more in chapter 5.4.3.

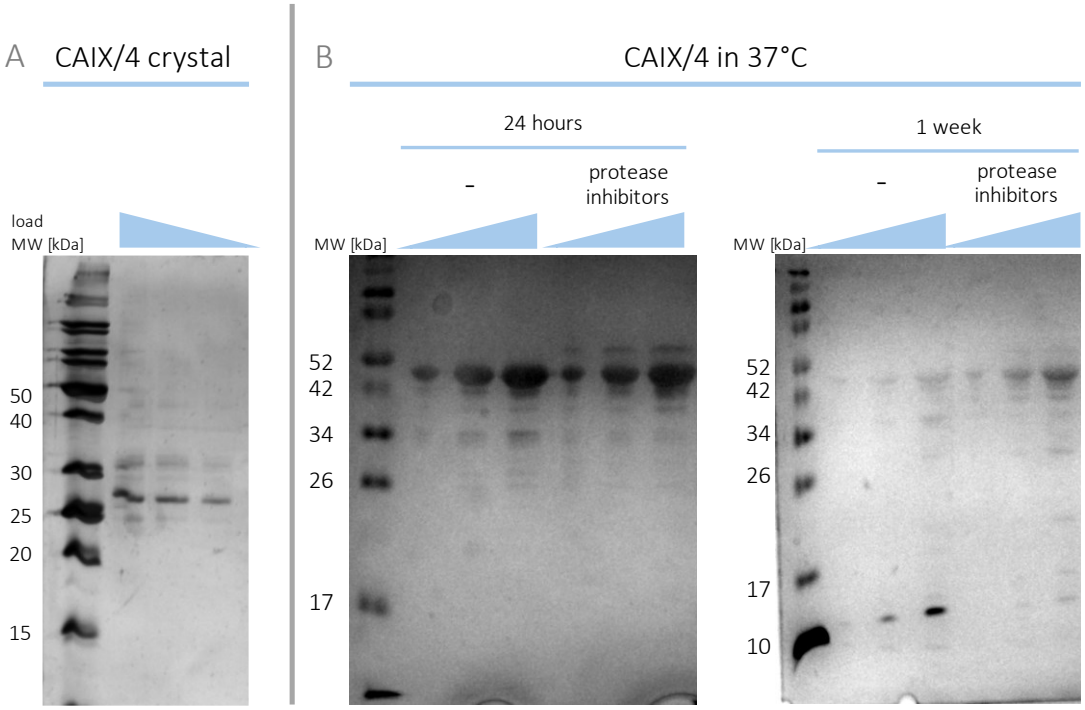


Figure 31: SDS-PAGE analysis of crystal content. A) Crystals of CAIX/4 were dissolved in electrophoresis sample buffer analyzed on gel, which clearly showed bands corresponding to MW of only the catalytic domain (25 kDa). B) To enhance the possible proteolysis the samples of CAIX/4 protein were stored for 24 hours and 1 weak in 37ºC. Although the final SDS-PAGE gel showed that majority of the sample migrated similarly to the purified protein containing both PG and catalytic domain there are bands of smaller MW evident in both cases with and without protease inhibitors. All SDS-PAGE gels were visualized by silver staining.

5.4.3 Structure determination of dimeric CA IX

Two crystals of CAIX/4 and one CAIX/5G crystal (depicted in Figure 29 and Figure 30) enabled collection of three complete datasets that were further processed. Crystal parameters and the data collection statistics are summarized in Table 18. All three crystals belong to monoclinic *P*2(1) space group. Two crystals of CAIX/4 (#1 and #2) obtained from different crystallization conditions exhibited similar cell size and angle parameters with four protein molecules in asymmetric unit (AU) according to calculated Matthews coefficient values (Kantardjieff & Rupp, 2003). On the other hand, the cell

parameters of CAIX/5G crystal refer to smaller AU with two protein chains. These diffraction datasets were used to determine structure of CA IX by molecular replacement using 3IAI PDB entry. The structure determination and refinement were done within CCP4i2 software package, and the statistics is stated in Table 18.

Table 18: Data collection and refinement statistics.

	CAIX/4 #1	CAIX/4 #2	CAIX/5G
Data collection statistics			
Wavelength (Å)	0.9786	0.9184	0.9184
Space group	*P*2(1)	*P*2(1)	*P*2(1)
Cell parameters (Å, o)	73.44 76.75 94.43 90.00 95.91 90.00	73.34 77.36 96.22 90.00 96.22 90.00	44.30 75.20 77.90 90.00 98.60 90.00
Resolution range (Å)	50.00-1.75 (1.85-1.75)	48.031-1.91 (2.03-1.91)	44.044-1.349 (1.43-1.35)
Number of unique reflections	100,209 (13,490)	82,650 (13,324)	110,870 (16,288)
Multiplicity[a]	3.0186 (2.47)	6.87 (7.02)	6.53 (5.03)
Completeness (%)	94.9 (79.6)	99.5 (99.6)	97.9 (89.4)
Rmeas[b]	9.7 (91.3)	21.6 (210.2)	8.1 (119.3)
CC(1/2) (%)	99.6 (53.7)	99.5 (41.1)	99.9 (50.7)
Average I/σ(I)	7.65 (0.92)	7.35 (0.91)	13.94 (1.20)
Wilson B (Å2)	33.942	34.749	21.093
Definition of AU content			
Number of molecules in AU	4	4	2
Solvent (%)	47.79	49.06	47.3
Matthews coefficient	2.35	2.41	2.33
Matthews probability	0.69	0.65	0.98
Refinement statistics			
Resolution range (Å)	36.53-1.75 (1.79-1.75)	47.87-1.91 (1.96-1.91)	38.11-1.35 (1.39-1.35)
No. of reflection in working set	98,202 (5385)	80,545 (5926)	108,536 (6873)
No. of reflection in the test set	2005 (110)	2101 (155)	2216 (141)
R_{work} value (%)[c]	18.44 (38.9)	19.87 (36.2)	14.61 (28.6)
R_{free} value (%)[d]	23.35 (38.8)	24.33 (38.6)	17.05 (28)
RMSD bond length (Å)	0.012	0.01	0.006
RMSD angle (o)	1.879	1.937	1.502
Mean B value	29.085	20.3	19.014

The data in parentheses refer to the highest-resolution shell for data collection statistics.

[a]Multiplicity = number of total reflections / number of unique reflections.

[b]Rmeas = $\sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ where I_i(hkl) is an individual intensity of the i[th] observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity of reflection *hkl* with summation over all data.

R-value = ||Fo| - |Fc||/|Fo|, where Fo and Fc are the observed and calculated structure factors.

[c]R_{work} is calculated as R-value for the reflections in working set.

[d]R_{free} is calculated as R-value for reflections in test set that were randomly chosen and omitted from the refinement process.

All structures were determined and refined to resolution below 2 Å, with the best resolution of 1.35 Å for CAIX/5G. The values of R_{work} and R_{free} for all three structures (Table 18) demonstrate that the models correspond to the experimentally measured data and the best refinement statistics was achieved for the high-resolution structure of CAIX/5G. The refined structural models were subjected for geometry and steric validation by MolProbity server (Williams *et al*, 2018). The outcome of the MolProbity validation is showed in Table 19 which displays that the solved structures are in 99th-100th percentile when compared to structures of the same resolution for both clashscore and overall MolProbity score.

Table 19: MolProbity geometry and steric analyzed for the refined structural models of CA IX.

		CAIX/4 #1		CAIX/4 #2		CAIX/5G	
All-Atom Contacts	Clashscore, all atoms	2.19	99th percentile* (N=932, 1.75Å ± 0.25Å)	2.98	99th percentile* (N=754, 1.91Å ± 0.25Å)	2.18	99th percentile* (N=466, 1.35Å ± 0.25Å)
Protein geometry	Poor rotamers	1	0.12%	9	1.10%	2	0.48%
	Favored rotamers	786	95.27%	766	93.76%	403	97.11%
	Ramachandran outliers	0	0.00%	0	0.00%	1	0.20%
	Ramachandran favored	983	98.50%	980	98.00%	492	98.40%
	Rama distrib. Z-score	-0.82 ± 0.24		-0.84 ± 0.24		-0.34 ± 0.34	
	MolProbity score ^	0.99	100th percentile* (N=11713, 1.75Å ± 0.25Å)	1.12	100th percentile* (N=11889, 1.91Å ± 0.25Å)	0.99	99th percentile* (N=3057, 1.35Å ± 0.25Å)
	Cβ deviations >0.25Å	2	0.22%	8	0.87%	0	0.00%
	Bad bonds	6 / 8072	0.07%	1 / 8014	0.01%	0 / 4048	0.00%
	Bad angles	36 / 11026	0.33%	42 / 10954	0.38%	7 / 5538	0.13%
Peptide omegas	Cis Prolines	16/100	16.00%	14/100	14.00%	8/50	16.00%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

^ MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

At the beginning of the structure determination, the initial electron density maps were investigated to find what domains are present in the AU. Specifically, it was examined if the map for the PG domain is present. No electron density was found for the PG domain, and this is also evident from the 2*Fo-Fc* electron density maps calculated after the CA domain was refined (Figure 32). Examples are given for chain A and chain C within CAIX/4 #1 crystal (Figure 32A,B) and chain A from CAIX/5G (Figure 32C). Clear electron density was observed from the N-terminus of the catalytic domain, starting from residue 140. In addition to the fact that there is no evident electron density for the PG domain, there is only limiting space for the domain as other chains of the catalytic domain from the same AU or surrounding AUs occupy the area. In conclusion, this analysis together with the crystal SDS-PAGE analysis described in chapter 5.4.2.3 confirmed that the crystals are composed of the catalytic domains and the N-terminal PG domain is not present due to proteolytic cleavage by contaminating proteases.

