WITS
UNIVERSITY

Understanding the inhibition and post-translational modification of soluble CLIC1 and its structural homologs

Roland Worth

598272

A thesis submitted in fulfilment of the requirements for the degree of a Doctor of Philosophy in Molecular and Cell Biology in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Professor H.W. Dirr

April 2020

Declaration

I, Roland Worth (598272), am a student registered for the degree of Doctor of Philosophy (PhD) in the academic year 2020.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this, or any other University.
- The information used in the Thesis has not been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature:  Date: 28-04-2020

Abstract

The chloride intracellular channel 1 (CLIC1) protein is a unique membrane protein considering that it can reversibly transit between a soluble and membrane-bound conformation, whereby the ion channel activity of the latter becomes inhibited by indanyloxyacetic acid-94 (IAA-94). It therefore remained uncertain if the constraint in the ion channel activity of CLIC1 was due to IAA-94 binding to its soluble conformation in order to prevent its structural transition into a membrane competent state and/or IAA-94 binding to the membrane conformation of CLIC1. Therefore, this study aimed at understanding the interaction between IAA-94 and soluble CLIC1 by performing isothermal titration calorimetry (ITC) experiments on CLIC1, together with human glutathione transferase P1-1 (hGSTP1-1) and *Escherichia coli* stringent starvation protein A (EcSspA), the structural homologs of CLIC1. Based on ITC, it was shown that IAA-94 did not bind to CLIC1, but instead it was shown to interact with hGSTP1-1 and EcSspA, the latter being a metamorphic protein whose ion channel activity is also inhibited by IAA-94. Interestingly, although IAA-94 is a structural analogue of ethacrynic acid (EAA) (an inhibitor for hGSTP1-1), the latter was shown to bind to CLIC1, but not to EcSspA. However, based on structural sequence alignment, it was apparent that the residues which coordinate the interaction between EAA and hGSTP1-1 were primarily conserved within CLIC1, whereas the residues from the buffer binding site (the proposed ligandin binding site (L-site)) of hGSTP1-1 were conserved within EcSspA. As a result, this provided insight as to why CLIC1 and EcSspA could not interact with IAA-94 and EAA, respectively. In fact, the crystal structure of hGSTP1-1 in complex with IAA-94 suggested that IAA-94 did bind at the buffer binding site, but due to crystal contacts, only the acetate moiety of IAA-94 could be solved. These findings therefore supported previous studies which suggested that the buffer binding site is also the L-site of hGSTP1-1. As a result, considering that the EAA and IAA-94 have distinct binding sites within hGSTP1-1 and that the EAA binding site is structurally conserved within CLIC1, the binding of EAA to the active site of CLIC1 was significant considering that it would inhibit the enzyme function of CLIC1 by masking the thiol group of Cys24. In fact, Cys24 was shown to be the main reactive cysteine residue in CLIC1 and in turn it was also shown to play a key role in maintaining the structural stability of CLIC1. Furthermore, Cys24 was shown to be a target site for *S*-nitrosylation and similar to the C24A mutation, it appeared that this PTM could influence the structural stability of CLIC1. Therefore, although CLIC1 did not bind to IAA-94, its ability to bind to EAA provides insight into the rational drug design of novel inhibitors that could effectively bind and stabilise the soluble conformation of CLIC1 in order to attenuate its structural transition into to a membrane-bound conformation.

Acknowledgments

Thank you to my supervisor, Professor H.W. Dirr, for your continuous support and guidance throughout my degree. Your enthusiasm for science was always inspirational and facilitated my curiosity for wanting to understand the fundamental principles of science.

Thank you to Dr Ramaesh Pandian for your involvement in solving the crystal structure of hGSTP1-1 in complex with IAA-94. This structure would not have been possible without your immense knowledge and experience in crystallography.

Thank you to Dr Vinuchakkaravarthy Thangaraj for your technical assistance in harvesting the protein crystals and for also coordinating the diffraction of the protein crystals at the DLS.

Thank you to my friends at the Protein Structure-Function Research Unit (PSFRU) for their continuous support and engaging scientific discussions. A great thanks also goes to Dr Ikechukwu Achilonu for your continuous enthusiasm and support in the laboratory.

Thank you to the National Research Foundation (NRF), WITS Postgraduate Research Award (PMA) and the Poliomyelitis Foundation (PRF) for offering me financial assistance throughout my degree. This work was also supported by the Global Challenge Research Fund (GCRF) through Science and Technology Facilities Council (STFC), grant number ST/R002754/1: Synchrotron Techniques for African Research and Technology (START).

Thank you to my family who have always been there to offer me love, patience and support. I treasured your continuous optimism and management throughout this journey.

Conference Outputs

1. The Molecular Bioscience Research Thrust (MBRT), University of the Witwatersrand (Johannesburg, South Africa), November 2019. Poster presentation. Regulating the redox state of soluble CLIC1 controls the propensity to transition to its membrane-bound conformation. Worth, R.W. and Dirr, H.W.
2. The 10th Cross-Faculty Postgraduate Symposium, University of the Witwatersrand (Johannesburg, South Africa), September 2019. Oral presentation. First place in the Faculty of Science. The interaction of IAA-94 with the soluble conformation of the CLIC1 protein and its structural homolog hGSTP1-1. Worth, R.W., Pandian, R., Thangaraj, V. and Dirr, H.W.
3. The 33rd annual symposium of the Protein Science Society, (Seattle, Washington), July 2019.
4. The Molecular Biosciences Research Thrust (MBRT), University of the Witwatersrand (Johannesburg, South Africa), November 2018. Oral presentation. Second place in the Faculty of Science. Understanding the inhibition of soluble CLIC1. Worth, R.W. and Dirr, H.W.
5. The European Molecular Biology Organisation (EMBO) membrane protein expression, purification and characterisation (mPEPC1) workshop, The Deutsches Elektronen-Synchrotron (DESY) (Hamburg, Germany), September 2018.
6. The Federation of the European Biochemical Societies (FEBS) hydrodynamic and thermodynamic analysis of biological macromolecules and their interactions workshop, Charles University (Prague, Czech Republic), September 2018.
7. The 5th European Crystallographic School (ECS5), Stellenbosch University (Stellenbosch, South Africa), July 2018.
8. The South Africa Association for Mass Spectrometry (SAAMS) introduction to liquid chromatography-mass spectrometry (LC-MS) workshop, Council for Scientific and Industrial Research (CSIR) (Pretoria, South Africa), February 2018.
9. The African Centre for Gene Technologies (ACGT) proteomics and structural mass spectrometry workshop, Council for Scientific and Industrial Research (CSIR) (Pretoria, South Africa), October 2017.

List of Abbreviations

$[M_{tot}]$	total macromolecule concentration
[Θ]	mean residue ellipticity
|F|	structure factor amplitude
$\|F_c\|$	calculated structure factor
$\|F_o\|$	observed structure factor
ΔC	change in the heat capacity
ΔC_p	change in the heat capacity under isobaric conditions
ΔG	change in Gibbs free energy
ΔH	change in enthalpy
ΔH_{bind}	change in binding enthalpy
ΔH_{ion}	change in ionisation enthalpy
ΔH_{obs}	observed change in enthalpy
ΔS	change in entropy
ΔS_{comp}	change in the translational, rotational and vibrational degrees of freedom of the molecules within a bound complex
ΔS_{conf}	change in the conformational degrees of freedom of a protein and ligand
ΔS_{solv}	change in entropy for the water molecules that are localised on the surface of the active site and ligand
ΔS_{sys}	total change in entropy for a system
ΔT	change in temperature
6-IAF	6-iodoacetamide fluorescein
A_{260}	absorbance at 260 nm
A_{280}	absorbance at 280 nm
A_{340}	absorbance at 340 nm
A9C	anthracene-9-carboxylic acid
AD	Alzheimer's disease
ANS	8-anilinonaphthalene-1-sulfonic acid
ASA	accessible surface area
AU	arbitrary unit
Aβ	β-amyloid protein
Bax	B-cell lymphoma 2 associated X protein

Bcl-2	B-cell lymphoma 2
Bcl-xL	B-cell lymphoma 2-extra-large
BITC	benzyl isothiocyanate
BLASTp	Basic Local Alignment Search Tool for proteins
c	concentration of a molecule
CCP4	Collaborative Computational Project Number 4
cDNA	complementary deoxyribonucleic acid
CDNB	2, 4-dinitrochlorobenzene
CFTR	cystic fibrosis transmembrane conductance regulator
cGMP	cyclic guanylate monophosphate
CLIC	chloride intracellular channel
CLIC1-C24A	C24A mutant of the chloride intracellular channel 1 protein
CLIC1-WT	wild-type chloride intracellular channel 1 protein
COOT	Crystallographic Object-Oriented Toolkit
CSC	cancer stem cell
c-value	Weismann c-constant value
DDA	data dependent acquisition
DEAE	diethylaminoethyl
DLS	Diamond Light Source
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNase I	deoxyriobonuclease I
DTT	dithiothreitol
E.coli	*Escherichia coli*
EAA	ethacrynic acid
EC	enzyme commission number
EcSspA	*Escherichia coli* stringent starvation protein A
EDTA	ethylenediaminetetraacetic acid
EHEC	enterohemorrhagic *Escherichia coli*
EM	energy minimisation
E_{pot}	potential energy
ER	endoplasmic reticulum
ESI	electrospray ionisation

ETC	electron transport chain
FI_{max}	maximum fluorescence intensity
F_{max}	maximum force
FRET	fluorescence resonance energy transfer
GABA	gamma-aminobutyric acid
GE_{model}	Glide model energy
GPS-SNO	group-based prediction system-*S*-nitrosothiol
GPS-YNO_2	group-based prediction system-tyrosine nitration
GROMACS	Groningen Machine for Chemical Simulations
G_{score}	Glide score
GS-DNB	1-(*S*-glutathionyl)-2,4-dinitrobenzene
GSH	glutathione (reduced)
G-site	glutathione binding site
GSNO	*S*-nitrosoglutathione
GSNOR	*S*-nitrosoglutathione reductase
GSSG	glutathione disulfide (oxidised glutathione)
GST	glutathione transferase
GUI	graphical user interface
HCV	hepatitis C virus
HDX	hydrogen-deuterium exchange
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl] ethanesulfonic acid
hGSTA1-1	human glutathione transferase alpha 1-1
hGSTO1-1	human glutathione transferase omega 1-1
hGSTP1-1	human glutathione transferase pi 1-1
HILIC	hydrophilic interaction chromatography
HOD	heats of dilution
H-site	hydrophobic binding site
IAA-94	indanyloxyacetic acid-94
IAM	iodoacetamide
IDA	iminodiacetic acid
IDOTP	isotope dot product
IFMD	induced-fit molecular docking
IMAC	immobilised metal affinity chromatography

iNOS	inducible nitric oxide synthase
Int.	intensity
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
K_d	dissociation constant
LC-MS/MS	liquid chromatography-mass spectrometry/mass spectrometry
L-site	ligandin binding site
m/z	mass-to-charge ratio
MAPEG	membrane-associated proteins in eicosanoid and glutathione metabolism
MBP	maltose binding protein
MD	molecular dynamics
MES	2-morpholin-4-ylethanesulfonic acid
MM	molecular mechanics
MR	molecular replacement
MSC	mesenchymal stem cell
MUP	major urinary protein
MWCO	molecular weight cut-off
MX	macromolecular crystallography
n	stoichiometry
N2RO	nitrous oxide reductase
NADH	nicotinamide adenine dinucleotide
NEM	N-ethylmaleimide
nH	number of protons released or taken up upon ligand binding
n_{HP}	Hill coefficient
NiR	nitrite reductase
NLS	nuclear localising signal
NMDA	N-methyl-D-aspartate
NMR	nuclear magnetic resonance
nNOS	neuronal nitric oxide synthase
NOS	nitric oxidase synthase
NOX	nicotinamide adenine dinucleotide phosphate oxidase
NPT	number of particles, pressure and temperature

NTA	nitrilotriacetic acid
NVT	number of particles, volume and temperature
OD_{600}	optical density at 600 nm
OPLS-AA	all atom-optimized potentials for liquid simulations
p53	tumour protein 53
p64	parchorin 64
PDB	Protein Data Bank
PEG	polyethylene glycol
PFT	pore-forming toxin
p*I*	isoelectric point
PIPES	1,4-piperazinediethanesulfonic acid
pK_a	acid dissociation constant
PMF	peptide mass fingerprinting
PMSF	phenylmethylsulfonyl fluoride
PTM	post-translational modification
QM	quantum mechanics
q_p	heat energy
R_g	radius of gyration
RMSD	root-mean-square standard deviation
RMSF	root-mean-square fluctuation
RNA	ribonucleic acid
RNAP	ribonuclease polymerase-associated protein
RNS	reactive nitrogen species
RONS	reactive oxygen species and reactive nitrogen species
ROS	reactive oxygen species
rpm	revolution per a minute
rRaHBP2	TIE2 angiopoietin receptor and histamine binding protein
R_s	resolution of separation
RSCB	Research Collaboratory for Structural Bioinformatics
RSLC	rapid system liquid chromatography
RyR	ryanodine receptor
S	tailing factor
SAXS	small angle X-ray scattering

SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SE-HPLC	size exclusion-high pressure liquid chromatography
sGC	guanylate cyclase
SjGST	*Schistosoma japonica* glutathione transferase
SNAP	*S*-nitroso-N-acetylpenicillamine
SOC	super optimal broth with catabolite repression
SOD	superoxide dismutase
SspA	stringent starvation protein A
TCEP	tris (2-carboxyethyl) phosphine-hydrogen chlroide
TEMED	tetramethylethylenediamine
TGF-β	transforming growth factor beta
TMD	transmembrane domain
TNF-α	tumour necrosis factor-α
T_{unfol}	thermal unfolding temperature
UV CD	ultraviolet circular dichroism
UV-Vis	ultraviolet-visible
Y.pestis	*Yersinia pestis*
YpSspA	*Yersinia pestis* stringent starvation protein A
YT	yeast extract and tryptone
ε	molar extinction coefficient
θ	ellipticity
θ_{fold}	ellipticity of folded protein
θ_{unfold}	ellipticity of unfolded protein
λ	wavelength
λ_{max}	maximum emission wavelength

Amino acids are named according to the IUPAC single and three letter codes.

List of Figures

List of Tables

Table of Contents

Chapter 1: Introduction

1.1 Membrane proteins

The functional properties of biological membranes are attributed predominately to the proteins associated with the lipid bilayer. In view of the central role membrane proteins contribute to numerous physiological functions, these proteins constitute the focus for ~60% of all approved drug targets. Therefore, ascertaining the three-dimensional conformation of membrane proteins will assist in further advancing structure-based drug design (Yildirim *et al*., 2007).

1.1.1 Structure and classification

The biogenesis of membrane proteins typically involves their synthesis on ribosomes which are bound to the endoplasmic reticulum (ER). During synthesis, a protein will be co-translationally inserted into the ER membrane and subsequently, through a series of vesicular transport events, it will be targeted to an appropriate membrane compartment. Importantly, throughout this process a membrane protein will not exist as a free molecule in solution, but instead it will remain associated with the lipid bilayer during synthesis. However, no protein is located solely within the hydrophobic core of a membrane and therefore the ability of membrane proteins to interact with or span the lipid bilayer, while still having some exposure to the aqueous environment, requires them to be amphiphilic (Alberts *et al*., 2008).

The interplay between lipid molecules and membrane proteins is imperative in maintaining the structure, stability and function of these proteins. Therefore, in view of the different ways that membrane proteins associate with the lipid bilayer, they can be classified as being either peripheral or integral membrane proteins. That is, peripheral membrane proteins associate reversibly with the lipid bilayer by forming electrostatic interactions with the membrane surface and/or by forming electrostatic and hydrophobic interactions with an integral membrane protein. In contrast, integral membrane proteins have strong, non-covalent membrane interactions and may either be monotopic or polytopic depending on whether they remain partially embedded or span the lipid bilayer, respectively (Alberts *et al*., 2008)

In contrast to monotopic integral proteins which depend upon a selected lipid binding domain, lipid covalent anchor and/or an amphiphilic α-helix in order to remain associated with the lipid bilayer, polytopic integral proteins contain a transmembrane domain (TMD) that will be able

to span the lipid bilayer and as a result they provide a link between the extracellular and intracellular environment of a cell. Therefore, considering the impermeability of biological membranes, polyprotic integral membrane proteins such as transporter proteins have ensured that polar solutes can move across the lipid bilayer (Alberts *et al*., 2008).

In particular, the movement of polar solutes across the lipid bilayer by transporter proteins can be seen to be facilitated by either (i) ion channels or (ii) carrier proteins, the latter being able to selectively bind and transport a polar molecule across the membrane after having undergone a series of conformational changes. In contrast, ion channel proteins, which may be voltage- or ligand-gated, rely on an aqueous pore that is lined with charged residues in order to ensure the transport of ions such as sodium (Na^+), potassium (K^+), calcium (Ca^{2+}) and chloride (Cl^-) ions across the lipid bilayer (Figure 1) (Alberts *et al*., 2008). However, considering the ubiquity of Cl^- ions within biological systems and the association of Cl^- channels with various pathological states, understanding the structure and function of Cl^- channels has come under immense focus within drug discovery (Alberts *et al*., 2008; Verkman and Galietta, 2009).

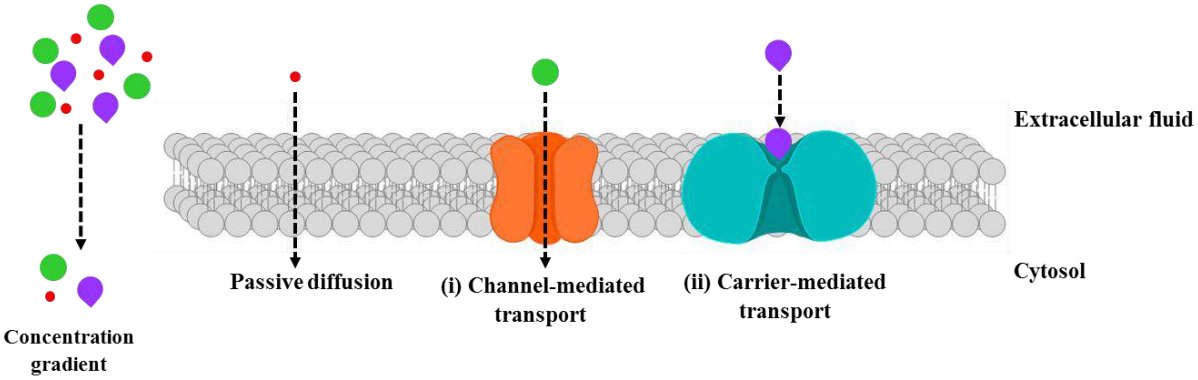


Figure 1. The transport of polar molecules across the lipid bilayer. Considering the impermeability of biological membranes to polar molecules, transporter proteins are required to ensure the effective movement of these molecules across the lipid bilayer. Transporter proteins, a type of polyprotic integral membrane protein, can be seen to comprise (i) ion channels or (ii) carrier proteins, the latter being able to selectively bind and transport a polar molecule across the membrane after having undergone a series of conformational changes. In contrast, ion channel proteins, which may be voltage- or ligand-gated, rely on an aqueous pore that is lined with charged residues in order to ensure the transport of ions across the lipid bilayer. The figure was adapted from Alberts *et al*., 2008 and generated by using ChemBioOffice®Ultra 14.0.

1.1.2 Cl^- channels- a transporter protein for the most abundant anion in a cell

Anion channel proteins such as Cl^- channels aid in the passive diffusion of negatively charged ions across the lipid bilayer by allowing these ions to move along their electrochemical gradient. In particular, although Cl^- channels have the potential to transport additional anions such bromide (Br^-) and iodide (I^-) ions across the lipid bilayer, they are often termed Cl^- channels owing to the fact that under most circumstances Cl^- are the most ubiquitous and predominant permeating anionic within a cell. Nevertheless, Cl^- channels are found to be localised to the plasma and intracellular membrane systems of a cell where they can be involved in regulating the pH and volume of a cell, the migration and differentiation of a cell as well as apoptotic pathway of a cell (Verkman and Galietta, 2009).

Currently, there are more than twenty members of Cl^- channels, and based on sequence homology, structure and functionality, these proteins have been classified into numerous protein families. Though in particular, ligand-gated receptor ion channel proteins such as the gamma-aminobutyric acid (GABA) receptors, the cystic fibrosis transmembrane conductance regulator (CFTR) proteins and the CLC proteins represent the Cl^- channel proteins which have been the best characterised to date. However, the identification of the chloride intracellular channel (CLIC) proteins has aided the characterisation of a new class of Cl^- channels at the molecular level (Al-Awqati, 1995). Consequently, the structural and functional characterisation of CLIC proteins would not only provide insight into the physiological functions that these proteins regulate, but it would also provide insight into additional and possible novel Cl^- channels that can be used as therapeutic targets.

1.2 The structural and functional properties of CLIC proteins

1.2.1 CLIC proteins share structural homology with the cytosolic GST superfamily.

The CLIC protein family, unlike most channel proteins, display enigmatic properties seeing that they are metamorphic and have the potential to reversibly transit between a soluble and integral membrane-bound conformation (Goodchild *et al*., 2011a; Tulk *et al*., 1998). This structural transition of CLIC1 can be similarly noted by (i) pore-forming toxins (PFTs) such colicins and diphtheria toxin (Parker and Feil, 2005) and (ii) apoptotic proteins such B-cell lymphoma 2 associated X protein (Bax), B-cell lymphoma 2 (Bcl-2) and B-cell lymphoma 2-extra-large (Bcl-xL) (Youle and Strasser, 2008). Furthermore, in contrast to typical membrane proteins, CLIC proteins lack a leader sequence required for membrane insertion and

instead they spontaneously auto-insert into a lipid bilayer to form a functional ion channel by undergoing an environmentally-induced conformational change (Tulk *et al*., 2002).

CLIC proteins have been identified both within vertebrate and invertebrate species (Littler *et al*., 2010), with the first identified CLIC protein being p64 from the microsomes of bovine kidney cells (Redhead *et al*., 1997). Collectively, with a pairwise sequence identity of 47-76% being shared among members, the CLIC family share a conserved sequence within vertebrate species (Cromer *et al*., 2002) and to date seven isoforms have been identified, namely (i) CLIC1 (nuclear Cl^- channel (NCC27)) (Valenzuela *et al*., 1997), (ii) CLIC2 (Heiss and Poustka, 1997), (iii) CLIC3 (Qian *et al*., 1999), (iv) CLIC4 (Duncan *et al*., 1997), (v) CLIC5 (Berryman and Bretscher, 2000) (vi) parchorin 64 (p64) (CLIC5B) (Redhead *et al*., 1997) and (vii) parchorin (Nishizawa *et al*., 2000).

Despite the sequence similarity among the members of the CLIC family, the sequence data of CLIC proteins additionally revealed that they share sequence homology with the glutathione (GSH) transferase (GST) superfamily (Dulhunty *et al*., 2001). Structurally, this similarity can be noted by the soluble conformation of CLIC proteins comprising a 240 amino acid core which in fact conforms to the canonical topology of the cytosolic GST superfamily by comprising both a conserved thioredoxin-like N-terminal domain (βαβαββα motif) and an all α-helical C-terminal domain (Figure 2) (Harrop *et al*., 2001). However, in contrast to the other CLIC protein members, p64 and parchorin are larger variants as they contain an additional N-terminal domain with repetitive sequences that are not evolutionary conserved (Dulhunty *et al*., 2001; Harrop *et al*., 2001; Littler *et al*., 2010). Interestingly though, unlike the members of the CLIC family, the members of the cytosolic GST superfamily cannot auto-insert into the lipid bilayer. However, microsomal GSTs, or otherwise membrane-associated proteins in eicosanoid and GSH metabolism (MAPEG), are the only GST superfamily to function as an integral membrane protein (Oakley, 2011).

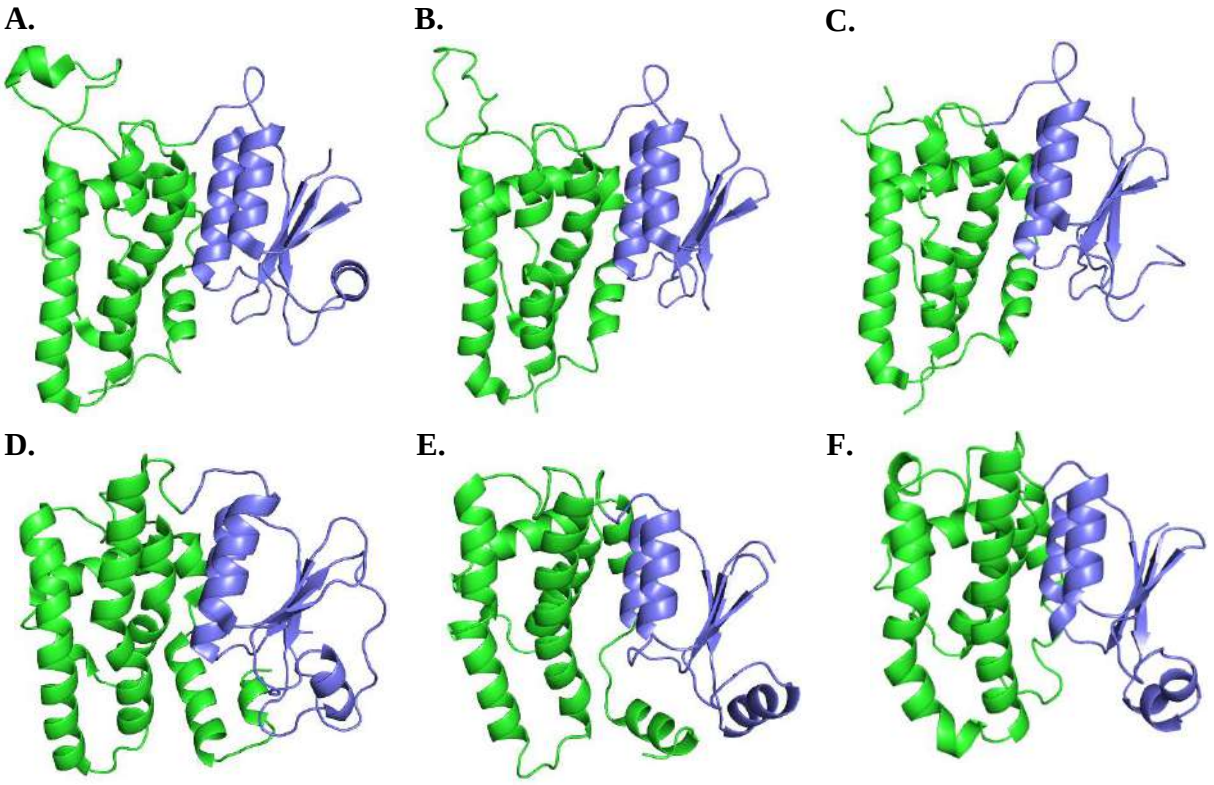


Figure 2. Structural homology between the members of the soluble CLIC and cytosolic GST protein family. Structurally, the soluble conformation of CLIC proteins conform to the canonical topology of the cytosolic GST superfamily by comprising both a conserved thioredoxin-like N-terminal domain (βαβαββα motif) (●) and an all α-helical C-terminal domain (●). Indicated are the crystal structures of (A) CLIC1 (PDB: 1K0M), (B) CLIC2 (PDB: 2PER), (C) CLIC4 (PDB: 2D2Z) as well as the monomeric conformations of (D) hGSTO1-1 (PDB: 1EEM), (E) GSTA1-1 (PDB: 1GSF) and (F) GSTP1-1 (PDB: 2GSS). All images were generated by using Pymol V1.3.

1.2.2 Physiological significance of CLIC proteins

CLIC proteins have a broad tissue distribution and have therefore been linked to a variety of essential physiological processes (Littler *et al*., 2010) such as the cell cycle (Valenzuela *et al*., 2000), apoptosis (Fernandez-Salas *et al*., 1999), cell motility (Ronnov-Jessen *et a*l., 2002) and bone resorption (Schlesinger *et al*., 1997). In particular, CLIC2 has for instance been reported to modulate the release of calcium ions from ryanodine receptors (RyRs) in cardiac and skeletal muscles, therefore suggesting that CLIC2 is effective in preventing an excess release of calcium ions which are commonly noted in apoptotic processes and conditions such as ischaemia (Board *et al*., 2004). In fact, the modulation of RyRs has similarly been observed by human GSTO1-1 (hGSTO1-1) (Dulhunty *et al*., 2001), therefore bearing in mind the shared sequence homology between CLIC2 and hGSTO1-1, the RyRs-binding site within CLIC and

GST proteins may share structural similarities (Board *et al*., 2004). Consequently, considering that cytosolic GSTs have not been reported to form ion channels, it was initially thought that the function of CLIC proteins was to regulate the activity of ion channel proteins as opposed to forming ion channels themselves (Littler *et al*., 2010).

The potential of CLIC proteins to insert into membranes as well as interact and regulate other proteins can be clearly exemplified by the member CLIC1. Present within CLIC1 is a negatively-charged, flexible loop (Pro147-Gln164) region formed by a series of acidic residues which have been suggested to interact with the positively charged regions of proteins, or facilitate membrane insertion by becoming hydrophobic when protonated at the membrane surface (Cromer *et al*., 2002; Harrop *et al*., 2001) (Figure 3). In particular, protein-protein interactions have be demonstrated by CLIC1, together with CLIC5, being able to interact with F-actin, which consequently can be linked to the regulation in cell volume, cell motility, cell division, intravesicular fusion along with endocytosis and exocytosis that is required to maintain cellular homeostasis (Sasaki *et a*l., 2014). In fact, the scaffolding proteins which are required to maintain the structure and dynamics of the membrane-actin interface is prevented in mice which are deficient in CLIC5 and as a result these mice suffer progressive deafness and vestibular dysfunction given the degeneration of hair cell stereocilia (Gagnon *et al*., 2006).

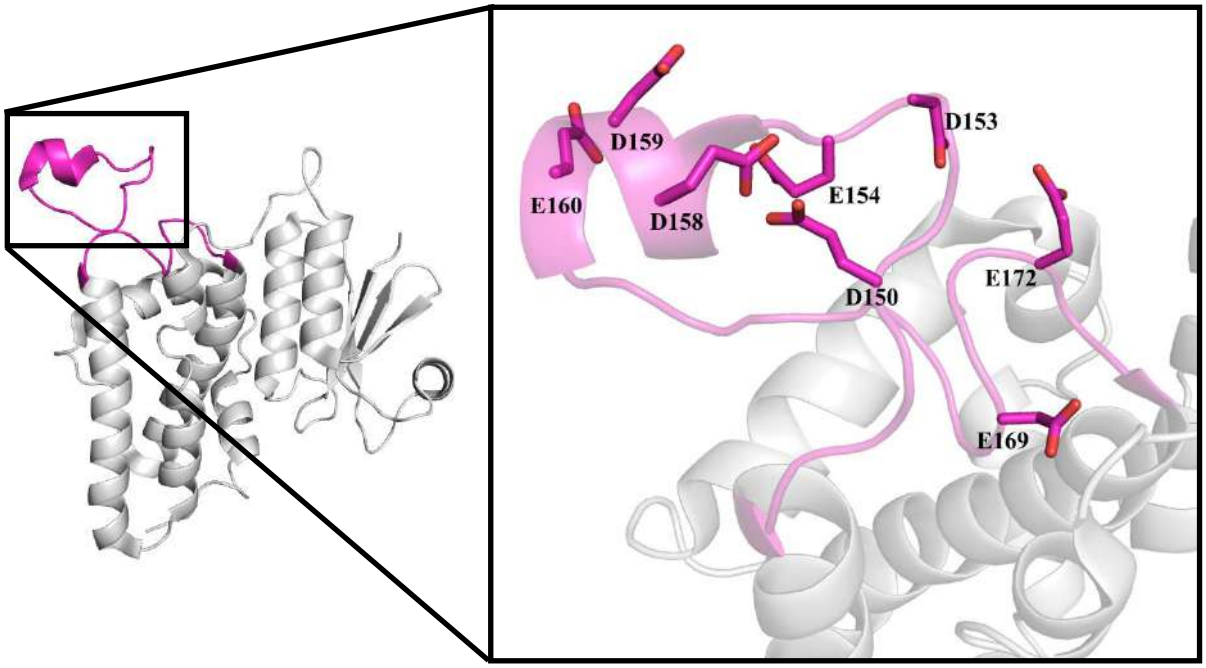


Figure 3. The negatively-charged and flexible loop region of CLIC1. Present at α-helix 7 within CLIC1 (PDB: 1K0M) is a negatively-charged and flexible loop region (●) formed by a series of acidic residues. This region has been suggested to be involved in protein-protein interactions as well as to possibly facilitate membrane insertion by becoming hydrophobic when protonated at the membrane surface. All images were generated by using Pymol V1.3.

Interestingly, the CLIC1 protein has also been shown to play a key role in the life cycle of the hepatitis C virus (HCV). During infection by HCV, it has been shown that indanyloxyacetic acid-94 (IAA-94), a well characterised Cl^- channel inhibitor, was able to inhibit the replication of the HCV genome (Figure 4), therefore suggesting that the replication of the HCV genome is dependent upon the function of Cl^- ion channels. In particular, it was demonstrated that the inhibition of Cl^- channels by IAA-94 did not inhibit virus entry or replication following the establishment of infection, instead IAA-94 was demonstrated to inhibit early post-entry virion trafficking and/or early replication events (Igloi *et al*., 2015).

During the life cycle of HCV, the virus is internalised into its target cell via a pH-dependent process following its attachment to specific receptors on the surface of hepatocytes. Once a HCV virus particle has been endocytosed, the resultant Cl^- channels within the membrane of the endosome will increase the influx of Cl^- and consequently result in endosomal acidification. The ability of Cl^- to regulate the pH of the endosome is extremely important considering that an acidic environment in these organelles is crucial for the induction of the HCV E1/E2 glycoprotein fusing with the membrane during its early post-entry events where the HCV genome will be released into its target cells. Importantly, upon investigating the molecular identity of the Cl^- channels required during the HCV life cycle it was revealed that there was an increase in the expression of CLIC1. Therefore, by inhibiting the Cl^- channel activity of CLIC1 in hepatocytes, this decreased the proportion of HCV ribonucleic acid (RNA) being released from the endosome and consequently the decreased synthesis of viral proteins prevented viral maturation and attenuated HCV spreading to neighbouring hepatocytes (Figure 4) (Igloi *et al*., 2015).

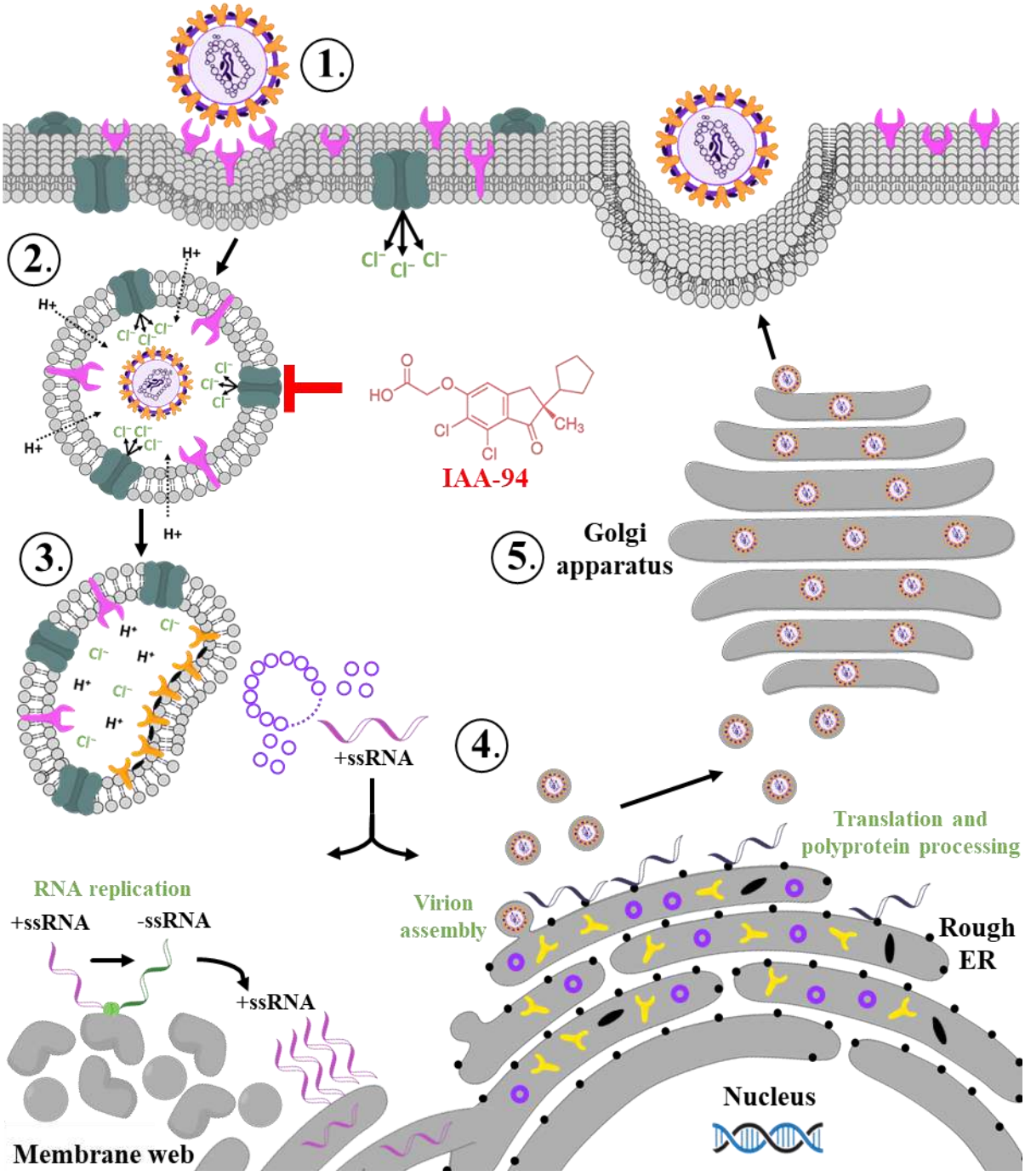


Figure 4. The HCV life cycle is regulated by the Cl^- channel activity of CLIC1. (Step 1) During the HCV life cycle the virus will first bind to target proteins (●) on the surface of hepatocytes and become endocytosed. (Step 2) CLIC1 (●) causes an influx of Cl^- into the endosome and results in an acidic environment due to the concomitant influx of protons down their electrochemical gradient into the endosome. (Steps 3) An acidic environment is crucial for the HCV E1/E2 glycoprotein to fuse with the endosome membrane and release the HCV genome into the cell. However, inhibiting the Cl^- channel activity of CLIC1 by IAA-94 (⊢) attenuates the release of the HCV genome from the endosome. (Step 4) Once released, the positive single stranded ribonucleic acid (ssRNA) will undergo (i) translation and (ii) RNA replication whereby an RNA-dependent RNA polymerase will produce a negative ssRNA strand. The latter serves as template for the synthesis of more positive ssRNA which will be used for the assembly of new HCV particles. (Steps 5) Once assembled, the virion will be transported to the Golgi apparatus and undergo virion maturation in order to be exocytosed. Therefore, by attenuating the release of the positive ssRNA, IAA-94 is be able to attenuate HCV spreading. The figure was generated by using ChemBioOffice®Ultra 14.0.

Despite CLIC1 being a well-characterised member of the CLIC protein family, there has also been a significant increase in the information surrounding the functional and structural characterisation of CLIC4. The intracellular localisation and subcellular distribution of CLIC4 varies among different cell lines and tissues, however, its cellular localisation often appears to be intimately linked to scaffold proteins from the cytoskeleton, chaperones, nuclear transporters and most notably the membrane of the ER and nucleus (Littler *et al*., 2005). Interestingly, CLIC4 is a direct downstream target gene for the tumour protein 53 (p53), a protein that leads to growth arrest, deoxyribonucleic acid (DNA) repair and apoptosis under conditions of stress such as DNA damage or oxidative stress (Williams and Schumacher, 2016). In fact, in cells such as mouse S1 keratinocytes and human osteosarcoma cells which are undergoing p53-mediated apoptosis due to cellular oxidative stress, CLIC4 is known to translocate to the nucleus the and subsequently increase the rate of apoptosis in these cells (Suh *et al*., 2004).

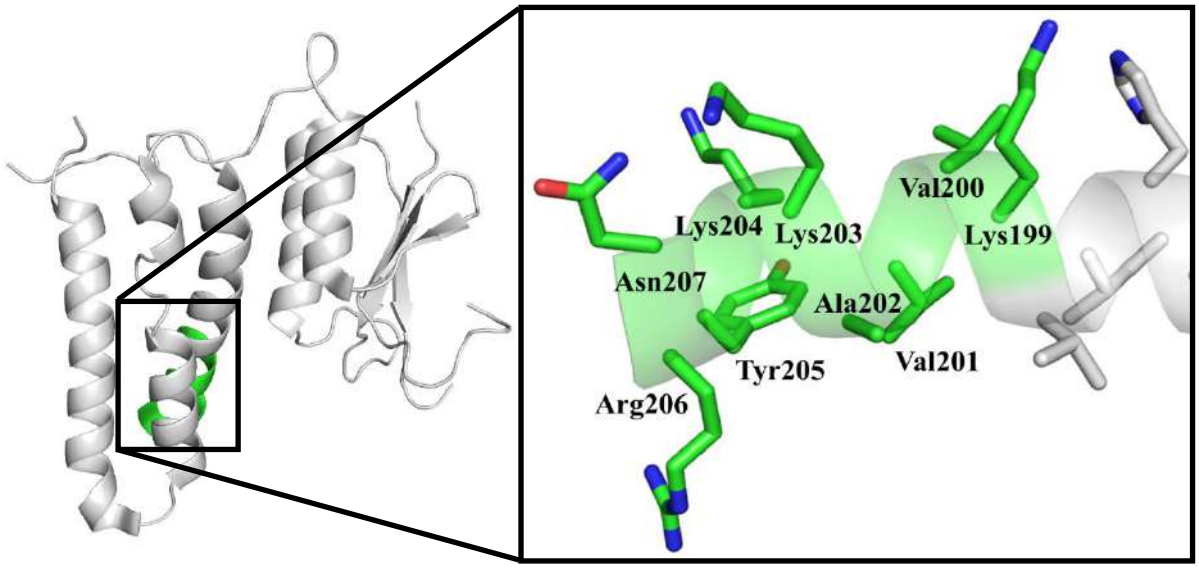
The nuclear translocation of CLIC4 under oxidative stress conditions is often accompanied by an increase in the transforming growth factor beta (TGF-β) signalling pathway which leads to growth arrest. Though in epithelial tumorigenic cells the expression levels of CLIC4 is often downregulated and consequently there is a diminished response to the TGF-β signalling pathway in these cancer cell lines. However, considering that TGF-β is still abundant within the microenvironment of these tumour cells, when CLIC4 was reconstituted to the nucleus through an adenoviral-mediated pathway, the TGF-β-dependent transcriptional activity and its downstream anti-tumour functions were restored. Therefore, CLIC4 plays a pivotal role in regulating the growth of tumorigenic cells (Suh *et al*., 2012).

Based on immunoprecipitation assays it has been reported that under oxidative stress conditions there is an interaction between CLIC4 and various members of the nuclear import machinery such as importin-α. Significantly, in the cell one of the best known nuclear trafficking systems is the importin-α: β nuclear transport pathway, whereby importin-α aids as a receptor that can recognise and bind directly to cargo proteins from the cytoplasm, while importin-β ensures the transport of this complex through the nuclear pore. However, this transport process is highly dependent upon importin-α firstly recognising a specific nuclear localising signal (NLS) on the cargo protein such as CLIC4 (Suh *et al*., 2004).

The NLS is a series of approximately six basic amino acid residues and based on sequence comparison, mutagenesis and crystallographic studies, a NLS site (^{199}KVVAKKYRN^{207}) has been identified within CLIC4 at the C-terminal domain (Suh *et al*., 2004) (Figure 5). In fact, this sequence is highly conserved across the CLIC family, but with respect to CLIC4, only the ^{202}AKKYRN^{207} residues appear to bind directly to importin-α. Typically, a NLS is located at the solvent-exposed sites of unstructured regions such as domain termini or flexible loops, which is important given that structural studies have concluded that a NLS binds to importin-α in an extended conformation and would therefore need to be unfolded and flexible within the cargo protein in order to be functional. However, the NLS in CLIC4 adopts an α-helical conformation that would otherwise preclude its binding to importin-α, but based on structural data it is clear that the NLS peptide from CLIC4 binds to importin-α in an extended conformation (Mynott *et al*., 2011) (Figure 5).

In fact, it has been observed that the *S*-nitrosylation of CLIC4 induces a conformational change that destabilises its native conformation and therefore triggers a structural change that is required to facilitate the interaction between its NLS and importin-α (Malik *et al*., 2010). Furthermore, Tyr205 can be seen to play an important role in the binding specificity of the NLS to importin-α (Figure 5), as a result the phosphorylation of Tyr205 could prevent the NLS from binding to importin-α. Therefore, the requirement for regulatory events to expose the NLS has offered insight into a new criteria required for the selection of cargo proteins by nuclear import receptors. In the case of CLIC4 where its nuclear translocation is linked to fundamental cellular events, it is apparent that substantial regulation in the folded state of the NLS is required in order to regulate its translocation across the nuclear membrane (Mynott *et al*., 2011).

A.



B.

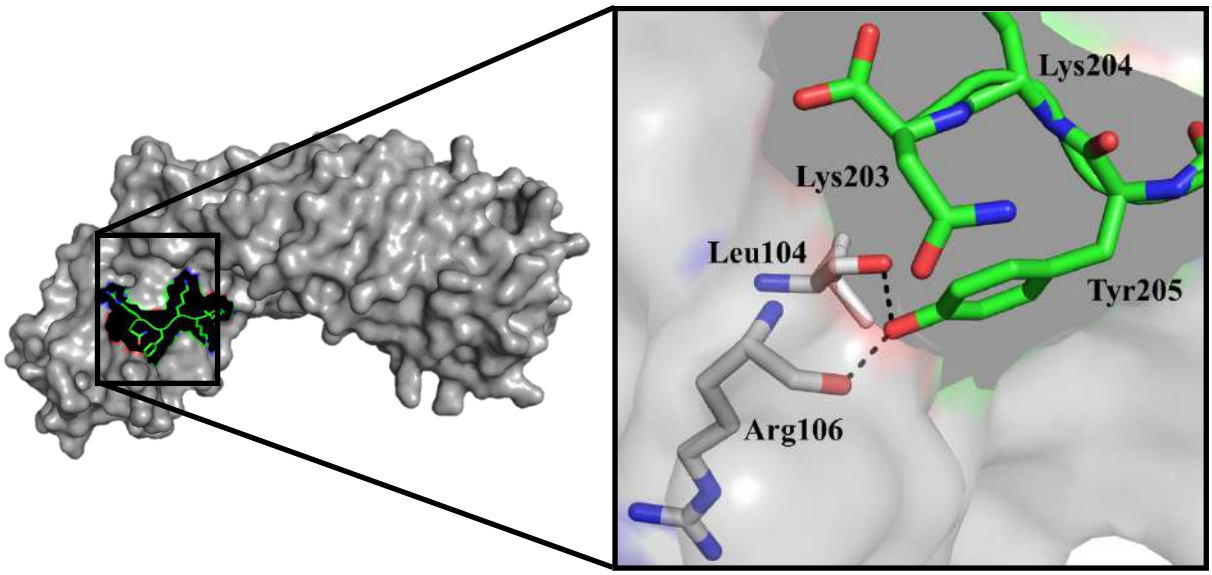


Figure 5. A NLS within CLIC4 is required for its interaction with importin-α. Based on immunoprecipitation assays it has been reported that under oxidative stress conditions there is an interaction between CLIC4 and importin-α from the importin-α: β nuclear transport pathway. However, this interaction is highly dependent upon importin-α recognising a specific NLS. (A) The NLS is a series of approximately six basic amino acid residues and based on sequence comparison, mutagenesis and crystallographic studies, a NLS site (^{199}KVVAKKYRN^{207}) (●) has been identified within CLIC4 at the C-terminal domain. (B) A NLS needs to be unfolded and flexible within the cargo protein in order to bind to importin-α in an extended conformation. However, the NLS within CLIC4 adopts an α-helical conformation that would otherwise preclude its binding to importin-α (●), but based on structural data it is clear that the peptide for the CLIC4 NLS binds to importin-α in an extended conformation. Furthermore, Tyr205 can be seen to play an important role in the binding specificity of the NLS to importin-α. All images were generated by using Pymol V1.3.

1.3 The structural and functional properties of the CLIC1 protein

The structure of CLIC1 was the first solved structure of the CLIC family and has therefore been the most extensively studied (Harrop *et al*., 2001; Valenzuela *et al*., 1997). Structurally, despite having a low sequence similarity (~15%) and varied orientations of selected α-helices (α4, α5 and α9), the soluble conformation of CLIC1 shares significant structural homology to hGSTO1-1 (Harrop *et al*., 2001) (Figure 6A). In particular, similar to the ^{32}CPY^{34} monothiol motif found within GST-Ω, native to the active site of CLIC1 is a ^{24}CPF^{26} monothiol motif which contains the catalytic cysteine residue required to bind and coordinate GSH when CLIC1 becomes glutathionylated by oxidised GSH (GSSG) (Figure 6B) (Board *et al*., 2000; Harrop *et al*., 2001). Collectively, these structural features support the glutaredoxin-like GSH-dependent oxidoreductase activity demonstrated by CLIC1 (Khamici *et al*., 2015).

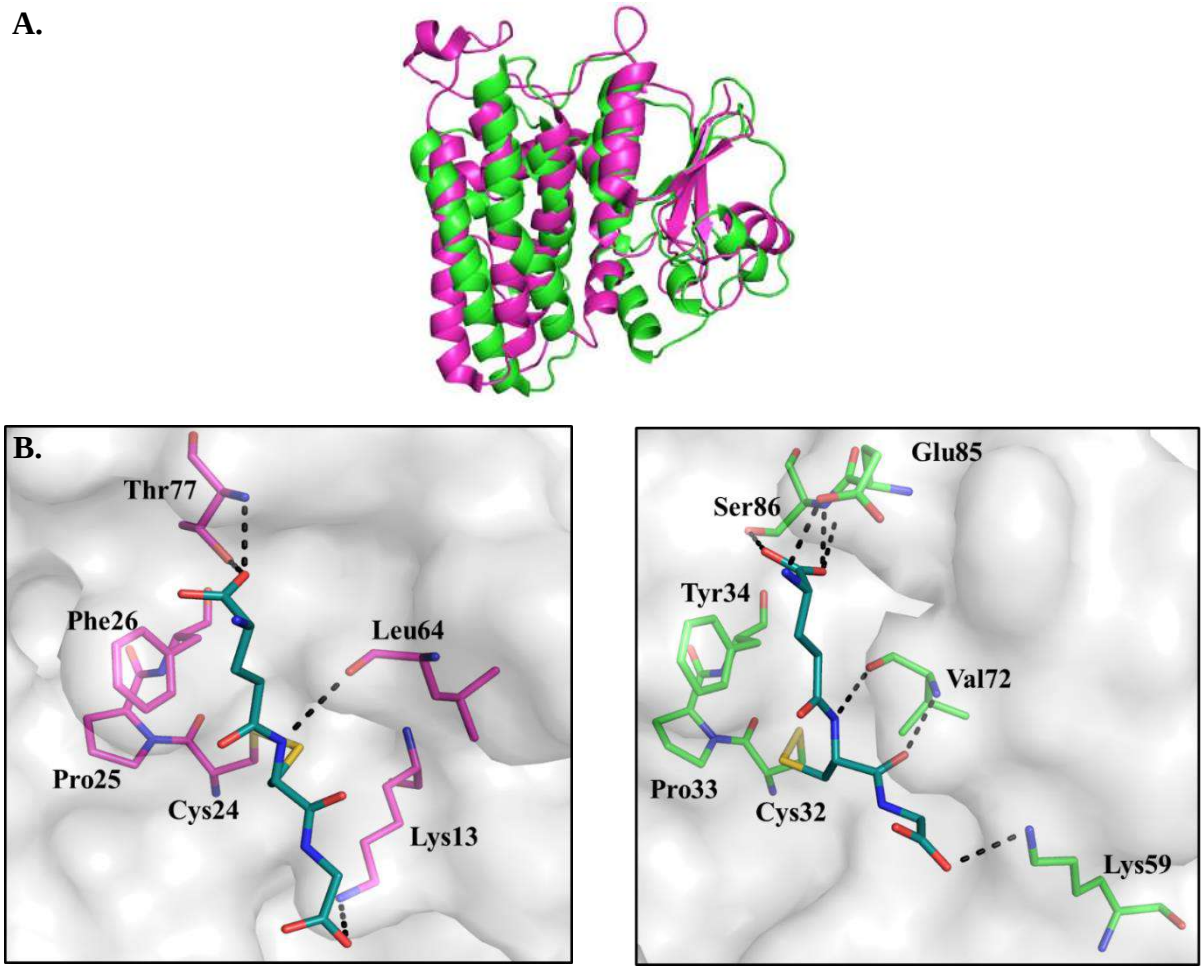


Figure 6. The structural homology between CLIC1 and hGSTO1-1. (A) The soluble conformation of CLIC1 (●) (PDB: 1K0M) is structurally similar to hGSTO1-1 (●) (PDB: 1EEM) (RMSD 4.342 Å) despite having a low sequence similarity (~15%) and varied orientations of selected α-helices. (B) Native to the active site of CLIC1 (●) and hGSTO1-1 (●) is a ^{24}Cys-Pro-Phe^{26} and ^{32}Cys-Pro-Tyr^{34} monothiol motif, respectively and both these motifs contain a catalytic cysteine residue required to bind and coordinate GSH (●). All images were generated by using Pymol V1.3.

However, although the physiological relevance together with several structural and functional features of CLIC1 has been characterised, the structure of the membrane-bound conformation along with the mechanism governing its membrane insertion still remains unclear. However, the metamorphic potential of CLIC1 is thought to accompany structural changes that are brought about by changes in its surrounding environment such as a change in pH (section 1.3.2) and the redox state of a cell (sections 1.3.4 and 1.3.5).

1.3.1 A single TMD governs the membrane insertion of CLIC1

A sequence comparison between the various members of the CLIC family have identified two regions of sufficient hydrophobicity that could account for a putative TMD (Cromer *et al*., 2002). The first region comprised α-helix 6 within the C-terminal domain (Figure 7A), whereas the second region comprised α-helix 1 and β-strand 2 within the N-terminal domain (Figure 7B) (Berryman and Bretscher, 2000). Though, considering the structural similarity of the N-terminal domain with the thioredoxin superfamily, it was unlikely that this domain would unfold in order to form a membrane-interacting domain, instead the similarity of the C-terminal domain to a number of pore-forming proteins proposed that the potential of CLIC1 to form a Cl^- channel resides within this domain (Cromer *et al*., 2002).

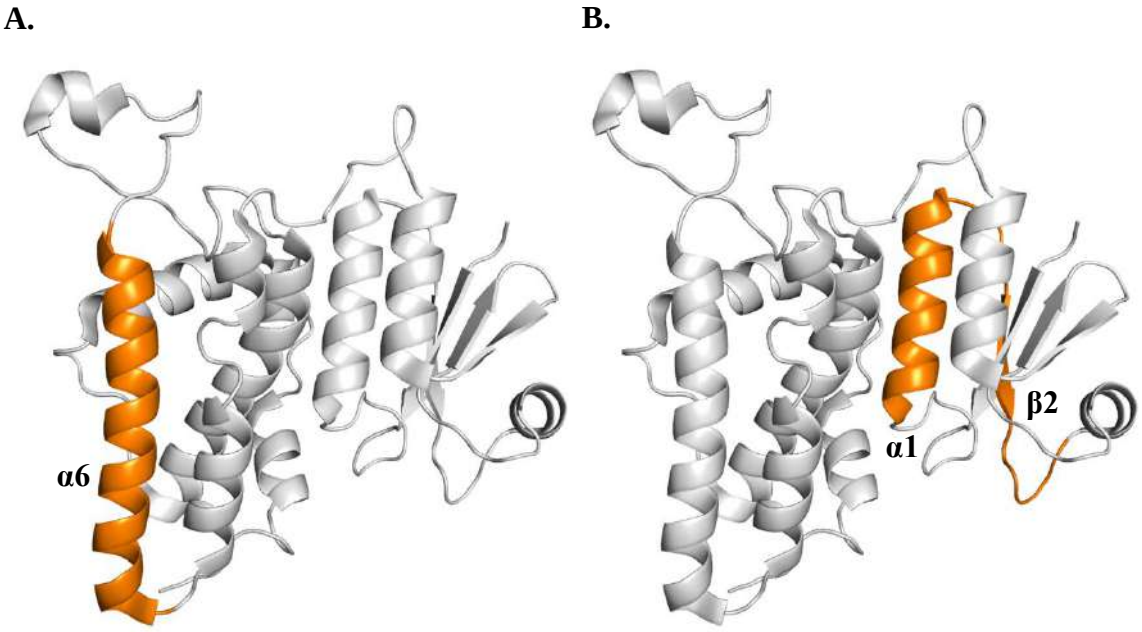


Figure 7. The putative regions for the TMD within CLIC1. Analysis of the hydrophobicity plots of soluble CLIC1 (PDB: 1K0M) indicated two regions of significant hydrophobicity that could constitute a TMD (●). (A) The first putative TMD region was predicted to be α-helix 6 within the C-terminal domain, whereas (B) the second putative TMD region was predicted to comprise both α-helix 1 and β-strand 2 within the N-terminal domain. However, various studies have indicated that the second putative TMD is instead required for the membrane insertion of CLIC1. All images were generated by using Pymol V1.3.

In fact, studies involving proteolytic digestion (Duncan *et al*., 1997), FLAG epitopes (Tonini *et al*., 2000) and terminal-directed antibodies (Proutski *et al*., 2002) against the membrane-bound conformation of CLIC1 revealed that the residues present within the second putative TMD remained buried within the lipid bilayer, while the C-terminal domain remained exposed to the cytoplasmic environment. Therefore, seeing that α-helix 1 and β-strand 2 constitute the TMD in soluble CLIC1, these secondary structural elements would need to undergo significant structural rearrangements in order to form a TMD with sufficient helical propensity and length to span the lipid bilayer.

Although CLIC1 lacks a membrane-targeting signal peptide, flanking the TMD of CLIC1 is a ^{49}KRR^{51} motif that would be involved in electrostatic interactions with the polar head groups at the membrane surface, therefore enabling the TMD to bind and subsequently insert into the membrane. In serving as an electrostatic plug that will anchor the TMD within the membrane, the ^{49}KRR^{51} motif functions to concentrate and subsequently facilitate the oligomerisation of CLIC1 within the lipid bilayer. Importantly, the ^{49}KRR^{51} motif has also been shown to be involved in orientating the TMD with respect to the *cis*- and *trans*-face of the membrane (Peter *et al*., 2014). Consequently, the orientation of the TMD for instance will preferentially result in Cys24 being exposed to the *trans*-face of the lipid bilayer where an oxidising environment persists (Peter *et al*., 2014; Singh and Ashley, 2006) (Figure 9).

Once inserted into the membrane, the single TMD of CLIC1 would need to oligomerise in order to form a pore of a functional ion channel. Initially, the self-association of the TMDs has been suggested to be thermodynamically driven by electrostatic and hydrophobic interactions which stabilise a dimeric species within the lipid bilayer environment (Peter *et al*., 2013). It has been suggested by Warton *et al*., 2002 that this dimeric species could represent a weakly active protopore, but due to the relatively weak hydrophobic forces that maintain this interaction, the dissociation of the dimer would be frequent and subsequently this would ensure the oligomerisation of the TMD into a fully active Cl^- channel. However, the oligomeric state of CLIC1 is rather controversial seeing that a protopore with either (i) four (Singh, 2010) or (ii) six to eight (Goodchild *et al*., 2011b) TMDs has been described (Figure 9). Therefore, further studies are needed to be performed in order to properly understand the true oligomeric state for the membrane-bound conformation of CLIC1.

Nevertheless, although being predominately hydrophobic, the TMD sequence of CLIC1 also contains two charged residues, namely Arg29 and Lys37. Therefore, considering that the TMD region is of sufficient length to span the lipid bilayer and that the TMD peptide alone is sufficient for Cl^- transport, it has been proposed that Arg29 will interact with the polar head groups at the membrane surface. However, Lys37 is located at the centre of the TMD and therefore it would need to interact with surrounding residues in order to overcome the energetic cost of being partitioned into the lipid bilayer. In fact, it has been reported by Peter *et al*., 2013 that this instability can be overcome by Lys37 and Trp35 forming a cation-π stacking interaction. In this interaction the positive charge of Lys37 will be neutralised and subsequently this will drive the formation of active dimer intermediates during the membrane insertion of CLIC1. However, during the formation of the active Cl^- channel, the cation-π interaction would need to be broken in order to allow these positively-charged residues to localise within the pore of the ion channel considering that Arg29 and Lys37 have been linked to the open probability and conductance of the ion channel, respectively (Averaimo *et al*., 2013).

The importance of Arg29 and Lys37 stems not only from their role in regulating the gating and conductance of Cl^- within the ion channel of CLIC1, but also from their ability to control how ligands such as metformin and IAA-94 interact with CLIC1 to subsequently inhibit its ion channel activity. Based on mutagenesis studies, a working model has been developed to suggest as to how metformin inhibits the Cl^- channel activity of CLIC1. It is likely that Arg29 is near the anion path at the extracellular surface (Figure 8 and 9), however, when the ion channel is in its open state, Arg29 is suggested to interact with a polar portion of the channel by means of its guanidinium group. Yet, due to the double guanidinium group on metformin, the side chain of Arg29 could become displaced from this polar site to possibly obstruct the pore of the ion channel. Yet, in the R29A mutant of CLIC1 where there is no side chain to make this destabilising interaction, metformin would be free to bind to the polar site and as reported, it would have no influence on the channel opening at high membrane potentials. These findings were further supported through inside-out electrophysiology experiments whereby metformin was unable to inhibit the single channel current of CLIC1 (Gritti *et al*., 2014). Perhaps this could be attributed to the ^{49}KRR^{51} motif which would bind to metformin, therefore preventing its translocation to Arg29 in order to inhibit the ion channel activity of CLIC1 (Figure 8).

However, with respect to IAA-94, it was shown that the dose response curves for both the wild-type and R29A mutant of CLIC1 were the same, indicating that unlike metformin, Arg29 did not have an influence on the binding of IAA-94 in the open state. On the other hand, when wild-type CLIC1 was incubated with metformin the membrane current was inhibited and no further decrease was noted upon the addition of IAA-94. However, metformin only partially blocked the membrane current in the R29A mutant, but the addition of IAA-94 resulted in full inhibition (Averaimo *et al*., 2013; Gritti *et al*., 2014). Therefore, it appears that Lys37 could be a target site for IAA-94 given that its positively charged amino group will be able to coordinate the acetate moiety of IAA-94. However, without a structure of CLIC1 in its membrane-bound conformation, this mode of inhibition still remains largely uncertain (Figure 8).

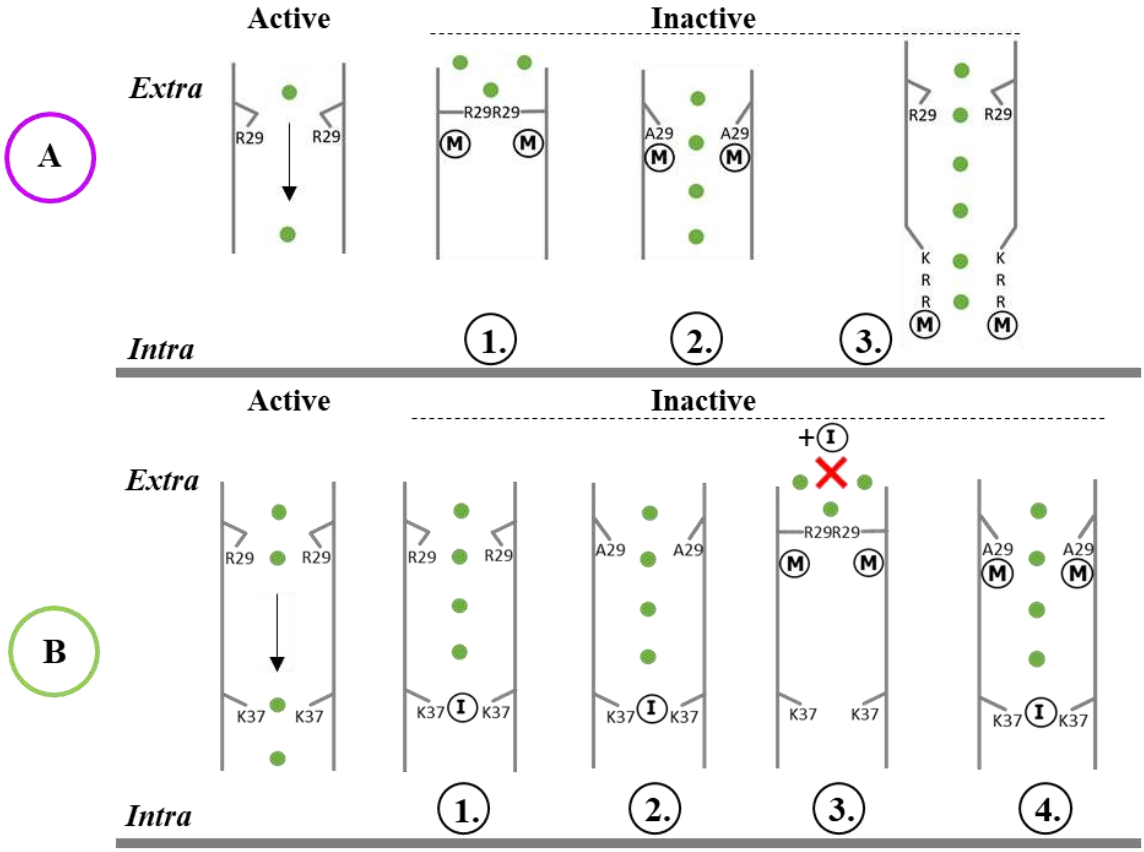


Figure 8. The inhibition of the Cl^- channel activity of CLIC1 by metformin and IAA-94. (A) In the open state of the ion channel it is suggested that Arg29 interacts with a polar site by means of its guanidinium group. (A1) However, the double guanidinium group of metformin (Ⓜ) will displace the side chain of Arg29 to block the influx of chloride ions. (A2) Metformin can therefore bind freely to this polar site in the R29A mutant and because there is no steric hindrance from the side chain of Arg29, there is an influx of chloride ions. (A4) This interaction is specific given that metformin could not inhibit the ion channel activity during inside-out electrophysiology studies. (B) The ion channel activity for both the (B1) wild-type and (B2) R29A mutant of CLIC1 can be inhibited by IAA-94 (Ⓘ), therefore confirming that Arg29 does not have an influence of the binding of IAA-94 in the open state. (B3) When wild-type CLIC1 is incubated with metformin, IAA-94 does not further inhibit the ion channel activity. (B4) Instead, IAA-94 results in complete inhibition in the R29A mutant of CLIC1, suggesting that Lys37 is involved in the binding of IAA-94.

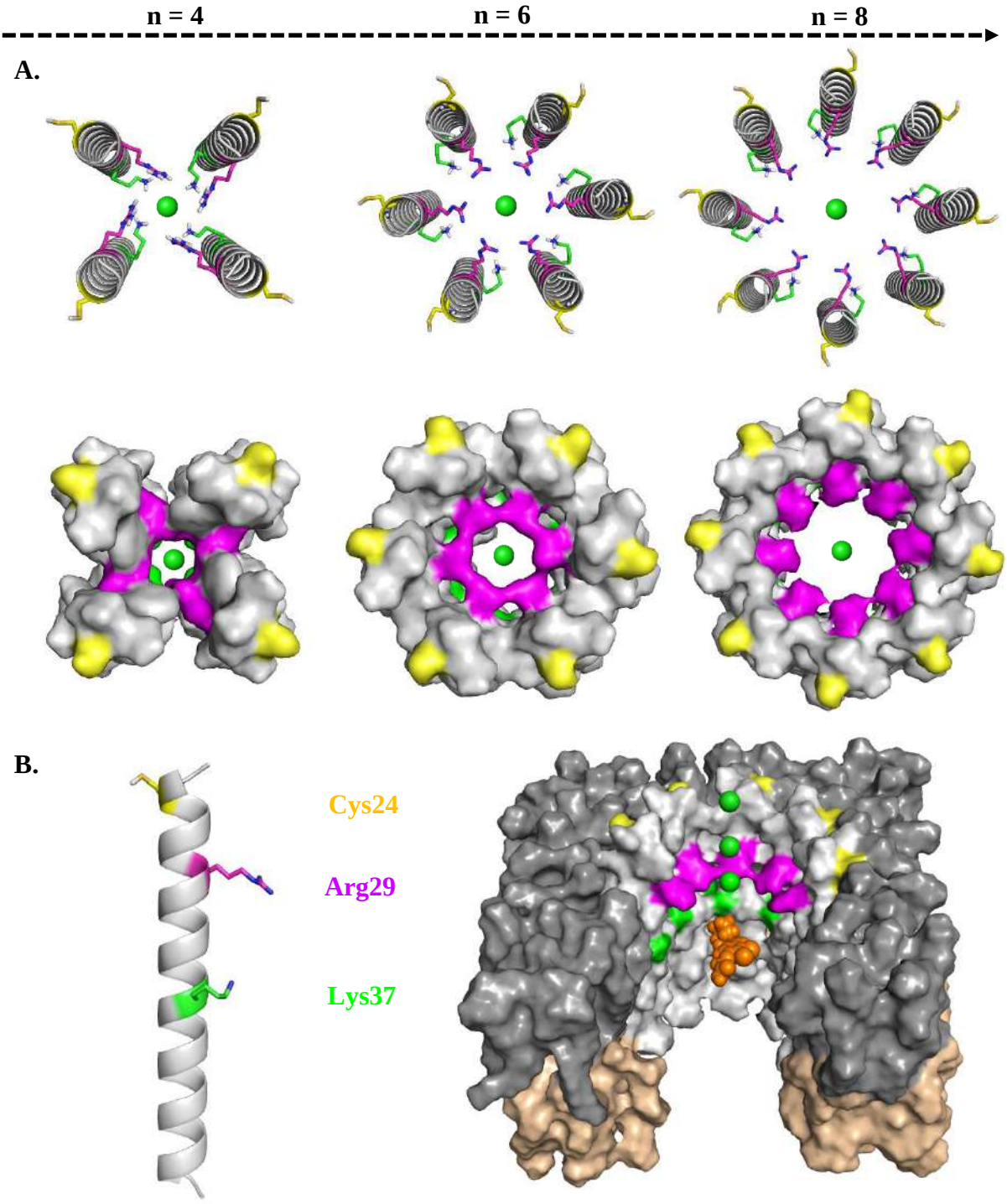


Figure 9. Modelling the putative membrane-bound conformation CLIC1. The TMD is sufficient to allow for the conductance of Cl^- across a membrane, however, the dimeric species of the TMD is not sufficient to represent an active ion channel and the single TMD of CLIC1 would therefore need to oligomerise in order to form a functional pore. (A) The TMD of CLIC1 was modelled with a stoichiometry of n = 4, n = 6 and n = 8, whereby Cys24 (●), Arg29 (●) and Lys37 (●) have been indicated. (B) Considering the sensitivity of Cys24 to the redox conditions of the cytosol, it is located on the outside of the lipid bilayer. However, Arg29 and Lys37 are located in the core of the ion channel so that their positive charge will coordinate the transport of incoming anions such as Cl^- (●), or possibly inhibitors such as IAA-94 (●). All images were generated by using Pymol V1.3.

1.3.2 The influence of pH on the membrane insertion potential of CLIC1

The appearance and activity of the CLIC1 ion channel is strongly dependent on the pH of the surrounding environment considering that its Cl^- channel activity was reported to be the lowest at pH 7.0, but with a decrease in pH its activity began to increase and was noted to be significantly enhanced at pH 5.0. (Tulk *et al*., 2002 and Warton *et al*., 2002). Therefore, considering an acidic environment (pH ~5.5) is prevalent at the membrane surface due to the attraction of protons to the negatively charged polar head groups in the lipid bilayer (McLaughlin, 1989), the enhanced ion channel activity of CLIC1 that was demonstrated under these conditions coincided with CLIC1 being in its functional membrane-bound conformation.

Collectively, fluorescence resonance energy transfer (FRET) (Goodchild *et al*., 2010; Goodchild *et al*., 2011b) along with hydrogen-deuterium exchange (HDX) studies (Stoychev *et al*., 2009) have provided insight into the dynamic nature of CLIC1 when present at the membrane interface. In particular, the N-terminal domain was noted to be the least stable region within CLIC1 and when present in an acidic environment, although remaining intact structurally, its stability was further reduced to prompt a large conformational change that would facilitate its separation from the C-terminal domain (Stoychev *et al*., 2009). Similar to PFTs, this apparent increase in flexibility of CLIC1 has been suggested to lower the activation energy barrier for the structural rearrangements required for membrane insertion (Fanucchi *et al*., 2008; Stoychev *et al*., 2009).

The increased flexibility of CLIC1 when at the membrane surface can be attributed to a collection of pH-sensitive residues at its domain interface (Fanucchi *et al*., 2008). These residues, based on their theoretical side chain acid dissociation constant (pK_a) values, would become protonated in an acidic environment to result in various hydrogen bonds or salt bridges being broken and consequently this would reduce the stability of CLIC1 (Fanucchi *et al*., 2008; Legg-E'silva *et al*., 2012; Stoychev *et al*., 2009). In particular, present within CLIC1 are three histidine residues, however, upon their protonation only His74 and His185 have been reported to influence the conformational stability of CLIC1 by stabilising the intermediate state formed during the unfolding pathway of CLIC1 (Achilonu *et al*., 2012). Furthermore, present within CLIC1 are a total of twenty-one glutamate residues, however, of particular interest is Glu81, Glu85 and Glu228 given their involvement in the formation of hydrogen bonds and salt bridges at the domain interface of CLIC1. At the domain interface Glu81 and Glu228 will form a hydrogen bond with Arg29 on the TMD and Ser16 on the N-terminal domain, respectively,

whereas Glu85 will form a salt bridge with Lys37 on the TMD along with a hydrogen bond with Leu96 on the C-terminal domain (Figure 10). Collectively, these bonds therefore maintain the integrity of CLIC1 and prime the N-terminal domain at a low pH for membrane insertion (Cross *et al*., 2015; Legg-E'silva *et al*., 2012; Stoychev *et al*., 2009).

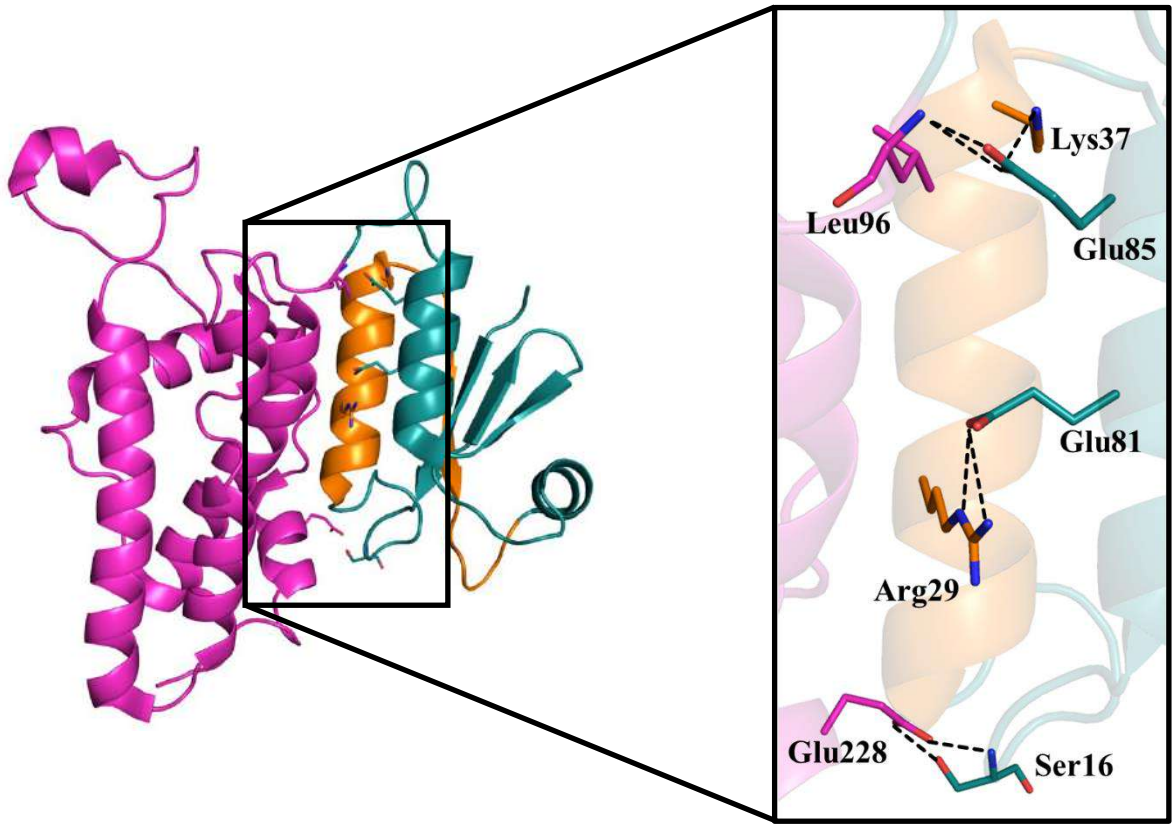


Figure 10. The pH-sensitive residues at the domain interface of CLIC1. At the domain interface of CLIC1 (PDB: 1K0M) are a collection of glutamate residues which are pH-sensitive. Therefore, when they are present at the membrane interface, they will become protonated to result in the breakage of hydrogen bonds and/or salt bridges (- - -). The residues Glu81, Glu85 and Glu228 are involved in the formation of hydrogen bonds and salt bridges at the domain interface. In particular, Glu81 and Glu228 will form a hydrogen bond with Arg29 on the TMD (●) and Ser16 on the N-terminal domain (●), respectively, whereas Glu85 will form a salt bridge with Lys37 on the TMD along with a hydrogen bond with Leu96 on the C-terminal domain (●). As a result, these interactions have proven to be essential in maintaining the integrity of CLIC1. All images were generated by using Pymol V1.3.

1.3.3 Post-translational modifications (PTMs) regulate the membrane insertion of CLIC1

Subsequent to the translation of messenger RNA (mRNA) into an amino acid sequence, the formation of a functional protein does not end at the ribosome. Instead, the completed polypeptide would need to fold correctly into its three-dimensional conformation, bind to required cofactors, assemble with other protein subunits and/or undergo post-translational modifications (PTMs) whereby various chemical moieties bind covalently to selected amino acid residues (Alberts *et al*., 2008).

PTMs are significantly prevalent in eukaryotes and include modifications such as phosphorylation, glycosylation, methylation, acetylation and *S*-nitrosylation. These modifications often change the electrostatic and/or structural properties of proteins and are therefore a source for protein diversity by further modulating and extending the range of possible protein functions. Collectively, the complexity initiated by these modifications provides the foundation for regulating cellular processes such as metabolism, signal transduction pathways and ion transport (Duan and Walther, 2014).

1.3.3.1 The production of oxidative chemical species and the redox regulation of proteins

Throughout the course of evolution, the controlled production of oxidative chemical species such as reactive oxygen species (ROS) and reactive nitrogen species (RNS), otherwise referred to as RONS, has ensured the expression of several transcription factors and signal transduction molecules by functioning as secondary messenger molecules in signal transduction pathways (Weidinger and Kozlov, 2015). As a result, the production of RONS has been linked to the regulation of cell adhesion, redox-mediated amplification of the immune response and apoptosis, however, the inability to produce antioxidants at sufficient levels can result in the overproduction of RONS. Therefore, considering that RONS are highly reactive free radicals, the overproduction of these chemical species can result in macromolecules becoming dysfunctional, together with multiple compromises in cellular homeostasis that cumulatively induce deleterious events implicated in carcinogenesis, neurodegeneration, atherosclerosis, diabetes and aging (Krumova and Cosa, 2016; Weidinger and Kozlov, 2015).

Biologically there exists several mechanisms under which RONS can be generated, with the electron transport chain (ETC) from the mitochondria for instance being the major source for intracellular ROS production. It has been reported that during aerobic respiration the transfer

of electrons along the ETC from Complex I (nicotinamide adenine dinucleotide (NADH) dehydrogenase) and Complex III (ubiquinone cytochrome c reductase) to molecular oxygen accounts for 1-2% of oxygen molecules being converted to superoxide anions ($O_2^{\cdot -}$), hydroxyl radicals ($OH^{\cdot}$) or hydrogen peroxide (H_2O_2) (Krumova and Cosa, 2016).

In addition, members of the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase (NOX) family represent another major superoxide-generating system. In particular, NOXs are membrane-bound enzymes that generate $O_2^{\cdot -}$ by using NADPH to ensure the transfer of electrons to oxygen molecules located on the extracellular side of the membrane (Figure 11). Yet, depending on which NOX isoform becomes expressed, they can elicit different cellular transformations and subsequently yield diverse biological outcomes, therefore, although representing a single protein family, NOX represent a model for how ROS generation is specifically governed to ensure normal cellular signalling and homeostasis. However, it is important to keep in mind that cellular enzymes such as xanthine oxidase, cytochrome p450, lipoxygenases and nitric oxidase synthase (NOS) represent additional intracellular sources for the production of RONS (Figure 11) (Krumova and Cosa, 2016).

Collectively, RONS can be classified as being either a primary or secondary RONS, whereby primary RONS such as $O_2^{\cdot -}$, H_2O_2 and nitric oxide (NO) are reported in most instances to reversibly react with their target molecules and have a weak damaging potential. However, secondary RONS such as $OH^{\cdot}$, hypochlorous acid (HOCL) and peroxynitrite ($ONOO^-$) are more toxic and linked to damaging host tissues. Nevertheless, a common factor among these oxidative molecules is that their formation often requires more than one reactive species to be present, which for instance can be noted during the synthesis of $ONOO^-$ (Weidinger and Kozlov, 2015). The biosynthesis of NO is catalysed by NOS, a homodimeric enzyme whereby each monomer comprises a C-terminal reductase domain and an N-terminal oxygenase domain that collectively facilitates the two-step oxidation of L-arginine to L-citrulline along with the concomitant production of NO. Yet, depending on selected conditions, NOS is also capable of catalysing the production of $O_2^{\cdot -}$. Therefore, NOS is able to provide insight as to how the chemical pathways of ROS and RNS are often not mutually exclusive, for instance an important mode in the inactivation of NO is its reaction with O_2^- to form the reactive oxidant $ONOO^-$ (Andrew and Meyer, 1999) (Figure 11).

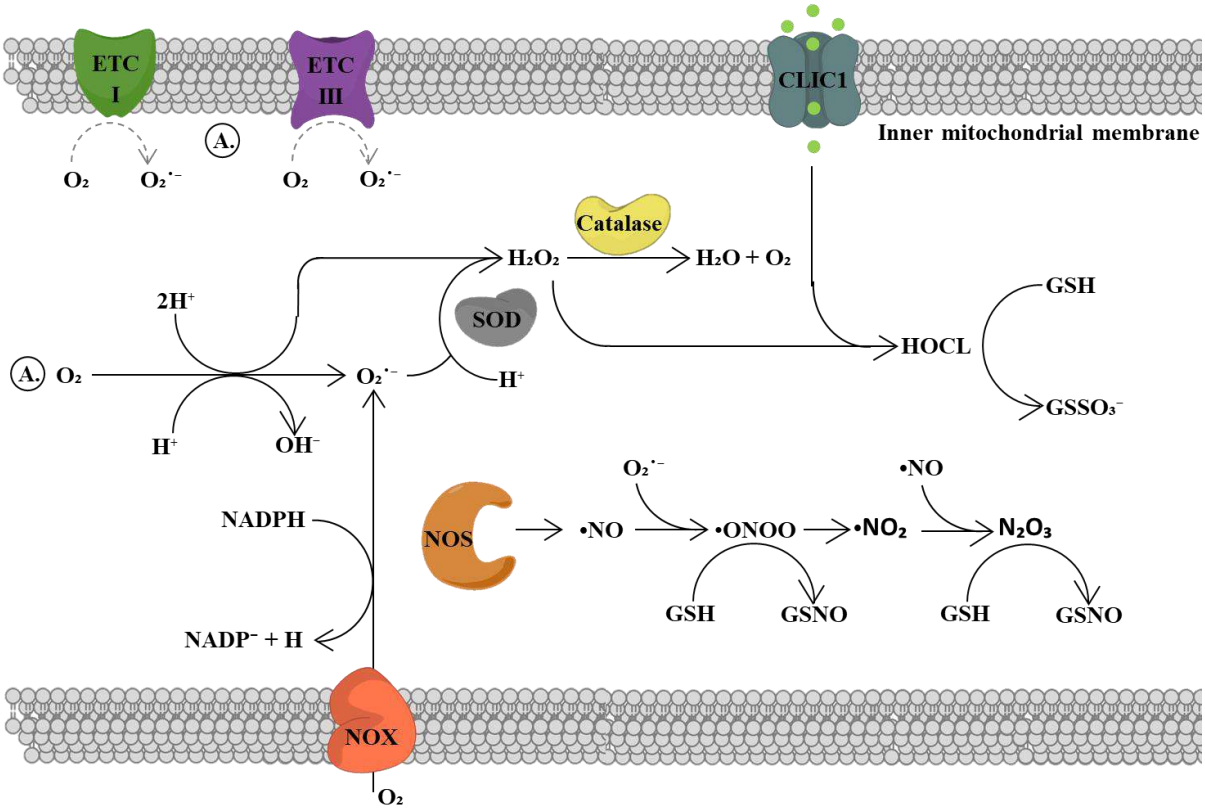


Figure 11. An overview on the production of ROS and RNS. Biologically there exists numerous mechanisms under which ROS and RNS, or otherwise referred to as RONS, can be generated. The major superoxide-generating systems include complex I and complex III from the ETC as well as NOX, which collectively ensure the production of primary RONS such as $O_2^{\cdot -}$, H_2O_2 and NO, as well as secondary RONS such as OH^- and HOCL. A common factor among these oxidative molecules is that their formation often requires more than one reactive species to be present as can be noted during the production of HOCL and $ONOO^-$. Indicated are (A) complex I (●) and III (●) from the ETC, (B) NOX (●), (C) NOS (●), (D) superoxide dismutase (SOD) (●) that catalyses the dismutation of $O^{2\cdot -}$ radical into either O_2 or H_2O_2, (E) catalase (●) that catalyses the decomposition of H_2O_2 to H_2O and (E) a Cl^- channel (●) that transports Cl^- (●) across the lipid bilayer. The figure was adapted from Krumova and Cosa, 2016 and was generated by using ChemBioOffice®Ultra 14.0.

1.3.3.2 Protein oxidation: a focus on cysteine residues

The oxidative modification of proteins by RONS can either lead to modifications in the side chains of amino acid residues, the cleavage of the protein backbone, the generation of carbonyl derivatives and/or the formation of crosslinked protein complexes. The specificity of these reactions is often diverse, with some reactions being limited to selected residues, while others often give rise to widespread non-specific modifications (Barelli *et al.*, 2008). It is important to note, however, that RONS not only modifies protein molecules, but it is also renowned for reacting with the sugar moieties of DNA to ensure the peroxidation of lipids, the by-products of which can further modify proteins (Weidinger and Kozlov, 2015).

A common mechanism in the redox regulation of protein molecules is the oxidation of cysteine residues. In contrast to methionine where the sulfur atom is incorporated into a less reactive thioether bond, the thiol group of cysteine is ionisable and therefore when deprotonated it can form the negatively charged and highly reactive thiolate group (Figure 12). However, although thiol and thiolate groups have lone pair of electrons and are nucleophilic, the chemical properties of these two chemical moieties are distinct and the equilibrium between them varies with pH and their location within a protein (Poole, 2014).

The thiol group of cysteine is reported to have a pK_a ~8.5, therefore considering that this is close to the physiological pH of the cytosol, this chemical property ensures that small variations within the physiological pH of a cell can easily trigger cysteine to function as a nucleophile (Figure 12) (Poole, 2014). The biochemical properties of cysteine, however, are often hard to characterise considering that solvent exposed cysteines tend to be polar, whereas due to the hydrophobic nature of amino acid packing inside a protein, buried cysteines tend to be hydrophobic (Marino and Gladyshev, 2010). Furthermore, depending on the molecular environment within a protein, reactive cysteines often have a lowered pK_a value should the negatively charged thiolate ion be stabilised by specific hydrogen bonds and positive residues, whereas cysteine residues located within a hydrophobic or electronegative local environment tend to have a higher pK_a value (Poole, 2014).

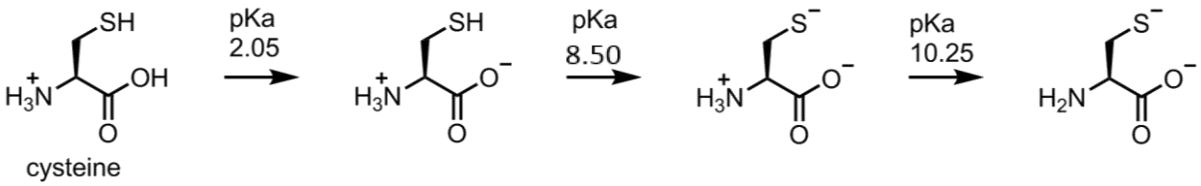


Figure 12. The ionisation of cysteine amino acids. The thiol group of cysteine is reported to have a pKa ~8.5 and therefore considering that this is close to the physiological pH of the cytosol, this chemical property ensures that small variations within the physiological pH of a cell can easily convert cysteine to function as a nucleophile. Although thiol and thiolate groups have lone pair of electrons and are nucleophilic, the chemical properties of these two chemical moieties are distinct and the equilibrium between them varies with pH and their location within a protein. The figure was adapted from Poole, 2014.

The majority of isolated and exposed cysteine residues would be susceptible to being directly targeted by a wide range of modifications and as a result newly evolved cysteine residues which are surface exposed are the least likely to be neutral. Consequently, when these cysteine residues are not involved in functional interactions, they are likely to be subject to negative selection in order to avoid unwanted reactions. Therefore, the anomalous hydrophobic-like behaviour of cysteine residues can be explained by the elimination of isolated cysteines rather than the enrichment of buried cysteine residues being based solely on the physical-chemical properties of this amino acid (Marino and Gladyshev, 2010). Reiteration of this process over a larger timescale throughout the course of evolution can therefore explain the controlled and limited overall abundance of cysteines on protein surfaces in order to circumvent certain diseases that would arise from the dysregulation in their reactivity. Instead, the clustering of cysteine residues that can partake in disulfide bond formation and metal-binding have been associated with a very high degree of conservation (Marino and Gladyshev, 2010). Therefore, apart from the sheer physical-chemical properties of free cysteine amino acids, the functional importance of these residues will dictate their abundance and location within a protein.

1.3.3.3 Protein *S*-nitrosylation

The production of NO *in vivo* is initiated by the endothelial, inducible and neuronal isoforms of NOS and therefore NO can be seen to have an influence on a variety of biological targets. In particular, the modification of a protein by NO can be seen to arise through different chemical processes that may either involve NO interacting directly with a protein, or indirectly when NO is involved in the synthesis of RONS (Förstermann and Sessa, 2012). A common example in the direct interaction of NO with a protein is the coordination between NO and the transition metals within heme groups. In fact, the primary receptor for NO is the soluble isoform of guanylate cyclase (sGC), an enzyme ensuring the synthesis of cyclic guanylate monophosphate (cGMP) upon its activation by NO interacting with the transition metal within its heme group (Förstermann and Sessa; 2012; Smith and Marletta, 2012).

The covalent modification of proteins by NO can also occur at cysteine, tryptophan or the amines of lysine and N-terminal residues. However, owing to its high reactivity and influence on numerous protein functions, the modification of the side chain of cysteine residues can be considered one of the most important PTMs of a protein. Yet, although thiol groups can be targets for NO, under physiological conditions NO is a poor oxidant and the direct interaction

of NO with cysteines occurs instead with thiolate ions to subsequently form *S*-nitrosothiols (Broniowska and Hogg, 2012). Collectively, in biological systems there are four main mechanisms under which *S*-nitrosylation can occur, namely: (i) an oxidative pathway, (ii) a radical-mediated pathway, (iii) transnitrosation pathway and a (iv) metal-catalysed RSNO formation pathway (Kovacs and Christian, 2013; Smith and Marletta, 2012) (Figure 13).