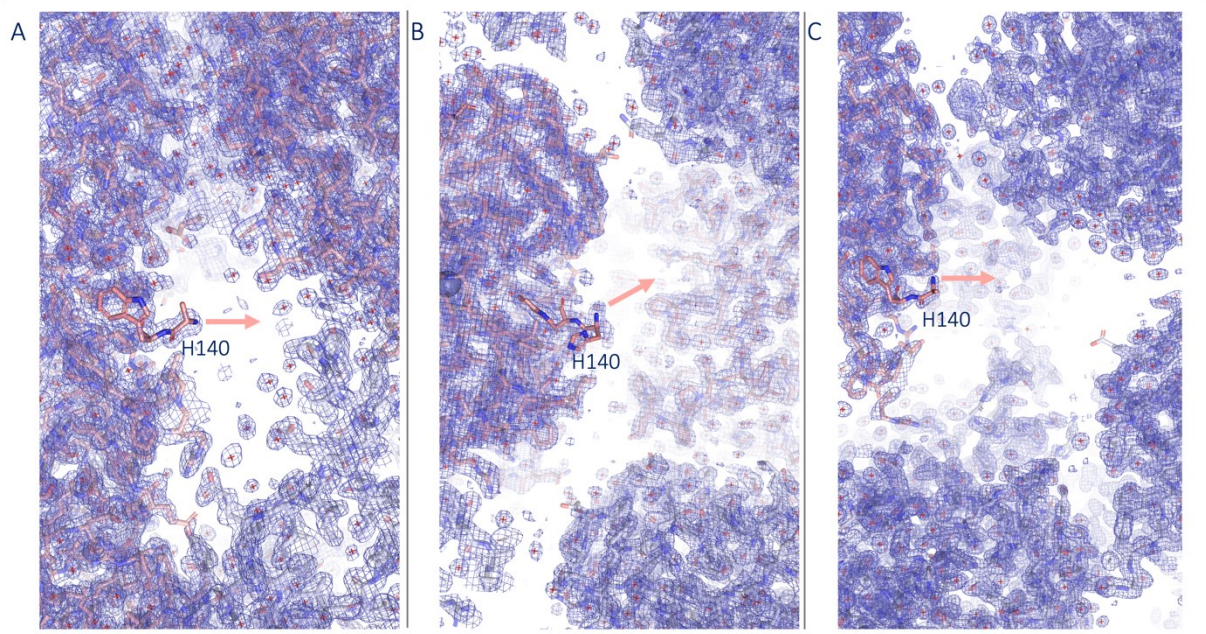


*Figure 32: Electron density maps for chain A (A) and chain C (B) in CAIX/4 #1 crystal, and chain A in CAIX/5G crystal (C). 2*Fo-Fc *electron density maps after refinement are contoured at 1 σ in radius of 70 Å around first two residues with clear electron density; H140 and W141 of catalytic domain (shown as sticks in pink). Evidently only in (B) figure the electron density enabled to model the side chain of the H140 residue, in the other instances (A,C) only the backbone of H140 is designated. The pink arrows suggest the prospective extension of the catalytic domain towards the PG domain and N-terminus of the protein variant. The figures were made in PyMol software.*

The electron density maps were explained by crystallographic model containing CA IX residues 140-391 corresponding to CA catalytic domain and consequently used for analysis of their dimeric assembly and interface.

The determined structures of the CA IX catalytic domain display the common fold of carbonic anhydrases as revealed by coloring based on the secondary structures. Similarly to other CAs, it consists of antiparallel β-sheet core with zinc atom situated in the active site, surrounded by short α-helices and surface loops, as shown on example of CAIX/5G (Figure 33).

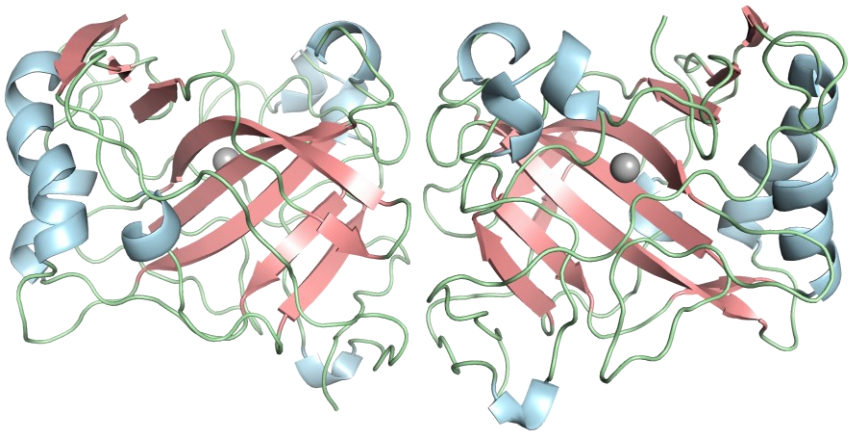


Figure 33: Structural model of CAIX/5G displayed as cartoon and colored based on the secondary structures with noticeable common fold of CA catalytic domain with β-sheet core (in pink) surrounded by short α-helices (in light pink) and numerous loops (in green). The zinc atoms located in the active sites are shown as grey spheres. The figure was made in PyMol software.

The structural models of CAIX/4 #1, CAIX/4 #2 and CAIX/5G were analyzed by PDBePISA server and in PyMol software to examine whether the CAIX/4 and CAIX/5G protein variants form the biologically relevant dimer as defined in (Alterio *et al*, 2009). This arrangement was defined as formation of H-bond between R268 and A258 of the two monomers and the interface formed by residues A172-A176. The same dimeric arrangement was identified in all three structures as shown by their structural superposition in Figure 34A. During this structure alignment of Cα atoms, 469, 449 and 455 atoms were aligned with RMSD = 0.202, 0.591 and 0.722 Å respectively for combinations of CAIX/4 #1 with CAIX/4 #2, CAIX/4 #1 with CAIX/5G, and CAIX/5G with CAIX/4 #2.

The first interface between two monomeric chains recognized by PDBePISA was analogous to the one defined in Table 14. Namely, all three structures have interface formed by five areas F160, A172-L177, P216-Y220, S256-G270, and N382-N387, all solvent-inaccessible residues were identified (except for L351 in both CAIX/4 and C336 in CAIX/5G), and H-bond between guanidium group of R268 side chain and carbonyl group of A258 backbone was present. Additionally, a complex salt bridge between positive charged residue R218 and negative charged E328 was recognized in CAIX/4 #1. The stated H-bond (R268 + A258) and salt bridge (R218 + E328) were confirmed in PyMol by

search of polar contacts between the two chains of the dimer (Figure 34C). All together six polar contacts were found within the distance range of 2.9-3.5 Å between non-hydrogen atoms, which is in accordance with the typical distance between H-bond donors and acceptors in proteins (2.5-3.5 Å). To confirm the position of the side chains involved within the proposed polar contacts, electron density of 2*Fo-Fc* map contoured at 1 σ around the highlighted residues is also displayed in Figure 34C.

These interactions are localized within the interfacing areas of S256-G270 and P216-Y220 (in Figure 34 B in pink and green, respectively). The third major interfacing region consisting of A172-L177 in both chains contains the intermolecular bridge in wild-type CA IX, instead of which two oppositely facing serins are located herein (in violet in Figure 34C).