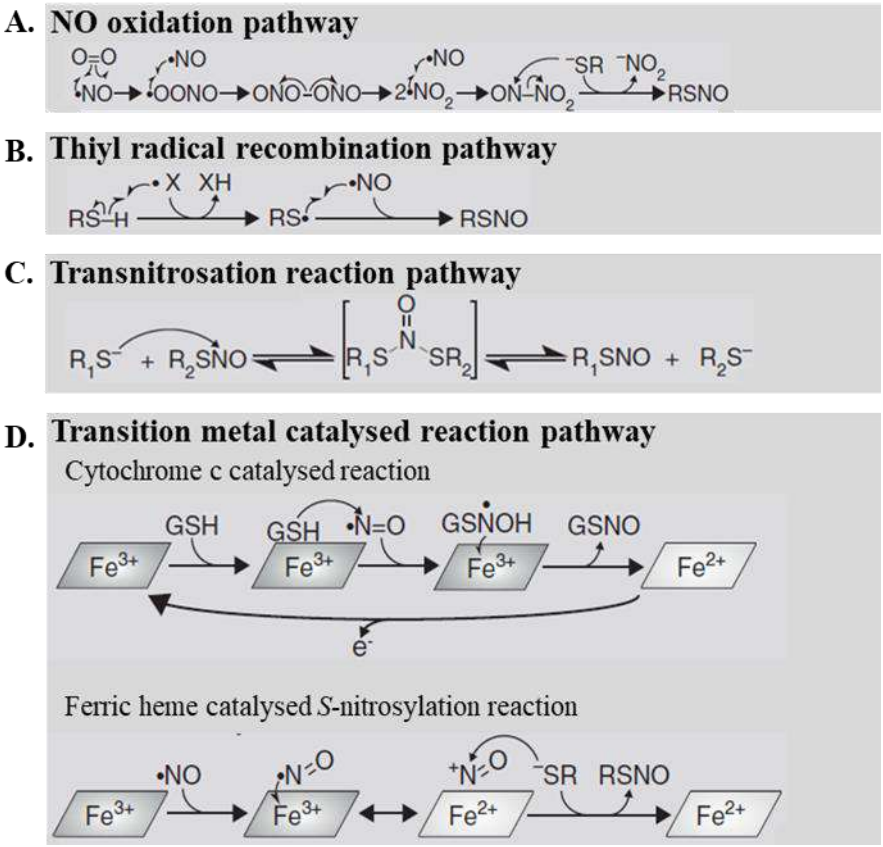


Figure 13. Types of reaction mechanisms underlying the formation of *S*-nitrosothiols. Biologically there exists numerous reaction mechanisms that are capable in driving the synthesis of *S*-nitrosothiols. (A) During the NO oxidation pathway the recombination of NO and NO_2 radicals results in the formation of *S*-nitrosothiols and NO^{2-}. (B) During the thiyl radical pathway, the abstraction of a hydrogen atom from a thiol group by another radical will form a thiolate ion/thiyl radical that can readily interact with NO to form *S*-nitrosothiols. (C) An *S*-nitrosothiol group can be transferred to other molecules through a transnitrosation reaction, whereby a thiolate ion will attack the nitrogen atom from an *S*-nitrosothiol in order to form a nitroxyl disulfide intermediate that will decay into a thiolate and new *S*-nitrosthiol. (D) The formation of *S*-nitrosothiols by transition metals can occur through several pathways, for instance ferric (Fe^{3+})-heme that is present in proteins such as cytochrome c can bind to NO and the resulting Fe^{3+}-NO complex can then react with a thiolate to form an *S*-nitrosothiol and ferrous (Fe^{2+})-heme. Alternatively, thiols such as GSH can associate with Fe^{3+} and react with NO to form a GSNOH radical, whereby the Fe^{3+} heme will accept an electron from GSNOH and subsequently form Fe^{2+} heme and GSNO. The following figure was adapted from Smith and Marletta, 2012.

As a free radical, NO can have multiple reaction pathways, for instance the oxidation or reduction of NO can form nitrosonium cations (NO^+) or nitroxyl anions (NO^-), respectively, both of which act as *S*-nitrosylating agents. Furthermore, the auto-oxidation of NO under aerobic conditions can generate nitrogen dioxide (NO_2), which due to its lone pair of electrons can oxidise thiols to thiyl radicals as well as react with NO to generate dinitrogen trioxide (N_2O_3), a highly reactive compound that can form nitrite (NO_2^-) ions or *S*-nitrosothiols when reacting with water or thiolates, respectively (Smith and Marletta, 2012) (Figure 13 and 14). Although the physiological relevance of *S*-nitrosothiol formation through the radical-mediated pathway remains controversial, it has been reported through pulse radiolysis and kinetic competition assays that the reaction between NO and thiyl radicals proceeds at a fast rate on the order of 2×10^9 $M^{-1}.s^{-1}$, therefore indicating a plausible route *in vivo* for the *S*-nitrosylation of proteins (Madej *et al*., 2008).

S-nitrosylation has also been shown to occur, even sometimes enhanced, under anaerobic conditions, therefore suggesting that oxygen may not be a prerequisite for *S*-nitrosylation (Smith and Marletta, 2012). Furthermore, considering that the physiological concentrations of NO and oxygen are within the nanomolar and micromolar range, respectively, the direct oxidation between N_2O_3 and thiols is suggested to be a rate-limiting reaction (Gaston *et al*., 2003). In support of this notion is also the fact that *in vivo* the formation of N_2O_3 from NO_2 is easily outcompeted by the reaction of NO_2 with free thiols and uric acid to form thiolates and urate, respectively (Ford *et al*., 2002; Silkstone *et al*., 2012). However, it has been suggested that the local concentration of NO and oxygen increases in the hydrophobic regions of membranes and lipoproteins. In fact, the rate of N_2O_3 formation in these hydrophobic environments has been reported to be at least 300 times higher in comparison to a more polar environment. Interestingly, the isoforms of NOS are noted to interact with caveolin, an integral membrane protein associated with caveole lipid rafts. Therefore, considering that NOS is capable of catalysing the synthesis of NO and $O_2^{\cdot -}$ under selected conditions, the association of NOS with biological membranes may support the enhanced formation of N_2O_3 (Zhang *et al*., 2009) (Figure 14).

Seeing that oxygen may not be a prerequisite in the formation of *S*-nitrosothiols, electron acceptors with large electrode potentials such as transition metals may instead prove to be beneficial in catalysing the reaction between NO and thiyl radicals. In fact, some of the fastest

reactions with NO occurs at proteins with iron-containing prosthetic groups such as cytochrome C oxidase (Sarti *et al*., 2000), sGC and haemoglobins (Brandish *et al*., 1998). Yet, depending on the redox environment, ferric ions (Fe^{3+}) or ferrous ions (Fe^{2+}) may either accept or donate electrons to NO to form NO^+ or NO^- ions, respectively, both of which are *S*-nitrosylating agents. Furthermore, iron ions are able to generate *S*-nitrosothiols by ensuring the oxidation of thiols from nearby and exposed cysteine residues (Kovacs and Christian, 2013) (Figure 14). When associated with iron ions and NO, low molecular weight thiols form a dinitrosyl iron complex that will aid as a carrier for NO to ensure that NO is transferred to the metal-centres of metalloproteins and/or that NO^+ ions are transferred to the thiyl radicals within other proteins. In fact, one of the most important *S*-nitrosylation reactions *in vivo* is transnitrosation, a phenomenon whereby *S*-nitrosothiols can transfer their nitroso moieties to other acceptor thiols (Kovacs and Christian, 2013) (Figure 13 and 14).

In human cells the most abundant thiol is GSH and depending on the cellular redox status it can reach concentrations as high as 10 mM, therefore it follows that the most abundant *S*-nitrosothiol will be *S*-nitrosoGSH (GSNO) within cells. However, although occurring *in vivo*, the transnitrosation of proteins by GSNO is likely to only achieve physiologically significant rates when the concentrations of GSNO are high, as is the case under conditions of intense nitrosative stress (Yap *et al*., 2010). Collectively, it is clear that there are numerous biological reaction mechanisms capable of enhancing the formation of thiolate ions to therefore drive the *S*-nitrosylation of proteins. However, *in vivo* the transnitrosation reactions between GSNO and free protein thiols proves to be one of the most significant reaction pathways. In fact, the treatment of human cells with GSNO *in vitro* has been demonstrated to be efficient in stimulating the *S*-nitrosylation of proteins (Paige *et al*., 2008) and furthermore mice whose GSNO reductase (GSNOR) has been knocked out show an increase in protein *S*-nitrosylation, therefore indicating that the *S*-nitrosylated state of proteins is in equilibrium with cellular GSNO (Liu *et al*., 2001).

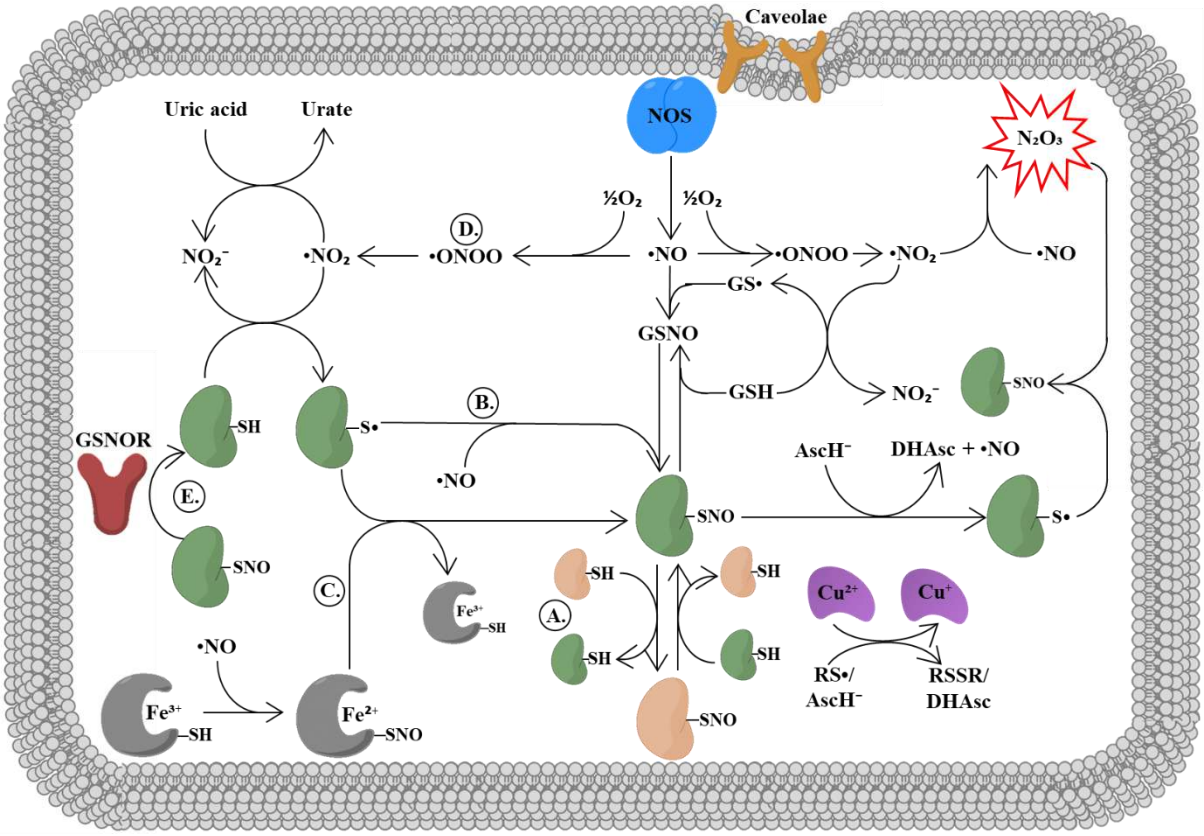


Figure 14. Overview of the reaction pathways involved in the formation and degradation of *S*-nitrosothiols. Upon the production of NO from NOS, the synthesis of GSNO or other *S*-nitrosothiols can occur either through a (A) transnitrosation pathway, (B) thiyl radical recombination pathway, (C) a transition-metal pathway and/or (D) an NO oxidation pathway. It has been suggested that the local concentration of NO and oxygen increases in the hydrophobic regions of membranes and lipoproteins. In fact, in hydrophobic environments such as caveolae where NOS can be found associated with caveolin, the rate of N_2O_3 formation has been reported to be at least 300 times higher( ) in comparison to a more polar environment. However, considering the physiological concentrations of NO and O_2, the formation of N_2O_3 is suggested to be a rate-limiting reaction that would otherwise also be easily outcompeted by the reaction between NO_2 and uric acid. Finally, although denistrosation can occur enzymatically by means of GSNOR, carbonyl reductase 1 (CBR1) and thioredoxin (Trx), denistrosation can also be catalysed by trace amounts of Cu^+, whose redox state is ultimately highly regulated by reductants such as ascorbic acid and GSH. Indicated are (i) NOS (●), (ii) two different *S*-nitrosylated proteins (●) and (●), (iii) metalloproteins that may have either a copper (●) or iron (●) ion as a cofactor and (iv) denitrosating enzymes (●). The figure was adapted from Smith and Marletta, 2012 and was generated by using ChemBioOffice®Ultra 14.0

1.3.3.4 The chemistry of *S*-nitrosothiols and their target sites

Although only a few structures of *S*-nitrosylated proteins have been solved, crystallographic studies on selected *S*-nitrosylated small molecules and proteins have been able to provide insight into the structure of *S*-nitrosothiols (Bartberger *et al.*, 2000). It is noted that *S*-nitrosothiols tend to show significant S-N double-bond characteristics due to the delocalisation of their N=O π electrons. Furthermore, although both *cis*- and *trans*-isoforms have been noted in small organic *S*-nitrosothiols, the *cis*-conformation appears to be prevalent in the solved structures of *S*-nitrosylated proteins. In relation to this, quantum mechanical calculations have placed the energy minima for the C-S-N-O dihedral angle at 0º and 180º, which corresponds to the *cis*- and *trans*-isoform conformations, respectively (Figure 15). Interestingly, it is for this reason that GSNO for instance is pink in colour while *S*-nitroso-N-acetylpenicillamine (SNAP) is green, since the structural conformation of their dihedral angles are in a *cis*- and *trans*-conformation, respectively (Zhao and Houk, 2006).

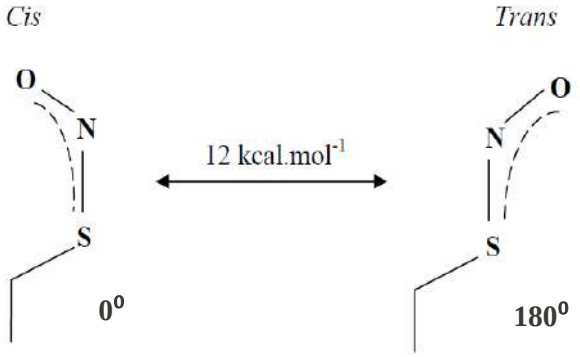


Figure 15. The chemical structure of *S*-nitrosothiols. It is noted that *S*-nitrosothiols tend to show significant S-N double-bond characteristics due to the delocalisation of their N=O π electrons (- - -). Quantum mechanical calculations have placed the energy minima for the C-S-N-O dihedral angle at 0^0 and 180^0, which corresponds to the *cis*- and *trans*-isoform conformations, respectively and the energy barrier for the rotation between these two conformations was calculated to be 12 kcal.mol^{-1}. Although both *cis*- and *trans*-isoforms have been noted in small organic *S*-nitrosothiols, the *cis*-conformation appears to be prevalent in the solved structures of *S*-nitrosylated proteins. The figure was adapted from Zhao and Houk, 2006.

S-nitrosothiols are subject to several mechanisms of decomposition such as S-N bond homolysis, photolytic S-N bond cleavage, metal ion-catalysed decomposition and hydrolysis. It is simply for this reason that the half-lives of *S*-nitrosothiols reported in literature vary dramatically, furthermore it is evident that the differences in the R-substituent of RSNO along with the double-bond character of the S-N linkage account for these variations. In proteins, these features are controlled by the local environment such as hydrophobicity, steric hindrance

and/or chemical groups which may, or may not, stabilise the products formed during the decomposition of the S-N bond (Broniowska *et al*., 2013).

While the spontaneous thermal homolysis of *S*-nitrosothiols has been recorded *in vitro*, the activation energies which have been reported do not seem feasible in order for this process to be physiologically relevant (Broniowska *et al*., 2013). Instead, the major route for the decomposition of *S*-nitrosothiols under physiological conditions is not spontaneous, but rather the decomposition of *S*-nitrosothiols will be catalysed by trace amounts of monovalent copper ions (Cu^{+}). Therefore, considering copper ions are essential cofactors for redox-regulated proteins such as superoxide dismutase (SOD), nitrite reductase (NiR) and nitrous oxide reductase (N2RO), the decomposition of *S*-nitrosothiols through a copper ion-linked pathway appears to be physiologically relevant (Broniowska *et al*., 2013; Holmes and Williams, 1998). In fact, it is has been found that the decomposition of GSNO is progressively attenuated in the presence of the specific Cu^{+} chelator, neocuproine, which *in vivo* is noted to reduce both the anti-platelet aggregation and vasodilation responses which are typically mediated by the release of NO from GSNO (Holmes and Williams, 1998; Williams, 1998).

In principle, any reducing agent capable of reducing divalent copper ions (Cu^{2+}) to Cu^{+} could be linked to the decomposition of *S*-nitrosothiols. Therefore, *in vivo* a more probable reducing agent of Cu^{2+} would be GSH and ascorbic acid considering their abundance within a cell. Though in particular, the reaction mechanism of ascorbic acid is very concentration dependent, that is, at high concentrations, ascorbic acid will undergo a nucleophilic reaction with the nitrogen of the nitroso functional group in order to liberate NO and dehydroascorbic acid by means of a homolytic cleavage reaction. The decomposition of GSNO by ascorbic acid can in fact occur through three different species of ascorbic acid, namely H_2A, HA^{-} and A^{2-}, with the release of NO via the A^{2-}pathway being the most effective. Though it should be kept in mind that the concentration of A^{2-} tends to only increase with an increase in pH and therefore under the physiological pH range of pH 5-8, the HA^{-} pathway is the most prominent pathway in the decomposition of GSNO into NO and its respective thiol. However, at low ascorbic acid concentrations the direct reaction with GSNO is too slow to be significant and instead denitrosation persists through the Cu^{+} reaction pathway. Furthermore, in contrast to the direct (high concentration) reaction pathway which leads to the formation of thiols, the Cu^{+} reaction

pathway appears to yield disulfides upon the decomposition of *S*-nitrosothiols (Holmes and Williams, 1998; Singh *et al.*, 1996; Smith and Dasgupta, 2000) (Figure 14).

It is evidently clear that the formation and decomposition of *S*-nitrosothiols *in vivo* is significantly complex and depends on a network of reaction pathways. However, while the copper-ion pathway has proven to be significant in the decomposition of *S*-nitrosothiols, the chelation of Cu^{2+} ions by GSSG has additionally been reported (Holmes and Williams, 1998). Therefore, considering the ubiquity of GSH in a cell, the aforementioned pathway may not be as significant under selected physiological conditions such as oxidative stress where the synthesis of GSSG can become apparent (Zitka *et al*., 2012). As previously mentioned, concomitant to the high intracellular concentrations of GSH, the most abundant *S*-nitrosothiol *in vivo* is GSNO and consequently its involvement in transnitrosation reactions with proteins presents an alternative pathway for its decomposition. In fact, following the transnitrosation reaction from a relatively unreactive *S*-nitrosothiol, the newly formed *S*-nitrosothiol can be much more reactive to NO formation in its new environment, which therefore supports the notion that transnitrosation reactions can provide a more rapid alternative pathway to NO formation (Singh *et al.*, 1996; Williams, 1998).

Overall, despite the reaction mechanisms underlying the formation of *S*-nitrosothiols being well understood, there is no universal rule to indicate which cysteine residue/s will become *S*-nitrosylated within a protein. However, it has been noted in selected structures that flanking reactive cysteine residues are motifs comprised of acidic or basic residues that can either suppress or favour the formation of thiolates through electrostatic interactions, respectively. Furthermore, these acid-base motifs have an influence on the electrostatic potential distribution of proteins and have therefore been shown to play a role in protein-protein interactions that can facilitate the transnitrosation of target proteins. Similarly, apolar residues such as alanine, glycine and isoleucine are often noted to be enriched in the flanking regions of modified cysteines and therefore it has been suggested that these apolar residues, along with aromatic side chains, play a role to sequester or stabilise radicals that can result in the formation of *S*-nitrosothiols. Furthermore, proteins with transition metal cofactors located near exposed cysteine residues are likely to facilitate *S*-nitrosylation considering their ability to lower the pK_a of thiol groups. Nevertheless, although much effort has been directed to identifying consensus structural features that can describe the specificity of *S*-nitrosylation, the accurate prediction of these sites in proteins still remains a challenge (Kovacs and Christian, 2013).

1.3.4 The influence oxidative stress has on the structure and function of CLIC1

Native to the TMD of CLIC1 is a conserved cysteine residue (Cys24) (Figure 9 and 16) which has been shown to play a functional role in the Cl^- channel activity of CLIC1 considering that the C24A or C24S mutants of CLIC1 reduced or completely eliminated its Cl^- channel activity, respectively (Littler *et al*., 2004; Singh and Ashley, 2006). Considering that there is often a change in the redox environment of a cell during various diseases, when soluble CLIC1 was placed in an oxidising environment an intramolecular disulfide bond formed between Cys24 and Cys59 and subsequently the conformation of CLIC1 changed to an all α-helical monomeric unit with an exposed hydrophobic surface (Figure 16B). It was therefore suggested that *in vitro* this hydrophobic surface would be masked by non-covalent dimerisation (Figure 16A), but *vivo* this region could represent a surface for membrane docking. Consequently, the membrane insertion and/or ion channel activity of CLIC1 has been suggested to be regulated by the redox environment of a cell (Littler *et al*., 2004).

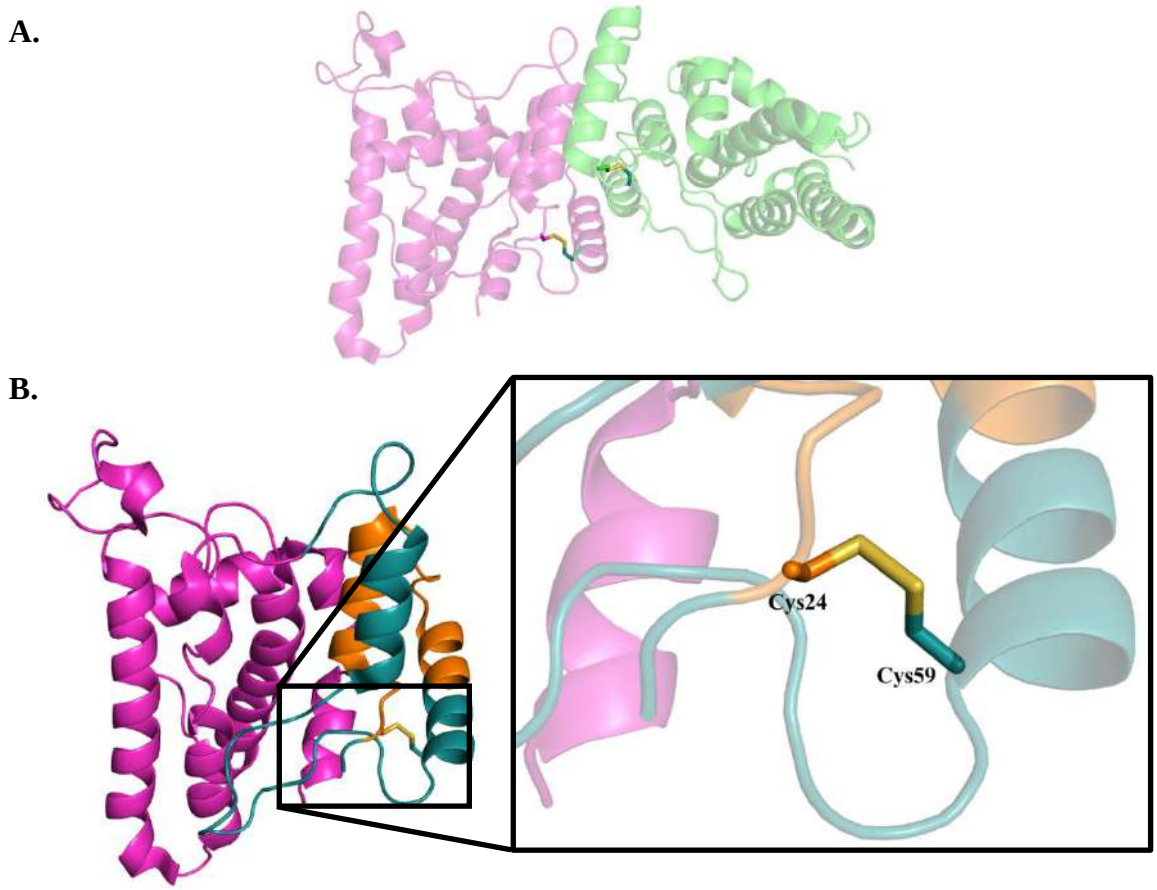


Figure 16. The structural conformation of oxidised CLIC1. (A) The oxidation of CLIC1 results in the exposure of a hydrophobic surface that will become masked by the non-covalent dimerisation of two oxidised CLIC1 monomers. (B) In contrast to the reduced conformation of CLIC1 which has an apparent all α-helical C-terminal domain (●) and a thioredoxin-like N-terminal domain (βαβαββα motif) (●) which comprises the TMD (●) region, the oxidised conformation of CLIC1 contains an intramolecular disulfide bond between Cys24 and Cys59 which causes its conformation to change to an all α-helical monomeric unit. All images were generated by using Pymol V1.3.

The intracellular redox environment of a cell is relatively reducing owing to the high concentration of GSH (10 mM) and under these conditions it has been demonstrated that the ion channel conductance of CLIC1 was lower and that the ion channel either remained open or closed for longer, therefore reiterating how the redox environment of a cell would influence the ion channel activity of CLIC1 (Singh and Ashley, 2006). However, given that the other CLIC proteins lack a cysteine residue which is structurally aligned to Cys59 and that the disulfide bond between Cys24 and Cys59 would need to be broken in order for the TMD to transverse the lipid bilayer, it has been proposed that the oxidation of CLIC1 to a dimeric state is not plausible prior to membrane insertion. Instead, it has been suggested that when the monomeric conformation of CLIC1 associates with the membrane surface it will produce an intermediate structure reminiscent of the dimeric state in order to induce the structural changes required for membrane insertion (Goodchild *et al*., 2009).

Interestingly, Varela *et al*. (2019) recently investigated by means of nuclear magnetic resonance (NMR) spectroscopy if the oxidation of CLIC1 was necessary for membrane insertion. In this study, both the soluble and membrane fractions from the cell lysate of C43 *Escherichia coli* were used considering that they represented the soluble and membrane-bound conformation of CLIC1, respectively. The rational of this study was that because the chemical shifts in nitrogen-15 (N^{15}) and hydrogen-1 (H^{1}) isotopes are sensitive to the dynamics of a protein, any large structural changes such as those noted during the oxidation of CLIC1 would have a significant impact on the N^{15}- and H^{1}-resonance. However, the NMR spectra for both the reduced conformations of CLIC1 showed no significant difference in their NH resonance, instead, there was only a substantial chemical shift when both conformations were oxidised. It was therefore suggested that CLIC1 was inserted into the membrane of *E.coli* in its reduced state, but that an oxidising environment can still influence the overall dynamics of the membrane-bound conformation of CLIC1 (Varela *et al*., 2019). The literature surrounding the oxidation of CLIC1 appears to be controversial, nevertheless, although oxidation may not be an absolute requirement for membrane insertion, when the surrounding environment becomes more oxidising as seen in many disease states, it could facilitate membrane insertion and also influence the ion channel activity of CLIC1.

1.3.5 *S*-nitrosylation regulates the membrane-insertion potential of CLIC1

Apart from CLIC1 being oxidised to form a disulfide bond, the PTM of CLIC1 could also be induced by RNS whereby the cysteine residues within CLIC1 could for instance become *S*-nitrosylated. In fact, the *S*-nitrosylation of CLIC proteins proves to be significant given that the *S*-nitrosylation of CLIC4 has previously been established to be a key regulatory mechanism for its nuclear translocation. Present within CLIC4 are four cysteine residues, however, based on mutagenesis studies it has been confirmed that Cys234 is a target site for *S*-nitrosylation, while Cys35 and Cys189 may act as allosteric regulators and protect Cys234 from becoming *S*-nitrosylated. In the crystal structure of CLIC4 it can be seen that Cys35, Cys189 and Cys234 are not only in close proximity to one another, but they are also closely associated with the NLS. Therefore, under normal physiological conditions Cys35 and Cys189 may protect Cys234 from becoming *S*-nitrosylated and consequently this may prevent the NLS from becoming exposed to the nuclear import machinery (Malik *et al*., 2010).

Structurally, CLIC1 and CLIC4 are very similar, but in particular it is noted that Cys24, Cys89, Cys178 and Cys223 in CLIC1 are structurally aligned to Cys35, Cys100, Cys189 and Cys234 in CLIC4, respectively (Figure 17). Therefore, considering the importance of Cys24 in the redox control of CLIC1, it appears that similar to Cys35, Cys189 and Cys234 in CLIC4, Cys24, Cys178 and Cys223 may be target sites for *S*-nitrosylation in CLIC1. In turn, this PTM could have an influence in regulating the conformational stability of CLIC1 that proves to be important for its membrane insertion. Interestingly, the NLS in CLIC4 is structurally aligned with a putative NLS site (^{188}QVVCKKYRG^{196}) in CLIC1, but in contrast to CLIC4, the putative NLS site in CLIC1 contains a cysteine residue (Cys191) (Figure 17). Therefore, in view of the fact that CLIC1 has been found to be localised in the nucleoplasm and outer nuclear membrane (Valenzuela *et al*., 1997), the *S*-nitrosylation of Cys191 may have a profound influence on the ability of CLIC1 to bind to nuclear transport proteins and/or its membrane insertion potential.

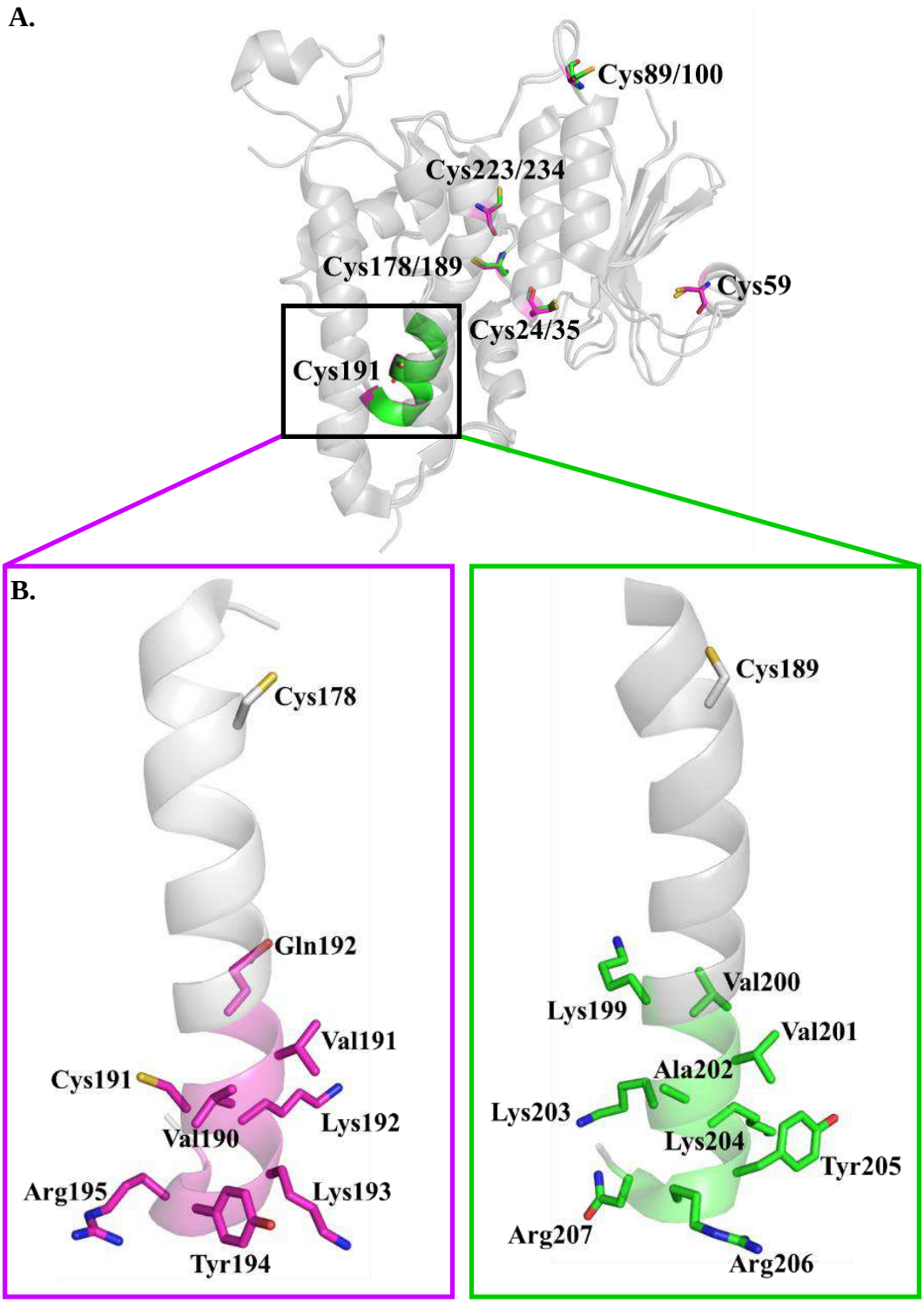


Figure 17. Structural homology between CLIC1 and CLIC4. (A) Structurally the soluble conformation of CLIC1 (PDB: 1K0M) is similar to CLIC4 (PDB: 1EEM) (RMSD 1.487 Å) (both of which are indicated in grey). However, in particular it is noted that Cys24, Cys89, Cys178 and Cys223 in CLIC1 (●) are structurally aligned to Cys35, Cys100, Cys189 and Cys234 in CLIC4 (●), respectively. Therefore, similar to Cys35, Cys189 and Cys234 in CLIC4, the *S*-nitrosylation of Cys24, Cys178 and Cys223 in CLIC1 may have an influence in regulating its conformational stability. (B) The NLS in CLIC4 is structurally aligned with a putative NLS site in CLIC1, but in contrast to CLIC4, the putative NLS site in CLIC1 contains a cysteine residue (Cys191). All images were generated by using Pymol V1.3.

The effect that *S*-nitrosylation may have on the membrane insertion and/or ion channel activity of CLIC1 can be noted in neurodegenerative disorders such as Alzheimer's disease (AD). During AD, the production of insoluble β-amyloid protein (Aβ) oligomers causes the activation of N-methyl-D-aspartate (NMDA)-type glutamate receptors, which consequently induce the influx of calcium ions. However, the activation of neuronal NOS (nNOS) within neuronal cells and inducible NOS (iNOS) within microglial cells are dependent on calcium ions. Therefore, owing to the production of Aβ-oligomers, the influx of Ca^{2+} ions will ensure the concomitant synthesis of NO. As a result, because of the high levels of NO generated during AD, cysteine residues which would not typically become *S*-nitrosylated under normal physiological concentrations of NO can now in fact become *S*-nitrosylated, consequently resulting in aberrant protein *S*-nitrosylation during AD (Nakamura *et al*., 2013).

Importantly, during AD it is also noted that Aβ-oligomers cause an increase in the influx of Cl^- into microglial cells due to the overexpression of CLIC1 and consequently this induces the production of neurotoxins and pro-inflammatory compounds such as RNOS and tumour necrosis factor-α (TNF-α). However, introducing these Aβ-exposed microglial cells to IAA-94 was noted to prevent neuronal apoptosis given that the decreased Cl^- channel activity of CLIC1 attenuated the production of neurotoxins by microglial cells (Novarino *et al*., 2004). Therefore, considering that aberrantly *S*-nitrosylated proteins play a crucial role in the pathogenesis of AD (Nakamura *et al*., 2013) and that *S*-nitrosylation is crucial in regulating the subcellular location and function of CLIC4, the *S*-nitrosylation of CLIC1 during AD is suggested to play a key role in regulating its membrane insertion potential and/or ion channel activity under these conditions. Consequently, being able to regulate the *S*-nitrosylation status of CLIC1, or being able to use the *S*-nitrosothiol moiety as a target site is promising for the development of new therapeutics that could attenuate the symptoms induced by the aberrant expression of CLIC1 during AD.

1.4 Stringent starvation protein A (SspA) is an important structural homolog to CLIC1

Apart from CLIC proteins, several GST-like proteins have been shown to contain the canonical topology of cytosolic GSTs, however, these proteins have largely been centred on eukaryotic organisms. Nonetheless, Hansen *et al*. (2005) provided insight on SspA from *E.coli*, a bacterial GST-like protein, which similar to CLIC proteins is metamorphic and can reversibly transit between a soluble and an integral membrane-bound conformation, the latter representing a functional Cl^- channel.

SspA is well conserved among Gram-negative bacteria, however, apart from forming a Cl^- channel, SspA functions as an RNA polymerase-associated protein (RNAP) in order to regulate gene expression (Ishihama and Saitoh, 1979). In particular, SspA is required for the transcriptional activation of phage P1 late genes during the lytic growth phase of the P1 bacteriophage (Hansen *et al*., 2003). Furthermore, SspA tends to be predominately expressed under stress conditions such as nutrient starvation and prolonged stationary-phase stress, where it will aid in conferring acid tolerance. The latter being important considering that during infection by enterohemorrhagic *E. coli* (EHEC) such as *E. coli* O157:H7, these pathogenic strains would need to tolerate the acidic environment of the gastrointestinal tract (Hansen *et al*., 2005).

In fact, similar to CLIC proteins, there is an increase in the ion channel activity of *E.coli* SspA (EcSspA) under these acidic conditions, but when exposed to IAA-94, the ion channel activity of EcSspA was inhibited. This susceptibility of *E.coli* to IAA-94 has only been observed during the early log phase of its growth and not the stationary phase, however, the growth of EcSspA null mutant *E.coli* were not shown to be susceptible to IAA-94, further supporting that EcSspA is a target for IAA-94. When exposed to IAA-94, the attenuation in the growth, but not the viability of *E.coli* has been attributed to the fact that under stress conditions EcSspA is significantly overexpressed, that is, the exposure of *E.coli* to IAA-94 causes the stress-induced overexpression of EcSspA which in turn will result in the increased survival of *E.coli* in order to reach the stationary phase (Rao *et al*., 2017).

When using planar lipid bilayer systems that were identified to be suitable for the optimum ion channel activity of CLIC1 (Singh *et al*., 2006), EcSspA was shown to auto-insert into the lipid bilayer and from a functional ion channel (Rao *et al*., 2017). Furthermore, as was the case

in vivo, when EcSspA was reconstituted into these lipid bilayers, IAA-94 was able to inhibit its Cl^- channel activity. In fact, the ion channel activity of EcSspA was shown to be anion selective, but opposed to CLIC1 whose ion channel was selective to $F^- > Cl^- > NO^- = I^-$ (Singh, 2010), it was noted that the EcSspA ion channel was instead selective to $I^- > Br^- > Cl^-$. Upon looking at the TMD found in CLIC1, it was shown to only share some sequence similarity with the putative TMD found in EcSspA, however, sequence analysis revealed that across various species, Leu29 in EcSspA was one of the main residues to be conserved at the TMD region for many CLIC and SspA proteins (Figure 18) (Rao *et al*., 2017). Therefore, similar to Leu34 in CLIC1 whereby the neighbouring Trp35 residue was noted to bind at the bilayer interface (Hare *et al*., 2016), Leu29 in EcSspA has been suggested to be located at a comparable protein-lipid interface. Interestingly, EcSspA containing a L29A mutation displayed a significant increase in the single channel amplitude and conductance, however, the channel no longer displayed any anion selectivity. It has been suggested that this could be attributed to the fact that neighbouring Leu29 is Arg26, therefore seeing that Leu29 interacts with the lipid bilayer, its disruption may have an influence on the conformation and consequently the local environment of Arg26, which would therefore have an impact in the conductance and selectively of the channel. In fact, the conformational change attributed to the L29A mutation coincided with the ion channel transitioning from a smaller pore to a wider pore (Rao *et al*., 2017).

Currently, there is no crystal structure available for either the soluble or membrane-bound conformation of EcSspA, instead only the crystal structure of SspA from *Yersinia pestis* (YpSspA) (PDB: 1YY7) has been solved. However, considering that YpSspA shares 91% sequence similarity to EcSspA, it is able to provide insight into the structure of EcSspA (Figure 18). The overall structure of YpSspA can be seen to exist as a dimer whereby each monomer adheres to the canonical topology of cytosolic GSTs by comprising an all α-helical C-terminal domain and a thioredoxin-like N-terminal domain (Figure 18). However, in contrast to cytosolic GSTs, EcSspA does not display the typical 2, 4-dinitrochlorobenzene (CDNB) enzyme activity seen by GSTs, nor does it bind to GSH given that EcSspA did not bind to a GSH sepharose matrix. Yet, apart from sharing structural and functional similarities with CLIC proteins, EcSspA also shares structural homology with GSTβ1-1 (PDB: 1A0F) and based on structural sequence alignment it can be seen that the catalytic cysteine residue (Cys10) found in GSTβ1-1 is replaced with Tyr21 in EcSspa. Upon looking at the putative GSH-binding site of EcSspA it has been suggested that Tyr21 cannot form a hydrogen bond with the thiol group

of GSH nor can it mediate the GSH conjugation with CDNB. Furthermore, it has been suggested that in contrast to Gln51 seen in GSTβ1-1, the shorter side of Ser58 in EcSspA would not be able to form a hydrogen bond with GSH and collectively this may account for EcSspA not being able to bind GSH and show GST activity (Hansen *et al*., 2005).

Instead, at the dimer interface of EcSspA is a hydrophobic surface pocket with a conserved ^{84}Pro-His-Pro^{86} motif that is suggested to be important for the activity of EcSspA (Figure 18). In fact, when these individual residues were mutated there was little to no impact on the ability of EcSspA to support acid tolerance and induce the lytic of P1 bacteriophages, however, the mutation of all three residues completely reduced the functional activity of EcSspA. Therefore, owing to the hydrophobicity of the surface pocket and that it is required for the activity of EcSspA, it has been suggested that this site may be required for protein-protein interactions such as the interaction between EcSspA and RNAP (Hansen *et al*., 2005). Therefore, apart from being able to understand the function of the surface pocket, investigating how EcSspA binds to GST inhibitors and/or their structural analogues would provide insight into the development of novel compounds that can inhibit the function of EcSspA in order to treat infections caused by pathogenic *E.coli* strains. However, similar to CLIC1, the structure of the membrane-bound conformation of EcSspA, along with the mechanism governing its membrane insertion still remains unclear.

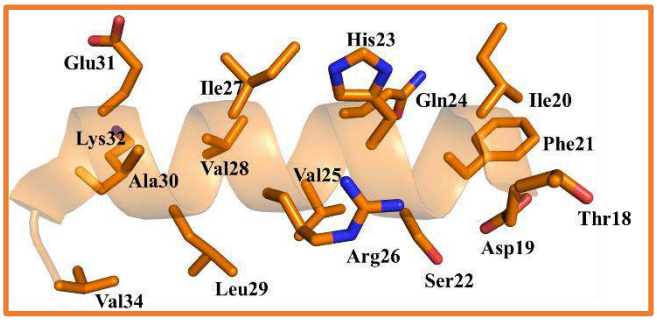
A.

```
EcSspA (WP_000257293.1)   MA--------VAANKRSVMTLFS--------GPTDIYSHQVRIVLAEKGVSFEIEHVEKDNPPQDLIDLN
YpSspA (WP_002210135.1)   MA--------VAANKRSVMTLFS--------GPTDIFSHQVRIVLAEKGVSVEIEQVEADNLPQDLIDLN
hCLIC1 (NP_001274522.1)   ----------MAEEQPQVELFVKAGSDGAKIGNCPFSQRLFMVLWLKGVTFNVTTVDTKRRTETVQKLC
hCLIC4 (NP_039234.1)      MALSMPLNGLKEEDKEPLIELFVKAGSDGESIGNCPFSQRLFMILWLKGVVFSVTTVDLKRKPADLQNLA
                                      ::.  : **          .  :*::: ::*  *** ..:  *: ..    : .*

EcSspA (WP_000257293.1)   PNQSVPTLVDRELTLWESRIIMEYLDERFPHPPLMPVYPVARGESRLYMHRIEKDWYTLMNT-IINGSAS
YpSspA (WP_002210135.1)   PYRTVPTLVDRELTLYESRIIMEYLDERFPHPPLMPVYPVARGSSRLMMHRIEHDWYSLLYK-IEQGNAQ
hCLIC1 (NP_001274522.1)   PGGQLPFLLYGTEVHTDTNKIEEFLEAVLCPPRYPKLAALNPESNTAGL-----DIFAKFSAYIKNSNPA
hCLIC4 (NP_039234.1       PGTHPPFITFNSEVKTDVNKIEEFLEEVLCPPKYLKLSPKHPESNTAGM-----DIFAKFSAYIKNSRPE
                          *    * :     .  : . * *:*:  :  *    :      ..   :     * :: :   * :.

EcSspA (WP_000257293.1)   EADAARKQLR------------------EELLAIAPVFGQKPYFLSDEFSLVDCYLAPLLWRLPQLGIEF
YpSspA (WP_002210135.1)   EAEAARKQLR------------------EELLSIAPVFNETPFFMSEEFSLVDCYLAPLLWRLPVLGIEF
hCLIC1 (NP_001274522.1)   LNDNLEKGLLKALKVLDNYLTSPLPEEVDETSAEDEGVSQRKFLDGNELTLADCNLLPKLHIVQVVCKKY
hCLIC4 (NP_039234.1       ANEALERGLLKTLQKLDEYLNSPLPDEIDENSMEDIKFSTRKFLDGNEMTLADCNLLPKLHIVKVVAKKY
                            :  .: *                    :*       ..   :: .:*::*.** * * *  :  :  ::

EcSspA (WP_000257293.1)   SGPGA----KELKGYMTRVFERDSFLASLTEAEREMRLGRS--------
YpSspA (WP_002210135.1)   TGAGS----KELKGYMTRVFERDAFLASLTEAEREMHLKTRS--------
hCLIC1 (NP_001274522.1)   RGFTIPEAFRGVHRYLSNAYAREEFASTCPDD-EEIELAYEQVAKALK-
hCLIC4 (NP_039234.1       RNFDIPKEMTGIWRYLTNAYSRDEFTNTCPSD-KEVEIAYSDVAKRLTK
                           .         :  *::..: *: *  :  .  .*:.:
```



B.

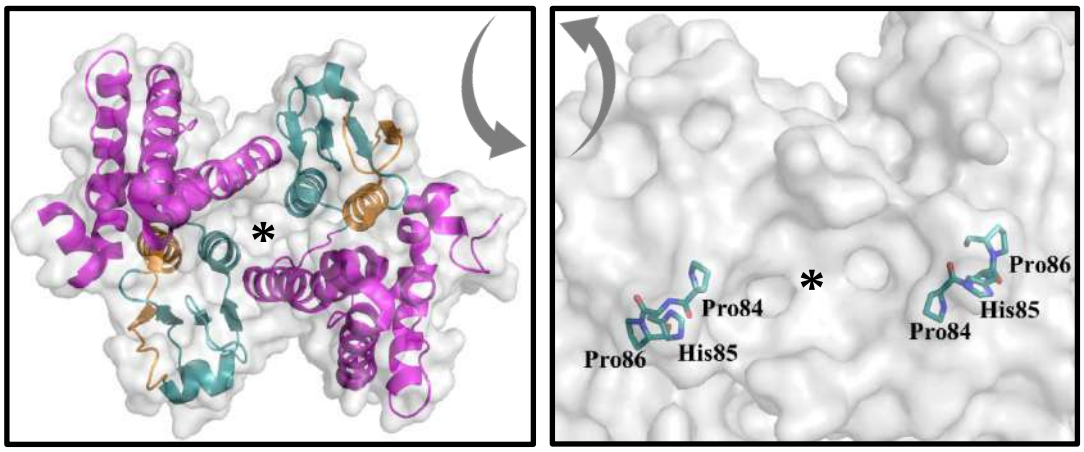


Figure 18. Soluble EcSspA is a structural homolog of soluble CLIC1. (A) The TMD found in CLIC1 and EcSspA share some sequence similarity, however, sequence analysis across various species indicated that Leu29 (□) in EcSspA was one of the main residues to be conserved at the TMD region (●) for many CLIC and SspA proteins. Sequence alignment was performed by using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and indicated are residues with perfect alignment (*), residues showing strong similarity (:) and residues showing weak similarity (.). (B) Only the crystal structure for YpSspA (PDB: 1YY7) has been solved, but it can provide insight into the structure of EcSspA given that YpSspA and EcSspA share 91% sequence similarity. The homology model of EcSspA can be seen to exist as a dimer whereby each monomer adheres to the canonical topology of cytosolic GSTs. At the back (⬇⬆) of the dimer interface (*) of EcSspA is a hydrophobic surface pocket with a conserved ^{84}Pro-His-Pro^{86} motif that is suggested to be important for the activity of EcSspA. Indicated is the all α-helical C-terminal domain (●) and the thioredoxin-like N-terminal domain (●) which contains the putative TMD (●) of EcSspA. All images were generated by using Pymol V1.3.

1.5 Inhibition of CLIC1 and its structural homologs

CLIC1 is expressed in various cell types where it exhibits a specific tissue and sub-cellular distribution, for instance CLIC1 has been found to be localised within the cytoplasm, nuclear membrane, plasma membrane, intracellular vesicular compartments and the apical domain of columnar epithelial cells (Ulmasov *et al*., 2007). The diverse distribution and concomitant physiological functions governed by CLIC1 has resulted in it being of particular interest *in vivo* considering that the dysregulation of CLIC1 has been linked notably to colorectal cancer (Petrova *et al*., 2008), cystic fibrosis (Edwards, 2006) and as previously mentioned AD (Novarino *et al*., 2004). Therefore, understanding the structural and functional characteristics for the varying conformational states of CLIC1 can present a target for novel therapeutics.

The renowned Cl^- channel inhibitor, IAA-94 (Landry *et al*., 1987), for instance has been demonstrated to inhibit the Cl^- channel activity of CLIC1 (Valenzuela *et al*., 2000). This inhibition in the ion channel activity of CLIC1 by IAA-94 has currently been linked to mitigating the metastasis of both colon cancer cells and glioblastoma cancer stem cells (CSCs), in addition to lowering the production of intracellular RONS which are typically associated with the cell cycle progression (Peretti *et al*., 2015) and in eliciting neurodegenerative disorders (Novarino *et al*., 2004). Although IAA-94 is an efficient Cl^- channel inhibitor, investigation into the specificity of additional compounds such as metformin (Gritti *et al*., 2014) and anthracene-9-carboxylic acid (A9C) (Valenzuela *et al*., 2000) (Figure 19) have come under focus given their potential to also inhibit the Cl^- channel activity of CLIC1.

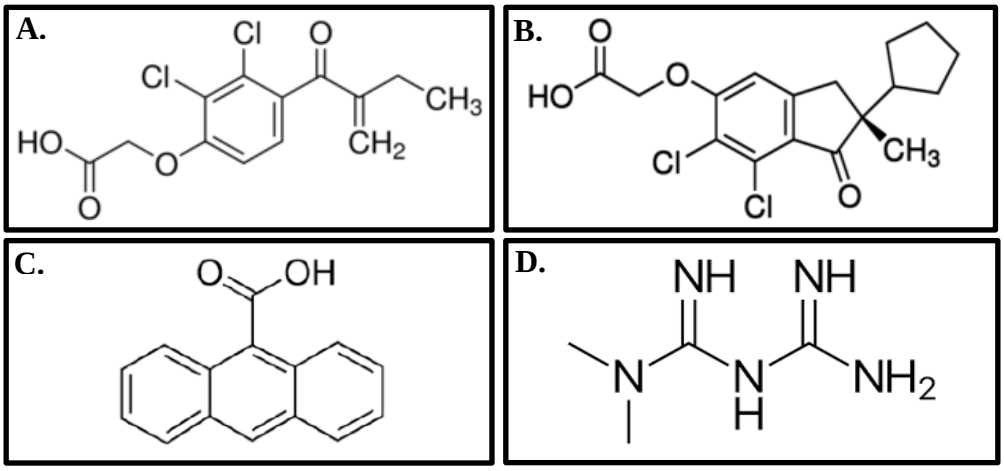


Figure 19. Compounds that inhibit the Cl^- ion channel activity of CLIC1. (A) EAA (PubChem CID: 3278), (B) IAA-94 (PubChem CID: 3667), (C) A9C (PubChem CID: 2201) and (D) metformin (PubChem CID: 4091). Despite having structural differences, the compounds B-D have been reported to inhibit the Cl^- channel activity of CLIC1.

Metformin for example is a known anti-diabetic drug, but has also been shown to influence the viability of CSCs by inhibiting the conductance of CLIC1. Importantly, however, metformin is noted to only inhibit the ion channel activity of CLIC1 in CSCs and not mesenchymal stem cells (MSCs). Consequently, this correlates with a reduction in the cell viability of tumorigenic stem cells and not normal stem cells (Gritti *et al*., 2014), therefore the possibility that CLIC1 could be used as a therapeutic target is promising given that tumour stem cells could be selectively targeted (Peretti *et al*., 2015).

Considering the structural homolog between CLIC1 and cytosolic GSTs, GST inhibitors could provide insight into additional compounds that could inhibit the enzyme activity and/or the ion channel activity of CLIC1. Looking at the crystal structure of cytosolic GSTs it can be observed that there are at least two binding sites per a monomer, namely (i) a GSH-binding site (G-site) and (ii) a hydrophobic-binding site (H-site), whereby both sites are adjacent to one another and are formed in a cleft between the N- and C-terminal domain (Figure 20). This in fact correlates with the enzyme activity of cytosolic GSTs whereby GSH from the G-site will be conjugated to a hydrophobic co-substrate at the H-site. However, in contrast to the G-site which shows great specificity for GSH, the H-site is less specific and considerably versatile and as a result GSTs are able to react with a broad range of xenobiotic compounds (Oakley, 2011; Wu and Dong, 2012). For instance, ethacrynic acid (EAA) has been demonstrated to be a potent reversible inhibitor for both human GSTP1-1 (hGSTP1-1) (Oakley *et al*., 1996) and human GSTA1-1 (hGSTA1-1) (Cameron *et al*., 1995). When bound to hGSTA1-1 both the H-site, and to some extent the G-site, are occupied (Cameron *et al*., 1995; Oakley, 2011). However, when EAA binds to hGSTP1-1 it is exclusively associated with the H-site and therefore it will compete with the hydrophobic co-substrate such as CDNB (Figure 20) (Oakley *et al*., 1996; Oakley, 2011).

The G-site from hGSTP1-1, however, shares structurally homology with the active site of CLIC1, therefore as to whether CLIC1 can bind EAA remains to be determined (Figure 20). Though, in relation to IAA-94, although the Cl^- channel activity of CLIC1 becomes inhibited by this compound, the enzyme activity for the soluble conformation of CLIC1 has interestingly also been demonstrated to be inhibited by IAA-94 (Khamici *et al*., 2015). Therefore, given the structural homology between soluble CLIC1 and hGSTP1-1, as to whether hGSTP1-1 can bind IAA-94 remains to be determined. Nevertheless, considering that IAA-94 can bind to the

soluble conformation of CLIC1, it remains uncertain if the constraint in the Cl^- conductance of CLIC1 is due to IAA-94 binding to its soluble conformation to then prevent its structural transition into a membrane competent state, and/or IAA-94 binding to the membrane-bound conformation of CLIC1 (Figure 21). Similarly, although EcSspA has not been reported to contain any GST-like enzyme activity nor has its soluble conformation been reported to bind IAA-94, given the fact that EcSspA is a metamorphic protein and that its Cl^- channel activity is inhibited by IAA-94 (Hansen *et al*., 2005), the mode underlining the constraint in the Cl^- conductance of EcSspA also remains uncertain.

Finally, although IAA-94 has been demonstrated to inhibit the ion channel activity of CLIC1, IAA-94 is not selective to CLIC1 as it can inhibit other Cl^- channels *in vivo*. Therefore, finding new compounds that can bind to the active site of CLIC1 is important to further improve the design of inhibitors that can selectively prevent the soluble to membrane transition of CLIC1. For that reason, owing to the similar monothiol motif at the active site of CLIC1 and hGSTO1-1, the inhibitors of hGSTO1-1 may offer additional insight for the design of selective CLIC1 inhibitors which are suitable for clinical testing. One such compound is 5-chloromethyl fluorescein diacetate, which has been demonstrated to selectively inhibit the enzyme activity of hGSTO1-1 by forming a covalent bond with Cys32 at its active site (Son *et al*., 2010). Therefore, as previously mentioned, given the importance of Cys24 in the enzyme activity, ion channel activity and possibly the *S*-nitrosylation of CLIC1, the interaction between CLIC1 and 5-chloromethyl fluorescein diacetate, together with its structural analogues such as 6-iodoacetamide fluorescein (6-IAF), could provide insight into developing new and effective therapeutics for treating diseases centred around the aberrant expression of CLIC1.

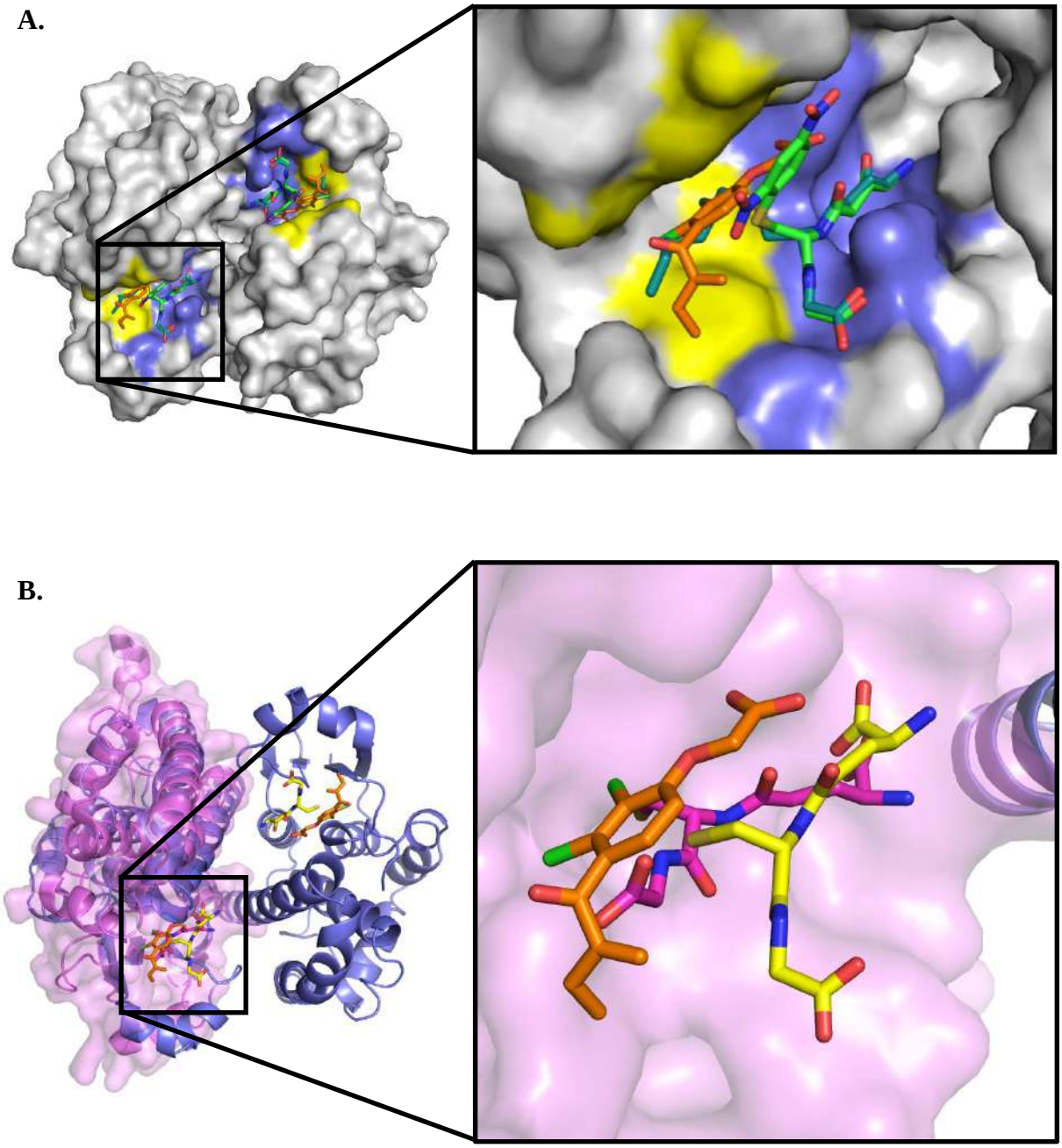


Figure 20. The structural homology between the active site of CLIC1 and hGSTP1-1 could provide insight into developing new inhibitors that can target CLIC1. (A) Present within hGSTP1-1 (PDB: 2GSS) and other cytosolic GSTs are at least two s per a monomer, namely (i) a GSH-binding site (G-site) (●) and (ii) a hydrophobic binding site (H-site) (●). Indicated are S-hexyl GSH (●) (PDB: 1PGT), CDNB-GSH (●) (PDB: 18GS) and EAA (●) (PDB: 2GSS). (B) The G-site from hGSTP1-1 (●) (PDB: 2GSS) can be seen to share structural homology with the active site of CLIC1 (●) (PDB: 1K0M), therefore as to whether CLIC1 can bind EAA remains to be determined. Indicated are EAA (●), GSH from CLIC1 (●) (PDB: 1K0N) and GSH from hGSTP1-1 (●) (PDB: 8GSS). All images were generated by using Pymol V1.3.

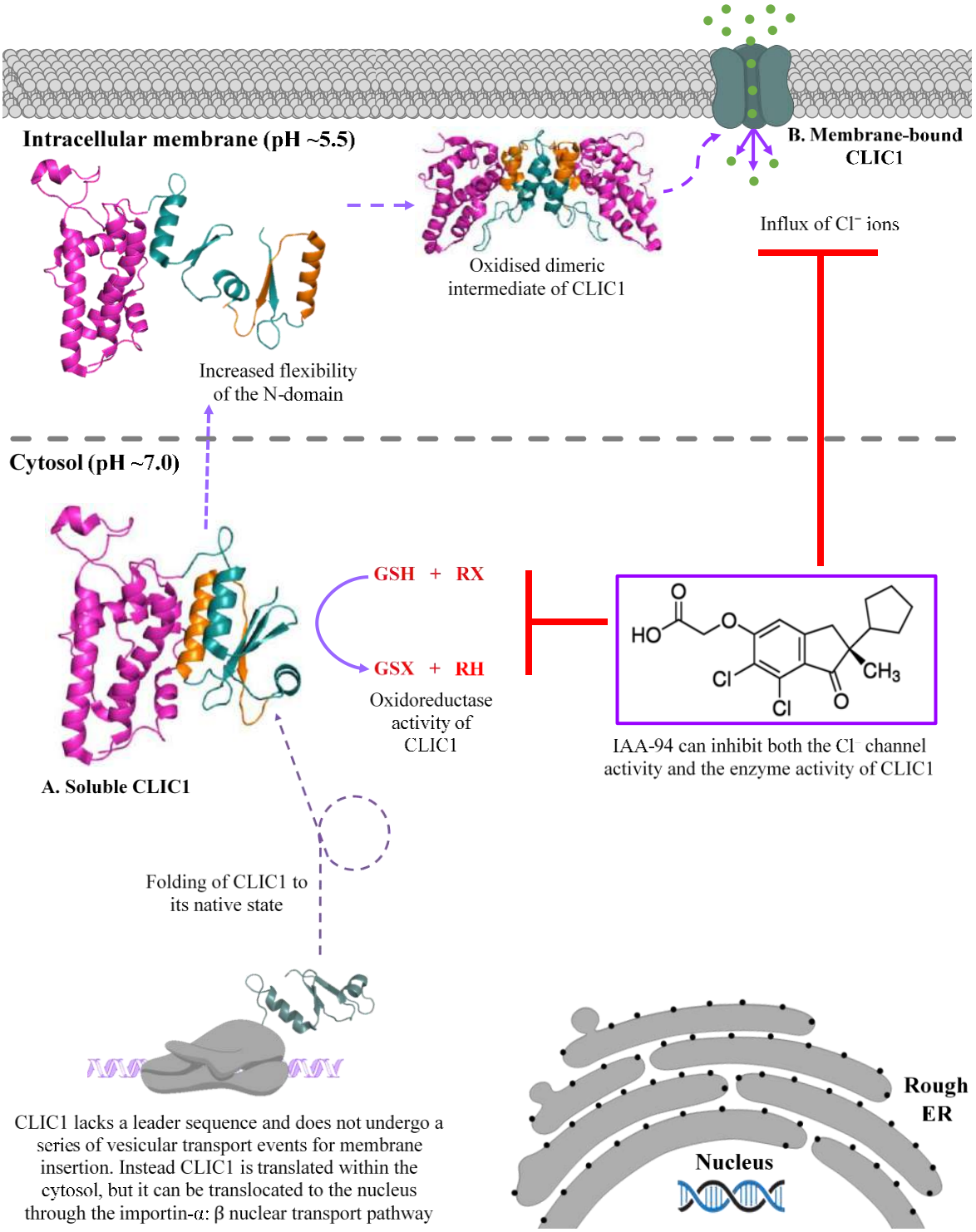


Figure 21. CLIC1 is a metamorphic protein where both of its conformational states are inhibited by IAA-94. Unlike most membrane proteins, CLIC1 is unique seeing that it can reversibly transit between a soluble and integral membrane-bound conformation depending on the conditions of the surrounding environment. However, considering that IAA-94 can inhibit both the enzyme activity and the Cl^- channel activity of CLIC1, it remains uncertain if the constraint in the Cl^- conductance of CLIC1 is due to IAA-94 binding to its soluble conformation to then prevent its structural transition into a membrane competent state, and/or IAA-94 binding to the membrane-bound conformation of CLIC1. The figure was generated by using ChemBioOffice®Ultra 14.0.

1.6 Study rationale: understanding the inhibition and PTM of soluble CLIC1

Membrane proteins continue to be one of the most challenging biological systems to study within structural biology, yet seeing that they represent the targets for up to 70% of current pharmaceutical drugs, understanding the structure and function of membrane proteins is imperative for future drug design. Unlike most membrane proteins which are exclusively associated with the plasma membrane, the CLIC1 protein is unique in that it can reversibly transit between a soluble and membrane-bound conformation depending on the conditions of the surrounding environment. However, considering that both the enzyme and ion channel activity of CLIC1 are inhibited by IAA-94, it remains uncertain if the constraint in the Cl^- conductance of CLIC1 is due to IAA-94 binding to its soluble conformation to then prevent its structural transition into a membrane competent state and/or IAA-94 binding to the membrane-bound conformation of CLIC1. However, seeing that soluble CLIC1 is a structural homolog of hGSTP1-1 and EcSspA, the latter also being a metamorphic protein, it remains uncertain if IAA-94 could also bind to these structural homologs. Similarly, it remains uncertain if EAA, an inhibitor for hGSTP1-1 as well as a structural analogue of IAA-94, can interact with CLIC1 and EcSspA.

However, native to the active site of CLIC1 is not only its catalytically reactive Cys24 residue, but also the TMD of CLIC1. In order for CLIC1 to form its membrane-bound conformation, the single TMD would need to unfold, insert into the membrane and then re-fold to form a functional Cl^- channel. Yet, seeing that a change in environmental conditions primes the structural changes required for the soluble to membrane transition of CLIC1, a change in the redox state of the surrounding environment could influence the reactivity of Cys24 to not only affect it enzyme activity, but also its structural stability. However, present within CLIC1 are six cysteine residues and while Cys24 clearly plays a critical role towards the function of CLIC1, it remains uncertain if this is the only reactive cysteine residue within CLIC1. The importance in studying these reactive cysteine residues stems from the fact that certain RONS such as GSNO are prominent in various disease states, for instance AD where the synthesis of GSNO is linked with the concomitant influx of Cl^- by CLIC1. Therefore, this study aims to provide context not only to the protein-ligand interactions for CLIC1 and its structural homologs, but also to PTMs such as *S*-nitrosylation which could influence both the structure and function of CLIC1.

1.7 Aims and Objectives

1.7.1 Aims

1. To determine the thermodynamic parameters and possible structure for the interactions between IAA-94 or EAA and soluble CLIC1-WT, hGSTP1-1 and soluble EcSspA
2. To determine the main reactive cysteine residues within soluble CLIC1, how they could influence the dynamics of CLIC1 and if they are a target site for *S*-nitrosylation.

1.7.2 Objectives

1. Overexpress and purify CLIC1-WT, CLIC1-C24A, hGSTP1-1 and EcSspA.
2. Assess the purity of all the target proteins by using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).
3. Assess the oligomeric state and purity of all the target proteins by using size exclusion-high pressure liquid chromatography (SE-HPLC).
4. Assess the secondary and tertiary structure of all the target proteins by using far-UV circular dichrosim (CD) spectropolarimetry and intrinsic tryptophan fluorescence, respectively.
5. Assess if IAA-94 and EAA inhibit the enzyme activity of hGSTP1-1 by using the GSH-CDNB conjugation assay.
6. Assess the displacement of ANS by IAA-94 and EAA from hGSTP1-1 by using extrinsic ANS fluorescence.
7. Assess the thermodynamics parameters for the interaction between IAA-94 or EAA and the target proteins by using isothermal titration calorimetry (ITC).
8. Crystallise and solve the structure of hGSTP1-1 in complex with IAA-94 by using hang-drop vapour diffusion and X-ray crystallography, respectively.
9. Assess the main reactive cysteine residue/s within CLIC1-WT and CLIC1-C24A by monitoring the interaction between 6-IAF and both the target proteins by using ITC.
10. Assess the thermal stability and tertiary structure of CLIC-WT and CLIC1-C24A by using far-UV CD and near-UV spectropolarimetry, respectively.
11. Asses the dynamics of CLIC1-WT and CLIC1-C24A by using molecular dynamic (MD) simulations.
12. Assess the stoichiometry and target site for *S*-nitrosylation in CLIC1-WT by using UV-Vis absorbance spectroscopy and liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS), respectively.

Chapter 2: Materials and Methods

2.1 Materials

A list of all the non-standard reagents used throughout this study can be found in the supplementary material (Table S1)

2.2 Methods

2.2.1 Preparation and transformation of competent T7 Express *E.coli* cells

2.2.1.1 Preparation of competent T7 Express *E.coli* cells

Initially, 50 µl of T7 Express *E.coli* cells from a frozen glycerol stock (50% (v/v) cell culture and 50% (v/v) glycerol (80% (v/v)) was first spread across an agar plate (1.5% (w/v) agar, 1.6% (w/v) tryptone, 1% (w/v) yeast extract and 0.5% (w/v) sodium chloride) containing 35 µg.ml^{-1} chloramphenicol. Then without flaming the plate spreader, the spreader was spread across five additional plates in order to dilute the number of cells from the glycerol stock. The plates, including one blank plate that functioned as a negative control to ensure that the plates and/or agar were not contaminated, were incubated for 12-16 h at 37 °C. A single bacterial colony was then selected and grown in 100 ml of 2x yeast extract/tryptone (YT) media (1.6% (w/v) tryptone, 1% (w/v) yeast extract and 0.5% (w/v) sodium chloride) containing 35 µg.ml^{-1} chloramphenicol for 12-16 h in a New Brunswick™ Excella® E24/E24R incubator (Eppendorf, Hamburg, Germany) at 250 rpm.

Thereafter, a 1:20 dilution of the cell culture was done in 100 ml of 2x YT media containing 35 µg.ml^{-1} chloramphenicol and the cells were left to grow at 37 °C in a New Brunswick™ Excella® E24/E24R incubator (Eppendorf, Hamburg, Germany) at 250 rpm until the mid-log growth phase (optical density $(OD)_{600}$ ~0.6) was achieved. The cells were then removed from solution by using a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer Scientific™, Massachusetts, USA) at 5000 x g for 10 min, 4 °C. In order to induce the *E.coli* cells to take up DNA, the permeability of the membrane needed to be increased and the first step to achieving this was to re-suspend the pellet in 10 ml of magnesium chloride (100 mM) followed by leaving the cells on ice for 10 min. Thereafter, the cells were removed from solution by using a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer Scientific™, Massachusetts, USA) at 5000 x g for 10 min, 4 °C. The cell pellet was then re-suspended in 10 ml of calcium chloride (100 mM), placed on ice for 4 h and once again the cells were removed from solution by using a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer

Scientific™, Massachusetts, USA) at 5000 x g for 10 min, 4 °C. The resultant competent cells were then re-suspended in 1 ml of calcium chloride (100 mM) and 1 ml glycerol (80% (v/v)), prepared into 100 µl aliquots and stored at - 80 °C.

2.2.1.2 Transformation of competent T7 Express *E.coli* cells

In this study the complementary deoxyribonucleic acid (cDNA) sequence for CLIC1-WT, CLIC1-C24A and EcSspA were inserted into a pET28-a vector, while the cDNA sequence for hGSTP1-1 was inserted into a pET15-b vector. Initially, 6 µl of the pET28a or pET15b vector (100 ng.µl^{-1}) in TE buffer (10 mM Tris and 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0) was added to a 100 µl stock of competent T7 Express *E.coli* cells and left on ice for 15 min. Thereafter, the cells were heat-shocked at 42 °C for 45-50 s to ensure that the membrane became porous for the uptake of DNA. The cells were then stored on ice for 2 min and thereafter they were transferred to 1 ml of super optimal broth with catabolite repression (SOC) media (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulfate and 20 mM glucose (which was added after being autoclaved to prevent the Millard reaction) and incubated for 1 h at 37 °C in a New Brunswick™ Excella® E24/E24R incubator (Eppendorf, Hamburg, Germany) at 250 rpm.

The cells were then spread across one agar plate containing 35 µg.ml^{-1} chloramphenicol as well 30 µg.ml^{-1} kanamycin or 100 µg.ml^{-1} ampicillin depending if the T7 Express *E.coli cells* were transformed with the pET28a or pET15b vector, respectively. Then without flaming the plate spreader, the spreader was spread across five additional plates in order to dilute the number of cells from the glycerol stock. The plates, including one blank plate that functioned as a negative control to ensure that the plates and/or agar was not contaminated, were incubated for 12-16 h at 37 °C. A single bacterial colony was then selected and grown in 100 ml of 2x YT media containing 35 µg.ml^{-1} chloramphenicol as well 30 µg.ml^{-1} kanamycin (or 100 µg.ml^{-1} ampicillin) for 12-16 h in a New Brunswick™ Excella® E24/E24R incubator (Eppendorf, Hamburg, Germany) at 250 rpm. Thereafter, 500 µl of the resultant cell culture was re-suspended in 1 ml glycerol (80% (v/v)) and stored at - 80 °C.

2.2.2 Heterologous protein overexpression

Expression of a target gene within the pET-11a vector or pET-15b is under the control of the *T7* promoter, but due to a repressor protein encoded by the *lacl* gene within the genomic and plasmid DNA, the T7 promoter, as well as the *lac* promoter required for the expression of T7 RNA polymerase, will be inactive. However, culturing the T7 Express *E.coli* cells in the presence of the lactose analogue isopropyl β-D-1-thiogalactopyranoside (IPTG) inactivates the repressor protein in order to allow the host RNA polymerase to express the T7 RNA polymerase required to bind to the *T7* promoter and consequently induce the overexpression of the target gene (Smith and Johnson, 1988). To confirm that the transformation protocol was successful and that the T7 Express *E.coli* cells therefore had the correct vector, the pET-11a and pET-15b vectors containing the cDNA for the respective target protein were isolated from the T7 Express *E.coli c*ells by using the GeneJet™ plasmid miniprep kit (Thermo Fischer Scientific™, Massachusetts, USA). The target genes were then sequenced by Sanger sequencing at inqaba biotec™ (Pretoria, South Africa) and upon successfully verifying the cDNA sequence, the overexpression of CLIC1-WT, CLIC1-C24A, EcSspA and hGSTP1-1 was performed.

Initially, the T7 Express *E.coli* cells that were transformed with the pET28-a vector that contained the cDNA sequence for either CLIC1-WT, CLIC1-C24A or EcSspA were grown at 37 °C for 12-16 h in 150 ml of 2x YT media that contained 30 µg.ml^{-1} kanamycin. However, for the T7 Express *E.coli* cells that were transformed with the pET15-b vector that contained the cDNA sequence for hGSTP1-1, the media instead contained 100 µg.ml^{-1} ampicillin. Thereafter, a 1:50 dilution of the cell culture was done in twelve 2 L Duran® Erlenmeyer flasks (Sigma Aldrich, Missouri, USA) which contained 500 ml of 2x YT media with 30 µg.ml^{-1} kanamycin (or 100 µg.ml^{-1} ampicillin) and the cells were then left to grow at 37 °C in a LOM-400 orbital shaking incubator (MRC, London, UK) while been shaken at 150 rpm until the mid-log growth phase (OD_{600} ~0.6) was achieved. Thereafter, 0.2 mM of IPTG was added to each flask to induce the heterologous protein overexpression of the target protein for a period 6-8 h, after which the cells were removed from solution by using a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer Scientific™, Massachusetts, USA) at 5000 x g for 10 min, 4 °C. The resultant pellet was then re-suspended in 30 ml of re-suspension buffer (50 mM potassium phosphate, 500 mM sodium chloride, 30 mM imidazole, 1 mM benzamidine, 1 mM phenylmethylsulfonyl fluoride (PMSF) and

0.02% (w/v) sodium azide, pH 8.0) per 1 L of cell culture and the cell lysate was then stored at -20 °C. Importantly, the addition of benzamaidine and PMSF to the cell lysate was always added freshly from frozen aliquots.

2.2.3 Protein purification

Achieving a purified protein product depends upon selectively isolating a protein from other miscellaneous cellular constituents by using an appropriate fractionation procedure with a high resolving power. One unique separation technique implemented in the purification of biomolecules for example is affinity chromatography. The fundamental principle of which relies on the potential for a biomolecule of interest to bind with a high affinity to an immobilised ligand, while allowing all contaminants to be eluted from the chromatography system (Sheehan, 2009).

To achieve this affinity, biomolecules are typically labelled with a peptide (fusion-tag) such as a maltose binding protein (MBP) or a GST fusion-tag. However, an alternative approach to these conventional fusion-tag expression systems is immobilised metal affinity chromatography (IMAC). In this technique metal ions such as ferrous (Fe^{2+}), cobalt (Co^{2+}) or nickel (Ni^{2+}) ions are immobilised onto a stationary phase due to their chelation by groups such as iminodiacetic acid (IDA) or nitrilotriacetic acid (NTA). In this arrangement the nitrogen and oxygen atoms of IDA or NTA will coordinate the metal ions, but still leave sites available to which the metal-binding regions of a protein or fusion-tag can bind. Importantly though, the amino acids which are likely to coordinate metal ions and which are therefore often found within the metal-binding regions of a protein or fusion-tag include histidine, cysteine and tryptophan residues (Block *et al.*, 2009; Sheehan, 2009).

2.2.3.1 Purification of the target proteins by means of IMAC

The pET-28a and pET-15b vectors encode for a hexa-histidine fusion-tag that will be expressed at the N-terminus of the target protein, therefore enabling the fusion-protein to become immobilised onto an IMAC column to ensure its purification. Initially, prior to the cell lysate being thawed in a beaker of water at 20 °C, 50 µl of magnesium chloride (1 M), 50 µl of bovine pancreatic deoxyribonuclease I (DNase I) (10 $mg.ml^{-1}$ in Milli-Q® water (18.2 $M\Omega.cm^{-1}$ at 25 ^{0}C) which was from a Millipore Direct-Q® 3 UV water purification system (MilliporeSigma, Massachusetts, USA)) and 500 µl of chicken egg white lysozyme

(10 mg.ml^{-1} in Milli-Q® water (18.2 MΩ.cm^{-1} at 25 ^{0}C) which was from a Millipore Direct-Q® 3 UV water purification system (MilliporeSigma, Massachusetts, USA)) were added per 30 ml of the cell lysate. Once thawed, the cell lysate was homogenised on ice by using a Bio-Gen series PRO200 homogeniser (Pro Scientific Inc., Connecticut, USA) and thereafter the cell lysate was incubated in the beaker of water for an additional 30 min in order to optimise cell lysis. The soluble fraction of the cell lysate was then obtained by using a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer Scientific™, Massachusetts, USA) at 23 000 x g for 20 min, 4 °C.

The supernatant was then filtered through a 0.45 µM syringe filter (cellulose acetate membrane) and loaded onto a 5 ml HiTrap® IMAC fast flow column (Sigma Aldrich, Missouri, USA) that had been connected to an ÄKTAprime plus™ liquid chromatography purification system (GE Healthcare Life Sciences, Chicago, USA) and pre-equilibrated with at least five column volumes of the IMAC equilibration buffer (50 mM potassium phosphate, 500 mM sodium chloride, 30 mM imidazole and 0.02% (w/v) sodium azide, pH 8.0) at a flow-rate of 5 ml.min^{-1}, 0.3 kPa. Importantly though, depending on whether (i) CLIC1-WT, CLIC1-C24A, EcSspA or (ii) hGSTP1-1 was to be purified, a (i) Ni^{2+} or (ii) Co^{2+}-IMAC column was used, respectively. Then in order to remove the unbound fusion-protein and/or contaminant proteins from the column, the IMAC equilibration buffer was passed through the column until the absorbance at 280 nm decreased to a stable baseline. Thereafter, the fusion-protein was removed from the column in a one-step elution with the IMAC elution buffer (50 mM potassium phosphate, 500 mM sodium chloride, 500 mM imidazole and 0.02% (w/v) sodium azide, pH 8.0) and 3 ml fractions of the fusion-protein were collected.

2.2.3.2 Purification of the target proteins by means of ion-exchange chromatography

Depending on the pH of the surrounding environment, amino acids may either have an overall positive, negative or zero-net charge. Therefore, considering that the surface charge of a protein is dependent upon its structure and amino acid composition, these physical properties prove to be beneficial in the separation of proteins from one another. The purification of proteins based on their charge is typically achieved by ion-exchange chromatography, which is a form of adsorption chromatography whereby small ions that are attached to ion exchangers on the stationary phase will be exchanged with biomolecules of a similar charge (Sheehan, 2009).

In analytical biochemistry, ion-exchangers can be classified as being either anion- or cation-exchangers depending on whether they bind to anions or cations, respectively. Furthermore, depending on the chemistry of their functional groups, ion exchangers may be regarded as being either weak or strong. That is, these terms do not refer to the strength at which a protein binds to the ion exchanger, but rather the pH range at which the functional groups of the ion exchangers will be ionised. Therefore, strong ion exchangers are completely ionised over a wide pH range and as a result proteins can adsorb to several ion exchanger sites, whereas weak ion exchangers on the other hand are only partially ionised over a narrow pH range and therefore this makes them more flexible in terms of their selectivity and also it makes them to be a more general option for the separation of proteins that retain their functionality over a pH range from pH 6-9 (Acikara, 2013; Sheehan, 2009). Though in particular, the purification of proteins by weak ion-exchangers is typically achieved by using either a carboxymethyl (CM) or a diethylaminoethyl (DEAE) column in order to bind and purify proteins with a net positive or negative charge, respectively.

Subsequent to the elution of the target protein from the IMAC column, several protein contaminants were often noted to co-elute with the target protein. However, present between the target protein and the hexa-histidine fusion-tag was a thrombin cleavage site, therefore incubating the fusion-protein in the presence of human plasma thrombin (EC 3.4.21.5) through an off-column cleavage approach ensured that the fusion-tag was cleaved from the target protein. This process was beneficial considering that the target protein can simply pass through the IMAC column, while allowing the histidine fusion-tag and contaminant proteins to rebind to the column. However, in order to ensure the separation of thrombin from the target protein, an additional chromatography technique will need to be implemented and this can often be achieved by coupling the IMAC column to a benzamidine, heparin or ion-exchange chromatography column.

The eluted fractions from the IMAC column for CLIC1-WT and CLIC1-C24A were pooled together, placed in SnakeSkin™ dialysis tubing (Thermo Fisher Scientific™, Massachusetts, USA) with a molecular weight cut-off (MWCO) of 10 kDa and dialysed against the DEAE equilibration buffer (50 mM potassium phosphate, 30 mM sodium chloride and 0.02% (w/v) sodium azide, pH 6.5) for 12-16 h at 4 °C. However, the eluted fractions for hGSTP1-1 and EcSspA were instead dialysed against the benzamidine equilibration buffer (50 mM potassium phosphate, 500 mM sodium chloride and 0.02% (w/v) sodium azide,

pH 8.0) for 12-16 h at 4 °C. The hexa-histidine-tag was then removed from the target protein by adding 1 µl of human plasma thrombin (1 U.µl^{-1} in 50 mM sodium citrate, 200 mM sodium chloride, 0.1% (w/v) polyethylene glycol (PEG)-6000, 50% (v/v) glycerol, pH 6.5) per a millilitre of protein sample and allowing thrombin cleavage to occur for 4-6 h at 20 °C. Although, calcium precipitates out of solution when in the presence of phosphate ions, the addition of calcium chloride (1 M) to a final concentration of 2.5 mM was required to induce thrombin cleavage of the hexa-histidine tag from hGSTP1-1 and EcSspA.

Subsequent to thrombin cleavage, a 5 ml HiTrap® IMAC fast flow column (Sigma Aldrich, Missouri, USA), which had been coupled to either a 5 ml HiTrap® DEAE or benzamidine fast flow column (Sigma Aldrich, Missouri, USA), was connected to an ÄKTAprime plus™ liquid chromatography purification system (GE Healthcare Life Sciences, Chicago, USA) and pre-equilibrated with at least five column volumes of either the DEAE or benzamidine equilibration buffer at a flow-rate of 5 ml.min^{-1}, 0.3 kPa. Thereafter, the fusion-tag, contaminant proteins and thrombin were removed from the target protein by re-applying the protein sample to the appropriate coupled IMAC column at a flow-rate of 5 ml.min^{-1}. However, with respect to CLIC1, its separation from thrombin relied on the fact that CLIC1-WT/C24A and thrombin have a p*I* of ~5.0 and ~8.0, respectively, as predicted by the ExPaSy ProtPARAM tool (Gasteiger *et al*., 2005) (https://web.expasy.org/protparam/). Therefore, at a pH of 6.5, CLIC1-WT/C24A and thrombin would have a net negative and positive charge, respectively and as a result CLIC1-WT/C24A would bind to the DEAE column, whereas thrombin would simply pass through to aid in its separation from CLIC1. The bound protein was then removed from the DEAE column in a one-step elution with the DEAE elution buffer (50 mM potassium phosphate, 500 mM sodium chloride and 0.02% (w/v) sodium azide, pH 8.0) and the target protein was collected in 3 ml fractions. However, with respect to hGSTP1-1 and EcSspA, both the target proteins did not bind to the benzamidine column and instead only thrombin became bound to the column, therefore aiding the separation of thrombin from both the target proteins.

The regeneration of the DEAE and IMAC columns was then achieved by the respective elution buffer used for each column and thereafter the columns were stored in the storage solution (20% (v/v) ethanol). However, with respect to the benzamidine column, thrombin was eluted from the column with the benzamidine elution buffer (50 mM glycine, 500 mM sodium

chloride and 0.02% (w/v) sodium azide, pH 3.0) in order to regenerate the column and thereafter the column was stored in the storage buffer (50 mM sodium acetate and 20% (v/v) ethanol, pH 4.0).

2.2.3.3 Concentration and storage of the purified target proteins

The purity of the collected fractions were then evaluated by SDS-PAGE (section 2.2.4) and those fractions which contained pure protein were pooled together and concentrated to a final concentration of ≥ 120 µM by using an Amicon® stirred cell (MilliporeSigma, Massachusetts, USA) that was fitted with a 10 kDa ultrafiltration membrane (MilliporeSigma, Massachusetts, USA). The concentrated target protein was then placed in SnakeSkin™ dialysis tubing (Thermo Fisher Scientific™, Massachusetts, USA) with a MWCO of 10 kDa and dialysed against the sample buffer (50 mM potassium phosphate, 1.5 mM EDTA and 0.02% (w/v) sodium azide, pH 7.0) at 4 °C. In particular, the protein sample was first dialysed against 2 L of the storage buffer for 12-16 h, followed by 1 L of the storage buffer for 3-4 h and then 2 L of the storage buffer for 3-4 h. However, following the last dialysis step the protein sample was then dialysed against 500 ml of the sample buffer containing 3 mM Tris (2-carboxyethyl) phosphine-HCl (TCEP) for 12-16 h. Importantly, although TCEP is a more useful reductant over a wider pH range (1.5-8.5) in comparison to dithiothreitol (DTT), the presence of phosphates and an increase temperature can adversely influence the stability of TCEP (Getz *et al*., 1999). Therefore, the dialysed protein sample and sample buffer were stored at 4 °C and used as soon as possible for downstream assays.

2.2.4 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

The implementation of SDS-PAGE has become an indispensable tool in the characterisation of protein species based on their migration and subsequent separation in the presence of an electric field. However, various proteins, although having a similar molecular weight, retain different native conformations that would otherwise result in their migration occurring at a dissimilar rate. To overcome this corollary, the anionic detergent SDS is used to disrupt the secondary and non-disulfide linked tertiary structure of proteins considering that it binds to both polar and hydrophobic regions with a uniform ratio of 1.4 g SDS. g protein^{-1}. Furthermore, considering its rod-like shape and negative charge, SDS aids to mask the intrinsic charge of a protein and provide all treated proteins with a uniform charge: mass ratio in order to ensure that the distance migrated by a protein is a function of its molecular weight (Laemmli, 1970; Wilson and Walker, 2010).

The purity of the purified target proteins were verified on a 12.5% (w/v) separating gel (41.6% (v/v) acrylamide/bis-acrylamide (30% (w/v)/0.8% (w/v)), 26.6% (v/v) Tris (1.5 M, pH 8.8), 1% (v/v) SDS (10% (w/v)), 0.2% (v/v) tetramethylethylenediamine (TEMED) and 0.7% (v/v) ammonium persulfate (APS) (10% (w/v)) that was overlaid with a 4% (w/v) stacking gel (13.5% (v/v) acrylamide/bis-acrylamide (30% (w/v)/0.8% (w/v)), 20.1% (v/v) Tris (0.5 M, pH 6.8), 1% (v/v) SDS (10% (w/v)), 0.2% (v/v) TEMED and 0.7% (v/v) APS (10% (w/v)). Importantly though, while the separating gel was left to polymerize at 20 °C in a 1 mm space plate by using a Mini-PROTEAN® Tetra hand cast system (Biorad, California, USA), a thin layer of isopropanol (98% (v/v)) was added on top of the gel to ensure that the gel was level.

All protein samples were mixed with the SDS-PAGE sample buffer (12.5% (v/v) Tris (0.5 M, pH 6.8), 20% (v/v) SDS (10 % (w/v)), 25% (v/v) glycerol (100 % (v/v)) and 2% (v/v) bromophenol blue (0.5% (w/v)) in a 1:2 ratio. However, with respect to the pellet from the cell lysate, before taking a sample for analysis it was important to first dilute the sample to its original volume prior to centrifugation. Thereafter, the protein samples were incubated at 95 °C for 5 min prior to being loaded on the gel, however, for viscus samples such as the cell lysate and pellet, they were additionally sonicated by using the Microson XL-2000® sonicator for 20-30 s on the highest setting. Subsequently, 2 µl of the protein molecular weight marker was loaded in the first well and 10 µl of the protein samples were loaded in the remaining wells of the stacking gel. The electrophoretic separation of the samples was then carried out at 160 V in the electrophoretic buffer (25 mM Tris, 192 mM glycine and 1% (w/v) SDS, pH 8.3 (not adjusted)) by using Mini-PROTEAN® Tetra vertical electrophoresis cell system (Biorad, California, USA).

The gel was then stained in the staining solution (0.25% (w/v) Coomassie Brilliant Blue R250, 10% (v/v) glacial acetic acid and 40% (v/v) methanol) for a period of 2-16 h and afterwards the gel was destained in destaining solution (10% (v/v) glacial acetic acid and 40% (v/v) methanol) until the background was clear. Images of the gel were then taken by using the Molecular Imager® Gel Doc™ XR+ System (Biorad, California, USA) and the data was processed by using the Image Lab™ Software (Biorad, California, USA). Thereafter, the relative migration distance (R_f) of the protein standards within the molecular weight marker were determined as follows:

$$R_f = \frac{\text{Migration distance of the protein}}{\text{Migration distance of the dye front}} \quad \textbf{(1)}$$

Therefore, by plotting the logarithm of the known molecular weight for the protein standards as of function of their R_f, the molecular weight for the target proteins could be determined from the linear regression model fitted to the data based on their distance migrated.

2.2.5 Protein concentration determination

Peptide bonds, aromatic amino acid residues and to a small extent disulfide bonds account for the ability of proteins to absorb UV light, thus aiding the concentration of proteins to be determined in solution. The protein concentration for the target proteins were determined in accordance to the Beer-Lambert Law as follows (Wilson and Walker, 2010):

$$A = \varepsilon c l \quad \textbf{(2)}$$

where A represents the absorbance (arbitrary units (AU)) by the molecule of interest, ε represents the molar extinction coefficient ($M^{-1}.cm^{-1}$) for the molecule of interest at a particular wavelength (for protein λ = 280 nm), c represents the concentration (M) for the molecule of interest and l represents the path length of the cuvette (cm). The ε_{280} used for the target proteins in this study was calculated by taking into account the number of absorbing chromophores and their respective extinction coefficients as follows (Perkins, 1986):

ε_{280} (CLIC1-WT) = 5550 ΣTrp + 1340 ΣTyr + 150 ΣCys **(3)**

= 5550 (1) + 1340 (8) + 150 (6)

= 17 170 M^{-1} cm^{-1}

ε_{280} (CLIC1-C24A) = 5550 ΣTrp + 1340 ΣTyr + 150 ΣCys

= 5550 (1) + 1340 (8) + 150 (5)

= 17 020 M^{-1} cm^{-1}

ε_{280} (hGSTP1-1) = 5550 ΣTrp + 1340 ΣTyr + 150 ΣCys

= 5550 (2) + 1340 (12) + 150 (4)

= 22 780 M^{-1} cm^{-1} (per a monomer)

ε_{280} (EcSspA) = 5550 ΣTrp + 1340 ΣTyr + 150 ΣCys

= 5550 (3) + 1340 (8) + 150 (1)

= 27 520 M^{-1} cm^{-1} (per a monomer)

An absorbance spectra for each target protein was obtained by mixing, in triplicate, 50 µl of the protein sample with 450 µl of the sample buffer and then measuring the absorbance of each sample in a Suprasil® quartz cuvette with a path length of 1 cm (Hellma®, Müllheim, Germany) by using a Jasco V-630 UV-Vis absorbance spectrophotometer (Jasco, Pfungstadt, Germany). All the spectra were measured at 20 °C across a wavelength range of 240-400 nm, data pitch of 1 nm and a scan speed of 200 $nm.min^{-1}$. The final protein concentration was then determined by correcting for contributions that arose due to the absorbance of the buffer and light scatter from protein aggregates as follows (McIntyre, 2008):

$$\text{Corrected } A_{280} = A_{280}\text{Protein} - A_{280}\text{Buffer} - \frac{[(A_{340}\text{Protein} - A_{340}\text{Buffer}) \cdot A_{280}\text{Buffer}]}{A_{340}\text{Buffer}} \quad \textbf{(4)}$$

2.2.6 Size exclusion-high pressure liquid chromatography (SE-HPLC)

Size exclusion chromatography (SEC), by depending on the molecular sieving properties of porous materials, is able to separate molecules based on their molecular size (Sheehan, 2009). A SEC column comprises a gel of beads with pores that have a defined and narrow size range, consequently molecules within the fraction range of the gel will be retained in a manner that is inversely proportional to their hydrodynamic volume. That is, molecules with a small size will be retained the longest, whereas molecules with a size exceeding the exclusion limit of the gel will pass freely though the column and appear to elute first. In the case of a SE-HPLC system, the use of high-pressure pumps aids the mobile phase in the SEC column to move at a high flow-rate, thus ensuring rapid separation, minimised band broadening and optimal resolution (Sheehan, 2009; Wilson and Walker, 2010).

Analysis of the hydrodynamic volume and consequently the oligomeric state and purity of each target protein by means of SE-HPLC was carried out on a TSKgel SuperSW2000 SEC column that had a resolution range of 5–150 kDa (Tosoh Corporation©, Tokyo, Japan). The SEC column was additionally conjugated with a TSKgel SW_{XL} guard column (Tosoh Corporation©, Tokyo, Japan) that were collectively connected to a Shimadzu ultrafiltration liquid chromatography (UFLC) SPD-20A SE-HPLC system. The SEC column was

equilibrated at 20 °C with the SE-HPLC buffer (50 mM potassium phosphate, 150 mM sodium chloride, 1.5 mM EDTA and 0.02% (w/v) sodium azide, pH 6.5) at a flow rate of 0.2 ml.min^{-1} and isobaric pressure of ~900 psi. Thereafter, in triplicate, 20 µl of the target protein (100 µM) was injected onto the column and allowed to pass through the column with a separation period of 30 min. Then by using a set of protein standards with a known molecular weight (thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B12 (1.35 kDa)), a standard curve was constructed by plotting the logarithm of the known molecular weight for the protein standards as a function of their retention time so that the molecular weight for the target proteins could be determined based on their retention time.

2.2.7 Secondary and tertiary structural characterisation

2.2.7.1 Far-UV circular dichroism (CD) spectropolarimetry

CD describes the phenomenon that when circularly polarised light at a specific wavelength passes through a chiral chromophore such as an amino acid, the left and right components of light will be differentially absorbed. In particular, the CD spectra of a protein in the far-UV region is dominated by the n→π* and π→π* transitions of peptide bonds, which can be seen to be centred around 222 nm and 190 nm, respectively. Importantly, these spectra are influenced by the geometries of the polypeptide backbone and will therefore reflect the various phi (Φ) and psi (Ψ) angles of the peptide bond to subsequently indicate the secondary structural elements present within a protein such as α-helices, β-sheets and random coils (Kelly *et al*., 2005; Whitmore *et al*., 2007). In contrast to β-sheet structural motifs which display a minimum and maximum ellipticity at 218 nm and 195 nm, respectively, proteins which are predominately α-helical, such as CLIC1 and its structural homologs, will display a trough at 208 nm and 222 nm along with a peak at 190 nm (Kelly *et al*., 2005).

The far-UV CD spectra were recorded, in triplicate, for 2 µM of the native or denatured (in the presence of 8 M urea) target protein in a Suprasil® quartz cuvette with a path length of 0.2 cm (Hellma®, Müllheim, Germany) by using the Jasco J-1500 spectropolarimeter that was connected to a Peltier temperature controller (Jasco, Pfungstadt, Germany). All spectra were measured at 20 °C in a continuous scanning mode over a wavelength range of 180-250 nm, bandwidth of 1 nm, data pitch of 0.5 nm and a scan speed of 200 nm.min^{-1}. The spectra were an average of ten accumulative scans, which then had the contributions from the buffer

subtracted from all the data collected as well as the resultant CD ellipticity signal converted to the mean residue ellipticity [Θ] (deg.cm².dmol⁻¹) as follows (Woody, 1995):

$$[\theta] = \frac{100.\theta}{c.n.l} \quad \textbf{(5)}$$

where θ is the measured ellipticity in millidegrees (mdeg) at a particular wavelength, c is the sample concentration (mM), n is the number of residues in the target protein and l is the path length (cm) of the cuvette used. The number of residues used for the native and biological unit of CLIC1-WT, CLIC1-C24A, EcSspA and hGSTP1-1 were 241, 241, 420 and 424, respectively.

2.2.7.2 Near-UV CD spectropolarimetry

In contrast to far-UV CD, the spectra from near-UV CD (250-290 nm) are dictated by the absorption from the side chains of phenylalanine, tyrosine and tryptophan. However, disulfide bonds and various co-factors may also absorb in this region and contribute to the observed signal in the near-UV CD spectra. Nevertheless, in the native conformation, the polypeptide chain can place the side chains of these aromatic residues in a chiral environment and seeing that the surrounding environment of these side chains has an influence on the near-UV CD spectra, similar to intrinsic tryptophan fluorescence, near-UV CD can therefore provide information about the tertiary, or otherwise native conformation of a protein (Kelly *et al*., 2005, Woody, 1995).

Importantly, each of these aromatic residues have a characteristic wavelength profile, namely (i) tryptophan has a peak close to 290 nm with fine structures seen between 290 nm and 305 nm, (ii) tyrosine has a peak between 275 and 282 nm, but any fine structure at longer wavelengths will be attenuated by the signal from tryptophan and (iii) phenylalanine has sharp fine structure between 255 nm and 270 nm. These fine structure in the bands arise from vibronic transitions in which different vibrational levels of the excited state are involved (Strickland, 1974). However, owing to the lower abundance and therefore concentration of these aromatic residues relative to the peptide bonds found in the protein backbone, the CD signal in the near-UV region is weaker than those in the far-UV region (Kelly *et al*., 2005). Furthermore, at room temperature individual features often meld into a spectral ensemble, yet when CD spectra are recorded at a low temperatures the information obtained can be more informative. As a

result, when measuring near-UV CD spectra, by lowering the temperature and in turn the dynamics of a protein, the signal-to-noise ratio would improve for the fine structural elements seen in the near-UV CD spectra (Gasymov *et al*., 2014; Kelly *et al*., 2005). Consequently, although depending on the composition of the primary sequence, in order to enhance the signal in near-UV CD spectra a higher protein concentration and/or longer path length should be used (Kelly *et al*., 2005).

The near-UV CD spectra were recorded for 40 µM of CLIC1-WT and CLIC1-C24A in a Suprasil® quartz cuvette with a path length of 0.5 cm (Hellma®, Müllheim, Germany) by using the Jasco J-1500 spectropolarimeter that was connected to a Peltier temperature controller (Jasco, Pfungstadt, Germany). All spectra were measured at 5 °C in a continuous scanning mode over a wavelength range of 250-360 nm, bandwidth of 0.5 nm, data pitch of 0.5 nm and a scan speed of 100 $nm.min^{-1}$. The spectra were an average of ten accumulative scans, which then had the contributions from the buffer subtracted from all the data collected as well as the resultant CD ellipticity signal converted to the [Θ] as described in section 2.2.7.1.

2.2.7.3 Monitoring the thermal stability of a protein by far-UV CD spectropolarimetry

The physiological functions initiated by proteins require that the nascent amino acid chain correctly folds into the three-dimensional structure of its biologically active native state. However, when subject to environmental changes such as an increase in temperature, change in pH, change in ionic strength or an increase in the concentration of chemical denaturants, the equilibrium between the native state of a protein and its unfolded state can be shifted to the latter. In contrast to its native conformation, the denatured state of a protein is thought to comprise a randomly coiled polypeptide chain given the loss in its secondary, tertiary and quaternary structure. Experimentally, however, the complete denaturation of a protein rarely occurs, but instead some residual structures are often noted and consequently a protein is considered to be unfolded as opposed to denatured. Therefore, the conformational stability of a protein can be described as its ability to remain in its native state, which thermodynamically represents the change in the Gibbs free energy (ΔG) between its native and unfolded state (Dobson, 2004; Dobson and Karplus, 1999).

Therefore, bearing in mind that hydrogen bonds, salt bridges, van der Waals forces and disulfide bonds are hallmark features in the stability of a protein, comparing the unfolding

profile between wild-type and variant proteins can provide insight as to how ligands, mutations or PTMs influence the stability of a protein (England and Haran, 2011). However, in contrast to the denaturation of proteins by chaotropic agents such as urea, the thermal unfolding of a protein is often irreversible given the formation of protein aggregates as a result of the hydrophobic interactions between neighbouring molecules. Therefore, the profile obtained from the thermal unfolding of a protein is merely qualitative, that is, the only insight into the conformational stability of a protein is through monitoring a shift in the thermal unfolding temperature (T_{unfol}), that is, the temperature which corresponds to an initial loss in the secondary structural content of a protein (Rocco *et al*., 2008).

The thermal stability of CLIC1-WT and CLIC1-C24A in sample buffer was monitored by far-UV CD spectropolarimetry by measuring a change in the CD ellipticity signal at 222 nm when 2 µM of protein was unfolded in a Suprasil® quartz cuvette with a path length of 0.2 cm (Hellma®, Müllheim, Germany). The unfolding of the protein was monitored by using the Jasco J-1500 spectropolarimeter that was connected to a Peltier temperature controller (Jasco, Pfungstadt, Germany) and the thermal unfolding profile was obtained by increasing the temperature from 10 °C to 80 °C at a rate of 1 °C.min^{-1}. All measurements were obtained with a bandwidth of 1 nm, data pitch of 0.5 °C and a data integration time (DIT) of 1 s. The unfolding of CLIC1-WT and CLIC1-C24A was not reversible and therefore it was not necessary to measure the change in the CD ellipticity signal at 222 nm when lowering the temperature. However, a far-UV CD spectra was performed pre- and post-heating of the protein sample as described in section 2.2.7.1. The unfolded data was then normalised by taking the fraction of unfolded protein as follows:

$$F_u = \frac{(\theta - \theta_{unfold})}{(\theta_{fold} - \theta_{unfold})} \quad \textbf{(6)}$$

where F_u represents the fraction of protein unfolded, θ represents the ellipticity at a particular temperature point, θ_{fold} represents the ellipticity for the folded protein and θ_{unfold} represents the ellipticity for the unfolded protein. The data was then fitted to a sigmoid, three parameter model in SigmaPlot® (SYSTAT, University of Illinois at Chicago, USA) and the data was plotted as the fraction of protein unfolded versus temperature. Importantly, considering that the thermal unfolding of CLIC1-WT and CLIC1-C24A was irreversible and that no thermodynamic parameters could be obtained, the data generated was used to qualitatively monitor the T_{unfol} for CLIC1-WT and its C24A mutant.

2.2.7.4 Intrinsic tryptophan fluorescence

The absorbance of light by a chromophore will cause its electrons to be raised to a higher energy state, however, upon returning to an energy minimised state, the energy that is lost by the electrons at a longer wavelength will be in a light emitting manner to induce the phenomenon of fluorescence (Sheehan, 2009). With respect to protein molecules, the fluorescence properties of tyrosine and tryptophan residues, owing to their aromaticity, aid as suitable intrinsic probes to verify the tertiary integrity of proteins (Lakowicz, 2006).

While tyrosine residues are optimally excited by light at 280 nm and tryptophan residues are selectively excited by light at 295 nm, considering that the energy emitted by tyrosine overlaps the excitation energy for tryptophan as a result of FRET, when proteins are exposed to light at 280 nm, the majority of the intrinsic protein fluorescence would be attributed to tryptophan residues. This is beneficial considering that the emission wavelength from the indole functional group of tryptophan is sensitive to the polarity of its surrounding environment and therefore, depending on where they are located, tryptophan residues function as a probe in order to monitor the local and/or global tertiary structural changes of a protein (Lakowicz, 2006). For instance, in proteins such as CLIC1 which contain a single conserved tryptophan (Trp35) residue, but also eight tyrosine residues distributed throughout the protein, exciting the protein at 280 nm will not only enhance the intensity of the emission spectra, but it while still remain dependent upon the polarity of the surrounding environment of Trp35 in order to provide insight into the tertiary structure of the protein.

The intrinsic fluorescence emission spectra were recorded, in triplicate, for 2 µM of the native or denatured (in the presence of 8 M urea) target protein in a Suprasil® quartz cuvette with a path length of 1 cm (Hellma®, Müllheim, Germany) by using the Jasco FP-6300 fluorimeter that was connected to a Peltier temperature controller (Jasco, Pfungstadt, Germany). The intrinsic fluorescence was monitored at 20 °C by using an excitation wavelength of 280 nm and all spectra were measured in a continuous scanning mode over a wavelength range of 280-500 nm, data pitch of 1 nm, scan speed of 200 $nm.min^{-1}$ and an excitation and emission bandwidth of 2.5 nm and 5.0 nm, respectively. The spectra were then corrected by subtracting the contributions from the buffer from all the data collected.

2.2.7.5 Extrinsic fluorescence: ANS displacement studies

While fluorescence spectroscopy often relies on the intrinsic fluorescence properties of proteins to aid in their structural characterisation, the use of extrinsic fluorescent dyes (fluorophores) can offer additional insight into the structural characterisation of proteins. Similar to tryptophan, a valuable property of many fluorophores is their sensitivity to the surrounding environment and therefore they can be used as probes to try and determine their location on a protein. Therefore, the use of extrinsic dyes has been used to detect molten globule intermediates during unfolding events, assess surface hydrophobicitiy and probe the active sites of enzymes (Hawe *et al*., 2008). For instance, one of the most commonly used extrinsic fluorophores used to probe hydrophobic surfaces in proteins is the compound 8-anilinonaphthalene-1-sulfonic acid (ANS). In an aqueous environment ANS has a low quantum yield, however, when it binds non-covalently within a non-polar environment of a protein its fluorescence will undergo a blue-shift along with an increase in intensity (Stryer, 1965; Uversky *et al*., 1998). Consequently, in proteins such as hGSTP1-1 which are known to bind a range of hydrophobic ligands, owing to the amphiphilic nature of ANS, the competitive or non-competitive displacement of ANS by hydrophobic ligands has been used to probe various protein-ligand interactions.

A 2 mM ANS ammonium salt stock solution was prepared in the sample buffer (section 2.2.3.3), however, the concentration of the ANS stock solution was confirmed by mixing, in triplicate, 50 µl of the ANS stock solution with 1 950 µl of the sample buffer and then measuring the absorbance of each sample across a wavelength range of 250-500 nm as described in section 2.2.5. The concentration of the ANS stock solution was then determined by using the Beer-Lambert Law (equation 2) (ε_{350} = 4950 $M^{-1}.cm^{-1}$) (Weber and Young, 1964). Thereafter, in triplicate, ANS was added to a final concentration of 100 µM ANS to samples that contained 2 µM of CLIC1-WT, hGSTP1-1 or EcSspA and the sample was left to incubate in the dark at 20 °C for a period of 30 min to ensure complete binding of ANS to the target protein. The extrinsic fluorescence emission spectra of ANS was then recorded in a Suprasil® quartz cuvette with a path length of 1 cm (Hellma®, Müllheim, Germany) by using the Jasco FP-6300 fluorimeter that was connected to a Peltier temperature controller (Jasco, Pfungstadt, Germany). The extrinsic fluorescence was monitored at 20 °C by using an excitation wavelength of 390 nm and all spectra were measured in a continuous scanning mode over a wavelength range of 380-700 nm, data pitch of 1 nm, scan speed of 200 $nm.min^{-1}$ and an excitation and emission bandwidth of 2.5 nm and 5.0 nm, respectively. The spectra were then

corrected by subtracting the contributions from the buffer and free ANS from all the data collected.

Thereafter, in order to monitor the potential displacement of ANS from hGSTP1-1 by IAA-94 and EAA, samples of hGSTP1-1 in the presence of 100 µM ANS were incubated with 2.1 mM IAA-94 and 2.5 mM EAA for a period of 30 min in the dark at 20 °C to ensure complete binding or displacement of ANS from hGSTP1-1. However, based on the concentration of IAA-94 and EAA that was used for the displacement ANS, the final concentration of dimethylsulfoxide (DMSO) in each sample was 3% (v/v) considering that IAA-94 and EAA were solubilised in DMSO. Therefore, the samples of free ANS and hGSTP1-1 with only ANS also contained DMSO to a final concentration of 3% (v/v). The extrinsic fluorescence for each sample was then monitored by using the same parameters stated above and the spectra were then corrected by subtracting the contributions from the buffer, ligand, DMSO and free ANS from all the data collected. Furthermore, the degree of ANS displacement was quantified as follows:

$$\text{ANS displacement (\%)} = \frac{FI_{max}\,(\text{Protein + ligand}) - FI_{max}\,(\text{Protein + solvent})}{FI_{max}\,(\text{ANS + ligand}) - FI_{max}\,(\text{Protein + solvent})} \times 100\% \quad \textbf{(7)}$$

where FI_{max} (Protein + ligand) represents the maximum fluorescence intensity of ANS in the presence of the protein and ligand, FI_{max} (Protein + solvent) represents the maximum fluorescence intensity of ANS in the presence of the protein and the solvent of the ligand and FI_{max} (ANS + ligand) represents the maximum fluorescence intensity of free ANS in the presence of the ligand.

2.2.8 Colorimetric-based enzyme activity assay

Monitoring a reaction that is catalysed by an enzyme would involve observing a change in the concentration of either the substrate or product overtime. Seeing that the substrate and/or product often have unique spectroscopic properties, it is possible to monitor the change in their concentration over time in order to determine the enzyme activity and consequently the underlying enzyme kinetics for a particular reaction. For instance, during an enzyme reaction the structural change in the substrate often results in the product containing a chromophore with a distinct absorbance spectra that can be monitored by visible absorbance spectroscopy. Therefore, if the ε of the product is known, it is possible, by means of the Beer-Lambert Law, to directly calculate the change in the concentration of the product overtime based on the

absorbance readings at a specific wavelength (Sheehan, 2009; Wilson and Walker, 2010). For instance, the enzyme activity of hGSTP1-1 can be monitored through its ability to catalyse the conjugation of GSH with its substrate CDNB in order to form the product 1-(S-glutathionyl)-2, 4-dinitrobenzene (GS-DNB) (Habig *et al.*, 1974), a chromophore which has a characteristic absorbance at 340 nm.

Initially, hGSTP1-1 (1 µM) and GSH (50 mM) were diluted in the assay buffer (50 mM potassium phosphate, 1 mM EDTA and 0.02% (w/v) sodium azide, pH 6.5) to a final concentration of 40 nM and 1 mM, respectively. The conjugation reaction was then initiated within a Herasil quartz cuvette with a path length of 1 cm (Hellma®, Müllheim, Germany) by adding CDNB (50 mM, 100% (v/v) ethanol) to a final concentration of 1 mM (2% (v/v) ethanol), while ensuring a final assay volume of 3 ml. The formation of GS-DNB (ε_{340} = 9600 $M^{-1}.cm^{-1}$) (Habig *et al.*, 1974) at 20 °C was then monitored at 340 nm for 60 s by using a Jasco V-630 UV-Vis absorbance spectrophotometer (Jasco, Pfungstadt, Germany) which had a bandwidth and data pitch of 1.0 nm and 0.1 s, respectively. The spectra were then corrected for all the data collected by subtracting the contributions from the buffer and the slow, but spontaneous reaction between GSH and CDNB. The linear progress curve was then plotted as a change in the absorbance at 340 nm over time and the gradient from the linear regression curve was used to represent that rate of the reaction.

To monitor the potential inhibition in the enzyme activity of hGSTP1-1 by IAA-94 and EAA, when hGSTP1-1 was added to each sample, the enzyme was incubated for 30 min in either IAA-94 or EAA over a concentration range of 0-200 µM. However, a sample containing 2.1 mM IAA-94 or 2.48 mM EAA was also prepared considering that this was the concentration of the inhibitor used to perform an ITC experiment (section 2.2.9.1). Furthermore, to ensure that the DMSO did not have an influence on the enzyme activity of hGSTP1-1, a control sample was prepared to contain an equivalent concentration of DMSO that was present in the sample that contained 200 µM of the inhibitor. The reaction rates for each sample were then converted to a percentage of enzyme activity as follows:

$$\text{Enzyme activity (\%)} = \frac{\text{(Reaction rate: inhibitor)}}{\text{(Reaction rate: no inhibitor)}} \times 100\% \quad \textbf{(8)}$$

The percentage of enzyme activity was then plotted against the logarithm of inhibitor concentration and the IC_{50} value was determined by fitting the data in SigmaPlot® (SYSTAT,

University of Illinois at Chicago, USA) to a sigmoidal dose-response (variable slope) function that used the following function:

$$y = \frac{\text{min} + (\text{max}/\text{min})}{1 + 10^{(\log IC_{50} - x)\cdot \text{Hillslope}}} \quad \textbf{(9)}$$

where min represents the minimum enzyme activity value, max represents the maximum enzyme activity value, IC_{50} represents the half maximal inhibitory concentration and the Hillslope represents the slope at the midpoint of the fitted curve and therefore a larger Hillslope value would result in a steeper curve and *vice versa*.

2.2.9 Isothermal titration calorimetry (ITC)

Biocalorimetry, the study of the heat transfer accompanying a biochemical reaction, depends upon techniques such as ITC in order to quantify the thermodynamic parameters associated with a biomolecular interaction (Sheehan, 2009). ITC is an ideal technique as it allows for high affinity binding constants within the range of 10^8-10^9 M^{-1} to be quantified without the need for labels such as fluorophores or radioisotopes. Consequently, by using ITC to quantify the energetics accompanying a molecular interaction at a constant temperature, thermodynamic parameters such as the change in enthalpy (ΔH), stoichiometry (n), dissociation constant (K_d), ΔG and the change in entropy (ΔS) can be quantified for an interaction (Freyer and Lewis, 2008). Therefore, ITC is able to provide insight into how favourable hydrophobic and/or electrostatic interactions mediate a particular biomolecular interaction in order to give better context to the relationship between the structure, thermodynamics and function of protein-ligand interactions, a comparison which cannot be deduced solely from the structure of a protein (Dutta *et al*., 2015).

An ITC instrument contains two cells, both of which are enclosed in an adiabatic jacket. The one cell contains the buffer of the system and therefore aids as a reference cell, while the other cell functions as a sample cell considering that this is where the biomolecular interaction of interest will occur. Consequently, in order to monitor the thermodynamic parameters underlying a particular biomolecular interaction, a typical ITC experiment involves the stepwise titration of a ligand into the sample cell which contains the protein of interest. However, due to a thermoelectric device which is present between both cells, any temperature difference that will occur during an interaction can be measured, and depending on whether the

binding reaction is endothermic or exothermic, the thermoelectric device will ensure that the difference between both cells remains zero by either adding or removing heat from the system, respectively. Therefore, the amount of power applied to maintain a uniform temperature will reflect the heat flow per an injection, which in turn is proportional to the stoichiometry as well as the binding enthalpy for the interaction. Finally, by using an appropriate mathematical model, the heat flow per an injection will be integrated to give the total heat per an injection and subsequently the thermodynamic parameters associated with a particular biomolecular interaction (Freyer and Lewis, 2008).

2.2.9.1 Determining the thermodynamic parameters for a protein-ligand interaction

Initially, a primary stock solution of 70 mM IAA-94 and 82.5 mM EAA were prepared by solubilising 25 mg of the lyophilised powder in 100% (v/v) DMSO to a final concentration of 25 $mg.ml^{-1}$ (the solubility limit of IAA-94 and EAA). The secondary stock solution of IAA-94 or EAA used for each ITC experiment was then prepared by mixing 30 µl of the primary stock solution with 970 µl of the sample buffer (section 2.2.3.3) given that the solution needed to have a final DMSO concentration of 3% (v/v) in order to avoid the organic solvent damaging the protein. Similarly, the stock solution of 100 µM CLIC1-WT, hGSTP1-1 or EcSspA used during each ITC experiment was prepared to have a final DMSO concentration of 3% (v/v) to avoid any form of buffer mismatch. Importantly though, when preparing the protein stock solution, DMSO was first mixed with the required volume of the sample buffer in order to reduce the concentration of DMSO and thereafter the required volume of the protein sample was added and mixed with the buffer-DMSO solution.

The thermodynamic parameters for all respective protein-ligand interactions were monitored by using the Nano ITC standard volume instrument (TA® Instruments, New Castle, USA) and therefore each ITC experiment required at least 2 ml of the protein sample and 1 ml of the respective ligand solution in order to perform not only the ligand into protein titration, but also to perform the ligand into buffer titration that was used to monitor the heats of dilution (HOD). Prior to performing each ITC experiment, the protein and ligand samples were degassed under a vacuum of 0.3–0.5 atm for a period of ~20 min. Unless stated otherwise, each ITC experiment was performed, in triplicate, at 20 °C, with a stir rate of 250 rpm, an initial and final baseline of 300 s and a duration period of 300 s between each injection. The volume of the ITC syringe used was 250 µl and during each run the initial injection volume was 2 µl, while the injection volume for all subsequent injections was 5 µl and therefore a total of 49 injections were

performed. Thereafter, the data from the isotherm was analysed by using the NanoAnalyze™ Software, during which the integrated heats per an injection (in microjoules) were corrected for the HOD followed by the data being fitted to an independent binding model.

2.2.9.2 The change in heat capacity (ΔC) for the interaction between IAA-94 and hGSTP1-1

In order to monitor the influence that temperature had on the thermodynamic parameters for the interaction between IAA-94 and hGSTP1-1, the ITC experiments were performed under the same experimental procedures as mentioned in section 2.2.9.1, but across a temperature range of 5-30 °C that varied by 5 °C increments. The observed change in enthalpy (ΔH_{obs}) for each experiment was then plotted as function of temperature in kelvin and a linear regression model was fitted to the data considering that the gradient for the slope represented the ΔC for the interaction between IAA-94 and hGSTP1-1 under isobaric conditions (ΔC_p).

2.2.9.3 Proton-linked effects during the interaction between IAA-94 and hGSTP1-1

To account for any protonation-linked effects which could have had an influence on the change in binding enthalpy (ΔH_{bind}) and therefore contributed to the ΔH_{obs}, ITC experiments were performed at 25 °C in buffer systems that had a different change in ionisation enthalpy (ΔH_{ion}) and under the same experimental procedures mentioned in section 2.2.9.1. While kept at a constant ionic strength of I = 75 mM, the buffers systems used for the different ITC experiments were: (i) 50 mM potassium phosphate (ΔH_{ion} = 3.52 kJ.mol^{-1}), (ii) 37.5 mM 1,4-piperazinediethanesulfonic acid (PIPES) (ΔH_{ion} = 8.70 kJ.mol^{-1}), (iii) 150 mM 2-morpholin-4-ylethanesulfonic acid (MES) (ΔH_{ion} = 12.70 kJ.mol^{-1}), (iv) 50 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl] ethanesulfonic acid (HEPES) (ΔH_{ion} = 20.38 kJ.mol^{-1}) and (v) 300 mM imidazole (ΔH_{ion} = 36.65 kJ.mol^{-1}) (Goldberg *et al*., 2002). The ΔH_{obs} was then plotted as a function of ΔH_{ion} and if there was a protonation-linked effect, a linear regression model was then fitted to the data considering that the y-intercept and gradient represented the true ΔH_{bind} and the number of protons exchanged, respectively.

2.2.9.4 Detecting the main reactive cysteine residue within soluble CLIC1

Similar to IAA-94 and EAA, a primary stock solution of 48.5 mM 6-IAF was prepared by solubilising 25 mg of the lyophilized powder in 100% (v/v) DMSO to a final concentration of 25 mg.ml^{-1}. The secondary stock solution of 6-IAF used for each ITC experiment was then prepared by mixing 10 µl of the primary stock solution with 990 µl of the sample buffer (section 2.2.3.3) which resulted in the solution having a final DMSO

concentration of 1% (v/v) and therefore considering that the concentration of the organic solvent was below 3-5% (v/v), it would not have damaged the protein.

Importantly, however, considering that 6-IAF is an alkylating agent, when CLIC1-WT/C24A was dialysed against the sample buffer as described in section 2.2.3.3, the last dialysis step was omitted whereby CLIC1-WT/C24A would have been dialysed against the sample buffer containing 3 mM TCEP. Instead, CLIC1-WT/C24A was first dialysed against the sample buffer containing 3 mM TCEP for 12-16 h at 4 °C to ensure that CLIC1-WT/C24A was reduced. Thereafter, the sample was dialysed at 4 °C against the sample buffer to remove the TCEP, whereby the sample was first dialysed against 2 L of the storage buffer for 12-16 h, followed by 1 L of the storage buffer for 3-4 h, 2 L of the storage buffer for 3-4 h and then 1 L of the storage buffer for 12-16 h. Thereafter, the stock solution of 100 µM CLIC1-WT/C24A used during each ITC experiment was prepared to have a final DMSO concentration of 1% (v/v) in order to avoid any form of buffer mismatch. Importantly though, when preparing the protein stock solution, DMSO was first mixed with the required volume of the sample buffer in order to reduce the concentration of DMSO and thereafter the required volume of the protein sample was added and mixed with the buffer-DMSO solution. From thereon, all the ITC experiments were performed under the same experimental conditions described in section 2.2.9.1.

2.2.10 Protein crystallography and structure determination

Under selected conditions proteins have the potential to separate from solution and self-assemble into a crystalline state, however, the growth of quality crystals is a formidable task and is often the rate-limiting step in protein crystallography. The formation of the crystalline state typically involves the controlled precipitation of a protein from a supersaturated solution in order to ensure that the protein molecules become arranged in a periodic crystal lattice that is held together through non-covalent bonds. (Rupp, 2009).

The hanging-drop vapour diffusion technique is routinely used for the growth of protein crystals considering that it is cost effective, allows for multiple drops (conditions) to be setup with a single reservoir solution and it offers simplicity in harvesting the protein crystals. In the hanging-drop method a micro-droplet of a protein and precipitant solutions will be mixed and placed on the underside of a glass coverslip which covers and seals a well containing the concentrated precipitant solution. In view that the precipitant concentration is lower within the

micro-droplet, the sealed system will allow for an equilibrium to be established between the micro-droplet and the reservoir precipitant solution by means of vapour diffusion. Therefore, the diffusion of water from the micro-droplet will cause the concentration of the protein to increase to a point that is ideal for initiating a nucleation event that is required for crystal growth (Rupp, 2009).

Once a protein crystal has formed, the lattice planes comprising the crystal can be seen to comprise a series of identical unit cells, which are the smallest and simplest volume element that represent the entire crystal. However, prior to understanding the structure of the molecular content present within a unit cell, it is important bear in mind that a typical C-C bond has a bond length of 1.5 Å, yet an X-ray has a typical wavelength of ~1 Å. Consequently, when an X-ray beam interacts with a protein crystal, the measured directions and intensities of the diffracted X-ray beams will be able to provide information about the electron density and consequently the structure of the molecular content present with the unit cell (Rhodes, 2006; Rupp, 2009).

Initially, a crystal is therefore mounted between an X-ray source and an X-ray detector so that the protein crystal lies in the path of the narrow incoming X-ray beam in order to produce a series of diffracted beams that will represent distinct reflections on the detector. Then by using an optical scanner, the position and intensity of each reflection will be quantified, thus allowing the direction at which a particular beam was diffracted by the crystal to be determined. However, unlike the protein crystal which represents real space, the diffraction pattern represents reciprocal space, hence determining the directions for the diffracted beams is important given that each reflection will be assigned a set of three-dimensional hkl coordinates. These coordinates will not only represent the geometry of the unit cells, but also the location of the atoms within the unit cell. Yet, the major challenge in determining the structure of a protein from the diffraction pattern lies in having to convert the experimentally accessible information in the reciprocal space of the diffraction pattern to the inaccessible information of the real space inside the unit cell (Rhodes, 2006; Rupp, 2009).

However, as described by the structure-factor equation, each reflection is the result of diffractive contributions from all atoms in the unit cell, that is, every atom in the unit cell contributes to every reflection in the diffraction pattern. This is significant considering that by performing a Fourier transformation, the structure factors will be able to generate the electron density present in the unit cell. Therefore, Fourier transformations are able to precisely describe

the mathematical relationship between the reflections on the diffraction pattern in reciprocal space to the atoms in a unit cell in real space. Finally, having indexed, integrated, scaled and merged the diffraction data, a model will be built into the electron density, refined and validated by using a series of well-documented algorithms in order to determine the crystal structure of the target protein (Rhodes, 2006; Rupp, 2009).

2.2.10.1 Protein crystallisation

Prior to setting up crystal trials, the concentrated and purified sample of hGSTP1-1 (section 2.2.3.3) was additionally purified through SEC to ensure that the protein sample was pure and homogenous, conditions required for succesful protein crystallography. Initially, a 120 ml HiLoad™ 16/60 Superdex™ 75 prep grade SEC column (GE Healthcare Life Sciences, Chicago, USA) was attached to a ÄKTAprime plus™ liquid chromatography purification system (GE Healthcare Life Sciences, Chicago, USA) and pre-equilibrated with the sample buffer (section 2.2.3.3) at a flow rate of 1 ml.min^{-1}. However, in order to calibrate the SEC column, a standard curve was constructed (section 2.2.6) by using the retention time for a series of gel filtration molecular weight markers, namely (i) cytochrome c (12.4 kDa, 2 mg.ml^{-1}), (ii) carbonic anhydrase (29 kDa, 3 mg.ml^{-1}), (iii) bovine serum albumin (66 kDa, 10 mg.ml^{-1}), (iv) alcohol dehydrogenase (150 kDa, 5 mg.ml^{-1}), (v) β-amylase (200 kDa, 4 mg.ml^{-1}) and (vi) blue dextran (2 000 kDa, 2 mg.ml^{-1} with 5% (v/v) glycerol (100% (v/v)). Thereafter, a 2 ml sample of hGSTP1-1 was injected onto the column and the single uniform peak that correlated to the moelcular weight of hGSTP1-1 was collected and concentrated to ~15 mg.ml^{-1} by using an Amicon® Ultra-15 10kDa (MilliporeSigma, Massachusetts, USA) centrifugal concentration device.

The succesful crystallisation of hGSTP1-1 in complex with various ligands has been reported in the Protein Data Bank (PDB) and therefore the crystal conditions from these experiments were adapted for the crystal trials of hGSTP1-1 in complex with IAA-94 (Table S5). Initially, 3.5 mM IAA-94 (5% (v/v) DMSO) was mixed with the required volume of the sample buffer (section 2.2.3.3) that contained both 10 mM DTT and 10 mM GSH. Thereafter, hGSTP1-1 was added to the solution to a final concentration of 10 mg.ml^{-1} and the protein-ligand solution was left to incubate at 4 °C for 30 min. The crystal trials were then setup within a Qiagen EasyXtal™ 15-well hanging-drop plate system (Qiagen®, Hilden, Germany) by using the Oryx8™ protein crystallisation robot (Doughlas Instruments, East Garston, UK). In particular, 500 µl of the reservoir solution was added to each well and then

by using the three-bore tip system on the Oryx8™ protein crystallisation robot, four drops of the protein and reservoir solution were dispensed to each cover slide at varying ratios of volume (in microlitres) as follows: (i) 2:2, (ii) 1:2, (iii) 1:1 and (iv) 0.5:1.

The crystal trials were then carried out at 20 °C and multiple protein crystals would often appear per a well within three to five days. The protein crystals were then harvested by using a CrystalCap™ SPINE HT goniometer base that was fitted with either a 0.05-0.1 mm, 0.1-0.2 mm or a 0.2-0.3 mm Mounted CryoLoop™ (Hampton Research, California, USA). However, bearing in mind that a protein crystal will be susceptible to radiation damage by an incoming X-ray beam during data collection, a protein crystal would need to be kept at ~100 K in order reduce secondary raditaiton damage. Yet, owing to the high water content within the crytsal lattice, as well as the reservoir solution that surronds a protein crytsal, ice can form when a protein crystal is cooled to crygogenic tempertures. Consequently, this will be reflected as ice rings on the diffraction pattern and result in a loss of diffraction data. Therefore, in order to achieve water verification as opposed to ice formation, a protein crystal is soaked in a cryoprotectant so that the reservoir solution will be exchanged with a solution that is more suitable for water verification when a protein crystal is cooled to cryogenic temperatures (Vera and Stura, 2014). As a result, the CryoLoop™ which contained the protein crystal was then immersed two to three times in 100% (v/v) Parabar 10312 (Paratone) (Hampton Research, California, USA), an oil-based cryoprotectant suitable for biomacromolecule crystals. The CryLoop™ was then briefly inspected under a darkfield protein crystallography microscope to ensure that the protein crystal was still present within the CryoLoop™ and that it was sufficently covered by the cryoprotectant. Thereafter, the loops were stored in a Cryo Puck which had been placed in a Spearlab FD-130 cryogenic foam dewar (Hampton Research, California, USA) containing liquid nitrogen and thereafter the samples were shipped to the Diamond Light Source (DLS) (Harwell Science and Innovation Campus, UK) for X-ray diffraction data collection.

2.2.10.2 Data collection, processing, refinment and validation

The X-ray diffraction data from the harvetesed crystals was generated by using the I03 beamline, one of the five macromolecular crystallography (MX) beamlines at the DLS. The I03 beamline is a tuneable beamline with a working wavelength of 0.6-2.48 Å, however, the standard working wavelength used was 0.915 Å (12.7 keV) with a focused beam size of 90 x 20 µm and 90 x 30 µm at the sample and on the detector, respectively. The diffraction

data was then collected by using the Dectris Eiger2 XE 16M detector (Dectris, Baden, Switzerland) at 100 K while collecting a total of 3500 images. As soon as the data collection was completed, automated data processing was initated in order to refine, index and intergrate the diffraction data by using the programs MOSFLM and POINTLESS from the Xia2 pipeline software (Winter, 2010). During this process the geometrical arrangment of the reflections were measured and each spot was assigned their respective Miller indicies in order to provide information on the properities of the unit cell such as the type of crytsal system, Bravais lattice and space group. Thereafter, by using programs in the Collaborative Computational Project Number 4 (CCP4) suite (Collaborative Computational Project Number 4, 1994), the intensities for symmetrically equivalent reflections were scaled by using AIMLESS and the intensities were converted into their structure factor amplitudes (|F|) by using TRUNCATE.

Thereafter, the initial phases required to generate the electron density map from the diffracted data were determined through molecular replacement (MR). During MR a known molecular model with a set of diffraction data and that also shares some structural homology to the target protein is used to model and solve the structure of the taregt protein. During this process all orientations and postions of the known model were assessed within the model of the target protein in order to observe were the predicted and observed diffracton data match. The phases from the known molecular model were then used for the model of the target protein to create an initial map for subsequent refinement (Evans and McCoy, 2007). The MR was performed by using the software PHASER (McCoy *et al.*, 2007) and the wild-type conformation of hGSTP1-1 (PDB ID: 1AQX) was used as the known molecular model for phasing.

Thereafter, successive rounds of refinement were performed by using the software PHENIX (Adams *et al.*, 2011) and the crystallographic object-oriented toolkit (COOT) (Emsley and Cowtan, 2004) for reciprocal refinement and real space optimisation, respectively. The fit of the modelled residues for hGSTP1-1 within the electron density were then evaluated by using the tools included within COOT. During each iterative step of refinment, the convergence between the observed structure factors ($|F_o|$) and calculated structure factors ($|F_c|$) was monitored by a decrease in the R-factor, but the similarity between the R-factor and R-free values was also assessed to ensure that there was no over-fitting of the experimental data. Thereafter, in order to detect any improbable bond lengths and angles, along with any Ramachandran deviations and steric clashes, the quality of the model was validated by using MolProbity (Chen *et al.*, 2010; Davis *et al.*, 2007)

(http://molprobity.biochem.duke.edu/). The final model was then submitted to the Research Collaboratory for Structural Bioinformatics (RCSB) PDB (Bernstein *et al.*, 1977) (https://www.rcsb.org/) and the PDB code 6Y1E was provided for the model. Thereafter, the crystal structure was evaluated by using PDBsum (Laskowski, 2001) (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) and Pymol (The PyMol Molecular Graphics System, Version 1.3 Schrödinger®, LLC).

2.2.11 *S*-nitrosylation of CLIC1 by GSNO

2.2.11.1 Synthesis of GSNO

The synthesis of GSNO was initiated by mixing GSH, sodium nitrite and EDTA to a final concentration of 100 mM GSH, 100 mM sodium nitrite and 1 mM EDTA in Milli-Q® water (18.2 MΩ.cm^{-1} at 25 ^{0}C) which was from a Millipore Direct-Q® 3 UV water purification system (MilliporeSigma, Massachusetts, USA) and which was kept cool at 4 °C. The pH of the solution was then lowered to 1.5 by using concentrated hydrochloric acid and the solution was then left to incubate at 4 °C for 30 min. However, considering that GSNO is a light-sensitive compound, the beaker was covered in aluminium foil throughout the preparation process. Thereafter, the solution was adjusted to either pH 5.5 or pH 7.0 by potassium hydroxide and the resulting solution was then passed through a 0.45 µM syringe filter (cellulose acetate membrane), prepared into 2 ml aliquots and stored at - 80 ^{0}C.

2.2.11.2 Determining the stoichiometry of *S*-nitrosylated soluble CLIC1

Considering that TCEP would have interfered with the *S*-nitrosylation reaction, when CLIC1-WT was dialysed against the sample buffer as described in section 2.2.3.3, the last dialysis step was omitted whereby CLIC1-WT would have been dialysed against the sample buffer containing 3 mM TCEP. Instead, CLIC1-WT was first dialysed against the sample buffer containing 3 mM TCEP for 12-16 h at 4 °C to ensure that CLIC1-WT was reduced. Thereafter, the sample was dialysed at 4 °C against the sample buffer to remove the TCEP, whereby the sample was first dialysed against 2 L of the storage buffer for 12-16 h, followed by 1 L of the storage buffer for 3-4 h, 2 L of the storage buffer for 3-4 h and then 1 L of the storage buffer for 12-16 h. However, considering that at the surface of the membrane the pH is ~5.5, the *S*-nitrosylation of CLIC1-WT was performed at both pH 5.5 and pH 7.0 and therefore the sample buffer used was either (i) 50 mM potassium phosphate, 1.5 mM EDTA and 0.02% (w/v) sodium azide, pH 7.0 or (ii) 150 mM sodium acetate, 1.5 mM EDTA and 0.02% (w/v) NaN_3, pH 5.5. The concentration of the GSNO stock solution was then

confirmed by mixing, in triplicate, 10 µl of the GSNO stock solution with 1990 µl of Milli-Q® water (18.2 MΩ.cm^{-1} at 25 ^{0}C) that was from a Millipore Direct-Q® 3 UV water purification system (MilliporeSigma, Massachusetts, USA). The absorbance of each sample was then measured across a wavelength range of 250-500 nm as described in section 2.2.5 and the concentration of the GSNO stock solution was determined by using the Beer-Lambert Law (equation 2) (ε_{335} = 922 M^{-1}.cm^{-1}) (Hart, 1985).

The *S*-nitrosylation of CLIC1-WT was then performed by incubating CLIC1-WT with GSNO in a 1:300 ratio at 20 °C for 1 h. Thereafter, the excess GSNO was removed by dialysing the protein sample at 4 °C against its respective sample buffer, whereby the sample was first dialysed against 2 L of the storage buffer for 12-16 h, followed by 1 L of the storage buffer for 3-4 h, 2 L of the storage buffer for 3-4 h and then 1 L of the storage buffer for 12-16 h. Thereafter, the samples for unmodified and *S*-nitrosylated CLIC1-WT (pH 5.5 and pH 7.0) were passed through a 0.45 µM syringe filter (cellulose acetate membrane) and an absorbance spectra for each protein sample was measured across a wavelength range of 250-500 nm as described in section 2.2.5. The concentration of CLIC1-WT and the *S*-nitrosothiol group was then determined by using the Beer-Lambert Law (equation 2) and the degree of *S*-nitrosylation was estimated by taking a ratio of the concentration for the *S*-nitrosothiol group (λ = 335 nm) to the concertation of CLIC1-WT (λ = 280 nm). Therefore, bearing in mind that CLIC1-WT contains six cysteine residues, this ratio could vary from one to six depending on which cysteine residue became *S*-nitrosylated.

2.2.11.3 Liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS)

In the presence of a magnetic field any moving charged particle can be deflected. However, the degree of this deflection is dependent on the momentum of the charged particle, that is, species with a large momentum will be deflected less than those with a small momentum. Therefore, when a stream of atoms with a similar charge and velocity, but different mass, pass through a magnetic field, the deflection experienced by each atom will be a function of its mass. The principle underlying MS therefore relies on a molecule being ionised into the gas phase so that the ions which are generated can be separated in a mass spectrometer in accordance to their mass-to-charge (m/z) ratio. As a result, MS is among one of the most accurate techniques for determining the molar mass of a molecule (Hoffmann and Stroobant, 2007; Sheehan, 2009).

The source for ionisation during an MS experiment can occur through numerous methods, yet, in the case of macromolecules such as proteins, these molecules present a challenge as they are often difficult to ionise without fragmentation. However, electrospray ionisation (ESI) is an evaporation-based ionisation technique which has been able to circumvent this issue. During ESI a sample is sprayed through a metal syringe at a high voltage and in turn this causes the solvent to evaporate, the droplet size to decrease and the charge density on the droplet surface to increase. Ultimately, as a result of the increased instability in the droplet, this technique causes a Columbic explosion that generates a number of sufficiently smaller droplets that allow ions to enter into the gas phase. Therefore, considering this ionisation method requires relatively small inputs of thermal energy and results in the gentle ionisation of the sample components, ESI is especially useful in the ionisation of intact macromolecules such as proteins (Hoffmann and Stroobant, 2007; Sheehan, 2009).

Consequently, in view of the extremely high accuracy in MS techniques, it is possible to obtain detailed information on the structural aspects of a protein such as its amino acid sequence as well as the presence of any PTMs (Sheehan, 2009). However, in relation to PTMs, a useful technique aiding in their detection is peptide mass fingerprinting (PMF), whereby during this technique a protein molecule will be subject to a specific protease such as trypsin and thereafter the mass of the individual peptides which are generated will be determined by MS. Consequently, by comparing the theoretical mass for a particular peptide sequence with the measured mass, any mass shift between the two values can be determined and assigned to a particular modification (Zhang *et al*., 2003). Therefore, PMF was performed at the Council for Scientific and Industrial Research (CSIR) (Pretoria, South Africa) in order to determine which cysteine residue became *S*-nitrosylated by GSNO.

Briefly, 100 µl of the unmodified and *S*-nitrosylated samples of CLIC1-WT (pH 5.5 and pH 7.0), at a concentration of 60 µM (2.8 $mg.ml^{-1}$), were incubated 5 µl of the labelling solution (200 mM N-ethylmaleimide (NEM) and 4% (w/v) SDS) at 20 °C for 1 hr in the dark in order to label any free and unmodified cysteine residues. Thereafter, the preparation of each sample relied on the KingFischer™ Duo Prime system (Thermo Fischer Scientific™, Massachusetts, USA), which was an automated liquid handling system with a magnetic bead isolation station that was configured for HILIC (hydrophilic interaction chromatography)-based clean-up and protein digestion.

Initially, MagReSyn® HILIC micro-particles (20 mg.ml^{-1}) were vortexed for ~5 s to ensure the micro-particles were re-suspended and that a homogenous solution was formed. Thereafter, 25 µl of the MagReSyn® HILIC micro-particles were transferred to a magnetic separator where they were diluted with the HILIC equilibration buffer (100 mM ammonium acetate and 15% (v/v) acetonitrile, pH 4.5) to a final concentration of 500 µg.ml^{-1}, followed by two successive washes in 250 µl of the HILIC equilibration buffer with gentle agitation for 20 s. Thereafter, the MagReSyn® HILIC micro-particles were placed in the wells of a row from a 96 deep-well microtiter plate, while 100 µl of the labelled protein, which had been diluted with the HILIC binding buffer (200 mM ammonium acetate and 35% (v/v) acetonitrile, pH 4.5) to a final concentration of 125 µg.ml^{-1}, was placed in the wells of another row that had 100 µl of HILIC binding buffer.

The automated system was then used to transfer the protein sample to the equilibrated MagReSyn® HILIC micro-particles in order to ensure a final protein concentration of 50 µg.µl^{-1} and ultimately a protein to MagReSyn® HILIC micro-particle ratio of 1:10. The protein and MagReSyn® HILIC micro-particles were then left to incubate together at 20 °C for 30 min, but due to the *S*-nitrosothiol groups being light-sensitive (Broniowska *et al*., 2013) all incubation steps were performed in the dark. However, in order to ensure a good protein to micro-particle interaction during the binding procedure, continuous and gentle mixing was important, yet excessive mixing was prevented to avoid poor protein recovery owing to the MagReSyn® HILIC micro-particles drying to the sides of the well. The protein-bound micro-particles were then transferred to two subsequent rows and washed in 200 µl of 95% (v/v) acetonitrile with gentle mixing. Thereafter, the protein-bound micro-particles were added to the final row containing 5 µg.µl^{-1} trypsin (50 mM ammonium bicarbonate, pH 8.0) to ensure that an enzyme to protein ratio of 1:10 was achieved. Protein digestion by trypsin was then performed at 37 °C for 4 h with continuous and gentle mixing to once again ensure a good protein to micro-particle interaction. Finally, the micro-particles were then isolated by using a magnetic separator and the supernatant containing the peptides was transferred to a microcentrifuge tube, lyophilised and stored at - 80 °C until analysis.

Prior to analysis, the protein digests were first re-suspended in 20 µl of the peptide re-suspension buffer (2% (v/v) acetonitrile and 0.2% (v/v) formic acid) and thereafter the samples were analysed by using a Dionex Ultimate 3000 rapid system liquid

chromatography (RSLC) system (Dionex, California, USA) that was coupled to a AB Sciex 6600 TripleTOF mass spectrometer (SCIEX, Framingham, USA). However, the re-suspended peptides were first desalted on an Acclaim™ PepMap™ C18 trap column (100 µm × 2 cm) (Thermo Fisher Scientific™, Massachusetts, USA) for 2 min at 15 µl.min^{-1} by using the peptide re-suspension buffer and thereafter they were separated on an Acclain™ PepMap™ C18 RSLC column (300 µm × 15 cm, 3 µm particle size) (Thermo Fisher Scientific™, Massachusetts, USA). Thereafter, the peptides were then eluted from the column at a flow-rate of 8 µl.min^{-1} by using two buffers to form a gradient, namely elution buffer A (0.1% (v/v) formic acid) and buffer B (80% (v/v) acetonitrile and 0.1% (v/v) formic acid), whereby the gradient was increased from 4-60% with respect to buffer B over a period of 15 min.

Then to start the analysis of the peptides, an electrospray voltage of 5.5 kV was applied to the emitter and the AB Sciex 6600 TripleTOF® mass spectrometer (SCIEX, Framingham, USA) was operated in Data Dependent Acquisition (DDA) mode for peptide identification. In DDA mode, precursor MS scans were acquired from m/z of 200-1500 by using an accumulation time of 250 ms, followed by 30 MS/MS product scans that were acquired from m/z of 100-1800 at 100 ms each, for a total scan time of 3.3 s. The MS/MS spectra obtained from the peptide pool for unmodified and *S*-nitrosylated CLIC1-WT at both pH 5.5 and pH 7.0 were then initially analysed by using the ProteinPilot™ 5.0.1 software in order to compare the data with a common database that contained the sequence of CLIC1-WT as well as a list of sequences of common contaminating proteins. However, the data was then analysed by using the Skyline-daily 3.7.1.11357 MacCoss Lab software © (MacLean *et al*., 2010), an open source software which provides an informative graphical user interface (GUI) to not only visualise the raw MS and chromatographic data, but to also refine and integrate the area under the chromatographic peak of each peptide.

Initially, the retention time for each peptide from the various samples was refined such that the retention time for a particular peptide was the same across all the samples. Then the main peak in the chromatogram for a particular peptide was weighted based on the intensity of the peak and the isotope dot product (IDOTP) score generated. In nature, associated with each element are a select number of isotopes, however, their distribution and abundance may differ. For instance, carbon is the most abundant atom within a protein and naturally there are three occurring isotopes of carbon, C^{12}, C^{13} and C^{14}, however, the relative abundance of C^{12} surpasses the other two isotopes (Brown *et al*., 2012). Therefore, based on the chemical formula for a

particular peptide, the IDOTP score will compare the predicted isotope distribution with the observed isotope distribution that was determined based on the integrated peak area for each isotope in the chromatogram. As a result, an IDOTP score of 1.0 indicated a perfect match and in this study an IDOTP score $\geq$ 0.95 was set as a cut-off to define the peptides that were successfully refined (MacLean *et al.*, 2010; Schilling *et al*., 2012).

To then quantify the degree at which a cysteine residue/s became *S*-nitrosylated by GSNO in CLIC1-WT, a ratio was taken between the integrated area of the modified and unmodified state for a particular peptide from each individual sample. Therefore, for unmodified CLIC1-WT, the proportion of the modified to unmodified state for a particular peptide should be $\leq$ 1, whereas the proportion of the modified to unmodified state for a particular peptide should be $>$ 1 for the *S*-nitrosylated conformation of CLIC1-WT.

2.2.12 Molecular dynamic (MD) simulations

While X-ray crystallography may offer insight into the structural conformation of a protein, it does not take into consideration the overall dynamic nature of a protein. However, although there are a variety of experimental techniques which can provide information about the dynamic nature of a protein, they only represent an ensemble of the average motions and not the motions of the individual protein molecules. Therefore, by using the laws of classical physical principles, *in silico* studies provide an attractive alternative method to model the atomic-level motions of a protein molecule and as a result the technique known as all-atom MD simulations has become the standard method for simulating the dynamic motion of protein molecules (Monticelli and Tielemann, 2014; Kukol, 2008)

MD simulations on CLIC1-WT and its corresponding C24A mutant were performed by using the crystal structure of CLIC1-WT (PDB ID: 1K0M) and CLIC1-C24A (PDB ID: 4K0N) that were solved to a resolution of 1.4 Å and 1.25 Å, respectively. The quality of the models were then validated by using PROCHECK in PDBsum (Laskowski, 2001) (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) and Molprobity (Chen *et al.*, 2010; Davis *et al.*, 2007) (http://molprobity.biochem.duke.edu/). However, it was observed that at the N-terminus of both proteins, the first five amino acids (^{1}MAEEQ^{5}) were missing, but these missing residues were added to both proteins by using Pymol (The PyMol Molecular Graphics System, Version 1.3 Schrödinger®, LLC). Thereafter, the MD simulations were carried out by using Groningen

Machine for Chemical Simulations (GROMACs) 5.07 that was operated on a Linux operating system. All MD simulations were carried out on an Intel core i7 5960 X extreme edition (3.3 GHz, 20 M cache 16 x cores) which was equipped with a GTX 750Ti graphics card and 32 GBDDR4-2133 MHz memory on a MSI X99 motherboard.

Initially, by using the complete protein structure of CLIC1-WT/C24A, a topology file was generated to contain all the information that was necessary to define the protein molecule within the simulation, which included non-bonded parameters such as atom types and charges as well as bonded parameters such as bond orders, lengths and angles. At this point a force field needed to be selected in order to process the structure of both proteins and in this study the all atom-optimised potentials for liquid simulations (OPLS-AA) force field was used to process both proteins. However, considering that the water molecules were initially removed from both crystal structures, solvent water molecules needed to be reintroduced in order to simulate CLIC1-WT/C24A in an aqueous system. In order to do this, a simple cubic unit cell was generated and the protein was centred 1 nm from the box edge, therefore at least 2 nm was kept between any two periodic images of the protein. Thereafter, the unit cell was filled with solvent by using the generic equilibrated three-point solvent model that was part of the standard GROMACS installation. Furthermore, in order to neutralise the net charge of CLIC1-WT/C24A, seven sodium ions were added to the solvated system by using the genion tool (Abraham *et al*., 2017; Lemkul, 2018).

However, the added hydrogens and broken hydrogen bond network among the water molecules would lead to large forces and structure distortion during MD simulations. Therefore, the system was relaxed through energy minimisation (EM) by using the steepest decent method in order to ensure that the solvated and electroneutral system had no steric clashes or inappropriate geometry. During EM the potential energy (E_{pot}) of the system was kept on the order of 10^5-10^6 kJ.mol^{-1}, while the maximum force (F_{max}) was kept at $F_{max} \leq 1000$ kJ.mol^{-1}.nm^{-1}. However, although EM ensured that the geometry and solvent orientation for the starting structure was reasonable, the kinetic energies of the solvent and ions surrounding the protein needed to be amended to ensure that these particles established a proper orientation around the protein. In order to achieve this, the temperature and therefore kinetic energy of the system was first brought to equilibrium by using the constant number of particles, volume and temperature (NVT) ensemble and the system was left to equilibrate for 100 ps until the temperature plateaued at the target value of 300 K. Thereafter, the pressure and hence the density of the

system was brought to equilibrium by using the constant number of particles, pressure and temperature (NPT) ensemble and the system was left to equilibrate for 100 ps until the density values for the system were stable over time (Abraham *et al*., 2017; Lemkul, 2018).

Once the system was well-equilibrated, a MD simulation run of 50 ns was performed for both proteins. Analysis of the resultant trajectories was then carried out by using the script-based functions in GROMACS (Abraham *et al*., 2017) and by using the molecular vision software UCSF Chimera Version 1.11.2 (Resource for Biocomputing, Visualisation, and Informatics (RBVI), UCSF, USA). Furthermore, all images of the proteins were generated by using Pymol (The PyMol Molecular Graphics System, Version 1.3 Schrödinger®, LLC).

2.2.13 Induced-fit molecular docking (IFMD) simulations

Understanding the mechanisms underlying protein-ligand interactions is essential for drug discovery. Yet, with limited or no structural information available for many protein-ligand complexes, alternative methods that are based on experimental data needed to be developed in order to predict the structure of these interactions. One such scaffold that provides an attractive means to study protein-ligand interactions is molecular docking. During molecular docking the induced-fit model is most often applicable, a model which stipulates that the binding site in a protein is flexible and will undergo a conformational change upon ligand-binding. Yet, because the induced-fit mechanism only takes into account the conformational flexibility of the ligand-binding site, it is not always suitable for proteins which undergo significant conformational changes upon ligand-binding. Nevertheless, IFMD will generate an optimised conformation of a protein-ligand complex which can be used to infer the possible mode of binding and which can be related to other experimental data (Dror *et al*., 2012; Du *et al*., 2016).

IFMD was performed by using the Schrödinger® Glide and Prime computational models in the Mastero 11.2 GUI (Schrödinger Release 2017-2: Maestro, Schrödinger, LLC, New York, NY, 2017). Initially, the target protein was pre-processed by using the Protein Preparation Wizard tool which assigned the necessary bond orders and hydrogen atoms to the target protein. Thereafter, (i) the solvent orientation was optimised, (ii) the water molecules that were less than three hydrogen bonds to non-water molecules were removed and (iii) the pK_a values of the ionisable groups within the target proteins were predicted by the PROPKA tool (Li *et al.*, 2005) in order to ensure that the charged states for all the ionisable residues were optimised in

accordance to a solution at pH 7.0. The protein model was then subjected to EM by using the OPLS3 force field (Harder *et al.*, 2015) until an average RMSD of 0.3 Å was reached.

The molecular structure of the ligands used for IFMD were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/) and these included: (i) IAA-94 (PubChem ID: 656717), (ii) EAA (PubChem ID: 3278), (iii) MES (PubChem ID: 78165), (iv) ANS (PubChem ID: 1369) and (v) 6-IAF (PubChem ID: 126346). Thereafter, by using the LigPrep tool, the ionisation state of a ligand was generated at pH 7.0 ± 2.0 by using Epik, a program used to predict the pK_a values for the ionisable groups within a ligand along with generating the appropriate ionised and tautomerised structures within a specific pH range. The Epik platform employs Jaguar pK_a calculations and the widely used Hammett and Taft empirical equations, whereby the values for the parameters in these equations are derived from experimental data reported in literature (Shelley *et al.*, 2007). Furthermore, in the case of a ligand molecule having a unique stereochemistry, it was selected to retain its specific chirality and at most 32 stereoisomers were generated per a ligand. The ligand molecule was then subject to EM by using the OPLS3 force field (Harder *et al.*, 2015).

The energy minimised protein and ligand structures were then used to perform the respective IFMD simulation. Initially, a residue at the binding-site of the protein was selected and surrounding this residue was a 20 Å cubic box that was used to define the region were IFMD could occur. During IFMD, the ring conformations of the ligand were sampled in an energy window of 2.5 kcal.mol^{-1} and in the case were a ligand contained an amide bond, the non-planar conformation of the bond was penalised. However, with respect to the target protein, the residues that were within 5.0 Å of the ligand poses were refined and their side chains optimised. Furthermore, the stages at which hydrogen bond constraints could be applied were both at the initial and re-docking stages of the IFMD simulation. However, Glide docking does not allow for receptor flexibility, instead the van der Waal radii of non-polar residues can be scaled to decrease the penalty scores for close contacts in order to provide some minor flexibility in the receptor and ligand (Schrödinger®, 2015). Therefore, during Glide docking a scaling factor of 0.50 was selected for the van der Waals radii of the ligand and protein. A maximum number of 20 poses were then generated during Glide docking and the ligand was re-docked into structures that were within 30 kcal.mol^{-1} of the best structure. Finally, the IFMD simulation was carried out by using the OPLS3 force field, a force field which was developed so that its parameters were optimised in order to reproduce liquid state

thermodynamic properties for a variety of small organic molecules and to provide information on the non-bonded parameters (Harder *et al.*, 2015).

The best docked structure was then selected based on the pose which had the lowest IFMD score, a parameter which represents the sum of the GlideScore (G_{score}) and 5% of the Prime energy (Schrödinger®, 2019a). Thereafter, the protein-ligand complex for the top pose was assessed by analysing the Glide model energy (GE_{model}) score and the G_{score}. The GE_{model} score takes into account the binding affinity predicted by the G_{score}, the non-bonded interaction energies and the excess internal energy that was generated by the ligand conformation during flexible docking (Schrödinger®, 2015). As a result the GE_{model} score has a more significant weighting on the force field components and it is used when comparing conformers of the same ligand, but not chemically-distinct species (Schrödinger®, 2019b). Although pose selection is often performed with accuracy, being able to accurately predict the binding affinity for a particular interaction is a challenge and often it goes beyond the scope of docking scoring functions given that they tend to omit essential thermodynamic principles that govern the free energy of binding. Nevertheless, Glide is focused on generating an accurate pose for each protein-ligand interaction followed by separating these complexes based on their binding affinity, however, it is important to keep in mind that Glide does not predict binding energies, instead it is primarily intended for database enrichment (Schrödinger®, 2019c).

The G_{score} is incorporated both within the GE_{model} and IFMD score and as a result it is a key parameter used to evaluate the success of an IFMD simulation. The G_{score} is an empirical scoring function, that is, by counting the number of various types of interactions between two binding partners, it intends to reproduce experimental affinity data (Guedes *et al.*, 2018; Schrödinger®, 2019b). Therefore, considering that the G_{score} aids at simulating the binding free energy of an interaction, a more negative G_{score} value will represent a protein-ligand interaction with a stronger binding affinity (Schrödinger®, 2019b). Consequently, by being an empirical scoring function, the G_{score} accounts for many interaction terms that were used to describe the protein-ligand complex for the top pose and it was calculated as follows:

$$G_{score} = 0.05 \cdot vdW + 0.15 \cdot Coul + Lipo + H_{bond} + Metal + BuryP + RotB + Site \quad \textbf{(10)}$$

where (i) vdW and (ii) Columb represent the van der Waals energy and Coulomb energy, respectively, both of which are calculated with reduced net ionic charges on groups with formal charges (e.g. metals, carboxylates and guanidiniums), (iii) Lipo represents the lipophilic term

and it will reward favourable hydrophobic interactions, (iv) H_{bond} represents the hydrogen bonding term which is separated into different components depending on whether both the donor and acceptor are neutral, one is neutral and the other is charged, or if both the donor and acceptor are charged, (v) Metal represents the metal-binding term and it only includes interactions with an anionic acceptor group, that is, if the net charge of the metal in the apo-protein is positive, then the preference for anionic ligands will be included in the score, but if the net charge is zero the preference will be suppressed, (vi) BuryP represents the penalty for buried polar groups, (vii) RotB represents the penalty for freezing rotatable bonds and (viii) Site represents the polar interactions in the active site whereby polar groups (with non-hydrogen-bonding atoms) are rewarded when they are in a hydrophobic region (Schrödinger®, 2015).

Having then used these parameters to select and analyse the pose which had the most favourably docked protein-ligand interaction, all images of the complex were then generated by using Pymol (The PyMol Molecular Graphics System, Version 1.3 Schrödinger®, LLC).

Chapter 3: Results and discussion

3.1 Protein purification and structural characterisation

3.1.1 Verifying the cDNA sequence of the target proteins

Sequencing of the CLIC1-WT cDNA sequence which had been inserted into a pET-28a vector confirmed that the nucleic acid sequence encoded for the hexa-histidine fusion-tag, thrombin cleavage site and wild-type sequence (Figure 22). In the case of the cDNA sequence for CLIC1-C24A, Sanger sequencing verified that the codon sequence ^{5}'TGC^{3}' at position 24 was mutated to ^{5}'GCT^{3}', therefore verifying that Cys24 was successfully mutated to Ala24 and that no further mutations occurred to the primary sequence (Figure 23). Furthermore, when sequencing the cDNA sequence for hGSTP1-1 and EcSspA which had been inserted into a pET-15b and pET-28a vector, respectively, it was confirmed that the nucleic acid sequence for both proteins encoded for the hexa-histidine fusion-tag, thrombin cleavage site and wild-type sequence (Figure 24 and 25). However, although the query sequence for hGSTP1-1 was noted to have 100% sequence similarity from the fourth amino acid residue to the sequence available on NCBI (NP_000843.1), it was noted that the query sequence contained an additional two residues (Leu2 and Asp3) subsequent to the first methionine residue (Figure 24). However, given that these two residues are located near the N-terminal cap it was reasoned that they would not have an impact on the overall structural integrity and function of the protein.

3.1.2 Purifying the target proteins by means of affinity chromatography

All recombinant target proteins were purified from the supernatant of homogenised T7 Express *E.coli* cells by means of either a Ni^{2+} or Co^{2+}-IMAC column as described in section 2.2.3. The supernatant was initially loaded onto the IMAC column which resulted in an increased absorbance at 280 nm due to various contaminant proteins which did not bind to the column passing through. Thereafter, any remaining unbound contaminant and fusion protein were removed from the IMAC column by using the respective wash buffer and this resulted in the decreased absorbance at 280 nm. Once the absorbance readings had decreased and became consistent the bound fusion protein was eluted from the column by means of a single step elution (Figure 26, 27, 28 and 29).

A.

```
ctc atc tag ata att ttg ttt act tta aga agg aga tat acc atg ggc agc agc cat cat cwt cat cat cac agc agc ggc ctg gtg ccg
 L   I   -   I   I   L   F   T   L   R   R   R   Y   T   M   G   S   S   H   H   X   H   H   H   S   S   G   L   V   P
cgc ggc agc cat atg gct gaa gaa caa ccg cag gtc gaa ttg ttc gtg aag gct ggc agt gat ggg gcc aag att ggg aac tgc cca
 R   G   S   H   M   A   E   E   Q   P   Q   V   E   L   F   V   K   A   G   S   D   G   A   K   I   G   N   C   P
ttc tcc cag aga ctg ttc atg gta ctg tgg ctc aag gga gtc acc ttc aat gtt acc acc gtt gac acc aaa agg cgg acc gag aca
 F   S   Q   R   L   F   M   V   L   W   L   K   G   V   T   F   N   V   T   T   V   D   T   K   R   R   T   E   T
gtg cag aag ctg tgc cca ggg ggg cag ctc cca ttc ctg ctg tat ggc act gaa gtg cac aca gac acc aac aag att gag gaa ttt
 V   Q   K   L   C   P   G   G   Q   L   P   F   L   L   Y   G   T   E   V   H   T   D   T   N   K   I   E   E   F
ctg gag gca gtg ctg tgc cct ccc agg tac ccc aag ctg gca gct ctg aac cct gag tcc aac aca gct ggg ctg gac ata ttt gcc
 L   E   A   V   L   C   P   P   R   Y   P   K   L   A   A   L   N   P   E   S   N   T   A   G   L   D   I   F   A
aaa ttt tct gcc tac atc aag aat tca aac cca gca ctc aat gac aat ctg gag aag gga ctc ctg aaa gcc ctg aag gtt tta gac
 K   F   S   A   Y   I   K   N   S   N   P   A   L   N   D   N   L   E   K   G   L   L   K   A   L   K   V   L   D
aat tac tta aca tcc ccc ctc cca gaa gaa gtg gat gaa acc agt gct gaa gat gaa ggt gtc tct cag agg aag ttt ttg gat ggc aac
 N   Y   L   T   S   P   L   P   E   E   V   D   E   T   S   A   E   D   E   G   V   S   Q   R   K   F   L   D   G   N
gag ctc acc ctg gct gac tgc aac ctg ttg cca aag tta cac ata gta cag gtg gtg tgt aag aag tac cgg gga ttc acc atc ccc
 E   L   T   L   A   D   C   N   L   L   P   K   L   H   I   V   Q   V   V   C   K   K   Y   R   G   F   T   I   P
gag gcc ttc cgg gga gtg cat cgg tac ttg agc aat gcc tac gcc cgg gaa gaa ttc gct tcc acc tgt cca gat gat gag gag atc
 E   A   F   R   G   V   H   R   Y   L   S   N   A   Y   A   R   E   E   F   A   S   T   C   P   D   D   E   E   I
gag ctc gcc tat gag caa gtg gca aag gcc ctc aaa taa gga tcc gaa ttc gag ctc cgt cga caa gct tgc ggc cgc act cga gca
 E   L   A   Y   E   Q   V   A   K   A   L   K   -   G   S   E   F   E   L   R   R   Q   A   C   G   R   T   R   A
```

B.

Score	Expect	Method	Identities	Positives	Gaps
496 bits(1278)	1e-177	Compositional matrix adjust.	241/241(100%)	241/241(100%)	0/241(0%)

```
Query  1    MAEEQPQVELFVKAGSDGAKIGNCPFSQRLFMVLWLKGVTFNVTTVDTKRRTETVQKLCP  60
Sbjct  1    ............................................................  60

Query  61   GGQLPFLLYGTEVHTDTNKIEEFLEAVLCPPRYPKLAALNPESNTAGLDIFAKFSAYIKN  120
Sbjct  61   ............................................................  120

Query  121  SNPALNDNLEKGLLKALKVLDNYLTSPLPEEVDETSAEDEGVSQRKFLDGNELTLADCNL  180
Sbjct  121  ............................................................  180

Query  181  LPKLHIVQVVCKKYRGFTIPEAFRGVHRYLSNAYAREEFASTCPDDEEIELAYEQVAKAL  240
Sbjct  181  ............................................................  240

Query  241  K  241
Sbjct  241  .  241
```

Figure 22. Verification of the cDNA sequence for CLIC1-WT which had been inserted within the pET-28a vector. (A) The cDNA sequence of CLIC1-WT was translated to its corresponding primary amino acid sequence by using the TRANSLATE tool in Expasy (https://web.expasy.org/translate/) and sequencing confirmed that the nucleic acid sequence encoded for the hexa-histidine fusion-tag (▢), thrombin cleavage site (- - -) and wild-type sequence (●), with no further insertions or mutations. Indicated is also the start (●) and stop (●) codon. (B) Sequence verification was then performed by using the Basic Local Alignment Search Tool for Proteins (BLASTp) (https://blast.ncbi.nlm.nih.gov/Blast), which showed that based on there being 100% sequence identities, all of the amino acids from the query sequence were perfectly aligned with the amino acids from the subject sequence. As a result, this confirmed that the query sequence of CLIC1-WT shared 100% sequence similarity to the subject sequence for wild-type CLIC1 from *Homo sapiens* (NP_001274522.1).

A.

```
gcc ccc cyy yyt aga aaw aaa ttt tgt tta act tta aga agg aga tat acc atg ggc agc agc cat cat cwt cat cat cac agc agc
A   P   X   X   R   X   K   F   C   L   T   L   R   R   R   Y   T   M   G   S   S   H   H   X   H   H   H   S   S
ggc ctg gtg ccg cgc ggc agc cat atg gct gaa gaa caa ccg cag gtc gaa ttg ttc gtg aag gct ggc agt gat ggg gcc aag att
G   L   V   P   R   G   S   H   M   A   E   E   Q   P   Q   V   E   L   F   V   K   A   G   S   D   G   A   K   I
ggg aac gct cca ttc tcc cag aga ctg ttc atg gta ctg tgg ctc aag gga gtc acc ttc aat gtt acc acc gtt gac acc aaa agg
G   N   A   P   F   S   Q   R   L   F   M   V   L   W   L   K   G   V   T   F   N   V   T   T   V   D   T   K   R
cgg acc gag aca gtg cag aag ctg tgc cca ggg ggg cag ctc cca ttc ctg ctg tat ggc act gaa gtg cac aca gac acc aac
R   T   E   T   V   Q   K   L   C   P   G   G   Q   L   P   F   L   L   Y   G   T   E   V   H   T   D   T   N
aag att gag gaa ttt ctg gag gca gtg ctg tgc cct ccc agg tac ccc aag ctg gca gct ctg aac cct gag tcc aac aca gct ggg
K   I   E   E   F   L   E   A   V   L   C   P   P   R   Y   P   K   L   A   A   L   N   P   E   S   N   T   A   G
ctg gac ata ttt gcc aaa ttt tct gcc tac atc aag aat tca aac cca gca ctc aat gac aat ctg gag aag gga ctc ctg aaa gcc ctg
L   D   I   F   A   K   F   S   A   Y   I   K   N   S   N   P   A   L   N   D   N   L   E   K   G   L   L   K   A   L
aag gtt tta gac aat tac tta aca tcc ccc ctc cca gaa gaa gtg gat gaa acc agt gct gaa gat gaa ggt gtc tct cag agg aag ttt
K   V   L   D   N   Y   L   T   S   P   L   P   E   E   V   D   E   T   S   A   E   D   E   G   V   S   Q   R   K   F
ttg gat ggc aac gag ctc acc ctg gct gac tgc aac ctg ttg cca aag tta cac ata gta cag gtg gtg tgt aag aag tac cgg gga
L   D   G   N   E   L   T   L   A   D   C   N   L   L   P   K   L   H   I   V   Q   V   V   C   K   K   Y   R   G
ttc acc atc ccc gag gcc ttc cgg gga gtg cat cgg tac ttg agc aat gcc tac gcc cgg gaa gaa ttc gct tcc acc tgt cca gat
F   T   I   P   E   A   F   R   G   V   H   R   Y   L   S   N   A   Y   A   R   E   E   F   A   S   T   C   P   D
gat gag gag atc gag ctc gcc tat gag caa gtg gca aag gcc ctc aaa taa gga tcc gaa ttc gag ctc cgt cga caa gct tgc ggc
D   E   E   I   E   L   A   Y   E   Q   V   A   K   A   L   K   -   G   S   E   F   E   L   R   R   Q   A   C   G
```

B.

Score	Expect	Method	Identities	Positives	Gaps
493 bits(1270)	2e-176	Compositional matrix adjust.	240/241(99%)	240/241(99%)	0/241(0%)

```
Query  1    MAEEQPQVELFVKAGSDGAKIGNAPFSQRLFMVLWLKGVTFNVTTVDTKRRTETVQKLCP  60
Sbjct  1    .......................C.....................................  60

Query  61   GGQLPFLLYGTEVHTDTNKIEEFLEAVLCPPRYPKLAALNPESNTAGLDIFAKFSAYIKN  120
Sbjct  61   ............................................................  120

Query  121  SNPALNDNLEKGLLKALKVLDNYLTSPLPEEVDETSAEDEGVSQRKFLDGNELTLADCNL  180
Sbjct  121  ............................................................  180

Query  181  LPKLHIVQVVCKKYRGFTIPEAFRGVHRYLSNAYAREEFASTCPDDEEIELAYEQVAKAL  240
Sbjct  181  ............................................................  240

Query  241  K  241
Sbjct  241  .  241
```

C.

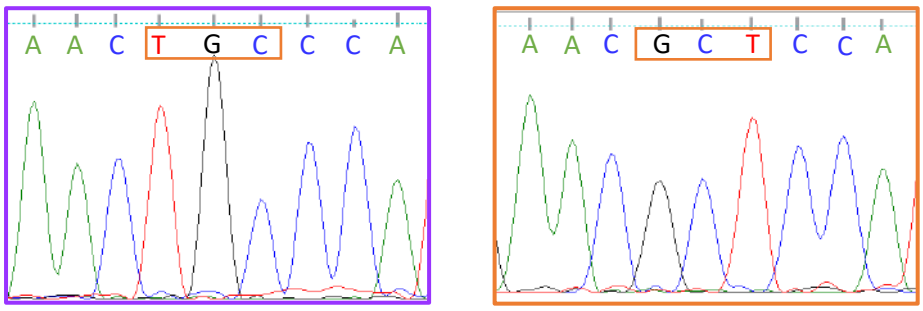


Figure 23. Verification of the cDNA sequence for CLIC1-C24A which had been inserted within the pET-28a vector. (A) Sequencing of the cDNA sequence of CLIC1-C24A confirmed that the nucleic acid sequence encoded for the hexa-histidine fusion-tag (●), thrombin cleavage site (- - -) and the C24A mutation sequence (□), with no further insertions or mutations. Indicated is also the start (●) and stop (●) codon. (B) Sequence verification showed 99% sequence identities, therefore confirming that apart from the C24A mutation (●), the sequence shared 100% sequence similarity to wild-type CLIC1 from *Homo sapiens* (NP_001274522.1). (C) Sanger sequencing verified that the codon sequence ⁵′TGC³′ at position 24 was mutated to ⁵′GCT³′, therefore verifying that Cys24 was successfully mutated to Ala24.

A.

```
ttc ccc tct aga taa ttt tgt tta act tta aga agg aga tat acc atg ggc agc agc cat cat cat cwt cat cac agc agc ggc ctg gtg
 F   P   S   R   -   F   C   L   T   L   R   R   R   Y   T   M   G   S   S   H   H   H   X   H   H   S   S   G   L   V
ccg cgc ggc agc cat atg ctc gag ccg ccg tac acc gtg gtc tat ttc cca gtt cga ggc cgc tgc gcg gcc ctg cgc atg ctg ctg
 P   R   G   S   H   M   L   E   P   P   Y   T   V   V   Y   F   P   V   R   G   R   C   A   A   L   R   M   L   L
gca gat cag ggc cag agc tgg aag gag gag gtg gtg acc gtg gag acg tgg cag gag ggc tca ctc aaa gcc tcc tgc cta tac ggg
 A   D   Q   G   Q   S   W   K   E   E   V   V   T   V   E   T   W   Q   E   G   S   L   K   A   S   C   L   Y   G
cag ctc ccc aag ttc cag gac gga gac ctc acc ctg tac cag tcc aat acc atc ctg cgt cac ctg ggc cgc acc ctt ggg ctc tat ggg
 Q   L   P   K   F   Q   D   G   D   L   T   L   Y   Q   S   N   T   I   L   R   H   L   G   R   T   L   G   L   Y   G
aag gac cag cag gag gca gcc ctg gtg gac atg gtg aat gac ggc gtg gag gac ctc cgc tgc aaa tac atc tcc ctc atc tac acc
 K   D   Q   Q   E   A   A   L   V   D   M   V   N   D   G   V   E   D   L   R   C   K   Y   I   S   L   I   Y   T
aac tat gag gcg ggc aag gat gac tat gtg aag gca ctg ccc ggg caa ctg aag cct ttt gag acc ctg ctg tcc cag aac cag gga
 N   Y   E   A   G   K   D   D   Y   V   K   A   L   P   G   Q   L   K   P   F   E   T   L   L   S   Q   N   Q   G
ggc aag acc ttc att gtg gga gac cag atc tcc ttc gct gac tac aac ctg ctg gac ttg ctg ctg atc cat gag gtc cta gcc cct ggc
 G   K   T   F   I   V   G   D   Q   I   S   F   A   D   Y   N   L   L   D   L   L   L   I   H   E   V   L   A   P   G
tgc ctg gat gcg ttc ccc ctg ctc tca gca tat gtg ggg  cgc ctc agt gcc cgg ccc aag ctc aag gcc ttc ctg gcc tcc cct gag tac
 C   L   D   A   F   P   L   L   S   A   Y   V   G   R   L   S   A   R   P   K   L   K   A   F   L   A   S   P   E   Y
gtg aac ctc ccc atc aat ggc aac ggg aaa cag tga ggg gga tcc ggc tgc taa caa agc ccg aaa gga agc tga gtt ggc yks ctg
 V   N   L   P   I   N   G   N   G   K   Q   -   G   G   S   G   C   -   Q   S   P   K   G   S   -   V   G   X   L
```

B.

Score	Expect	Method	Identities	Positives	Gaps
429 bits(1104)	5e-152	Compositional matrix adjust.	209/209(100%)	209/209(100%)	0/209(0%)

```
Query  4    PPYTVVYFPVRGRCAALRMLLADQGQSWKEEVVTVETWQEGSLKASCLYGQLPKFQDGDL  63
Sbjct  2    ............................................................  61

Query  64   TLYQSNTILRHLGRTLGLYGKDQQEAALVDMVNDGVEDLRCKYISLIYTNYEAGKDDYVK  123
Sbjct  62   ............................................................  121

Query  124  ALPGQLKPFETLLSQNQGGKTFIVGDQISFADYNLLDLLLIHEVLAPGCLDAFPLLSAYV  183
Sbjct  122  ............................................................  181

Query  184  GRLSARPKLKAFLASPEYVNLPINGNGKQ  212
Sbjct  182  ............................  210
```

Query sequence: MLEPPYTV
NP_000843.1: M··PPYTV

Figure 24. Verification of the cDNA sequence for hGSTP1-1 which had been inserted within the pET-15b vector. (A) The cDNA sequence of hGSTP1-1 was translated to its corresponding primary amino acid sequence by using the TRANSLATE tool in Expasy (https://web.expasy.org/translate/) and sequencing confirmed that the nucleic acid sequence encoded for the hexa-histidine fusion-tag (●), thrombin cleavage site (- - -) and wild-type sequence (●), with no further insertions or mutations. Indicated is also the start (●) and stop (●) codon. (B) Sequence verification was then performed by using the Basic Local Alignment Search Tool for Proteins (BLASTp) (https://blast.ncbi.nlm.nih.gov/Blast), which showed that based on there being 100% sequence identities, all of the amino acids from the query sequence were perfectly aligned with the amino acids from the subject sequence. As a result, this confirmed that the query sequence of hGSTP1-1 shared 100% sequence similarity from the fourth amino acid to the subject sequence for wild-type GSTP1-1 from *Homo sapiens* (NP_000843.1). However, despite having 100% sequence identities, it was noted that the query sequence contained an additional two residues (Leu2 and Asp3) subsequent to the first methionine residue.

A.

```
ccc cty cct wrg aaw taa ttt tgt tta act tta aga agg aga tat acc atg ggc agc agc cay cmt ywt cat cat cac agc agc ggc
 P  X  P  X  X  -  F  C  L  T  L  R  R  R  Y  T  M  G  S  S  X  X  X  H  H  H  S  S  G
ctg gtg ccg cgc:ggc agc cat atg gct gtc gct gcc aac aaa cgt tcg gta atg acg ctg ttt tcc ggt cct act gac atc tat agc cat
 L  V  P  R :G  S  H  M  A  V  A  A  N  K  R  S  V  M  T  L  F  S  G  P  T  D  I  Y  S  H
cag gtc cgc att gtg ctg gct gag aaa ggt gta agt ttc gag atc gaa cac gtg gaa aag gac aat ccg cct cag gat ctg att gac
 Q  V  R  I  V  L  A  E  K  G  V  S  F  E  I  E  H  V  E  K  D  N  P  P  Q  D  L  I  D
ctc aac ccg aat cag agc gtt ccg acc ctg gtg gat cgt gag ctg acc ctg tgg gaa tct cgc atc att atg gaa tat ctg gat gag cgt
 L  N  P  N  Q  S  V  P  T  L  V  D  R  E  L  T  L  W  E  S  R  I  I  M  E  Y  L  D  E  R
ttc ccg cat ccg cca ctg atg cct gtt tac ccg gta gct cgc ggt gaa agc cgt ctg tac atg cat cgc atc gaa aaa gac tgg tac acg
 F  P  H  P  P  L  M  P  V  Y  P  V  A  R  G  E  S  R  L  Y  M  H  R  I  E  K  D  W  Y  T
ctg atg aac acc atc atc aac ggt tca gct tct gaa gca gat gcc gca cgt aag caa ctg cgc gaa gaa ctg ctg gcg att gcg ccg
 L  M  N  T  I  I  N  G  S  A  S  E  A  D  A  A  R  K  Q  L  R  E  E  L  L  A  I  A  P
gtc ttc ggt cag aag ccg tac ttc ctg agc gat gag ttc agc ctg gtc gat tgc tat ctt gct ccg ctg ctg tgg cgt ctg ccg caa ctg
 V  F  G  Q  K  P  Y  F  L  S  D  E  F  S  L  V  D  C  Y  L  A  P  L  L  W  R  L  P  Q  L
ggc atc gag ttc agc ggc ccg ggt gcg aaa gag ctg aaa ggc tat atg acc cgc gtc ttt gag cgt gac tct ttc ctt gct tct tta act
 G  I  E  F  S  G  P  G  A  K  E  L  K  G  Y  M  T  R  V  F  E  R  D  S  F  L  A  S  L  T
gaa gca gaa cgt gaa atg cgt ctg ggc cgg agt taa gcg gcc gca ctc gag cac cac cac cac cac cac tga gat ccg gct gct
 E  A  E  R  E  M  R  L  G  R  S  -  A  A  A  L  E  H  H  H  H  H  H  -  D  P  A  A
```

B.

Score	Expect	Method	Identities	Positives	Gaps
437 bits(1124)	4e-155	Compositional matrix adjust.	212/212(100%)	212/212(100%)	0/212(0%)

```
Query  1    MAVAANKRSVMTLFSGPTDIYSHQVRIVLAEKGVSFEIEHVEKDNPPQDLIDLNPNQSVP  60
Sbjct  1    ............................................................  60

Query  61   TLVDRELTLWESRIIMEYLDERFPHPPLMPVYPVARGESRLYMHRIEKDWYTLMNTIING  120
Sbjct  61   ............................................................  120

Query  121  SASEADAARKQLREELLAIAPVFGQKPYFLSDEFSLVDCYLAPLLWRLPQLGIEFSGPGA  180
Sbjct  121  ............................................................  180

Query  181  KELKGYMTRVFERDSFLASLTEAEREMRLGRS  212
Sbjct  181  ................................  212
```

Figure 25. Verification of the cDNA sequence for EcSspA which had been inserted within the pET-28a vector. (A) The cDNA sequence of EcSspA was translated to its corresponding primary amino acid sequence by using the TRANSLATE tool in Expasy (https://web.expasy.org/translate/) and sequencing confirmed that the nucleic acid sequence encoded for the hexa-histidine fusion-tag (●), thrombin cleavage site (- - -) and wild-type sequence (●), with no further insertions or mutations. Indicated is also the start (●) and stop (●) codon. (B) Sequence verification was then performed by using the Basic Local Alignment Search Tool for Proteins (BLASTp) (https://blast.ncbi.nlm.nih.gov/Blast), which showed that based on there being 100% sequence identities, all of the amino acids from the query sequence were perfectly aligned with the amino acids from the subject sequence. As a result, this confirmed that the query sequence of EcSspA shared 100% sequence similarity to the subject sequence for wild-type SspA from *E.coli* (WP_000257293.1).

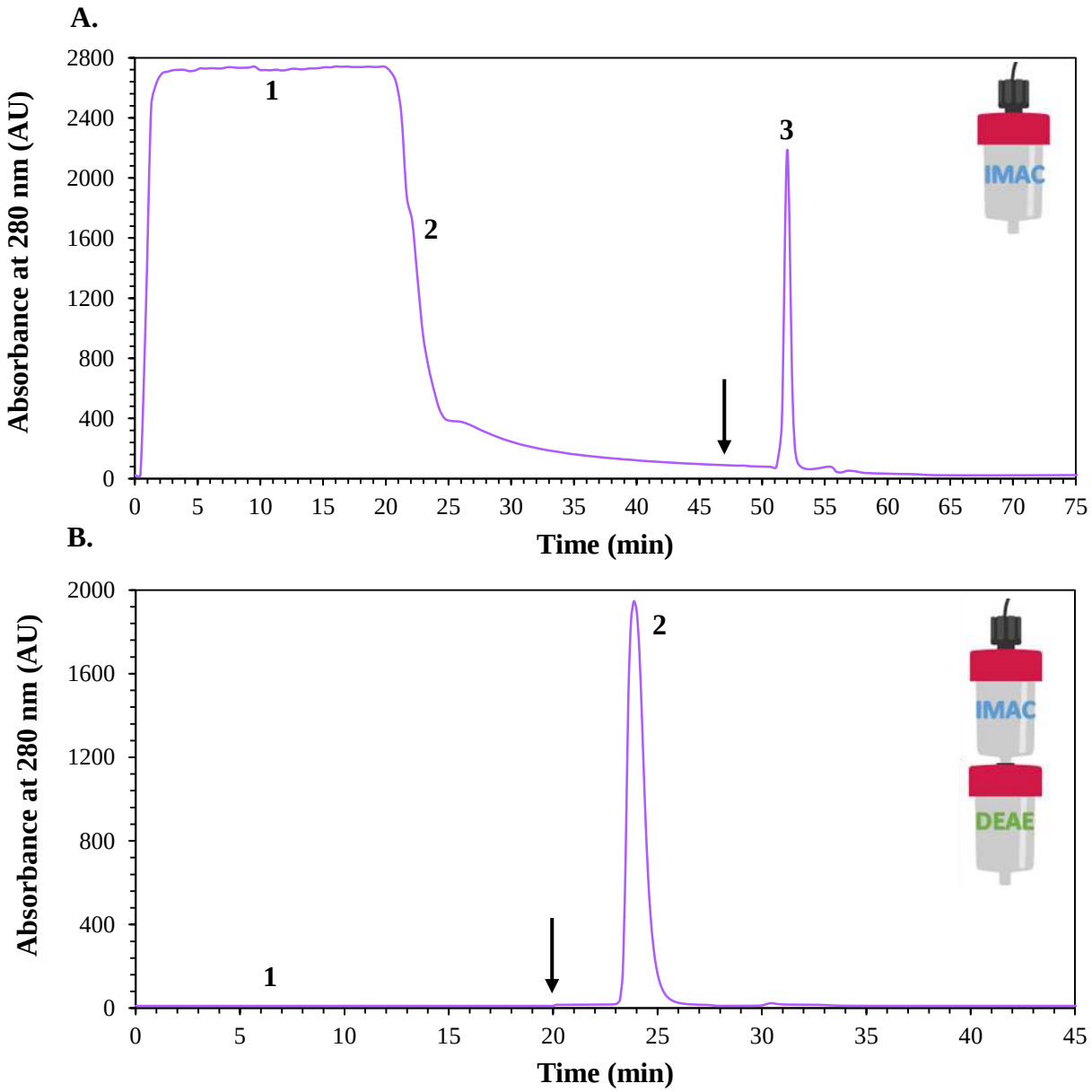


Figure 26. Purification of CLIC1-WT by means of a Ni^{2+}- IMAC and DEAE column. (A1) The supernatant of homogenised T7 Express *E.coli* cells was loaded onto a Ni^{2+}-IMAC column at a flow rate of 5 $ml.min^{-1}$ which resulted in an increased absorbance at 280 nm due to various contaminant proteins which did not bind to the column passing through. (A2) Any remaining unbound contaminant and fusion protein were removed from the IMAC column by using the equilibration buffer and this resulted in the decrease in absorbance. (A3) Once the absorbance readings had decreased and became consistent the bound fusion protein was eluted from the column through a single step elution (→). (B1) Subsequent to thrombin cleavage, the sample was loaded onto a conjugated Ni^{2+}-IMAC/DEAE column at a flow rate of 5 $ml.min^{-1}$. Initially, there was no change in absorbance at 280 nm given that the histidine-tag and contaminant proteins bound to the IMAC column, while CLIC1-WT had bound to the DEAE column. Once the sample was loaded onto the column and washed with equilibration buffer, CLIC1-WT was eluted from the column through a single step elution. *See section 2.2.3 for the experimental details pertaining to this method.*

.