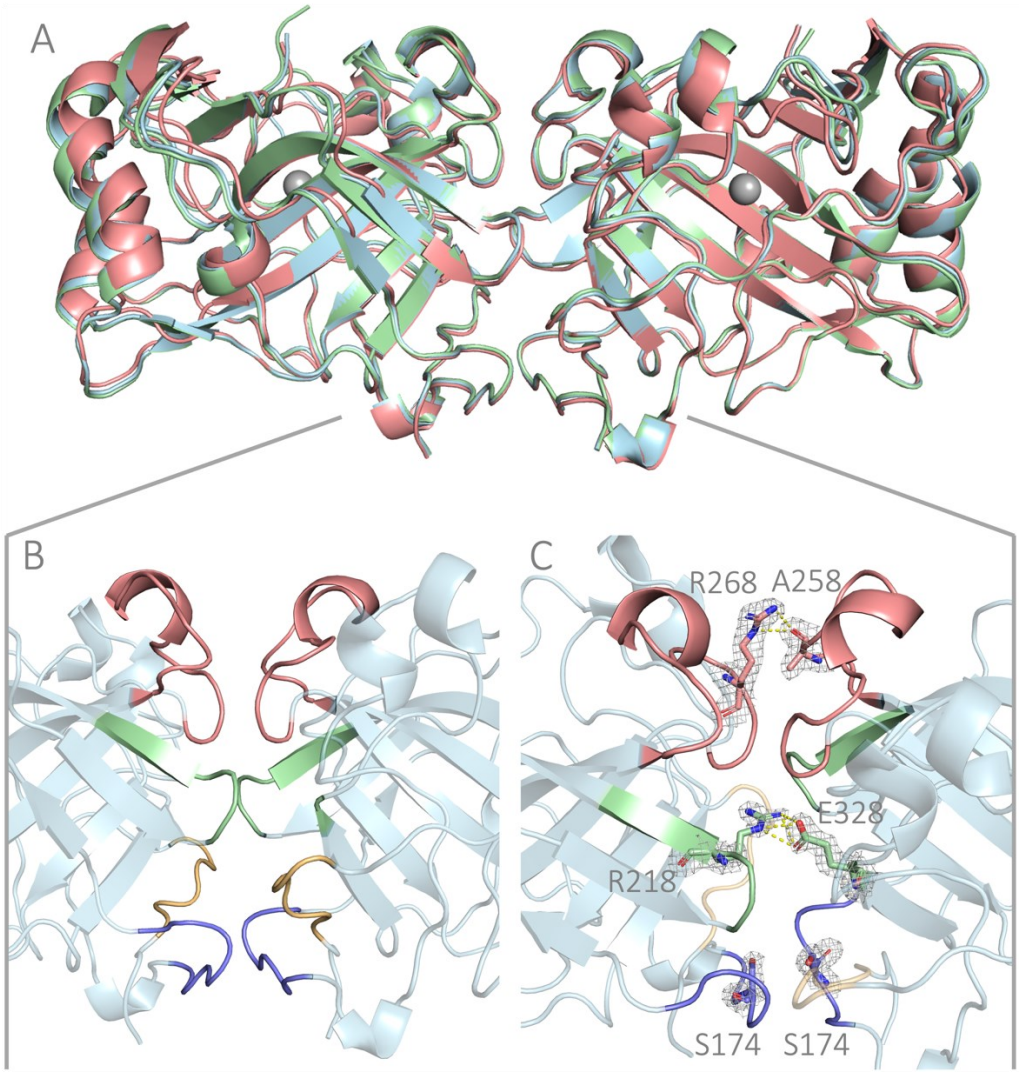


Figure 34: Dimeric structures of CAIX/4 #1 and #2 and CAIX/5G. A) Superposition of all three solved structures with RMSD range of 0.202-0.722 where 449/445/469 Cα atoms were aligned. CAIX/4 #1 in light blue, CAIX/4 #2 in green, CAIX/5G in pink. B) Interfacing regions consisting of residues A172-L177 (violet), P216-Y220 (green), S256-G270 (pink), N382-E387 (yellow) and shown on CAIX/4 #1 example. C) Detail of polar contacts formed between two chains. The H-bond between R268 side chain and A258 backbone (displayed as sticks in pink) was identified in all three structures, herein on example of CAIX/4 #1. Complex salt bridge between charged residues of R218 and E328 was identified within the central region of the CAIX/4 #1 interface (displayed as sticks in green). In addition, the serins in position 174 are displayed (as sticks in violet) to address where the intermolecular disulfide bond is localized in wild-type CA IX. The figures were made in PyMol software.

6 Discussion

The presented diploma thesis reports structural studies of cancer-related enzyme carbonic anhydrase IX. The project included design of protein variants with different amino acid substitutions, preparation of highly pure protein samples and experimental structure determination at high resolution. The CA IX is overexpressed in cells of various solid tumors as a transmembrane protein with the major part exposed to the extracellular environment. Recombinant protein preparation of the extracellular region has been published before, typically for the common catalytic domain, or for the PG domain as individual (Langella *et al*, 2018). There are two instances where protein containing both PG and catalytic domain were reported in structural studies (Alterio *et al*, 2009; Koruza *et al*, 2020). Number of amino acid substitutions was introduced to numerous CA IX protein variants summarized in Table 4 at page 22. These variants were designed and prepared to help in recombinant protein production and crystallization as exemplified by protein variant bearing six amino acid substitutions (C174S, L183S, A213K, A258K, F259Y, M350S) allowing high-yield production in *E. coli* and reproducible crystallization (Mahon *et al*, 2016). This variant, referred as CAIX/6 in this work, has amino acid substitutions that affect the native dimeric interface of CA IX (residues 258-259) and the aim of this work was to restore the dimeric arrangement of CA IX and still have protein variant with favorable production and crystallization properties.

Thus, three novel protein variants were designed and produced together with the parental variant CAIX/6 (Mahon *et al*, 2016). Those four variants were expressed in two sets comprising only the catalytic domain (ΔPG-CAIX/x) or full extracellular region containing PG and CA domains (CAIX/x). All together eight recombinant CA IX protein variants were expressed in *E. coli* competent strains BL21 (DE3) or NiCo21 (DE3). Ultimately seven protein variants were prepared at high purity by different purification strategies (see Figure 12). The purification protocol (described in chapter 5.2) partially followed protocol for CA IX published before (Mahon *et al*, 2016), but was optimized according to the purity of protein variants after each step. The protein variant ΔPG-CAIX/6 which corresponds to the one characterized previously, was purified in one-step protocol by affinity chromatography with *p*-aminomethylbenzenesulfonamide-agarose. For ΔPG-CAIX/5G and CAIX/6 the following purification step was anion-exchange chromatography. Other protein variants (ΔPG-CAIX/4, CAIX/4, CAIX/5S, CAIX/5G) were purified by size-exclusion chromatography in the second step. What distinguishes the heterologous production described herein and in Mahon *et al* (2016) from other protocols used for structural and biochemical studies of CA IX is the expression within bacterial system and initial capture of CA IX protein by affinity chromatography based on the specific binding of CA to sulfonamide moiety. Originally, the purification of CA IX protein for biochemical and structural studies was established in insect- or yeast-based eukaryotic systems with CA IX protein variant in conjunction with His-tag and Ni-affinity column as first purification step (Hilvo *et al*, 2008; Alterio *et al*, 2009; Leitans *et al*, 2015).

The yield of purified protein samples in this thesis varied within range of 0.3-4.4 mg per 1 L of bacterial cultures cultivated in LB broth (Table 15). These values are lower than the yield claimed for ΔPG-CAIX/6 protein variant (10-20 mg/L) in previous research (Mahon *et al*, 2016; Koruza *et al*, 2019). Interestingly, when the CAIX/6 protein variant was produced in minimal media for $^{15}N/^{13}C$-labelled sample, it manifested higher yield of 4.6 mg/L, which is uncommon as lower yield is typical for recombinant production in MM. On the other hand, the yield of CAIX/4 variant as ^{15}N-labelled sample was as low as 0.3 mg/L.

The purification steps and final purity were examined by SDS-PAGE electrophoresis during which the protein variants with both PG and catalytic domain exhibited abnormal migration within the gel (as described in chapter 5.2.2.2). This phenomenon of apparent higher molecular weight was observed formerly for CA IX protein analyzed directly from cancer cell lines (Opavský *et al*, 1996; Li *et al*, 2011) and during recombinant protein production (Hilvo *et al*, 2008; Koruza *et al*, 2020), where they attributed it to the possible post-translational modifications such as glycosylation. However, that is highly improbable explanation for the results presented herein as protein glycosylation does not occur in bacterial systems (Walsh, 2010). Therefore, different reasons for the abnormal migration might be considered. First, unproper laboratory practice and protocol, such as poor SDS-PAGE gel preparation or insufficiently denaturated proteins can lead to odd results. However, this reasoning is not probable because the abnormal migration results were consistent and obtained only for CAIX protein variants and never for any ΔPG-CAIX. Hence, the apparent higher MW of protein variants with PG domain may be caused by its amino acid content with 15 proline residues and 36 acidic residues (when residues 38-137 are considered). These amino acid residues cause decreased electrophoretic mobility by structural rigidity formed by prolines with cyclic side chain (Bio Rad tech note 3133), next to that both prolines and acidic residues can affect the binding of SDS to the protein (Tiwari *et al*, 2019).

Similarly, an atypical migration of CAIX/x protein variants was observed during size exclusion chromatography procedure, which was used simultaneously as final purification step and for analysis of the oligomeric state. Since the biologically relevant oligomerization arrangement of CA IX on cancer cell is proposed to be homodimeric (Švastová *et al*, 2004; Li *et al*, 2011), one of the aims of this thesis was to elucidate the oligomerization state of recombinantly prepared proteins in solution. Previously reported biochemical and structural studies declared both monomeric and dimeric populations within their protein samples, demonstrated as two peaks in SEC chromatograms. Results presented herein exhibited different elution curves for distinct protein variants. All protein variants without PG domain (ΔPG-CAIX/6, ΔPG-CAIX/4, ΔPG-CAIX/5G) were eluted within one peak that properly corresponded to their MW weight of monomer by calculation from the calibration curve (~28 kDa). While this agrees with the experimental results discussed for ΔPG-CAIX/6 variant in Mahon *et al* (2016), the actual chromatograms are not reported and thus could not be compared with our results.

On the other hand, protein variants with PG domain showed three or two peaks on the corresponding chromatograms of CAIX/6 (Figure 16) or CAIX/4, CAIX/5G, CAIX/5S (Figure 17), respectively. However, in all instances only one peak from each experiment contained the desirable protein as confirmed by SDS-PAGE electrophoresis. These peaks apparent MW were about 110 kDa for CAIX/4, CAIX/5G, CAIX/5S, and about 210 kDa for CAIX/6 by calculation from the SEC calibration curves. Considering the expected molecular weight for different oligomeric states, any CAIX/x protein variant is theoretically about 38.9 kDa as monomer, therefore about 78 kDa for dimer, about 116 kDa for trimer and so on. When these values are taken into account together with SEC analysis published for CA IX form of both PG and catalytic domain (Hilvo *et al*, 2008; Koruza *et al*, 2020), it may be suggested that these protein variants appear with exaggerated MW due to non-globular features of the PG domain.

It should be clarified that the previously published SEC results, for which monomeric-dimeric equilibrium was concluded, are of CA IX protein variant without any amino acid substitutions. It was identified that the dimer is stabilized by covalent bond between two cysteine residues (C174) forming covalent disulfide bond both in recombinant protein (Hilvo *et al*, 2008) and in CA IX isolated from cancer cells (Švastová *et al*, 2004; Li *et al*, 2011). Subsequently, this interaction was confirmed in structural study, when the dimer of CA IX catalytic domains was resolved by X-ray crystallography. Even when the residue C174 was mutated to serin the same interface was preserved by polar interactions within region 267-270, and was declared to be the biological dimeric arrangement (Alterio *et al*, 2009).

Since the protein variant CAIX/6 has two mutations within this dimerization region we hypothesized that these may interrupt formation of the dimer. However, monomeric nature of CAIX/6 was not confirmed by the SEC analysis herein, even if the unusual migration is considered, since the protein seems to form high MW aggregates (Figure 16). On the other hand, the apparent MW of protein variants CAIX/4, CAIX/5G, CAIX/5S is indicative of the dimer. But as this result was not shown for the same protein variants without PG domain (ΔPG-CAIX/4, ΔPG-CAIX/5G), it may be argued that this domain may play a role in the dimerization. That is in contrast with monomeric and dimeric populations declared for CA IX catalytic domain variants (Hilvo *et al*, 2008; Koruza *et al*, 2018, 2019), although it should be noted that the SEC chromatograms are either not shown at all, the peaks content is not confirmed by SDS-PAGE analysis and the apparent MW is not stated in these reports.