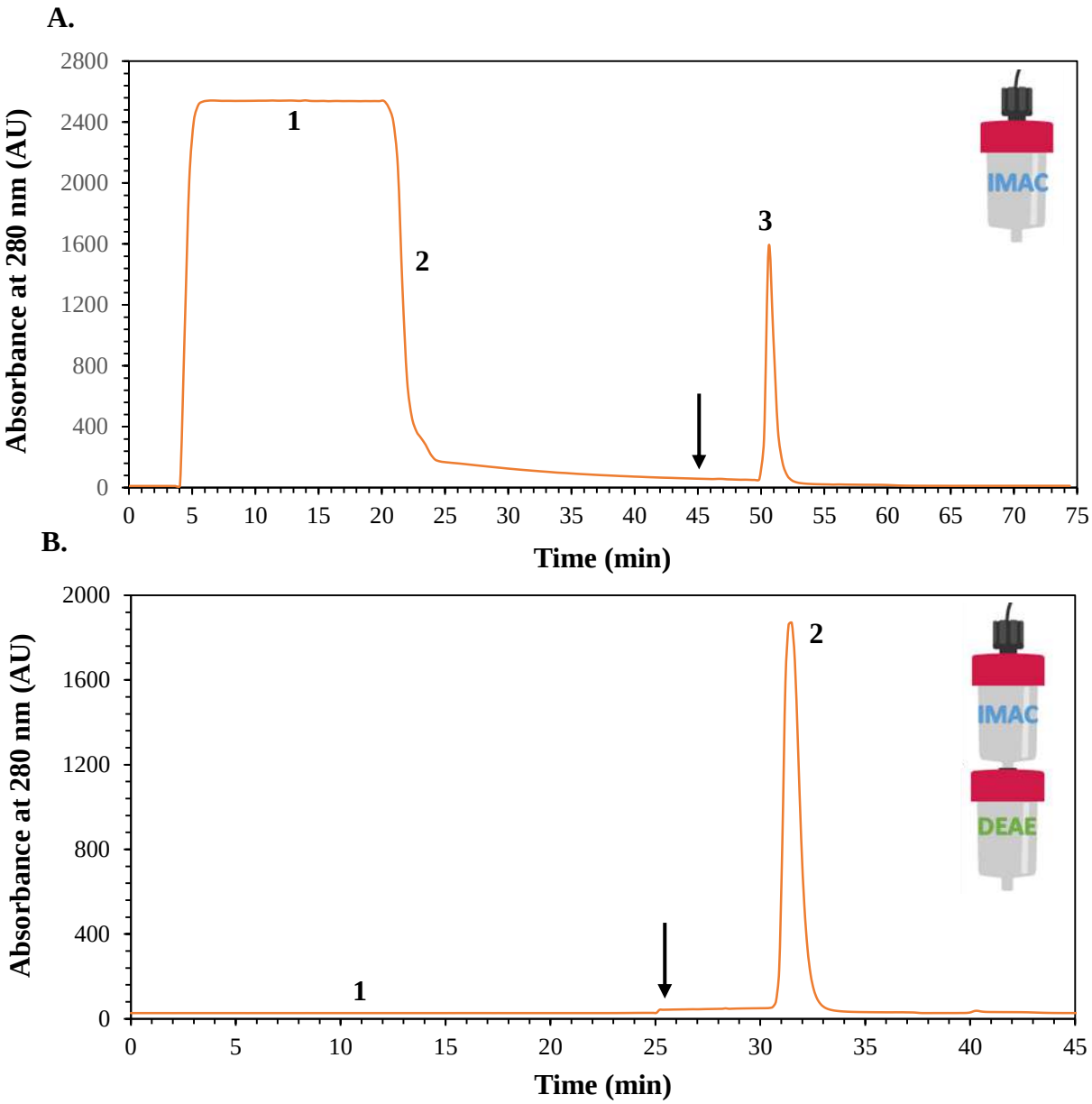


Figure 27. Purification of CLIC1-C24A by means of a Ni^{2+}- IMAC and DEAE column. (A1) The supernatant of homogenised T7 Express *E.coli* cells was loaded onto a Ni^{2+}-IMAC column at a flow rate of 5 $ml.min^{-1}$ which resulted in an increased absorbance at 280 nm due to various contaminant proteins which did not bind to the column passing through. (A2) Any remaining unbound contaminant and fusion protein were removed from the IMAC column by using the equilibration buffer and this resulted in the decrease in absorbance. (A3) Once the absorbance readings had decreased and became consistent the bound fusion protein was eluted from the column through a single step elution (→). (B1) Subsequent to thrombin cleavage, the sample was loaded onto a conjugated Ni^{2+}-IMAC/DEAE column at a flow rate of 5 $ml.min^{-1}$. Initially, there was no change in absorbance at 280 nm given that the histidine-tag and contaminant proteins bound to the IMAC column, while CLIC1-C24A had bound to the DEAE column. Once the sample was loaded onto the column and washed with equilibration buffer, CLIC1-C24A was eluted from the column through a single step elution. *See section 2.2.3 for the experimental details pertaining to this method.*

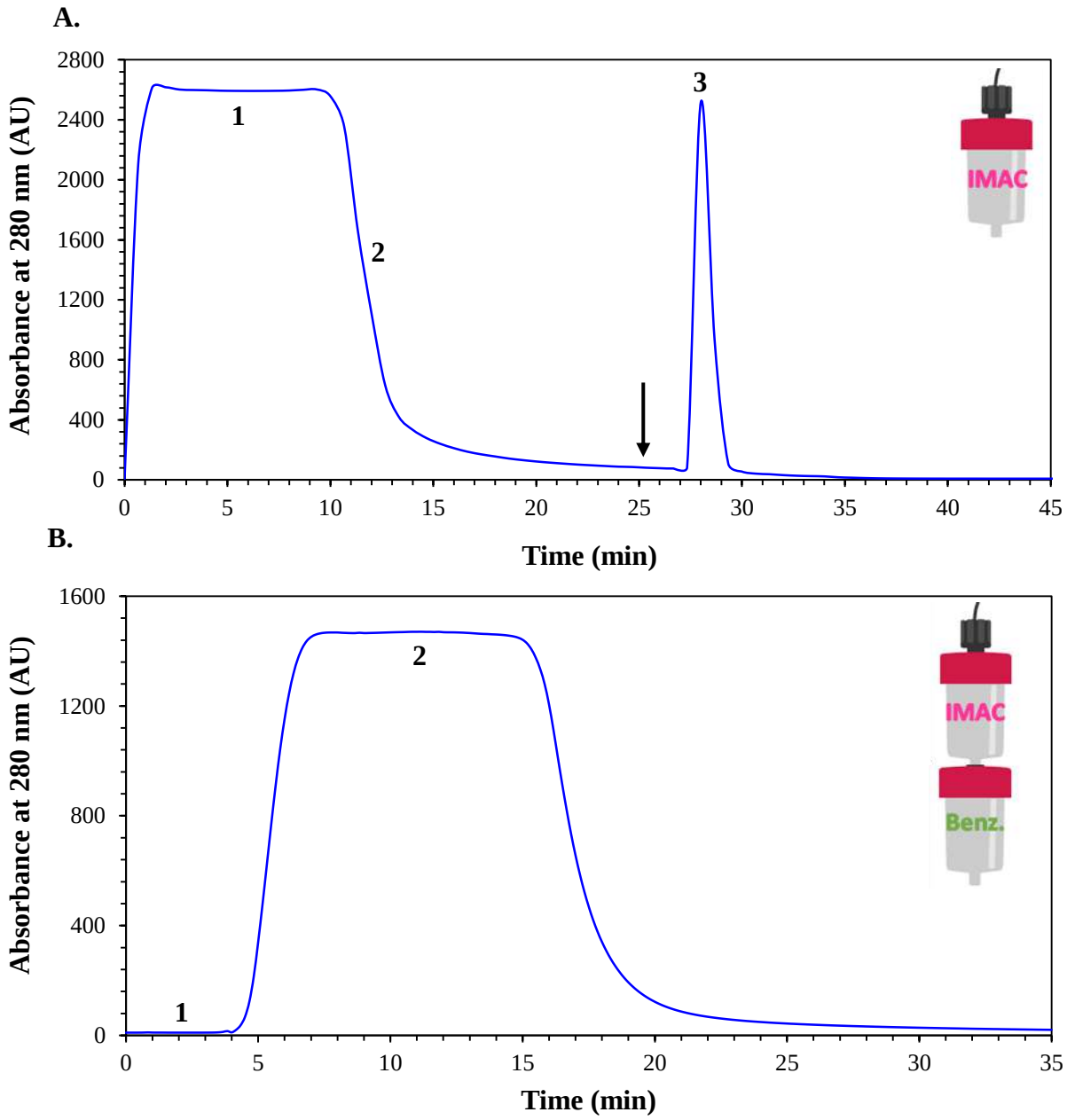


Figure 28. Purification of hGSTP1-1 by means of a Co^{2+}- IMAC and benzamidine column. (A1) The supernatant of homogenised T7 Express *E.coli* cells was loaded onto a Co^{2+}-IMAC column at a flow rate of 5 $ml.min^{-1}$ which resulted in an increased absorbance at 280 nm due to various contaminant proteins which did not bind to the column passing through. (A2) Any remaining unbound contaminant and fusion protein were removed from the IMAC column by using the equilibration buffer and this resulted in the decrease in absorbance. (A3) Once the absorbance readings had decreased and became consistent the bound fusion protein was eluted from the column through a single step elution (→). (B1) Subsequent to thrombin cleavage, the sample was loaded onto a conjugated Co^{2+}-IMAC/benzamidine column at a flow rate of 5 $ml.min^{-1}$. Initially, the histidine-tag and contaminant proteins bound to the IMAC column, whereas thrombin bound to the benzamidine column. However, hGSTP1-1 did not bind to either column and was instead collected in the flow-through, therefore resulting in an increased absorbance at 280 nm. *See section 2.2.3 for the experimental details pertaining to this method.*

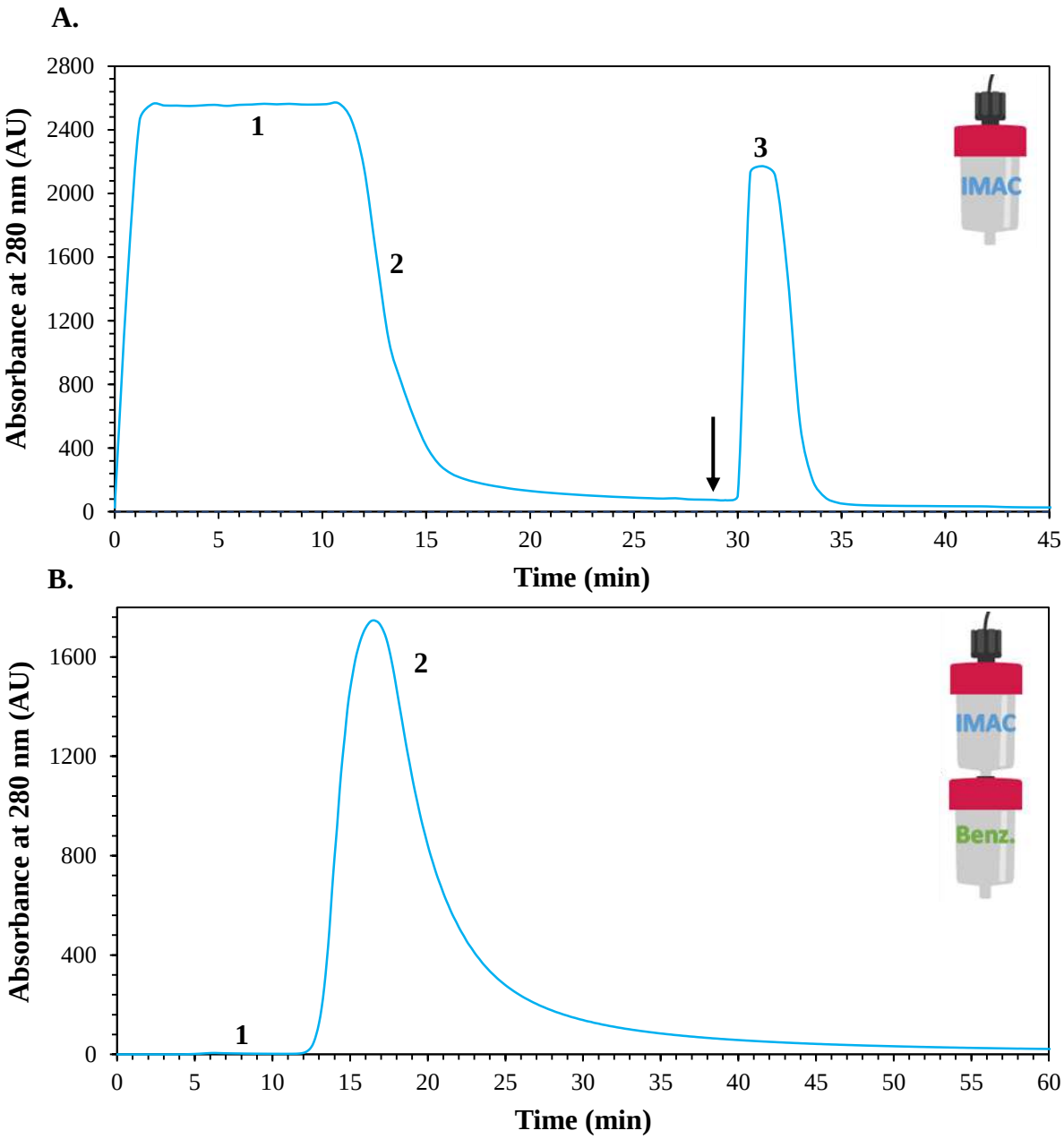


Figure 29. Purification of EcSspA by means of a Ni^{2+}- IMAC and benzamidine column. (A1) The supernatant of homogenised T7 Express *E.coli* cells was loaded onto a Ni^{2+}-IMAC column at a flow rate of 5 $ml.min^{-1}$ which resulted in an increased absorbance at 280 nm due to various contaminant proteins which did not bind to the column passing through. (A2) Any remaining unbound contaminant and fusion protein were removed from the IMAC column by using the equilibration buffer and this resulted in the decrease in absorbance. (A3) Once the absorbance readings had decreased and became consistent the bound fusion protein was eluted from the column through a single step elution (→). (B1) Subsequent to thrombin cleavage, the sample was loaded onto a conjugated Ni^{2+}-IMAC/benzamidine column at a flow rate of 5 $ml.min^{-1}$. Initially, the histidine-tag and contaminant proteins bound to the IMAC column, whereas thrombin bound to the benzamidine column. However, EcSspA did not bind either column and was instead collected in the flow-through, therefore resulting in an increased absorbance at 280 nm. *See section 2.2.3 for details pertaining to the composition of the buffers used.*

3.1.3 Assessing the purity of the target proteins through SDS-PAGE

The purity of the protein samples were assessed by means of SDS-PAGE as described in section 2.2.4. It was noted that the fractions collected from the IMAC column for both the CLIC1-WT and CLIC1-C24A fusion proteins were initially not pure given the presence of other contaminant proteins (Figure S1 and S2). Consequently, in order to obtain a pure protein sample, the fusion-tag was cleaved from the protein to ensure that the hexa-histidine-tag and other contaminant proteins would rebind to the IMAC column, while allowing the purified target protein to pass through the column (Figure S1 and S2). In relation to hGSTP1-1 and EcSspA, although the fractions collected from the IMAC column for both proteins was ≥ 95% pure based on densitometry (Figure 30; Table S2), considering that comparative studies were to be done between all the target proteins, the hexa-histidine-tag was also removed from both hGSTP1-1 and EcSspA. It was therefore important to note that the final purified target proteins had an additional two amino acid residues from the thrombin cleavage site at the N-terminus, namely glycine and serine.

The overall SDS-PAGE purification profile (Figure S1-S4) for each of the target proteins indicated that while the majority of the protein was soluble and found to be present in the supernatant of the cell lysate, some of the target protein was incorporated as inclusion bodies by being present within the pellet (insoluble) sample of the cell lysate. This could be expected considering that during the high level expression of heterologous proteins in *E.coli*, the high local concentration of polypeptide chains emerging from ribosomes along with inefficient folding and the unavailability of chaperones can lead to the aggregation of proteins (Upadhyay *et al*., 2012). However, the isolation of bioactive protein from inclusion bodies is often very cumbersome, but considering that the protein concentration within the supernatant was high enough for downstream studies the presence of the target proteins within the pellet (insoluble) sample was not of much concern. Furthermore, when loaded onto the IMAC column, the fusion-protein was efficiently isolated from the supernatant given the absence of the target protein in the flow-through from the supernatant that was loaded onto the IMAC column as well as in the flow-through from the wash step that was done to remove any unbound contaminant and fusion-protein from the column (Figure S1-S4).

The purity of the purified and concentrated target proteins was noted to be ≥ 94% based on densitometry analysis (Figure 30). Furthermore, the predicated molecular weight of the target

proteins were 2.5-3.5 kDa higher than the expected molecular weight recorded in NCBI (https://www.ncbi.nlm.nih.gov/) (Figure 30), however, a slight deviation between the molecular weight predicted from SDS-PAGE and the expected molecular weight can be expected. Although the work by Reynolds and Tanford, 1970, indicated that the maximum SDS-binding levels to a protein was 1.4 g SDS.g protein^{-1}, this saturation limit has subsequently been refined to 0.7-2 g SDS.g protein^{-1} depending on the amino acid composition of a protein. Consequently, the relationship between the amino acid composition and the SDS-binding levels to a protein could have contributed to the anomalous behaviour in the electrophoretic mobility of the target proteins during SDS-PAGE (Rath *et al*., 2009).

A. Protein type	Main protein band intensity (Int.) (AU)	Relative band abundance (%)	Molecular weight (kDa)	
			Predicted	Expected
CLIC1-WT (△)	21 813 273	98.4	28.7	26.2
CLIC1-C24A (△)	23 093 400	97.3	29.7	26.2
hGSTP1-1 (△)	51 800 805	94.0	25.7	23.4
EcSspA (△)	25 144 056	97.2	26.7	24.3

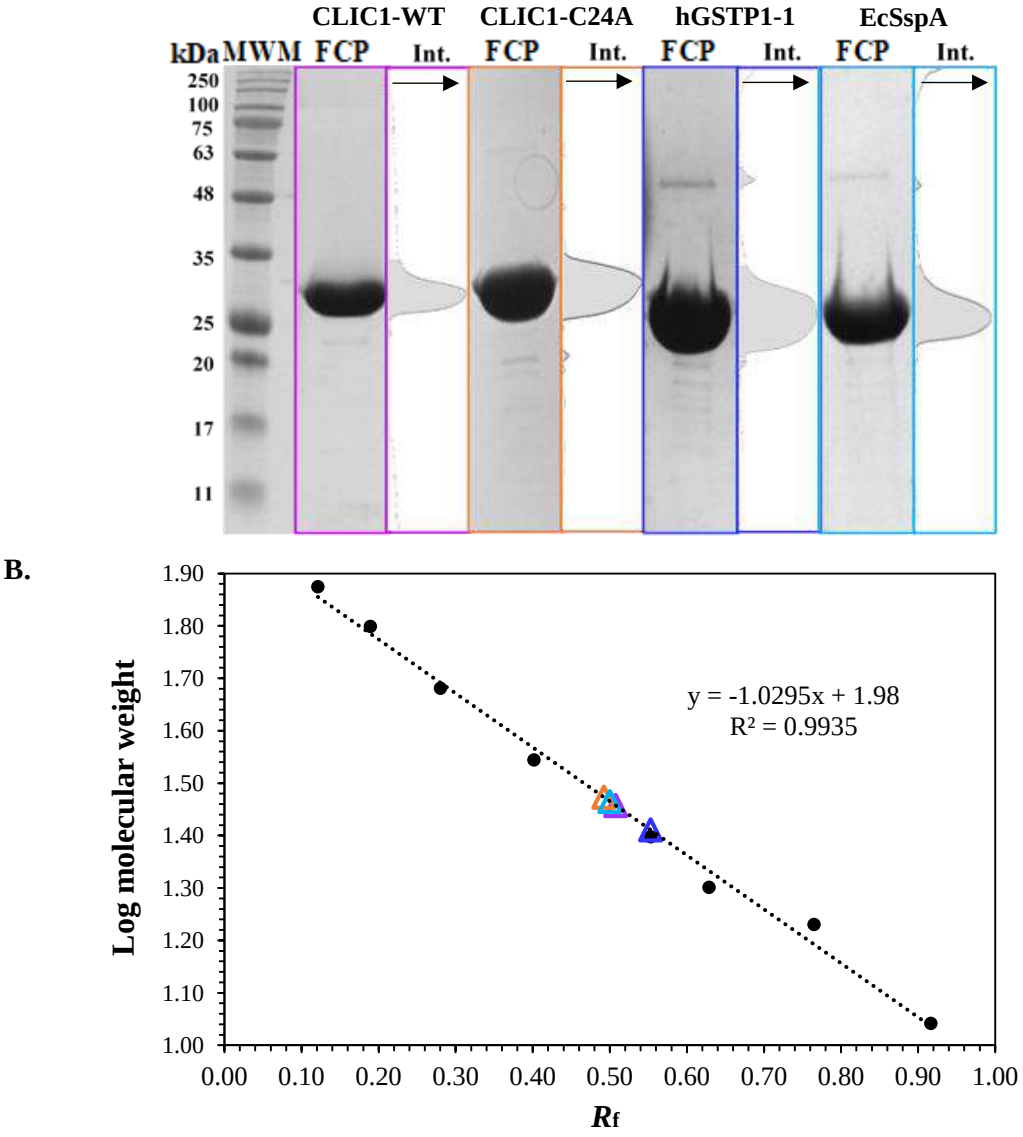


Figure 30. Verifying the purity of the purified target proteins by SDS-PAGE. (A) All gel images were analysed in the Image Lab™ software and based on the intensity (Int.) for the main protein band of the final concentrated protein (FCP), it was observed that the target proteins were ≥ 94% pure. (B) The predicted molecular weight for each target protein was determined from a calibration curve that was constructed by plotting the R_f of the protein molecular weight marker (MWM) standards against the logarithm of their molecular weight. When constructing the standard curve, the first four bands from the protein MWM standards were removed considering that their migration on the gel did not aid in their separation and hence they skewed the linear regression model that was used to fit the data. *See section 2.2.4 for the experimental details pertaining to this method.*

3.1.4 Determining the concentration of the purified target proteins

The concentration of the purified target proteins was verified by means of UV-Vis absorbance spectroscopy. Absorbance is a particularly appropriate scale for measuring the interactions between light and molecules in solution, consequently an absorbance spectrum of a protein sample is able to provide not only information on the concentration of the sample based of the Beer-Lamberts Law, but it can also be used to assess the purity of the sample from nucleic acids and also measure the presence of protein aggregates within solution (Sheehan, 2009; Wilson and Walker, 2010).

The average protein yield per a purification for all the target proteins was often 150-250 µM (4-6 $mg.ml^{-1}$) with a volume of 10-15 ml and this proved to be sufficient to perform numerous downstream assays, especially for ITC which required large volumes of concentrated protein per an experiment. A protein sample that is free of nucleic acids should have an A_{280}/A_{260} ratio which is $\geq$ 1.7 (Kelly *et al*., 2005), therefore the samples for all the target proteins were considered to be free of nucleic acid contamination (Figure 31). However, although disulfide bonds between two cysteine residues can also contribute to the absorbance signal at 260 nm (Schmid, 2001), the absorbance at 260 nm for the recorded absorbance spectra was not attributed to disulfide bonds considering that the protein samples were dialysed in a reducing environment brought about by TCEP (section 2.2.2.3). Instead, given that the maximum absorbance for phenylalanine residues is at 257 nm, the phenylalanine residues present in CLIC1-WT/C24A (F = 12), hGSTP1-1 (F = 14) and EcSspA (F = 18) would have contributed to the overall absorbance at 260 nm. However, the absorbance seen between 240-260 nm can be attributed to the peptide bonds within the target proteins (Figure 31) (Schmid, 2001).

Furthermore, when monitoring the absorbance signal above 320 nm the presence of protein aggregates would give rise to Rayleigh scatter that would be visible by a significant slope in the region from 320-500 nm (Raynal *et al*., 2014; Sheehan, 2009). However, the absorbance readings above 320 nm, and in particular 340 nm, where in the range of 0.002-0.004 AU for the target protein samples. Therefore, given the low value of these readings, combined with the flat trajectory for this region on the absorbance profile (Figure 31), it was concluded that there was little to no aggregation within the target protein samples and consequently the protein concentration could be determined more accurately.

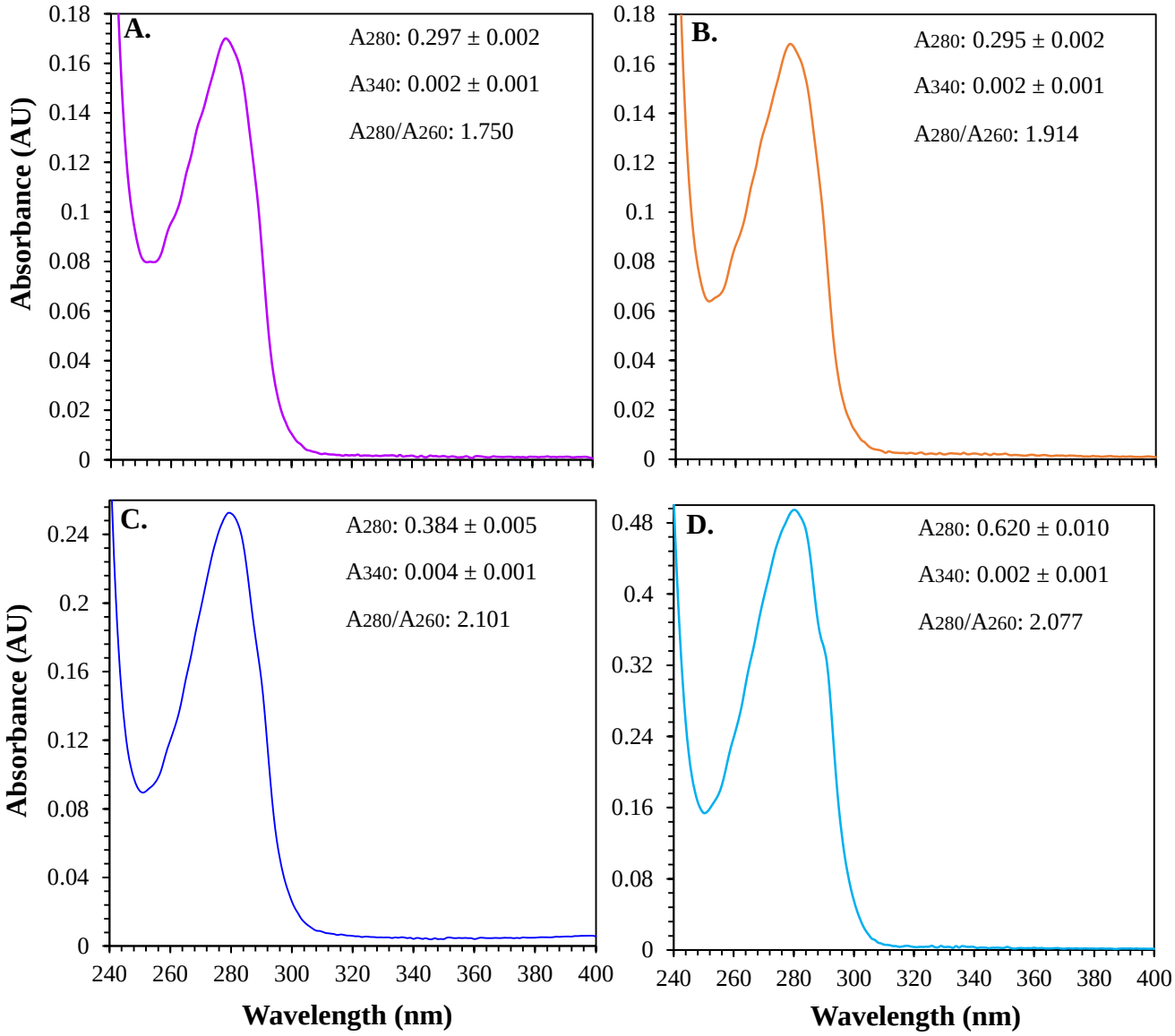


Figure 31. Determining the protein concentration of the purified target proteins. An absorbance spectra ranging from 240-400 nm was measured for (A) CLIC1-WT, (B) CLIC1-C24A, (C) hGSTP1-1 and (D) EcSspA. The average protein yield per a purification for all target proteins was often 150-250 µM (4-6 mg.ml^{-1}) and this proved to be sufficient to perform numerous downstream assays. The samples for all the target proteins were also free of nucleic acid contamination given that the A_{280}/A_{260} ratio was ≥1.7. However, although disulfide bonds between two cysteine residues can also contribute to the absorbance signal at 260 nm, the absorbance at 260 nm reading for the recorded absorbance spectra was not attributed to disulfide bonds considering that the protein samples were dialysed in a reducing environment brought about by TCEP. Instead, the phenylalanine residues present within each target protein would have contributed to the overall absorbance at 260 nm. However, the absorbance seen between 240-260 nm can be attributed to the peptide bonds within the target proteins. Furthermore, the absorbance readings above 320 nm, and in particular 340 nm, where in the range of 0.002-0.004 AU for the target protein samples. Therefore, given the low value of these readings, combined with the flat trajectory for this region on the absorbance profile, it was concluded that there was little to no aggregation within the target protein samples and consequently the concentration of the target proteins could be determined more accurately. *See section 2.2.5 for the experimental details pertaining to this method.*

3.1.5 Evaluating the oligomeric state and purity of the target proteins by SE-HPLC

The hydrodynamic radius, or otherwise Stokes radius, describes the radius of a spherical particle that would have the same diffusion coefficient in a solution with the same viscosity. The hydrodynamic radius factors not only the volume of a particle, but it also incorporates the volume of solvent associated with the particle. Therefore, in the case of protein molecules which are not strictly spherical, the hydrodynamic radius will incorporate both the shape, conformation and dynamics of the protein molecule (Sheehan, 2009; Wilson and Walker, 2010). Consequently, the hydrodynamic volume can provide information not only on the oligomeric state of a protein, but also on the purity of a protein sample should the hydrodynamic volume of the contaminant proteins be dissimilar, characteristics which were directly assessed by SE-HPLC for the target protein samples.

The elution time for both CLIC1-WT and CLIC1-C24A was reported to be 17.26 ± 0.01 min and consequently the molecular weight was predicted to be 40 kDa (Figure 32; Table S3). Although the molecular weight for CLIC1-WT/C24A was expected to be ~26 kDa, this discrepancy can be accounted for by the anomalous behaviour of CLIC1-WT/C24A with the stationary phase of the SEC column. In particular, CLIC1-WT/C24A contains a very dynamic and negatively-charged loop region (Pro147-Gln164) (Cromer *et al*., 2002; Harrop *et al*., 2001) and consequently this would increase its hydrodynamic volume. Furthermore, it appeared that the molecular weight of CLIC1-WT/C24A was closer to the molecular weight of hGSTP1-1, a dimeric protein. Yet, when assessing the accessible molecular surface area (ASA) of CLIC1-WT (PDB: 1K0M) it was noted to be greater than hGSTP1-1 (PDB: 2GSS) (Figure 33), consequently the increased hydrodynamic volume of CLIC1-WT/C24A accounted for its anomalous behaviour on the SEC column.

The SE-HPLC profile for both CLIC1-WT and CLIC1-C24A indicated the presence of four peaks, however, the fourth peak was noted to be an artefact during each run and did not contain any protein based on SDS-PAGE (Figure 32A) and UV-Vis absorbance spectroscopy. However, the predicted molecular weight for peak one and peak two was ~187 kDa and ~78 kDa, respectively, yet because both peaks represented CLIC1 (Figure 32A), the two peaks could therefore possibly account for a higher ordered oligomer (or soluble aggregate) and a non-specific covalently-linked dimer of CLIC1, respectively. Yet, it was estimated based on the integrated area under each peak that the protein species within the third and main peak

represented 95.7 ± 1.3% of the sample (Figure 32A; Table S3), therefore, the sample of CLIC1-WT and CLIC1-C24A was considered to be sufficiently pure for downstream studies.

The third and main peak, however, did not follow a perfect Gaussian distribution and therefore the peak was not completely symmetrical, instead based on a tailing factor (*S*) of S = 1.65 ± 0.03 the peak represented a slight tailing peak (Figure 32A; Table S3). The profile of a peak is important in indicating efficient separation, however, tailing peaks can create issues with resolution, integration and reproducibility. There are numerous reasons that contribute to the tailing of a peak, however, one of the most notable reasons is the composition of the mobile phase which can cause unwanted secondary interactions between an analyte and the stationary phase. The resin of the TSK Gel SuperSW2000 SEC column was silica-based, consequently when the pH of the mobile phase is more basic, hydrogen bond interactions can form between the polar functional groups of the analyte and silanol groups of the stationary phase. To overcome this one can decrease the pH and/or increase the ionic strength of the mobile phase (Harris, 2010), yet for reasons pertaining to the effect the local environment has on protein stability, optimising the tailing affect can present various challenges and was not done in this study (Harris, 2010).

The efficiency of separation can further be quantified by the resolution of the SE-HPLC profile. There are two main factors which contribute to the resolution of separation (R_s): (i) the difference in elution time between two peaks and (ii) the peak broadness. In chromatography a resolution of ≥ 1.5 is considered desirable for quantitative analysis (Harris, 2010). Yet, the separation of CLIC1-WT and CLIC1-C24A yielded a resolution of R_s = 0.58 ± 0.04 (Table S3), indicating that although the protein species within the third peak occupied the majority of the sample, the separation of the protein species between peak two and peak three was not that efficient given that there was an overlap in their elution time. Although this could be overcome by changing the composition of the mobile phase, for reasons stated previously this was not done in this study.

In relation to both hGSTP1-1 and EcSspA, the SE-HPLC profile indicated the elution time was 23.89 ± 0.02 min and 24.72 ± 0.08 min, respectively and consequently their molecular weight was predicted to be ~47 kDa and ~30 kDa, respectively. The molecular weight for hGSTP1-1 correlated with the expected molecular weight of ~46 kDa, yet the predicted molecular weight

for EcSspA was 19 kDa smaller than the expected molecular weight of ~48 kDa (Figure 32; Table S3). As was the case with CLIC1-WT/C24A, EcSspA showed anomalous behaviour on the SEC column such that its hydrodynamic volume appeared to be smaller than expected. Yet, upon observing the ASA and volume of the crystal structure of YpSspA, it was expected that EcSspA would also be dimeric given their high sequence similarity.

The tailing factors for hGSTP1-1 and EcSspA were reported to be S = 1.35 ± 0.00 and S = 2.16 ± 0.62, respectively, therefore indicating that EcSspA had a larger tailing peak which in turn correlated with the anomalous behaviour noted by EcSspA on a SEC column (Table S3). This anomalous behaviour could possibly be attributed to the fact that present within EcSspA is a hydrophobic surface pocket that was suggested to play a key role in protein-protein interactions (Figure 33) (Hansen, 2005), therefore should this site be involved in any secondary interactions with the stationary phase of the column, this would retard the protein to subsequently influence its retention time and consequently molecular weight. Yet, depending on the stability of EcSspA at higher ionic strengths, one could alter the ionic strength of the mobile phase to prevent these secondary interactions. However, based on the integrated area under the first and main peak for both hGSTP1-1 and EcSspA it was estimated that the protein species represented 99.20 ± 0.19% and 98.77 ± 0.70% of the sample, respectively (Table S3). Therefore, the samples of hGSTP1-1 and EcSspA were considered to be sufficiently pure for downstream studies.

Therefore, based on the SE-HPLC data it was concluded that hGSTP1-1 was dimeric. However, in relation to EcSspA, considering its structural homology to YpSspA and that the predicted molecular weight was larger than the molecular weight for an individual monomer (~24 kDa), it was suggested that EcSspA was dimeric. Yet, although the predicted molecular weight for CLIC1-WT/C24A was also shown to be greater than the molecular weight for an individual monomer, the crystal structure of CLIC1-WT and its associated mutants have confirmed that CLIC1 is monomeric. Therefore, the irregularity in the predicted molecular size for both EcSspA and CLIC1-WT/C24A could be attributed to their anomalous behaviour when passing through the SEC column.

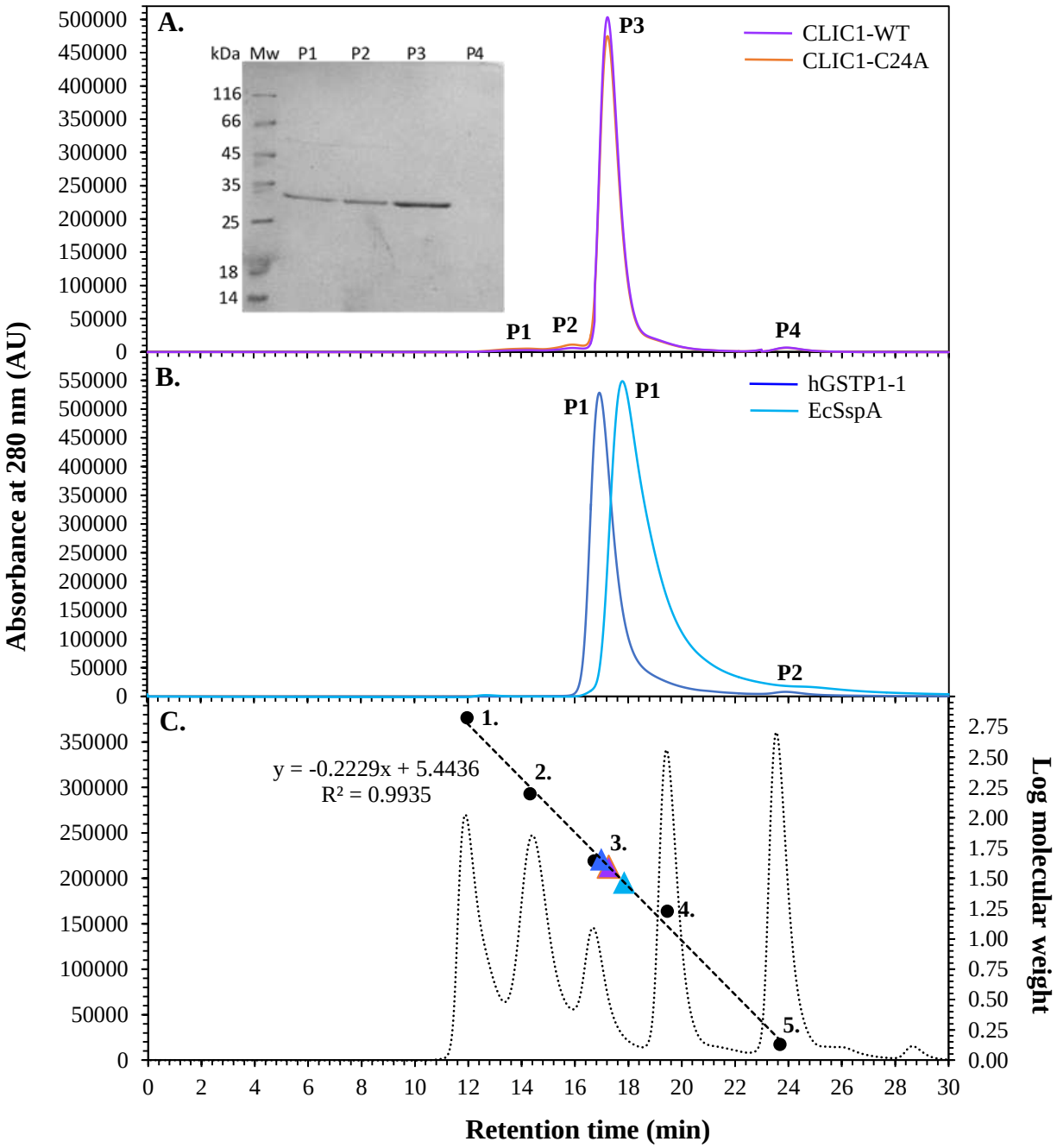


Figure 32. Verifying the oligomeric state and purity of the target proteins by SE-HPLC. The oligomeric state of the target proteins were verified by injecting 100 µM of the target protein onto a SEC column at flow-rate of 0.2 ml.min^{-1}. (A) The retention time for CLIC1-WT (●) and CLIC1-C24A (●) was ~17.23 min, therefore reporting a predicted molecular weight of ~40 kDa. Furthermore, based on SDS-PAGE, the first additional two peaks in the SEC profile were CLIC1 and could possibly represent a higher ordered oligomer (or soluble aggregate) (P1) and a non-specific covalently-linked dimer (P2). (B) The retention time for hGSTP1-1 (●) and EcSspA (●) was 16.92 min and 17.84 min, therefore reporting a predicted molecular weight of 47 kDa and 30 kDa, respectively. The expected and predicted molecular weight for hGSTP1-1 were similar and this indicated that it was dimeric. However, EcSspA and CLIC1-WT/C24A, despite their anomalous behaviour on a SEC column, were reported to be dimeric and monomeric, respectively. (C) The molecular weight was determined from a calibration curve that was constructed by using: (1) thyroglobulin (670 kDa), (2) γ-globulin (158 kDa), (3) ovalbumin (44 kDa), (4) myoglobin (17 kDa) and (5) vitamin B12 (1.35 kDa). *See section 2.2.6 for the experimental details pertaining to this method.*

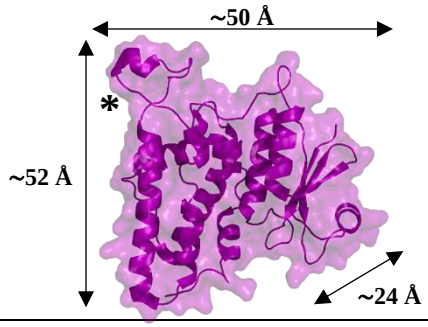
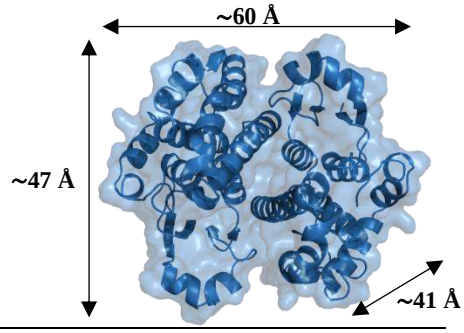
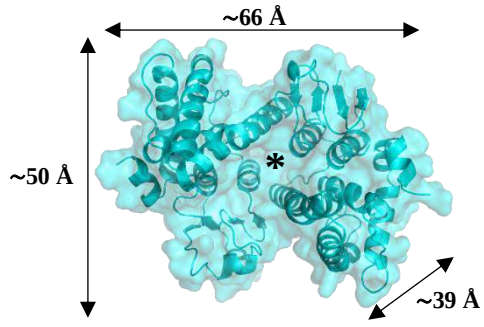
Protein type	Retention time (min)	Molecular weight (kDa)		Molecular surface	
		Predicted	Expected	Area (Å²)	Volume (Å³)
CLIC1-WT and C24A (▲/▲)	17.23	40.09	26.2		
				21 820	58 270
hGSTP1-1 (▲)	16.92	47.0	46.8		
				16 920	52 420
EcSspA (▲)	17.84	30.23	48.6		
				17 370	52 910

Figure 33. Protein crystal structures provide insight into the oligomeric state. The expected and predicted molecular weight for hGSTP1-1 were similar, indicating that the protein was dimeric as seen in the crystal structure. However, CLIC1-WT/C24A displayed anomalous behaviour on a SEC column which could be attributed to its very dynamic and negatively charged loop (*) that consequently increased its hydrodynamic volume. More importantly, the total ASA for CLIC1-WT/C24A was similar to hGSTP1-1, a dimeric protein, and would therefore account for a larger predicted molecular weight. Similarly, EcSspA showed anomalous behaviour on a SEC column such that its hydrodynamic volume appeared to be smaller than expected. This could be attributed to the fact that present within EcSspA is a hydrophobic surface pocket (*) that was suggested to play a key role in protein-protein interactions, therefore depending on the ionic strength this site could interact non-specifically with the SEC column to retard its motion and consequently report a lower molecular weight. All images were generated by using Pymol V1.3.

3.1.6 Evaluating the global secondary and tertiary structure of the target proteins

3.1.6.1 Far-UV CD spectropolarimetry and its inference on secondary structure

The far-UV CD spectra for all the target proteins indicated two prominent troughs at 208 nm and 222 nm (Figure 34), a profile characteristic of a predominantly α-helical protein. This correlated well with the crystal structure of (i) CLIC1-WT (PDB: 1K0M), (ii) hGSTP1-1 (PDB: 2GSS) and (iii) YpSspA (1YY7) which respectively showed that the α-helical and beta-sheet content was (i) 47% and 8%, respectively, (ii) 60% and 8%, respectively and (iii) 56% and 9%, respectively. However, the spectra for hGSTP1-1 had a larger ellipticity at 208 nm as opposed to CLIC1-WT/C24A and EcSspA which showed a larger ellipticity at 222 nm. In the case of proteins which are predominately α-helical the difference in ellipticty at 208 nm and 222 nm can provide insight to the nature of the α-helices, that is, when the (222/208) nm ratio is either ≤ 0.86 or ≥ 1 the helices are considered to be mainly isolated helices or coiled-coil α-helices, respectively (Lau *et al*., 1984). The (222/208) nm ratio for CLIC1-WT, hGSTP1-1 and EcSspA was reported to be 0.83, 1.12 and 0.92, respectively, therefore indicating that the secondary structure of hGSTP1-1 comprised more coiled-coil α-helices.

The solvent and buffer system that is used during a study should keep the requirements of both the protein and the experimental technique in mind. For example, urea (8 M) is routinely used to ensure the denaturation of proteins, yet even if a shorter path-length of the cuvette is to be used, urea tends to have a high absorbance at lower wavelengths ($\leq$ 210 nm) in the far-UV region. Consequently, this reduces the optical transparency of the sample and makes the data unreliable for analysis (Kelly *et al*., 2005). However, considering that the ellipticity at 222 nm is a suitable probe to monitor the secondary structure of a protein, one can still evaluate the affect these chaotropic agents have on the structural stability of a protein. In this study when the protein samples were incubated in the presence of 8 M urea it resulted in their denaturation and consequently a loss in their secondary structure (Figure 34), therefore further suggesting that the target proteins were in their native α-helical conformations.

However, with respect to CLIC1-WT and CLIC1-C24A, it was noted that the C24A mutation resulted in an 11% decrease in the ellipticity at 222 nm (Figure 34 and 36). Therefore, it was suggested that the C24A mutation induced a change in some of the secondary structural content present within CLIC1. This observation correlated with the thermal stability of both proteins,

which showed that the T_{unfol} was 50 °C and 41 °C for CLIC1-WT and CLIC1-C24A, respectively, therefore indicating that the C24A mutation resulted in CLIC1 being less stable. The thermal unfolding of both proteins, however, was irreversible and consequently no thermodynamic parameters could be calculated from data fitting. Yet, it can be noted that the gradients from the thermal unfolding data appeared to be similar and therefore this suggested that the C24A mutation induced no change in the unfolding cooperativity of CLIC1 (Figure 36).

3.1.6.2 Intrinsic fluorescence spectroscopy and its inference on tertiary structure

The UV-absorbance and consequently intrinsic fluorescence of proteins is attributed to three main aromatic amino acid residues, namely phenylalanine, tyrosine and tryptophan. However, based on (i) the phenomena of FRET where resonance energy transfer can occur from phenylalanine to tyrosine and then to tryptophan, along with (ii) tryptophan having the greatest molar extinction coefficient, the fluorescence emission of proteins is dominated by tryptophan residues. This is significant seeing that tryptophan is a unique and complex fluorophore with two nearby isoenergetic transitions that are strongly dependent on the polarity of the surrounding solvent (Lakowicz, 2006). Therefore, by monitoring the average environment of the tryptophan residues, the global and/or local tertiary structure of the target proteins were determined by using this fluorophore as a probe.

The fluorescence emission spectra for both CLIC1-WT/C24A, hGSTP1-1 and EcSspA indicated fluorescence emission maxima at 342 nm, 338 nm and 339 nm, receptively (Figure 34). These emission maxima reflected the phenomena that when tryptophan is in an environment surrounded by water, it has a maximum fluorescence emission near 350 nm (Lakowicz, 2006) and therefore the tryptophan residues within the target proteins were more solvent exposed. For instance, present within CLIC1-WT is a single tryptophan residue (Trp35) and based on its crystal structure (PDB: 1K0M) it was predicted by PDBePISA (https://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) that Trp35 has a total ASA of 89.28 $Å^2$ (Figure 35), therefore confirming that Trp35 indeed was significantly more solvent exposed. However, when CLIC1-WT/C24A was denatured in the presence of urea there was both a decrease and a red-shift to 347 nm in the maximum fluorescence emission intensity (Figure 34), therefore confirming that Trp35 was no longer slightly buried by its neighbouring residues, but instead that it was completely exposed to the surrounding solvent.

Present in hGSTP1-1, however, are two tryptophan residues (Trp28 and Trp38) and they are noted to have an ASA of 52.67 Å² and 26.87 Å², respectively (Figure 35). Therefore, the predominately solvent exposed nature of Trp28 would account for the fluorescence emission maxima noted at 338 nm. However, although there was a red-shift to 346 nm when hGSTP1-1 was denatured in the presence of urea, there was also an increase in the intensity of the fluorescence emission maxima (Figure 34). Variations in the emission of tryptophan fluorescence are due to the structure of proteins, consequently the quenching of indole may not only be due to the environment of the surrounding solvent, but it can also be due to the residues surrounding tryptophan. For instance, the quenching of tryptophan can be influenced by charged amino groups and by electron acceptors such as protonated carboxyl groups, however, the extent and stabilisation of the electron transfer quenching state depends strongly on the location and distance between the interacting partners (Lakowicz, 2006). In the case of hGSTP1-1 it was noted that the side chain amine of Lys44 was in direct contact with the imino nitrogen of Trp38 (Figure 35), consequently accounting for the quenched and enhanced fluorescence of Trp38 in the folded and unfolded conformation of hGSTP1-1, respectively. This kind of phenomena whereby the unfolded state of a protein will relieve the quenching of tryptophan is not uncommon and has been previously documented in the case of cytochrome c and myoglobin (Eftink, 1994).

In contrast, EcSspA was noted to have two distinct fluorescence emission maxima (Figure 34), yet present in EcSspA are three tryptophan residues (Trp70, Trp110 and Trp166). While there is no crystal structure currently available for EcSspA, based on the crystal structure for YpSspA, the structural homology model of EcSspA indicated that Trp70, Trp110 and Trp166 have a putative ASA of 65.95 Å², 21.27 Å² and 21.73 Å², respectively (Figure 35). Therefore, the predominately solvent exposed nature of Trp70 would account for the maximum fluorescence emission intensity noted at 339 nm, while the more buried Trp110 and Trp166 residues would account for the blue-shifted fluorescence maxima noted at 332 nm (Figure 34). Therefore, even though there are charged residues surrounding both Trp110 and Trp116, they are not in direct contact with tryptophan and its imino group, furthermore at these sites there are more non-polar residues that bring about the blue-shifted fluorescence spectra (figure 35). However, when EcSspA was denatured in the presence of urea there was both a decrease and a red-shift to 347 nm in the maximum fluorescence emission intensity (Figure 34), confirming that the tryptophan residues were no longer buried by its neighbouring residues and instead were completely exposed to the surrounding solvent.

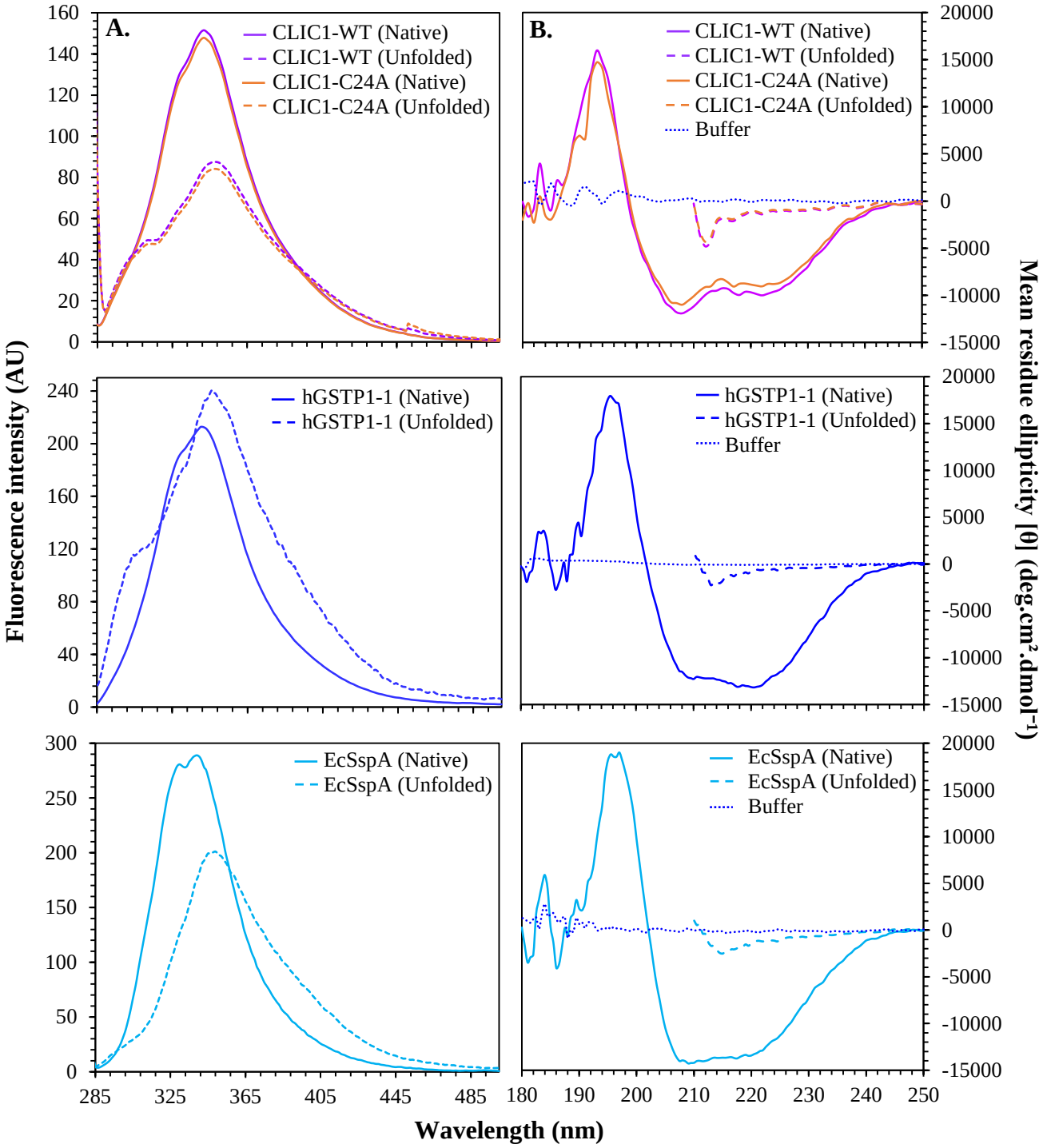


Figure 34. Verifying the secondary and tertiary structure of the target proteins. (A) The secondary structure of the target proteins were determined at 20 °C by far-UV CD spectropolarimetry by using 2 µM of the native or denatured (in the presence of 8 M urea) target protein in storage buffer. All spectra were measured in triplicate and the CD spectra for all the target proteins indicated two prominent troughs at 208 nm and 222 nm, a profile characteristic of a predominantly α-helical protein. (B) Intrinsic tryptophan fluorescence was monitored at 20 ^{0}C with an excitation wavelength of 280 nm by using 2 µM of the native or denatured (in the presence of 8 M urea) target protein in storage buffer. The fluorescence spectra indicated that the tryptophan residues (or at least one of them in the case of hGSTP1-1 and EcSspA) were predominately solvent exposed. *See section 2.2.7.1 and 2.2.7.4 for the experimental details pertaining to these methods.*

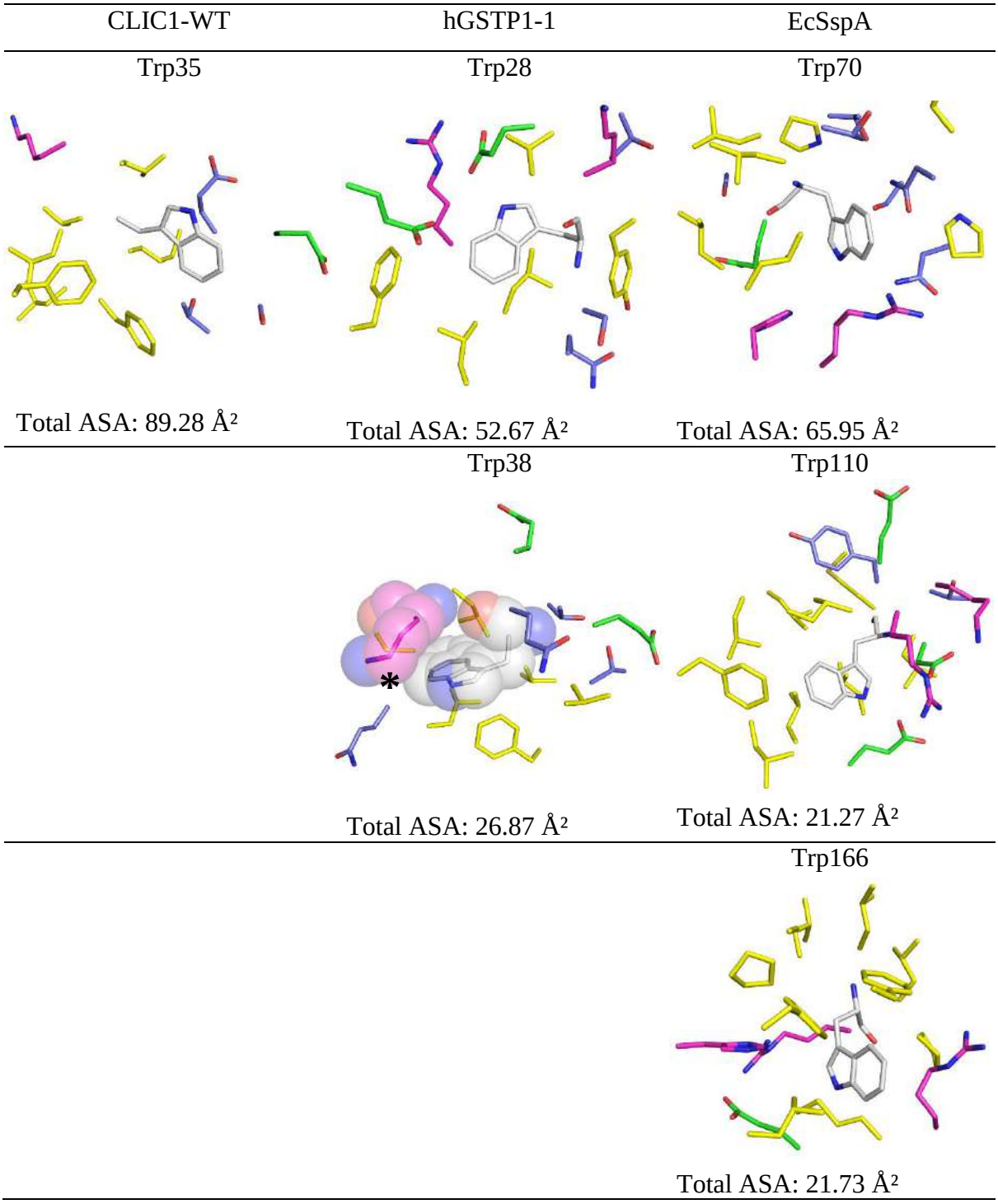


Figure 35. The crystal structure of a protein can provide insight into the observed fluorescence. The single tryptophan residue (Trp35) within CLIC1-WT is predominately solvent exposed. However, within hGSTP1-1 only Trp28 is predominately solvent exposed, while Trp38 has a lower total ASA. More importantly, in the crystal structure of hGSTP1-1 it can be seen that Trp38 is in direct contact with Lys44 (*), therefore accounting for Trp38 being quenched in the native state of hGSTP1-1. However, upon unfolding, this interaction would be omitted to result in the observed increase in fluorescence. On the other hand, the two characteristic fluorescence emission maxima of EcSspA can be attributed to the fact that Trp70 is predominately solvent exposed, whereas Trp110 and Trp116 are in a more non-polar environment where their fluorescence will not be quenched. Indicated are the polar (●), non-polar (●), positively charged (●) and negatively charged (●) residues surrounding each tryptophan (●) residue. All images were generated by using Pymol V1.3.

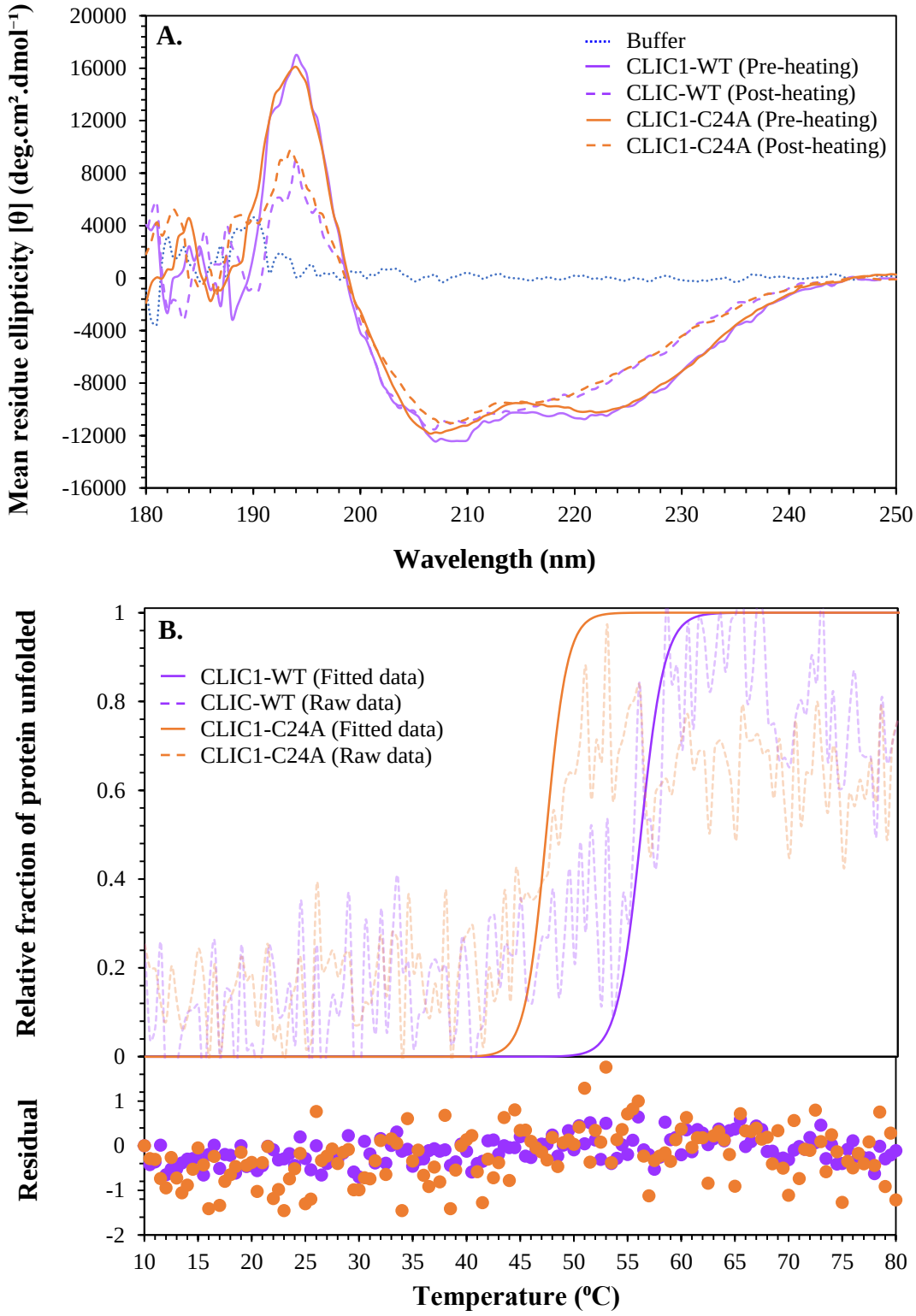


Figure 36. Thermal stability of CLIC1-WT and CLIC1-C24A. (A) The secondary structure for both target proteins was determined pre- and post-heating. (B) The unfolding of CLIC1-WT and its C24A mutant was monitored by far-UV CD spectropolarimetry by monitoring a change in the ellipticity at 222 nm for 2 µM of the target protein in sample buffer across a temperate range of 10-80 ºC. Indicated are both the raw data (- - -), the data which were fitted (—) to a sigmoid three parameter model in SigmaPlot® and the residuals from the fit. The T_{unfol} value, corresponding to an initial loss in the secondary structure content, was 50 °C and 41 °C for CLIC1-WT and CLIC1-C24A, respectively. *See section 2.2.7.1 and 2.2.7.3 for the experimental details pertaining to this method.*

3.1.6.3 Near-UV CD spectropolarimetry and its inference on tertiary structure

Even though the intrinsic fluorescence of proteins is a suitable probe to monitor the local or global tertiary structure of a protein, an additional and complementary probe used to monitor the tertiary structure of a protein is near-UV CD spectropolarimetry. Since near-UV CD is sensitive to the environment of the aromatic residues, this technique is useful as it is able to provide information about the packing environment of these chromophores and consequently insight into the tertiary structure of a protein. The near-UV CD spectra for CLIC1-WT and CLIC1-C24A indicated the presence of negative peaks at 260 nm, 280 nm and 289 nm along with characteristic positive peaks at 261 nm, 263 nm, 270 nm and 275 nm (Figure 37). The characteristic troughs seen at 280 nm and 289 nm, along with the fine structural changes between 290 nm and 305 nm, is typical of tryptophan residues and can be attributed to Trp35 seeing that CLIC1-WT/C24A contains a single tryptophan residue. However, the signal observed at 260 nm, 261 nm and 263 nm was attributed to the phenylalanine residues within CLIC1-WT/C24A, while the signal at 270 nm and 275 nm was attributed to its tyrosine residues.

The near-UV CD spectra confirmed that based on the overall magnitude and shape of the peaks, the mobility, packing environment and subsequently the spatial disposition (Kelly *et al*., 2005) of the phenylalanine and tyrosine residues was significantly different between CLIC1-WT and CLIC1-C24A. However, there appeared to be no significant loss in the packing environment and tertiary structure surrounding Trp35 in both proteins (Figure 37). The latter coincided with the intrinsic fluorescence of CLIC1-WT and CLIC1-C24A, whereby CLIC1-C24A showed only a 2.5% decrease in the maximum fluorescence emission intensity (Figure 34), therefore confirming that the C24A mutation did not significantly alter the local environment surrounding Trp35.

Therefore, based on the collective CD and fluorescence data for CLIC1-WT/C24A, hGSTP1 and EcSspA, it was determined that their secondary structure was predominately α-helical and that the location of their tryptophan residues, as suggested by their fluorescence spectra, correlated with the crystal structure or homology model for each target protein. For that reason, based on the local and global structural analysis of each protein, it was suggested that the target proteins were in their native conformation.

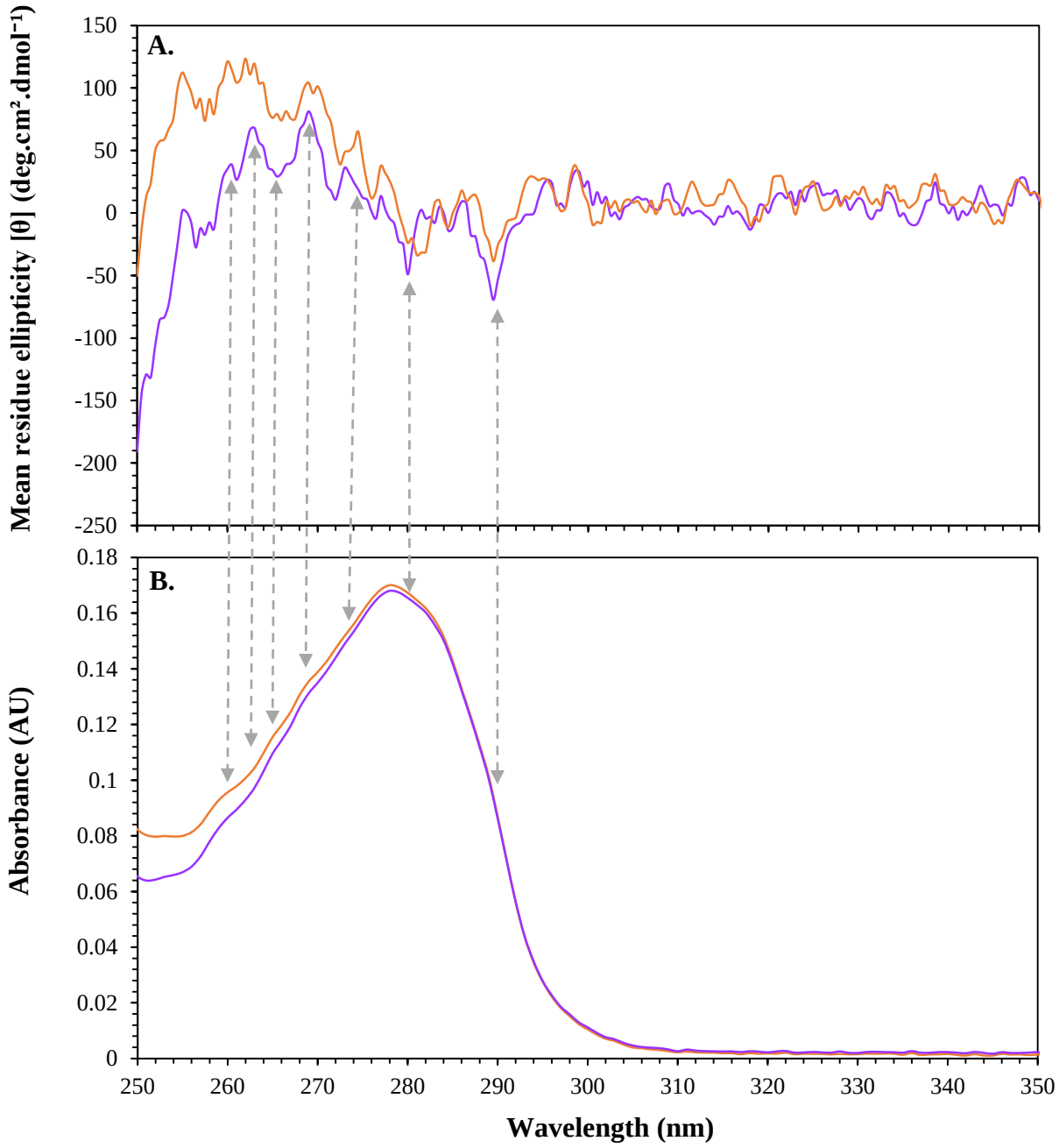


Figure 37. Tertiary structure of CLIC1-WT and its C24A mutant. (A) The tertiary structure for CLIC1-WT and its C24A mutant was monitored at 5 °C by near-UV CD spectropolarimetry by using 40 µM of the target protein in sample buffer. (B) The absorbance spectra for both CLIC1-WT and CLIC1-C24A provided a correlation between the absorbance attributed by particular aromatic residues with the signal observed in the near-UV spectra. Each aromatic residue had a characteristic wavelength profile within the near-UV CD spectra whereby (i) tryptophan had a peak close to 290 nm with fine structure between 290 nm and 305 nm, (ii) tyrosine had a peak between 275 nm and 282 nm and (iii) phenylalanine has sharp fine structure between 255 nm and 270 nm. The mobility, packing environment and subsequently spatial disposition of the phenylalanine and tyrosine residues was significantly different between both proteins. However, there appeared to be no significant loss in the packing environment and tertiary structure surrounding Trp35. *See section 2.2.7.2 for the experimental details pertaining to this method.*

3.2 Probing protein-ligands interactions through spectroscopic-based techniques

3.2.1 Monitoring the interaction between IAA-94 and hGSTP1-1 by the CDNB assay

Although the enzyme activity for CLIC1-WT was shown to be inhibited by IAA-94, in contrast to hGSTP1-1, its structural homolog, CLIC1-WT does not display any CDNB activity (Khamici *et al*., 2015). Similarly, EcSspA does not display any CDNB activity (Hansen *et al*., 2005) and therefore the ability for both CLIC1-WT and EcSspA to interact with IAA-94 was investigated by ITC (section 3.3.1 and 3.3.4). Instead, however, the enzyme activity of hGSTP1-1 can be measured by using the CDNB colorimetric assay. In this assay hGSTP1-1 catalyses the reaction between reduced GSH and the substrate CDNB, resulting in a coloured GS-DNB product with an apparent absorbance at 340 nm (Habig, 1974). Therefore, the increase in absorbance will be directly proportional to the specific activity of hGSTP1-1. However, Oakley *et al*. (1996) indicated that EAA bound to the H-site and therefore competitively inhibited the binding of CDNB at the active site. As a result, considering that CLIC1-WT and hGSTP1-1 are structural homologs and that IAA-94 is a structural analogue of EAA (section 1.5), which was shown to bind to hGSTP1-1 (section 3.3.3), the ability of IAA-94 to inhibit the CDNB activity of hGSTP1-1 was investigated.

Initially, as a control, hGSTP1-1 was incubated in the presence of EAA to confirm that the CDNB enzyme activity for hGSTP1-1 did become inhibited, but also this aided to verify that hGSTP1-1 was folded in its native and active conformation. It was noted that EAA did in fact inhibit the enzyme activity of hGSTP1-1 with an IC_{50} value of 12.37 µM, which compared to the IC_{50} value of 4.9 µM reported by Musdal *et al*. (2013) indicated that the concentration of EAA required to inhibit 50% of the CDNB enzyme activity of hGSTP1-1 was in the low micromolar range. In contrast, it was noted that IAA-94 did not inhibit the CDNB enzyme activity of hGSTP1-1, even at a concentration of 2.1 mM IAA-94 that was used during the ITC experiments (section 3.3.3) (Figure 38). Therefore, although IAA-94 did bind to hGSTP1-1, its mode of binding was different to that of EAA to not have induced inhibition of the CDNB enzyme activity of hGSTP1-1.

It is important to note that apart from CDNB, hGSTP1-1 can initiate a nucleophilic substitution reaction with various other xenobiotic compounds, for instance benzyl isothiocyanate (BITC) which is ubiquitous in cruciferous vegetables and is a common substrate for hGSTP1-1. In fact, Ralat and Colman, 2004 reported that BITC is an efficient substrate for hGSTP1-1 with a lower

apparent Michaelis-Menten constant (K_m) compared to CDNB. Furthermore, when hGSTP1-1 was incubated in S-methyl GSH the enzyme maintained near full activity with respect to CDNB, but was nearly inactive with respect to BITC. Therefore, although BITC and CDNB are substrates for hGSTP1-1, they exhibit different modes of binding which is concomitantly related to a different mode of inhibition. Therefore, although IAA-94 did not inhibit the CDNB enzyme activity for hGSTP1-1, it may still be able to inhibit the enzyme activity of hGSTP1-1 for other substrates that could appear to be more physiologically relevant.

However, hGSTP1-1 has been reported to also have a non-substrate, or ligandin-binding site (L-site) (Oakley *et al.*, 1999) and therefore it was possible that IAA-94 could have become bound at this site. Despite it being uncertain where the location of the L-site is within hGSTP1-1, it has been suggested that based on the fact that ligands for GST from *Schistosoma japonica* can bind at the dimer interface, the L-site for hGSTP1-1 may be at the dimer interface (McTigue *et al*., 1995). However, owing to the very polar environment of this region it appears to be an unlikely location for the L-site (Oakley *et al.*, 1999). In fact, it was confirmed that in contrast to hGSTP1-1, IAA-94 could inhibit the enzyme activity of SjGST (Dr Ikechukwu Achilonu, unpublished data) and furthermore based on an ITC experiment it was confirmed that IAA-94 bound to SjGST with a stoichiometry of n = 1.04, therefore confirming that IAA-94 most likely bound to the dimer interface of SjGST (Figure S5-S7; Table S4). Although the dimer interface of hGSTP1-1 may not be a plausible region for the L-site, Ji *et al*. (1997) and Prade *et al*. (1997) had instead suggested that L-site for hGSTP1-1 could be located at the buffer-binding site for HEPES and MES, and for reasons that will be made apparent in subsequent sections it was proposed that the binding of IAA-94 to hGSTP1-1 occurred at the L-site, or in particular at/near the MES-binding site.

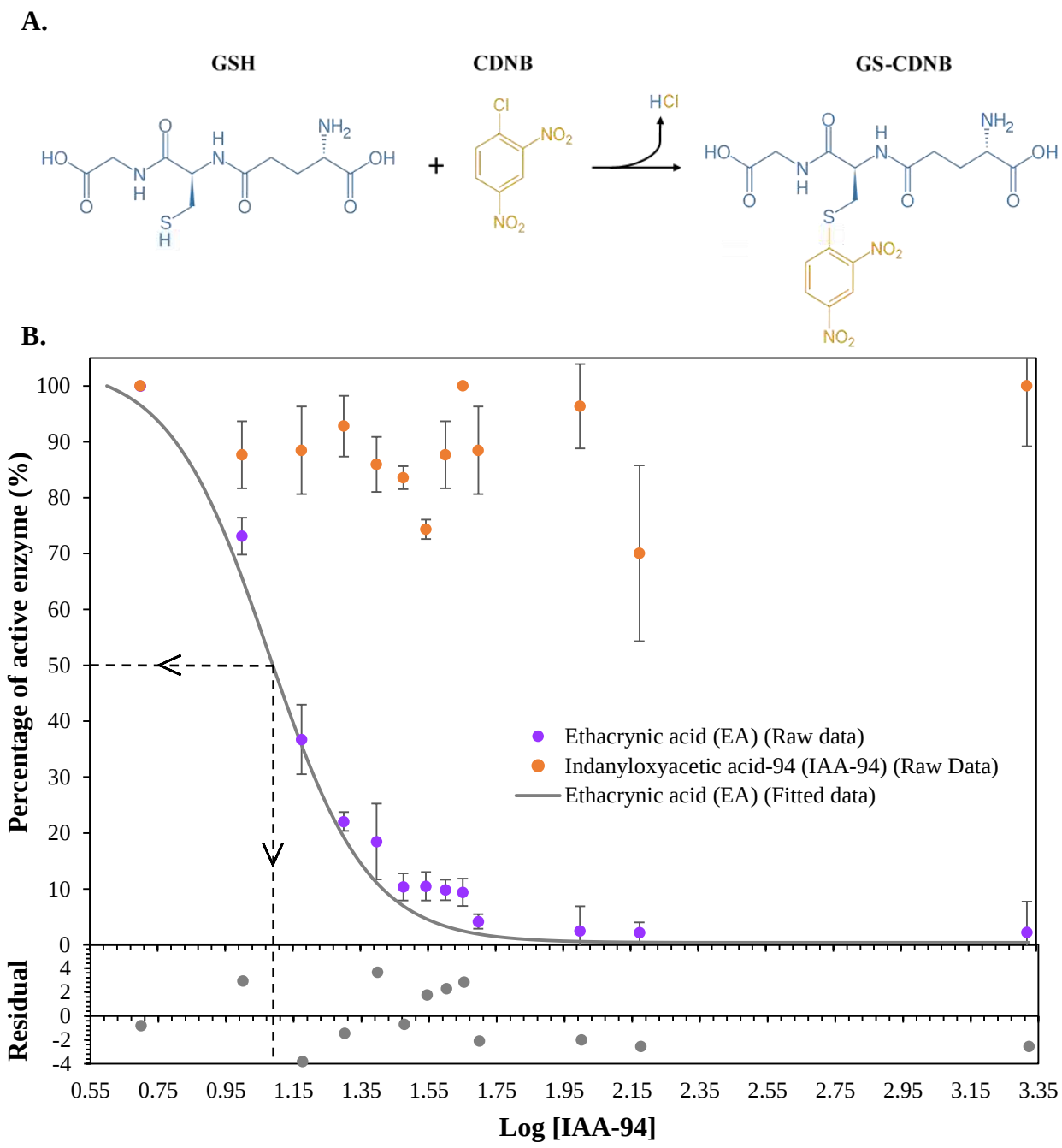


Figure 38. Inhibition in the CDNB enzyme activity of hGSTP1-1 by EAA and IAA-94. (A) In the CDNB assay hGSTP1-1 catalyses the reaction between reduced GSH and CDNB, resulting in a coloured GS-DNB product that has an apparent absorbance at 340 nm. Therefore, the increase in absorbance will be directly proportional to the specific activity of hGSTP1-1. (B) The reaction rates for each sample were then converted to a percentage of enzyme activity that was plotted against the logarithm of inhibitor concentration. The IC_{50} value was then determined by fitting the data to a sigmoidal dose-response (variable slope) function in SigmaPlot®. It was noted that EAA competitively inhibited the binding of CDNB at the active site and therefore it inhibited the CDNB enzyme activity of hGSTP1-1 with an IC_{50} value of 12.37 µM. However, it was noted that IAA-94 did not inhibit the CDNB enzyme activity of hGSTP1-1, therefore indicating a different mode of binding to its structural analog EAA. *See section 2.2.8 for the experimental details pertaining to this method.*

3.2.2 Monitoring the interaction between IAA-94 and hGSTP1-1 by ANS displacement

Extrinsic fluorescence probes are routinely applied during protein analysis to characterise features such as folding intermediates, surface hydrophobicity and to detect aggregation events (Hawe *et al*., 2008). For instance, the fluorophore ANS is one of the most commonly used extrinsic fluorescence probes to investigate the hydrophobicity of native proteins because of its enhanced quantum yield upon binding to solvent-exposed hydrophobic patches. That is, similar to tryptophan, in an aqueous environment the fluorescence of ANS is quenched, but when placed in a hydrophobic environment the fluorescence is greatly enhanced and the maximum fluorescence emission wavelength will be blue-shifted (Stryer, 1965; Uversky *et al*., 1998).

In this study, it was confirmed that both hGSTP1-1 and EcSspA could bind ANS, therefore confirming the presence of exposed hydrophobic patches on the surface of these proteins in their native conformation. However, the ANS fluorescence spectrum in the presence of EcSspA indicated a larger blue-shift and a larger increase in the fluorescence intensity, therefore suggesting that a greater proportion of the surface area of the naphthalene moeity became buried upon ANS binding to EcSspA. The fluorescence profile for ANS in the presence of CLIC1-WT, however, was similar to free ANS (Figure 39), therefore confirming that in the native conformation of CLIC1-WT there are no significant surface-exposed hydrophobic patches, which coincided with the findings reported by Fanucchi *et al*. (2008).

Since ANS binds to hGSTP1-1, considering that IAA-94 and EAA are hydrophobic molecules, should they bind with a greater affinity and at a similar or allosteric site to ANS, these ligands would displace ANS from its binding site in a competitive or non-competitive manner, respectively as indicated by a shift in the fluorescence towards the spectrum of free ANS in solution. In this study it was noted that both IAA-94 and EAA could displace ANS from hGSTP1-1 with an efficiency of 51% and 44%, respectively (Figure 40). Yet, the absence of complete ANS displacement could have arisen due to hGSTP1-1 having a stronger binding affinity for ANS compared to IAA-94 and EAA, or alternatively hGSTP1-1 could have other high and low affinity ANS-binding sites located at hydrophobic patches throughout the enzyme (Ralat and Colman, 2004).

The displacement of ANS from hGSTP1-1 by EAA has previously been reported by Abdalla *et al*. (2002), however, the ability of EAA to displace ANS has also been reported for

GSTA1-1 (Dirr *et al*., 2005) and GSTM1-1 (Kinsley, 2005) and collectively it has been suggested that ANS binds at the H-site. Yet, the binding of MES at its buffer-binding site, or otherwise the putative L-site, has also been reported to influence the binding of ANS (Prade *et al*., 1997). Therefore, although the displacement of ANS from hGSTP1-1 could adhere to a non-competitive mechanism, due to the absence of complete ANS displacement by both IAA-94 and EAA it was suggested that there was more than one binding site for ANS in hGSTP1-1. Furthermore, compared to EAA, the apparent increase in the displacement of ANS by IAA-94 could infer a binding site that has a lower binding affinity for ANS. Therefore, in association with the results obtained from the CDNB assay, it appeared that the binding of ANS can occur at more than one site within hGSTP1-1 and that the binding of IAA-94 and EAA to hGSTP1-1 occurs at distinct sites.

Sample	Maximum fluorescence intensity (AU)		F/F$_o$ ratio at maximum fluorescence intensity
	Magnitude	Wavelength (nm)	
Free ANS	154.84 ± 0.82	518	1.00
hGSTP1-1 with ANS	266.63 ± 11.81	501	1.72
CLIC1-WT with ANS	181.10 ± 2.61	514	1.17
EcSspA with ANS	369.73 ± 10.69	485	2.39

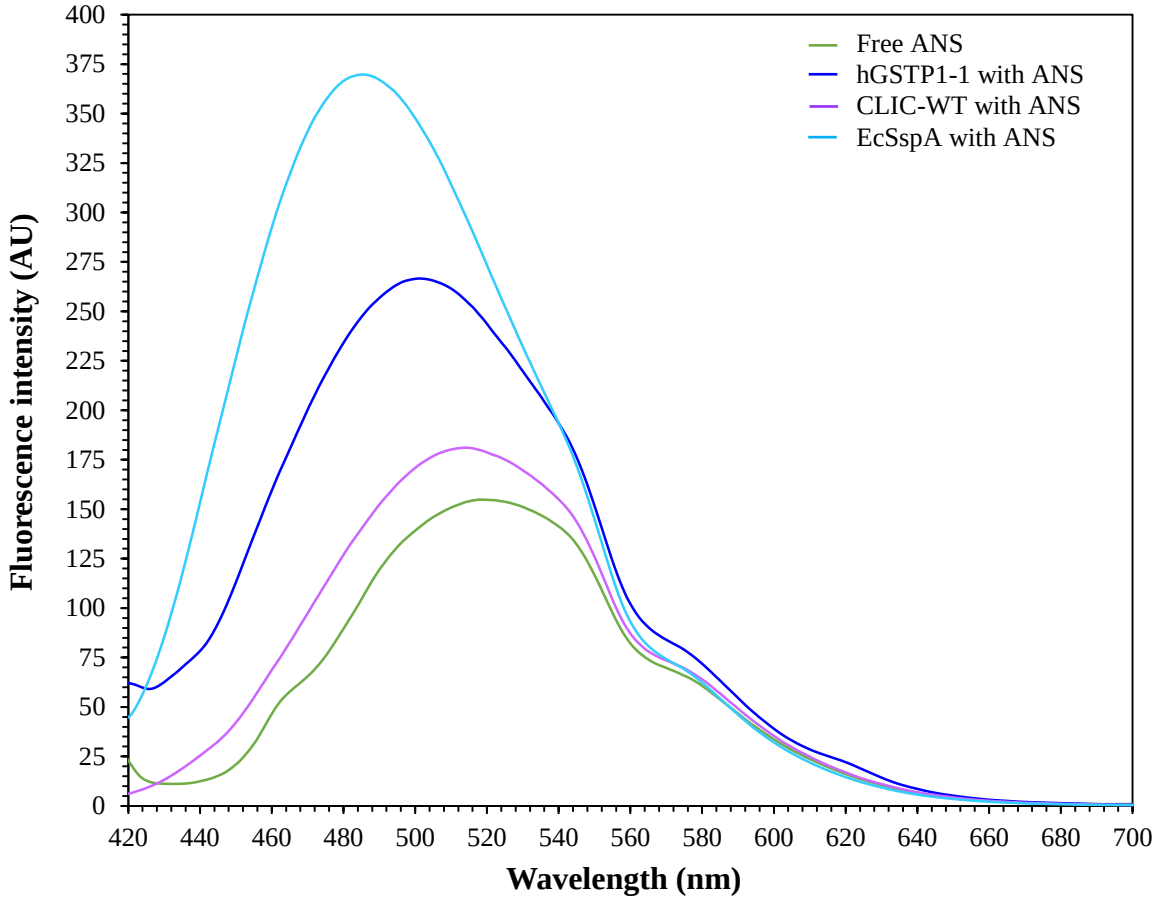


Figure 39. Binding of ANS to CLIC1-WT, hGSTP1-1 and EcSspA. Extrinsic ANS fluorescence was monitored at 20 ^{0}C with an excitation wavelength of 390 nm by using 2 µM of the target protein that had been incubated in 100 µM ANS. Relative to free ANS, the binding of ANS to the target protein was accompanied by an increase in intensity and a blue-shift in the wavelength of the maximum fluorescence intensity. The table above indicates the maximum fluorescence intensity for free ANS as well as for ANS in the presence of the target protein. It also indicates the degree of ANS binding by taking a ratio of the maximum fluorescence intensity for ANS in the presence of protein (F) to the maximum fluorescence intensity for free ANS (F$_o$). It was observed that in the presence of hGSTP1-1 and EcSspA, the maximum fluorescence of ANS showed an intensity increase and a blue-shift in the wavelength, therefore indicating that ANS had bound to both proteins. In contrast, when ANS was in the presence of CLIC1-WT the F/F$_o$ ratio was ~1, therefore indicating that CLIC1-WT did not bind ANS. *See section 2.2.7.5 for the experimental details pertaining to this method.*

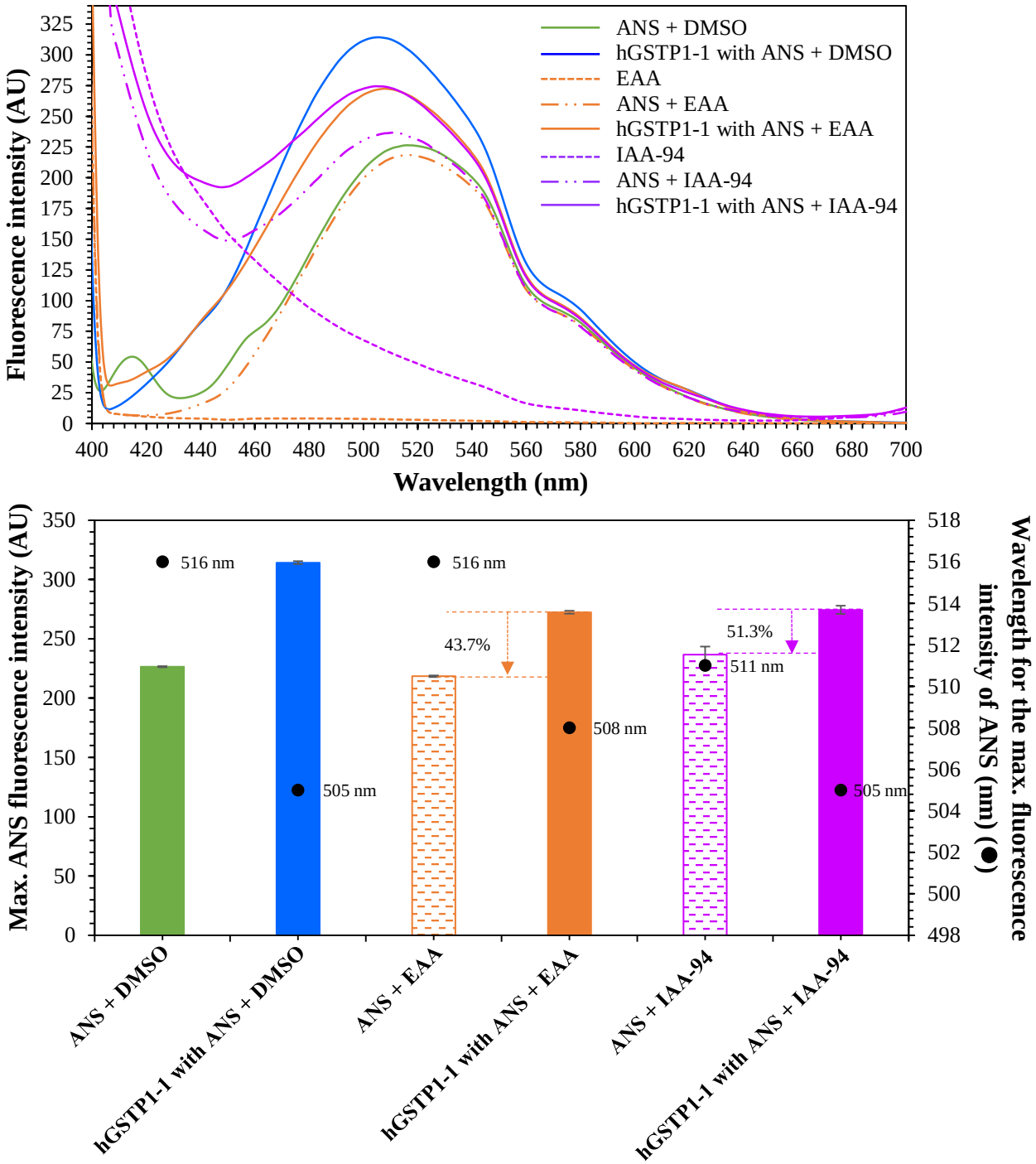


Figure 40. The displacement of ANS from hGSTP1-1 by IAA-94 and EAA. Extrinsic ANS fluorescence was monitored at 20 ^{0}C with an excitation wavelength of 390 nm by using 2 µM of the target protein that had been incubated in 100 µM ANS. Relative to free ANS, the binding of ANS to the target protein was accompanied by an increase in intensity and a blue-shift in the wavelength from 516 nm to 505 nm for the maximum fluorescence intensity. However, it was noted that both IAA-94 and EAA could displace ANS from hGSTP1-1 with an efficiency of 51% and 44%, respectively. *See section 2.2.7.5 for the experimental details pertaining to this method.*

3.3 Evaluating the thermodynamic parameters underlying protein-ligand interactions

3.3.1 The interaction between IAA-94 and the soluble conformation of CLIC1-WT

It was reported by Khamici *et al*. (2015) that the enzyme activity of CLIC1-WT was inhibited by IAA-94, therefore ITC was implemented to better understand this interaction from a thermodynamic perspective. A successful ITC experiment firstly requires that the concentrations of both the macromolecule and ligand are appropriately determined to ensure that the magnitude of the heat involved in an interaction is sufficient to ensure a reasonable signal-noise ratio. Secondly, defining the binding affinity for an interaction is important considering that it governs the profile of the isotherm (Freyer *et al*., 2008; Perozzo *et al.*, 2004). Collectively, these parameters contribute to the Weismann or c-value, a term used to predict the overall profile of an isotherm and it is defined as follows (Wiseman *et al.*, 1989):

$$c = \frac{n \cdot [M_{tot}]}{K_d} \quad \textbf{(11)}$$

whree n represents the number of ligands that will bind to the macromolecule, $[M_{tot}]$ represents the total macromolecule concentration of the titrand and the K_d represents the dissociation constant for the interaction. To obtain an ideal sigmoidal isotherm that will accurately provide the thermodynamic parameters for an interaction, the c-value should be $10 \leq c \leq 100$ considering that when $c < 10$ the isotherm will result in a shallower slope with less of the reaction occurring per an injection, while when $c > 100$ it will shift the isotherm profile to correspond to a step-like function given that either all or none of the species are reacting before or after stoichiometric equivalence of the titrant and titrand is achieved, respectively (Hansen *et al*., 2011; Perozzo *et al.*, 2004; Wiseman *et al*., 1989)

There was, however, no apparent binding constant for the interaction between soluble CLIC1-WT and IAA-94, yet apart from ITC there are other methods that can be used to investigate protein-ligand interactions. For instance, fluorescence spectroscopy can be used to determine the binding affinity by monitoring (i) the quenching of protein fluorescence induced by a ligand or (ii) the enhanced fluorescence of the bound ligand should it be a fluorophore. The absorbance spectroscopic properties of IAA-94 had not been reported previously, however, it was noted that IAA-94 had a strong absorbance at both 280 nm and 295 nm along with a lower absorbance near 340 nm (Figure 41), as a result this would present a problem due to the

primary and secondary inner-filter effect, respectively (Fonin *et al*., 2014). Therefore, owing to the inner-filter effect and the challenges it can present, an approximate binding affinity constant for the interaction between CLIC1-WT and IAA-94 was not determined by fluorescence spectroscopy. Furthermore, due to the limited solubility of IAA-94, the presence of DMSO had an impact on other spectroscopic techniques, for instance the impact IAA-94 may have had on the secondary structural thermal stability of CLIC1-WT could not be monitored by CD spectropolarimetry given that DMSO has a high absorbance and therefore low optical transparency. Therefore, ITC was used a probe in order to monitor the interaction between CLIC1-WT and IAA-94.