Since the identification of protein oligomeric state in solution is quite a complex problem, as was demonstrated on number of protein examples from *E. coli* (Marciano *et al*, 2022), we sought to investigate the quaternary structure of CA IX by additional methods. Thus, two protein variants CAIX/6 and CAIX/4 were chosen to be produced as isotopically labelled samples that can be examined by NMR spectroscopy. The resulting rotational correlation time and subsequent calculation of MW were indicative of CAIX/6 being monomer and CAIX/4 forming dimer (Figure 21). Next to that mass photometry was utilized to examine the oligomerization at low protein concentration. Comparably,

CAIX/6 was found only as population of monomers, and CAIX/4 as two populations consisting of monomers and dimers suggesting equilibrium of these two quaternary forms similarly to the one described for non-mutated CAIX. These results support the initial hypothesis that the six mutations hinder the dimeric arrangement. However, to further explore the oligomeric state of CA IX recombinant proteins in solution, more accurate methods for determination of protein molecular weight and particles size could be used in future studies, such as SEC-MALS, native mass spectrometry or SAXS.

Regardless their oligomeric state five protein variants (ΔPG-CAIX/6, CAIX/6, CAIX/4, CAIX/5S and CAIX/5G) were subjected to crystallization with the aim to perform X-ray diffraction analysis. Protein crystallization is an intricate process during which the supersaturated protein molecules form crystals thanks to specific conditions within the experiment. These conditions cannot be predicted and the success of crystallization experiment is given by the distinct technique, precipitant solution composition and other factors (reviewed in (McPherson & Gavira, 2014). During this project vapor-diffusion techniques with both sitting- and hanging-drop were employed. At the beginning, the precipitant solution that yielded crystalline material was searched by experiments with commercial screening kits in sitting-drop arrangement. To investigate as many conditions as possible with the limiting amount of protein sample, robotic systems were utilized in combination with four screening kits (JSCG Plus, Morpheus®, PEGs Suite, SG1), later restricted to usage of only one (JSGC Plus). This approach resulted in both crystals of insufficient quality for X-ray analysis that were further optimized, and well-formed crystals necessary to structure determination.

Although the six mutations were originally designed to ease the crystallization process (Mahon *et al*, 2016; Koruza *et al*, 2019), neither of the two forms presented herein (ΔPG-CAIX/6, CAIX/6) yielded crystals adequate for X-ray experiment. The initial crystallization screening produced some positive crystallization events (clustered needles or plates) in one condition which was further exploited during optimization process by microseeding, streak seeding, buffer change based on differential scanning fluorimetry experiment, precipitant composition adjustments and additive screen. All together these approaches ended up in more than 2000 experimental setups in which the best outcomes were similarly to the initial result clustered plates or needles. That is in contrast with the previous experiments, when crystals of ΔPG-CAIX/6 were obtained both in sitting- and hanging-drop together with microseeding optimization (Mahon *et al*, 2016; Koruza *et al*, 2019). It should be noted that it is quite common that previously reported crystallization cannot be reproduced. The crystallization, as described above, is affected by many factors that need to be optimized to achieve analogous results.

Three protein variants (CAIX/4, CAIX/5G, CAIX/5S) which were designed as part of this thesis to resemble the wild-type CA IX and remain the biologically relevant dimer formed three-dimensional crystals within the first screening experiment with commercial kit JSCG Plus. The positive crystallization events were found in numerous crystallization conditions and have not been further optimized in this work. The crystals typically appeared as rod clusters which were separated into single

crystals during mounting into loops for X-ray experiment. Ideally, high number of single crystals is desired for structural studies, that is even more profound for CA IX as cancer-target examined in the structure-based drug design research. Hence, the ability of these protein variants to form single crystals may be subject of future effort aimed at optimization of crystallization conditions. Any component of the crystallization solution impact the crystal quality, accordingly concentration of buffers, salt, ions, protein, as well as pH value, temperature and other variables might be changed to fine-tune the conditions which will repeatedly give rise to crystals (McPherson & Gavira, 2014). Considering the conditions that produced crystals and which are stated for each crystal example in chapter 5.4.2.2, it can be expected that the PEG molecules (PEG 3350 or PEG MME 2000) have major effect on the successful crystallization of CAIX protein variants.

The protein variants that formed crystals originally consisted of both PG and catalytic domain. But the fold characterization by NMR spectroscopy showed in this thesis and published before (Langella *et al*, 2018, 2021) identified the PG domain to be intrinsically disordered. The disordered features of PG domain were observed on 1D 1H NMR spectra of all prepared protein samples and on 2D 1H-15N HSQC spectra of $^{15}N/^{13}C$-CAIX/6 (Figure 23). The main difference from the previously published results is that herein the PG domain was prepared and evaluated in the context of the catalytic domain, while the previous studies worked with individual PG domain (Langella *et al*, 2018).

Generally, the disordered protein regions are improbable to form crystals (Oldfield *et al*, 2013), and therefore we explored whether the PG domain is present within the crystals that were formed for variants containing PG and CA domains. Selected crystals of CAIX/6 were subjected to SDS-PAGE electrophoretic analysis which displayed evident protein band at the MW of the catalytic domain only (Figure 31A). Based on that result a proteolytical cleavage was suspected to occur during the crystallization procedure. Trying to mimic this process and to distinguish if the proteolysis is spontaneous or caused by remaining proteases within the protein sample, CAIX/6 was stored at 37º C in the presence or absence of protease inhibitors. The result of this experiment was not conclusive as some cleavage was evident in both instances (Figure 31B). Eventually, the absence of PG domain was confirmed in the electron density maps when the structures were determined (Figure 32). The loss of the PG domain during crystallization and storage over time has been recognized before (Koruza *et al*, 2020).

The designated crystals of CAIX/4, CAIX/5G and CAIX/5S were subjected to X-ray experiment on synchrotron facilities. Ultimately three complete diffraction datasets were collected, processed, and used for structure determination of CA IX catalytic domain (CAIX/4 #1, CAIX/4 #2, CAIX/5G, see Table 18). The crystals exhibited diffraction up to resolution of 1.35-1.91 Å, which is comparable to already published structures (Table 3). All three structure were solved in space group $P2_1$ by molecular replacement with monomer from structure 3IAI (Alterio *et al*, 2009). The same space group was recognized for ΔPG-CAIX/6 protein variant in four structures 6RQN, 6RQQ, 6RQW, 6RQU

(Koruza *et al*, 2019). On the contrary, the only remaining structure of CAIX recombinant protein produced in *E. coli* (5DVX) was of space group $P2_12_12_1$ (Mahon *et al*, 2016). The discussed structures 6RQQ and 5DVX have two molecules of the catalytic domain inside the asymmetric unit, however, these dimeric assemblies do not correspond to the biological dimer as discussed by authors (Koruza *et al*, 2019).

Collectively the literature and analysis of interfaces showed in this thesis (Table 14) define the biological dimer of CA IX by interface consisting of residues A172-L177, P216-Y220, S256-G270, N382-E387, hydrogen bonds between side chain of R268 and backbone of A258 and possible interaction of side chains of residue174 (either of H-bond by serins, or covalent disulfide bond in wild-type CAIX). The three crystal structures of CAIX/4 and CAIX/5G determined in this work were analyzed to evaluate the dimeric interface. The same dimeric arrangement was found in all structures (Figure 34A), even though the structures CAIX/4 #1 and CAIX/4 #2 belonged to monoclinic crystal form with four chains within AU and CAIX/5G crystallized with two chains in the AU. The structural superposition of the solved structures with 3IAI (Figure 35A) displays high structural similarity of the dimers with RMSD value 0.426-0.749 Å for all Cα atoms, a value within the range observed for different crystal structures of identical proteins (Betts & Sternberg, 1999). On the contrary, when dimer of CAIX/4 was structurally aligned with the two chains of PDB entries 5DVX and 6RQQ, no overlay was observed (Figure 35B) because the dimeric arrangement is completely different.

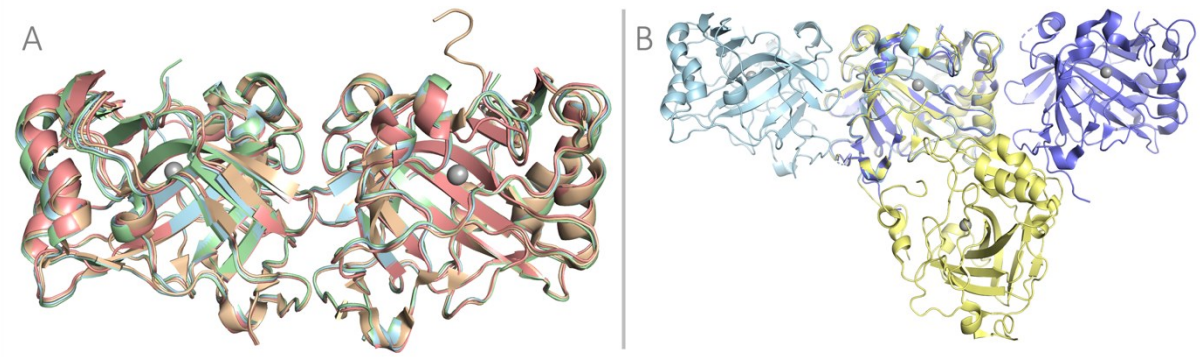


Figure 35: Comparison of CAIX catalytic domain structures solved as part of this thesis and previously published structures. A) Superposition of three solved structures with the biological assembly of structure 3IAI. Cα atoms 3IAI (gold) were aligned to CAIX/4 #1 (light blue), CAIX/4 #2 (green) and CAIX/5G (pink) with RMSD = 0.626, 0.749, 0.426 Å with number of atoms 486, 495, 450, respectively. B) Superposition of chains A from structure CAIX/4 #1 solved herein and 5DVX (violet) and 6RQQ (yellow). This figure shows that even though the monomeric structures of the catalytic domain are the same, their relevant dimeric arrangements are not. The figures were made in PyMol software.

The dimers of CAIX/4 and CAIX/5G were subjected to analysis of interfaces and interactions by PDBePISA and PyMol tools. As described in results, analysis confirmed the interfacing regions analogous to those defined for the biological interface (highlighted in Figure 34B). Moreover, the H-bond between R268 and A258 was found in all of them, with an additional salt bridge interaction identified in CAIX/4 #1 structure (shown in Figure 34C).

To further explore the impact of the six amino acid substitutions originally designed for expression in *E. coli* and crystallization (in 5DVX and 6RQQ) on the dimeric arrangement, these positions were highlighted and compared with CAIX/4 #1 bearing only four of these mutations (Figure 36). To begin with position 174, where the disulfide bond is present in wild-type CAIX, the serines in the structures presented herein are interfacing each other, contrarily in both 5DVX and 6RQQ they are not in close contact. Therefore, even if there was the cysteine, the covalent bond confirmed for native CA IX from cancer cells could not be formed in the 5DVX or 6RQQ dimeric arrangement. Regarding the residues S183, K213 and S350, these appear on the surface in all cases and seem to neither participate nor hinder the dimeric arrangement of CAIX/4, and they do not engage in interchain interactions of 5DVX and 6RQQ as described in Koruza *et al* (2019).

Finally, the residues at positions 258 and 269 shall be discussed. Initially during this project, the bulky residues of lysine (A258K) and tyrosine (F259Y) in the original protein variant (CAIX/6) were hypothesized to obstruct the formation of biological dimer both in solution and in crystal structures. From the structure of CAIX/4 where these positions were returned to the wild-type amino acid residues (A258, F259), it is evident that the interactions within this region are necessary to preserve the biological dimer of CA IX. To display the differences between the dimeric arrangements even more, the residues which are interfacing in the biological dimer were highlighted as well (Figure 36). Also of note, the biological dimeric arrangement given for structures solved herein and 3IAI were not found in any structure from *E. coli* even when symmetry mates were generated and searched.

To further support the native nature of the dimer found in our structures, we might consider its arrangement in the context of full-length CAIX. When the CAIX dimer is situated at plasma membrane, the N-termini of both chains should be at the same face of the dimer, and the C-termini on the opposite face (Alterio *et al*, 2009), that is not true for the protein variants with six mutations and specifically for the structure 5DVX (denoted in Figure 36). Indeed, the protein variant with six mutations originally designed for production in *E. coli* crystallizes as dimer but is truly monomer as declared by the author-assigned biological assembly within PDB (remark 350 in pdb file) (Mahon *et al*, 2016; Koruza *et al*, 2019).

To conclude, the amino acid substitutions to bulky residues at positions 258 and 259 impair the formation of biological dimer in CA IX protein variant ΔPG-CAIX/6 or CAIX/6. The physiological dimeric interface can be restored by re-introduction of wild-type amino acid residues A258 and F259. The unsettled question is whether the remaining four amino acid substitutions are truly needed for heterologous expression of the extracellular soluble part of CA IX in bacterial system of *E. coli*.

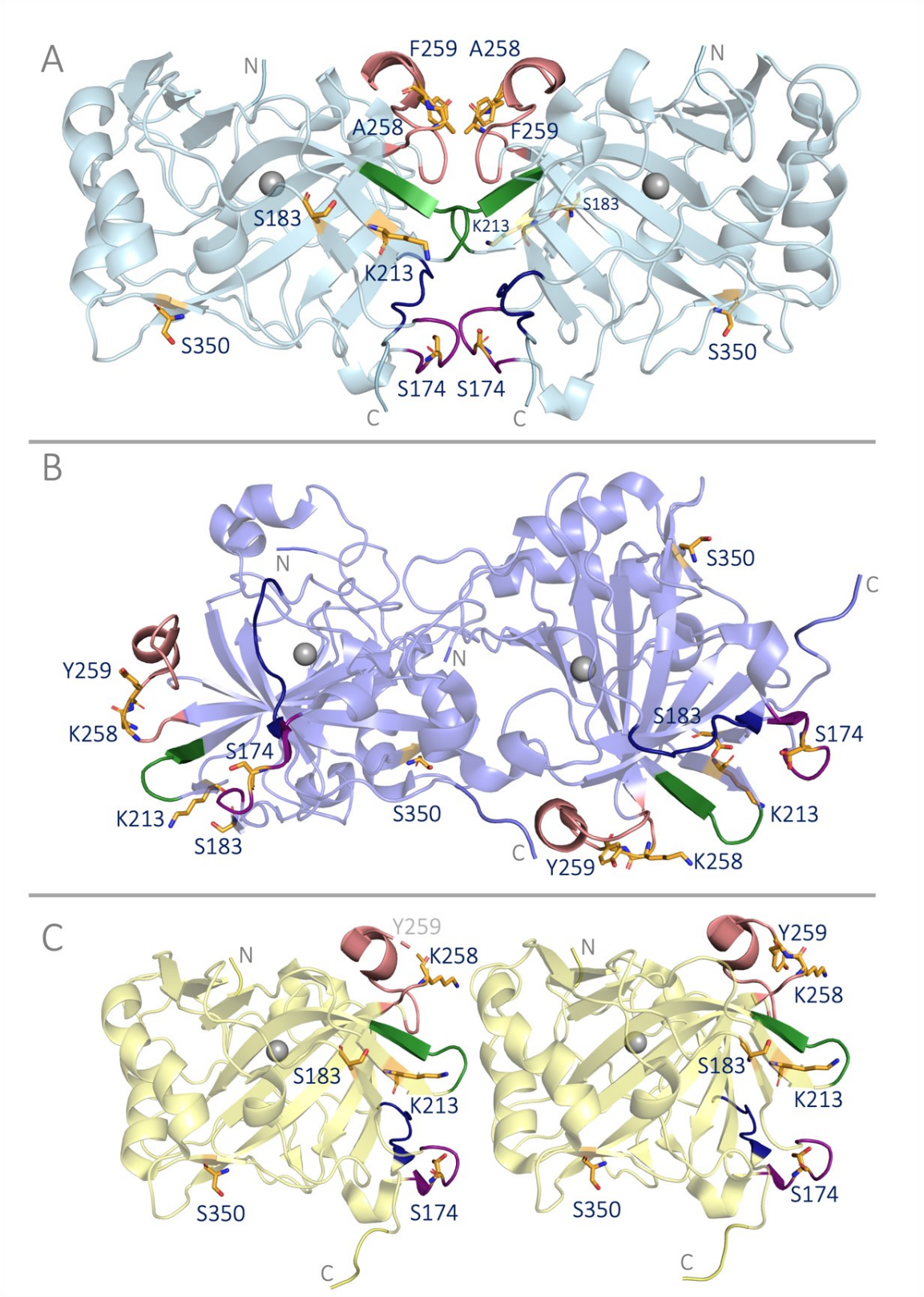


Figure 36: Examination of the impact of six amino acid substitutions in protein variant CAIX/6 in comparison to CAIX/4 bearing only four of them. Catalytic domains are shown as cartoons for CAIX/4 #1 in light blue (A), 5DVX in violet (B) and 6RQQ in yellow (C). The positions of amino acid residues originally mutated for CA IX expression in bacterial system (C174S, L183S, A213K, A258K, F259Y, M350S) and are showed as orange sticks and labelled according to the respective amino acid residue. The defined interfacing regions of biological dimer are highlighted; residues 172-177 (purple), 216-220 (green), 256-270 (pink), 382-387 (dark blue). Additionally, N-terminus and C-terminus are labelled. The figures were made in PyMol software.

In the end, it should be remembered that besides the discussed soluble part, the enzyme CA IX consists of transmembrane helix and short cytoplasmic tail. Since structural determination of transmembrane proteins by X-ray crystallography is rather challenging, one may consider other methods of structural biology to explore the whole homodimeric complex of transmembrane CA IX. Since this complex is theoretically about 100 kDa in MW, single-particle cryo-electron microscopy may be utilized in future studies. This technique could be potentially used to elucidate CA IX interactions with its proposed partners which cooperate on the surface of cancer cells (as reviewed in literature section, chapter 2.2.2). Apart from the experimentally determined structures, we can explore the CA IX by structure prediction software, such as such as ColabFold using AlphaFold2 and AlphaFold2-multimer (Mirdita *et al*, 2022). When discussing predicted structures, one should always consider the given statistics about per-residue model confidence (pLDDT; local Distance Difference Test) and the inter-domain accuracy (PAE; Predicted Aligned Error). From the CA IX predicted dimer consisting of residues 38-459 (Figure 37A) we can see that besides the catalytic domain, all other regions are predicted with very low confidence and the estimation of their relative orientation is poor. Nevertheless, as the protein regions of low confidence were showed to be usually disordered, it can be mentioned that this agrees with the described features of PG domain. Additionally, as the dimeric interface is predicted with high but not very high confidence, it can be speculated that some dynamics may occur there.

All together the outcome of this diploma thesis can be simplified into schematic of CA IX dimer consisting of well-folded catalytic domains which were resolved at nearly atomic level in combination with PG domain showed to be disordered, depicted within the context of cellular plasma membrane (Figure 37B).