When setting up an ITC experiment it has been suggested that the concentration of the macromolecule (titrand) should be ten times the K_d value for an interaction, while the concentration of the ligand (titrant) should be ten to twenty times the macromolecule concentration. However, in the case where there is no prior knowledge on the K_d value, one can make the assumption that the K_d value is ~10 µM, consequently ensuring that not only an appropriate macromolecule concentration is used for a sufficient heat signal, but that also a c-value of 10 is achieved (Freyer *et al*., 2008; Perozzo *et al.*, 2004). Using these experimental parameters it was noted that when IAA-94 was titrated into CLIC1-WT the heat produced correlated with the heat produced when IAA-94 was titrated into buffer (Figure 42), therefore suggesting that there was no interaction between IAA-94 and CLIC1-WT. Unfortunately, the highest concentration of the ligand that could be achieved was 2.1 mM IAA-94 due to the limited solubility of IAA-94 and also because the concertation of DMSO needed to be ≤ 3% (v/v) in order to avoid damaging in the structural integrity of CLIC1-WT.

In general the magnitude for the ΔH for molecular interactions varies considerably and in some cases the ΔH_{obs} at a particular temperature may be very small and consequently not detected (Filloux and Ramos, 2014). Under isobaric conditions, the ΔH is equal to the change in heat energy for a particular system, however, the flow in heat energy is also directly correlated to the change in temperature for a particular system. Consequently, the change in the amount of heat a system experiences when exposed to a temperature change is determined by the C_p for the system as follows (Brown *et al.*, 2012):

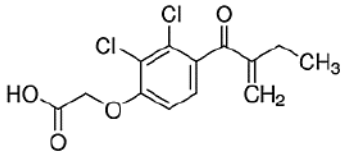
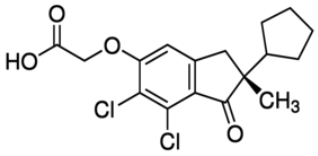
$$\Delta H = q_p = C_p \cdot m \cdot \Delta T \qquad \textbf{(12)}$$

where ΔH represents the change in enthalpy, q_p represents the exchange of heat energy, C_p represents the heat capacity for a system under isobaric conditions, m represents the mass in grams of the molecules within the system and ΔT represents the change in temperature. Therefore, by changing the temperature for a particular ITC experiment, without inducing any phase change, this will change the magnitude for ΔH by changing the amount of heat energy transferred. The relationship between the ΔH and ΔT can be further understood and supported by Kirchhoff's Law which states that (Atkins and Paula, 2010):

$$\Delta H^\circ (T_2) = \Delta H^\circ (T_1) + \Delta C_p^\circ (T_2 - T_1) \quad \textbf{(13)}$$

where ΔH° (T_1) and ΔH° (T_2) represent the standard change in enthalpy at two different temperatures and ΔC_p° represents the standard heat capacity difference between the reactants and products for a particular interaction. Kirchhoff's Law indicates that when the temperature of the system changes, the enthalpy of the reactants and products will also change, but to different extents and in each case the ΔH depends on the heat capacities of the apo- and ligand-bound conformations of the macromolecules (Atkins and Paula, 2010). In general, an increase in temperature tends to favour exothermic signals, whereas a reduction in temperature favours endothermic signals (Filloux and Ramos, 2014).

Therefore, to conclude that the negligible ΔH_{obs} corresponded to no binding upon titrating IAA-94 into CLIC1-WT at 20 °C, the ITC experiment needed to be repeated at different temperatures. However, when the titration was repeated at both 10 °C and 30 °C the heat generated correlated once again with the HOD (figure 42). Accordingly, based on ITC, it was inferred that IAA-94 did not interact with the soluble conformation of CLIC1-WT and if it did, it did so very weakly. As a result, it can be concluded that the constraint in the Cl^- conductance of CLIC1 is not due to IAA-94 binding to its soluble conformation in order to prevent its structural transition into a membrane competent state, but instead the inhibition of the ion channel activity of CLIC1 is due to IAA-94 binding to its membrane-bound conformation.

Chemical properties	EAA	IAA-94
Chemical structure		
Chemical formula	$C_{13}H_{12}Cl_2O_4$	$C_{17}H_{18}Cl_2O_4$
Molecular weight (g.mol^{-1})	302.01	356.06
Molecular shape volume (Å^3)	218.50	257.70
Polar surface area (Å^2)	63.60	63.60
ALogP	3.97	4.82
No. hydrogen bond acceptors	4	4
No. hydrogen bond donors	1	1
Stereo centre count	0	1

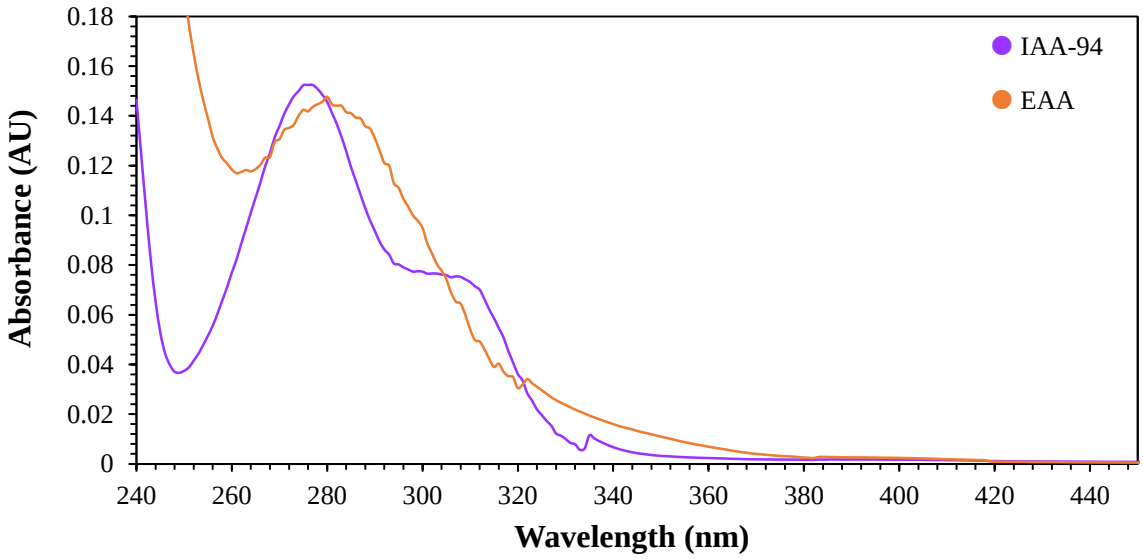


Figure 41. Spectroscopic and physical-chemical properties of IAA-94 and EAA. The UV-Vis absorbance spectroscopic properties of IAA-94 (●) and EAA (●) were monitored by using 15 µM of either compound in storage buffer (section 2.2.3.3). All data were corrected by subtracting the contributions caused by the buffer which had an equal concentration of DMSO (0.02% (v/v)). Both IAA-94 and EAA had an absorbance maxima at 280 nm, however, IAA-94 had an additional characteristic strong absorbance between 295-320 nm. It is evident that IAA-94 and EAA are structural analogues and consequently they share similar physical-chemical properties, for instance they both have four hydrogen bond acceptors, one hydrogen bond donor and a total polar surface area of 63.60 Å^3. However, the AlogP value, where P is the partition coefficient that defines the ratio of the concentration of a solute between two solvents with different polarities, was different between EAA and IAA-94. The AlogP value for IAA-94 was larger than EAA and therefore indicated that it was more hydrophobic. Furthermore, opposed to EAA, IAA-94 has a stereo centre which may account for the different UV-Vis absorbance spectroscopic properties.

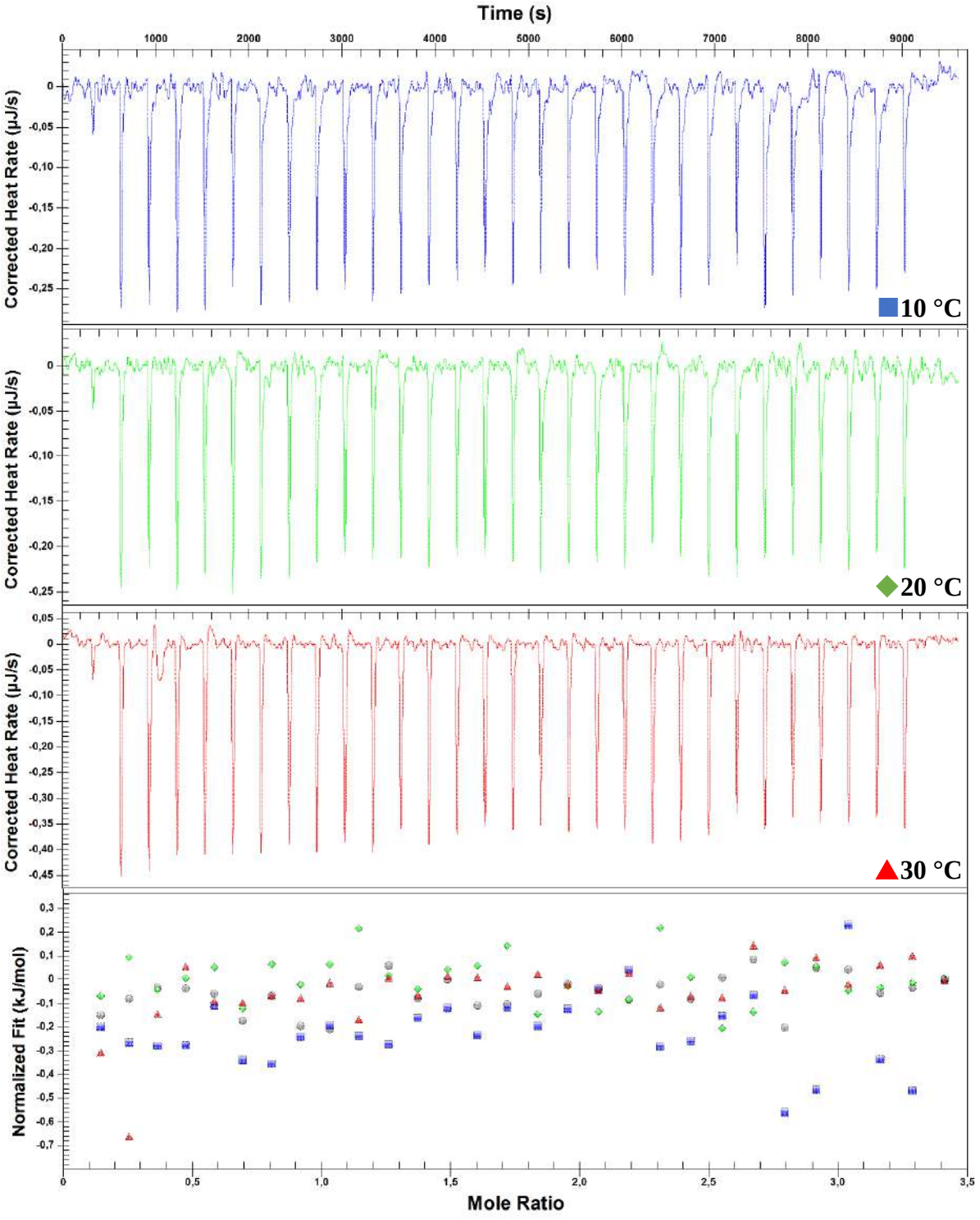


Figure 42. Monitoring the thermodynamic parameters for the interaction between IAA-94 and CLIC1-WT. Using ITC, the interaction between IAA-94 and CLIC1-WT was monitored by titrating 2.1 mM IAA-94 into 100 µM CLIC1-WT. Each titration was performed, in triplicate, at 20 °C (◆) with a stir rate of 250 rpm, an initial and final baseline of 300 s and a duration period of 300 s between each injection. The volume of the ITC syringe used was 250 µl and during each run the initial injection volume was 2 µl, but for all subsequent injections it was 5 µl and therefore a total of 49 injections were performed. In the event where no interaction was apparent the titration were repeated at both 10 °C (■) and 30 °C (▲). Indicated is also the HOD (●). *See section 2.2.9.1 for the experimental details pertaining to this method.*

3.3.2 The interaction between EAA and the soluble conformation of CLIC1-WT

Although IAA-94 appeared to not show an interaction with soluble CLIC1-WT, considering CLIC1-WT is a structural homolog of hGSTP1-1 and that IAA-94 is a structural analogue to EAA, a renowned inhibitor to the enzyme activity of hGSTP1-1 (section 1.5), the ability of soluble CLIC1-WT to interact with EAA was similarly investigated by ITC in order to understand this interaction from a thermodynamic perspective.

In contrast to IAA-94, EAA was shown to interact with the soluble conformation of CLIC1-WT with a stoichiometry of n = 1.18, therefore indicating that one molecule of EAA bound to one molecule of CLIC1-WT. The interaction was noted to be enthalpically and entropically favourable, yet the overall interaction was noted to be entropically driven (Figure 43). The binding enthalpy is related to the breaking and/or formation of bonds, whereby the latter indicates a favourable ΔH that is often associated to the formation of hydrogen bonds, van der Waals interactions and various electrostatic interactions. Furthermore, the ΔH can provide context towards ligand specificity considering that a favourable ΔH is an indication of specific interactions between the binding partners (Bronowska, 2011; Chen *et al.*, 2016; Du *et al.*, 2016; Ross and Subramanian, 1981). Upon examining the interaction between CLIC1-WT and EAA, the ΔH_{obs} was noted to be smaller in magnitude (ΔH_{obs} = - 2.22 kJ.mol^{-1}) (Figure 43) compared to the interaction between hGSTP1-1 and EAA (ΔH_{obs} = - 22.18 kJ.mol^{-1}) (Table 1) (Quesada-Soriano *et al.*, 2010). Therefore, indicating that although the interaction between CLIC1-WT and EAA was sufficient to overcome the interactions that the individual molecules had made with the solvent in order to facilitate binding, hGSTP1-1 showed a greater degree of specificity for EAA that can be attributed to a greater number and/or strength of non-covalent interactions between hGSTP1-1 and EAA.

On the other hand, the change in entropy represents the change in the dynamics, or disorder, within a system and is derived from numerous components, namely (Perozzo *et al.*, 2004):

$$\Delta S_{sys} = \Delta S_{solv} + \Delta S_{conf} + \Delta S_{comp} \quad \textbf{(14)}$$

where ΔS_{sys} represents the total change in entropy for the system, ΔS_{solv} represents the change in entropy for the water molecules that are localised on the surface of the active site and ligand, ΔS_{conf} represents the change in the conformational degrees of freedom of the protein and

ligand, while ΔS_{comp} represents the change in the translational, rotational and vibrational degrees of freedom of the molecules in the bound complex. Upon ligand binding, the loss in the ΔS_{conf} and ΔS_{comp} comes from the fact that the side chains can no longer sample different rotameric states and although this may be entropically unfavourable, the ΔS_{solv} is a major driving force in allowing a favourable entropic change (Perozzo *et al*., 2004).

The overall ΔS for the interaction between CLIC1 and EAA ($T\Delta S$ = 21.97 kJ.mol^{-1}) (Figure 43) was larger compared to the interaction between hGSTP1-1 and EAA ($T\Delta S$ = 4.60 kJ.mol^{-1}) (Table 1) (Quesada-Soriano *et al*., 2010). Due to the change in a systems free energy needing to be negative for a favourable protein-ligand interaction, complementary changes between enthalpy and entropy have been frequently observed in thermodynamic binding studies of biological systems, a phenomena termed enthalpy-entropy compensation (Bronowska, 2011; Chodera and Mobley, 2013; Du *et al.*, 2016). Therefore, although the larger ΔH_{obs} by hGSTP1-1 was attributed to the number and/or strength of non-covalent interactions between the interacting partners, this came at the cost of a lower ΔS due to their being less conformational degrees of freedom for both interacting partners. Furthermore, based on the crystal structure for the apo- (PDB: 16GS) and EAA-bound (PDB: 2GSS) conformation of hGSTP1-1, there was a small difference in the number of water molecules that become displaced at the active site upon ligand binding (Figure 44). Therefore, although the net entropy change was favourable, the ΔS_{conf} and ΔS_{comp} terms attenuated the $\Delta S_{solv,}$ ultimately leading to the relatively small net entropy change.

However, in the case of CLIC1-WT, based on the crystal structure for its apo- (PDB: 1K0M) and GSH-bound (PDB 1K0N) conformation it was noted that the binding of GSH at the active site resulted in a significant displacement of water molecules to the bulk solvent (Figure 44). Therefore, considering that EAA is a hydrophobic molecule, owing to the hydrophobic effect, the desolvation of both EAA and CLIC1-WT upon binding would have significantly contributed to the state of disorder in the surrounding environment in a similar manner to when GSH bound to CLIC1-WT. This therefore compensated for the enthalpic penalty of ligand binding and made the overall interaction to be entropically driven. Yet, collectively the binding free energy and consequently the binding affinity when EAA interacted with CLIC1-WT and hGSTP1-1 was similar with a reported K_d of 57.87 µM and 47.41 µM, respectively.

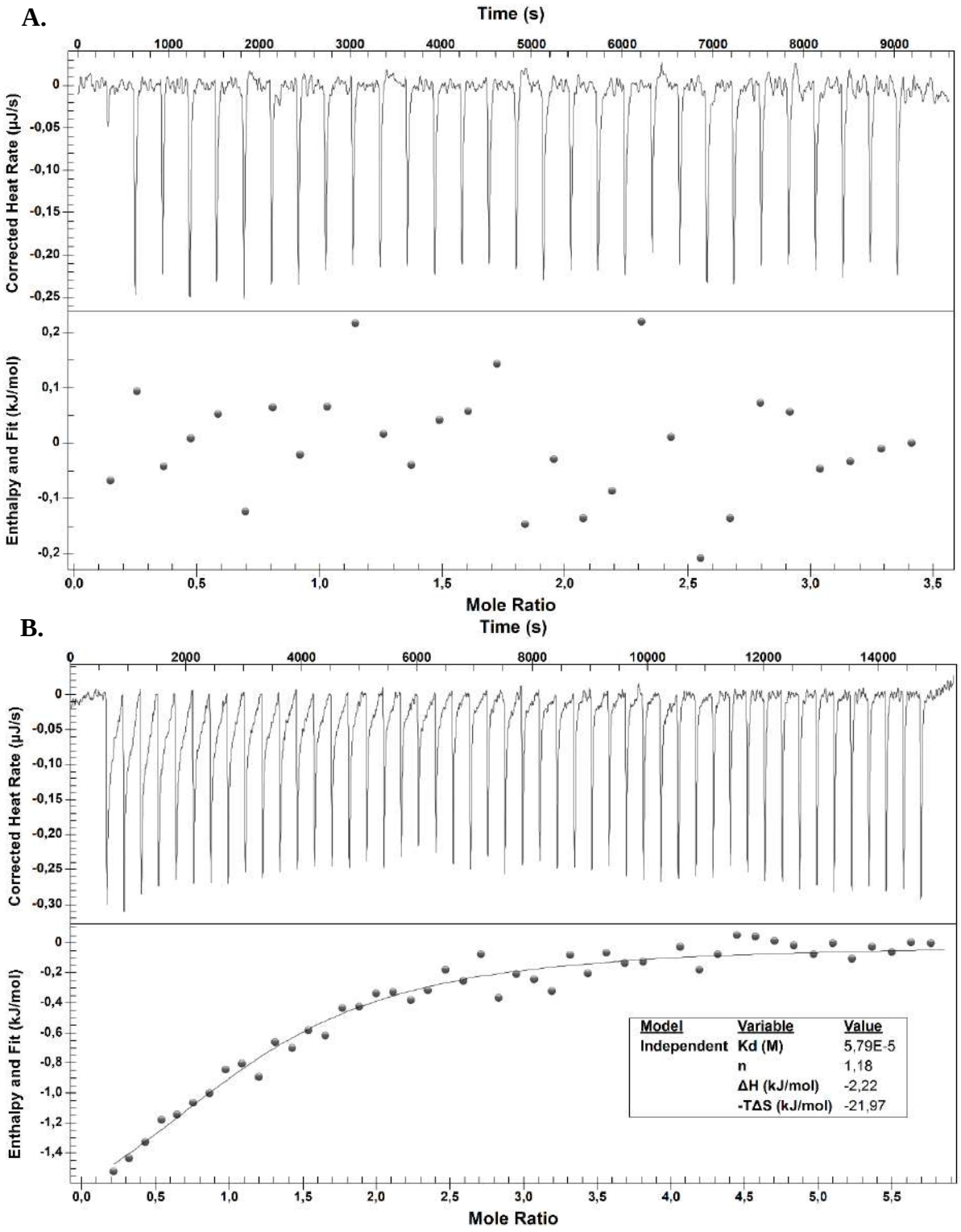


Figure 43. Monitoring the thermodynamic parameters for the interaction between IAA-94 or EAA and CLIC1-WT. (A) No interaction was observed between IAA-94 and CLIC1-WT. (B) Using ITC, the interaction between EAA and CLIC1-WT was monitored by titrating 2.48 mM EAA into 100 µM CLIC1-WT. Each titration was performed, in triplicate, at 20 °C with a stir rate of 250 rpm, an initial and final baseline of 300 s and a duration period of 300 s between each injection. The volume of the ITC syringe used was 250 µl and during each run the initial injection volume was 2 µl, but for all subsequent injections it was 5 µl and therefore a total of 49 injections were performed. The interaction occurred with a stoichiometry of $n \sim 1$, was entropically driven and had a moderate K_d value in the micromolar range *See section 2.2.9.1 for the experimental details pertaining to this method.*

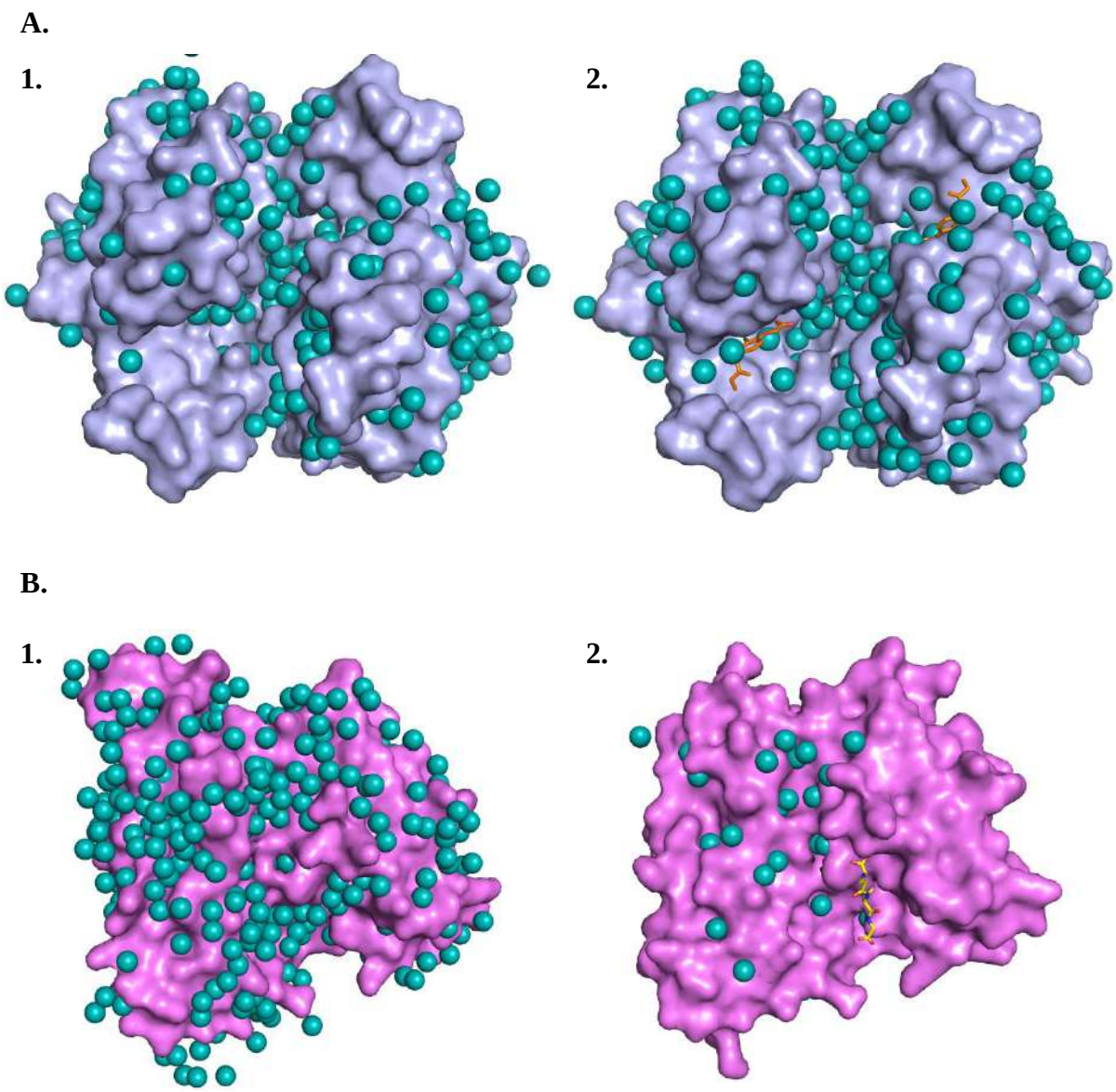


Figure 44. Water molecules associated with the crystal structure of hGSTP1-1 and CLIC1 in their apo- and ligand-bound conformations. (A) Based on the crystal structure for the (1) apo- (PDB: 16GS) and (2) EAA-bound (PDB: 2GSS) conformation of hGSTP1-1 (●), there was a small difference in the number of water molecules (●) that become displaced upon EAA (●) binding at the active site. Therefore, although the net ΔH_{obs} for the interaction between hGSTP1-1 and EAA was favourable, the ΔS_{conf} and ΔS_{comp} terms attenuated the $\Delta S_{solv,}$ ultimately leading to the relatively small net ΔS. (B) Based on the crystal structure for the apo- (PDB: 1K0M) and GSH-bound (PDB: 1K0N) conformation of CLIC1-WT, it was noted that the binding of GSH (●) at the active site resulted in a significant displacement of water molecules (●) to the bulk solvent. Therefore, considering that EAA is a hydrophobic molecule, owing to the hydrophobic effect, the desolvation of both EAA and CLIC1-WT upon binding would contribute significantly to the state of disorder in the surrounding environment in a similar manner to when GSH bound to CLIC1-WT. This therefore compensated for the enthalpic penalty of ligand binding and made the overall interaction to be entropically driven. All images were generated by using Pymol V1.3.

3.3.3 The interaction between IAA-94 and hGSTP1-1

In view of the fact that soluble CLIC1-WT has recently been shown to bind EAA, but not IAA-94, and that hGSTP1-1 has previously been shown to interact with EAA (Oakley *et al*., 1996), the ability of hGSTP1-1 to interact with IAA-94 was similarly investigated by ITC in order to understand this interaction from a thermodynamic perspective. Initially, a titration of EAA into hGSTP1-1 was performed as a control and it was confirmed that hGSTP1-1 could bind EAA with a stoichiometry of n = 2.02 (Figure 45), therefore confirming that one molecule of EAA bound to each subunit as previously reported by Oakley *et al*. (1996). Furthermore, the obtained thermodynamic parameters correlated with the results reported by Quesada-Soriano *et al*. (2009) and Quesada-Soriano *et al*. (2010) (Table 1).

Table 1: Summary of the thermodynamic parameters for the binding of EAA to hGSTP1-1

Enzyme	$\Delta G°$	ΔH_{obs}	$T\Delta S°$	K_d	$\Delta C_p°$
	(kJ.mol^{-1})			(µM)	(kJ.mol^{-1}.K^{-1})
Experiment	-24.27	- 18.65	5.62	47.41	-
hGSTP1-1[a]	- 27.82 ± 0.2	- 14.27 ± 0.2	13.55 ± 0.3	13.0 ± 1.5	- 1.23 ± 0.3
hGSTP1-1[b]	- 26.78 ± 0.2	- 22.18 ± 0.6	4.60 ± 0.1	22.2 ± 0.2	- 1.59 ± 0.0

[a] The interaction between hGSTP1-1 and EAA as reported by Quesada-Soriano *et al*. (2009).

[b] The interaction between hGSTP1-1 and EAA as reported by Quesada-Soriano *et al*. (2010).

Similarly, the interaction between IAA-94 and hGSTP1-1 indicated a stoichiometry of n = 2.13, therefore confirming that one molecule of IAA-94 bound to each subunit. The interaction was noted to be enthalpically and entropically favourable, yet the overall interaction was noted to some extent to be more enthalpically driven (Figure 45). Therefore, similar to EAA, the specificity of hGSTP1-1 for IAA-94 can be attributed to the number and/or strength of non-covalent interactions between hGSTP1-1 and IAA-94. However, in contrast to EAA, the desolvation of both IAA-94 and hGSTP1-1 upon binding contributed more to the state of disorder in the surrounding environment. Therefore, this would have compensated better for the entropic penalty that arose from the loss in the conformational degrees of freedom upon ligand binding in order to make the overall interaction entropically favourable. Furthermore, the binding free energy and consequently binding affinity when IAA-94 interacted with hGSTP1-1 yielded a K_d value of 12.74 µM (Figure 45), indicating that compared to its structural analogue EAA, hGSTP1-1 has a better affinity for IAA-94.

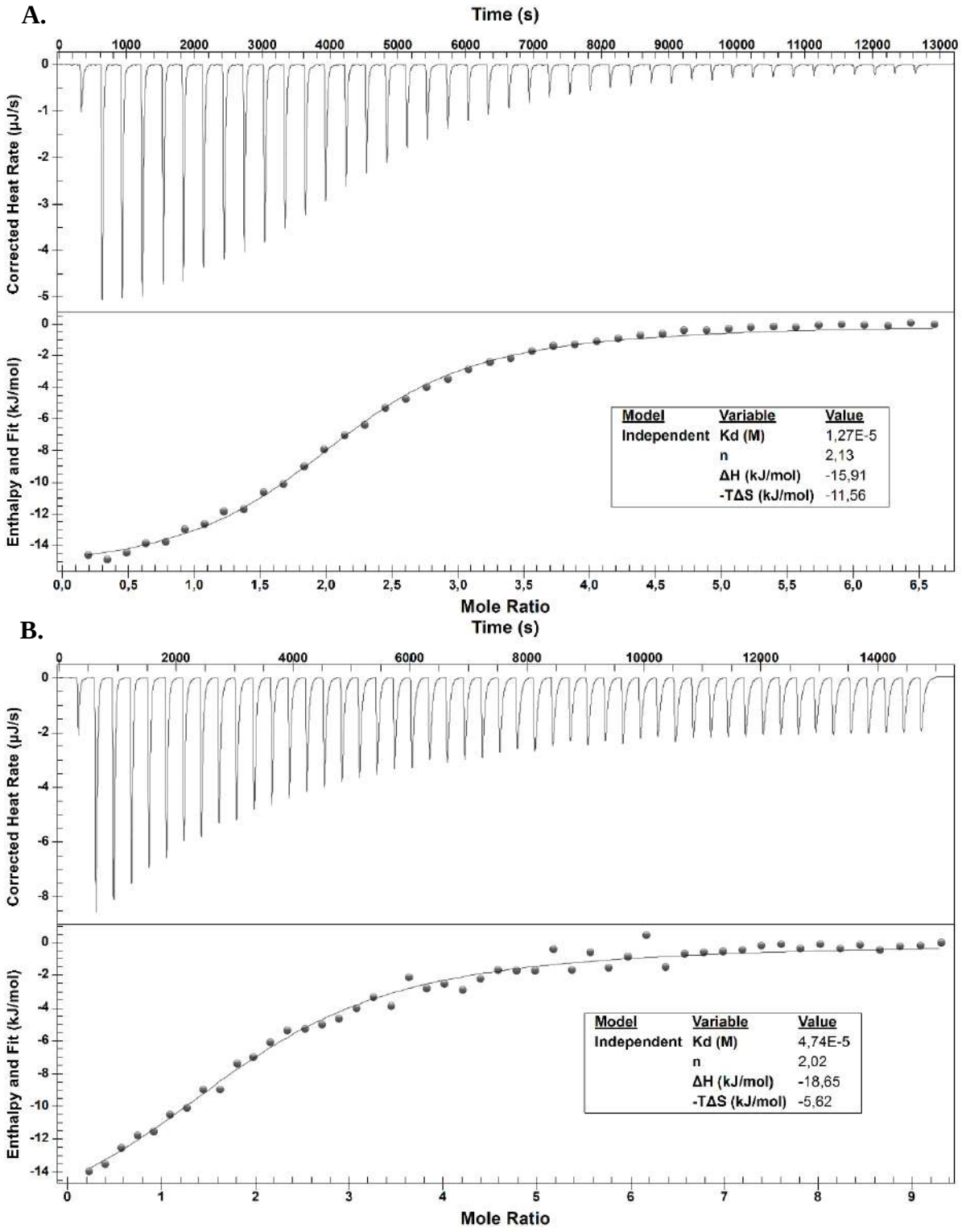


Figure 45. Monitoring the thermodynamic parameters for the interaction between IAA-94 or EAA and hGSTP1-1. Using ITC, the interaction between (A) IAA-94 or (B) EAA and hGSTP1-1 was monitored by titrating either 2.1 mM IAA-94 or 2.48 mM EAA into 100 µM hGSTP1-1. Each titration was performed, in triplicate, at 20 °C with a stir rate of 250 rpm, an initial and final baseline of 300 s and a duration period of 300 s between each injection. The volume of the ITC syringe used was 250 µl and during each run the initial injection volume was 2 µl, but for all subsequent injections it was 5 µl and therefore a total of 49 injections were performed. The interaction for both ligands occurred with a stoichiometry of $n \sim 2$, was enthalpically driven and had a moderate K_d value in the micromolar range. *See section 2.2.9.1 for the experimental details pertaining to this method.*

3.3.3.1 Monitoring the cooperativity in the interaction between IAA-94 and hGSTP1-1

Considering hGSTP1-1 is a homodimeric enzyme, there lies the possibility that the components within both subunits act collectively and consequently when a ligand binds to the active site of one subunit it may induce a conformational change that could influence how the ligand binds to the second subunit, a phenomenon described as cooperativity. In biological systems cooperativity describes a mechanism for transferring information and depending whether the binding affinity for a ligand is increased or decreased, the cooperativity can be described as being positive or negative, respectively. Yet, when there is no interacting system and the individual subunits work independently, then there is no cooperativity. One simple way to assess the cooperativity for a ligand binding system is to construct a Hill plot, a linear graph reflecting the relationship between the fraction of ligand bound to a marcomolecule as a function of the ligand concentration. Having then fitted a linear regression model to the data, the gradient will represent the Hill coefficient (n_{HP}), a parameter used to describe the cooperativity for a system. When n_{HP} is > 1, < 1 or = 1, a system is said to have positive, negative or no cooperativity, respectively (Brown, 2009; Stefan and Novere, 2013). In the interaction between IAA-94 or EAA and hGSTP1-1 the n_{HP} was reported to be n_{HP} = 0.86 and n_{HP} = 0.81, respectively (Figure 46). Therefore, although the n_{HP} was less than one for both compounds, based on the magnitude and overall trajectory of the plot, it was noted that n_{HP} ~1, therefore indicating that the binding of IAA-94 and EAA to hGSTP1-1 occurred in a non-cooperative manner.

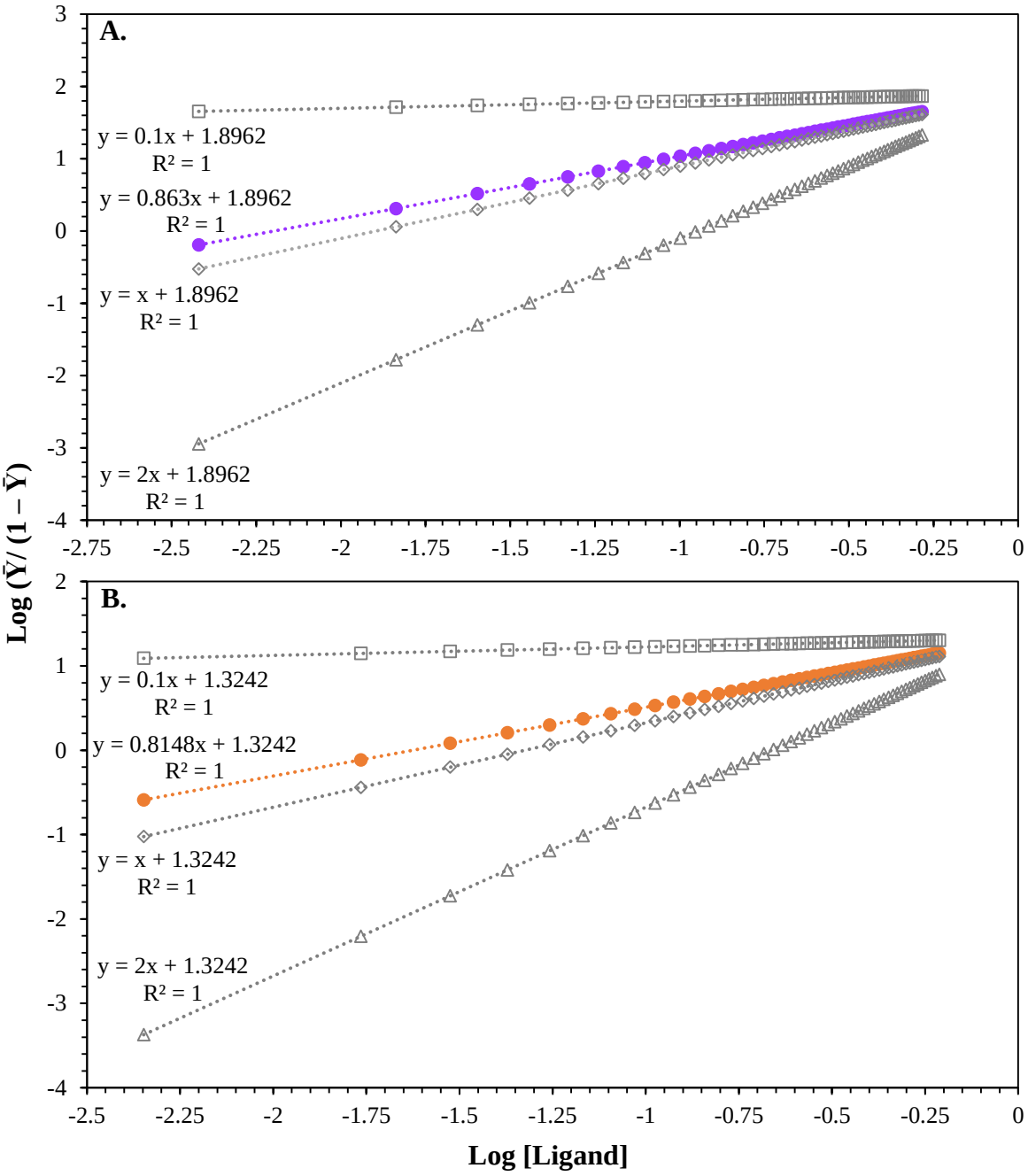


Figure 46. Monitoring the cooperativity for the interaction between IAA-94 or EAA and hGSTP1-1 by means of a Hill plot. Considering hGSTP1-1 is a homodimeric enzyme, there lies the possibility that the components within both subunits act collectively and consequently when a ligand binds to the active site of one subunit it may induce a conformational change that could influence how the ligand binds to the second subunit. A Hill plot is a simple way to assess the cooperativity for a ligand binding system by reflecting the relationship between the fraction of ligand bound to a marcomolecule as a function of the ligand concentration. The Hill coefficient (n_{HP}), a parameter used to describe the cooperativity for a system was reported to be n_{HP} = 0.86 and n_{HP} = 0.81 for the interaction between (A) IAA-94 (●) or (B) EAA (●) and hGSTP1-1, respectively. Therefore, considering that n_{HP} ~1, this indicated that the binding of IAA-94 or EAA to hGSTP1-1 occurred in a non-cooperative manner. Indicated are also the theoretical traces when n_{HP} = 0.1 (□), n_{HP} = 1 (◇) and n_{HP} = 2 (△).

3.3.3.2 The ΔC_p for the interaction between IAA-94 and hGSTP1-1

To further comprehend the thermodynamics underlying the interaction between IAA-94 and hGSTP1-1, a temperature-dependent study was performed. Consequently, a series of ITC experiments were done at different temperatures and the ΔH_{obs} was plotted in relation to the change in temperature. The trajectory of these plots for most binding interactions appear to be linear, consequently by fitting a linear regression model to the data set, the gradient will be able to represent the ΔC_p for a system when going from one state of equilibrium to another (Duff and Howell, 2015; Perozzo *et al.*, 2004). The ΔC_p is a quantitative measure as to how an object will take up heat energy, therefore on a molecular level this heat energy will be portioned as kinetic energy (e.g. vibrational, translational and rotational motion) and potential energy (e.g. interatomic energies, bond stretching and breaking). As a result, the more ways that a molecule can distribute heat energy the higher will be its heat capacity (Cooper, 2010; Prabhu and Sharp, 2005).

The ΔC_p primarily originates from the changes in the ASA involved in binding and therefore describes the changes in the hydration of non-polar and polar molecular surfaces. A negative ΔC_p is ascribed to the hydrophobic effect, that is, an increase in hydrophobic interactions, the burial of a non-polar surface area from water and the presence of buried water molecules at an interface due to binding are common factors that contribute to a negative ΔC_p (Perozzo *et al.*, 2004). This phenomenon therefore involves a solvent reorganisation component considering that the ordered water molecules that form a clathrate structure near the hydrophobic surface will be displaced and become more disordered to be able to absorb more energy when in the bulk solvent. In contrast, water molecules on a hydrophilic surface are less ordered and tend to not release as much energy when displaced into the solvent, consequently a positive ΔC_p is noted for a hydrophilic binding event (Bronowska, 2011; Du *et al*., 2016; Duff and Howell, 2015; Perozzo *et al*., 2004;. Snyder *et al*., 2013).

The gradient from the linear regression model, and hence the ΔC_p value, was reported to be - 0.76 $kJ.mol^{-1}.K^{-1}$ for the interaction between IAA-94 and hGSTP1-1 (Figure 47). Yet, compared to the ΔC_p of - 1.59 $kJ.mol^{-1}.K^{-1}$ reported for the interaction between EAA and hGSTP1-1 (Table 1) (Quesada-Soriano *et al*., 2010), the ΔC_p for this interaction was lower indicating that the binding of EAA showed a greater reduction in its hydrophobic surface area and that a more specific interface was formed. Upon exploring the binding of other

ligands such as GSH and S-hexyl GSH to hGSTP1-1, a ΔC_p of - 1.23 kJ.mol^{-1}.K^{-1} and - 1.85 kJ.mol^{-1}.K^{-1} (Quesada-Soriano *et al*., 2009) was reported, respectively. This would be expected considering that EAA and S-hexyl GSH are both occupied by the H-site and therefore a large portion of their non-polar surface area would become buried in the hydrophobic H-site. However, for reasons that will be made clear in subsequent sections, it was suggested that IAA-94 did not bind at the active site and therefore no inferences could be made between the ΔC_p value and the binding of IAA-94 to the H-site of hGSTP1-1. Instead, it was suggested that IAA-94 bound to the L-site, or otherwise the MES-binding site, and compared to the H-site it was observed that the L-site was more polar (Figure 51 and 53). Therefore, the increased polarity at the L-site could have accounted for the lower ΔC_p value for the interaction between IAA-94 and hGSTP1-1 given that the proportion of the hydrophobic surface area of IAA-94 that became buried would have been less when binding to hGSTP1-1. In fact, this observation was further supported by the IFMD data (Table S4) which illustrated that in comparison to the IFMD of EAA to the H-site, the Glide ligand efficiency score for the IFMD of IAA-94 to the L-site was less negative, therefore indicating that the surface area of IAA-94 had less of an affect towards the favourable binding interaction between IAA-94 and hGSTP1-1.

Additionally, with an increase in temperature the ΔG of binding was fairly consistent, however, owing to the enthalpy-entropy compensation phenomena, there was a notable decrease in the ΔH and ΔS (Figure 47), therefore indicating a shift in the interaction becoming more enthalpically driven with an increase in temperature. Upon examining the physical-chemical properties of IAA-94, the compound was shown to be non-polar with an ALogP value of 4.82 (Figure 41), hence the negative ΔC_p coincided with the desolvation of the binding site. Although the change in solvation is a significant driving factor in governing a favourable entropic change, the decrease in the ΔS with an increase in temperature would have still accounted for the desolvalation of the binding site, however, the ΔS_{conf} and ΔS_{comp} would have attenuated the ΔS_{solv} by becoming more unfavourable, ultimately leading to the relatively small net ΔS with an increase in temperature. This comes from the fact that with an increase in temperature there would be an increase in the overall dynamics of hGSTP1-1, hence its backbone and sides chains would experience an increase in their conformational degrees of freedom that may therefore result in a larger number and/or strength of non-covalent interactions with IAA-94, interactions that would stabilise the binding of IAA-94 to hGSTP1-1.

It is becoming increasingly apparent that protein-ligand interactions are not just solely driven by the release of entropically unfavourable waters, but rather that they are also driven by a dominant favourable ΔH. For instance, the active site for the major urinary protein (MUP) is occluded and not well solvated. Consequently, since there are little water molecules to remove upon binding, the desolvation penalty is minimal and the "dry active site" does not significantly contribute to the observed ΔS, therefore allowing the binding enthalpy to predominate. A hydrophobic effect which is enthalpically driven can be attributed mainly to the desolvation of the ligand, with only a minor contribution from the desolvation of the protein, a phenomenon termed the non-classical hydrophobic effect. In fact, when the water network that forms on the surface of the bound ligand becomes larger and organised, the greater will be the binding enthalpy. Furthermore, the dipole moment of the ligand can influence the arrangement of the water network at a binding site, therefore if a strong hydrogen-bond network can be formed with the first hydration shell of the protein-ligand complex, then this would contribute considerably to a favourable enthalpy change associated with binding (Duff and Howell, 2015). Therefore, with an increase in temperature a series of intra- and intermolecular bonds may have changed the topography of the IAA-94-binding site in hGSTP1-1 to ensure that not only favourable protein-ligand interactions were formed, but importantly that also favourable ligand-solvent interactions were established to make the overall interaction enthalpically driven (Figure 45 and 47).

Temperature (K)	Thermodynamic parameters				
	Stoichiometry (n)	ΔH_{obs} (kJ.mol^{-1})	TΔS (kJ.mol^{-1})	K_d (μM)	ΔG (kJ.mol^{-1})
278.15	2.12 ± 0.01	- 4.70 ± 0.11	22.48 ± 1.56	36.97 ± 19.37	- 27.19 ± 0.32
283.15	2.01 ± 0.05	- 8.31 ± 0.05	18.62 ± 0.02	10.79 ± 0.23	- 26.93 ± 0.05
288.15	2.01 ± 0.06	- 12.08 ± 0.20	15.15 ± 0.31	11.67 ± 0.57	- 27.23 ± 0.12
293.15	2.02 ± 0.03	- 17.03 ± 0.46	9.87 ± 0.66	17.38 ± 2.42	- 26.90 ± 0.22
298.15	2.07 ± 0.03	- 20.28 ± 0.54	7.15 ± 0.65	15.67 ± 0.69	- 27.43 ± 0.11
303.15	2.15 ± 0.05	- 23.48 ± 0.63	4.41 ± 0.73	15.73 ± 0.67	- 27.88 ± 0.10

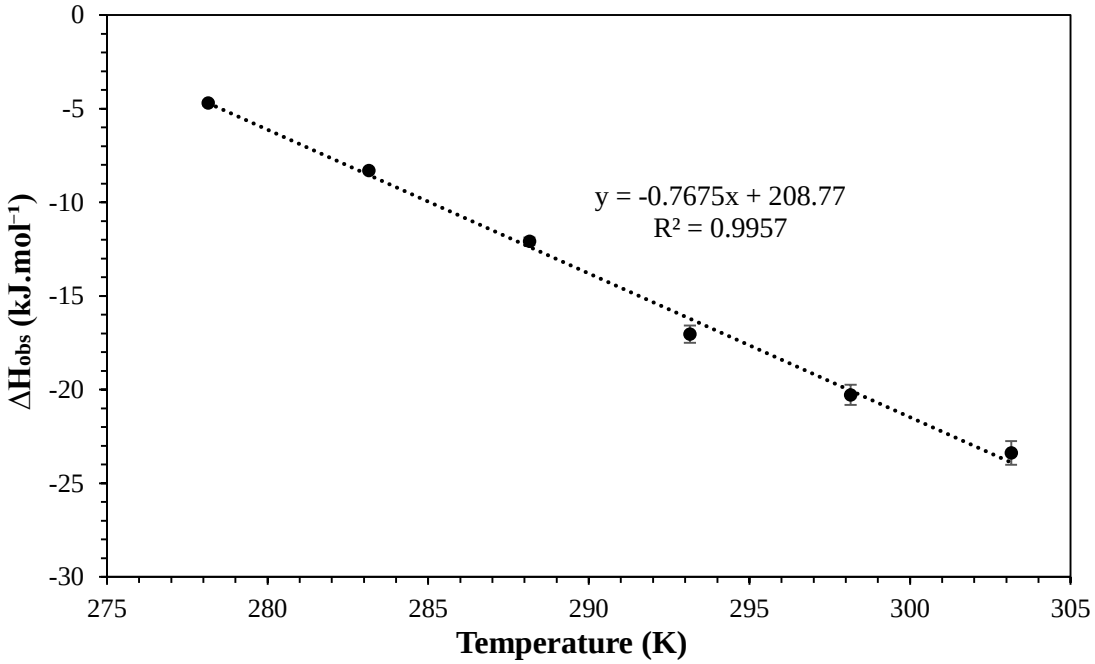


Figure 47. The ΔC_p for the interaction between IAA-94 and hGSTP1-1. Using ITC, the dependence of the interaction between IAA-94 and hGSTP1-1 on temperature was monitored by titrating, in triplicate, 2.1 mM IAA-94 into 100 μM hGSTP1-1 across a temperature range of 5-30 °C. Each titration was performed at a stir rate of 250 rpm, an initial and final baseline of 300 s and a duration period of 300 s between each injection. The volume of the ITC syringe used was 250 μl and during each run the initial injection volume was 2 μl, but for all subsequent injections it was 5 μl and therefore a total of 49 injections were performed. The ΔH_{obs} for each experiment was plotted as function of temperature and the ΔC_p was determined from the gradient of the linear regression model fitted to the data. The ΔC_p for the interaction was observed to be - 0.76 kJ.mol^{-1} which correlated with the non-polar surface area of IAA-94 becoming buried when interacting with hGSTP1-1. Upon an increase in temperature the overall reaction becomes more exothermic, however, owing to the enthalpy-entropy compensation phenomenon, this was accompanied by a decrease in ΔS which consequently kept ΔG consistent. However, the n and K_d value did not change significantly, indicating that the stoichiometry and binding affinity remained constant. *See section 2.2.9.2 for the experimental details pertaining to this method.*

3.3.3.3 Proton-linked effects during the interaction between IAA-94 and hGSTP1-1

Although the ΔH_{obs} represents the change in the heat for an interaction, it is important to keep in mind that an interaction can involve either the uptake or release of a proton. However, the ionisation for some buffer systems can also involve the release of a proton into the solvent, therefore based on Le Chatelier's Principle the change in the protonation state brought about by an interaction can shift the equilibrium and therefore contribute to the heat of ionisation (ΔH_{ion}) for the buffer system (Perozzo *et al.*, 2004). To evaluate if the binding of IAA-94 to hGSTP1-1 was coupled to any protonation events and therefore if the ΔH_{obs} was due to the protein-ligand interaction and the ΔH_{ion} of the buffer system, the ITC experiment was repeated by using buffer systems with a different ΔH_{ion}. Consequently, the ΔH_{bind} that has been corrected for the ΔH_{ion} can be established as follows (Perozzo *et al*., 2004):

$$\Delta H_{obs} = \Delta H_{bind} + n\mathrm{H} + \Delta H_{ion} \quad \textbf{(15)}$$

where ΔH_{obs} represents the change in enthalpy directly observed from an ITC experiment, ΔH_{bind} represents the true change in enthalpy for the interaction without any contribution form the buffer system, nH represents the number of protons released or taken up upon complex formation, and ΔH_{ion} represents the change in entropy for the ionisation of the buffer system. However, in this study it was evident that the thermodynamic parameters for the interaction between IAA-94 and hGSTP1-1 were similar across the various buffer systems (Figure 48). Therefore, indicating that the ΔH_{obs} was the true ΔH for the interaction (i.e. $\Delta H_{obs} = \Delta H_{bind}$) and that the interaction between IAA-94 and hGSTP1-1 was therefore not accompanied by a protonation-linked effect.

Thermodynamic parameters	Heat of ionisation for different buffer systems (ΔH_{ion} (kJ.mol^{-1}))				
	Phosphate	PIPES	HEPES	MES	Imidazole
	3.52	8.70	12.70	20.38	36.65
Stoichiometry (*n*)	2.07	1.93	1.92	2.02	2.09
ΔH_{obs} (kJ.mol^{-1})	- 20.28	- 17.86	- 20.62	- 17.08	- 16.12
$T\Delta S$ (kJ.mol^{-1})	7.15	8.83	6.63	8.65	9.64
K_d (µM)	15.67	21.12	16.82	31.14	30.62
ΔG (kJ.mol^{-1})	- 27.43	- 26.69	- 27.25	- 25.73	- 25.76

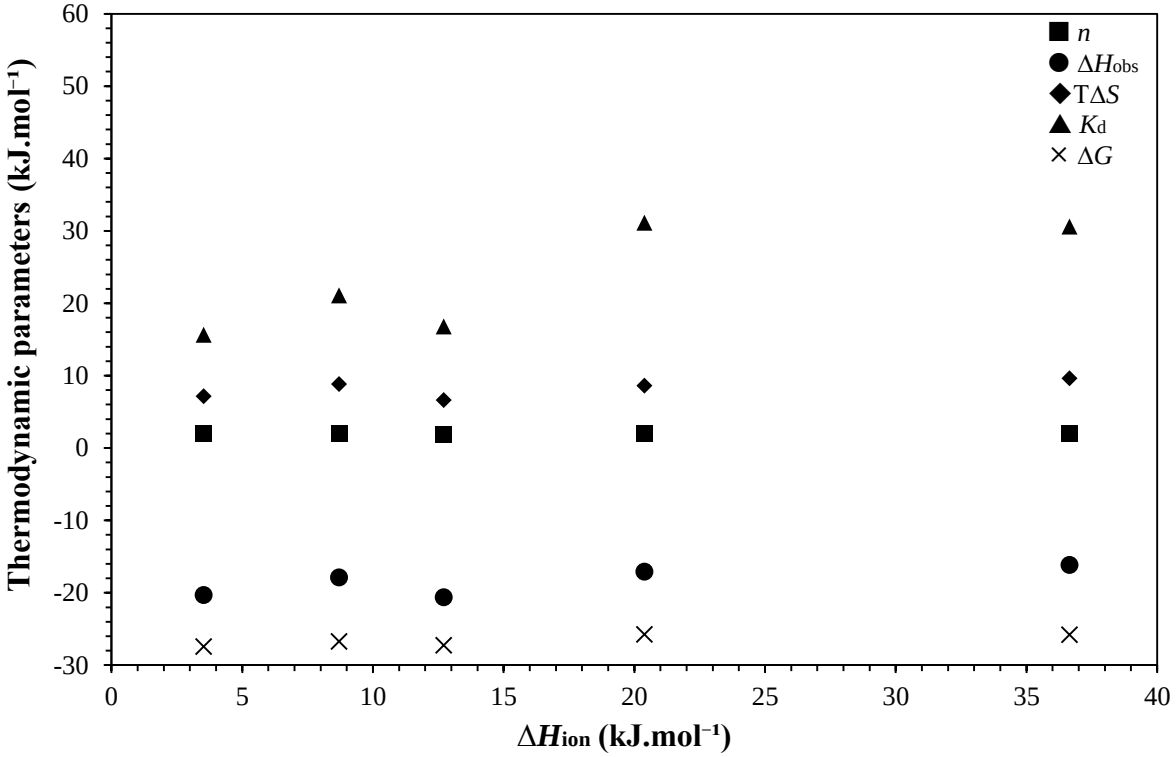


Figure 48. Proton-linked effects during the interaction between IAA-94 and hGSTP1-1. To evaluate if the binding of IAA-94 to hGSTP1-1 was coupled to any protonation events and therefore if the ΔH_{obs} was due to the protein-ligand interaction and the ΔH_{ion} of the buffer system, the ITC experiment was repeated by titrating 2.1 mM IAA-94 into 100 µM hGSTP1-1 within various buffer systems that had a different ΔH_{ion}. Each titration was performed at 20 °C with a stir rate of 250 rpm, an initial and final baseline of 300 s and a duration period of 300 s between each injection. The volume of the ITC syringe used was 250 µl and during each run the initial injection volume was 2 µl, but for all subsequent injections it was 5 µl and therefore a total of 49 injections were performed. It was observed that the thermodynamic parameters for the interaction between IAA-94 and hGSTP1-1 were similar across the various buffer systems, therefore indicating that ΔH_{obs} was the true ΔH for the interaction (i.e. $\Delta H_{obs} = \Delta H_{bind}$) and that the interaction between IAA-94 and hGSTP1-1 was therefore not accompanied by a protonation-linked effect. *See section 2.2.9.3 for the experimental details pertaining to this method.*

3.3.4 The interaction between IAA-94 and the soluble conformation of EcSspA

Although soluble CLIC1-WT is structurally homologous with the cytosolic members from the GST superfamily, a particular focus was given to hGSTP1-1 in this study. However, soluble CLIC1-WT also shares structural homology with EcSspA, and similar to CLIC1-WT, EcSspA is a metamorphic protein that can reversibly transition from a soluble to a membrane-bound conformation where it will form a functional Cl^- channel. Furthermore, the ion channel activity of EcSspA has been shown to be inhibited by IAA-94 (Rao *et al*., 2017) (section 1.4). Therefore, it remains uncertain if the constraint in the chloride conductance of EcSspA is due to IAA-94 binding to its soluble conformation to prevent its structural transition into a membrane competent state, and/or IAA-94 binding to the membrane-bound conformation of EcSspA. However, similar to soluble CLIC1-WT and hGSTP1-1, the ability of soluble EcSspA to interact with IAA-94, as well as its structural analogue EAA, was investigated by ITC in order to understand these interactions from a thermodynamic perspective.

It was confirmed that EcSspA could bind IAA-94 with a stoichiometry of $n = 2.17$, therefore indicating that one molecule of IAA-94 bound to each subunit. The interaction was noted to be enthalpically favourable, but entropically unfavourable and therefore the overall interaction was enthalpically driven (Figure 49). This confirmed that the binding specificity of EcSspA for IAA-94 can be attributed to the number and/or strength of non-covalent protein-ligand interactions and also possibly ligand-solvent interactions. In the absence of structural data of EcSspA in complex with IAA-94 it is difficult to correlate how the obtained thermodynamic parameters reflect the structural changes upon ligand binding. However, it was noted that the L-site, or otherwise the MES-binding site, of hGSTP1-1 was structurally homologous in EcSspA (Table 4). Therefore, considering that the L-site was suggested to be the putative binding site for IAA-94, for reasons that will be made clear in subsequent sections, it was suggested that the interaction between EcSspA and IAA-94 occurred at this homologous site.

Importantly though, considering that the primary sequence for EcSspA shares 91% sequence similarity to YpSspA, the crystal structure from YpSspA (PDB: 1YY7) therefore provided good insight into the putative crystal structure of EcSspA. As a result, based on the homology model of EcSspA, it was observed that the putative IAA-94-binding site in EcSspA was located at Phe196 and when observing this site in YpSspA it was noted that not many water molecules surrounded this site. Therefore, like the binding of hydrophobic ligands to MUP (Duff and Howell, 2015), this desolvated binding site would not have contributed to a favourable entropic

change when IAA-94 interacted with EcSspA. Consequently, the unfavourable ΔS for the interaction between IAA-94 and EcSspA indicated that there was a reduction in the conformational degrees of freedom EcSspA and IAA-94 upon binding and as a result $\Delta S_{conf} + \Delta S_{comp} > \Delta S_{solv}$. Therefore, this marginal desolavtion event could possibly reflect not only the environment of the binding site (i.e. a "dry active site"), but also a decrease in the proportion of the hydrophobic surface area for IAA-94 being buried. In fact, this observation was further supported by the IFMD data (Table S4) which illustrated that in comparison to the IFMD of EAA to the H-site of hGSTP1-1, the Glide ligand efficiency score for the IFMD of IAA-94 to the putative IAA-94-binding site of EcSspA was less negative, therefore indicating that the surface area of IAA-94 had less of an affect towards the favourable binding interaction between IAA-94 and EcSspA.

However, without any structural data it cannot be discounted that the interaction between IAA-94 and EcSspA was also mediated by water molecules and therefore should this network of water molecules have become more ordered during the interaction, this would have contributed towards the entropic penalty. For instance, the binding of inhibitors to the TIE2 angiopoietin receptor and histamine binding protein (rRaHBP2) indicated that the formation of several hydrogen bonds and salt bridges resulted in a favourable ΔH that compensated for the entropic penalty of ordered water molecules within the buried pocket (Syme *et al*., 2010). Nevertheless, the entropic penalty for the interaction between IAA-94 and EcSspA renders the weak binding affinity with a K_d of 1 mM for this interaction (Figure 49).

Therefore, considering that IAA-94 can bind to the soluble conformation of EcSspA, it is possible that the constraint in the Cl^- conductance of EcSspA could be due to IAA-94 binding to both its soluble and membrane-bound conformations. However, the binding affinity for the interaction between IAA-94 and the soluble conformation of EcSspA is weak, therefore *in vivo* it appears that the interaction between IAA-94 and the membrane-bound conformation of EcSspA is a more plausible interaction to result in the inhibition of the ion channel activity of EcSspA. However, on that note, considering that EcSspA can function under acidic conditions, the effect that a change in pH would have on the interaction between IAA-94 and the soluble conformation of EcSspA needs to be further investigated.

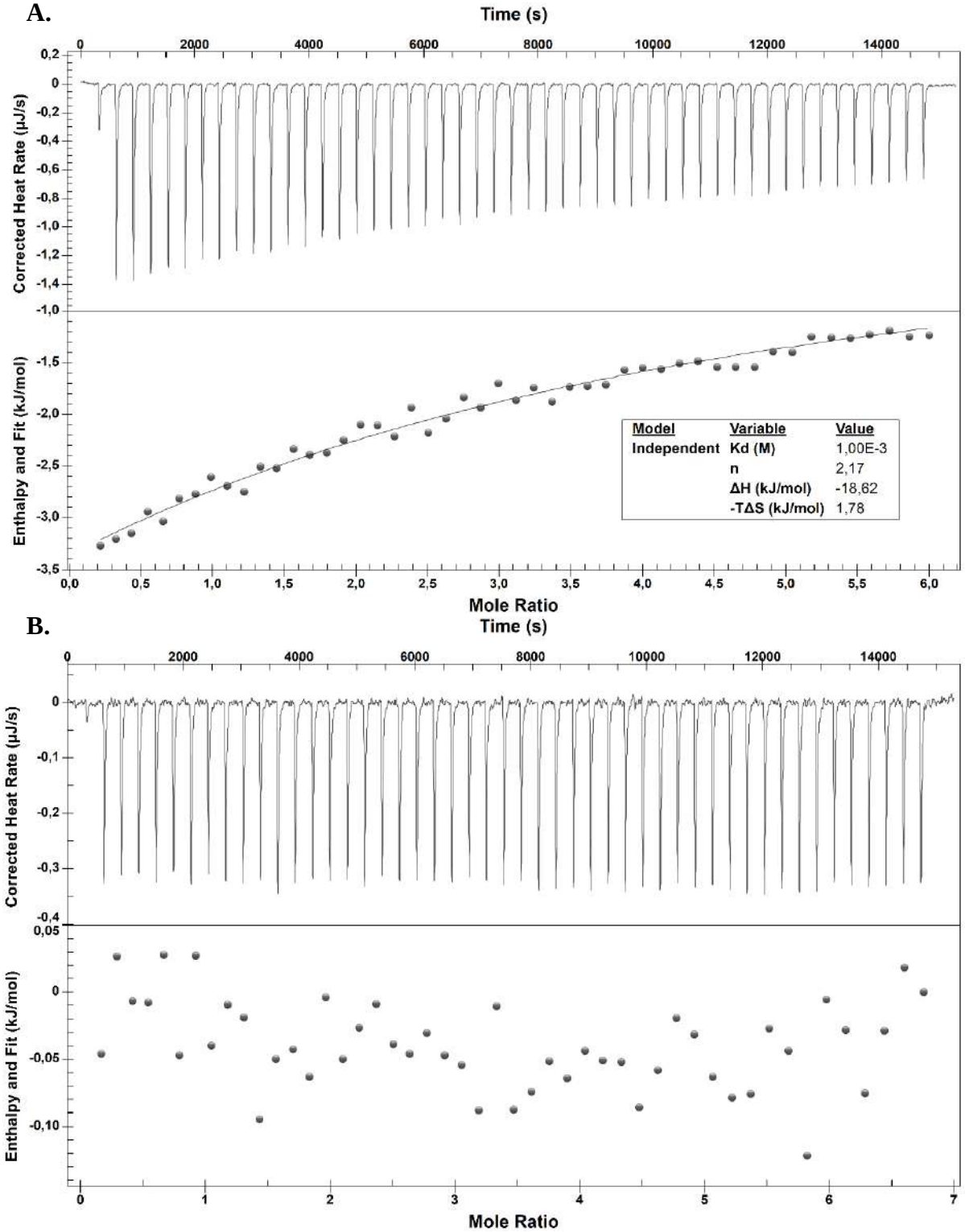


Figure 49. Monitoring the thermodynamic parameters for the interaction between IAA-94 or EAA and EcSspA. (A) Using ITC, the interaction between IAA-94 and EcSspA was monitored by titrating 2.1 mM IAA-94 into 100 µM EcSspA. Each titration was performed, in triplicate, at 20 °C with a stir rate of 250 rpm, an initial and final baseline of 300 s and a duration period of 300 s between each injection. The volume of the ITC syringe used was 250 µl and during each run the initial injection volume was 2 µl, but for all subsequent injections it was 5 µl and therefore a total of 49 injections were performed. The interaction occurred with a stoichiometry of $n \sim 2$, was entropically driven and had a weak K_d value in the millimolar range. (B) No interaction was observed between EAA and EcSspA. *See section 2.2.9.1 for the experimental details pertaining to this method.*

3.3.5 The interaction between EAA and the soluble conformation of EcSspA

Since the ITC experiments for CLIC1-WT and hGSTP1-1 involved a comparative study on the binding of IAA-94 and EAA to these target proteins, the same approach was adopted for EcSspA were the ability of EcSspA to interact with EAA was similarly investigated by ITC in order to understand this interaction from a thermodynamic perspective. It was noted that when EAA was titrated into EcSspA the heat produced correlated with the heat produced when EAA was titrated into buffer (Figure 50), therefore suggesting that there was no interaction between EAA and EcSspA. However, due to the phenomena of entropy-enthalpy compensation, the apparent absence of binding could have arisen due to the net ΔH and ΔS attenuating their contribution to the ΔG of binding by being equal in magnitude. Depending on the magnitude and whether the enthalpic and entropic terms for a reaction are favourable or unfavourable, various combinations of these thermodynamic terms can lead to a favourable ΔG of binding. Therefore, depending on whether (i) $\Delta H < 0$ and $\Delta S > 0$ or (ii) $\Delta H > 0$ and $\Delta S < 0$ an interaction will be spontaneous or nonspontaneous at all temperatures, respectively. However, when (i) $\Delta H < 0$ and $\Delta S < 0$ or (ii) $\Delta H > 0$ and $\Delta S > 0$ an interaction will only be spontaneous at low and high temperatures, respectively (Brown *et al.*, 2012). However, owing to the conformational stability of a protein being dependent on temperature, the latter thermodynamic parameters may not often be feasible for numerous biological systems.

Consequently, to conclude that the negligible ΔH_{obs} corresponded to no binding, the ITC experiment needed to be repeated at different temperatures considering that the ΔH for a system is directly correlated with a change in temperature. However, when the titration was repeated at both 10 °C and 30 °C the heat generated correlated once again with the HOD (Figure 50). Accordingly, based on ITC, it can be inferred that EAA did not interact with the soluble conformation of EcSspA and if it did, it did so very weakly.

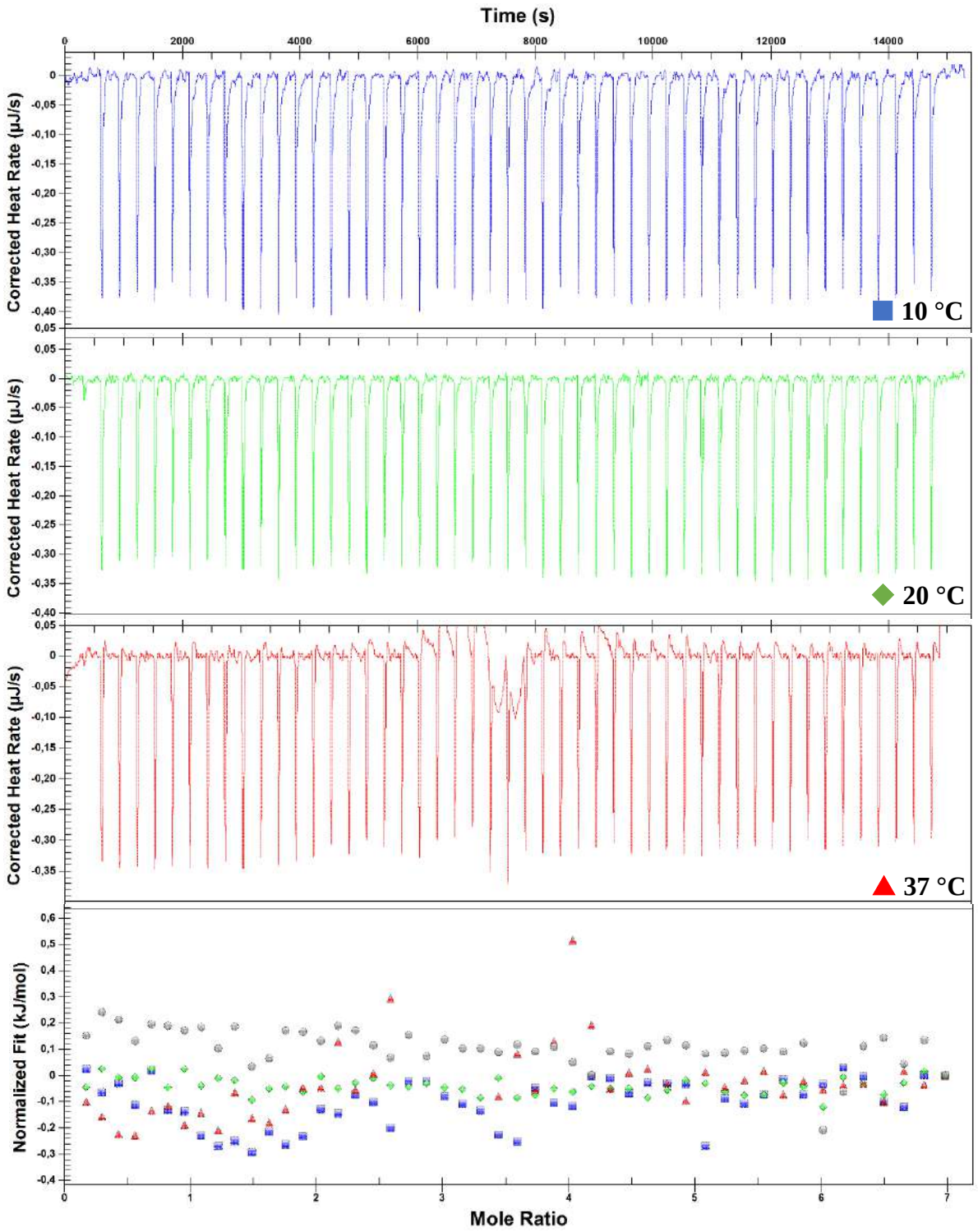


Figure 50. Monitoring the thermodynamic parameters for the interaction between EAA and EcSspA. Using ITC, the interaction between EAA and EcSspA was monitored by titrating 2.48 mM EAA into 100 µM EcSspA. Each titration was performed, in triplicate, at 20 °C (◆) with a stir rate of 250 rpm, an initial and final baseline of 300 s and a duration period of 300 s between each injection. The volume of the ITC syringe used was 250 µl and during each run the initial injection volume was 2 µl, but for all subsequent injections it was 5 µl and therefore a total of 49 injections were performed. In the event where no interaction was apparent the titration was repeated at both 10 °C (■) and 30 °C (▲). Indicated is also the HOD (●). *See section 2.2.9.1 for the experimental details pertaining to this method.*

3.4 Understanding protein-ligand interactions by IFMD and X-ray crystallography

3.4.1 Predicting the binding site for the interaction between hGSTP1-1 and IAA-94 by IFMD

It was concluded that hGSTP1-1 could bind to both EAA and IAA-94, whereas CLIC1-WT and EcSspA could only bind to EAA and IAA-94, respectively (section 3.3). Furthermore, CLIC1-WT, hGSTP1-1 and EcSspA displayed varying properties with regards to their CDNB enzyme activity, oligomeric state and ability to bind ANS and GSH (Table 2). Yet, considering that (i) the active site of soluble CLIC1-WT is structurally homologous with the active site of hGSTP1-1 (section 1.5), and that (ii) the MES-binding site of hGSTP1-1 was suggested to be the putative binding site for IAA-94, in order to try and understand the different mode of ligand binding at these sites, a comparative study of the structural homology between all three target proteins was undertaken.

Table 2: Summary of the functional characteristics of CLIC1, hGSTP1-1 and EcSspA

Characteristic properties	CLIC1-WT	hGSTP1-1	EcSspA
Oligomeric state	Monomeric[a]	Dimeric[bfgh]	Dimeric[c]
CDNB enzyme activity	✗[d]	✓[b]	✗[c]
Interacts with GSH	✗[e]	✓[f]	✗[c]
Interacts with ANS	✗[e]	✓[g]	✓[c]
Interacts with EAA	✓	✓[h]	✗
Interacts with IAA-94	✗	✓	✓

[a] Harrop *et al*. (2001).
[b] Oakley *et al*. (1999).
[c] Hansen *et al*. (2005).
[d] Khamici *et al*. (2015).
[e] Fanucchi *et al*. (2008).
[f] Oakley *et al*. (1997).
[g] Ralat *et al*. (2004).
[h] Oakley *et al*. (1996).

3.4.1.1 Predicating the EAA-binding site within the target proteins

In the crystal structure of hGSTP1-1 in complex with EAA (PDB: 2GSS), it was noted that Arg13 and Tyr108 were the two prominent residues involved in coordinating the interaction by forming a hydrogen bond and a π-π stacking interaction with EAA, respectively. At the active site of hGSTP1-1, Tyr108 not only played a pertinent role in the binding of EAA, but it also played a significant role in the catalytic function of hGSTP1-1 by stabilising the activated thiolate ion from GSH (Oakley *et al*., 1996). In fact, mutating Tyr108 had a significant reduction in the substrate specific activity for CDNB and EAA (Quesada-Soriano *et al*., 2009;

Quesada-Soriano *et al*., 2010). Therefore, considering the functional importance of Tyr108 and that a tyrosine residue is conserved at the active site across the pi-, mu-, alpha- and sigma-classes of cytosolic GSTs (Oakley, 2011), it was important to identify a site within CLIC1-WT and EcSspA that was structurally homologous to the EAA-binding site of hGSTP1-1 and that also contained a tyrosine (or a similar) residue which was structurally aligned or in close proximity to Tyr108. In fact, based on structural sequence alignment it was confirmed that Tyr233 and Tyr21 were within 8 Å of the putative EAA-binding site of CLIC1-WT and EcSspA, respectively. However, at this putative EAA-binding site there were no arginine residues present, but instead there were lysine residues which could be able to coordinate the negatively charged acetate moiety of EAA (Table 3).

Considering that there was no crystal structure available for CLIC1-WT in complex with EAA, IFMD was able to provide some insight into this interaction. The docked structure of CLIC1-WT in complex with EAA suggested that Lys183 formed a hydrogen bond with the carbonyl group of the 2-methylidenebutanoyl moiety of EAA, whereas Lys49 coordinated the acetate moiety of EAA. However, it was noted that the aromatic moiety of EAA was not coordinated by any π-π stacking interactions from Tyr233 as the distance between the aromatic rings of Tyr233 and EAA was found to be ∼12 Å. Yet, at the binding site for EAA there was a phenylalanine residue (Phe26) that could be a source for potential π-π stacking interactions (Figure 51). In fact the tenth pose out of the thirty that were generated during the IFMD suggested that Lys183 could possibly also form a cation-π stacking interaction. Therefore, although docking methods are efficient for predicting the mode of binding and the affinity for an interaction, the quality and reliability of docking predictions are limited by several shortcomings that are attributed to difficulties in accurately modelling the flexibility of a protein upon ligand binding and also due to the simplified scoring functions used in quantifying the protein-ligand binding energetics (Xu and Lill, 2013). Therefore, although IFMD did not indicate that Phe26 was involved in a π-π stacking interaction with EAA, in solution the protein backbone and side chains would be more dynamic and consequently this kind of interaction could occur to ensure the favourable interaction between EAA and CLIC1-WT.

At the putative EAA-binding site of EcSspA there was, however, only a single lysine (Lys43) and tyrosine (Tyr21) residue (Table 3), which therefore suggested that EcSspA could have interacted with EAA. However, based on the structure homology model of EcSspA, these residues are spatially not in close proximity to one another, but rather they are ∼16 Å apart.

Therefore, if Lys43 were to bind and coordinate the negatively acetate moiety of EAA, then Tyr21 would not be in close enough to stabilise EAA through a π-π stacking interaction and *vice versa*. Furthermore, based on a hydropathy plot it was illustrated that compared to CLIC1-WT and hGSTP1-1, this site was more polar in EcSspA (Figure 52). Therefore, due to defined geometric constraints of the peptide backbone and side chain resides, along with other physical-chemical properties such as the polarity of the surrounding site, it was suggested that the conformational change that would be required to bring Lys43 and Tyr21 in close proximity to one another would be unfavourable and may therefore account for the data obtained from ITC which indicated the absence of an interaction between EAA and EcSspA (Figure 50). This observation was further supported by the IFMD data (Table S4) which illustrated that compared to the IFMD of EAA to the active site of CLIC1-WT and hGSTP1-1, the IFMD of EAA to EcSspA produced an E_{model} score which was less negative (more unfavourable), but most importantly the G_{score} was less negative, therefore predicting that the binding of EAA to EcSspA occurred with a lower affinity.

Table 3: Structural sequence alignment of the residues that are within 8 Å of the EAA-binding site of hGSTP1-1

hGSTP1-1 (PDB: 2GSS)		hCLIC-WT (PDB: 1K0M)		YpSspA (PDB: 1YY7)		EcSspA (Homology model)	
AA No.	AA type	AA No.	AA type	AA No.	AA type	AA No.	AA type
7	Y	12	V	10	V	10	V
8	F	13	K[2]	11	M	11	M
9	P	14	A	12	T	12	T
10	V	15	G	13	L	13	L
11	R	16	S	14	F	14	F
12	G	22	G	15	S	15	S
13	R[2]	23	N	16	G	16	G
14	C	24	C	17	P	17	P
		25	P	18	T	18	T
		26	F[1]	19	D	19	D
		27	S	20	I	20	I
				21	F[1]	21	Y[1]
				22	S	22	S
33	V	46	V		-	41	V
34	T	-	-		-	42	E
35	V	-	-		-	43	K[2]
36	E	47	D		-	44	D
37	T	48	T		-	45	N
38	W	49	K	41	V	46	P
39	Q	50	R	42	E	47	P
				43	A	56	N
				44	D	57	Q
				45	N	58	S
				46	L		
				57	R[2]		
51	Q	63	E	58	T	-	-
52	L	64	L	59	V	59	V
53	P	65	P	60	P	60	P
65	S	77	T	72	S	72	S
97	E	108	L	104	H	104	H
100	R	111	F[1]	107	E	107	E
101	C	112	A	108	H	108	K
104	I	115	S	111	Y	111	Y
105	S	-	-	112	S	112	T
107	I	-	-	114	L	114	M
108	Y[1]	-	-	115	Y[1]	115	N
		-	-				
		183	K[2]				
		231	L				
202	P	232	A	-	-	209	L
203	I	233	Y[1]	201	L		
204	N	234	E	202	T		
205	G	235	Q	203	E		
206	N	236	V	204	A		
207	G	237	A	205	E		
RMSD (Å)		4.086		6.037		5.021	

[1]: Tyrosine (or similar) residues structurally aligned or in close proximity to Tyr108.

[2]: Arginine (or similar) residues structurally aligned or in close proximity to Arg13.

■: Residues involved in coordinating the interaction between EAA and hGSTP1-1.

□: Structurally conserved residues with respect to the EAA-binding site of hGSTP1-1.

■: Structurally aligned residues which are within 8 Å of the EAA-binding site of hGSTP1-1.

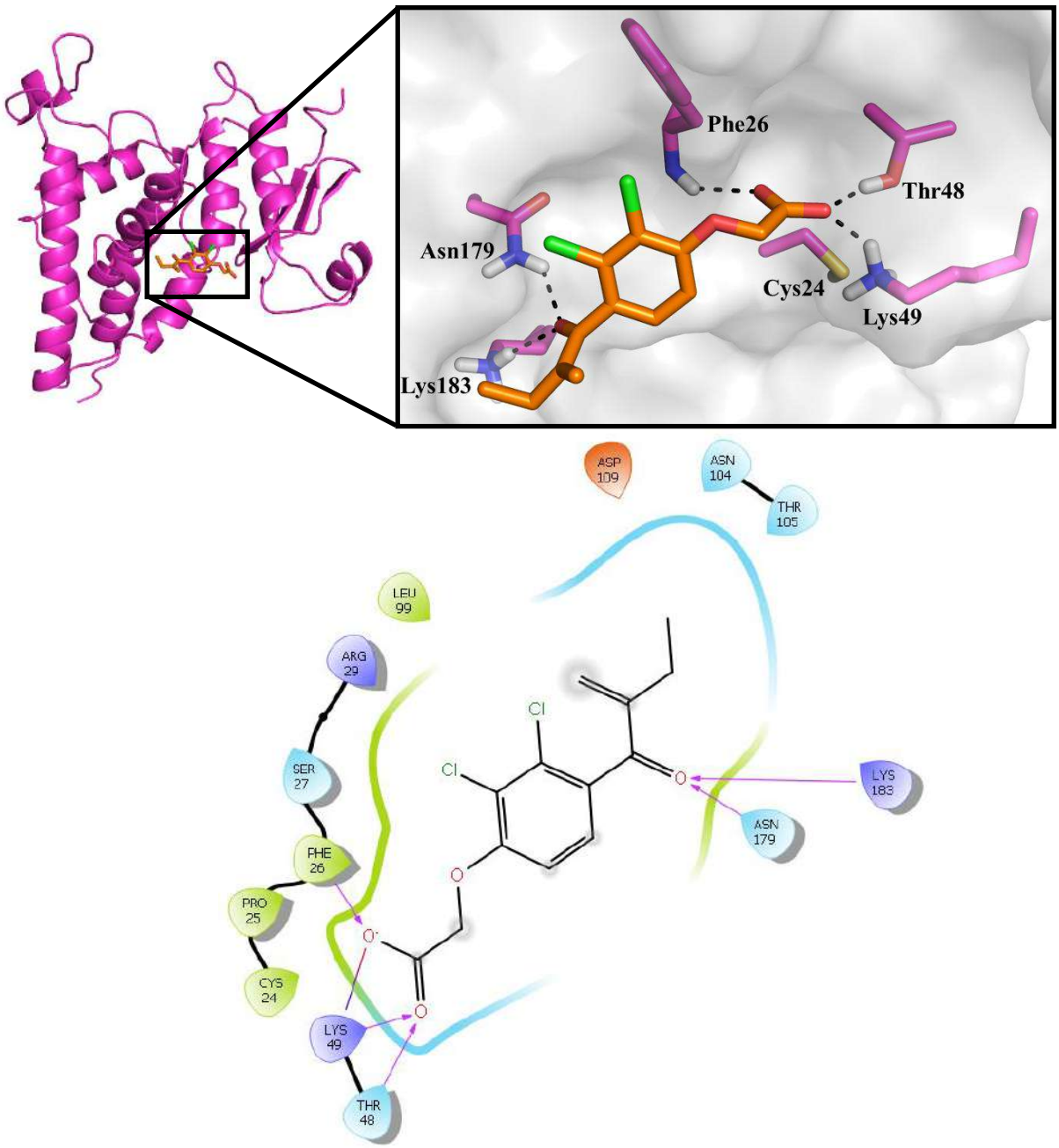
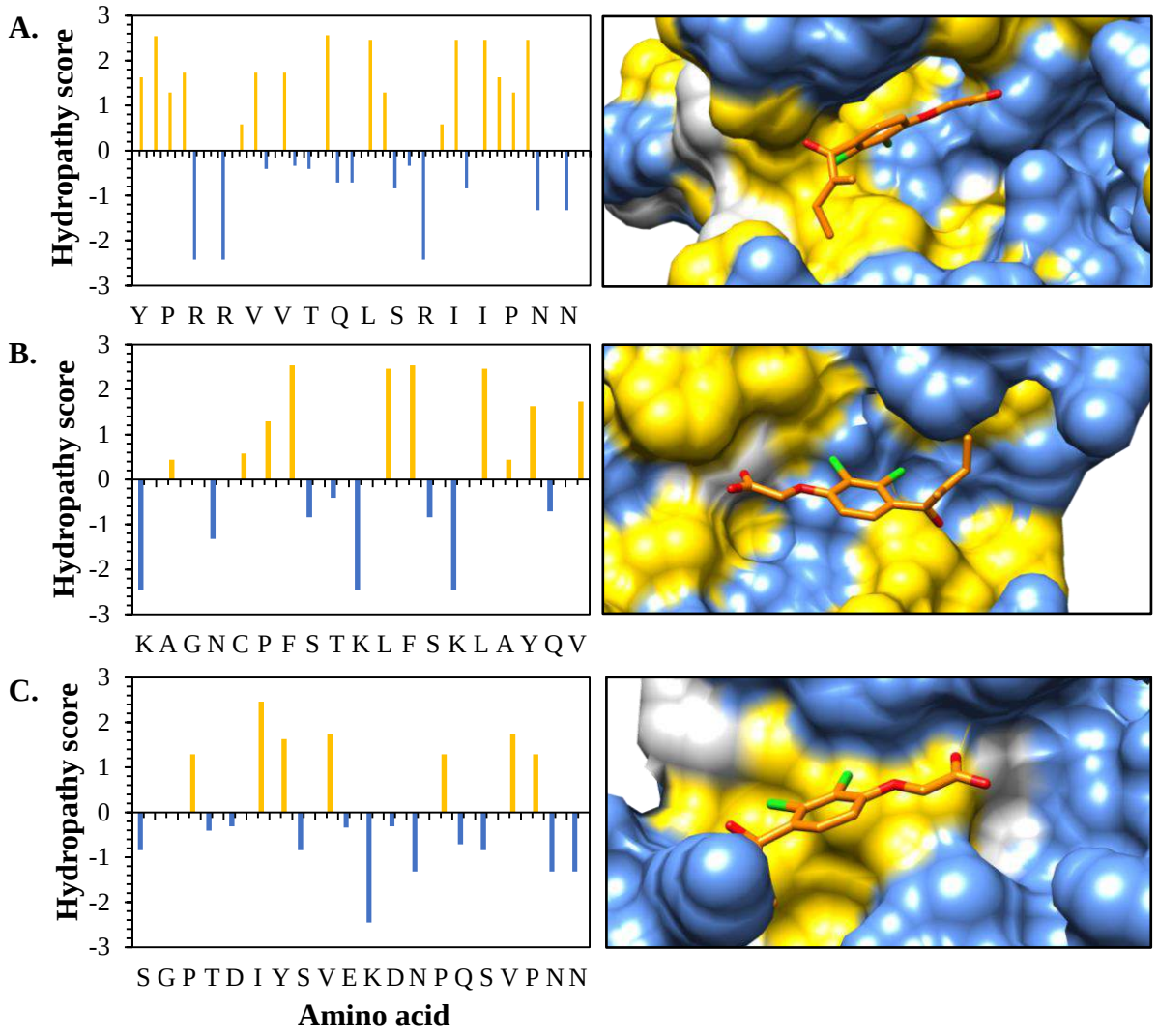


Figure 51. IFMD of EAA at the putative EAA-binding site of CLIC1-WT. The IFMD of EAA to the putative EAA-binding site of CLIC1-WT was validated by selecting the docked pose which has the lowest IFMD score. IFMD of CLIC1-WT in complex with EAA suggested that Lys183 formed a hydrogen bond with the carbonyl group of the 2-methylidenebutanoyl moiety of EAA, whereas Phe26, Thr48 and Lys49 coordinated the acetate moiety of EAA. The aromatic moiety of EAA, however, was not coordinated by any π-π stacking interactions from Tyr233 as the distance between the aromatic rings of Tyr233 and EAA was found to ~12 Å. However, in solution where CLIC1-WT will be more dynamic, Phe26 could be involved in a π-π stacking interaction with EAA. The GE_{model} and components of the G_{score} for the top pose can be found in Table S4. Indicated are positively charged residues (●), negatively charged residues (●), hydrophobic residues (●), polar residues (●), hydrogen bonds (→) and salt bridges (—). The top (3D diagram) and bottom (2D diagram) images for the protein-ligand interaction were generated by using Pymol V1.3 and the 2D protein-ligand interaction diagram tool in Mastero 11.2, respectively. *See section 2.2.13 for the experimental details pertaining to this method.*



Protein type	Polar residue hydropathy score	Non-polar residue hydropathy score	Ratio polar/non-polar residue
A. hGSTP1-1	- 14.50	28.42	0.51
B. CLIC1-WT	- 11.47	15.57	0.74
C. EcSspA model	- 10.17	11.42	0.89
D. YpSspA	- 7.07	15.57	0.46

Figure 52. Hydropathy plot of residues that are structurally aligned and within 8 Å of the EAA-binding site of hGSTP1-1. The polarity of the EAA (putative EAA)-binding site was assessed by an Abraham and Leo hydropathy plot (https://web.expasy.org/protscale/). The polar and non-polar residue hydropathy scores represent the sum of the hydropathy scores for all the individual polar and non-polar residues that were within 8 Å of the binding site, respectively. The polarity of the site was then assessed by taking a ratio of the polar residue hydropathy score to the non-polar residue hydropathy score. Although this site was polar both in (A) hGSTP1-1 and (B) CLIC1-WT, it was noted to be more polar in (C) the homology model of EcSspA. Indicated next to each hydropathy plot is the EAA (putative EAA)-binding site which shows EAA (●) and its surrounding polar (●) and non-polar (●) residues. All images were generated by using Chimera V1.11.2.

3.4.1.2 Predicating the putative IAA-94-binding site within the target proteins

In relation to the MES, or otherwise the putative IAA-94-binding site of hGSTP1-1, it was observed that there was significant structural homology for this site in EcSspA, but not in CLIC1-WT (Table 4). Furthermore, based on a hydropathy plot, it was noted that compared to hGSTP1-1 and EcSspA, this putative binding site was more non-polar in the case CLIC1-WT (Figure 53). Therefore, the difference in the amino acid composition at this site could account for the data obtained from ITC, which indicated the absence of an interaction between IAA-94 and CLIC1-WT (Figure 42). Furthermore, these findings correlated with the data obtained from IFMD, which indicated that compared to the IFMD of IAA-94 to the active site of CLIC1-WT, the scoring functions associated with the IFMD of IAA-94 to the putative IAA-94-binding site of CLIC1-WT were less negative (more unfavourable), therefore indicating that the interaction occurred with a lower affinity. However, the data obtained from the IFMD of IAA-94 to the active site of CLIC1-WT was in conflict with these findings considering that the scoring functions were more negative in comparison to the scoring functions associated with the IFMD of IAA-94 to the putative IAA-94-binding site of hGSTP1-1 and EcSspA (Table S4). Though bearing in mind that a major caveat in IFMD is being able to accurately model the flexibility of a protein upon ligand binding (Xu and Lill, 2013), the IFMD did not accurately account for the dynamic nature of CLIC1-WT and hence the interaction between CLIC1-WT and IAA-94 was overstated considering that this interaction was experimentally shown to not occur.

Based on the crystal structure of hGSTP1-1 in complex with MES (PDB: 2GSS), it was noted that MES formed a hydrogen bond with Glu197 at the MES-binding site of hGSTP1-1. Therefore, although the MES-binding site has been suggested to be the putative binding site for IAA-94, the binding of IAA-94 and MES to hGSTP1-1 was not competitive considering that in the presence of MES the ITC data for the interaction between IAA-94 and hGSTP1-1 did not show a significant difference in the thermodynamic parameters (Figure 48). However, as was the case for the interaction between CLIC1-WT and EAA, there was no crystal structure available of hGSTP1-1 in complex with IAA-94 and therefore IFMD was performed to provide some insight into this interaction. The top pose for the docked structure of IAA-94 at the MES-binding site of hGSTP1-1 suggested that IAA-94 was coordinated by Trp28 and Ser27 forming a hydrogen bond with the acetate moiety of IAA-94, while the indanyl group of IAA-94 was involved in a π-π stacking interaction with Phe192 (Figure 54). However, based on the other poses generated during IFMD it was also observed that the acetate functional group of IAA-94 could be coordinated by forming a hydrogen bond and/or salt bridge with either

Arg18, Lys29, Glu30 or Lys188, whereas Trp28 could also coordinate IAA-94 by either being involved in a π-π stacking interaction with IAA-94 or by forming a hydrogen bond with the carbonyl oxygen form the cyclopentyl moiety of the indanyl moiety present in IAA-94.