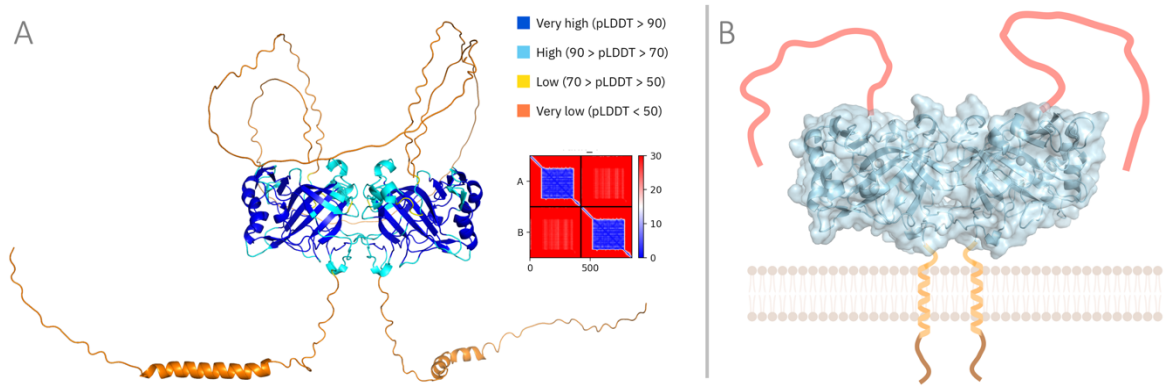


Figure 37: A) Structure of CA IX dimeric structure (residues 38-459) as predicted by ColabFold; the pLDDT code and PAE diagram (in red) are showed. B) Illustrative overview of CA IX dimer on plasma membrane depicting the outcome of this thesis consisting of the structurally determined catalytic domain (CAIX/4 #1 in light blue cartoon and transparent surface) and schematic PG domain characterized to be intrinsically disordered (pink). The figure was made in PyMol and BioRender software.

7 Conclusion

This diploma thesis aimed to structurally characterized extracellular region of cancer-related enzyme carbonic anhydrase IX. During this project three novel protein variants were developed and subjected to numerous experiments to characterize their biochemical properties. The submitted results elucidate amino acid residues which are needed to maintain the biologically relevant dimeric arrangement of CA IX. The gained knowledge may be utilized for future development in structure-based drug design project, which is addressed in the host Laboratory of Structural Biology at IOCB, Prague. The following outcomes have been achieved in this thesis:

- Three protein variants (CAIX/4, CAIX/5S, CAIX/5G) were designed based on *in silico* analysis of CA IX structures in PDB database.
- The designed protein variants and one protein variant published before (CAIX/6) were expressed in *E. coli* in two sets – consisting only of the catalytic domain (ΔPG-CAIX/x), or both catalytic and PG domain (CAIX/x).
- Purification protocols were optimized to obtain protein samples of high purity convenient for structural studies.
- Two protein variants (CAIX/6, CAIX/4) were prepared as isotopically labelled protein samples for NMR spectroscopy experiments ($^{15}N/^{13}C$-CAIX/6, ^{15}N-CAIX/4).
- Protein variants were investigated for their oligomeric state by size-exclusion chromatography, NMR spectroscopy and mass photometry. These experiments identified protein variants (ΔPG-CAIX/6, ΔPG-CAIX/5G, ΔPG-CAIX/4, CAIX/6) to be monomeric and protein variants (CAIX/4, CAIX/5G, CAIX/5S) to be dimeric in solution.
- All protein variants were examined for their overall fold by 1D 1H NMR spectroscopy experiment which revealed the well-folded globular catalytic domain and intrinsically disordered properties of the PG domain.
- Five protein variants (ΔPG-CAIX/6, CAIX/6, CAIX/4, CAIX/5G, CAIX/5S) were subjected for crystallization procedure. Three dimensional crystals suitable for X-ray diffraction experiment were gained for protein variants CAIX/6, CAIX/5G, CAIX/5S.
- The crystals were subjected for X-ray diffraction experiments at synchrotron facilities and three protein structures of the catalytic domain were determined and refined to resolution 1.35-1.9 Å (CAIX/4 #1, CAIX/4 #2, CAIX/5G).
- The biological dimeric arrangement was recognized in all three determined structures. The regions of interface were identified, and it was showed that H-bond between residues R268 and A258 is crucial to maintain the dimer.

In conclusion, the protein variants prepared and characterized in this work represent a great tool for structural studies of CA IX as they can be produced in good yields from *E. coli* expression system, and they crystallize reproducibly.

8 References

* References marked with asterisk are reviews

Agirre J, Atanasova M, Bagdonas H, Ballard CB, Baslé A, Beilsten-Edmands J, Borges RJ, Brown DG, Burgos-Mármol JJ, Berrisford JM, *et al* (2023) The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr Sect Struct Biol* 79: 449–461

Alterio V, Hilvo M, Di Fiore A, Supuran CT, Pan P, Parkkila S, Scaloni A, Pastorek J, Pastorekova S, Pedone C, *et al* (2009) Crystal structure of the catalytic domain of the tumor-associated human carbonic anhydrase IX. *Proc Natl Acad Sci U S A* 106: 16233–16238

Alterio V, Pan P, Parkkila S, Buonanno M, Supuran CT, Monti SM & De Simone G (2014) The structural comparison between membrane-associated human carbonic anhydrases provides insights into drug design of selective inhibitors. *Biopolymers* 101: 769–778

Ames S, Pastorekova S & Becker HM (2018) The proteoglycan-like domain of carbonic anhydrase IX mediates non-catalytic facilitation of lactate transport in cancer cells. *Oncotarget* 9: 27940–27957

*Aspatwar A, Tolvanen MEE, Barker H, Syrjänen L, Valanne S, Purmonen S, Waheed A, Sly WS & Parkkila S (2022) Carbonic anhydrases in metazoan model organisms: molecules, mechanisms, and physiology. *Physiol Rev* 102: 1327–1383

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN & Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242

Betts MJ & Sternberg MJ (1999) An analysis of conformational changes on protein-protein association: implications for predictive docking. *Protein Eng* 12: 271–283

Buanne P, Renzone G, Monteleone F, Vitale M, Monti SM, Sandomenico A, Garbi C, Montanaro D, Accardo M, Troncone G, *et al* (2013) Characterization of carbonic anhydrase IX interactome reveals proteins assisting its nuclear localization in hypoxic cells. *J Proteome Res* 12: 282–292

Buonanno M, Langella E, Zambrano N, Succoio M, Sasso E, Alterio V, Di Fiore A, Sandomenico A, Supuran CT, Scaloni A, *et al* (2017) Disclosing the Interaction of Carbonic Anhydrase IX with Cullin-Associated NEDD8-Dissociated Protein 1 by Molecular Modeling and Integrated Binding Measurements. *ACS Chem Biol* 12: 1460–1465

Burley SK, Bhikadiya C, Bi C, Bittrich S, Chao H, Chen L, Craig PA, Crichlow GV, Dalenberg K, Duarte JM, *et al* (2023) RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res* 51: D488–D508

Chiche J, Ilc K, Laferrière J, Trottier E, Dayan F, Mazure NM, Brahimi-Horn MC & Pouysségur J (2009) Hypoxia-inducible carbonic anhydrase IX and XII promote tumor cell growth by counteracting acidosis through the regulation of the intracellular pH. *Cancer Res* 69: 358–368

Christianson HC, Menard JA, Chandran VI, Bourseau-Guilmain E, Shevela D, Lidfeldt J, Månsson A-S, Pastorekova S, Messinger J & Belting M (2017) Tumor antigen glycosaminoglycan modification regulates antibody-drug conjugate delivery and cytotoxicity. *Oncotarget* 8: 66960–66974

Csaderova L, Debreova M, Radvak P, Stano M, Vrestiakova M, Kopacek J, Pastorekova S & Svastova E (2013) The effect of carbonic anhydrase IX on focal contacts during cell spreading and migration. *Front Physiol* 4: 271

Debreczeni JÉ & Emsley P (2012) Handling ligands with Coot. *Acta Crystallogr D Biol Crystallogr* 68: 425–430

Debreova M, Csaderova L, Burikova M, Lukacikova L, Kajanova I, Sedlakova O, Kery M, Kopacek J, Zatovicova M, Bizik J, *et al* (2019) CAIX Regulates Invadopodia Formation through Both a pH-Dependent Mechanism and Interplay with Actin Regulatory Proteins. *Int J Mol Sci* 20: 2745

*Di Fiore A, Supuran CT, Scaloni A & De Simone G (2022) Post-translational modifications in tumor-associated carbonic anhydrases. *Amino Acids* 54: 543–558

Dierks K, Meyer A, Oberthür D, Rapp G, Einspahr H & Betzel C (2010) Efficient UV detection of protein crystals enabled by fluorescence excitation at wavelengths longer than 300 nm. *Acta Crystallograph Sect F Struct Biol Cryst Commun* 66: 478–484

Ditte P, Dequiedt F, Svastova E, Hulikova A, Ohradanova-Repic A, Zatovicova M, Csaderova

L, Kopacek J, Supuran CT, Pastorekova S, *et al* (2011) Phosphorylation of carbonic anhydrase IX controls its ability to mediate extracellular acidification in hypoxic tumors. *Cancer Res* 71: 7558–7567

Domsic JF, Avvaru BS, Kim CU, Gruner SM, Agbandje-McKenna M, Silverman DN & McKenna R (2008) Entrapment of carbon dioxide in the active site of carbonic anhydrase II. *J Biol Chem* 283: 30766–30771

Dorai T, Sawczuk IS, Pastorek J, Wiernik PH & Dutcher JP (2005) The role of carbonic anhydrase IX overexpression in kidney cancer. *Eur J Cancer Oxf Engl 1990* 41: 2935–2947

Doyen J, Parks SK, Marcié S, Pouysségur J & Chiche J (2012) Knock-down of hypoxia-induced carbonic anhydrases IX and XII radiosensitizes tumor cells by increasing intracellular acidosis. *Front Oncol* 2: 199

Dudutienė V, Zubrienė A, Kairys V, Smirnov A, Smirnovienė J, Leitans J, Kazaks A, Tars K, Manakova L, Gražulis S, *et al* (2020) Isoform-Selective Enzyme Inhibitors by Exploring Pocket Size According to the Lock-and-Key Principle. *Biophys J* 119: 1513–1524

Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501

Fujikawa-Adachi K, Nishimori I, Taguchi T & Onishi S (1999) Human carbonic anhydrase XIV (CA14): cDNA cloning, mRNA expression, and mapping to chromosome 1. *Genomics* 61: 74–81

Gerlach M, Mueller U & Weiss MS (2016) The MX beamlines BL14.1-3 at BESSY II. *J Large-Scale Res Facil JLSRF* 2: A47–A47

Gieling RG, Babur M, Mamnani L, Burrows N, Telfer BA, Carta F, Winum J-Y, Scozzafava A, Supuran CT & Williams KJ (2012) Antimetastatic Effect of Sulfamate Carbonic Anhydrase IX Inhibitors in Breast Carcinoma Xenografts. *J Med Chem* 55: 5591–5600

*Hanahan D & Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646–674

Hilvo M, Baranauskiene L, Salzano AM, Scaloni A, Matulis D, Innocenti A, Scozzafava A, Monti SM, Di Fiore A, De Simone G, *et al* (2008) Biochemical characterization of CA IX, one of the most active carbonic anhydrase isozymes. *J Biol Chem* 283: 27799–27809

Hulikova A, Zatovicova M, Svastova E, Ditte P, Brasseur R, Kettmann R, Supuran CT, Kopacek J, Pastorek J & Pastorekova S (2009) Intact intracellular tail is critical for proper functioning of the tumor-associated, hypoxia-regulated carbonic anhydrase IX. *FEBS Lett* 583: 3563–3568

Ilardi G, Zambrano N, Merolla F, Siano M, Varricchio S, Vecchione M, De Rosa G, Mascolo M & Staibano S (2014) Histopathological determinants of tumor resistance: a special look to the immunohistochemical expression of carbonic anhydrase IX in human cancers. *Curr Med Chem* 21: 1569–1582

Innocenti A, Pastorekova S, Pastorek J, Scozzafava A, De Simone G & Supuran CT (2009) The proteoglycan region of the tumor-associated carbonic anhydrase isoform IX acts as anintrinsic buffer optimizing CO2 hydration at acidic pH values characteristic of solid tumors. *Bioorg Med Chem Lett* 19: 5825–5828

Ivanova JN, Nocentini A, Ta Rs K, Leita Ns JN, Dvinskis E, Kazaks A, Domračeva I, Supuran CT & Žalubovskis R (2023) Atropo/Tropo Flexibility: A Tool for Design and Synthesis of Self-Adaptable Inhibitors of Carbonic Anhydrases and Their Antiproliferative Effect. *J Med Chem* 66: 5703–5718

Jamali S, Klier M, Ames S, Barros LF, McKenna R, Deitmer JW & Becker HM (2015) Hypoxia-induced carbonic anhydrase IX facilitates lactate flux in human breast cancer cells by non-catalytic function. *Sci Rep* 5: 13605

*Kanchanawong P & Calderwood DA (2023) Organization, dynamics and mechanoregulation of integrin-mediated cell-ECM adhesions. *Nat Rev Mol Cell Biol* 24: 142–161

Kantardjieff KA & Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci Publ Protein Soc* 12: 1865–1871

Kazokaitė J, Niemans R, Dudutienė V, Becker HM, Leitāns J, Zubrienė A, Baranauskienė L, Gondi G, Zeidler R, Matulienė J, *et al* (2018) Novel fluorinated carbonic anhydrase IX inhibitors reduce hypoxia-induced acidification and clonogenic survival of cancer cells. *Oncotarget* 9: 26800–26816

Koruza K, Lafumat B, Nyblom M, Mahon BP, Knecht W, McKenna R & Fisher SZ (2019) Structural comparison of protiated, H/D-exchanged and deuterated human carbonic anhydrase IX. *Acta*

Crystallogr Sect Struct Biol 75: 895–903

Koruza K, Lafumat B, Végvári Á, Knecht W & Fisher SZ (2018) Deuteration of human carbonic anhydrase for neutron crystallography: Cell culture media, protein thermostability, and crystallization behavior. *Arch Biochem Biophys* 645: 26–33

Koruza K, Murray AB, Mahon BP, Hopkins JB, Knecht W, McKenna R & Fisher SZ (2020) Biophysical Characterization of Cancer-Related Carbonic Anhydrase IX. *Int J Mol Sci* 21: 5277

Krissinel E (2010) Crystal contacts as nature's docking solutions. *J Comput Chem* 31: 133–143

Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372: 774–797

*Langella E, Buonanno M, De Simone G & Monti SM (2021) Intrinsically disordered features of carbonic anhydrase IX proteoglycan-like domain. *Cell Mol Life Sci CMLS* 78: 2059–2067

Langella E, Buonanno M, Vullo D, Dathan N, Leone M, Supuran CT, De Simone G & Monti SM (2018) Biochemical, biophysical and molecular dynamics studies on the proteoglycan-like domain of carbonic anhydrase IX. *Cell Mol Life Sci CMLS* 75: 3283–3296

Lee D, Hilty C, Wider G & Wüthrich K (2006) Effective rotational correlation times of proteins from NMR relaxation interference. *J Magn Reson San Diego Calif 1997* 178: 72–76

Lee S-H, McIntyre D, Honess D, Hulikova A, Pacheco-Torres J, Cerdán S, Swietach P, Harris AL & Griffiths JR (2018) Carbonic anhydrase IX is a pH-stat that sets an acidic tumour extracellular pH in vivo. *Br J Cancer* 119: 622–630

Leitans J, Kazaks A, Balode A, Ivanova J, Zalubovskis R, Supuran CT & Tars K (2015) Efficient Expression and Crystallization System of Cancer-Associated Carbonic Anhydrase Isoform IX. *J Med Chem* 58: 9004–9009

Leitans J, Kazaks A, Bogans J, Supuran CT, Akopjana I, Ivanova J, Zalubovskis R & Tars K (2023) Structural Basis of Saccharin Derivative Inhibition of Carbonic Anhydrase IX. *ChemMedChem* 18: e202300454

Li Y, Wang H, Tu C, Shiverick KT, Silverman DN & Frost SC (2011) Role of hypoxia and EGF on expression, activity, localization and phosphorylation of carbonic anhydrase IX in MDA-MB-231 breast cancer cells. *Biochim Biophys Acta BBA - Mol Cell Res* 1813: 159–167

*Linder S, Cervero P, Eddy R & Condeelis J (2023) Mechanisms and roles of podosomes and invadopodia. *Nat Rev Mol Cell Biol* 24: 86–106

Liu Y, Engelman DM & Gerstein M (2002) Genomic analysis of membrane protein families: abundance and conserved motifs. *Genome Biol* 3: research0054.1-research0054.12

Lou Y, McDonald PC, Oloumi A, Chia S, Ostlund C, Ahmadi A, Kyle A, Auf dem Keller U, Leung S, Huntsman D, *et al* (2011) Targeting tumor hypoxia: suppression of breast tumor growth and metastasis by novel carbonic anhydrase IX inhibitors. *Cancer Res* 71: 3364–3376

Mahon BP, Bhatt A, Socorro L, Driscoll JM, Okoh C, Lomelino CL, Mboge MY, Kurian JJ, Tu C, Agbandje-McKenna M, *et al* (2016) The Structure of Carbonic Anhydrase IX Is Adapted for Low-pH Catalysis. *Biochemistry* 55: 4642–4653

Marciano S, Dey D, Listov D, Fleishman SJ, Sonn-Segev A, Mertens H, Busch F, Kim Y, Harvey SR, Wysocki VH, *et al* (2022) Protein quaternary structures in solution are a mixture of multiple forms. *Chem Sci* 13: 11680–11695

Maupin CM, McKenna R, Silverman DN & Voth GA (2009) Elucidation of the proton transport mechanism in human carbonic anhydrase II. *J Am Chem Soc* 131: 7598–7608

McDonald PC, Chafe SC, Brown WS, Saberi S, Swayampakula M, Venkateswaran G, Nemirovsky O, Gillespie JA, Karasinska JM, Kalloger SE, *et al* (2019) Regulation of pH by Carbonic Anhydrase 9 Mediates Survival of Pancreatic Cancer Cells With Activated KRAS in Response to Hypoxia. *Gastroenterology* 157: 823–837

*McPherson A & Gavira JA (2014) Introduction to protein crystallization. *Acta Crystallogr Sect F Struct Biol Commun* 70: 2–20

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S & Steinegger M (2022) ColabFold: making protein folding accessible to all. *Nat Methods* 19: 679–682

*Mishra CB, Tiwari M & Supuran CT (2020) Progress in the development of human carbonic anhydrase inhibitors and their pharmacological applications: Where are we today? *Med Res Rev* 40: 2485–2565

Mueller U, Darowski N, Fuchs MR, Förster R, Hellmig M, Paithankar KS, Pühringer S, Steffien

M, Zocher G & Weiss MS (2012) Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J Synchrotron Radiat* 19: 442–449

Mueller U, Förster R, Hellmig M, Huschmann FU, Kastner A, Malecki P, Pühringer S, Röwer M, Sparta K, Steffien M, *et al* (2015) The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. *Eur Phys J Plus* 130: 141

Müller I (2017) Guidelines for the successful generation of protein-ligand complex crystals. *Acta Crystallogr Sect Struct Biol* 73: 79–92

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F & Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67: 355–367

Oldfield CJ, Xue B, Van Y-Y, Ulrich EL, Markley JL, Dunker AK & Uversky VN (2013) Utilization of protein intrinsic disorder knowledge in structural proteomics. *Biochim Biophys Acta* 1834: 487–498

Opavský R, Pastoreková S, Zelník V, Gibadulinová A, Stanbridge EJ, Závada J, Kettmann R & Pastorek J (1996) Human MN/CA9 gene, a novel member of the carbonic anhydrase family: structure and exon to protein domain relationships. *Genomics* 33: 480–487

Pastoreková S, Závadová Z, Kostál M, Babusíková O & Závada J (1992) A novel quasi-viral agent, MaTu, is a two-component system. *Virology* 187: 620–626