Therefore, it appeared that similar to Tyr108 which interacted with EAA, Phe192 played a key role to ensure the binding of IAA-94 to hGSTP1-1. In fact, although Phe192 is not structurally aligned with Phe196 from EcSspA, these two residues are structurally adjacent to one another. Therefore, considering that Phe192 is structurally aligned with Phe196 from YpSspA and that Phe192 is structurally adjacent to Phe196 from EcSspA (Table 4), it appears as if this conserved aromatic residue plays a key role in the binding of IAA-94 to hGSTP1-1 and EcSspA. Yet, in view that MES also binds to this site, IFMD was performed to assess how the docked poses of MES compared to those found in the crystal structure of hGSTP1-1 (PDB: 2GSS). Based on IFMD, it was found that the docked structure of MES had its 2-(N-morpholino) moiety coordinated by a salt bridge and hydrogen bond with Glu197, along with a π-cation interaction with Trp28 (Figure 55), therefore correlating with previous findings in the crystal structure of hGSTP1-1 (PDB: 2GSS). However, in the docked structure of MES it was observed that its sulfonate moiety was coordinated by a hydrogen bond with Glu30 (Figure 55) as opposed by water molecules which were identified in the crystal structure of hGSTP1-1 (PDB: 2GSS).

In view of the fact that IAA-94 could still bind to hGSTP1-1 in the presence of MES, it was therefore also important to evaluate the binding conformation of IAA-94 and MES when in complex together with hGSTP1-1. In the presence of MES that had been previously docked to hGSTP1-1, the indanyl group of IAA-94 was coordinated by the 2-(N-morpholino) moiety of MES through a cation-π stacking interaction as opposed to Phe192, whereas the acetate moiety was coordinated by Ser27, Glu197 and/or the sulfonate group of MES through a hydrogen bond (Figure 56). The absence of a competitive mode of binding therefore correlated with the ITC data given that the thermodynamic parameters for the interaction between IAA-94 and hGSTP1-1 did not change significantly (Figure 48). Interestingly, however, when the conformation of MES from the crystal structure of hGSTP1-1 (PDB: 2GSS) was used during IFMD, it was found that IAA-94 was displaced for the putative IAA-94-binding site and instead the acetate moiety of IAA-94 was coordinated to hGSTP1-1 by forming a hydrogen bond with Lys29, the 2-(N-morpholino) moiety of MES and/or through a water network (Figure 57). Therefore, to reliably resolve the interaction between hGSTP1-1 and IAA-94, a crystal structure of hGSTP1-1 in complex with IAA-94 needed to be obtained.

Table 4: Structural sequence alignment of the residues that are within 8 Å of the MES, or otherwise the putative IAA-94-binding site of hGSTP1-1

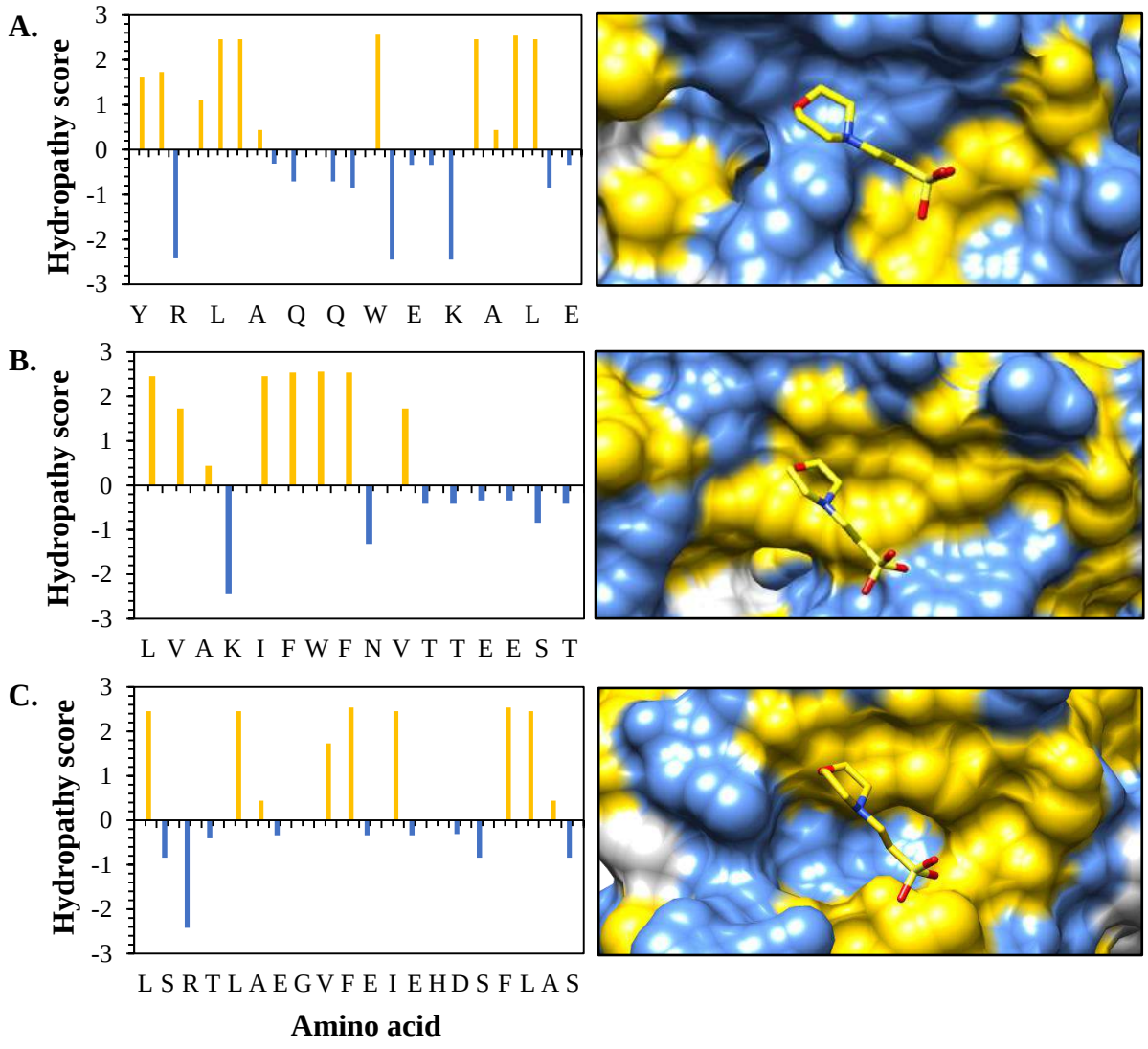
hGSTP1-1 (PDB: 2GSS)		hCLIC-WT (PDB: 1K0M)		YpSspA (PDB: 1YY7)		EcSspA (Homology model)	
AA No.	AA type	AA No.	AA type	AA No.	AA type	AA No.	AA type
3	Y	8	V	6	N	6	N
5	V	10	L	8	R	8	R
		12	V			11	M
		19	A	13	L	13	L
		20	K	15	S		S
		21	I				
18	R	31	F	26	R	26	R
19	M	32	M	27	I	27	I
20	L	33	V	28	V	28	V
21	L	34	L	39	L	29	L
22	A	35	W	30	A	30	A
23	D	36	L	31	E	31	E
24	Q	37	K	32	K	32	K
25	G	38	G	33	G	33	G
26	Q	39	V	34	V	34	V
27	S	40	T	35	S	35	S
28	W	41	F	36	V	36	F
29	K	42	N	37	E	37	E
30	E	43	V	38	I	38	I
31	E	44	T	39	E[1]	39	E[1]
		45	T	40	Q	40	H
		217	E[1]	194	D	194	D
188	K	218	E[1]	-	-	195	S
189	L	219	F	-	-	196	F
						197	L
191	A	221	S	195	A	198	A
192	F	222	T	196	F	199	S
193	L	223	C	197	L	200	L
				198	A	201	T
195	S	225	D	199	S	202	E
197	E[1]	227	E	-	-	204	E
				200	L	205	R
				201	T		
				205	R		
RMSD (Å)		4.086		6.037		5.021	

[1]: Glutamate (or similar) residues structurally aligned or in close proximity to Glu197.

■: Residues involved in coordinating the interaction between MES and hGSTP1-1.

□: Structurally conserved residues with respect to the MES-binding site of hGSTP1-1.

■: Structurally aligned residues which are within 8 Å of the MES-binding site of hGSTP1-1.



Protein type	Polar residue hydropathy score	Non-polar residue hydropathy score	Ratio polar/non-polar residue
A. hGSTP1-1	- 11.75	20.28	0.58
B. CLIC1-WT	- 6.52	16.46	0.40
C. EcSspA model	- 9.52	17.53	0.54
D. YpSspA	- 9.38	17.16	0.55

Figure 53. Hydropathy plot of residues that are structurally aligned and within 8 Å of the MES (putative IAA-94)-binding site. The polarity of the MES (putative IAA-94)-binding site was assessed by an Abraham & Leo hydropathy plot (https://web.expasy.org/protscale/). The polar and non-polar residue hydropathy scores represent the sum of the hydropathy scores for all the individual polar and non-polar residues that were within 8 Å of the binding site, respectively. The polarity of the site was then assessed by taking a ratio of the polar residue hydropathy score to the non-polar residue hydropathy score. This site was found to be polar both within (A) hGSTP1-1 and (C) the homology model of EcSspA, whereas it was non-polar in (B) CLIC1-WT. Indicated next to each hydropathy plot is the MES (putative IAA-94)- binding site which shows MES (●) and its surrounding polar residues (●) and non-polar residues (●). All images were generated by using Chimera V1.11.2.

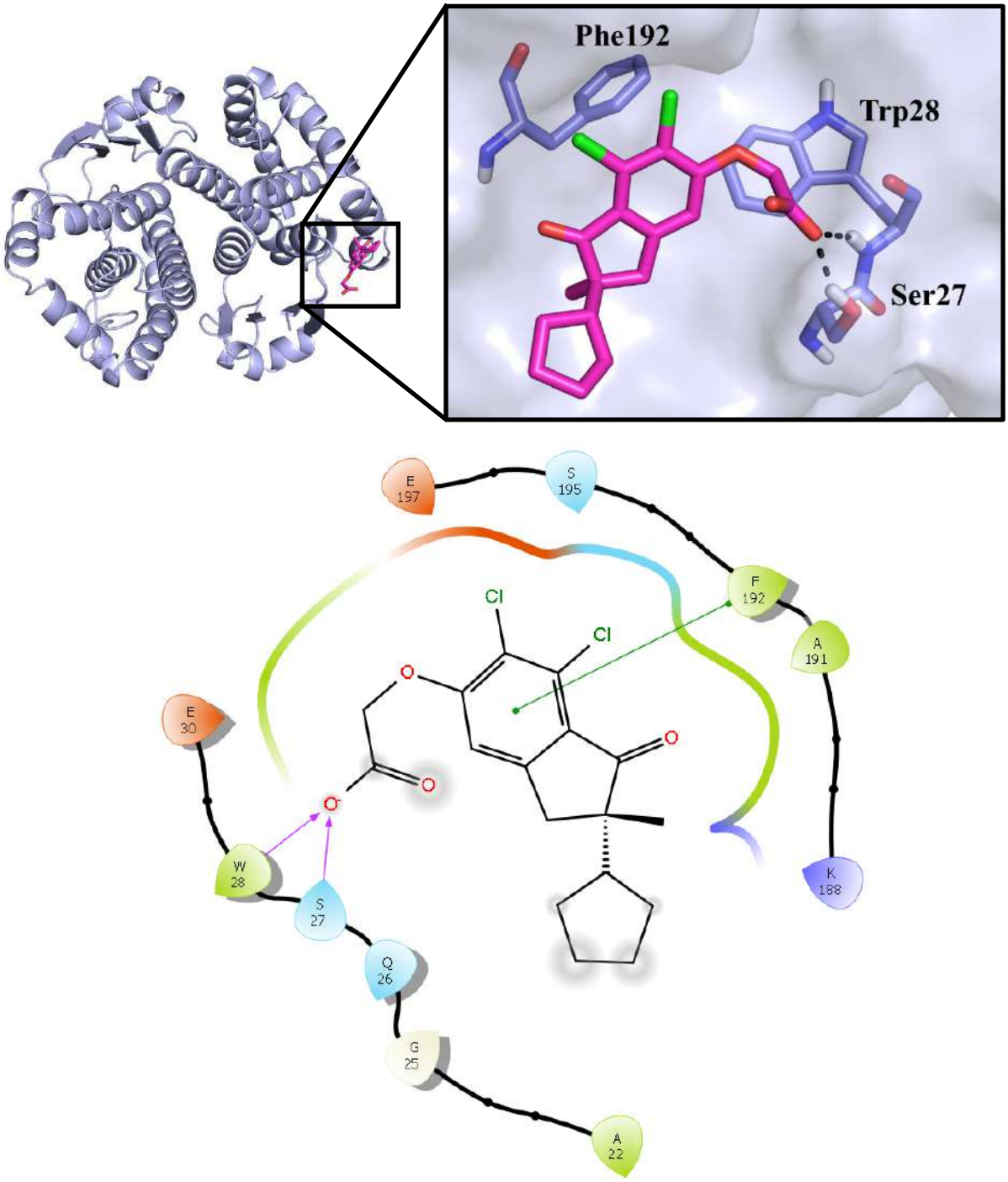


Figure 54. IFMD of IAA-94 at the MES (putative IAA-94)-binding site of hGSTP1-1. The IFMD of IAA-94 at the putative IAA-94-binding site of chain B from hGSTP1-1 was validated by selecting the docked pose which had the lowest IFMD score. IFMD indicated that the binding of IAA-94 to hGSTP1-1 was coordinated by Trp28 and Ser27 forming a hydrogen bond with the acetate moiety of IAA-94, while the indanyl group of IAA-94 was involved in a π-π stacking interaction with Phe192. The GE_{model} and G_{score} for the top pose can be found in Table S4. Indicated are positively charged residues (●), negatively charged residues (●), hydrophobic residues (●), polar residues (●), hydrogen bonds (→) and π-π interactions (—). The top (3D diagram) and bottom (2D diagram) images for the protein-ligand interaction were generated by using Pymol V1.3 and the 2D protein-ligand interaction diagram tool in Mastero 11.2, respectively. *See section 2.2.13 for the experimental details pertaining to this method.*

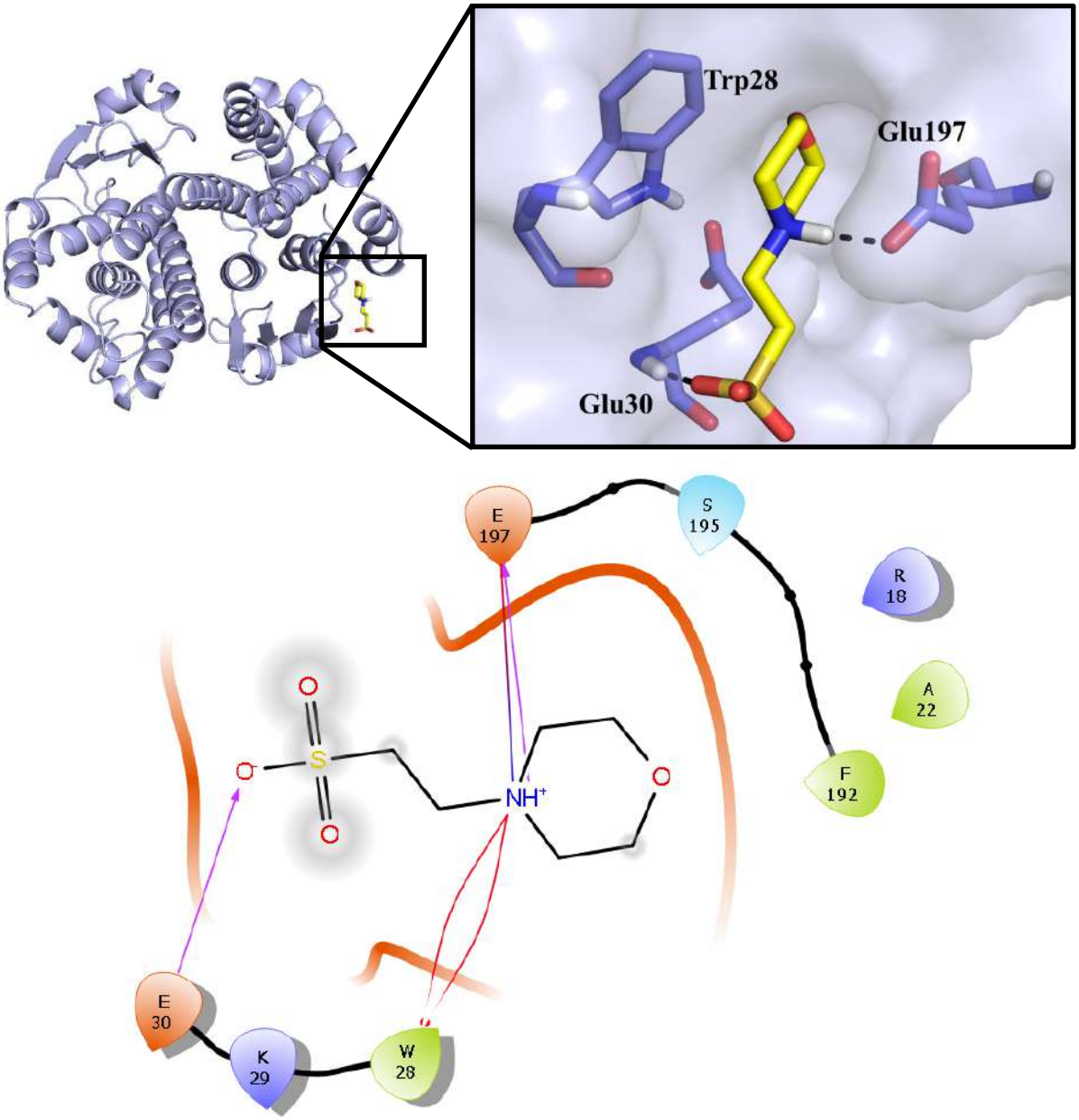


Figure 55. IFMD of MES at the MES-binding site of hGSTP1-1. The IFMD of MES at the MES-binding site of chain B from hGSTP1-1 was validated by selecting the docked pose which had the lowest IFMD score. IFMD of hGSTP1-1 in complex with MES suggested that the 2-(N-morpholino) moiety of MES was coordinated by a salt bridge and hydrogen bond with Glu197, along with a π-cation interaction with Trp28, therefore correlating with previous findings in the crystal structure of hGSTP1-1. Yet, in the docked structure the sulfonate moiety of MES was coordinated by a hydrogen bond with Glu30 as opposed by water molecules which were identified in the crystal structure of hGSTP1-1 (PDB: 2GSS). The GE_{model} and components of the G_{score} for the top pose can be found in Table S4. Indicated are positively charged residues (●), negatively charged residues (●), hydrophobic residues (●), polar residues (●), hydrogen bonds (→), salt bridges (—), π-π interaction (—) and π-cation interaction (—).The top (3D diagram) and bottom (2D diagram) images for the protein-ligand interaction were generated by using Pymol V1.3 and the 2D protein-ligand interaction diagram tool in Mastero 11.2, respectively. *See section 2.2.13 for the experimental details pertaining to this method.*

Pose	IFMD scoring functions		
	GE_{model}	IFMD score	G_{score}
	(kcal.mol^{-1})		
1	- 43.672	- 892.568	- 5.367
2	- 38.109	- 892.236	- 5.170
3	- 41.979	- 892.113	- 4.917
4	- 35.490	- 892.072	- 5.053
5	- 33.383	- 891.297	- 3.809
6	- 32.173	- 891.213	- 3.966

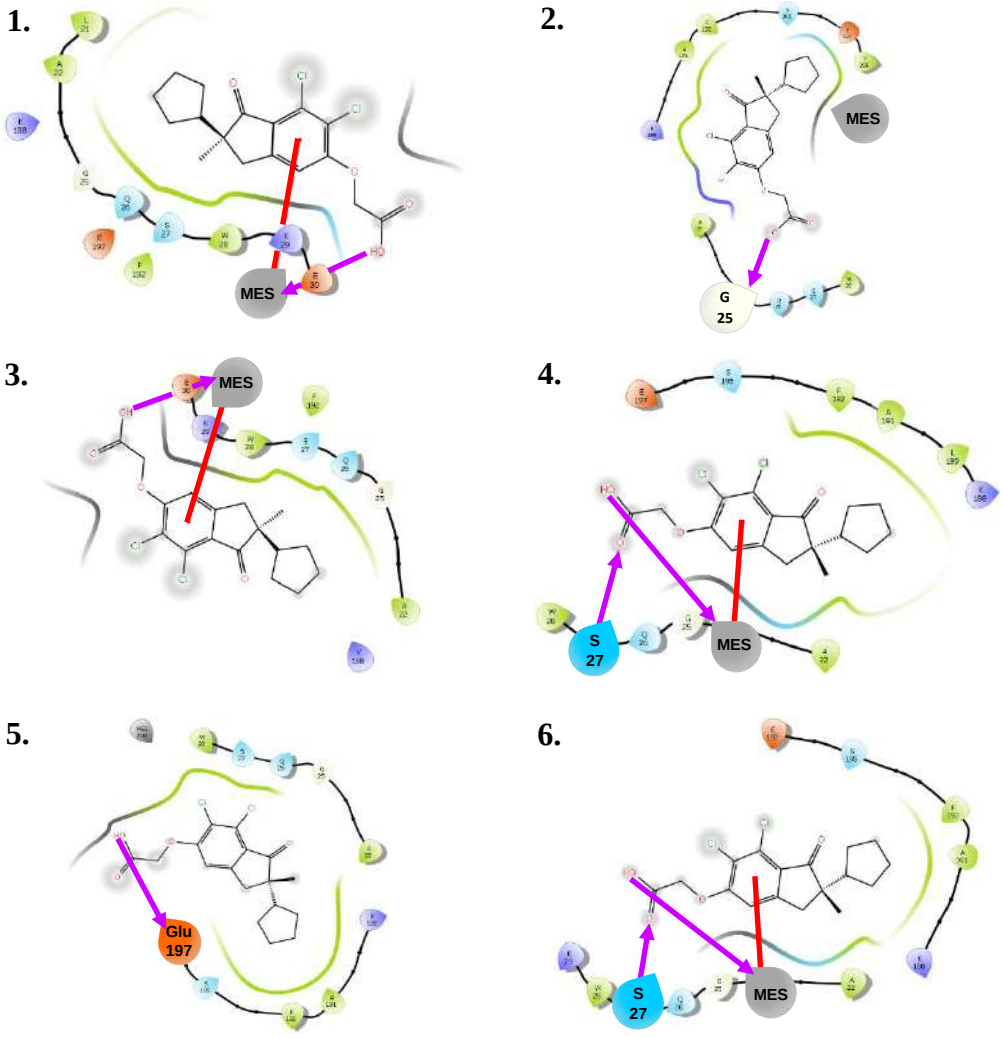


Figure 56. Top six IFMD poses for IAA-94 in complex with MES which had previously been docked to hGSTP1-1. The binding conformation of IAA-94 together with MES which had previously been docked to hGSTP1-1 indicated that the the indanyl group of IAA-94 could be coordinated by the 2-(N-morpholino) moiety of MES through a cation-π stacking interaction as opposed to Phe192, whereas the acetate moiety was coordinated by Ser27, Glu197 and/or the sulfonate group of MES through a hydrogen bond. The images were generated by using the 2D protein-ligand interaction diagram tool in Mastero 11.2. *See section 2.2.13 for the experimental details pertaining to this method.*

Pose	IFMD scoring functions		
	GE_{model}	IFMD score	G_{score}
	(kcal.mol^{-1})		
1	- 32.461	- 897.228	- 4.346
2	- 42.859	- 897.100	- 4.998
3	- 34.853	- 896.767	- 4.555
4	- 33.491	- 896.659	- 4.182
5	- 31.747	- 896.483	- 3.720
6	- 30.685	- 896.323	- 3.640

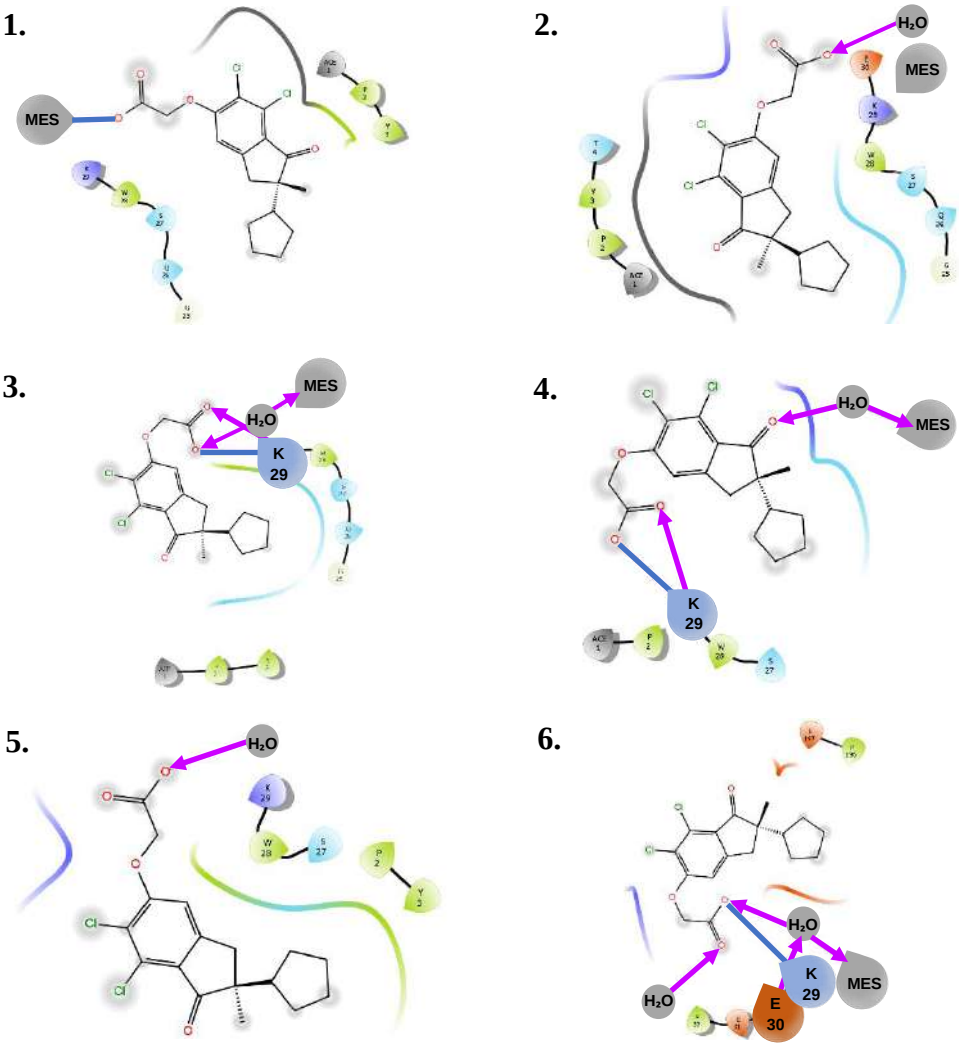


Figure 57. Top six IFMD poses for IAA-94 in complex with MES from the crystal structure of hGSTP1-1. The binding conformation of IAA-94 together with MES from the crystal structure of hGSTP1-1 indicated that IAA-94 was displaced for the putative IAA-94-binding site and instead the acetate moiety of IAA-94 was coordinated to hGSTP1-1 by forming a hydrogen bond with Lys29, the 2-(N-morpholino) moiety of MES and/or through a water network. The images were generated by using the 2D protein-ligand interaction diagram tool in Mastero 11.2. *See section 2.2.13 for the experimental details pertaining to this method.*

3.4.2 Determining the structure of IAA-94 bound to hGSTP1-1 by X-ray crystallography

3.4.2.1 The crystallisation of IAA-94 in complex with hGSTP1-1

The crystal structure of hGSTP1-1 in complex with various ligand molecules has been widely reported in the PDB, therefore by modifying these pre-existing crystallisation conditions, the aim was to obtain a crystal structure of hGSTP1-1 in complex with IAA-94. However, due to the limited solubility of IAA-94 ($\leq$ 25 mg.ml^{-1}) in an aqueous environment, DMSO was used to solubilise IAA-94, but the drawback to this method was the limited ligand concentration given that the concertation of DMSO needed to be kept $\leq$ 5% (v/v) in order to avoid damaging the structural integrity of hGSTP1-1, but also because this organic solvent could have interfered with the crystallisation process. In an attempt to overcome this problem, hGSTP1-1 was incubated with the solid compound of IAA-94 (25 mg.ml^{-1}) for an hour and although the solution was clear with no visible precipitate, when this solution was used to perform crystal trials it immediately resulted in the precipitation of hGSTP1-1 and/or IAA-94 out of solution and therefore this method was unsuccessful. As a result, hGSTP1-1 needed to be either co-crystallised with IAA-94 or the resultant crystals needed to be soaked with IAA-94 which had been solubilised in DMSO.

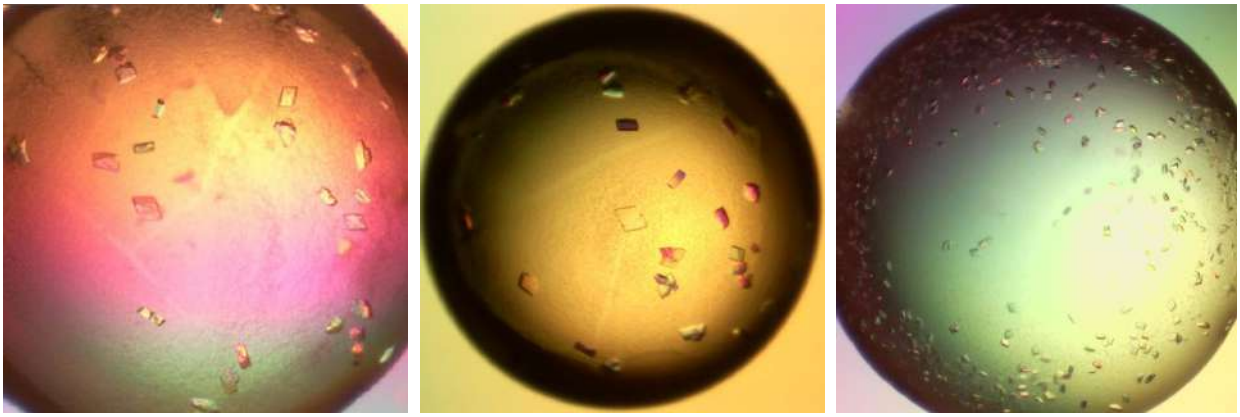
The successful complexation of a ligand to its target protein depends on the protein and ligand concentration and ultimately the K_d value for the respective protein-ligand complex. Therefore, considering that the K_d value for the interaction between hGSTP1-1 and IAA-94 was observed to be 47 µM and that the concertation of hGSTP1-1 used during the crystal trials was 215 µM (10 mg.ml^{-1}), a concentration of 3.5 mM IAA-94 (5% (v/v) DMSO) used during the crystal trials was sufficient to saturate the protein and achieve full occupancy. However, the crystallisation of hGSTP1-1 often required the presence of GSH and this could be attributed to the highly dynamic α-helix 2 at the active site of hGSTP1-1. That is, when the G-site of hGSTP1-1 becomes occupied by GSH it results in Cys47 becoming unreactive as a result of its thiol group becoming buried in a hydrophobic patch when α-helix 2 is in its closed and more stable conformation (Hitchens *et al.*, 2001; Oakley *et al.*, 1998). Therefore, in an attempt to promote the crystallisation process of hGSTP1-1 in complex with IAA-94, the protein stock solution was incubated initially with 10 mM GSH.

The crystallisation of hGSTP1-1 in the presence of IAA-94 was observed across various conditions (Figure 58, Table S5), however, possibly due to DMSO and/or IAA-94 having an

influence on the crystallisation process, the success rate was lower compared to the apo-conformation of hGSTP1-1 whereby protein crystals were present for most of the conditions. However, in an attempt to soak the protein crystals for the apo-conformation of hGSTP1-1 with the solution of IAA-94, the protein crystals dissolved and therefore crystals of hGSTP1-1 that were co-crystallised with IAA-94 had to be used for diffraction. A total of 16 crystals were harvested from various conditions and sent to the DLS for diffraction, however, the protein crystal harvested from condition B (100 mM MES, 28% (w/v) PEG-6000, 20 mM calcium chloride, 10 mM DTT and 10 mM GSH, pH 5.5) provided suitable diffraction data with a resolution of 1.40 Å and therefore the data was used downstream for solving the structure of hGSTP1-1 in complex with IAA-94 (Figure 58 and 59).

Condition A

Condition B



Condition C

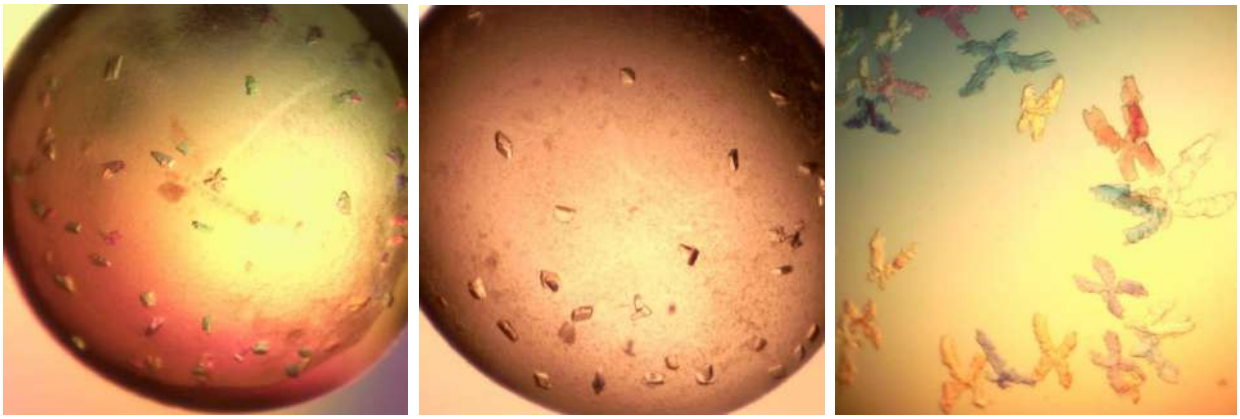


Figure 58. Crystallisation of hGSTP1-1 in complex with IAA-94. Crystal trials of hGSTP1-1 in the presence of IAA-94 was performed at 20 ° C by using 215 µM (10 mg.ml^{-1}) hGSTP1-1 and 3.5 mM IAA-94 (5% (v/v) DMSO). The formation of protein crystals could be noted across various crystal conditions, however, those which yielded suitably sized crystal for diffraction could be seen in condition A (100 mM MES, 16% (w/v) PEG-6000, 10 mM DTT and 10 mM GSH, pH 6.5), condition B (100 mM MES, 28% (w/v) PEG-6000, 20 mM calcium chloride, 10 mM DTT and 10 mM GSH, pH 5.5) and condition C (100 mM MES, 22% (w/v) PEG-6000, 20 mM calcium chloride, 10 mM DTT and 10 mM GSH, pH 6.0). The protein crystals were harvested by using the CrystalCap™ SPINE HT goniometer base that was fitted with either a 0.05-0.1 mm, 0.1-0.2 mm or a 0.2-0.3 mm Mounted CryoLoop™. The crystals were then immersed two to three times in 100% (v/v) Parabar 10312 (Paratone) and frozen in liquid nitrogen. *See section 2.2.10.1 for the experimental details pertaining to this method.*

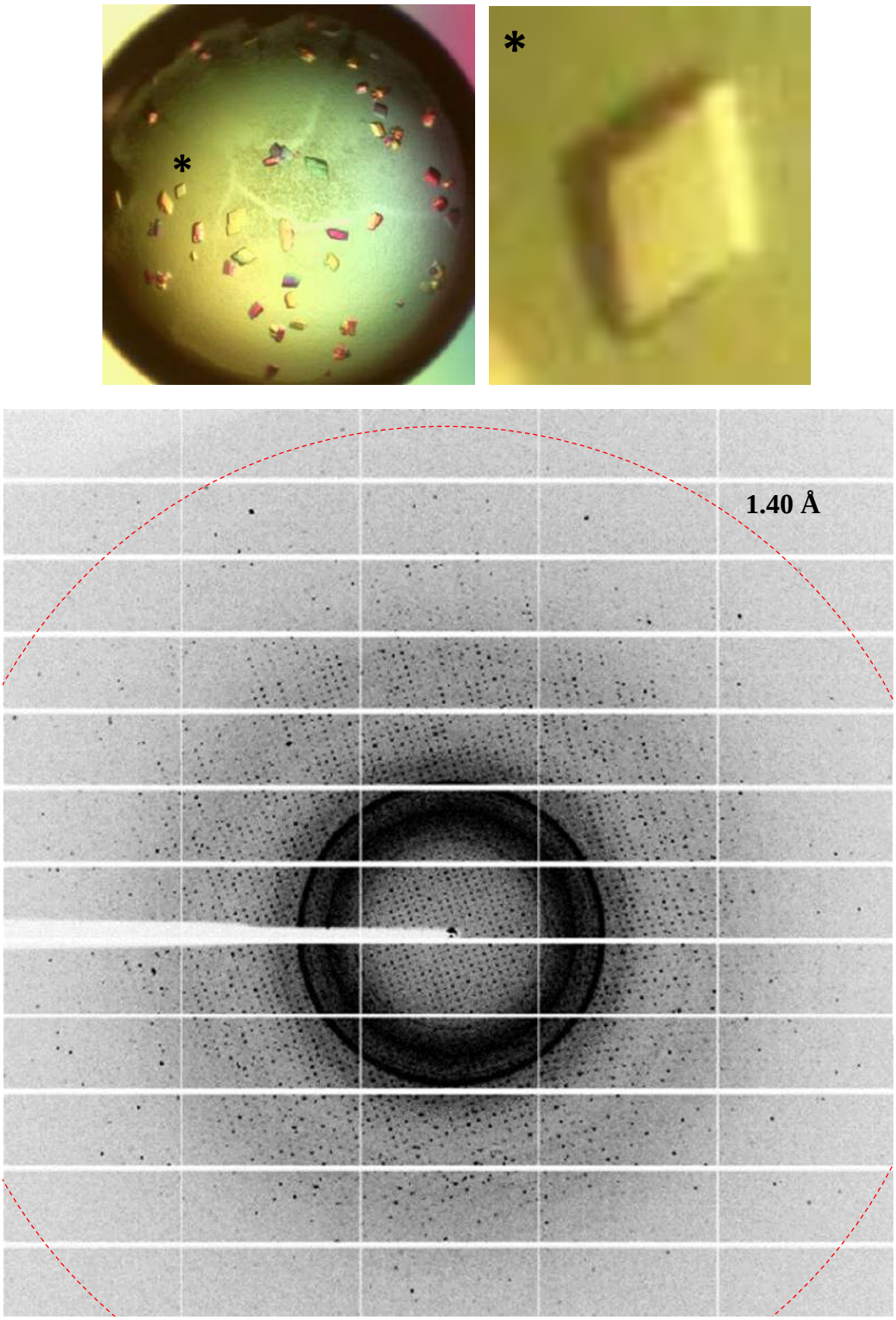


Figure 59. Diffraction of the protein crystals for hGSTP1-1 in complex with IAA-94. Protein crystals that were harvested from various conditions were used for diffraction, however, the diffraction data obtained from the protein crystal harvested from condition A2 were used for subsequent data processing. The X-ray diffraction data from the harvetesed crystals were generated by using the I03 beamline at the DLS which has a working wavelength of 0.915 Å and a focused beam size of 90 x 20 µm and 90 x 30 µm at the sample and detector, respectively. The data were collected by using the Dectris Eiger2 XE 16M detector at 100 K while collecting a total of 3500 images. A total of 163 749 unique reflections were collected with a completeness of 96.8% and the resolution of the lowest resolution shell was reported to be 1.4 Å with I/σ (I) = 0.6 for the highest resolution shell.

3.4.2.2 Evaluating the quality of the diffraction data and the refined protein structure

The crystallographic unit cell by having the dimensions of $a \neq b \neq c$ and $\alpha = \gamma = 90° \neq \beta$ formed part of the monoclinic crystal system and had a primitive lattice type (Table 5). Prior to analysing the structure of a protein molecule, it was essential to know how many copies of the asymmetric unit were present in unit cell, which relied on knowing both the crystal solvent content as well as the protein density. It is known that 40-60% of the volume for most protein crystals is occupied by solvent and that the protein density, which is reflected by the partial specific volume ($\overline{V}_p$) that defines the increase in volume of the solution per a unit mass of protein, is constant and typically 0.74 $cm^3.g^{-1}$ for most protein molecules. Collectively, the Matthew's coefficient (V_M), or otherwise the crystal volume per unit of protein molecular weight, showed a straightforward relationship to the fraction of the unit cell that was occupied by solvent and was therefore able to give information on the asymmetric unit copy number as follows (Sherwood and Cooper, 2011):

$$V_M = \frac{V_s}{V} = 1 - \frac{n M \overline{V}_p}{NV} \quad \textbf{(16)}$$

where n represents the number of symmetry-related molecules in the unit cell (asymmetric unit copy number), M represents the molecular weight (Da), $\overline{V}_p$ represents the partial specific volume, V_s represents the volume of the unit cell that is occupied by solvent (cm^3), V represents the volume of the unit cell (cm^3) and N represents Avogadro's number (6.02×10^{23} particles.mol^{-1}).

The V_M value and solvent content for the diffracted protein crystal was reported to be 2.33 $Å^3.Da^{-1}$ and 47.1%, respectively and collectively the asymmetric unit copy number was reported to be n = 4, indicating that present within the unit cell were four copies of the asymmetric unit. Furthermore, the symmetrical orientation of these molecules were related by a $P12_11$ space group (Table 5), therefore indicating that there was no symmetry element in the x-direction and z-direction, but instead there was a two-fold screw axis (2_1) that was parallel to the y-direction, meaning that one copy of the asymmetric unit was first rotated 180 ° about a two-fold axis and then it was translated along half of the axis in the y-direction.

When the diffraction data was collected from the protein crystal, there were a number of practical factors that could have influenced the measured intensity for each reflection, for

instance the primary beam intensity, the exposure time and very importantly radiation damage owing to ionisation radiation in the protein crystal that was brought about by the intense energy of the incoming X-ray beam. Initially, the diffraction data from the area detector was interpreted by converting the value of each pixel into the corresponding number of X-ray photons. The software importantly was then able to detect the pixel value for each reflection by monitoring the variation of the pixel values in the background region, therefore when the background was not significantly constant it presented noise in the measured intensities for each reflection (Sherwood and Cooper, 2011).

In summary, the intensity for each spot was therefore given by the difference in the total peak count (P) and the background counts (B), however, taking into account the propagation of errors and that P and B adhered to a Poisson distribution, the standard deviations of the spot intensities (σ (I)) could simply be determined by taking the square root to the sum of P and B. The data-processing programs had to then be able to estimate σ (I) given that by σ (I) indicating the noise in the measured intensities, it was used as a weighting factor when the data from the different diffraction images were scaled and merged together. Therefore, given the fact that estimating the uncertainties within the diffraction data is often a challenge, the signal-to-noise (I/σ (I)) ratio may not always be completely informative (Powell, 2017; Sherwood and Cooper, 2011). Nevertheless, the diffraction limit was defined at a resolution where the I/σ (I) value decreased to 2.0, therefore considering I/σ (I) was reported to be 0.6 at the highest resolution shell in this study (Table 5), it can be assumed that the true resolution at this resolution shell was not as good (Wlodawer *et al*., 2008). Yet, it has been postulated that these criteria may be too conservative and that even weaker high-resolution data can still improve the refined model (Karplus and Diederichs, 2012).

Due to the fact that there were various factors which could have influenced the measured intensity for a particular reflection, the consequence was that the measured intensity for a particular reflection at the beginning of diffraction was not the same at the end of the experiment. Therefore, seeing that during data processing the intensity of symmetrically related reflections were averaged and that these reflections should be of equal magnitude based on their symmetry on the diffraction pattern (Laue group), it was clear that the decrease in intensity overtime could have presented a challenge. Therefore, it was imperative to correct for these interference factors prior to the intensities of these symmetry related reflections being scaled and merged. To achieve this it was assumed that because the factors affecting the intensity of

a reflection are correlated with time, a scale and temperature factor were ascribed to each diffraction image. Therefore, taking into account these factors, the intensity for each reflection was then assigned a mean intensity $\langle I_h \rangle$, a factor incorporated into the R_{merge} parameter that was used to quantify the spread of individual intensities of all symmetry equivalent reflections within the diffraction data (Sherwood and Cooper, 2011). In this study, the overall R_{merge} value was reported to be 0.131% (Table 5), therefore since the value was < 4-5%, the diffraction data obtained was considered to be of good quality. However, this parameter has its own limitations considering that it does not take into account the multiplicity, or otherwise the number of measurements made per each unique reflection (Wlodawer *et al*., 2008). Instead, the multiplicity-weighted R_{factor}, or R_{meas}, is an improved term of R_{merge} given that it will take into account the number of measurements per a reflection. Consequently, R_{merge} is more advantageous in reflecting the data more accurately when there is high redundancy. In this study, the Rmeas value was reported to be 0.142, therefore similar to Rmerege, seeing that this value was < 4-5%, the spread of the diffraction data was considered to be of good quality (Wlodawer *et al*., 2013).

Subsequent to the scaling and merging process, it was important to determine what proportion of unique reflections had actually been measured. Owing to the nature of Fourier transforms, it was important that all reflections were accounted for so that each value of the electron density map could be correctly calculated (Sherwood and Cooper, 2011; Wlodawer *et al*., 2008). In this study, however, the data set was only 96.8% complete and the remaining 3.2% reflections that were missing could be attributed to radiation damage whereby a significant portion of the spots became unmeasurably weak towards the end of data collection (Table 5). If these missing reflections were from the lowest resolution shells, which are important for phasing procedures and the proper appearance of electron density maps, this would have had a significant impact on phasing and model building (Wlodawer *et al.*, 2008), yet the latter two parameters did not present any caveats in this study and it was therefore suggested that the missing reflections were mainly from the higher resolution shells. Ideally, the dataset used to solve a crystal structure should be 90-95% complete (Wlodawer *et al.*, 2008) therefore the 96.8% completeness in the diffraction data was sufficient to solve the crystal structure of hGSTP1-1 in complex with IAA-94.

An additional parameter that was considered when assessing the confidence in the diffraction data was the resolution, a parameter which is defined as the minimum distance between two

crystal lattice planes that can still provide a measurable diffracted X-ray that will define the level of detail that can be distinguished between structural features in the electron density map (Rhodes, 2006; Wlodawer *et al*., 2008). In this study the lowest resolution shell represented an atomic resolution of 1.40 Å (Table 5), therefore considering that the C-C covalent bond length is 1.5 Å, the carbon atoms from the protein backbone and side chain of hGSTP1-1 could easily be fitted into the electron density map during model building and refinement (Wlodawer *et al*., 2008). During refinement the initial Fourier map was interpreted to derive a model of hGSTP1-1, however, this refined model produced a new and more accurate set of phases. Therefore, by using the newly calculated phases, along with the observed structure factor magnitudes $|F_o|$, a second Fourier synthesis generated a new electron density function with an improved representation for the molecular structure. During this process, to minimise bias in the resultant model, the $2|F_o|-|F_c|$ map was used during model building considering that by multiplying |F| by 2 it bolstered the electron density for missing atoms, whereas subtracting $|F_c|$ reduced the electron density for incorrectly placed atoms in the model (Sherwood and Cooper, 2011).

The relative global discrepancy between $|F_o|$ and $|F_c|$, or otherwise how well the refined structure predicted the observed data, was quantified by means of the residual factor (R_{factor}) parameter to provide insight into model errors. A well-refined protein structure would have an $R_{factor} < 20\%$ (Sherwood and Cooper, 2011; Wlodawer *et al*., 2008), however, in this study the R_{factor} was reported to be 21.66% (Table 5), indicating that the model of hGSTP1-1 was well refined into the electron density. Although informative, seeing that the $|F_o|$ and $|F_c|$ data were used to calculate shifts that were applied to the model, if all the reflections contributed to the calculated shifts then by definition $|F_o| \approx |F_c|$ and the R_{factor} would have decreased during refinement. That is, increasing the number of model parameters during refinement would have over-fitted the diffraction data. Instead, this problem of bias was avoided by omitting 5-10% of the reflection data at the start of refinement. These reflections, which formed the test set, were randomly selected throughout the reciprocal space and the remainder of the data then formed the work-set that underwent refinement as normal (Sherwood and Cooper, 2011; Wlodawer *et al*., 2008).

To then see how well the model predicted the test set that was not used in refinement, the R_{factor} of the test set, known as the free R_{factor} (R_{free}), was monitored to asses if the refinement

was genuinely improving the accuracy of the model or if there was over-fitting of the data (Sherwood and Cooper, 2011). For instance, considering that the solvent is highly dynamic and disordered, the electron density is often noisy and although adding an unreasonable number of water molecules to this region may reduce the R_{factor}, it does not decrease the R_{free}, indicating over-fitting of the experimental data (Wlodawer *et al*., 2008). In this study the R_{free} value for the refined data was 25.11% (Table 5), therefore considering that the difference of 3.45% between the R_{factor} and R_{free} was less than the 7% limit, it was concluded that the data was not over-interpreted and therefore the model of hGSTP1-1 was well refined into the electron density so that $|F_o| \approx |F_c|$. The goodness of fit for the refined model was further validated by looking at the correlation coefficient (cc) between $|F_o|$ and $|F_c|$, that is, when cc = 1 the structure would be deemed to be perfect, however, when cc = 0 then there would be no relationship between $|F_o|$ and $|F_c|$. In this study, when looking at the cc value for the diffraction data set that was randomly split in half ($cc_{1/2}$), it was reported that $cc_{1/2} = 0.996$, therefore indicating that there was a strong correlation between $|F_o|$ and $|F_c|$ and the structure of hGSTP1-1 in complex with IAA-94 was well refined (Sherwood and Cooper, 2011; Karplus and Diederichs, 2012). Furthermore, the RMSD for the bond lengths and angles were reported to be 0.006 Å and 0.934°, respectively (Table 5). Therefore, in view that these values are below the accepted limit of 0.02 Å and 4° for bond lengths and angles, respectively (Wlodawer *et al*., 2008), the refined model of hGSTP1-1 in complex with IAA-94 was in good agreement with the accepted structural geometry for protein molecules.

Structurally, the quality of the refined model of hGSTP1-1 was evaluated by monitoring the torsion angles of the polypeptide backbone conformation. The peptide bond of the backbone has a rigid planar structure, however, the backbone conformation is governed by the torsion angles about the C_α-N (φ) and C_α-C (ψ) bonds between residues (Voet and Voet, 2011). Yet, considering that the bond angles surrounding these torsion angles are typically not restricted during refinement, these angles offered a strong validation tool when assessing the stereochemical parameters of the refined structure of hGSTP1-1 (Wlodawer *et al*., 2008). The sterically allowed values of φ and ψ are determined by calculating the distance between the atoms of a tripeptide at all values of φ and ψ from the central peptide carbon. However, when the non-bonding interatomic distances are considered to be less than their van der Waals distance, the torsion angles are then considered to be in a sterically forbidden conformation (Voet and Voet, 2011). Upon examining the Ramachandran diagram (Ramachandran and

Sasisekharan, 1968) for the refined structure of hGSTP1-1 it was noted that only 0.24% of the amino acids were found in the disallowed or unfavourable regions on the Ramachandran plot and therefore there were no outliers. Instead it was noted that 97.70% of the residues were found in the favourable region and therefore they experienced no steric clashes, however, in the case where the van der Waals radii of atoms were slightly decreased so that they were allowed to come a little closer (Voet and Voet, 2011) it was observed that only 2.06% of the residues were found in this generously allowed region (Table 5; Figure S8; Table S6). As a result, considering the majority of the torsion angles of the polypeptide backbone were in their allowed and favourable stereochemical conformation, the final refined model of hGSTP1-1 in complex with IAA-94 was of good quality.

Table 5: Summary of the diffraction data collected and the statistics for the refinement of hGSTP1-1 in complex with IAA-94

Data collection	
Light source	I03 beamline
Detector	Dectris Eiger2 XE 16M
Wavelength (Å)	0.915 Å
Beam size (µm)	90 x 20 (at sample) and 90 x 30 (at detector)
Temperature (K)	100
Resolution (Å)	68.638 (1.402)
Space group	$P12_11$
Unit cell dimensions: Length (Å)	*a* = 63.64, b = 72.18, c =88.62
Angle (°)	α = γ = 90°, β = 90.14°
Asymmetric unit content	4
Number of non-hydrogen atoms: Total	7649
Protein	6621
Solvent	841
GSH	80
MES	72
PEG	20
O7Z	15
Matthew's coefficient ($Å^3.Da^{-1}$)	2.33
Solvent content (%)	47.1
Number of observations	1 023 676 (36 264)
Number of unique reflections	163 749 (7 637)
Completeness (%)	96.80 (90.9)
Multiplicity	6.3 (4.7)
Rmerge (%)	0.131 (2.251)
Rmeas (%)	0.142 (2.539)
CC1/2	0.996 (0.327)
I/σ (I)	7.1 (0.6)
Refinement	
Final overall R-factor (%)	21.83
Rwork (%)	21.66
Rfree (%)	25.11
B-factor ($Å^2$): Mean B-factor (total)	22.06
Mean B-factor (protein)	20.60
Mean B-factor (solvent)	32.18
RMSD bond lengths (Å)	0.006
RMSD bond angles (°)	0.934
Ramachandran analysis: Favoured (%)	97.70
Allowed (%)	2.06
Outliers (%)	0.24
All-atom clash score	4.34

[a] Rmerge = ΣhklΣi|Ii-<I>|/|<I>|, where Ii is the intensity for the ith measurement of an equivalent reflection with indices h, k and l.

[b] Rwork = Σ||Fobs|-|Fcalc||/Σ|Fobs|, where Fobs and Fcalc are the observed and calculated structure factor amplitudes, respectively.

[c] Rfree was calculated with 5% of the diffraction data that were selected randomly and not used through refinement.

[d] The values in parenthesis are from the highest resolution (outer) shell.

3.4.2.3 Interpreting the crystal structure of hGSTP1-1 in complex with IAA-94

3.4.2.3.1 The organic molecules which were co-crystallised with hGSTP1-1

In this study the electron density for the refined crystal structure of hGSTP1-1 indicated the presence of four monomeric chains (A-D), but in particular chain-A and chain-B comprised the dimeric biological unit of hGSTP1-1, whereas chain-C and chain-D represented an individual monomeric subunit of hGSTP1-1 given that the adjacent symmetrical subunits were present in the neighbouring unit cell. Although there was no distinct electron density associated with hGSTP1-1 that resembled the molecular geometry of IAA-94, bound to each chain at the G-site and buffer-binding site was a GSH and MES molecule, respectively, yet, bound only to chain-C and chain-D was a PEG molecule at the dynamic α-helix 2. Furthermore, bound to hGSTP1-1 were three acetate molecules, that is, the one acetate molecule was bound to chain-B, while the other two acetate molecules were located on chain-D (Figure 60).

The interaction of GSH with chain-A and chain-B was demonstrated to be coordinated by hydrogen bonds formed with Arg13, Trp38, Lys44, Gln51, Leu52, Gln64 and Ser65 (Figure 61). Additionally, when GSH was bound to chain-A or chain-B, a hydrogen bond was formed with Asp98 from either chain-B or chain-A, respectively. This interaction with GSH was the same in both chain-C and chain-D (Figure 62) and therefore the overall conformation of GSH in all four chains was noted to be the same. Collectively, these interactions correlated with previous crystal structures of hGSTP1-1 in complex with GSH (PDB: 2GSS), therefore further supporting that the G-site of hGSTP1-1 was folded in its native conformation.

Previously, when hGSTP1-1 had been crystallised in the presence of MES it was noted that MES bound to each subunit by forming a hydrogen bond with Glu197 at the buffer-binding site, or otherwise the L-site as suggested by Ji *et al*. (1997) and Prade *et al*. (1997). In this study, however, the dihedral angel of the nitrogen atom from the 2-(N-morpholino) moiety of MES was 114.4° when bound to chain-A as opposed to 109.4° when bound to chain-B, chain-C and chain-D and consequently no hydrogen bond was predicted to form between MES and Glu197 in chain-A. Furthermore, a third MES molecule was noted to form a hydrogen bond with Arg74 on the outer surface of the dimer interface and hence a stoichiometry of three was reported for the binding of MES to hGSTP1-1. Interestingly, the binding of MES at this site has only been observed in some crystal structures of hGSTP1-1 and the overall significance of this sight still remains unknown (Table 6), yet given that it is not present within all the crystal structures of hGSTP1-1, it can be inferred that this site is not the leading MES-binding site.

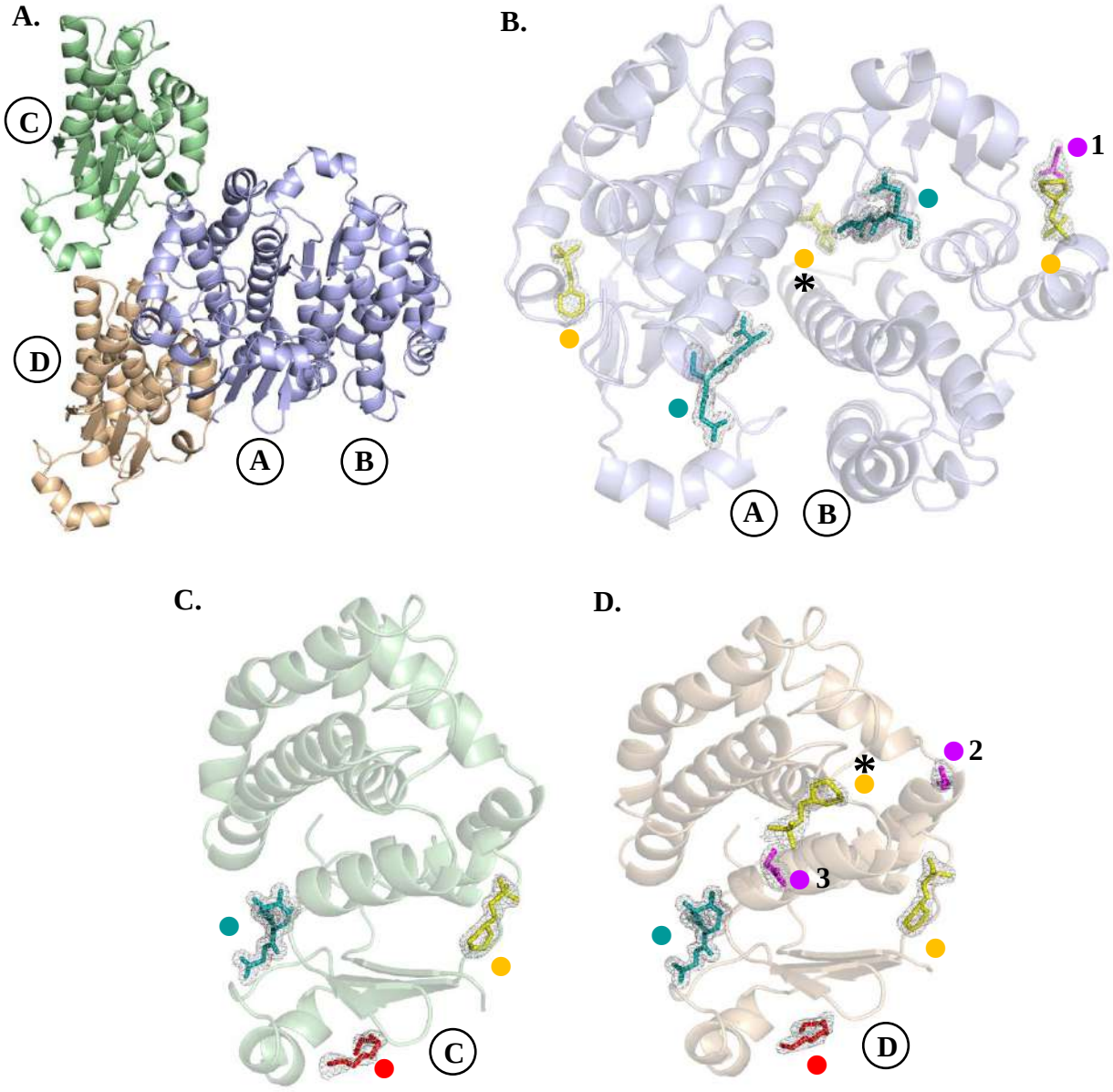


Figure 60. The asymmetric unit of hGSTP1-1 and its associated ligands. (A) The asymmetric unit copy number was reported to n = 4, therefore indicating that per a unit cell there were four monomeric chains (A-D) of hGSTP1-1 that were symmetrically related by the $P12_11$ space group. (B) chain-A and chain-B (●) can be seen to represent the complete dimer and therefore the biological unit of hGSTP1-1. Bound to both chains was GSH (●) and MES (●) at the G-site and buffer-binding site, respectively. However, a third MES molecule (*) was bound to the outer surface of the dimer interface, while an acetate molecule (●1) was bound to only chain-B near the buffer-binding site. (C) chain-C (●) and (D) chain-D (●) represent an individual monomeric subunit of hGSTP1-1 considering that the adjacent symmetrical subunit was present in the neighbouring unit cell. Bound to both chains was GSH, MES and PEG (●) at the G-site, buffer-binding site and α-helix 2, respectively. However, bound to chain-D was a third MES molecule (*) that could be seen to bind at the outer surface of the dimer interface. Furthermore, an acetate molecule was found to be bound at the outer surface of the dimer interface adjacent to MES (●2), in addition to the crystal-packing interface formed between chain-A and chain-D (●3). All images were generated by using Pymol V1.3.

Table 6: Summary of the organic molecules reported to be in complex with hGSTP1-1

PDB code	Resolution (Å)	Space group	Organic ligand molecules bound to hGSTP1-1		
			Type of organic molecule type	Stoichiometry (per dimer)	Chain location
6Y1E	1.4	$P12_11$	GSH	2	Chains A, B, C and D
			MES	3*	Chains A, B, C and D
			PEG	2	Chain C and D
			O7Z	3*	Chain B and C
1AQW	1.8	$P12_11$	GSH	2	Chains A, B, C and D
			MES	2	Chains A, B, C and D
1AQX	2	$P12_11$	MES	2	Chains A, B, C and D
			GTD	2	Chains A, B, C and D
			GSH	2	Chains A, B, C and D
3CSI	1.9	$P12_11$	MES	2 & 1	Chain A, B and D
			LZ6	2	Chains A, B, C and D
3HJM	2.1	$P12_11$	MES	2 & 1	Chains A, B and D
3CSH	1.9	C121	MES	2	Chains A and B
			CBL	1	Chains B
2GSS	1.9	C121	MES	3*	Chains A and B
			EAA	2	Chains A and B

*: Represents organic molecules that bind to both individual subunits of hGSTP1-1 as well as to the outer surface of the dimer interface.

O7Z: Represents the acetate moiety of IAA-94

GTD: Represents S-(p-bromobenzyl) glutathione

LZ6: Represents chlorambucil-GSH conjugate

CBL: Represents chlorambucil

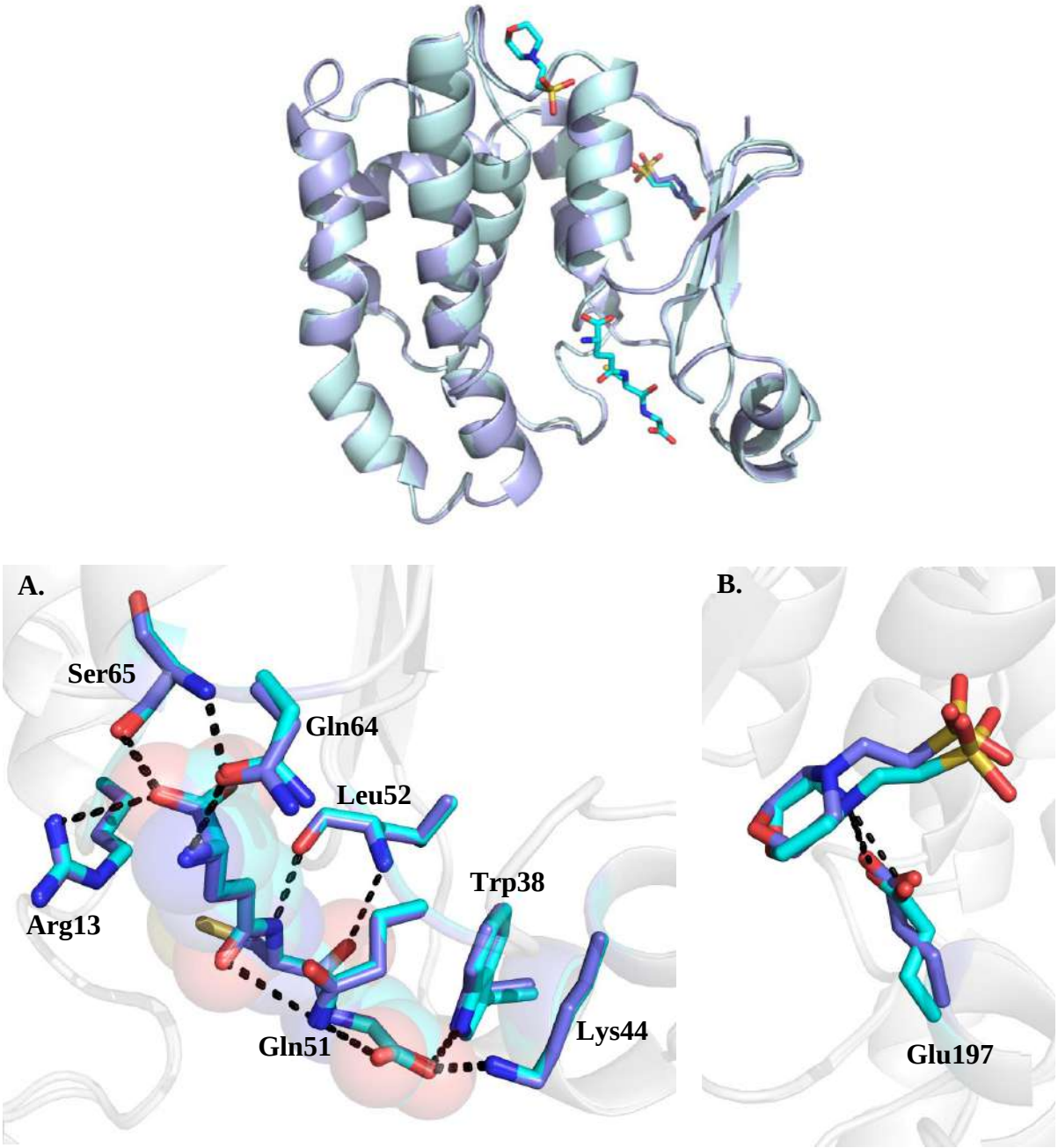


Figure 61. The binding of GSH and MES to chain-A and chain-B of hGSTP1-1. hGSTP1-1 is a homodimeric protein and as would be expected the structural alignment of chain-A with chain-B indicated that the peptide backbone for both chains were structurally identical (RMSD = 0.194 Å). (A) The interaction of GSH with either chain-A (●) or chain-B (●) was coordinated by GSH forming a hydrogen bond with Arg13, Trp38, Lys44, Gln51, Ley52, Gln64 and Ser65. This corresponded with the interactions previously observed in the crystal structure of hGSTP1-1 in complex with GSH, therefore supporting that the G-site of hGSTP1-1 was folded in its native conformation. (B) At the buffer-binding site, the interaction of MES with hGSTP1-1 was coordinated by a hydrogen being formed between MES and Glu197. However, the dihedral angel of the nitrogen atom from the 2-(N-morpholino) moiety of MES was 114.4° when bound to chain-A as opposed to 109.4° when bound to chain-B and consequently no hydrogen bond was predicted to have formed between MES and Glu197 in chain-A. All images were generated by using Pymol V1.3.

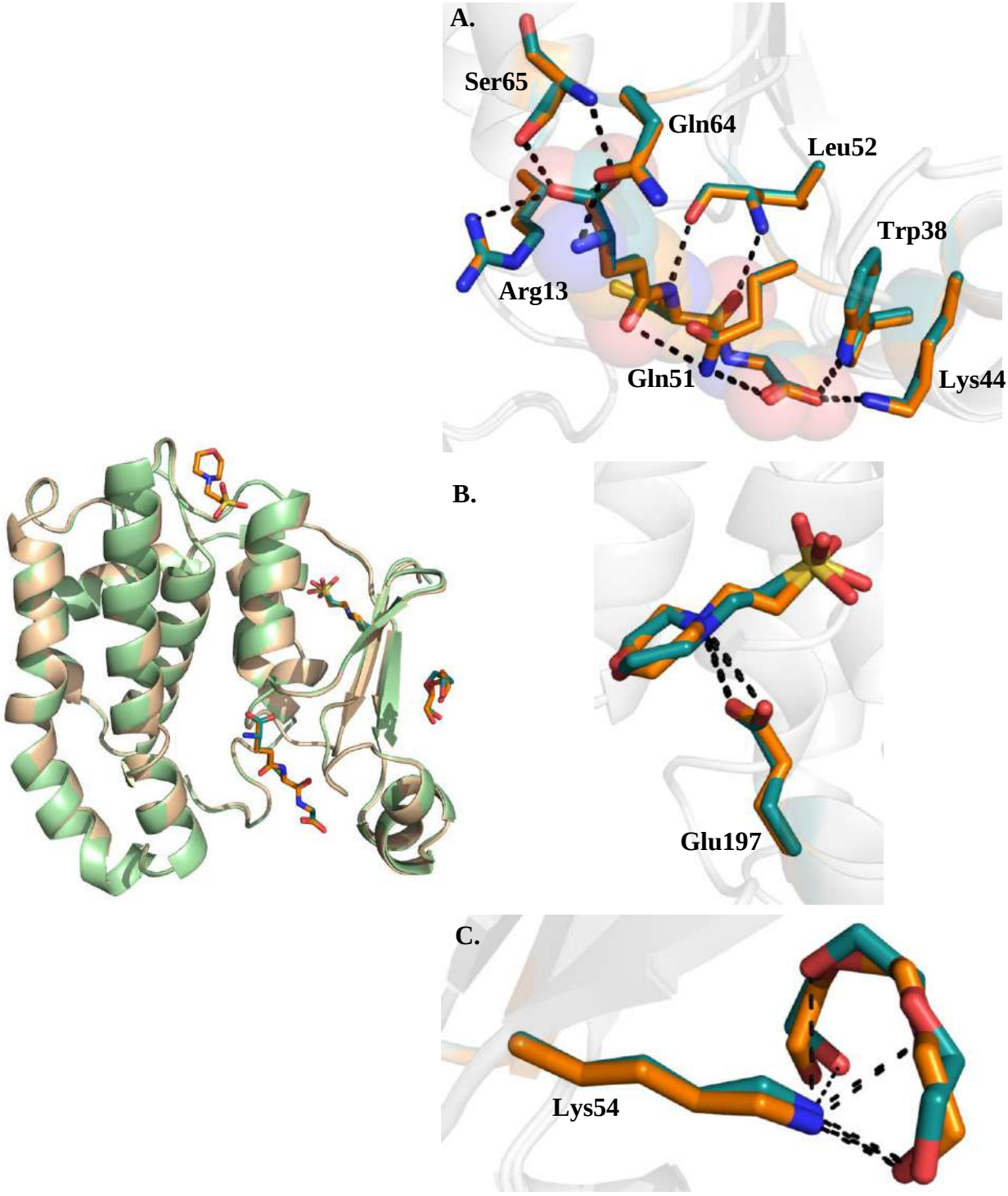


Figure 62. The binding of GSH, MES and PEG to chain-C and chain-D of hGSTP1-1. The structural alignment of chain-C with chain-D indicated that the peptide backbone for both chains were structurally identical (RMSD = 0.131 Å). (A) The interaction of GSH with either chain-C (●) or chain-D (●) was coordinated by GSH forming a hydrogen bond with Arg13, Trp38, Lys44, Gln51, Ley52, Gln64 and Ser65. This corresponded with the interactions previously observed in the crystal structure of hGSTP1-1 in complex with GSH, therefore supporting that the G-site of hGSTP1-1 was folded in its native conformation. (B) At the buffer-binding site, the interaction of MES with hGSTP1-1 was coordinated by a hydrogen bond being formed between MES and Glu197. (C) The interaction of PEG with both chains was coordinated by PEG forming a hydrogen bond with Lys54 at α–helix2. All images were generated by using Pymol V1.3.

3.4.2.3.2 The interaction between the acetate moiety of IAA-94 and hGSTP1-1

The three acetate molecules that were present in the crystal structure of hGSTP1-1 were of particular interest, specifically because the buffers that were used for the purification of hGSTP1-1 and for the conditions of the crystals that were used for diffraction did not contain any acetate-based salts. In particular, the acetate molecule that was associated with chain-B was found to bind near the MES-binding site, therefore bearing in mind that the MES-binding site was suggested to be the putative binding site for IAA-94, it was proposed that this acetate molecule could in fact be the acetate moiety of IAA-94. In contrast, although the second acetate molecule was in close proximity to the putative IAA-94-binding site, its interaction with hGSTP1-1 instead was found to be mediated at a crystal-packing interface formed between chain-A and chain-D. Whereas, the third acetate molecule that was bound to chain-D was not located near the putative IAA-94-binding site, but instead was found bound at the same site as the third MES molecule on the outer surface of the dimer interface (Figure 63).

The acetate molecule bound to chain-B was noted to interact with hGSTP1-1 through a hydrogen bond being formed between its oxygen anion and the backbone of Trp28 (Figure 64). Interestingly, considering that this acetate molecule was proposed to be the acetate moiety of IAA-94, this correlated with the findings reported by IFMD whereby Trp28 and Ser27 formed a hydrogen bond with the acetate moiety of IAA-94 in order to coordinate ligand binding (Figure 54). However, because the crystallisation of hGSTP1-1 with IAA-94 was in the presence of MES, when IAA-94 was docked to hGSTP1-1 which had previously been docked with MES, it was noted that the acetate moiety of IAA-94 only formed a hydrogen bond with Ser27 and not Trp28 (Figure 56). This discrepancy in the mode of binding could be described by the Debye-Waller factor (DWF), or otherwise the B-factor, which in crystallography is used to describe the oscillation amplitudes of the atoms around their equilibrium positions and is therefore able to provide insight into atomic disorder. Therefore, the more dynamic an atom becomes the larger the displacement will be and consequently this flexibility eventually lowers the electron density to bring about uncertainty in the atomic positions (Liu *et al.,* 2014). The B-factor of the atoms for the acetate moiety of IAA-94 bound to chain-B were noted to have a magnitude of 27.30-47.49 $Å^2$, whereas the B-factor for the atoms from Ser27 and Trp28 were 19.54-37.45 $Å^2$ and 16.02-22.99 $Å^2$, respectively. Therefore, due to the atoms from the acetate moiety of IAA-94 and Ser27 being more dynamic, this could coincide as to why no hydrogen bonds were predicted between these two molecules in the crystal structure.

On the other hand, at the crystal packing interface formed between chain-A and chain-D, the second acetate moiety of IAA-94 was noted to interact with hGSTP1-1 through a hydrogen bond being formed between its carbonyl functional group and the side chain of Arg186 from chain-A. In addition, the oxygen atom that would be flanked by the acetate and indanyl moiety of IAA-94 formed a hydrogen bond with the side chains of Arg186 and Lys188 as well as the backbone of Asp23 from chain-D. This in fact also correlated with the IFMD data considering that some of the poses predicted that the Arg188 could form a salt bridge and/or hydrogen bond the with the acetate moiety of IAA-94. However, with respect to the third acetate moiety of IAA-94 which bound to the outer surface of the dimer interface, it did not form a direct interaction with any residues from hGSTP1-1 and instead it only formed a hydrogen bond with the sulfonate group of MES (Figure 64). Therefore, as was the case for MES, the importance of this site in the binding of IAA-94 remains unknown, yet the displacement of IAA-94 to this site could be attributed to an anomaly caused by crystal packing.

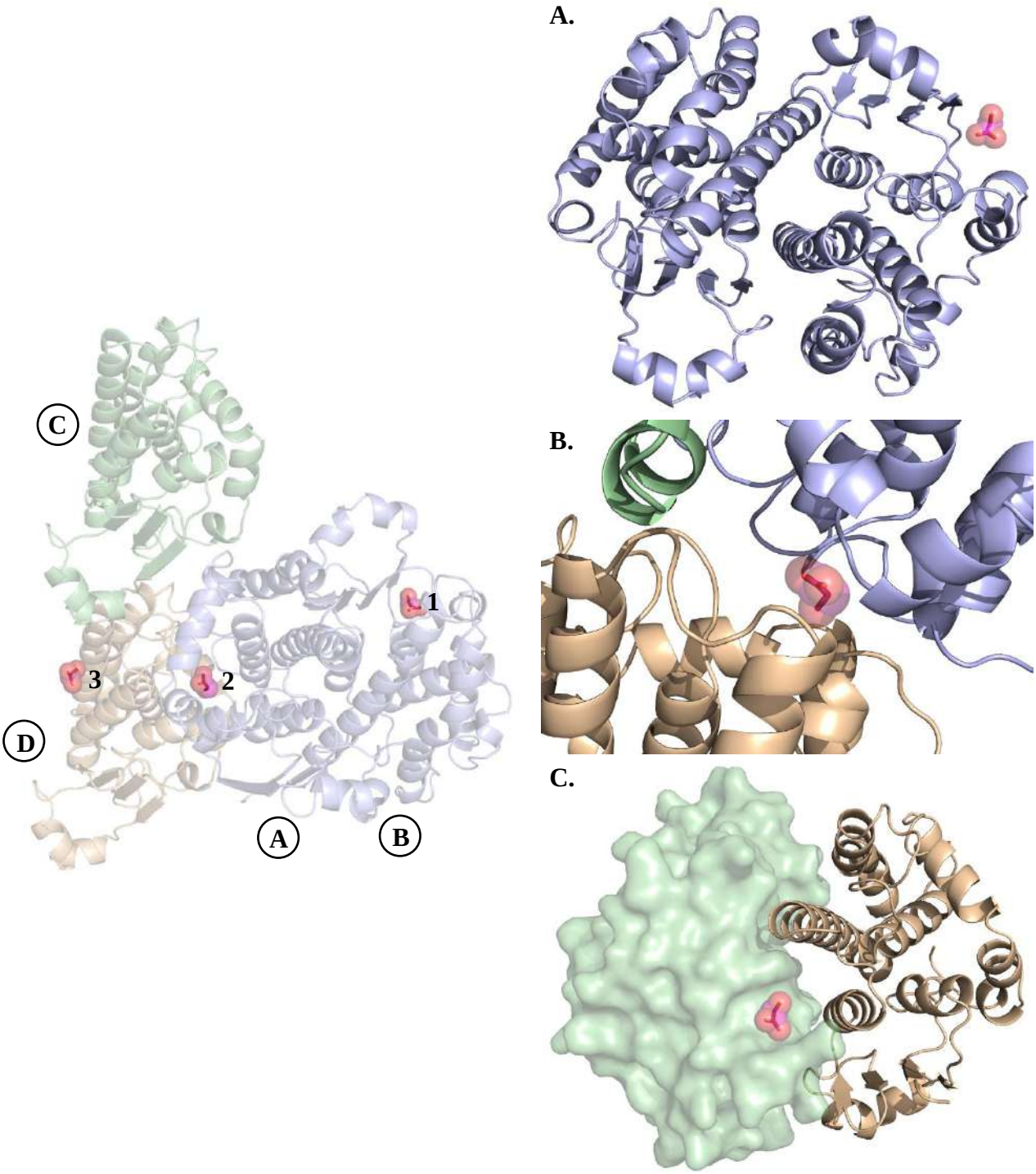


Figure 63. The binding of the acetate moiety of IAA-94 to chain-B and chain-D of hGSTP1-1. A total of three acetate molecules were present per a unit cell, whereby the one acetate molecule bound to chain-B (●), while the other two acetate molecules were bound to chain-D (●). (A) The acetate molecule associated with the B-chain was found to bind near the MES-binding site, therefore bearing in mind that the MES-binding site was suggested to be the putative binding site for IAA-94, it was proposed that this acetate molecule could in fact be the acetate moiety of IAA-94. (B) The second acetate molecule was found to be mediated at a crystal-packing interface formed between chain-A and chain-D. (C) The third acetate molecule that was bound to chain-D was found bound at the same site as the third MES molecule on the outer surface of the dimer interface. All images were generated by using Pymol V1.3.

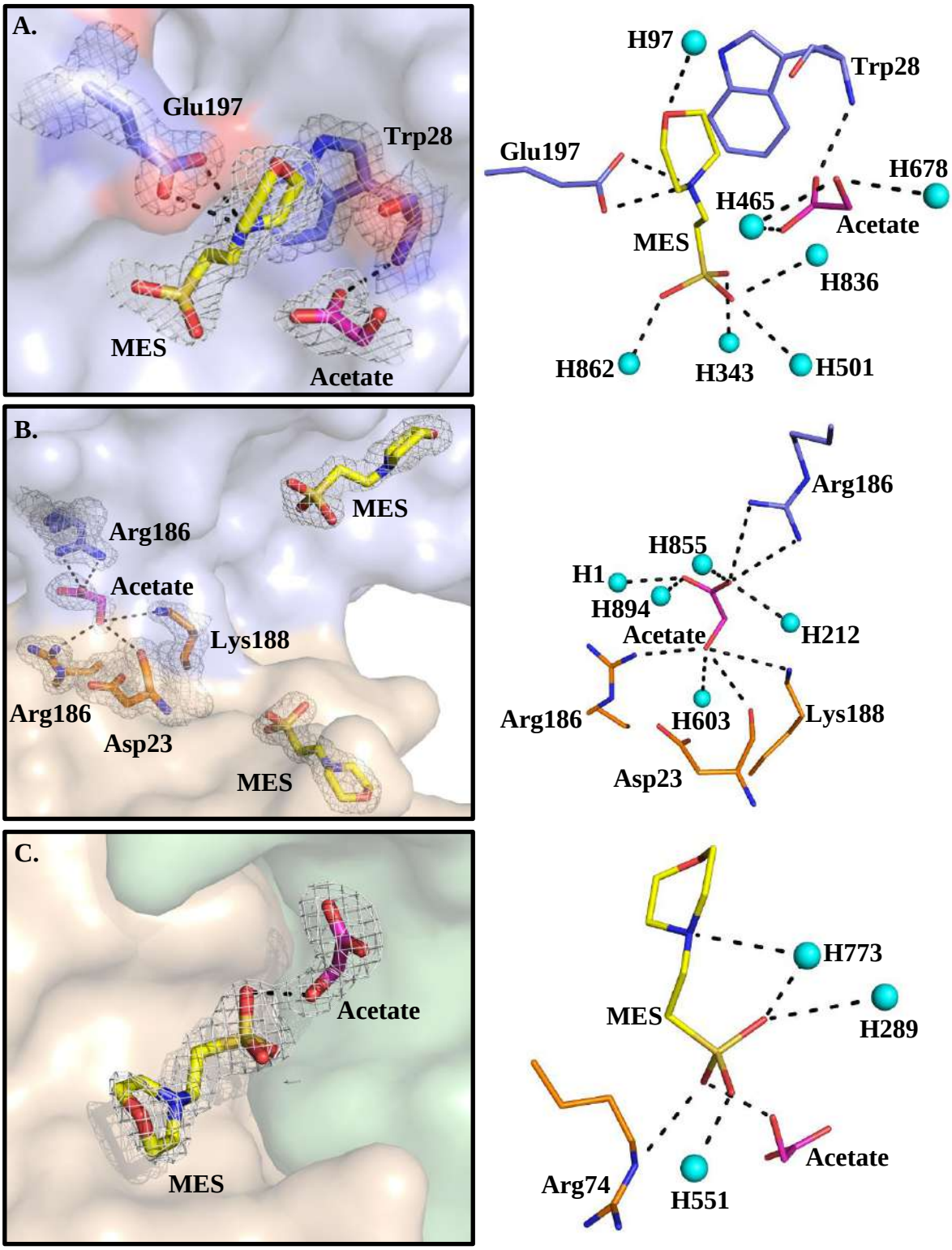


Figure 64. The interactions formed between the acetate moiety of IAA-94 and hGSTP1-1. (A) The acetate moiety of IAA-94 which bound to chain-B (●) was noted to interact with hGSTP1-1 through a hydrogen bond being formed between its oxygen anion and the backbone of Trp28. (B) The second acetate moiety of IAA-94 formed a hydrogen bond with side chain of Arg186 from chain-A (●), in addition to forming a hydrogen bond with the side chains of Arg186 and Lys188 along with the backbone of Asp23 from chain-D (●). (C) The third acetate moiety of IAA-94 which bound to the outer surface of the dimer interface only formed a hydrogen bond with the sulfonate group of MES. The residues within the left panel have been placed in a sigma-weighted 2Fo - Fc electron density map (σ = 1), therefore indicating how the refined structures fitted into their electron density map. All images were generated by using Pymol V1.3.

3.4.2.3.3 Crystal contacts interfered with the binding of IAA-94 to hGSTP1-1

The contacts made during crystal packing have been shown to have a similar, or even larger, interface area compared to the interfaces formed during protein-protein and protein-ligand interactions. Therefore, when these crystal contacts are around a ligand binding site, or regions expected to undergo a conformational change upon ligand binding, the presence of a ligand molecule could prevent the formation of protein crystals or alternatively, depending on their affinity, the crystal contacts could displace the ligand molecule (Luo *et al*., 2015). Interestingly, the MES-binding site was in close proximity to where crystal contacts occurred and therefore these contacts could have had an influence on the binding of IAA-94 to hGSTP1-1. When evaluating the residues that were involved in the crystal contacts near the MES-binding site, together with the residues that were within 8 Å of the MES-binding site it was noted that Ser27, Lys29 and Glu30 from the MES-binding site were directly involved in crystal (Figure 65). Furthermore, when evaluating the residues involved in the crystal contacts between chain-A and chain-D, it was noted that Tyr3, Gln24, Gly25, Gln26, Ser27 and Glu197 from the MES-binding site were directly involved in crystal contacts (Figure 66).

Therefore, considering that Ser27 was suggested to play a key role in the interaction of IAA-94 with hGSTP1-1 (Figure 54 and 56), when this residue became involved in the crystal contacts between chain-C and chain-D (Figure 65) as well as at the crystal packing interface between chain-A and chain-D (Figure 66) it would have impacted the overall binding of IAA-94. However, apart from the residues that are directly involved in crystal contacts, these contacts would also influence the van der Waal contacts made by their neighbouring residues. For instance, this could be noted whereby both Lys29 and Glu30 from chain-A and chain-D were directly involved in crystal contacts with chain-C and chain-B, respectively (Figure 65). Therefore, the residues surrounding these crystal contact sites would have induced some conformational change and/or steric hindrance around Trp28 and Ser27 that would have interrupted the binding of IAA-94 to hGSTP1-1. However, with respect to chain-B, only Lys29 was directly involved in crystal contacts (Figure 65), therefore the apparent decrease in steric hindrance at this site, combined with the fact that Ser27 had a high B-factor, could account for why the acetate moiety of IAA-94 was able to form a hydrogen bond with Trp28 (Figure 64). Furthermore, considering the importance of Phe192 in forming a π-π stacking interaction with IAA-94 (Figure 54), it has become clear how the likes of Pro196 and Glu197, which are directly involved in crystal contacts, could influence the conformation of Phe192 to hinder the π-π stacking between IAA-94 and hGSTP1-1 .

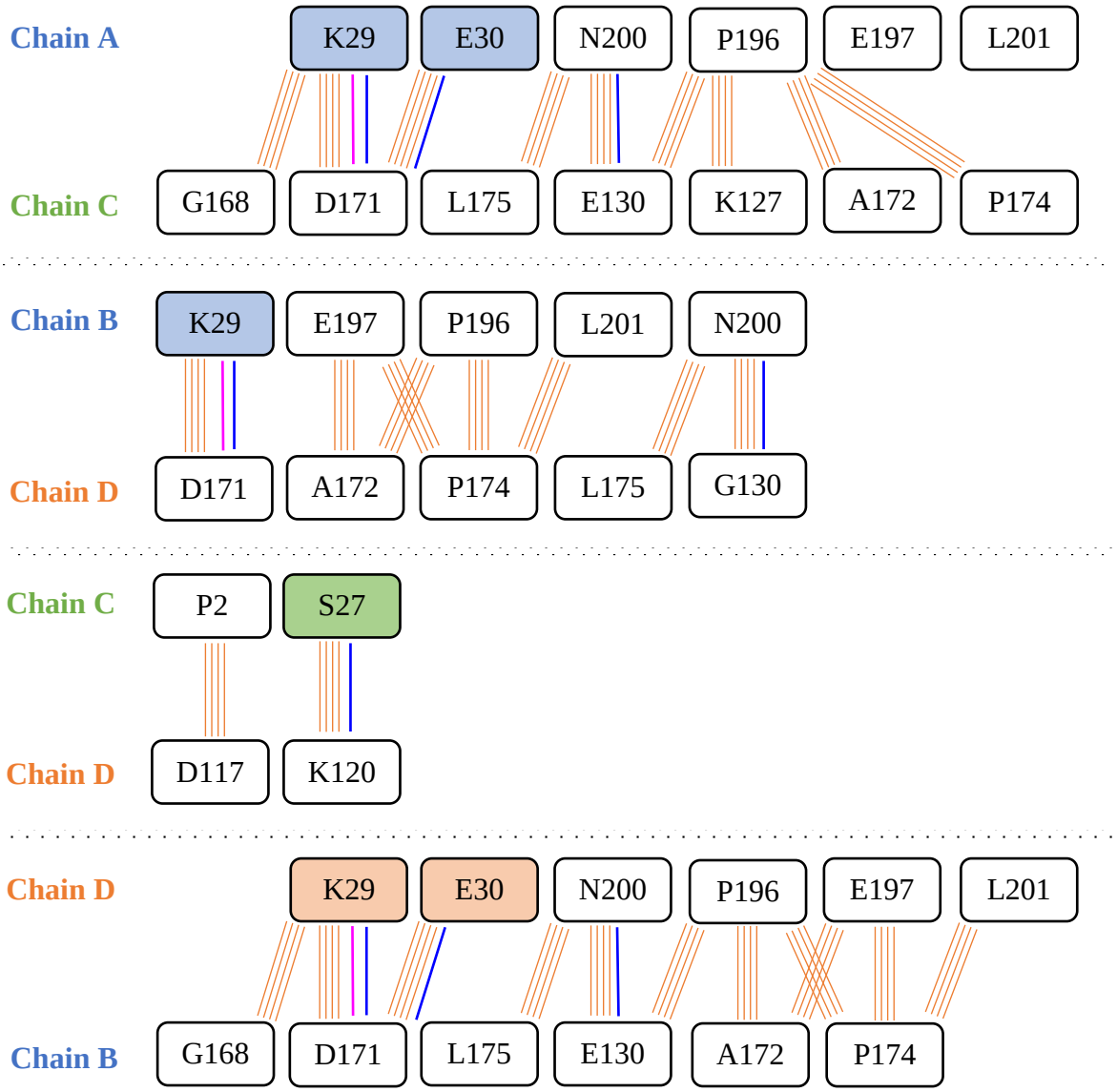


Figure 65. The crystal contacts made at the MES (putative IAA-94)-binding site of hGSTP1-1. The MES-binding site was in close proximity to where crystal contacts occurred and therefore it was suggested that these contacts could have an influence on the binding of IAA-94 to hGSTP1-1. The neighbouring unit cells and corresponding asymmetrical units were generated in Pymol and those chains which were not involved in crystal contacts at the MES-binding site for a particular chain were deleted. All crystal contacts were then generated by using the protein-protein interaction tool in PDBsum (Laskowski, 2001) (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html). Upon evaluating the residues involved in the crystal contacts near the MES-binding site, together with the residues that were within 8 Å of the MES-binding site, it was noted that Ser27, Lys29 and Glu30 from the MES-binding site were directly involved in crystal contacts. Indicated are the van der Waal interactions (—), hydrogen bonds (—) and salt bridges (—) involved in the crystal packing contacts.

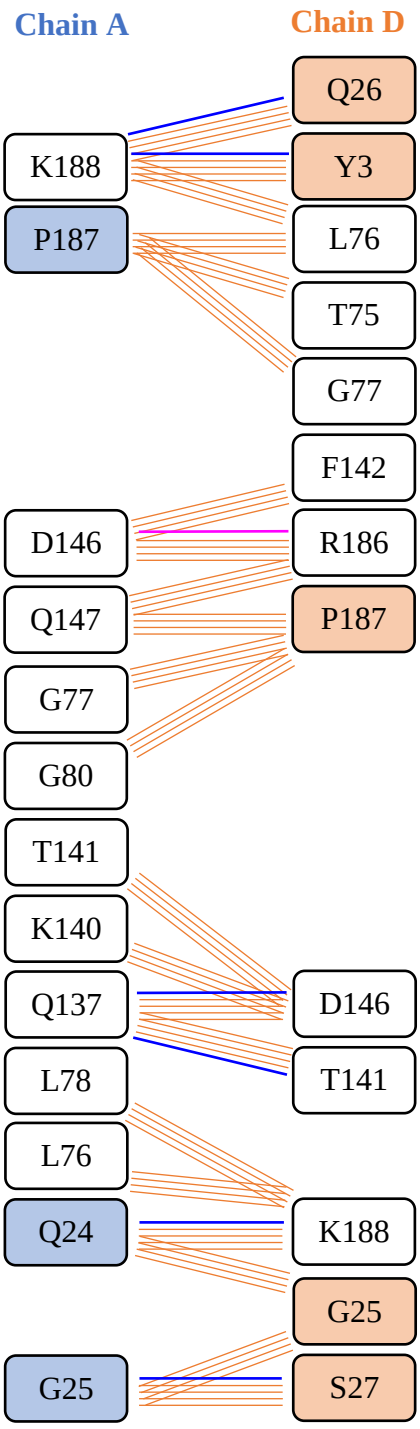


Figure 66. The crystal contacts made at the crystal-packing interface formed between chain-A and chain-D. The second acetate moiety of IAA-94 was found to be mediated at a crystal-packing interface formed between chain-A and chain-D and it was suggested that these contacts could have an influence on the binding of IAA-94 to hGSTP1-1. All crystal contacts were then generated by using the protein-protein interaction tool in PDBsum (Laskowski, 2001) (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html). Upon evaluating the residues involved in the crystal contacts at this site, together with the residues that were within 8 Å of the MES-binding site, it was noted that Tyr3, Gln24, Gly25, Gln26, Ser27 and Glu197 from the MES-binding site were directly involved in crystal-packing contacts. Indicated are the van der Waal interactions (—), hydrogen bonds (—) and salt bridges (—) involved in the crystal packing contacts.

However, considering that MES was bound at the putative IAA-94-binding site, the crystal contacts could have also had a direct influence on the binding of MES, which in turn could have influenced the binding of IAA-94. Although the 2-(N-morpholino) moiety for the docked and crystallised structure of MES shared a similar location, the orientation of the sulfonate group was different for the docked structure of MES considering that the sulfonate group formed a hydrogen bond with Glu30 (Figure 67). This was interesting to observe considering that Glu30 was noted to be directly involved in crystal contacts by chain-A and chain-D (Figure 65). Therefore, upon crystallisation, the interaction between Glu30 and the sulfonate group of MES would have been disrupted to result in MES changing its orientation to what was observed in the crystal structure of hGSTP1-1 where the sulfonate group of MES was no longer coordinated by Glu30 (Figure 67 and 68), but instead it was coordinated by water molecules (Figure 64). However, in the case of chain-B, although Lys29 was only involved in crystal contacts (Figure 65), it would have indirectly influenced the binding of MES by influencing the Van der Waal contacts of its neighbouring residues through some conformational change and/or steric hindrance effect.

In fact, when looking at the docked structure of IAA-94 in relation to the crystal structure of MES, it was clear that during crystal-packing the change in the orientation of MES will result in the sulfonate group of MES sterically interfering with the indanyl moiety of IAA-94 (Figure 67 and 68). Therefore, apart from Pro196 and Glu197 being involved in crystal contacts and having an influence on the π-π stacking interaction between Phe192 and the indanyl group of IAA-94 (Figure 54 and 65), it was evident that given the sulfonate group of MES became more dynamic, it would have had an influence on the π-π stacking interaction that was required to coordinate the binding of IAA-94 to hGSTP1-1. To further provide insight into this phenomenon, when IAA-94 was docked with MES from the crystal structure of hGSTP1-1, it resulted in IAA-94 being displaced to a site where the indanyl group of IAA-94 was instead in contact with the surface of Ser27 (Figure 67). This conformation of IAA-94 would still not have been favourable seeing that Ser27 and its neighbouring residues are involved in crystal contacts (Figure 65) and therefore this would have displaced IAA-94 from this site during crystal packing (Figure 67).

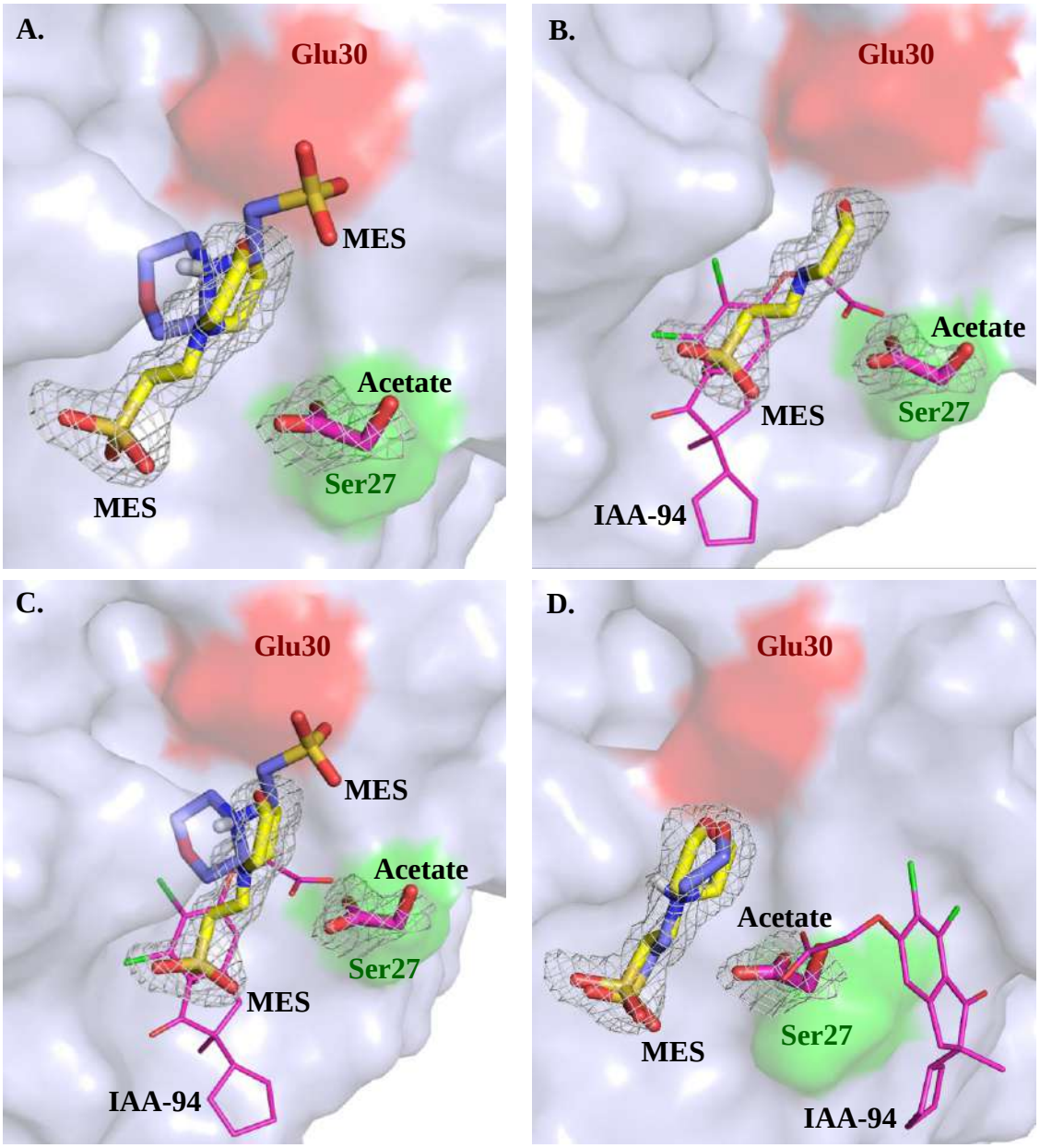


Figure 67. The displacement of IAA-94 by MES due to crystal contacts. (A) The IFMD of MES to hGSTP1-1 indicated that Glu30 was a key residue involved in coordinating the sulfonate group of MES. However, Glu30 and its neighbouring residue (Lys29) are involved in crystal contacts, therefore upon crystallisation, MES would change its orientation to what was observed in the crystal structure of hGSTP1-1 where the sulfonate group of MES is no longer coordinated by Glu30. The IFMD of (B) IAA-94 or (C) IAA-94 in complex with MES which had previously been docked to hGSTP1-1, indicated that when MES changes it orientation during crystallisation, it would result in the sulfonate group of MES sterically interfering with the indanyl moiety of IAA-94. (D) IFMD of IAA-94 with the crystal structure of MES resulted in IAA-94 being displaced to a site where the indanyl group of IAA-94 was instead in contact with the surface of Ser27. This conformation of IAA-94 would still not have been favourable seeing that Ser27 and its neighbouring residues are also involved in crystal contacts. Indicated is Ser27 (●), Glu30 (●), the docked structure of MES (●), the docked structure of IAA-94 (●) as well as the crystal structure of MES (●) and the acetate moiety of IAA-94 (●), both of which have been placed in a sigma-weighted $2F_o$ - F_c electron density map (σ = 1), therefore indicating how the refined structures fitted into the electron density map. All images were generated by using Pymol V1.3.

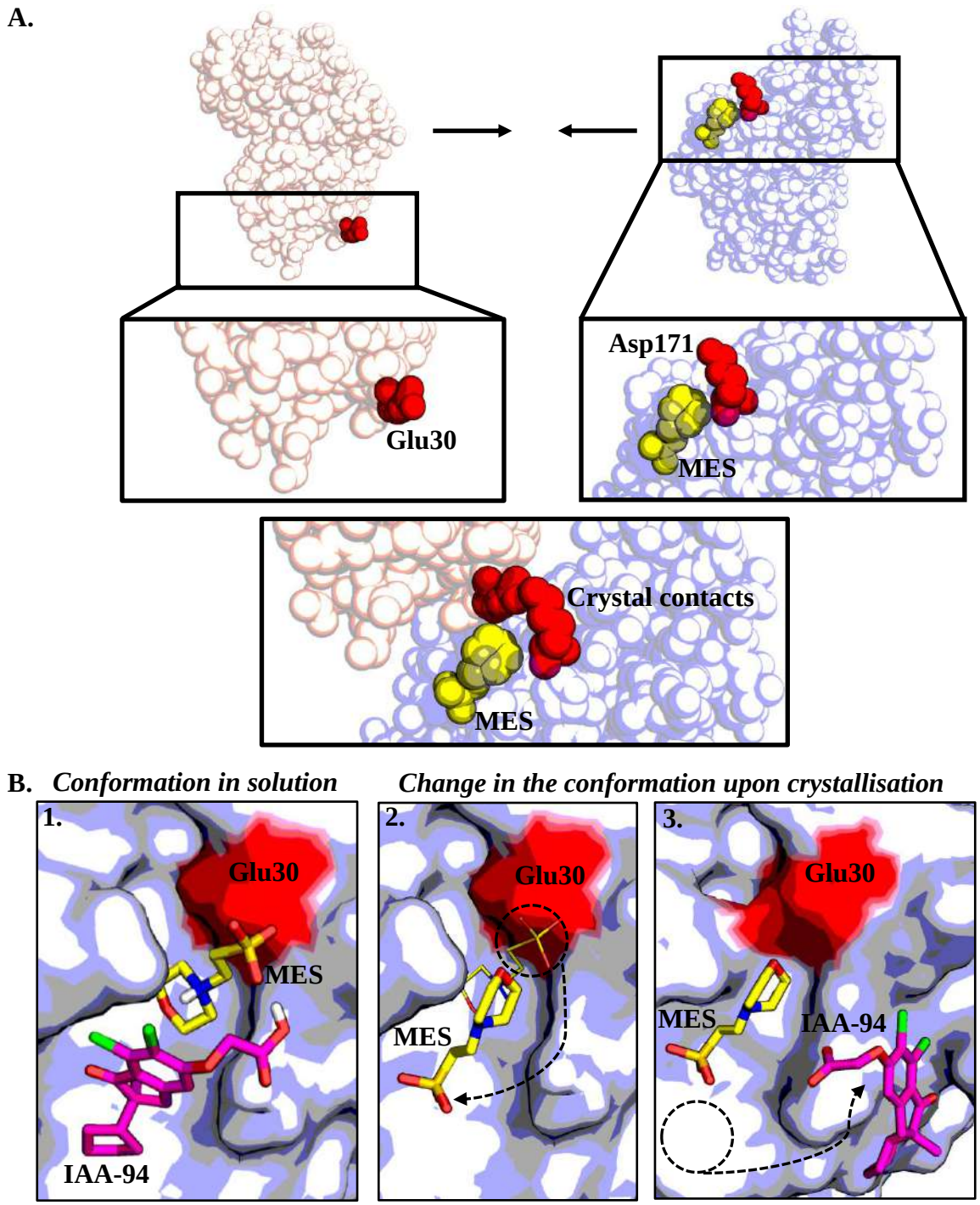


Figure 68. A schematic representing the change in the orientation of MES due to crystal contacts. (A) Upon crystallisation, Glu30 (●) and its neighbouring residue (Lys29) were involved in crystal contacts with Asp171. (B1-2) The interaction between Glu30 and the sulfonate group of MES would have therefore become disrupted to result in MES changing its orientation (⇢) to what was observed in the crystal structure of hGSTP1-1, where the sulfonate group of MES was no longer coordinated by Glu30, but instead it was coordinated by water molecules. In the case of chain-B, although Lys29 was involved in crystal contacts as opposed to Glu30, it would have still indirectly influenced the binding of MES through some conformational change and/or steric hindrance effect. (B3) During crystal-packing, the change in the orientation of MES would have resulted in its sulfonate group becoming more dynamic. Consequently, it would have sterically interfered with the π-π stacking interaction formed between IAA-94 and hGSTP1-1 and as a result IAA-94 (⇢) was displaced from its binding site. All images were generated by using Pymol V1.3.

Although the acetate molecules identified in the crystal structure of hGSTP1-1 were ascribed to being the acetate moiety of IAA-94, it was important to search the PDB to identify if there were any crystal structures of hGSTP1-1 that had previously been solved in complex with an acetate molecule and to then identify were these acetate molecules had bound. In fact, currently there are two available crystal structures of hGSTP1-1 in the presence of acetate (PDB ID: 6AP9 and 5JCW) (Table 7) and these molecules were noted to bind to hGSTP1-1 by forming a hydrogen bond only with the side chain of Arg186 (Figure 69). Interestingly, this correlated with the findings reported in this study whereby the second acetate moiety of IAA-94 also formed a hydrogen bond with the side chain of Arg186 at the crystal-packing interface formed between chain-A and chain-D (Figure 63 and 64). Therefore, bearing in mind the affect crystal contacts would have had on the binding of IAA-94, it was suggested that upon crystallisation IAA-94 would have been displaced. Yet, owing to the affinity that Arg186 has for an acetate molecule, the displaced IAA-94 molecule from either chain-A or chain-D would have been shifted to the crystal-packing interface where Arg186 would have coordinated the binding of its acetate moiety. However, as previously mentioned (section 3.4.2.3.2), the interaction of the third acetate moiety of IAA-94 with MES at the outer surface of the dimer interface could have been attributed as an anomaly caused by crystal packing. In fact, it is now clear that this indeed could have been the case considering that between the third acetate moiety of IAA-94 and the crystal contact sites are solvent channels that would have allowed the second displaced IAA-94 molecule from either chain-A or chain-D to diffuse through these channels and form a hydrogen bond with MES (Figure 64 and 70).

Table 7: Summary of the molecules from the precipitant solution which have been reported to co-crystallise with hGSTP1-1

PDB code	Resolution (Å)	Space group	Precipitant ions co-crystallised with hGSTP1-1		
			Type of ion molecule	Stoichiometry (per dimer)	Chain location
6Y1E	1.4	$P12_11$	PEG	2	Chain C and D
3CSI	1.9	$P12_11$	SO_4^{2-}	1	Chain B
			Cl^-	2	Chain A, B, C and D
			Ca^{2+}	1	Chain B and D
			CO_3^{2-}	1	Chain B
3HJM	2.1	$P12_11$	PO_4^{2-}	1	Chain D
			Cl^-	2 & 1	Chain A, B and D
			Ca^{2+}	2	Chain A, B, C and D
			CO_3^{2-}	2	Chain A, B, C and D
3CSJ	1.9	C121	Cl^-	1	Chain A
2GSS	1.9	C121	SO_4^{2-}	1	Chain A
8GSS	1.9	C121	SO_4^{2-}	1	Chain A and C
9GSS	1.9	C121	SO_4^{2-}	1	Chain A
6AP9	1.5	C121	Ca^{2+}	2	Chain A and B
			CH_3COO^-	2	Chain A and B
5JCW	1.9	C121	Ca^{2+}	2	Chain A and B
			CH_3COO^-	2	Chain A and B

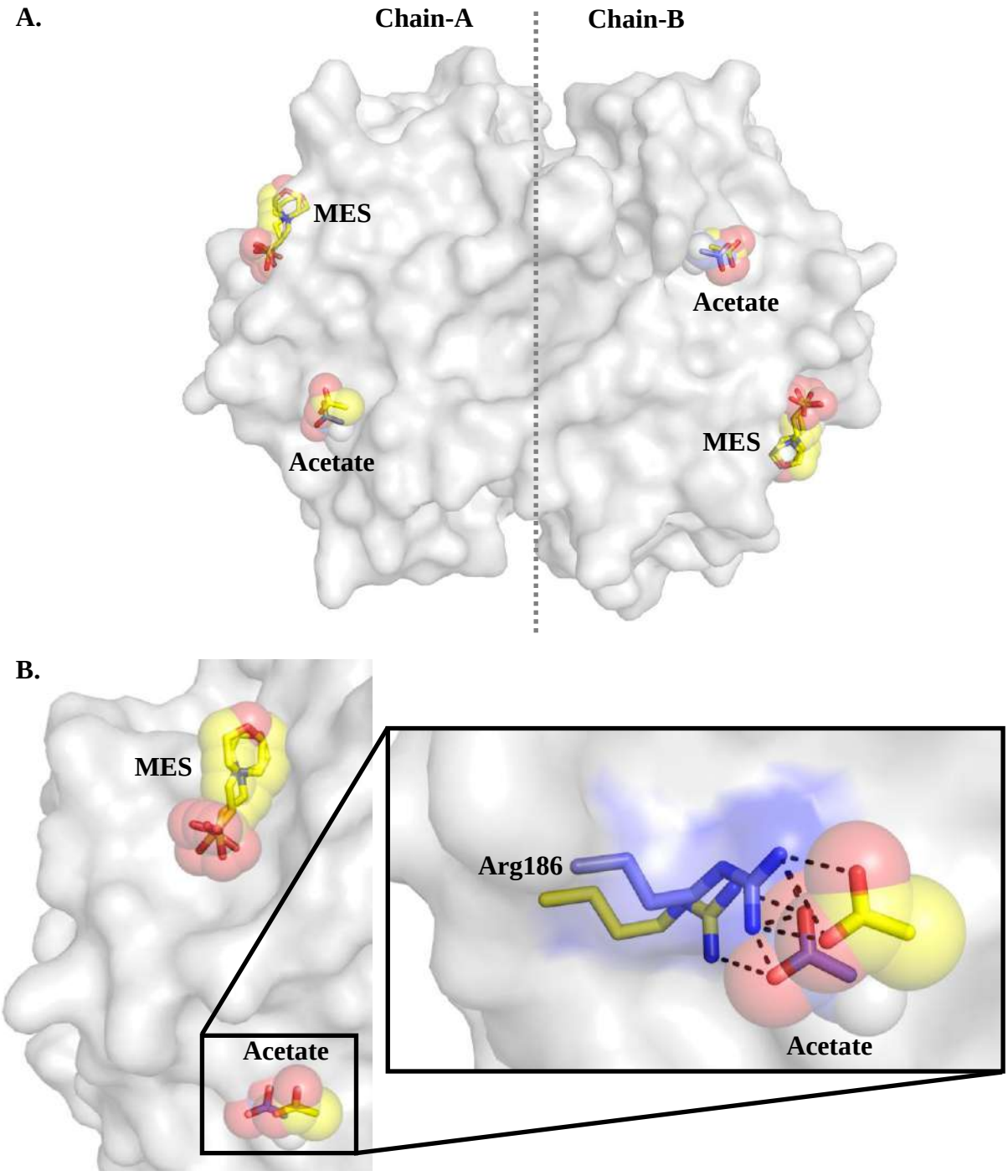


Figure 69. The binding of acetate to hGSTP1-1 is specific and occurs at a defined site. (A) To date, there are only two crystal structures available in the PDB of hGSTP1-1 in complex with an acetate ion bound to each monomer (PDB ID: 6AP9 (●) and 5JCW (●)). (B) In these structures, it was noted that the acetate ion interacted with hGSTP1-1 by forming a hydrogen bond only with the side chain of Arg186. This correlated with the findings in this study whereby the second acetate moiety of IAA-94 also formed a hydrogen bond with the side chain of Arg186 at the crystal-packing interface formed between chain-A and chain-D. All images were generated by using Pymol V1.3.

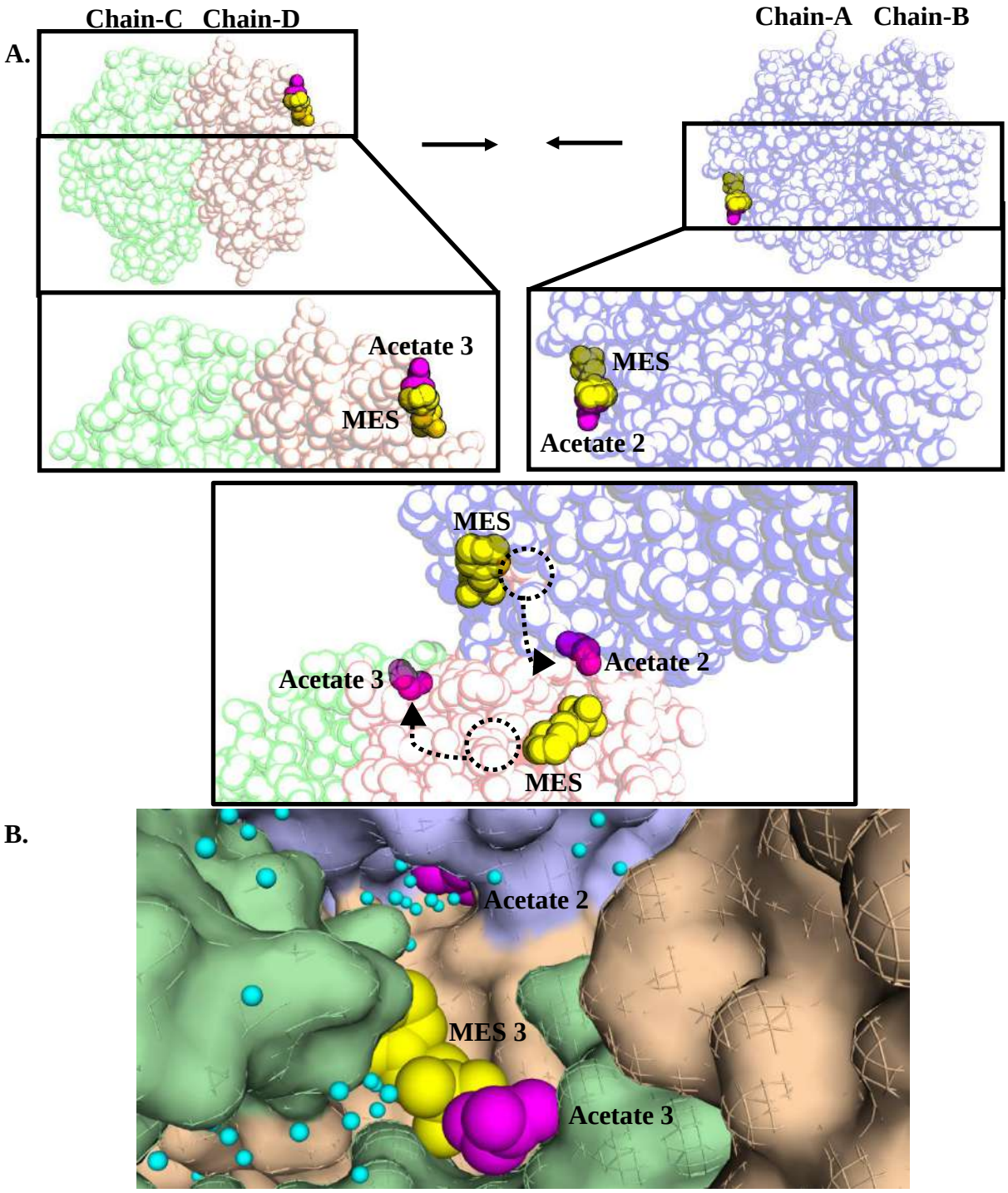


Figure 70. A schematic representing the displacement of IAA-94 from chain-A and chain-D due to crystal packing contacts. (A) The crystal contacts formed between chain-A and chain-D resulted in the displacement (⇢) of IAA-94 from both chains during the crystallisation of hGSTP1-1. Yet, owing to the affinity that Arg186 has for an acetate molecule, the displaced IAA-94 molecule from either chain-A or chain-D would have been shifted to the crystal-packing interface where Arg186 would have coordinated the binding of its acetate moiety. (B) However, the third acetate moiety of IAA-94 which interacted with MES on the outer surface of the dimer interface could have accounted for the second displaced molecule of IAA-94 considering that the solvent channels within the crystal structure of hGSTP1-1 would have allowed the second displaced molecule of IAA-94 to diffuse through the crystal structure and form a hydrogen bond with MES. Indicated above is chain-A (●), chain-C (●), chain-D (●) and water molecules (●) that filled the solvent channels in the crystal structure of hGSTP1-1. All images were generated by using Pymol V1.3.

Although the crystal contacts offered insight into the displacement of IAA-94 from hGSTP1-1 during crystallisation, the crystal contacts also provided further insight as to why there was no interaction observed between the acetate moiety of IAA-94 and chain-C in the crystal structure. During the crystal packing of hGSTP1-1 it was noted that the MES-binding site of chain-B was adjacent to the MES-binding site of chain-C from a neighbouring unit cell. Therefore, due to the fact that crystal contacts interfered with the binding of IAA-94 to hGSTP1-1, the region formed between both MES sites would not have been able to accommodate two highly dynamic IAA-94 molecules due to steric hindrance and consequently the acetate moiety of IAA-94 was only observed on chain-B (Figure 71).

In the study by Prade *et al*. (1997), they were able to show that MES had an influence on the binding of cholic acid and when attempting to soak the crystals of hGSTP1-1 with cholic acid in order to exchange MES with cholic acid, it led to the destruction of the crystals. Therefore, it was suggested that the crystal contacts were being disrupted by the binding of cholic acid at or near the MES-binding site. Consequently, in order to successfully understand the interaction between IAA-94 and hGSTP1-1, the protein-ligand complex would possibly need to be crystallised in a different space group. Though, in the PDB, hGSTP1-1 has typically been crystallised in the $P12_11$, $P2_12_12_1$ or C121 space group, all of which have the MES binding site being closely associated with the crystal contacts. However, two crystal structures of hGSTP1-1 (PDB: 1GSS and 2J9H) have been crystallised in the absence of MES and in a $P4_32_12$ space group. Interestingly, the MES-binding site for these crystal structures were not involved in crystal contacts and therefore this crystal system may be able to accommodate the binding of IAA-94 to the L-site in the crystal structure of hGSTP1-1.

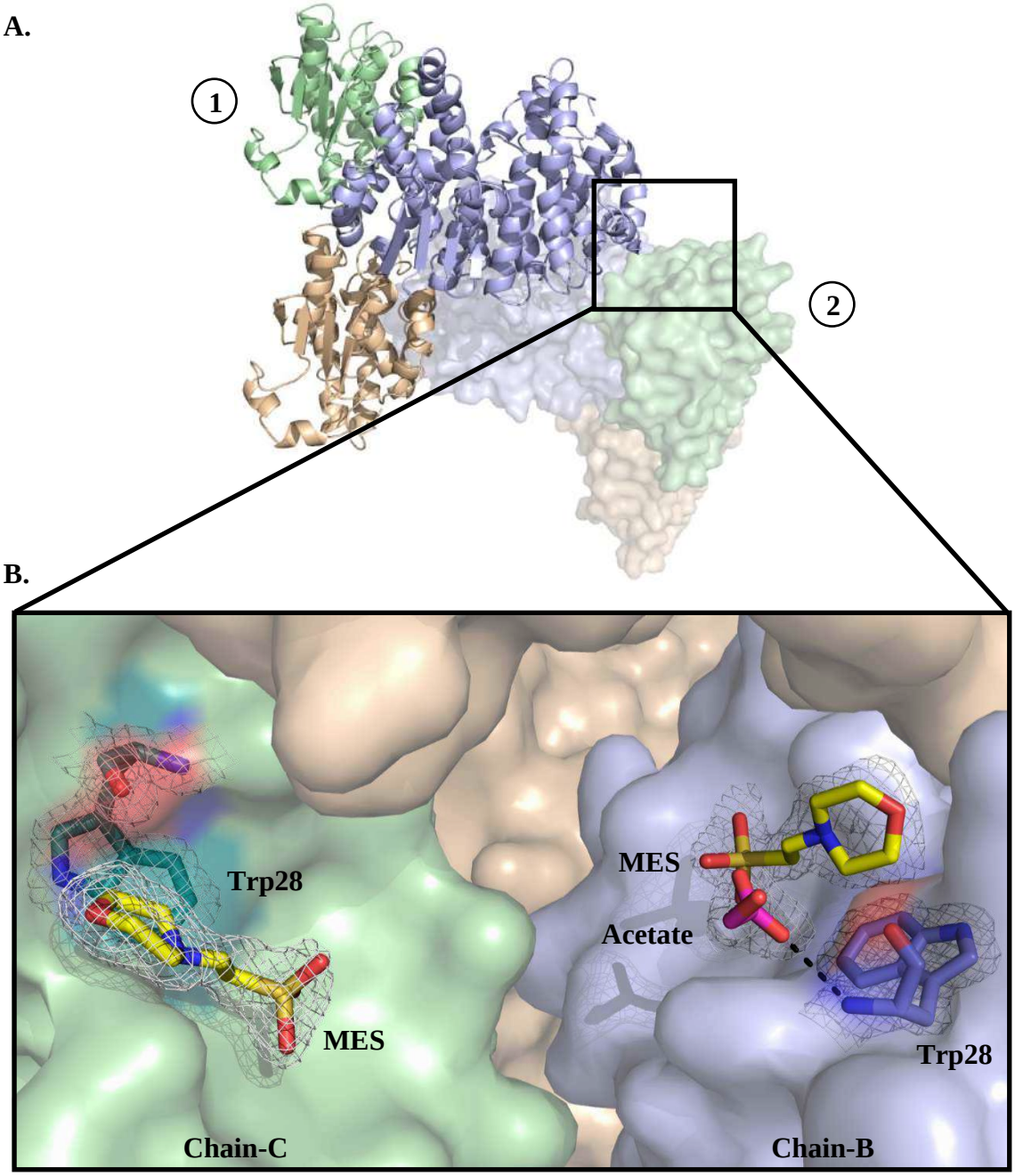


Figure 71. Crystal contacts provide insight as to why the acetate moiety of IAA-94 was absent from chain-C. (A) Indicated are two neighbouring unit cells (1 and 2), both of which contain chain-A (●), chain-B (●), chain-C (●) and chain-D (●) of hGSTP1-1. (B) In the crystal packing arrangement of hGSTP1-1 it was noted that the MES-binding site of chain-B was adjacent to the MES-binding site of chain-C from a neighbouring unit cell. Therefore, due to the fact that crystal contacts interfered with the binding of IAA-94 to hGSTP1-1, the region formed between both MES sites would not have been able to accommodate two highly dynamic IAA-94 molecules due to steric hindrance and consequently the acetate moiety of IAA-94 was only observed on chain-B. The residues have been placed in a sigma-weighted 2Fo - Fc electron density map ($\sigma = 1$), therefore indicating how the refined structures fitted in their electron density map. All images were generated by using Pymol V1.3.

3.4.2.3.4 The binding of IAA-94 to hGSTP1-1 describes the displacement of ANS

Collectively, these findings coincided with the findings obtained from the spectroscopic-based techniques used to probe the interaction between hGSTP1-1 and IAA-94 (section 3.2). That is, because IAA-94 was found to bind near to the MES-binding site, or otherwise the L-site, IAA-94 did not compete with CDNB at the H-site and hence there was no inhibition in the CDNB enzyme activity of hGSTP1-1 by IAA-94 (Figure 38). In addition, due to the absence of complete ANS displacement by both IAA-94 and EAA (Figure 40), it appeared that either the binding affinity for the interaction between hGSTP1-1 and ANS was greater than IAA-94 and EAA, or that there could simply be more than one binding site for ANS. In the latter case, considering that EAA binds at the H-site, but IAA-94 binds at the L-site, these two sites could account for the different binding sites of ANS.

In fact, IFMD data was able to show that when ANS was bound at the H-site, its sulfonate group was coordinated by Arg13, while its naphthalene moiety was involved in a π-π stacking interaction with Tyr108 (Figure 72) . Therefore, bearing in mind that these residues are directly involved in coordinating the interaction between EAA and hGSTP1-1, it was evidently clear as to how EAA was able to displace ANS from hGSTP1-1. On the other hand, the IFMD of ANS at the L-site indicated that its binding was coordinated by Lys29 and Glu30 forming a hydrogen bond with the sulfonate group of ANS (Figure 73). However, although these residues were not directly involved in the interaction with IAA-94, their proximity to Trp28 indicated that steric hindrance caused by the binding of IAA-94 would have disrupted this interaction and caused the displacement of ANS from hGSTP1-1. Therefore, although having distinct binding sites, the ability for both IAA-94 and EAA to displace ANS indicated that ANS bound to more than one site within hGSTP1-1.

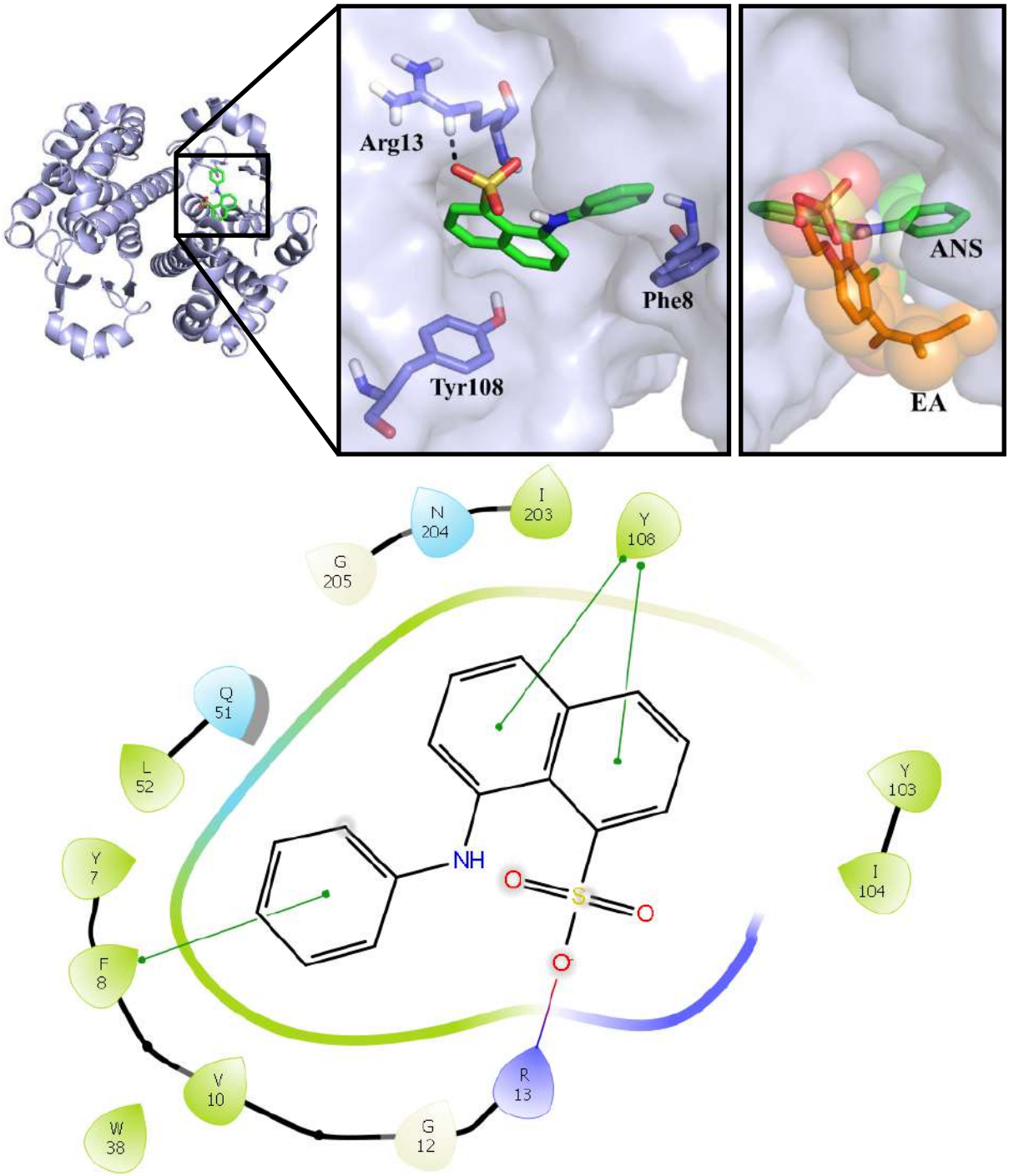


Figure 72. IFMD of ANS at the EAA-binding site of hGSTP1-1. The IFMD of ANS at the EAA-binding site of chain-B from hGSTP1-1 was validated by selecting the docked pose which had the lowest IFMD score. IFMD of hGSTP1-1 in complex with ANS at its H-site suggested that the sulfonate moiety of ANS was coordinated by a hydrogen bond with Arg13, while the naphthalene moiety of ANS was coordinated by a π-π stacking interaction with Tyr108. Therefore, bearing in mind that these residues are directly involved in coordinating the interaction between EAA and hGSTP1-1, it is clear that ANS will displace EAA from hGSTP1-1. The GE_{model} and components of the G_{score} can be found in Table S4. Indicated are positively charged residues (●), hydrophobic residues (●), polar residues (●), salt bridges (—) and π-π interactions (—).The top (3D diagram) and bottom (2D diagram) images for the protein-ligand interaction were generated by using Pymol V1.3 and the 2D protein-ligand interaction diagram tool in Mastero 11.2, respectively. *See section 2.2.13 for the experimental details pertaining to this method.*

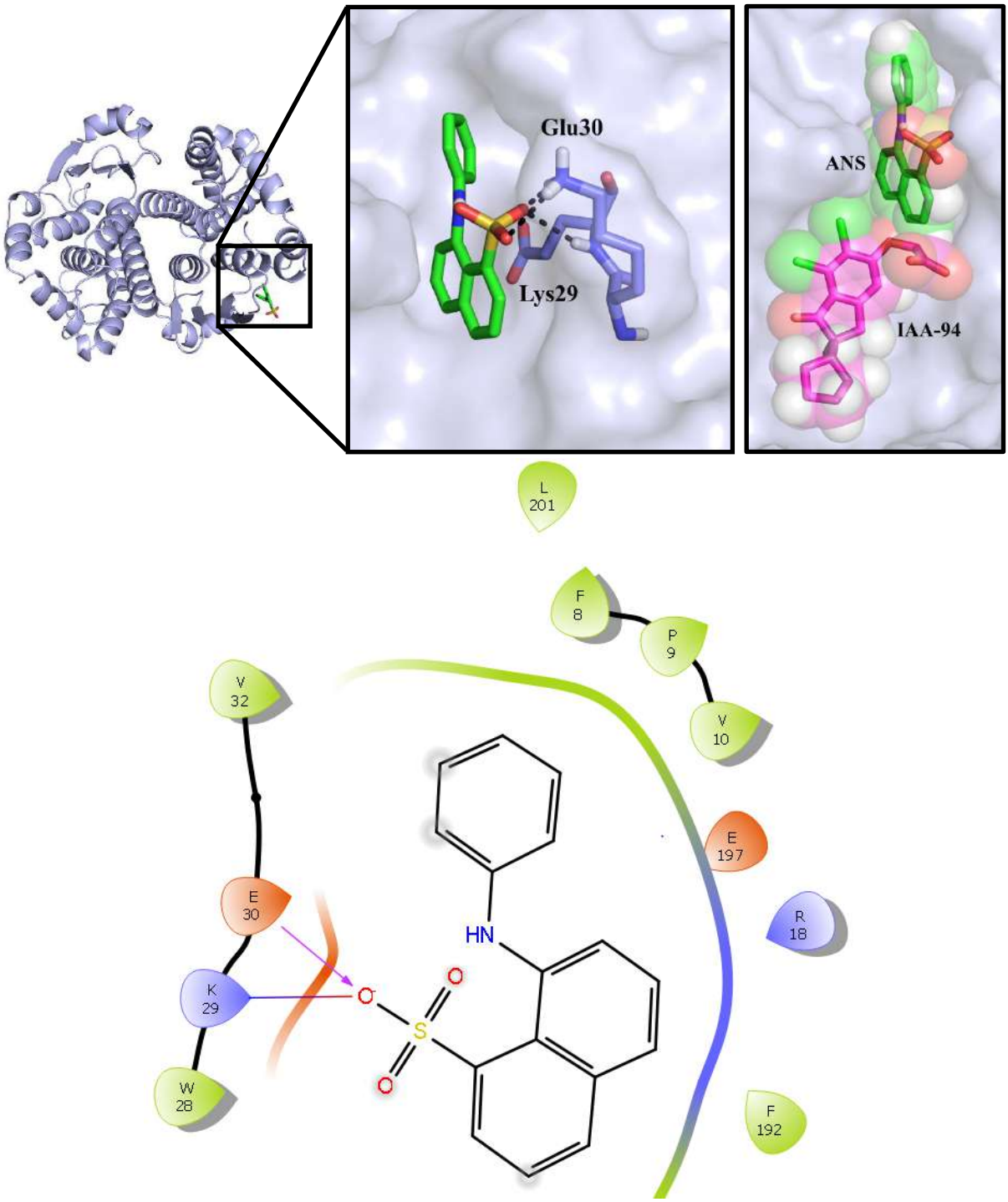


Figure 73. IFMD of ANS at the MES (putative IAA-94)-binding site of hGSTP1-1. The IFMD of ANS at the MES (putative IAA-94)-binding site of chain-B from hGSTP1-1 was validated by selecting the docked pose which had the lowest IFMD score. IFMD of hGSTP1-1 in complex with ANS suggested that the sulfonate moiety of ANS was coordinated by a hydrogen bond with Lys29 and Glu30. Although these residues were not directly involved in the interaction with IAA-94, their proximity to Trp28 indicated that steric hindrance caused by the binding of IAA-94 would have interfered with this interaction and caused the displacement of ANS from hGSTP1-1. The GE_{model} and components G_{score} for the top pose can be found in Table S4. Indicated are positively charged residues (●), negatively charged residues (●), hydrophobic residues (●), hydrogen bonds (→) and salt bridges (—).The top (3D diagram) and bottom (2D diagram) images for the protein-ligand interaction were generated by using Pymol V1.3 and the 2D protein-ligand interaction diagram tool in Mastero 11.2, respectively. *See section 2.2.13 for the experimental details pertaining to this method.*

3.4.2.4 The IFMD of IAA-94 to EcSspA

In relation to the MES-binding site of hGSTP1-1, it was noted that in contrast to CLIC1-WT, both EcSspA and YpSspA shared significant structural homology for this site (Table 4). Therefore, considering that the MES-binding site was suggested to also be the putative IAA-94-binding site of hGSTP1-1, it was suggested that IAA-94 also bound to this site in EcSspA. In fact, it has recently been demonstrated that ANS was displaced from EcSspA by both MES and IAA-94, therefore indicating that ANS, IAA-94 and MES shared a similar binding site (Rethabile Mokoena MSc dissertation, in preparation). Consequently, considering that MES and IAA-94 were found to bind at a similar site in hGSTP1-1, it was suggested that this structurally homologous site was also the putative IAA-94-binding site of EcSspA. Structural sequence alignment indicated that the ^{191}AFLS^{194} motif at the MES (putative IAA-94)-binding site of hGSTP1-1 was conserved in both YpSspA and EcSspA. Therefore, considering the importance of Phe192 in coordinating the interaction between IAA-94 and hGSTP1-1 through a π-π stacking interaction, it was suggested that Phe196 would be important for the coordination of IAA-94 to the putative IAA-94-binding site of EcSspA. Therefore, IFMD was performed in order to provide some insight into this interaction.

Based on IFMD, it was found that the docked structure of IAA-94 had its acetate moiety coordinated by forming a hydrogen bond with Arg208. However, Phe196 was not involved in any π-π stacking interaction, instead it was observed in some poses that Arg208 could also be involved in a π-cation stacking interaction with the indanyl group of IAA-94 (Figure 74). Therefore, similar to the interaction between EAA and CLIC1-WT, although IFMD did not indicate that Phe196 was involved in a π-π stacking interaction with IAA-94, in solution the protein backbone and side chains of EcSspA could be more dynamic and consequently this kind of interaction could occur to ensure the favourable interaction between IAA-94 and EcSspA. Yet, given that the interaction between IAA-94 and EcSspA was entropically unfavourable (Figure 49), it is possible that the proportion of the hydrophobic surface area of IAA-94 which became buried upon binding was not sufficient to ensure this kind of interaction. Therefore, without a crystal structure of EcSspA in complex with IAA-94, the true binding site for IAA-94 remains unknown. However, based on molecular modelling, structural sequence alignment and spectroscopic work it appears that the putative IAA-94-binding site of EcSspA and hGSTP1-1 are structurally homologous. In fact, this was interesting considering that Leu29 from the putative TMD of EcSspA was one of the conserved residues at this site, therefore if IAA-94 stabilises this binding site it could attenuate the membrane insertion of EcSspA.

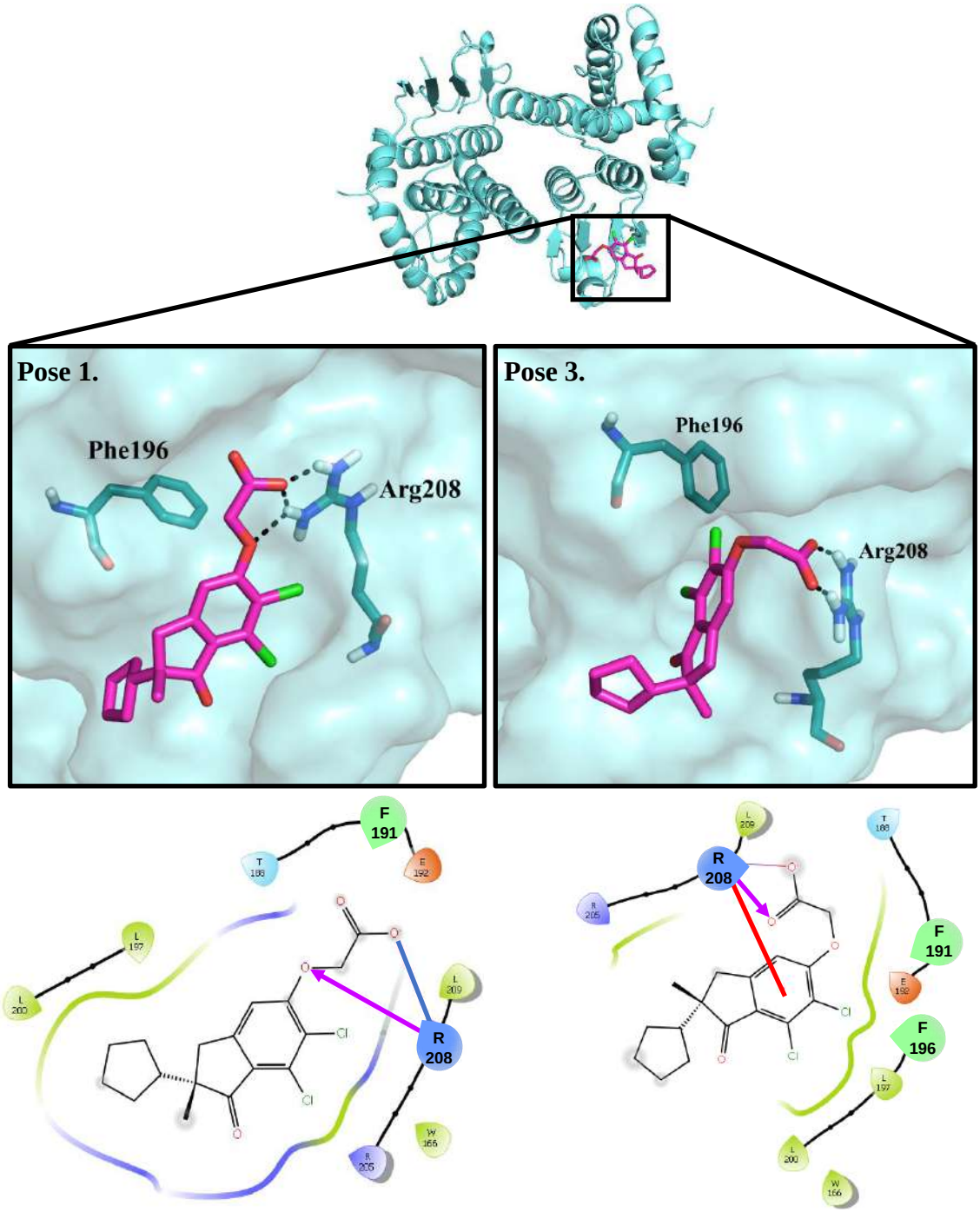


Figure 74. IFMD of IAA-94 at the putative IAA-94-binding site of EcSspA. The IFMD of IAA-94 to the putative IAA-94-binding site of chain A from EcSspA was validated by selecting the docked pose which had the lowest IFMD score. IFMD of EcSspA in complex with IAA-94 suggested that the acetate moiety of IAA-94 was coordinated by a hydrogen bond with Arg208. Furthermore, the indanyl group of IAA-94 was not involved in any π-π stacking interaction with Phe196, instead it was observed in some poses that Arg208 could also be involved in a π-cation stacking interaction with the indanyl group of IAA-94. The GE_{model} and components for the G_{score} for both pose 1 and pose 3 can be found in Table S4. Indicated are positively charged residues (●), negatively charged residues (●), hydrophobic residues (●), polar residues (●), hydrogen bonds (→), salt bridges (—) and π-cation interactions (—). The top and bottom images were generated by using Pymol V1.3 and the 2D protein-ligand interaction diagram tool in Mastero 11.2, respectively. *See section 2.2.13 for the experimental details pertaining to this method.*

3.5 The importance of cysteine residues in the structural stability and PTM of CLIC1

The binding site for GSH within cytosolic GSTs is formed by a conserved thioredoxin-like domain, however, based on sequence and structural similarity, there are two major subgroups and these include (i) Y-GSTs which utilise a tyrosine residue to activate GSH and (ii) S/C-GSTs which utilise a serine or cysteine residue in their interaction with GSH (Oakley, 2011). In relation to CLIC1 (Harrop *et al.*, 2001) and hGSTO1-1 (Board, 2000), it was noted that Cys24 and Cys32 formed a mixed disulfide bond with GSH at their active sites, respectively (Figure 7). Interestingly though, the omega class of GSTs have a dehydroascorbate (DHA) reductase activity, that is, they catalyse the GSH-dependent reduction of DHA (oxidised ascorbate) to ascorbate, a well-known cofactor for numerous enzymes which are involved in the defence mechanisms against oxidative stress (Board, 2000). Therefore, similar to hGSTO1-1, in view of the fact that Cys24 governs the enzyme activity of CLIC1 (Khamici *et al.*, 2015), *in vivo* CLIC1 may be involved in catalysing a number of critical reduction reactions.

However, although Cys24 plays a critical role in the enzyme activity of CLIC1, it also forms part of the single TMD of CLIC1 (Figure 9) and therefore depending on the redox state of the cell, Cys24 may be involved in regulating the membrane insertion and/or ion channel activity of CLIC1. In fact, compounds such as NEM which target cysteine residues have been reported to attenuate the Cl^- channel activity of CLIC1 (Singh and Ashley, 2006). Therefore, considering that in certain disease states such as in AD where the increased membrane insertion of CLIC1 is accompanied by an increase in RNOS such as GSNO (Novarino *et al*., 2004), a compound renowned for the PTM of cysteine residues, the importance of Cys24 in potentially regulating the membrane insertion and/or ion channel activity of CLIC1 cannot be underestimated.

Even though the findings from the ITC experiments indicated that IAA-94 did not interact with CLIC1-WT (Figure 42), EAA was instead shown to interact with CLIC1-WT (Figure 43) and based on IFMD studies it was apparent that when EAA bound to CLIC1-WT, it sterically hindered the thiol group of Cys24 and as a result EAA would be able to attenuate the enzyme activity of CLIC1-WT (Figure 52). Therefore, in relation to future drug design whereby CLIC1 can be implemented as a target protein, an important approach to inhibiting the enzyme activity of CLIC1 could rely on using the 2, 3-dichlorophenoxyacetate moiety of EAA as a

pharmacophore for the rational drug design of novel inhibitors that can effectively mask Cys24. Furthermore, depending on the role that Cys24 may have in the membrane insertion of CLIC1, targeting Cys24 could play a role in stabilising soluble CLIC1 in order to decrease the proportion of CLIC1 that is found in its membrane-bound conformation. However, even in the membrane-bound conformation of CLIC1, Cys24 in its reduced and/or *S*-nitrosylated state can become a suitable target site in order to regulate the ion channel activity of CLIC1. Yet, although it is apparent that Cys24 is an important residue within CLIC1, considering that there are six cysteine residues in CLIC1, it remains uncertain if Cys24 is the main reactive cysteine residue, how it influences the stability of soluble CLIC1 and if it is in fact a target site for *S*-nitrosylation. Therefore, by using various biophysical chemistry techniques, additional insight was provided to the relevance of Cys24 in soluble CLIC1.

3.5.1 Probing the main reactive cysteine residue within soluble CLIC1

It was shown by Son *et al*. (2010) that when halide derivatives of fluorescein became covalently bound to Cys32 in hGSTO1-1, they not only influenced the enzyme activity of hGSTO1-1, but they also labelled the protein. Therefore, considering the structural homology between the active sites of CLIC1 and hGSTO1-1 (Figure 7), the fluorescein derivate 6-IAF was used to label the main reactive cysteine residue/s within CLIC1. Interestingly, it was shown through ITC that 6-IAF interacted with CLIC1-WT with a stoichiometry of n = 0.97 (Figure 75), therefore indicating that one molecule of 6-IAF had bound to one molecule of CLIC1-WT and that there is one main reactive cysteine residue in CLIC1-WT. Consequently, considering the importance of Cys24, it was suggested that it became labelled with 6-IAF, however, it needs to be kept in mind that fluorescein is also a bulky molecule and therefore the other reactive cysteine residues within CLIC1-WT may not have been labelled by the iodoacetamide (IAM) moiety of 6-IAF due to steric hindrance. Nevertheless, in order to confirm that Cys24 was the main reactive cysteine residue, a titration of 6-IAF with CLIC1-C24A was performed. During this titration the magnitude of the heat produced correlated with the magnitude of the heat produced when 6-IAF was titrated into buffer (Figure 75 and 76), therefore indicating that there was no interaction between 6-IAF and CLIC1-C24A, but more importantly that Cys24 was the main reactive cysteine residue within CLIC1.

The interaction between CLIC1-WT and 6-IAF was noted to be enthalpically favourable, but entropically unfavourable and therefore the overall interaction was enthalpically driven

(Figure 75). This confirmed that the binding specificity of CLIC1-WT for 6-IAF can be attributed to the number and/or strength of non-covalent protein-ligand interactions and also possibly ligand-solvent interactions. In the absence of structural data of CLIC1-WT in complex with 6-IAF it is difficult to correlate how the obtained thermodynamic parameters reflect the structural changes upon ligand binding. However, based on IFMD it was observed that the specificity for the interaction between CLIC1-WT and 6-IAF was governed by Asp109 and Tyr233 forming a hydrogen bond with the hydroxyl groups of the xanthene moiety of 6-IAF. Furthermore, and most importantly, the covalent bond formed between Cys24 and the IAM moiety of 6-IAF was the main bond to contribute to the specificity and therefore the favourable ΔH for this interaction (Figure 77). Interestingly, based on IFMD, the xanthene moiety of 6-IAF was predicted to be ~5.4 Å apart from Phe111. Therefore, in solution where the protein backbone and side chains of CLIC1-WT would be more dynamic, Phe111 may be involved in a π-π stacking interaction with 6-IAF to further ensure the favourable interaction between CLIC1-WT and 6-IAF.

However, although the xanthene moiety of 6-IAF may become buried enough in order to form a π-π stacking interaction with Phe111, the overall interaction between CLIC-WT and 6-IAF was entropically unfavourable (Figure 75). Therefore, upon binding, the proportion of the hydrophobic surface area of 6-IAF that became buried to result in a desolvation event, and subsequently a favourable ΔS, was minimal. In fact, this observation can be related to the IFMD data (Table S4) which illustrated that in comparison to the IFMD of EAA to the active site of CLIC1-WT, the Glide ligand efficiency score for the IFMD of 6-IAF to the active site of CLIC1-WT was less negative, therefore indicating that the surface area of 6-IAF had less of an affect towards the favourable binding interaction between 6-IAF and CLIC1-WT. Furthermore, based on IFMD, it was clear that the xanthene moiety of 6-IAF did become slightly buried in order to interact with CLIC1-WT, but a significant region of the xanthene ring, and in particular the isobenzofuran moiety of 6-IAF was still exposed to the solvent (Figure 77), therefore correlating with the findings reported by ITC. However, without any structural data it cannot be discounted that the interaction between CLIC1-WT and 6-IAF could also mediated by water molecules and therefore should this network of water molecules have become more ordered during the interaction, this would have contributed towards the entropic penalty.

The degree of specificity for this interaction can be further related to the tight binding given that the K_d value was reported to be in the nanomolar range (Figure 75), or perhaps even lower

since the Nano-ITC can only detect K_d values in the millimolar to low nanomolar range. Consequently the c-value for this interaction was c > 1000 and therefore the isotherm profile corresponded to a step-like function. In molecular interactions were very tight binding occurs, as the ligand is titrated into the sample it will all become bound until saturation occurs and consequently a rectangular isotherm forms. Importantly, in these type of isotherms, the steep transition will occur at the stoichiometric equivalence point along the molar ratio and as a result only the ΔH and the n-value can be accurately determined for tight binding interactions such as the interaction between CLIC1-WT and 6-IAF (Chilom *et al.*, 2006).

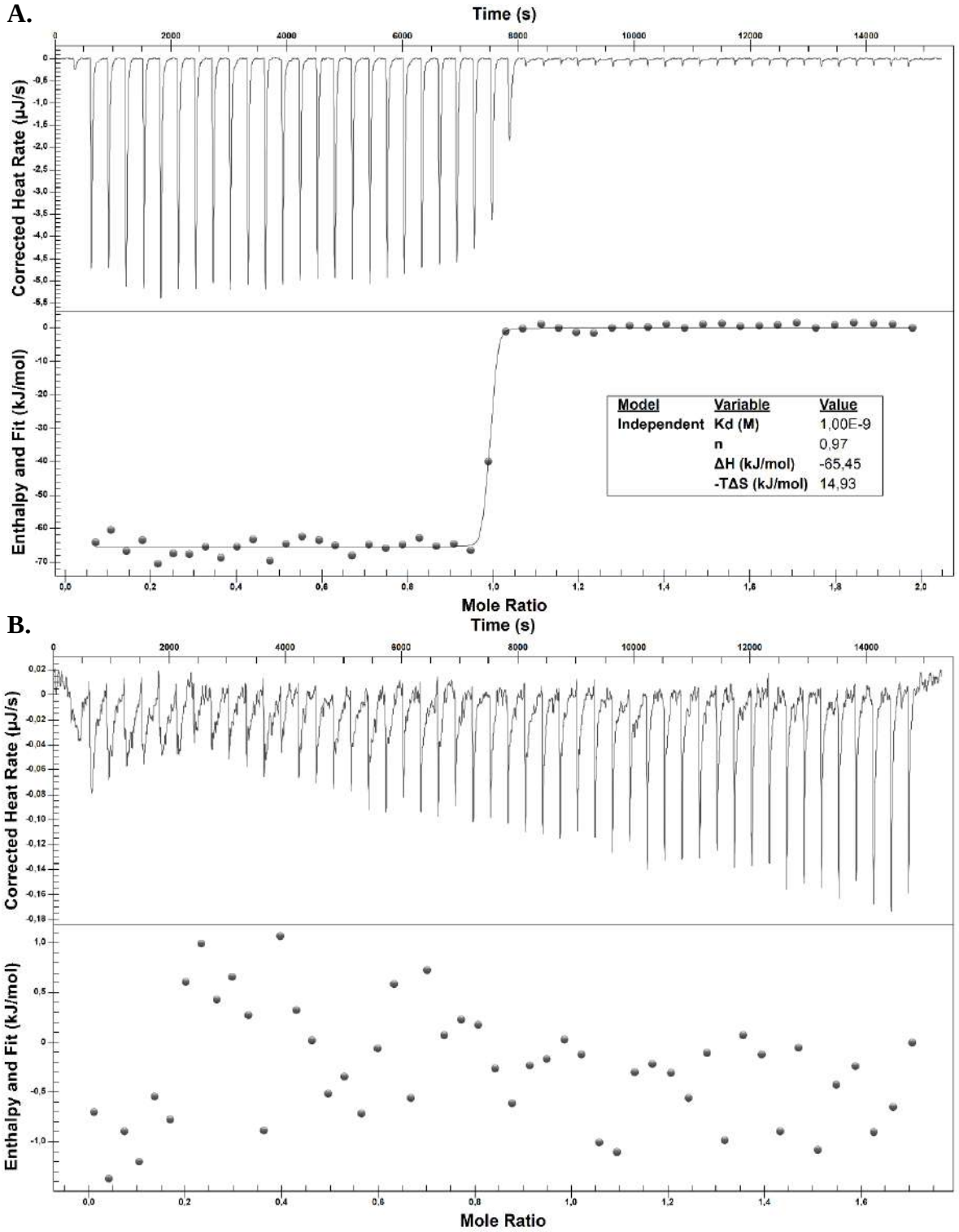


Figure 75. Monitoring the thermodynamic parameters for the interaction between 6-IAF and either CLIC1-WT or its C24A mutant. (A) Using ITC, the interaction between 6-IAF and CLIC1-WT was monitored by titrating 825 µM 6-IAF into 100 µM CLIC1-WT. Each titration was performed, in triplicate, at 20 °C with a stir rate of 250 rpm, an initial and final baseline of 300 sec and a duration period of 300 sec between each injection. The volume of the ITC syringe used was 250 µl and during each run the initial injection volume was 2 µl, but for all subsequent injections it was 5 µl and therefore a total of 49 injections were performed. The interaction occurred with a stoichiometry of $n \sim 1$, was entropically driven and had a K_d value in the nanomolar range. (B) No interaction was observed between 6-IAF and CLIC1-C24A. *See section 2.2.9.4 for the experimental details pertaining to this method.*

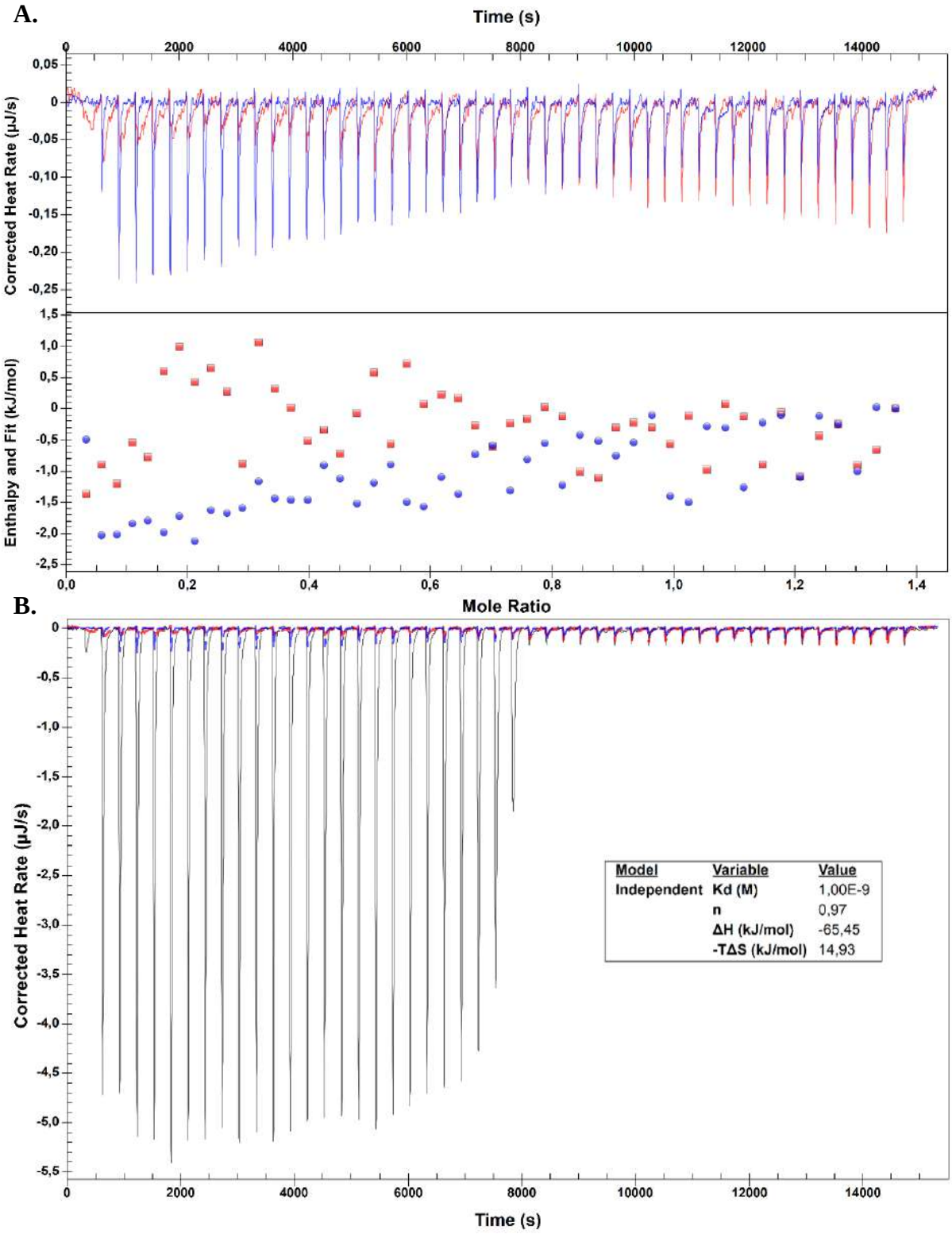


Figure 76. Cys24 is the main reactive cysteine residue within CLIC1. (A) The heat generated from the titration of 6-IAF into CLIC1-C24A (■) have been overlaid with the heat generated from the titration of 6-IAF into buffer (●), or otherwise the HOD. (B) The heat generated from the titration of 6-IAF into CLIC1-WT (●), CLIC1-C24A (●) and buffer (●) have all been overlaid. It was apparent that the heat generated during the titration of 6-IAF into CLIC1-C24A coincided with the HOD, therefore indicating that the C24A mutation prevented the binding of 6-IAF to CLIC1 and that Cys24 was the main reactive cysteine residue within CLIC1.

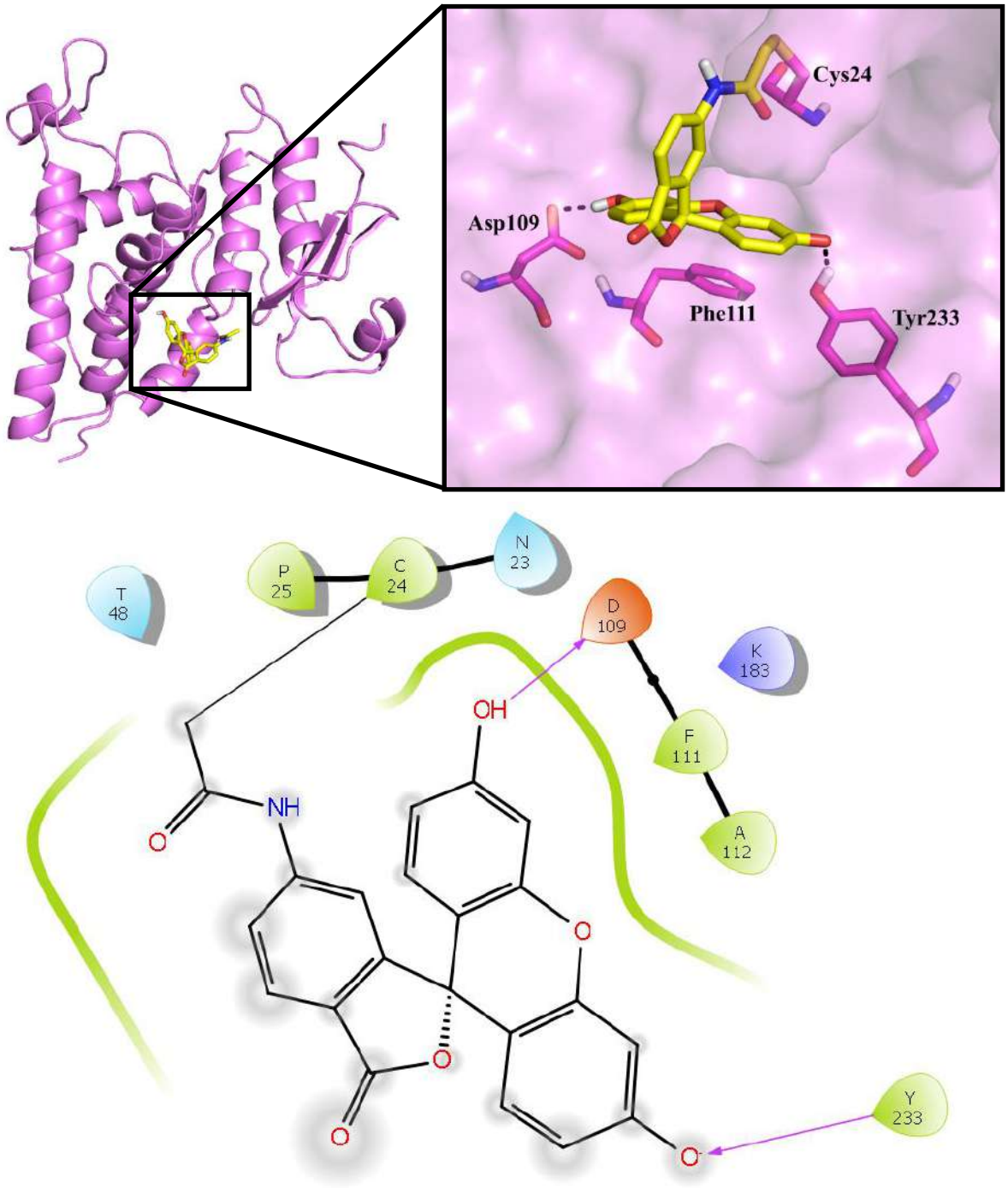


Figure 77. IFMD of 6-IAF to Cys24 at the active site of CLIC1-WT. The IFMD of 6-IAF to Cys24 at the active site of CLIC1-WT was validated by selecting the docked pose which had the lowest IFMD score. IFMD of CLIC1-WT in complex with 6-IAF suggested that Asp109 and Tyr233 formed a hydrogen bond with the hydroxyl groups of the xanthene moiety of 6-IAF, whereas Cys24 formed a covalent bond with the IAM moiety of 6-IAF. Furthermore, Phe111 was ~5.4 Å apart from the xanthene moiety of 6-IAF, therefore when taking into account the true global nature of CLIC1-WT, Phe111 could be involved in a π-π stacking interaction with 6-IAF to further ensure the favourable interaction between CLIC1-WT and 6-IAF. The GE_{model} and components of the G_{score} can be found in Table S4. Indicated are positively charged residues (●), negatively charged residues (●), hydrophobic residues (●), polar residues (●) and hydrogen bonds (→).The top and bottom images were generated by using Pymol V1.3 and the 2D protein-ligand interaction diagram tool in Mastero 11.2, respectively. *See section 2.2.13 for the experimental details pertaining to this method.*

3.5.2 The role of Cys24 in regulating the MD of soluble CLIC1

3.5.2.1 Monitoring the influence that the C24A mutation has on the MD of CLIC1

It was already shown that although the C24A mutation did not influence the local environment of Trp35 in CLIC1, it did have an influence on the mobility, packing environment and subsequently the spatial disposition of the phenylalanine and tyrosine residues (Figure 37). In addition, the C24A mutation reduced the thermal stability of CLIC1 (Figure 38). However, to better understand as to how the C24A mutation had an influence on the overall dynamics of CLIC1, MD simulations were performed on both CLIC1-WT and its C24A mutant.

One of the main parameters when performing a MD simulation is to select an appropriate force field that can accurately describe the native state of a protein. A force field is simply a mathematical expression that is used to describe the dependence of the potential energy of a system on the coordinates of its particles. The parameters used in these force field calculations are derived from quantum mechanical calculations which have been fitted to experimental data from NMR spectroscopy, small angle X-ray scattering (SAXS), X-ray and electron diffraction as well as various other spectroscopic techniques (Gonzalex, 2011). However, one of the main challenges for studying chemical systems *in silico* is being able to accurately describe the interaction between atoms, however, this can be achieved either through molecular mechanics (MM) or quantum mechanics (QM). The latter would be the most accurate approach to describe chemical systems given that it takes into account the electronic degrees of freedom as well as the nuclear motion between particles in order to describe the shifts in the electron density and the charge-transfer (polarizability) that accompany complex chemical phenomena such bond formation and breakage. Unfortunately, due to the large size of biological macromolecules, the benefits of these calculations come at the cost of requiring high performance computational resources (Hofer and de Visser, 2018).

Instead, most simulation protocols therefore employ MM given that it describes a molecular system by using classical (Newtonian) mechanical energy functions. In MM the Born-Oppenheimer approximation is assumed and in turn the potential energy of a system is calculated as a function of its nuclear coordinates. In summary, an all-atom MM simulation will (i) treat each atom as one particle, (ii) treat bond interactions as springs and (iii) assign a van der Waals radius, polarizability and constant net charge to each particle, as opposed to a united-atom MM simulation approach where the interatomic potentials between atoms are

treated as one interaction centre. Therefore, by using all-atom MM, all common force fields subdivide potential functions into two classes, namely (i) bonded interactions which account for bond-stretching, bond-angle bending and bond-torsional potential, and (ii) non-bonded interactions which constitute Lennard-Jones attraction and repulsion interactions (van der Waals interactions) as well as Coulomb electrostatic interactions. Therefore, in order for a MD simulation to have diverse applications in structural biology it is imperative that high quality force fields are utilised if accurate predictions are to be obtained (Hofer and de Visser, 2018; Kukol, 2008).

In this study the MD simulations of CLIC1-WT and its C24A mutant were implemented by using the OPLS-AA force field (section 2.2.12). The OPLS-AA force field is a broadly applicable and extensively developed force field that has been documented to give a highly accurate description of molecules within a fluidic system (Jorgensen *et al*., 1996). The bonded parameters of the OPLS-AA force field were mainly adopted from the AMBER-AA force field, however, the main focus by Jorgensen *et al*. (1996) in the development of the OPLS-AA force field was to focus on the non-bonded parameters by computing experimentally well-determined thermodynamic and structural properties from several organic liquids. Consequently, the OPLS-AA force field has been well parametrised for the MD simulations of protein and small organic molecules and was therefore well suited to monitor the MD of CLIC1-WT and its C24A mutant in this study.

Analysis of the data from the MD simulations of CLIC1-WT and CLIC1-C24A was initially performed by looking at the root-mean-square standard deviation (RMSD) of the α-carbon from the protein backbone in order to provide insight into the structural stability of both proteins throughout the run (Abraham *et al*., 2017; Lemkul, 2018). The RMSD for both proteins was noted to be different throughout the run, whereby the difference in the RMSD was $\geq$ 15% notably between 4-8 ns, 22-25 ns and 38-50 ns during the simulation (Figure 78), therefore indicating that compared to the wild-type conformation of CLIC1, the backbone of the C24A mutant was more dynamic. This observation was additionally related to the radius of gyration (R_g) for both proteins, a parameter used to measure the compactness of a protein. That is, if a protein is stably folded it will maintain a relatively steady R_g value, but should the protein become more dynamic its R_g value will therefore change over time (Abraham *et al*., 2017; Lemkul, 2018). Throughout the simulation both proteins could be considered to be stably folded considering that their R_g values were relatively constant (stable), however, the R_g values

for the C24A mutant were at times 1-5% different to the R_g values reported for CLIC1-WT (Figure 78), therefore reiterating that compared to CLIC1-WT, the C24A mutation increased the overall dynamics of CLIC1.

Furthermore, the data from the MD simulation was analysed by looking at the root-mean-square fluctuation (RMSF) which described the average deviation of an amino acid residue from a reference position over time. Therefore, the RMSF provided insight into the regions of CLIC1-WT/C24A that fluctuated from their mean structure the most (or least) throughout the simulation (Abraham *et al*., 2017; Lemkul, 2018). It was observed for CLIC1-WT that the amino acid residues 144-167 from α-helix 7 displayed the greatest RMSF (Figure 79), however, this correlated with the crystal structure for selected variants of CLIC1 (PDB ID: 1K0N, 4JZQ and 4IQA) whereby this region could not be solved owing to a lack of electron density that was attributed to the highly dynamic nature of this region. Similarly, this region was noted to be highly dynamic in CLIC1-C24A, but with no substantial difference to the RMSF of CLIC1-WT. However, in the C24A mutant there was a significant increase in the RMSF and hence the dynamics between residues 14-23 (loop between α1β1), 31-37 (α1) and 47-75 (α2β3β4) on the N-terminal domain of CLIC1-C24A (Figure 79).

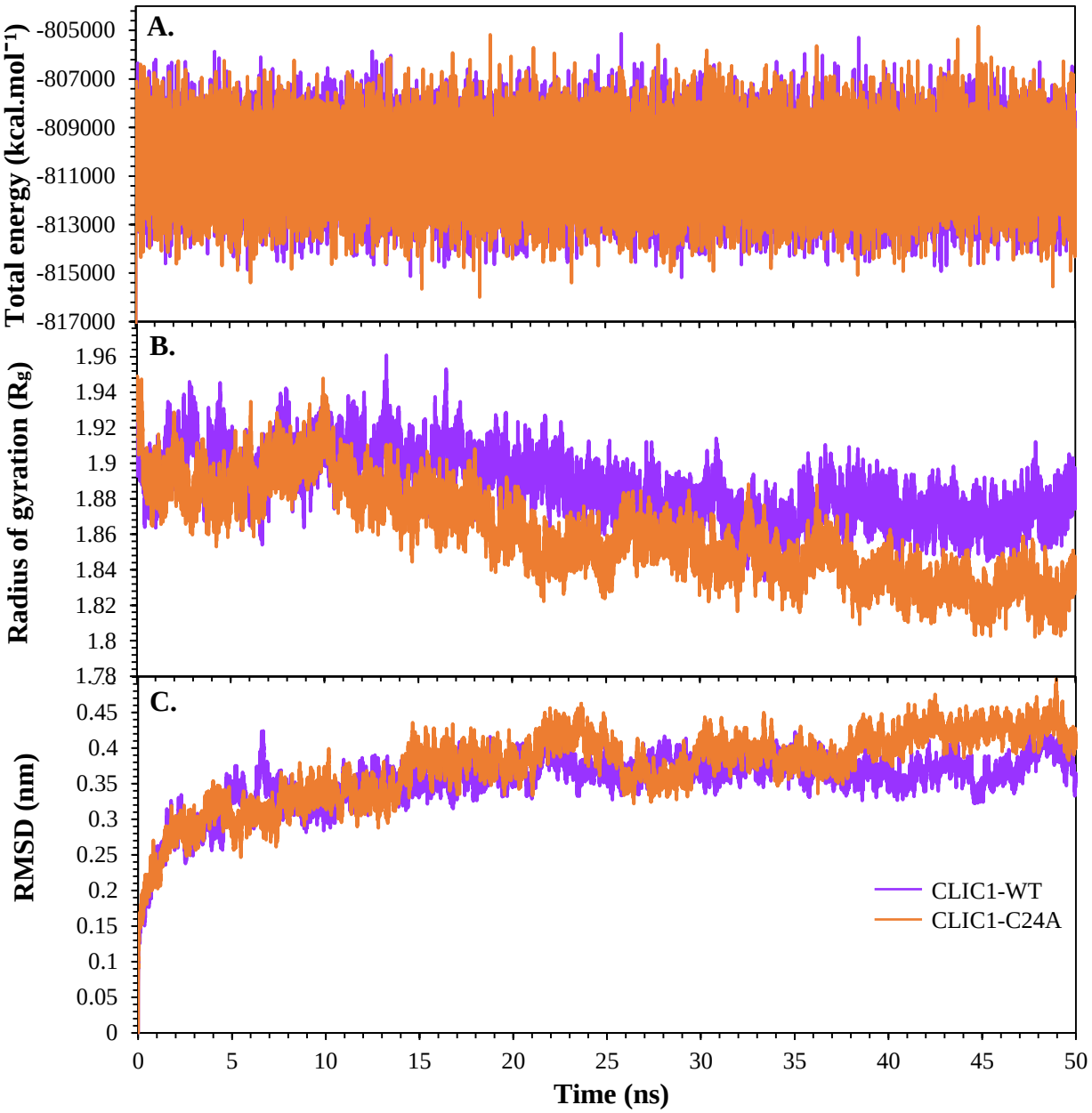


Figure 78. The MD simulations of CLIC1-WT and its C24A mutant. The MD simulations of CLIC-WT and CLIC1-C24A were performed by using the OPLS-AA force field. (A) Prior to the MD simulation run the system underwent EM and the system maintained a stable E_{pot} throughout the run on the order of 10^5-10^6 kJ.mol^{-1}. (B) The R_g for both proteins, a parameter used to measure the compactness of a protein, was relatively constant (stable) throughout the simulation and therefore both proteins were considered to be stably folded. However, the R_g values for the C24A mutant were at times 1-5% different to the R_g values reported for CLIC1-WT, therefore indicating that the C24A mutant increased the overall dynamics of CLIC1. (C) Monitoring the RMSD of the α-carbon from the protein backbone provided insight into the structural stability of both proteins throughout the simulation. Relative to CLIC1-WT, the RMSD for CLIC1-C24A was ≥ 15% different between 4-8 ns, 22-25 ns and 38-50 ns during the simulation, therefore indicating that the backbone of the C24A mutant was more dynamic. *See section 2.2.12 for the experimental details pertaining to this method.*

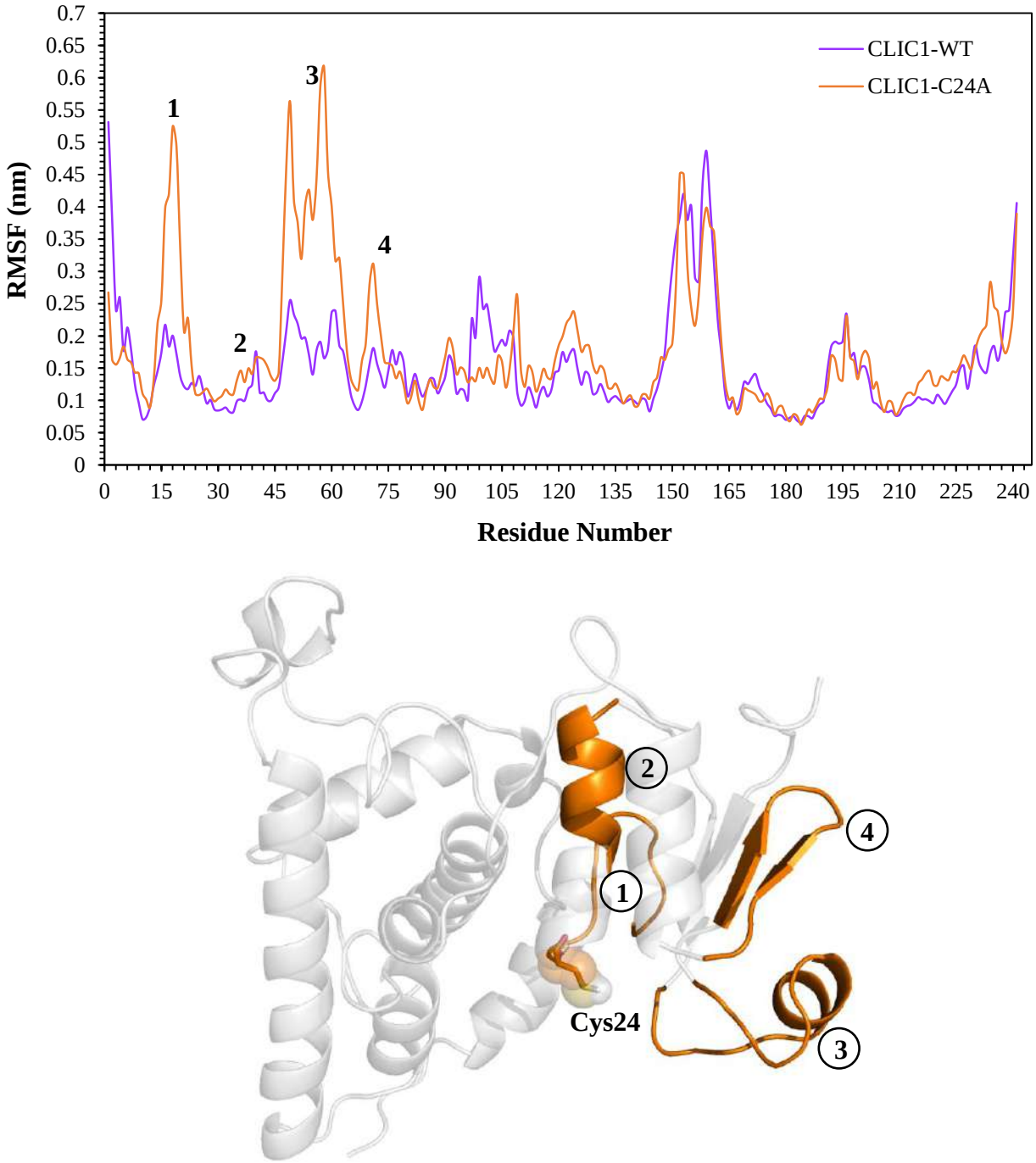


Figure 79. The C24A mutant influenced the dynamics of CLIC1 at the N-terminal domain. The RMSF described the average deviation of an amino acid residue from a reference position over time and therefore it provided insight into the regions of CLIC1-WT/C24A that were fluctuating from their mean structure the most (or least) throughout the simulation. It was observed for CLIC1-WT that the amino acid residues 144-167 from helix α7 displayed the greatest RMSF, however, based on previous crystal structures for selected variants of CLIC1, this region could not be solved owing to a lack of electron density that was attributed to the highly dynamic nature of this region. Similarly, this region was noted to be highly dynamic in CLIC1-C24A, but with no substantial difference to the RMSF of CLIC1-WT. However, in the C24A mutant there was a significant increase in the RMSF and hence dynamics between residues (1) 14-23 (loop between α1β1), (2) 31-37 (α1) and (3-4) 47-75 (α2β3β4) at the N-terminal domain of CLIC1. All images were generated by using Pymol V1.3. *See section 2.2.12 for the experimental details pertaining to this method.*

3.5.2.2 Understanding how the C24A mutation changes the MD of CLIC1

Apart from van der Waals interactions, the side chain of Cys24 was not involved in any hydrogen bonds or salt bridges with its neighbouring residues throughout the MD simulation. Therefore, in order to understand how the C24A mutation influenced the MD of CLIC1 at the N-terminal domain, the hydrogen bond network formed by the backbone of Cys24 and Ala24 with their neighbouring residues was investigated. In particular, it was noted that the protein backbone of Cys24 in CLIC1-WT mainly interacted with Gly22, Asn23, Ser27 and Glu28, however, in CLIC1-C24A it was noted that the protein backbone of Ala24 had an increased number of interactions with the backbone of Ser27, the side chain of Gln28 and also an additional interaction with the backbone of Lys13 (Figure 80).

When looking at the initial energy minimised structure of CLIC1-C24A that was used for MD simulations, it was noted that when the side chain of Cys24 was mutated, a void was created that consequently resulted in the neighbouring residues, most notably Lys13, changing their orientation (Figure 80). Therefore, considering that during protein folding a protein molecule would typically move down an energy landscape to assume the lowest energy conformation, during the folding of CLIC1-C24A the resultant change in the orientation of the residues that surround Ala24 could have induced a cooperative affect that increased the dynamics at the N-terminal domain by causing a change in the hydrogen bond network. Therefore, to better understand this phenomenon, the hydrogen bond network formed by the backbone and side chain of Lys13, Gly22, Asn23, Ser27 and Gln28 with their neighbouring residues was investigated.

It was initially noted that accompanying the decreased distance between the backbone of Ala24 and Ser27 was an increase in the number of hydrogen bonds formed between these two residues. Similarly, the decreased distance between the backbone of Ala24 and the backbone of Asn23 and Lys13 resulted in an increase in the number of hydrogen bonds being formed. In contrast, however, although there was an increase in the distance between the backbone of Ala24 and Gln28, these residues were brought in close proximity to one another from 10-21 ns of the simulation and this resulted in the formation of a hydrogen bond between the backbone of Ala24 and the side chain of Gln28. However, the overall distance between Ala24 and Gln28 had increased throughout the simulation (Figure 81 and 83).

However, it was only upon examining the interactions that Asn23 made with its surroundings residues that the increased dynamics at the N-terminal domain of CLIC1-C24A was made clear. It was observed that Asn23 in CLIC1-WT formed interactions with Lys13 and Gln28 along with Asp225 and Glu228 which are located on the all-α helical C-terminal domain near the domain interface of CLIC1-WT. Therefore, by Asn23 forming a hydrogen bond with Asp225 and Glu228 it appeared that Asn23 played a key role in regulating the stability of the domain interface of CLIC1-WT (Figure 82 and 83). In fact, it was reported by Cross *et al*. (2013) that because Glu228 formed a hydrogen bond with Ser16 at the N-terminal domain in the crystal structure of CLIC1-WT (PDB: 1K0M), when Glu228 was mutated to E228L it decreased the overall stability of CLIC1, therefore highlighting the significance of Glu228 in stabilising the domain interface of CLIC1. However, it is important to bear in mind that although protein crystals contain solvent channels with dynamically filled solvent molecules that allow proteins to have some degree of flexibility, the X-ray diffraction data collected from protein crystals only depict the average structural conformation of the protein molecules within the crystal. Therefore, while informative, a protein crystal structure may only represent one state in the overall dynamics of a protein (Rupp, 2009), hence the need to perform MD simulations in order to better understand the true dynamic nature of a protein. In fact, with respect to CLIC1-WT it was noted that Glu228 only formed a hydrogen bond with Ser16 once throughout the simulation, but instead Glu228 was noted to predominately form a hydrogen bond with Asn23 (Figure 82 and 83). Nevertheless, Glu228 is still imperative in stabilising the domain interface of CLIC1.

When Cys24 was mutated to Ala24, the distance between Asn23 and Glu228 along with the distance between Asn23 and Asp225 increased. Consequently, the number of hydrogen bonds formed between the backbone and side chain of these residues decreased, therefore it was suggested that the domain interface would become less stable. Interestingly, however, the number of hydrogen bonds formed between Glu228 and Ser16 increased in CLIC1-C24A given that the backbone and side chain for both residues moved closer together. As a result, this appeared to contradict the previous findings seeing that Ser16 compensated for the decreased number of hydrogen bonds that were formed between Glu228 and Asn23. Instead, the distance between Asp225 and the residues Ile21 and Ser16 appeared to increase and consequently the number of hydrogen bonds formed between these residues decreased (Figure 82 and 83).

Finally, it was also noted that accompanying Lys13 moving closer to Ala24 and Ser27 in CLIC1-C24A, the backbone and side chains of Asn23 and Lys13 moved closer to one another and hence there was an increase in the number of hydrogen bonds being formed between these two residues. However, due to the distance between Ala24 and Gln28 moving further apart during the simulation, this was also associated with Asn23 and Gln28 moving further apart and therefore there was a decrease in the number of hydrogen bonds being formed between these two residues (Figure 82 and 83). In view of these findings it was therefore apparent that when Cys24 was mutated to Ala24, the overall change in the dynamics of Lys13 and Asn23 decreased the number of hydrogen bonds that were formed between Asp225 and Asn23 together with the number of hydrogen bonds that were formed between Glu228 and Lys13. Furthermore, due to there being a change in the dynamics of Ser16 and Ile21, there was a decrease in the number of hydrogen bonds that Asp225 made with Ser16 and Iln21. Therefore, owing to the decrease in the number of hydrogen bonds being formed at the domain interface of CLIC1 (Figure 83), the C24A mutation induced a cooperative change in the hydrogen bond network of CLIC1 and ultimately this decreased the stability of CLIC1-C24A by causing the N-terminal domain to become more dynamic.

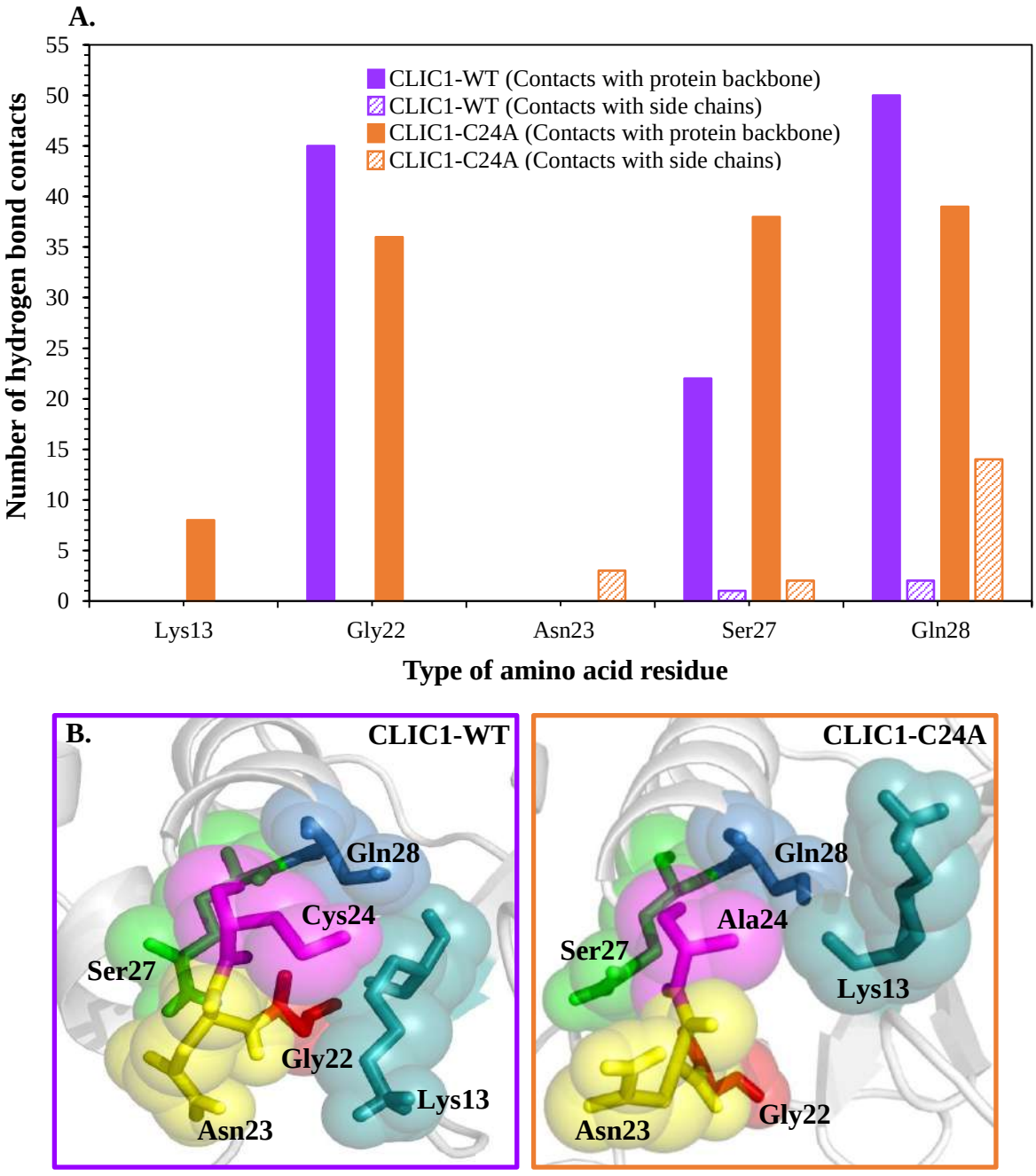


Figure 80. The interactions that Cys24 and Ala24 make with their surrounding residues. (A) Apart from van der Waals interactions, the side chain of Cys24 was not involved in any hydrogen bonds or salt bridges with its neighbouring residues. Therefore, to understand how the C24A mutation influenced the dynamics at the N-terminal domain of CLIC1, the hydrogen bond network formed by the backbone of Cys24 and Ala24 with their neighbouring residues was investigated. This was achieved by taking the sum for the number of hydrogen bond contacts that the backbone of Cys24 or Ala24 made with the backbone (■) and/or side chain (▨) of its neighbouring residues. The backbone of Cys24 mainly interacted with Gly22, Asn23, Ser27 and Gln28, however, the backbone for Ala24 also showed an increase in number of interactions with the backbone of Ser27, the side chain of Gln28 and also an additional interaction with the backbone of Lys13. (B) It was observed in the initial energy minimised structure of CLIC1-C24A that when the side chain of Cys24 was mutated, a void was created that consequently resulted in the neighbouring residues, most notably Lys13, changing their orientation. All images were generated by using Pymol V1.3.