Pérez-Sayáns M, Suárez-Peñaranda JM, Pilar G-D, Supuran CT, Pastorekova S, Barros-Angueira F, Gándara-Rey JM & García-García A (2012) Expression of CA-IX is associated with advanced stage tumors and poor survival in oral squamous cell carcinoma patients. *J Oral Pathol Med Off Publ Int Assoc Oral Pathol Am Acad Oral Pathol* 41: 667–674

Pinard MA, Mahon B & McKenna R (2015) Probing the surface of human carbonic anhydrase for clues towards the design of isoform specific inhibitors. *BioMed Res Int* 2015: 453543

Písačková J, Procházková K, Fábry M & Řezáčová P (2013) Crystallization of the Effector-Binding Domain of Repressor DeoR from Bacillus subtilis. *Cryst Growth Des* 13: 844–848

Radvak P, Repic M, Svastova E, Takacova M, Csaderova L, Strnad H, Pastorek J, Pastorekova S & Kopacek J (2013) Suppression of carbonic anhydrase IX leads to aberrant focal adhesion and decreased invasion of tumor cells. *Oncol Rep* 29: 1147–1153

Robertson N, Potter C & Harris AL (2004) Role of Carbonic Anhydrase IX in Human Tumor Cell Growth, Survival, and Invasion. *Cancer Res* 64: 6160–6165

Robson SA, Dağ Ç, Wu H & Ziarek JJ (2021) TRACT revisited: an algebraic solution for determining overall rotational correlation times from cross-correlated relaxation rates. *J Biomol NMR* 75: 293–302

Rohani N, Hao L, Alexis MS, Joughin BA, Krismer K, Moufarrej MN, Soltis AR, Lauffenburger DA, Yaffe MB, Burge CB, *et al* (2019) Acidification of Tumor at Stromal Boundaries Drives Transcriptome Alterations Associated with Aggressive Phenotypes. *Cancer Res* 79: 1952–1966

Sáenz-de-Santa-María I, Bernardo-Castiñeira C, Secades P, Bernaldo-de-Quirós S, Rodrigo JP, Astudillo A & Chiara M-D (2017) Clinically relevant HIF-1α-dependent metabolic reprogramming in oropharyngeal squamous cell carcinomas includes coordinated activation of CAIX and the miR-210/ISCU signaling axis, but not MCT1 and MCT4 upregulation. *Oncotarget* 8: 13730–13746

*Schofield CJ & Ratcliffe PJ (2004) Oxygen sensing by HIF hydroxylases. *Nat Rev Mol Cell Biol* 5: 343–354

Shin H-J, Rho SB, Jung DC, Han I-O, Oh E-S & Kim J-Y (2011) Carbonic anhydrase IX (CA9) modulates tumor-associated cell migration and invasion. *J Cell Sci* 124: 1077–1087

Silverman DN & Lindskog S (1988) The catalytic mechanism of carbonic anhydrase: implications of a rate-limiting protolysis of water. *Acc Chem Res* 21: 30–36

Silverman DN & McKenna R (2007) Solvent-mediated proton transfer in catalysis by carbonic anhydrase. *Acc Chem Res* 40: 669–675

Soltermann F, Foley EDB, Pagnoni V, Galpin M, Benesch JLP, Kukura P & Struwe WB (2020) Quantifying Protein-Protein Interactions by Molecular Counting with Mass Photometry. *Angew Chem Int Ed Engl* 59: 10774–10779

Sparta KM, Krug M, Heinemann U, Mueller U & Weiss MS (2016) XDSAPP2.0. *J Appl Crystallogr* 49: 1085–1092

Stams T, Nair SK, Okuyama T, Waheed A, Sly WS & Christianson DW (1996) Crystal structure

of the secretory form of membrane-associated human carbonic anhydrase IV at 2.8-A resolution. *Proc Natl Acad Sci U S A* 93: 13589–13594

Stillebroer AB, Mulders PFA, Boerman OC, Oyen WJG & Oosterwijk E (2010) Carbonic anhydrase IX in renal cell carcinoma: implications for prognosis, diagnosis, and therapy. *Eur Urol* 58: 75–83

*Supuran CT (2023) A simple yet multifaceted 90 years old, evergreen enzyme: Carbonic anhydrase, its inhibition and activation. *Bioorg Med Chem Lett* 93: 129411

Švastová E, Hulíková A, Rafajová M, Zat'ovičová M, Gibadulinová A, Casini A, Cecchi A, Scozzafava A, Supuran CT, Pastorek J, *et al* (2004) Hypoxia activates the capacity of tumor-associated carbonic anhydrase IX to acidify extracellular pH. *FEBS Lett* 577: 439–445

Svastova E, Witarski W, Csaderova L, Kosik I, Skvarkova L, Hulikova A, Zatovicova M, Barathova M, Kopacek J, Pastorek J, *et al* (2012) Carbonic anhydrase IX interacts with bicarbonate transporters in lamellipodia and increases cell migration via its catalytic domain. *J Biol Chem* 287: 3392–3402

Švastová E, Žilka N, Zat'ovičová M, Gibadulinová A, Čiampor F, Pastorek J & Pastoreková S (2003) Carbonic anhydrase IX reduces E-cadherin-mediated adhesion of MDCK cells via interaction with β-catenin. *Exp Cell Res* 290: 332–345

Swayampakula M, McDonald PC, Vallejo M, Coyaud E, Chafe SC, Westerback A, Venkateswaran G, Shankar J, Gao G, Laurent EMN, *et al* (2017) The interactome of metabolic enzyme carbonic anhydrase IX reveals novel roles in tumor cell migration and invadopodia/MMP14-mediated invasion. *Oncogene* 36: 6244–6261

*Swietach P, Boedtkjer E & Pedersen SF (2023) How protons pave the way to aggressive cancers. *Nat Rev Cancer* 23: 825–841

Swietach P, Patiar S, Supuran CT, Harris AL & Vaughan-Jones RD (2009) The Role of Carbonic Anhydrase 9 in Regulating Extracellular and Intracellular pH in Three-dimensional Tumor Cell Growths *. *J Biol Chem* 284: 20299–20310

Swietach P, Wigfield S, Cobden P, Supuran CT, Harris AL & Vaughan-Jones RD (2008) Tumor-associated carbonic anhydrase 9 spatially coordinates intracellular pH in three-dimensional multicellular growths. *J Biol Chem* 283: 20473–20483

Tiwari P, Kaila P & Guptasarma P (2019) Understanding anomalous mobility of proteins on SDS-PAGE with special reference to the highly acidic extracellular domains of human E- and N-cadherins. *Electrophoresis* 40: 1273–1281

Türeci O, Sahin U, Vollmar E, Siemer S, Göttert E, Seitz G, Parkkila AK, Shah GN, Grubb JH, Pfreundschuh M, *et al* (1998) Human carbonic anhydrase XII: cDNA cloning, expression, and chromosomal localization of a carbonic anhydrase gene that is overexpressed in some renal cell cancers. *Proc Natl Acad Sci U S A* 95: 7608–7613

*Uversky VN (2013) A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci* 22: 693–724

Vagin A & Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* 66: 22–25

Waheed A & Sly WS (2017) Carbonic anhydrase XII functions in health and disease. *Gene* 623: 33–40

*Walsh G (2010) Post-translational modifications of protein biopharmaceuticals. *Drug Discov Today* 15: 773–780

Whittington DA, Waheed A, Ulmasov B, Shah GN, Grubb JH, Sly WS & Christianson DW (2001) Crystal structure of the dimeric extracellular domain of human carbonic anhydrase XII, a bitopic membrane protein overexpressed in certain cancer tumor cells. *Proc Natl Acad Sci U S A* 98: 9545–9550

Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, *et al* (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* 27: 293–315

Wu D & Piszczek G (2020) Measuring the affinity of protein-protein interactions on a single-molecule level by mass photometry. *Anal Biochem* 592: 113575

Wykoff CC, Beasley NJ, Watson PH, Turner KJ, Pastorek J, Sibtain A, Wilson GD, Turley H, Talks KL, Maxwell PH, *et al* (2000) Hypoxia-inducible expression of tumor-associated carbonic anhydrases. *Cancer Res* 60: 7075–7083

Zakšauskas A, Čapkauskaitė E, Jezepčikas L, Linkuvienė V, Paketurytė V, Smirnov A, Leitans J, Kazaks A, Dvinskis E, Manakova E, *et al* (2020) Halogenated and di-substituted benzenesulfonamides as selective inhibitors of carbonic anhydrase isoforms. *Eur J Med Chem* 185: 111825

Zakšauskas A, Čapkauskaitė E, Paketurytė-Latvė V, Smirnov A, Leitans J, Kazaks A, Dvinskis E, Stančaitis L, Mickevičiūtė A, Jachno J, *et al* (2021) Methyl 2-Halo-4-Substituted-5-Sulfamoyl-Benzoates as High Affinity and Selective Inhibitors of Carbonic Anhydrase IX. *Int J Mol Sci* 23: 130

Zavada J, Zavadova Z, Machon O, Kutinova L, Opavsky R & Pastorek J (1997) Transient transformation of mammalian cells by MN protein, a tumor-associated cell adhesion molecule with carbonic anhydrase activity. *Int J Oncol* 10: 857–863

Závada J, Závadová Z, Pastorek J, Biesová Z, Jezek J & Velek J (2000) Human tumour-associated cell adhesion protein MN/CA IX: identification of M75 epitope and of the region mediating cell adhesion. *Br J Cancer* 82: 1808–1813

Závada J, Závadová Z, Pastoreková S, Ciampor F, Pastorek J & Zelník V (1993) Expression of MaTu-MN protein in human tumor cultures and in clinical specimens. *Int J Cancer* 54: 268–274

Zhu XL & Sly WS (1990) Carbonic anhydrase IV from human lung. Purification, characterization, and comparison with membrane carbonic anhydrase from human kidney. *J Biol Chem* 265: 8795–8801