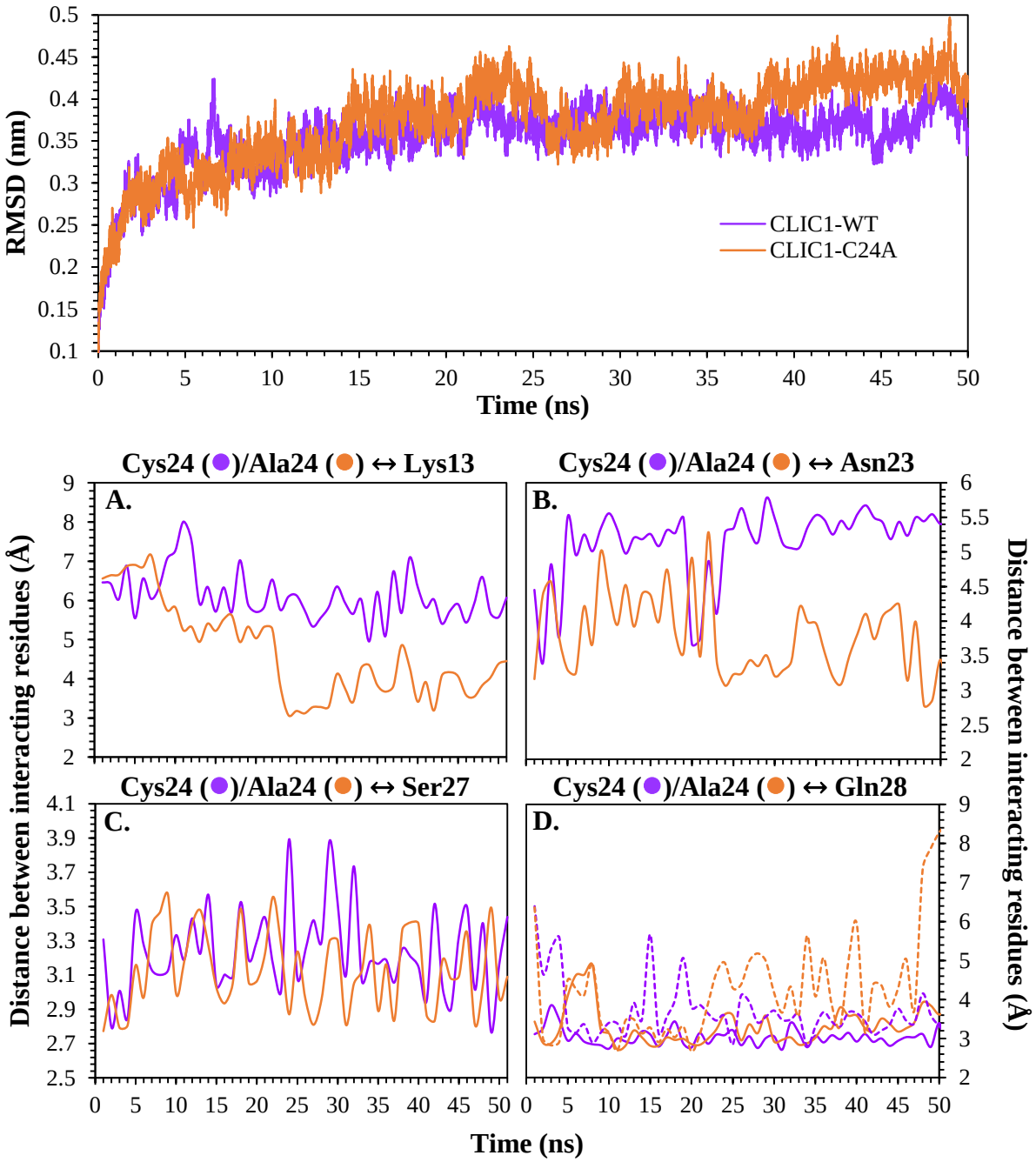


Figure 81. The dynamic nature of the interactions that Cys24 and Ala24 make with their surrounding residues. The change in the distance between the backbone of Ala24 and its neighbouring residues gave insight as to how the C24A mutation influenced the overall dynamics and therefore hydrogen bond network between these residues. (A) There was an overall decrease in the distance between Ala24 and Lys13 that was accompanied by an increase in the number of hydrogen bond being formed. (B-C) There was a decrease in the distance, and hence an increase in the number of hydrogen bonds formed, between the backbone of Ala24 and the backbone of Asn23 and Ser27. (D) During 10-21 ns of the simulation, Ala24 and Gln28 were brought in close proximity to one another and this resulted in the formation of a hydrogen bond between the two residues. However, the overall distance between Ala24 and Gln28 had increased throughout the simulation. Indicated are the interactions that the backbone of Cys24 (●) and Ala24 (●) makes with the backbone (—) and/or the side chain (- - -) of its neighbouring residues. The data for the distance plots were generated by using Chimera 1.11.2.

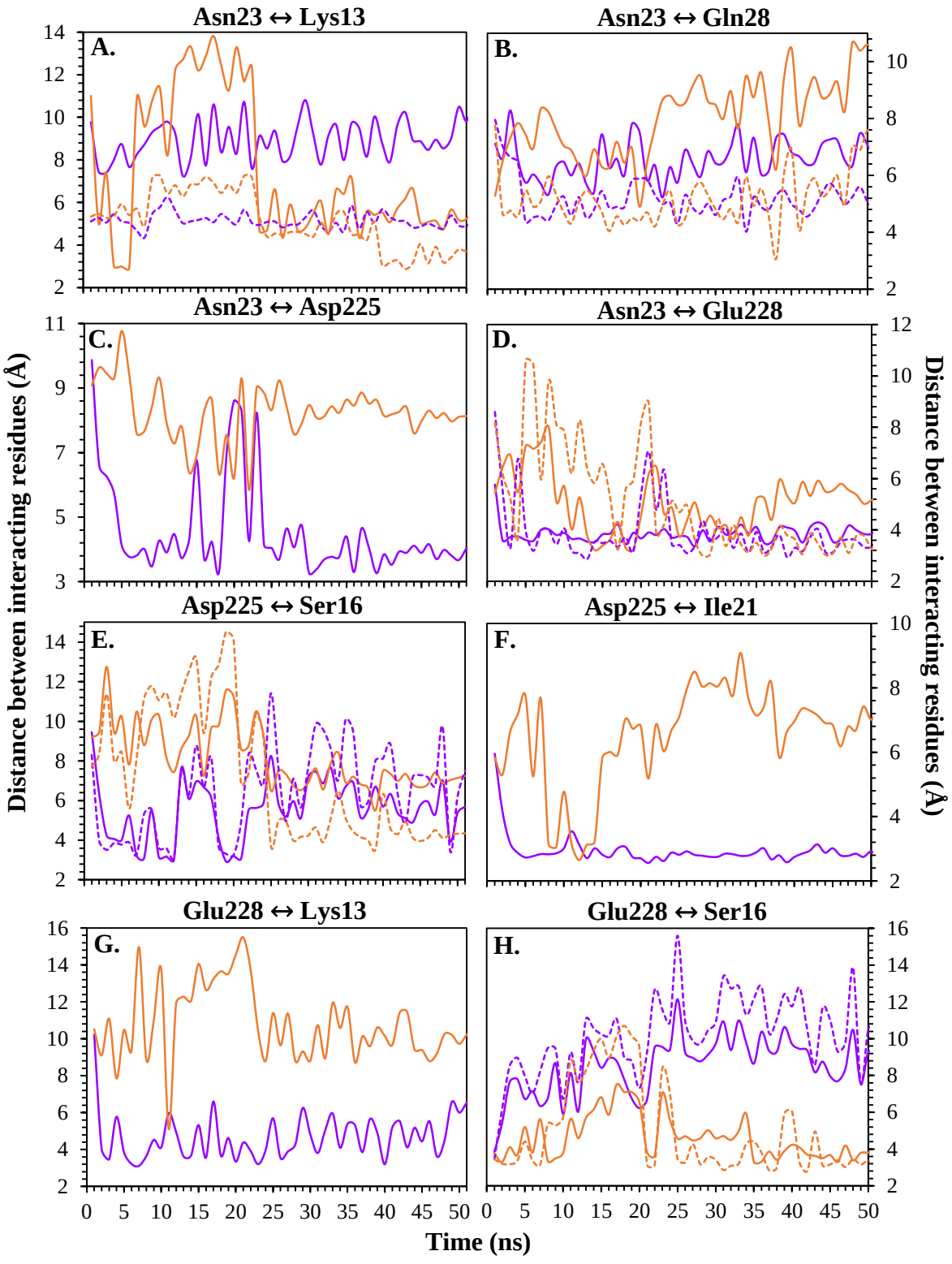


Figure 82. The dynamic nature of the interactions that the neighbouring residues of Cys24 and Ala24 make with their surrounding residues. (A-D) There was an overall increase in the distance between Asn23 and its neighbouring residues in CLIC1-C24A. (E-G) There was an overall increase in the distance between Asp225 and E228 and some of their neighbouring residues. (H) There was a decrease in the distance, and hence an increase in the number of hydrogen bonds formed, between Glu228 and Ser16 in CLIC1-C24A. Indicated are the interactions that the neighbouring residues of Cys24 (●) and Ala24 (●) makes with the backbone (—) and/or the side chain (- - -) of their surrounding residues. The data for the distance plots were generated by using Chimera 1.11.2.

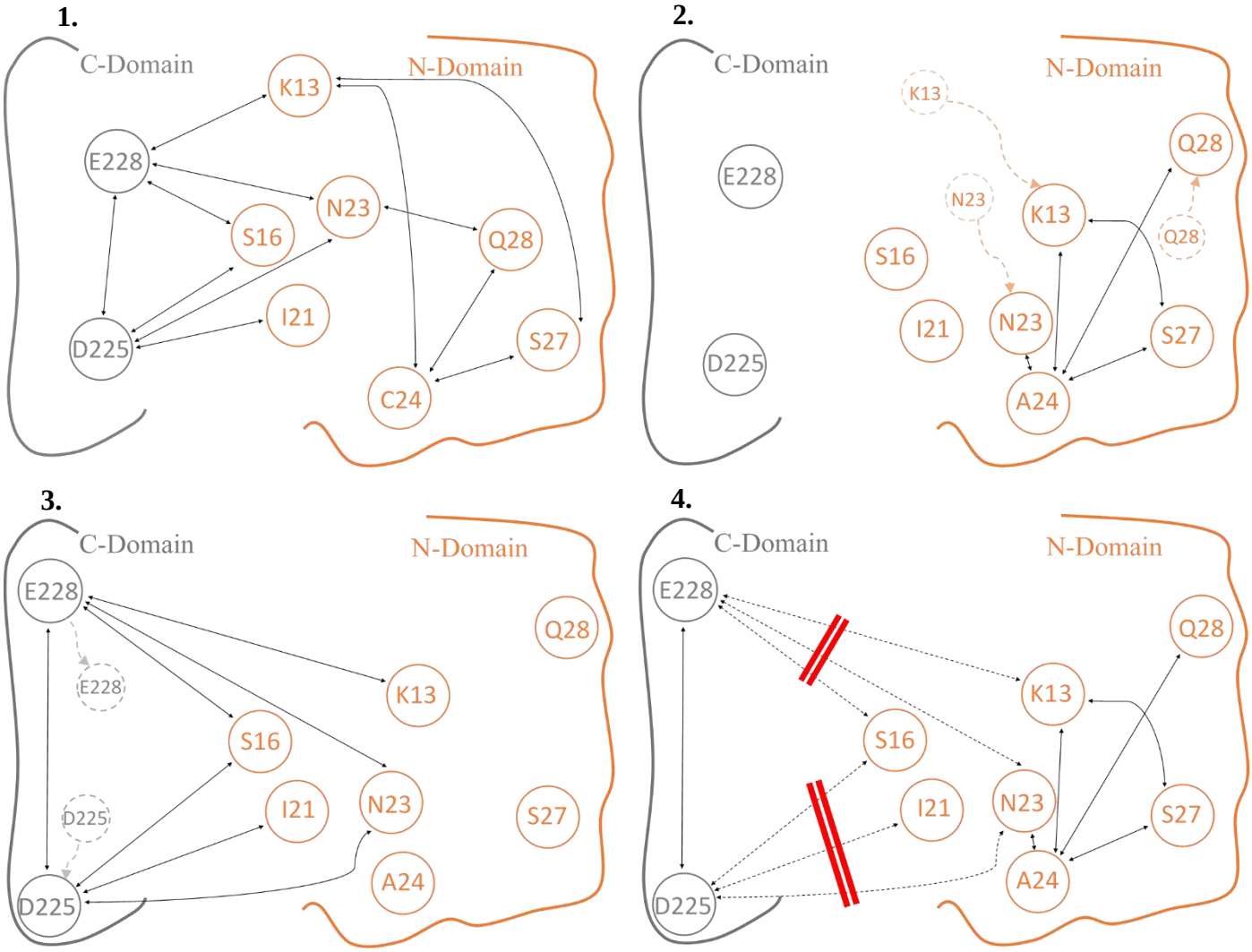


Figure 83. A schematic representing how the C24A mutation destabilises CLIC1 by causing the N-terminal domain to become more dynamic. (1) The backbone of Cys24 is involved in forming interactions with various residues, most notably Lys13, Ser27 and Glu28. (2) When Cys24 was mutated to Ala24 there was a change in the orientation of its surrounding residues. In particular, accompanying the decreased distance between Ala24 and Ser27 was also a decrease in the distance between the backbone of Ala24 and the backbone of Asn23 and Lys13. However, there was an overall increase in the distance between Ala24 and the side chain of Gln28. (3) At the all α-helical C-terminal domain, Asp225 and Glu228 played a key role in stabilising the domain interface considering that Asp225 formed a hydrogen bond with Ser16, Ile21 and Asn23, whereas Glu228 formed a hydrogen bond with Lys13, Ser16 and Asn23. However, in CLIC1-C24A the distance between Asp225, Glu228 and their interacting residues increased and consequently the number of hydrogen bonds formed between these interacting residues decreased. However, in CLIC1-C24A, the decreased distance between Glu228 and Ser16 compensated for the decreased number of hydrogen bonds that would form between Glu228 and Asn23 in CLIC1-WT. (4) Collectively, the C24A mutation changed the dynamics of its neighbouring residues and consequently this induced a cooperative affect that changed the hydrogen bond network and ultimately the dynamics of the N-terminal domain. In turn, this decreased the number of hydrogen bonds which are formed at the domain interface and therefore it resulted in CLIC1-C24A becoming more dynamic and less stable.

3.5.3 Determining the site and stoichiometry for the *S*-nitrosylation of soluble CLIC1

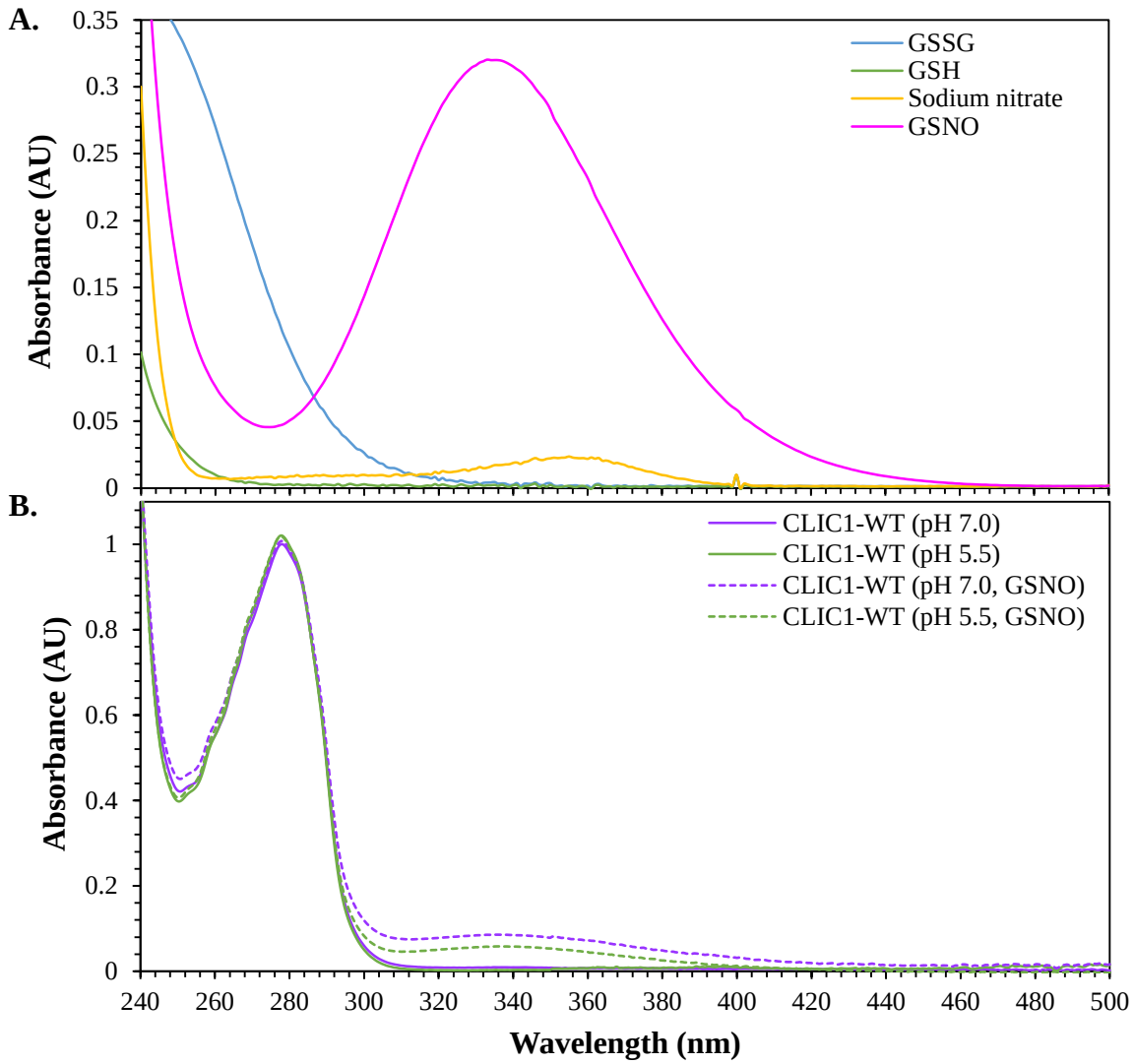
3.5.3.1 Using absorbance spectroscopy to determine the stoichiometry for *S*-nitrosylation

The *S*-nitrosylation of CLIC proteins has previously been reported for CLIC4, whereby Cys234 was reported to be *S*-nitrosylated in response to GSNO, while Cys35 and Cys189 were suggested to act as allosteric regulators and protect the protein from *S*-nitrosylation (Malik *et al*., 2010). Structurally, CLIC1 and CLIC4 are very similar, but in particular it was noted that Cys24 and Cys223 in CLIC1 were structurally aligned with Cys35 and Cys234 in CLIC4, respectively (Figure 17). Therefore, considering that Cys24 is the main reactive cysteine residue within CLIC1 and plays a key role in maintaining its structural stability, as to whether this residue is a target site for *S*-nitrosylation and as to how this PTM would influence the overall structural stability of CLIC1 remained to be investigated.

The reactivity of a cysteine residue is governed by the formation of a negatively charged thiolate anion and therefore its reactivity would be dependent on the pH of the surrounding environment. However, the physical-chemical properties of cysteine residues can often be hard to categorise, but for cysteine residues which are typically solvent accessible, they tend to have an intrinsic pK_a of ~8.6. However, there are many factors which can influence the pK_a value for the thiol group of cysteine, for instance the pK_a value will be lowered by intermolecular interactions that destabilise the thiol group, but stabilise the thiolate ion through a hydrogen bond and/or an electrostatic interaction from a positively charged residue. Whereas, intermolecular interactions that stabilise the thiol group, but destabilise the thiolate ion (for instance by a negatively charged residue) will increase the pK_a value for the thiol group of cysteine (Marino and Gladyshev, 2012; Williams and Rowley, 2016).

Although all previous studies were performed at pH 7.0 it was important to perform the *S*-nitrosylation of CLIC1-WT at both pH 5.5 and pH 7.0 considering that the pH of the former coincides with the acidic environment found at the surface of a membrane (McLaughlin, 1989). Hence, given that the formation of a thiolate ion is pH-dependent, the reactivity of the cysteine residues within CLIC1, and hence the degree of *S*-nitrosyaltion, may differ under these two conditions. Since GSNO has its own characteristic absorbance profile with a maximum absorbance at 335 nm, the stoichiometry of the *S*-nitrosthiol group/s within CLIC1-WT could be determined (section 2.2.11.2). It was observed that at pH 5.5 and pH 7.0, the *S*-nitrosthiol stoichiometry was n = 1.10 (~1) and n = 1.55 (~2), respectively (Figure 84), which suggested

that at pH 5.5 one cysteine residue became *S*-nitrosylated, while at pH 7.0 two cysteine residues became *S*-nitrosylated. Therefore, in order to predict which cysteine residues could have become *S*-nitrosylated, the Group-Based Prediction System-*S*-nitrosothiol (GPS-SNO) 1.0 software was used (Xue *et al.*, 2010). It was predicted that both Cys24 and Cys223 could become *S*-nitrosylated within CLIC1-WT (Figure 85) and therefore this could account for the stoichiometry of n ~2 which was predicted by UV-Vis absorbance spectroscopy (Figure 83). This was interesting considering that Cys234, which becomes *S*-nitrosylated in CLIC4, is structurally homologous to Cys223, therefore supporting the notion that Cys223 could possibly be a target site for *S*-nitrosylation in CLIC1. Therefore, LC-MS/MS (section 2.2.11.3) was performed in order to conclude which cysteine residues become *S*-nitrosylated in CLIC1-WT at both pH 5.5 and 7.0.



Condition	CLIC1-WT (pH 5.5)	CLIC1-WT (pH 7.0)	CLIC1-WT (pH 5.5, GSNO)	CLIC1-WT (pH 7.0, GSNO)
Stoichiometry	0.01	0.10	1.10	1.55

Figure 84. The stoichiometry of the *S*-nitrosthiol group/s present within CLIC1-WT. (A) The *S*-nitrosylating agent GSNO has its own characteristic absorbance profile with a maximum absorbance at 335 nm. However, sodium nitrate, a key reactant during the synthesis of GSNO, also has its own characteristic absorbance profile with a maximum absorbance at 355 nm. Therefore, although this resulted in the concentration of GSNO being over-estimated, this was not an issue considering that GSNO was used in excess in order to ensure the complete *S*-nitrosylation of CLIC1-WT. Importantly, however, this meant that the GSNO stock also contained nitrate, an endogenous compound that can often lead to the nitration of tyrosine residues. (B) The degree of *S*-nitrosylation was estimated by taking a ratio of the concentration for the *S*-nitrosothiol group/s to the concentration of CLIC1-WT. It was observed that at pH 5.5 and pH 7.0, the *S*-nitrosthiol stoichiometry was $n = 1.10$ (~1) and $n = 1.55$ (~2), respectively, therefore suggesting that at pH 5.5 one cysteine residue became *S*-nitrosylated, while at pH 7.0 two cysteine residues became *S*-nitrosylated. *See section 2.2.11.1-2.2.11.2 for the experimental details pertaining to this method.*

Position	Peptide sequence	Score	Threshold cut-off	Cluster
A. *S*-nitrosylation of cysteine residues				
	GSNO + Cysteine → Cysteine-SNO + GSH			
24	^{17}DGAKIGN**C**PFSQRLF^{31}	0.814	0.797	A
223	^{216}REEFAST**C**PDDEEIE^{230}	1.798	0.797	A
B. Nitration of tyrosine residues				
	Tyrosine + Nitrite → Tyrosine-NO_2			
233	^{226}DEEIELA**Y**EQVAKAL^{240}	2.99	1.16	D

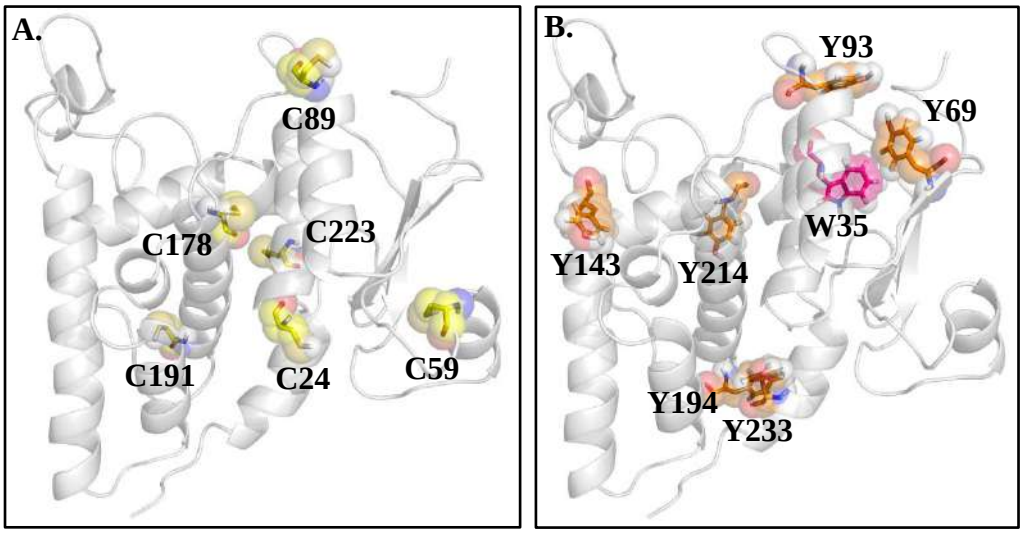


Figure 85. Predicting which residues could become *S*-nitrosylated and/or nitrated in to CLIC1-WT. (A) The GPS-SNO 1.0 software (http://sno.biocuckoo.org/) was used to predict which cysteine residues in CLIC-WT could become *S*-nitrosylated. It was predicted that within CLIC-WT both Cys24 and Cys223 could become *S*-nitrosylated and therefore this could account for the stoichiometry of n~2 which was predicted by UV-Vis absorbance spectroscopy. This was interesting considering that Cys234, which becomes *S*-nitrosylated in CLIC4, is structurally homologous to Cys223, therefore supporting the notion that Cys223 could be a target site for *S*-nitrosylation. (B) The GPS-YNO_2 1.0 software (http://sno.biocuckoo.org/) was then used to predict which tyrosine residues in CLIC-WT could become nitrated. It was predicted that only Tyr233 could become nitrated, but this was also interesting considering that Tyr233 is a neighbouring residue to Cys24. All images were generated by using Pymol V1.3.

3.5.3.2 Determining the *S*-nitrosylation site by means of LC-MS/MS

The addition of the N=O moiety from GSNO to the thiol group of a cysteine residue would be expected to be accompanied by a 30 Da increase in the mass of a cysteine residue. The LC-MS/MS spectra for CLIC1-WT under both pH conditions indicated that the cysteine residue from the ^{20}KIGNCPFSQRL^{30} peptide was the only cysteine residue within CLIC1-WT to be accompanied by a 30 Da increase in mass. Furthermore, the IDOTP scores for this particular peptide was ≥ 0.97 for each individual sample (Figure S9), therefore indicating that the observed isotope distribution correlated with the predicted isotope distribution and that the LC-MS/MS spectra obtained were reliable for downstream analysis. As a result, it was confirmed that Cys24 was a target site for *S*-nitrosylation by GSNO considering that the ratio between the integrated area for the modified and unmodified states of this peptide were ~2 at pH 5.5 (Figure 86; Table 8) and ~3 at pH 7.0 (Figure 87; Table 8), respectively.

Although *S*-nitrosothiols have been implicated in a variety of biological functions, due to the labile and therefore unstable nature of the N=O moiety, this has made detecting and studying *S*-nitrosylated protein molecules difficult (Hao *et al.*, 2006). For instance, when *S*-nitrosylated proteins are put into an ionisation state during MS, the nitroso functional group could become short-lived and result in the homolytic cleavage of the S-N bond to produce NO and a thiyl radical (Broniowska *et al.*, 2013). However, it has been reported that by initially labelling free cysteines with NEM, followed by reducing the *S*-nitrosothiol bonds and labelling the cysteines with IAM, this could ensure that *S*-nitrosylated cysteine residues are detected more efficiently considering that these alkylating agents are more stable (Klapoetke and Xie, 2017). Therefore, although Cys233 was predicted to become *S*-nitrosylated (Figure 85) and could account for the stoichiometry of n ~2 as predicted by UV-Vis absorbance spectroscopy (Figure 84), it was possible that the *S*-nitrosylation of this residue could not be detected due to the labile nature of the nitroso functional group

However, when observing the absorbance spectra of the chemical components used in the synthesis of GSNO, it was noted that sodium nitrate had a characteristic absorbance profile with a maximum absorbance at 355 nm (Figure 84). Therefore, considering that sodium nitrate also had an absorbance at 335 nm, this resulted in the concentration of GSNO being over-estimated, but this was not an issue considering that GSNO was used in excess in order to ensure the complete *S*-nitrosylation of CLIC1-WT. Importantly, however, this meant that

the GSNO stock also contained nitrate, an endogenous compound that can often leads to the nitration of tyrosine residues. Using the GPS-tyrosine nitration (GPS-YNO_2) 1.0 software (Liu *et al.*, 2011), it was predicted that only Tyr233 could become nitrated, which was interesting considering that Tyr233 is a neighbouring residue to Cys24 (Figure 85).

During the nitration of tyrosine one of the two aromatic hydrogens which are ortho with respect to the hydroxyl group become replaced by a nitro group, and in some instances a nitrated tyrosine can further react and replace the second hydrogen atom with another nitro group. However, although much effort has been made to determine the factors that promote the selective nitration of tyrosine, owing to the data for this PTM often being ambiguous, a universal and reliable model that can predict the mechanism of tyrosine nitration has unfortunately not been developed (Bayden *et al.*, 2011). Nevertheless, it was reported by Elfering *et al*. (2004) that a generalised sequence surrounding nitrated tyrosine residues was H-X-[DE]-H-X (2,3)-H-H-X(2,4)-Y, whereby X represents any amino acid, H represents a hydrophobic residue and Y represents the target tyrosine residue. Furthermore, it has been observed that tyrosine for instance is more likely to become nitrated when it is present on a loop, located near a transition metal co-factor and/or when it is in close proximity to a basic or acidic residue. The role of cysteine in the nitration of tyrosine on the other hand is controversial seeing that some evidence has suggested that by competing for the nitrating species, cysteine residues would impede the nitration of tyrosine, whereas some findings have suggested that cysteine residues have facilitated tyrosine nitration (Radi, 2013).

Therefore, considering that Tyr233 is in close proximity to Cys24, is at the start of a loop region and is surrounded by Asp230 and Asp234, it is possible that Tyr233 could be a target site for nitration. However, no nitrated tyrosine residues were detected by LC-MS/MS, and similar to the nitroso group, this possibly could be attributed to the labile nature of the nitro group. Therefore, it still remains unclear if the stoichiometry of n ~2 at pH 7.0 (Figure 84) was due to the *S*-nitrosylation of Cys233 or due to the nitration of Tyr233. Nevertheless, it was apparent that at both pH 5.5 (Figure 86, Table 8) and pH 7.0 (Figure 87, Table 8) Cys24 was a target site for *S*-nitrosylation.

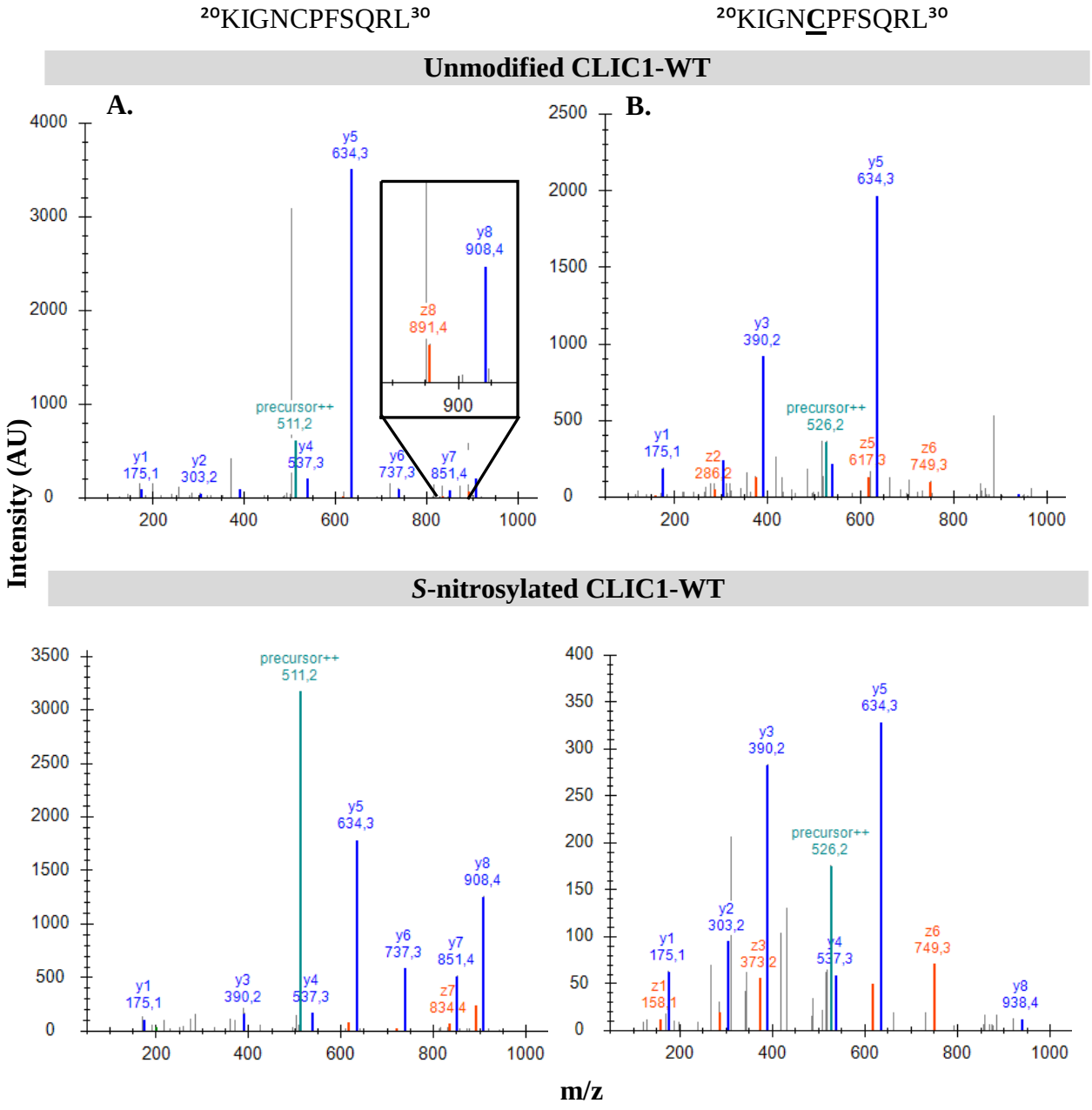


Figure 86. The product ion MS/MS spectra of a peptide from CLIC1-WT which was identified to become *S*-nitrosylated at pH 5.5. The addition of the N=O moiety from GSNO to the thiol group of a cysteine residue would be expected to be accompanied by a 30 Da increase in the mass of a cysteine residue. In contrast to the (A) unmodified conformation of CLIC1-WT, a single positively charged peptide (²⁰KIGNCPFSQRL³⁰) was identified to have a 30 Da mass shift at Cys24 in the (B) *S*-nitrosylated conformation of CLIC1-WT. Indicated are the y and z ions and the m/z for each of these precursor ions. All data was initially analysed by using the ProteinPilot™ 5.0.1 software in order to compare the data with a common database that contained the sequence of CLIC1-WT as well as a list of sequences from common contaminating proteins. Thereafter, all data was refined and integrated by using the Skyline-daily 3.7.1.11357 MacCoss Lab software ©.

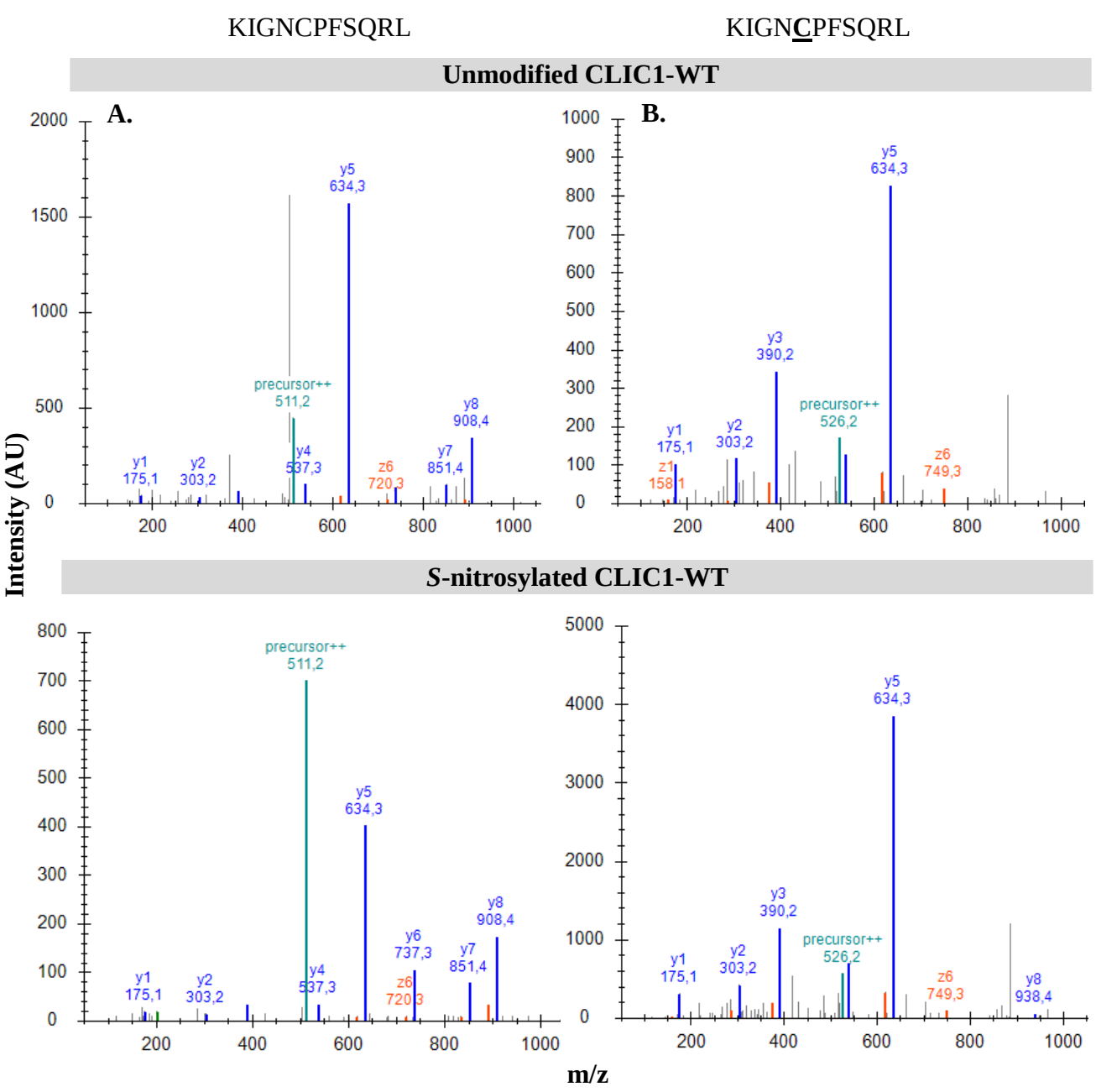


Figure 87. The product ion MS/MS spectra of a peptide from CLIC1-WT which was identified to become *S*-nitrosylated at pH 7.0. The addition of the N=O moiety from GSNO to the thiol group of a cysteine residue would be expected to be accompanied by a 30 Da increase in the mass of a cysteine residue. In contrast to the (A) unmodified conformation of CLIC1-WT, a single positively charged peptide (^{20}KIGNCPFSQRL^{30}) was identified to have a 30 Da mass shift at Cys24 in the (B) *S*-nitrosylated conformation of CLIC1-WT. Indicated are the y and z ions and the m/z for each of these precursor ions. All data was initially analysed by using the ProteinPilot™ 5.0.1 software in order to compare the data with a common database that contained the sequence of CLIC1-WT as well as a list of sequences from common contaminating proteins. Thereafter, all data was refined and integrated by using the Skyline-daily 3.7.1.11357 MacCoss Lab software ©.

Table 8: Normalised integrated area for the three isotopic peaks from the LC-MS/MS spectra for the *S*-nitrosylated peptide of CLIC1-WT under both pH conditions.

	A. Peptide area (^{20}KIGNCPFSQRL^{30})		B. Peptide area (^{20}KIGN**C**PFSQRL^{30})		Normalised peptide area (B/A)	
Unmodified CLIC1-WT						
Samples	pH 7.0	pH 5.5	pH 7.0	pH 5.5	pH 7.0	pH 5.5
1.1	111 120	53 124	220 490	72 192	2.0	1.4
1.2	498 600	82 795	131 260	88 365	0.3	1.1
1.3	216 300	70 849	218 540	91 736	1.0	1.3
Average	275 340	68 923	190 097	84 098	1.1	1.2
***S*-nitrosylated CLIC1-WT**						
Samples	pH 7.0	pH 5.5	pH 7.0	pH 5.5	pH 7.0	pH 5.5
1.1	197 360	81 762	529 060	128 990	2.7	1.6
1.2	266 020	85 331	832 070	141 900	3.1	1.7
1.3	185 280	84 743	761 770	239 500	4.1	2.8
Average	216 220	83 945	707 633	170 130	3.3	2.0

3.5.3.3 Evaluating the secondary and tertiary structure of soluble *S*-nitrosylated CLIC1

The CD spectra for both the unmodified and *S*-nitrosylated conformations of CLIC1-WT under both pH conditions indicated two prominent troughs at 208 nm and 222 nm (Figure 88), therefore indicating that CLIC1-WT was pre-dominantly α-helical under these conditions. This correlated well with the crystal structure of CLIC1-WT (PDB: 1K0M) which showed that the α-helical and beta-sheet content was 47% and 8%, respectively. However, although the CD spectra for the unmodified conformation of CLIC1-WT indicated that there was no change in its secondary structural content under both pH conditions, when monitoring the ellipticity at 222 nm, it was noted that the secondary structural content for *S*-nitrosylated CLIC1-WT decreased by 6.5% and 11.2% at pH 5.5 and pH 7.0, respectively (Figure 88). Therefore, similar to the C24A mutation which resulted in a decrease in the secondary structural content and the thermal stability of CLIC1 (Figure 36), it appeared that the *S*-nitrosylation of Cys24 influenced the secondary structure of CLIC1-WT and therefore it may have also influenced the overall structural stability of CLIC1-WT.

On the other hand, the fluorescence emission spectra for both the unmodified and *S*-nitrosylated conformations of CLIC1-WT at both pH conditions indicated a fluorescence emission maxima at ~342 nm (Figure 88). This emission maxima reflected the phenomena that when tryptophan is in an environment surrounded by water it has a maximum fluorescence emission near 350 nm (Lakowicz, 2006), therefore indicating that Trp35 was predominantly solvent exposed as seen in the crystal structure of CLIC1-WT (PDB: 1K0M) were Trp35 has a total ASA of 89.28 Å^2 (Figure 35). Furthermore, a change in pH had a negligible change in the local tertiary structure surrounding Trp35 considering that there was only a 2.3% difference in the maximum fluorescence emission intensity for unmodified CLIC1-WT under both pH conditions. However, the *S*-nitrosylation of CLIC1-WT caused a decrease in the maximum fluorescence emission intensity of CLIC1-WT, but to a greater extent at pH 7.0 (Figure 88). This decrease in the maximum fluorescence emission intensity did not necessarily indicate any significant structural changes around Trp35, instead this decrease was most likely attributed to the primary inner-filter effect (Fonin *et al*., 2014) considering that the maximum emission and absorbance wavelength for CLIC1-WT and GSNO was 342 nm (Figure 34) and 335 nm (Figure 84), respectively. As a result, considering that the *S*-nitrosthiol stoichiometry was n ~2 at pH 7.0 (Figure 84), this accounted for the additional quenching seen at pH 7.0. Therefore, once again, similar to the C24A mutant (Figure 34 and 37), it appeared that the *S*-nitrosylation and change in pH had no influence on the local tertiary structure at Trp35 of CLIC1-WT.

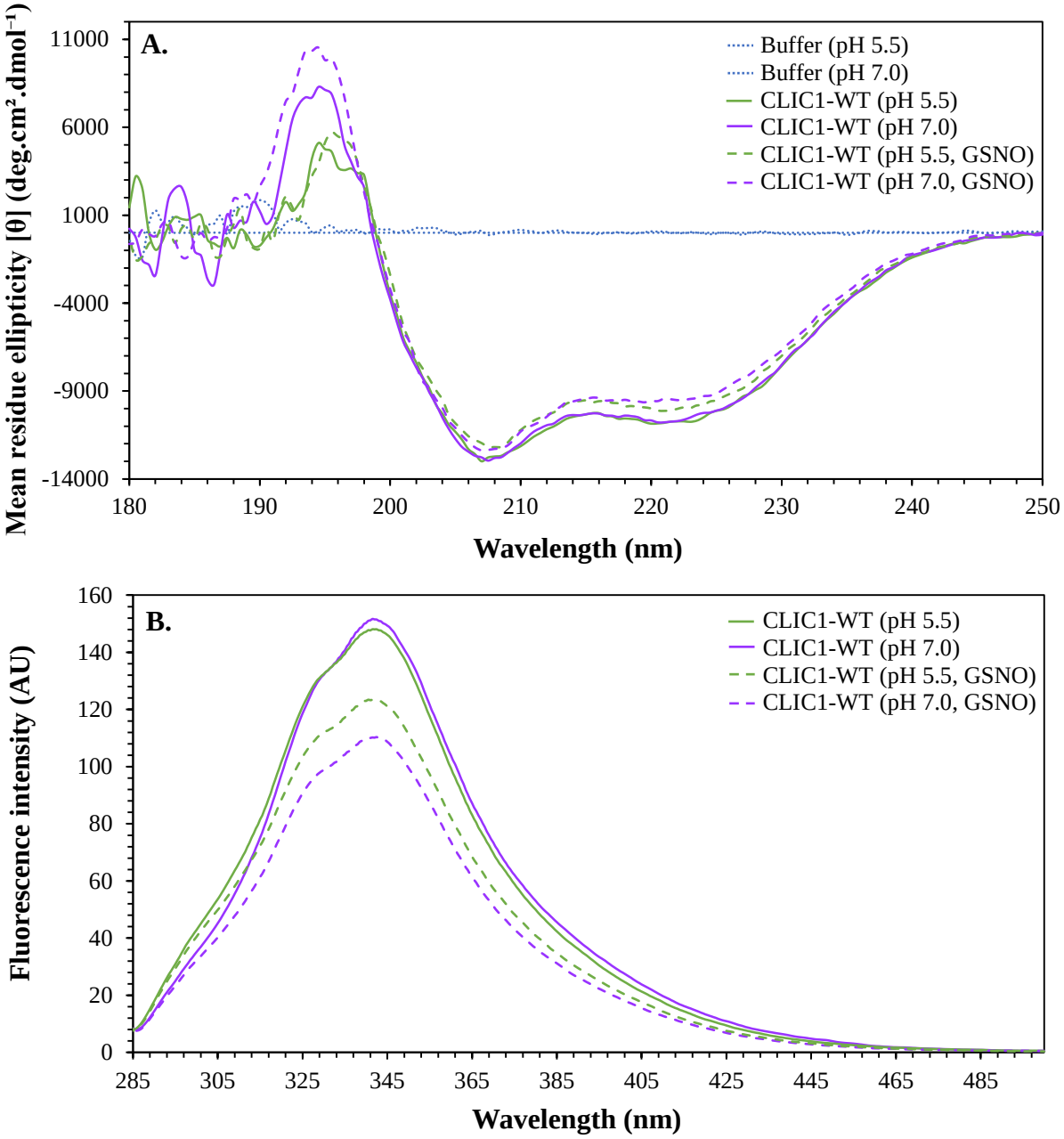


Figure 88. Verifying the secondary and tertiary structure of *S*-nitrosylated CLIC1-WT. (A) The secondary structure for unmodified and *S*-nitrosylated CLIC1-WT under both pH conditions were determined at 20 °C by far-UV CD spectropolarimetry while using 2 µM of the target protein in storage buffer. All spectra were measured in triplicate and the resultant CD ellipticity signal in millidegrees was converted to the [Θ]. The CD spectra for unmodified and *S*-nitrosylated CLIC1-WT indicated two prominent troughs at 208 nm and 222 nm, a profile characteristic of a predominantly α-helical protein. (B) Intrinsic tryptophan fluorescence was monitored at 20 ^{0}C with an excitation wavelength of 280 nm while using 2 µM of the target protein in storage buffer. The fluorescence spectra indicated that Trp35 was predominately solvent exposed and therefore a change in pH and the *S*-nitrosylation of CLIC1-WT had a negligible change in the local tertiary structure surrounding Trp35. *See section 2.2.7 for the experimental details pertaining to this method.*

Chapter 4: Conclusion

The CLIC1 protein is unique in that it can reversibly transit between a soluble and membrane-bound conformation, where the ion channel activity of the latter becomes inhibited by IAA-94. Therefore, it remained uncertain if the constraint in the Cl^- conductance of CLIC1 was due to IAA-94 binding to its soluble conformation in order to prevent its structural transition into a membrane competent state and/or IAA-94 binding to the membrane conformation of CLIC1. In this study, it was shown that IAA-94 did not bind to the soluble conformation of CLIC1, but instead it was shown to interact with hGSTP1-1 and the soluble conformation of EcSspA, the structural homologs of CLIC1. Interestingly, although IAA-94 is a structural analogue of EAA, the latter was shown to bind to both CLIC1 and hGSTP1-1, but not to EcSspA. Furthermore, unlike EAA, IAA-94 did inhibit the CDNB enzyme activity of hGSTP1-1, therefore indicating that they occupied different sites within hGSTP1-1.

Based on structural sequence alignment, it was apparent that the residues which coordinated the interaction between EAA and hGSTP1-1 were primarily conserved within CLIC1, therefore possibly accounting for why EcSspA did not interact with EAA. However, the residues from EcSspA, but not CLIC1, were structurally conserved with the residues from the MES, or otherwise putative IAA-94-binding site of hGSTP1-1, therefore possibly accounting for why CLIC1 did not interact with IAA-94. In fact, the crystal structure of hGSTP1-1 in complex with IAA-94 suggested that IAA-94 did bind at the MES binding site, but due to crystal contacts only the acetate moiety of IAA-94 could be solved. These findings therefore supported previous studies which suggested that the MES binding site is also the L-site of hGSTP1-1.

The fact that EAA can bind CLIC1 was significant considering that EAA would inhibit the enzyme function of CLIC1 by masking the thiol group of Cys24. In fact, Cys24 was shown to be the main reactive cysteine residue in soluble CLIC1 and in turn it was also shown to play a key role in maintaining the structural stability of the domain interface by regulating the dynamics of the N-terminal domain. Furthermore, Cys24 was shown to be a target site for *S*-nitrosylation and similar to the C24A mutation, it appeared that this PTM could influence the structural stability of soluble CLIC1. Therefore, although soluble CLIC1 did not bind to IAA-94, given that it could bind to EAA, the 2, 3-dichlorophenoxyacetate moiety of EAA could serve as a pharmacophore for the rational drug design of novel inhibitors that can effectively bind to CLIC1, mask the reactivity of Cys24 and stabilise the soluble conformation of CLIC1 in order to regulate its dimorphic nature and attenuate its structural transition into to a membrane-bound conformation.

Chapter 5: References

Abdalla, A.M., Bruns, C.M., Tainer, J.A., Mannervik, B. and Stenberg, G. (2002). Design of a monomeric human glutathione transferase GSTP1, a structurally stable but catalytically inactive protein. *Protein Eng. Des. Sel.*, 15: 827-834.

Abraham, M.J., van der Spoel, D., Lindahl, E. and Hess, B. (2017). GROMACS user manual version 2016.3. The GROMACS development team, www.gromacs.org.

Achilonu, I., Fanucchi, S., Cross, M., Fernandes, M. and Dirr, H.W. (2012). Role of individual histidines in the pH-dependent global stability of human chloride intracellular channel 1. *Biochem*., **51**: 995-1004.

Acikara, O.B. (2013). Ion-exchange chromatography and its applications, column chromatography. *InTech.*, DOI: 10.5772/55744.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.W., Jain, S., Kapral, G.J., Kunstleve, R.W.G., McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. and Zwart, P.H. *Methods*, **55**: 94-106.

Al-Awqati, Q. (1995). Chloride channels of intracellular organelles. *Curr. Opin. Struct. Biol*., **5**: 504-508.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (2008). *Molecular biology of the Cell*. 5th ed. New York (USA): Garland Science.

Andrew, P.J. and Meyer, B. (1999). Enzymatic function of nitric oxide synthases. *Cardiovasc. Res*., **43**: 521-531.

Atkins, P. and Paula, J. (2010). *Atkins' Physical Chemistry*. England (UK): Oxford University Press.

Averaimo, S., Abeti, R., Savalli, N., Brown, L.J., Curmi, P.M.G., Breit, S.N. and Mazzanti, M. (2013). Point mutations in the transmembrane region of the CLIC1 ion channel selectively modify its biophysical properties. *PLoS One*, DOI: 10.1371/journal.pone.0074523.

Barelli, S., Canellini, G., Thadikkaran, L., Crettaz, D., Quadroni, M., Rossier, J.S., Tissot, J.D. and Lion, N. (2008). Oxidation of protein: basic principles and perspectives for blood proteomics. *Proteomics Clin. Appl*., **2**: 142-157.

Bartberger, M.D., Houk K.N., Powell S.C., Mannion, J.D., Lo, K.Y., Stamler, J. and Toone, E.J. (2000). Theory, spectroscopy and crystallographic analysis of *S*-nitrosothiols: conformational distribution dictates spectroscopic behaviour. *J. Am. Chem. Soc*., **122**: 5889-5890.

Bayden, A.S., Yakovlev, V.A., Graves, P.R., Mikkelsen, R.B. and Kellogg, G.E. (2011). Factors influencing protein tyrosine nitration: structure-based predictive models. *Free. Radic. Biol. Med.,* **50**: 749-762.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **25**: 535-542.

Berryman, M. and Bretscher, A. (2000). Identification of a novel member of the chloride intracellular channel gene family (CLIC5) that associates with the actin cytoskeleton of placental microvilli. *Mol. Biol. Cell*, **11**: 1509-1521.

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J. and Schäfer. (2009). Immobilised-metal affinity chromatography (IMAC): a review. *Methods Enzymol*., **463**: 439-472.

Board, P.G., Coggan, M., Chelvanayagam, G., Easteal, S., Jermiin, L.S., Schulte, G.K., Danley, D.E., Hoth, L.R., Griffor, M.C., Kamath, A.V., Rosner, M.H., Chrunyk, B.A., Perregaux, D.E., Gabel, C.A., Geoghehan, K.F. and Pandit, J. (2000). Identification, characterisation and crystal structure of the omega class glutathione transferases. *J. Biol. Chem.*, **275**: 24798-24806.

Board, P.G., Coggan, M., Watson, S., Gage, P.W. and Dulhunty, A.F. (2004). CLIC-2 modulates cardiac ryanodine receptor Ca^{2+} release channels. *Int. J. Biochem. Cell Biol.*, **36**: 1599-1612.

Brandish, P.E., Buechler, W. and Marletta, M.A. (1998). Regeneration of the ferrous heme of soluble guanylate cyclase from the nitric oxide complex: acceleration by thiols and oxyhemoglobin. *Biochem.*, **37**: 16898-16907.

Broniowska, K.A. and Hogg, N. (2012). The chemical biology of *S*-nitrosothiols. *Antioxid. Redox Sign.*, **17**: 969-980.

Broniowska, K.A., Diers, A.R. and Hogg, N. (2013). *S*-nitrosoglutathione. *Biochim. Biophys. Acta*, **1830**: 3173-3181.

Bronowska, A.K. (2011). Thermodynamics of ligand-protein interactions: implications for molecular design. Thermodynamics - interaction studies - solids, liquids and gases. Available from: http://www.intechopen.com/books/thermodynamicsinteraction-studies-solids-liquids-andgases/thermodynamics-of-ligand-protein-interactions-implications-formolecular-design.

Brown, A. (2009). Analysis of cooperativity by isothermal titration calorimetry. *Int. J. Mol. Sci.*, DOI: 10.3390/ijms10083457.

Brown, T.L., LeMay, H.E., Bursten, B.E., Murphy, C.J. and Woodward, P.M. (2012). *Chemistry: The Central Science*. 12th ed. New Jersey (USA): Pearson Prentice Hall.

Cameron, A.D., Sinning, I., L'Hermite, G., Olin, B., Board, P.G., Mannervik, B. and Jones, T.A. (1995). Structural analysis of human alpha-class glutathione transferase A1-1 in the apo-form and in complexes with ethacrynic acid and its glutathione conjugate. *Struct.*, **3**: 717-727.

Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S.M. and Savidge, T.C. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Sci. Adv.*, **2**: 1-16.

Chen, V.B., Arendall III, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallogr. D. Biol. Crystallogr.*, **66**: 12-21.

Chilom, C.G., Craescu, C.T. and Popescu, A.I. (2004). Parameters of interaction between proteins and their specific ligands, deduced by isothermal titration calorimetry. *Rom. Journ. Phys.*, **51**: 443-457.

Chodera, J.D. and Mobley, D. (2013). Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys.*, **42**: 121-142.

Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta. Crystallogr. D. Biol. Crystallogr.*, **50**: 760-763.

Cooper, A. (2010). Protein heat capacity: an anomaly that maybe never was. *J. Phys. Chem. Lett.*, **1**: 3298-3304.

Cromer, B.A., Morton, C.J., Board, P.G., and Parker, M.W. (2002). From glutathione transferase to pore in a CLIC. *Eur. Biophys. J.*, **31**: 356-364.

Cross, M., Fernandes, M., Dirr, H. and Fanucchi, S. (2015). Glutamate 85 and glutamate 228 contribute to the pH-responses of the soluble form of chloride intracellular channel 1. *Mol. Cell Biochem.*, **398**: 83-93.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall III, W.B., Snoeyink, J., Richardson, J.S. and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**: 375-383.

Dirr, H.W., Little, T., Kuhnert, D.C. and Sayed, Y. (2005). A conserved N-capping motif contributes significantly to the stabilisation and dynamics of the C-terminal region of class alpha glutathione *S*-transferases. *J. Biol. Chem.*, **280**: 19480-19487.

Dobson, C.M. (2004). Principles of protein folding, misfolding and aggregation. *Sem. Cell Dev. Biol.*, **15**: 3-16.

Dobson, C.M. and Karplus, M. (1999). The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.*, **9**: 92-101.

Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H. and Shaw, D.E. (2012). Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.*, **41**: 429-452.

Du, X., Li, Y., Xia, Y., Ai, S., Liang, J., Sang, P., Ji, X. and Liu, S. (2016). Insights into protein-ligand interactions-mechanisms, models and methods. *Int. J. Mol. Sci.*, **17**: 1-34.

Duan, G. and Walther, D. (2014). The roles of post-translational modifications in the context of protein interaction networks. *PLOS Comput. Biol*., DOI: 10.1371/JOURNAL.PCBI.1004049.

Duff, M.R. and Howell, E.E. (2015). Thermodynamics and solvent linkage of macromolecule-ligand interactions. *Methods*, **1**: 51-60.

Dulhunty, A., Gage, P., Curtis, S., Chelvanayagam, G. and Board, P. (2001). The glutathione transferase structural family includes a nuclear chloride channel and a ryanodine receptor calcium release channel modulator. *J. Biol. Chem.*, **276**: 3319-3323.

Duncan, R.R., Westwood, P.K., Boyd, A. and Ashley, R.H. (1997). Rat brain p64H1, expression of a new member of the p64 chloride channel protein family in endoplasmic reticulum. *J. Biol. Chem.*, **272**: 23880-23886.

Dutta, A.K., Rösgen, J. and Rajarathnam, K. (2015). Using isothermal titration calorimetry to determine thermodynamic parameters of protein-glycosaminoglycan interactions. *Methods Mol. Biol.*, **1229**: 315-324.

Edwards, J.C. (2006). The CLIC1 chloride channel is regulated by the cystic fibrosis transmembrane conductance regulator when expressed in *Xenopus* Oocytes. *J. Membr. Biol.*, **213**: 39-46.

Eftink, M.R. (2010). The use of fluorescence methods to monitor unfolding transitions in proteins. *Biophys. J.*, **66**: 482-501.

Elfering, S.L., Haynes, V.L., Traaseth, N.J., Ettl, A. and Giulivi, C. (2004). Aspects, mechanism and biological relevance of mitochondrial protein nitration sustained by mitochondrial nitric oxide synthase. *Am. J. Physiol. Heart Circ. Physiol*., **286**: H22–H29.

Emsley, P. and Cowtan, K. (2004). COOT: model-building tools for molecular graphics. *Acta. Crystallogr. D. Biol. Crystallogr.*, **60**: 2126-2132.

England, J.L. and Haran, G. (2011). Role of solvation effects in protein denaturation: from thermodynamics to single molecules and back. *Annu. Rev. Phys. Chem*., **62**: 257-277.

Evans, P. and McCoy, A. (2007). An introduction to molecular replacement. *Biol. Crystallogr*., DOI: 10.1107/S0907444907051554.

Fanucchi, S., Adamson, R.J. and Dirr, H.W. (2008). Formation of an unfolding intermediate state of soluble chloride intracellular channel protein CLIC1 at acidic pH. *Biochem.*, **47**: 11674-11681.

Fernandez-Salas, E., Sagar, M., Cheng, C., Yuspa, S.H. and Weinberg, W.C. (1999). p53 and tumor necrosis factor alpha regulate the expression of a mitochondrial chloride channel protein. *J. Biol. Chem*., **274**: 36488-36497.

Filloux, A. and Ramos, J.L. (2014). *Pseudomonas methods and protocols*. New York (USA): Humana Press.

Fonin, A.V., Sulatskaya, A.I., Kuznetsova, I.M. and Turoverov, K.K. (2014). Fluorescence of dyes in solutions with high absorbance. Inner filter effect correction. *PLoS ONE*, DOI: 10.1371/journal.pone.0103878.

Ford, E., Hughes, M.N. and Wardman, P. (2002). Kinetics of the reactions of nitrogen dioxide with glutathione, cysteine and uric acid at physiological pH. *Free Radic. Biol. Med.*, **32**: 1314-1323.

Förstermann, U. and Sessa, W.C. (2012). Nitric oxide synthases: regulation and function. *Euro. Heart J.*, **33**: 829-837.

Freyer, M.W. and Edwin, A.L. (2008). Isothermal titration calorimetry: experimental design, data analysis and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol.*, **84**: 79-113.

Gagnon, L.H., Longo-Guess, C.M., Berryman, M., Shin, J.B., Saylor, K.W., Yu, H., Gillespie, P.G. and Johnson, K.R. (2006). The chloride intracellular channel protein CLIC5 is expressed at high levels in hair stereocilia and is essential for normal inner ear function. *J. Neurosci.*, **26**: 10188-10198.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Baroch, A. (2005*). The Proteomics Protocols Handbook.* New York (USA): Humana Press.

Gaston, B.M., Carver, J., Doctor, A. and Palmer, L.A. (2003). *S*-nitrosylation signalling in cell biology. *Mol. Interv.*, **3**: 253-263.

Gasymov, O.K., Abduragimov, A.R. and Glasgow, B.J. (2014). Probing tertiary structure of proteins using single Trp mutations with circular dichroism at low temperature. *J. Phys. Chem.*, **118**: 986-995.

Getz, E.B., Xiao, M., Chakrabarty, T., Cooke, R. and Selvin, P.R. (1999). A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phophine and dithiothreitol for use in protein biochemistry. *Anal. Biochem.*, **273**: 73-80.

Goldberg, R.N., Kishore, N. and Lennen, R.M. (2002). Thermodynamic quantities for the ionization reactions of buffers. *J. Phys. Chem.*, **31**: 231-360.

Gonzalez, M.A. (2011). Force fields and molecular dynamics simulations. *Collection SFN*, DOI: 10.1051/sfn/201112009.

Goodchild, S.C., Angstmann, C.N., Breit, S.N., Curmi, P.M.G. and Brown, L.J. (2011b). Transmembrane extension and oligomerisation of the CLIC1 chloride intracellular channel protein upon membrane interaction. *Biochem.*, DOI: 10.1021/bi2012564.

Goodchild, S.C., Curmi, P.M.G. and Brown, L.J. (2011a). Structural gymnastics of multifunctional metamorphic proteins. *Biophys. Rev.*, DOI: 10.1007/s12551-011-0053-8.

Goodchild, S.C., Howell, M.W., Cordina, N.M., Littler, D.R., Samuel, N.B., Curmi, P.M.G. and Brown, L.J. (2009). Oxidation promotes insertion of the CLIC1 chloride intracellular channel into the membrane. *Eur. Biophys. J.*, **39**: 129-138.

Goodchild, S.C., Howell. M.W., Littler, D.R., Mandyam, R.A., Sale, K.L., Mazzanti, M., Breit, S.N., Curmi, P.M.G. and Brown, L.J. (2010). Metamorphic response of the CLIC1 chloride intracellular channel protein upon membrane interaction. *Biochem.*, **49**: 5278-5289.

Gritti, M., Wurth, R., Angelini, M., Barnieri, F., Peretti, M., Pizzi, E., Pattarozzi, A., Carra, E., Sirito, R., Daga, A., Curmi, P.M., Mazzanti, M. and Florio, T. (2014). Metformin repositioning as antitumoral agent: selective antiproliferative effects in human glioblastoma stem cells, via inhibition of CLIC1-mediated ion current. *Oncotarget*, **5**: 11252-11268.

Guedes, I.A., Pereira, F.S.S. and Dardenne, L.E. (2018). Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. *Front. Pharmacol.*, DOI.org/10.3389/fphar.2018.01089.

Habig, W. H., Pabst, M. J. and Jakoby, W. B. (1974). Glutathione *S*-transferases. The first enzymatic step in mercapturic acid formation. *J. Biol. Chem.*, **249**: 7130–7139.

Hansen, A.M., Gu, Y., Li, M., Andrykovitch, M., Waugh, D.S., Jin, D.J. and Ji, X. (2005). Structural basis for the function of stringent starvation protein A as a transcription factor. *J. Biol. Chem.*, **280**: 17380-17391.

Hansen, A.M., Lehnherr, H., Wang, X., Mobley, V. and Jin, D.J. (2003). *Escherichia coli* SspA is a transcription activator for bacteriophage P1 late genes. *Mol. Microbiol*., **48**: 1621-1631.

Hansen, L.D., Fellingham, G.W. and Russell, D.J. (2011). Simultaneous determination of equilibrium constants and enthalpy changes by titration calorimetry: methods, instruments and uncertainties. *Anal. Biochem.*, **409**: 220-229.

Hao, G. and Gross, S.S. (2006). Electrospray tandem mass spectrometry analysis of *S*- and *N*-nitrosopeptides: facile loss of NO and radical-induced fragmentation. *J. Am. Soc. Mass Spectrom.*, **17**: 1725-1730.

Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y., Wang, L., Lupyan, D., Dahlgren, M.K., Knight, J.L., Kaus, J.W., Cerutti, D.S., Krilov, G., Jorgensen, W.L., Abel, R. and Friesner, R.A. (2015). OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.*, **12**: 281-296.

Hare, J.E., Goodchild, S.C., Breit, S.N., Curmi, P.M.G. and Louise, B. (2016). Interaction of human chloride intracellular channel protein 1 (CLIC1) with lipid bilayers: a fluorescence study. *Biochem*., DOI: 10.1021/acs.biochem.6b00080.

Harris, D.C. (2010). Quantitative chemical analysis. 8th ed. New York (USA): W.H. Freeman and Company.

Harrop, J.S., DeMaere, M.Z., Fairlie, W.D., Reztsova, T., Valenzuela, S.M., Mazzanti, M., Tonini, R., Qiu, M.R., Jankova, L., Warton, K., Bauskin, A.R., Wu, W.M., Pankhurst, S., Campbell, T.J., Breit, S.N. and Curmi, P.M.G. (2001). Crystal structure of a soluble form of the intracellular chloride ion channel CLIC1 (NCC27) at 1.4 Å Resolution. *J. Biol. Chem*., **276**: 44993-45000.

Hart T.W. (1985). Some observations concerning the *S*-nitroso and *S*-phenylsulphonyl derivatives of L-cysteine and glutathione. *Tetrahedron Lett.*, **26**: 2013-2016.

Hawe, A., Sutter, M. and Jiskoot, W. (2008). Extrinsic fluorescent dyes as tools for protein characterisation. *Pharm. Res.*, **25**: 1487-1499.

Heiss, N.S. and Poustka, A. (1997). Genomic structure of a novel chloride channel gene, CLIC2, in Xq28. *Genomics*, **45**: 224-228.

Hitchens, T. K., Mannervik, B. and Gordon, S. (2001). Disorder-to-order transition of the active site of human class Pi glutathione transferase, GST P1-1. *Biochem.,* **40**: 11660-11669.

Hofer, T.S. and de Visser, S.P. (2018). Editorial: quantum mechanical/molecular mechanical approaches for the investigation of chemical systems-recent developments and advanced applications. *Front. Chem.,* **6**: 1-5.

Hoffmann, E. and Stroobant, V. (2007). Mass spectrometry: principles and applications. 3rd ed. England (UK): John Wiley and Sons, Ltd.

Holmes, A.J. and Williams, D.L.H. (1998). Reaction of *S*-nitrosothiols with ascorbate: clear evidence of two reactions. *Chem. Commun.,* **0**: 1711-1712.

Igloi, Z., Mohl, B.P., Lippiat, J.D., Harris, M. and Mankouri, J. (2015). Requirement for chloride channel function during the hepatitis C virus life cycle. *J. Virol.,* **89**: 4023-4029.

Ishihama, A. and Saitoh, T. (1979). Subunits of RNA polymerase in function and structure. *J. Mol. Biol.*, **129**: 517-530.

Ji, X., Tordova, M., O'Donnell, R., Parsons, J. F., Hayden, J. B., Gilliland, G. L. and Zimniak, P. (1997). Structure and function of the xenobiotic substrate binding site and the location of a potential non-substrate-binding site in a class Pi glutathione *S*-transferase. *Biochem.,* **36:** 9690-9702.

Jorgensen, W.L., Maxwell, D.S. and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.,* **118**: 11225-11236.

Karplus, P.A. and Diederichs, K. (2012). Linking crystallographic model and data quality. *Sci.*, DOI: 10.1126/science.1218231.

Kelly, S.M., Thomas, J.J. and Price, N.C. (2005). How to study protein by circular dichroism. *Biochim. Biophys. Acta*, **1751**: 119-139.

Khamici, H.A., Brown, L.J., Hossain, K.R., Hudsom, A.L., Sinclair-Burton, A.A., Ng, J.P.M., Daniel, E.L., Hare, J.E., Cornell, B.A., Curmi, P.M., Davey, M.W. and Valenzuela, S.M. (2015). Members of the chloride intracellular ion channel protein family demonstrates glutaredoxin-like enzymatic activity. *PLoS ONE*, **10**: 1-19.

Kinsley, N.M. (2005). Energetics of ligand binding to the active site of glutathione transferase M1-1. *MSc Dissertation*. University of the Witwatersrand, Molecular and Cell Biology. http://wiredspace.wits.ac.za/jspui/handle/10539/1715.

Klapoetke, S. and Xie, M.H. (2017). Disulfide bond characterisation of human factor Xa by mass spectrometry through protein-level partial reduction. *J. Pharm. Biomed. Anal.*, **132**: 238-246.

Kovacs, I. and Christian, L. (2013). Nitric oxide-based protein modification: formation and site-specificity of protein *S*-nitrosylation. *Front. Plant Sci.*, DOI: 10.3389/fpls.2013.00137.

Krumova, K. and Cosa, G. (2016). Overview of reactive oxygen species. *Photochem. Photobiol. Sci.*, DOI: 10.1039/9781782622208-00001.

Kukol, A. (2008). *Molecular modelling of proteins*. New York (USA): Humana Press.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**: 680-685.

Lakowicz, J.R. (2006). *Principles of fluorescence spectroscopy*. 3rd ed. New York (USA): Springer Science + Business Media, LLC.

Landry, D.W., Reitman, M., Cragoe, E.J. and Al-Awqati, Q. (1987). Epithelial chloride channel. Development of inhibitory ligands. *J. Gen. Physiol.*, **90**: 779-798.

Laskowski, R.A. (2001). PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.*, **29**: 221-222.

Legg-E'silva, D., Achilonu, I., Fanucchi, S., Stoychev, S., Fernandes, M. and Dirr, H.W. (2012). Role of arginine 29 and glutamic acid 81 interactions in the conformational stability of human chloride intracellular channel 1. *Biochem.*, **51**: 7854-7862.

Lemkul, J.A. (2018). GROMACS tutorial: lysozyme in water. Available from: http://www.mdtutorials.com/gmx/lysozyme/index.html.

Li, H., Robertson, A.D. and Jensen, J.H. (2005). Very fast empirical prediction and rationalization of protein pK_a values. *Proteins*, **61**: 704-721.

Littler, D.R., Assaad, N.N., Harrop, S.J., Brown, L.J., Pankhurst, G.J., Luciani, P., Aguilar, M.I., Mazzanti, M., Berryman, M.A., Breit, S.N. and Curmi, P.M.G. (2005). Crystal structure of the soluble form of the redox-regulated chloride ion channel protein CLIC4. *FEBS J.*, **272**: 4996-5007.

Littler, D.R., Harrop, S.J., Goodchild, S.C., Phang, J.M., Mynott, A.V., Jiang, L., Valenzuela, S.M., Mazzanti, M., Brown, L.J., Breit, S.N. and Curmi, P.M. (2010). The enigma of the CLIC proteins: ion channels, redox proteins, enzymes, scaffolding proteins? *FEBS Lett.*, **584**: 2093-2101.

Littler, D.R., Stephen, J.H., Fairlie, W.D., Brown, L.J., Pankhurst, G.J., Pankhurst, S., DeMaere, M.Z., Campbell, T.J., Bauskin, A.R., Tonini, R., Mazzanti, M., Breit, S.N. and Curmi, P.M.G. (2004). The intracellular chloride ion channel protein CLIC1 undergoes a redox-controlled structural transition. *J. Biol. Chem.*, **279**: 9298-9305.

Liu, L., Hausladen, A., Zeng, M., Que, L., Heitman, J. and Stamler, J.S. (2001). A metabolic enzyme for *S*-nitrosothiol conserved from bacteria to humans. *Nature*, **410**: 490-494.

Liu, Q., Li, J. and Li, J. (2014). Use B-factor related features for accurate classification between protein binding interfaces and crystal packing contacts. *BMC Bioinform.*, **15**: 2-11.

Liu, Z., Cao, J., Ma, Q., Gao, X., Ren, J. and Xue, Y. (2011). GPS-YNO2: computational prediction of tyrosine nitration sites in proteins. *Mol. BioSyst.*, **7**: 1197-1204.

Luo, J., Liu, Z., Guo, Y. and Li, M. (2015). A structural dissection of large protein-protein crystal packing contacts. *Sci. Rep.*, DOI: 10.1038/srep14214.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C. and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analysing targeted proteomics experiments. *Bioinform.*, **26**: 966-968.

Madej, E., Folkes, L.K., Wrdman, P., Czapski, G. and Goldstein, S. (2008). Thiyl radicals react with nitric oxide to form *S*-nitrosothiols with rate constants near the diffusion-controlled limit. *Free Radic. Biol. Med.*, **44**: 2013-2018.

Malik, M., Shukla, A., Amin, P., Niedelman, W., Lee, J., Jividen, K., Phang, J.M., Ding, J., Suh, K.S., Curmi, P.M. and Yuspa, S.H. (2010). *S*-nitrosylation regulates nuclear translocation of chloride intracellular channel protein CLIC4. *J. Biol. Chem.*, **285**: 23818–23828.

Marino, S.M. and Gladyshev, V.N. (2010). Cysteine function governs its conservation and degeneration and restricts its utilisation on protein surfaces. *J. Mol. Biol.*, **404**: 902-916.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007). PHASER crystallographic software. *J. Appl. Crystallogr.*, **40**: 658-674.

McIntyre, S. (2007). Effects of the environment on the conformational stability of the chloride intracellular channel protein CLIC1. *PhD thesis*. University of the Witwatersrand, Molecular and Cell Biology. http://mobile.wiredspace.wits.ac.za/handle/10539/4853.

McLaughlin, S. (1989). The electrostatic properties of membranes. *Annu. Rev. Biophys. Biophys. Chem.*, **18**: 113-136.

McTigue, M.A., Williams, D.R. and Tainer, J.A. (1995). Crystal structures of a schistosomal drug and vaccine target: glutathione *S*-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug praziquantel. *J. Mol. Biol.*, **246**: 21-27.

Monticelli, L. and Tieleman, P. (2014). Force fields for classical molecular dynamics. *Methods Mol. Biol*., DOI: 10.1007/978-1-62703-017-5_8.

Musdal, Y., Hegazy, U.M., Aksoy, Y. and Mannervik, B. (2013). FDA-approved drugs and other compounds tested as inhibitors of human glutathione transferase P1-1. *Chem. Biol. Interact.*, **205**: 53-62.

Mynott, A.V., Harrop, S.J., Brown, L.J., Breit, S.N., Kobe, B. and Curmi, P.M.G. (2011). Crystal structure of importin-α bound to a peptide bearing the nuclear localisation signal from chloride intracellular channel 4. *FEBS J.*, **278**: 1662-1675.

Nakamura, T., Tu,S., Akhtar, M.W., Sunico, C.R., Okamoto, S. and Lipton, S.A. (2013). Aberrant protein *S*-nitrosylation in neurodegenerative diseases. *Neuron*, **78**: 596-614.

Nishizawa, T., Nagao, T., Iwatsubo, T., Forte, J.G. and Urushidani, T. (2000). Molecular cloning and characterisation of a novel chloride intracellular channel-related protein, parchorin, expressed in water-secreting cells. *J. Biol. Chem.*, **275**: 11164-11173.

Novarino, G., Fabrizi, C., Tonini, R., Denti, M.A., Malchiodi-Albedi, F., Lauro, G.M., Saccetti, B., Paradisi, S., Ferroni, A., Curmi, P.M., Breit, S.N. and Mazzanti, M. (2004). Involvement of the intracellular ion channel CLIC1 in microglia-mediated β-amyloid-induced neurotoxicity. *J. Neurosci.*, **24**: 5322-5330.

Oakley, A. (2011). Glutathione transferase: a structural perspective. *Drug Metab. Rev.*, **43**: 138-151.

Oakley, A.J., Bello, M.L., Nuccetelli, M., Mazzetti, A.P. and Parker, M.W. (1999). The ligandin (non-substrate) binding site of human pi class glutathione transferase is located in the electrophile binding site (H-site). *J. Mol. Biol.*, **291**: 913-926.

Oakley, A.J., Lo Bello, M., Battistoni, A., Ricci, G., Rossjohn, J., Villar, H.O. and Parker, M.W. (1997). The structures of human glutathione transferase P1-1 in complex with glutathione and various inhibitors at high resolution. *J. Mol. Biol.*, **274**: 84-100.

Oakley, A.J., Lo Bello, M., Ricci, G., Federici, G. and Parker, M.W. (1998). Evidence for an induced-fit mechanism operating in pi class glutathione transferases. *Biochem.*, **37**: 9912-9917.

Oakley, A.J., Rossjohn, J., Bello, M.L., Caccuri, A.M., Federici, G. and Parker, M.W. (1996). The three-dimensional structure of human pi class glutathione transferase P1-1 in complex with the inhibitor ethacrynic acid and its glutathione conjugate. *Biochem.*, **36**: 576-585.

Paige, J.S., Xu, G., Stancevic, B. and Jaffrey, S.R. (2008). Nitrosothiol reactivity profiling identifies *S*-nitrosylated proteins with unexpected stability. *Chem. Biol.*, **15**: 1307-1316.

Parker, M.W. and Feil, S.C. (2005). Pore-forming protein toxins: from structure to function. *Prog. Biophys. Mol. Biol.*, **88**: 91-142.

Peretti, M., Angelini, M., Savalli, N., Florio, T., Yuspa, S.H. and Mazzanti, M. (2015). Chloride channels in cancer: focus on chloride intracellular channel 1 and 4 (CLIC1 and CLIC4) proteins in tumour development and as novel therapeutic agents. *Biochim. Biophys. Acta*, **1848**: 2523-2531.

Perkins, S.J. (1986). Protein volume and hydration effects. *Eur. J. Biochem.*, **319**: 749-754.

Perozzo, R., Folkers, G. and Scapozza, L. (2004). Thermodynamics of protein-ligand interactions: history, presence and future aspects. *J. Recept. Sig. Transd.*, **24**: 1-52.

Peter, B., Fanucchi, S. and Dirr, H.W. (2014). A conserved cationic motif enhances membrane binding and insertion of the chloride intracellular channel protein 1 transmembrane domain. *Eur. Biophys. J.*, **43**: 405-414.

Peter, B., Polyansky, A.A., Fanucchi, S. and Dirr, H.W. (2013). A Lys-Trp cation-π interaction mediates the dimerisation and function of the chloride intracellular channel protein 1 transmembrane domain. *Biochem.*, **53**: 57-67.

Petrova, D.T., Asif, A.R., Armstrong, V.W., Dimova, I., Toshev, S., Yaramo, N., Oellerich, M. and Toncheva, D. (2008). Expression of chloride intracellular channel protein 1 (CLIC1)

and tumour protein D52 (TPD52) as potential biomarkers for colorectal cancer. *Clin. Biochem.*, **41**: 1224-1236.

Poole, L. (2014). The basic of thiols and cysteines in redox biology and chemistry. *Free Radic. Biol. Med.*, **80**: 148-157.

Powell, H.R. (2017). X-ray data processing. *Biosci. Rep.*, DOI: 10.1042/BSR20170227.

Prabhu, N.V. and Sharp, K.A. (2005). Heat capacity in proteins. *Annu. Rev. Phys. Chem.*, **56**: 521-548.

Prade, L., Huber, R., Manoharan, T.H., Fahl, W.E. and Reuter, W. (1997). Structures of class Pi glutathione *S*-transferase from human placenta in complex with substrate, transition-state analogue and inhibitor. *Struct.*, **5**: 1287-1295.

Proutski, I., Karoulias, N. and Ashley, R.H. (2002). Overexpressed chloride intracellular channel protein CLIC4 (p64H1) is an essential component of novel plasma membrane anion channels. *Biochem. Biophys. Res. Commun.*, **297**: 317-322.

Qian, Z., Okuhara, D., Abe, M.K. and Rosner, M.R. (1999). Molecular cloning and characterisation of a mitogen-activated protein kinase-associated intracellular chloride channel. *J. Biol. Chem.*, **274**: 1621-1627.

Quesada-Soriano, I., Parker, L.J., Primavera, A., Casas-Solvas, J.M., Vargas-Berenguel, A., Baron, C., Morton, C.J., Mazzetti, A.P., Bello, M.L., Parker, M.W. and Garcia-Fuentes, L. (2009). Influence of the H-site residue 108 on human glutathione transferase P1-1 ligand binding: structure-thermodynamic relationships and thermal stability. *Protein Sci.*, **18**: 2454-2470.

Quesada-Soriano, I., Parker, L.J., Primavera, A., Wielens, J., Holien, J.K., Casas-Solvas, J.M., Vargas-Berenguel, A., Aguilera, A.M., Nuccetelli, M., Mazzetti, A.P., Bello, M.L., Parker, M.W. and Garcia-Fuentes, L. (2010). Diuretic drug binding to human glutathione transferase P1-1: potential role of Cys-101 revealed in the double mutant C47S/Y108V. *J. Mol. Recognit.*, **24**: 220-234.

Radi, R. (2013). Protein tyrosine nitration: biochemical mechanisms and structural basis of its functional effects. *Acc. Chem. Res.*, **46**: 550-559.

Ralat, L.A. and Colman, R.F. (2004). Glutathione *S*-transferase Pi has at least three distinguishable xenobiotic substrate sites close to its glutathione-binding site. *J. Biol. Chem.*, **279**: 50204-50213.

Ramachandran, G.N. and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Prot. Chem.*, **23**: 283-438.

Rao, S.G., Ponnalagu, D., Sukur, S., Singh, H., Sanghvi, S., Mei, Y., Jin, D.J. and Singh, H. (2017). Identification and characterisation of a bacterial homolog of chloride intracellular channel (CLIC) protein. *Sci. Rep*., **7**: 1-13.

Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G. and Deber, C.M. (2008). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *PNAS*, **106**: 1760-1765.

Raynal, B., Lenormand, P., Baron, B., Hoos, S. and England, P. (2014). Quality assessment and optimisation of purified proteins samples: why and how? *Microb. Cell. Fact.*, **13**: 2-10.

Redhead, C., Sullivan, S.K., Koseki, C., Fujiwara, K. and Edwards, J.C. (1997). Subcellular distribution and targeting of the intracellular chloride channel p64. *Mol. Biol. Cell.*, **8**: 691-704.

Reynolds, J.A. and Tanford, C. (1970). Binding of dodecyl sulfate to proteins at high binding ratios: Possible implications for the state of proteins in biological membranes. *Proc. Natl. Acad. Sci.*, **66**: 1002–1007.

Rhodes, G. (2006). *Crystallography made crystal clear: a guide for users of macromolecular models*. 3rd ed. Massachusetts (USA): Academic Press.

Rocco, A.G., Mollica, L., Ricchiuto, P., Baptista, A.M., Gianazza, E. and Eberini, I. (2008). Characterisation of the protein unfolding process induced by urea and temperature. *Biophys. J.*, **94**: 2241-2251.

Ronnov-Jessen, L., Villadsen, R., Edwards, J.C. and Petersen, O.W. (2002). Differential expression of a chloride intracellular channel gene, CLIC4, in transforming growth factor beta1-mediated conversion of fibroblasts to myofibroblasts. *Am. J. Pathol.*, **161**: 471- 480.

Ross, P.D. and Subramanian, S. (1981) Thermodynamics of protein association reactions: forces contributing to stability. *Biochem.*, **20**: 3096-3102.

Rupp, B. (2009). *Biomolecular crystallography-principles, practice and application to structural biology*. 1st ed. New York (USA): Garland Science.

Sarti P., Giuffre A., Forte E., Mastronicola D., Barone M. C., Brunori M. (2000). Nitric oxide and cytochrome c oxidase: mechanisms of inhibition and NO degradation. *Biochem. Biophys. Res. Commun.*, **274**: 183–187.

Sasaki, S., Yui, N. and Noda, Y. (2014). Actin directly interacts with different membrane channel proteins and influences channel activities: AQP2 as a model. *Biochim. Biophys. Acta*, **1838**: 514-520.

Schilling, B., Rardin, M.J., MacLean, B.X., Zawadzka, A.M., Frewen, B.E., Cusack, M.P., Sorensen, D.J., Bereman, M.S., Jing, E., Wu, C.C., Verdiny, E., Kahn, R., MacCoss, M.J. and Gibson, B.W. (2012). Platform independent and label-free quantification of proteomic data MS1 extracted ion chromatograms in Skyline. Applications to protein acetylation and phosphorylation. *Mol. Cell. Proteomics*, **11**: 202-214.

Schlesinger, P.H., Blair, H.C., Teitelbaum, S.L. and Edwards, J.C. (1997). Characterisation of the osteoclast ruffled border chloride channel and its role in bone resorption. *J. Biol. Chem.*, **272**: 18636-18643.

Schmid, F.X. (2001). *Biological macromolecules: UV-Visible spectrophotometry*. New York (USA): Macmillan Publishers Ltd.

Schrödinger®, LLC. (2015). *Glide 6.7: user manual*. Available from: https://www.schrodinger.com/maestro.

Schrödinger®, LLC. (2019a). *How the IFD Score calculated and what is its units?*. Available from: https://www.schrodinger.com/kb/307.

Schrödinger®, LLC. (2019b). *What is the difference between Glide$_{Score}$ and E$_{model}$, and which should I use for ranking poses?* Available from: https://www.schrodinger.com/kb/1027.

Schrödinger®, LLC. (2019c). *Glide$_{Score}$/Docking Score doesn't correlate with my known activities. What is wrong?*. Available from: https://www.schrodinger.com/kb/144.

Sheehan, D. (2009). *Physical biochemistry: Principles and applications*. England (UK): John Wiley and Sons Ltd.

Shelley, J.C., Cholleti, A., Frye, L.L., Greenwood, J.R., Timlin, M.R. and Uchimaya, M. (2007). Epik: a software program for pK_a prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.*, **21**: 681-691.

Sherwood, D. and Cooper, J. (2011). *Crystals, X-rays and proteins: comprehensive protein crystallography*. 1st ed. England (UK): Oxford University Press.

Silkstone, R.S., Mason, M.G., Nicholls, P. and Cooper, C.E. (2012). Nitrogen dioxide oxidizes mitochondrial cytochrome c. *Free Radic. Biol. Med.*, **52**: 80-87.

Singh, H. (2010). Two decades with dimorphic chloride intracellular channels (CLICs). *FEBS Lett.*, **584**: 2112-2121.

Singh, H. and Ashley, R.H. (2006). Redox regulation of CLIC1 by cysteine residues associated with the putative channel pore. *Biophys. J.*, **90**: 1628-1638.

Singh, R.J., Hogg, N., Joseph, J. and Kalyanaraman, B. (1996). Mechanism of nitric oxide release from S-nitrosothiols. *J. Biol. Chem.*, **271**:18596-18603.

Smith, B.C. and Marletta, M.A. (2012). Mechanisms of *S*-nitrosothiol formation and selectivity in nitric oxide signalling. *Curr. Opin. Chem. Biol.*, **16**: 498-506.

Smith, D.B. and Johnson, K.S. (1988). Single-step purification of polypeptides in *Escherichia coli* as functions with glutathione *S*-transferase. *Gene*, **67**: 31-40.

Smith, J.N. and Dasgupta, T.P. (2000). Kinetics and mechanism of the decomposition of *S*-nitrosoglutathione by L-ascorbic acid and copper ions in aqueous solution to produce nitric oxide. *Nitric Oxide-Biol. Chem.*, **4**: 57-66.

Snyder, P.W., Lockett, M.R., Moustakas, D.T. and Whitesides, G.M. (2013). Is it the shape of the cavity, or the shape of the water in the cavity? *Eur. Phys. J. Spec. Top.*, DOI: 10.1140/epjst/e2013-01818-y.

Son, J., Lee, J.J, Lee, J.S., Schüller, A. and Chang, Y.T. (2010). Isozyme-specific fluorescent inhibitor of glutathione transferase omega 1. *ACS Chem. Biol.*, **5**: 449-453.

Stefan, M.I. and Novere, N. (2013). Cooperative binding. *PLoS ONE*, DOI: 10.1371/journal.pcbi.1003106.

Stoychev, S.H., Nathaniel, C., Fanucchi, S., Brock, M., Li, S., Asmus, K., Woods, V.L. and Dirr, H.W. (2009). Structural dynamics of soluble chloride intracellular channel protein CLIC1 examined by amide hydrogen-deuterium exchange mass spectrometry (DXMS). *Biochem.*, **48**: 8413-8421.

Strickland, E.H. (1974). Aromatic contributions to circular dichroism spectra of proteins. *Crit. Rev. Biochem.*, **2**: 113-175.

Stryer, L.S. (1965). The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J. Mol. Biol.*, **13** (2): 482-495.

Suh, K.S., Malik, M., Shukla, A., Ryscavage, A., Wright, L., Jividen, K., Crutchley, J.M., Dumont, R.A., Fernandez-Salas, E., Webster, J.D., Simpson, R.M. and Yuspa, S.H. (2012). CLIC4 is a tumor suppressor for cutaneous squamous cell cancer. *Carcinogenesis*, **0**: 1-10.

Suh, K.S., Mutoh, M., Nagashima, K., Fernandez-Salas, E., Edwards, L.E., Hayes, D.D., Crutchley, J.M., Marin, K.G., Dumont, R.A., Levy, J.M., Cheng, C., Garfield, S. and Yuspa,

S.H. (2004). The organellular chloride channel protein CLIC4 / mtCLIC translocates to the nucleus in response to cellular stress and accelerates apoptosis. *J. Biol. Chem.*, **279**: 4632–4641.

Syme, N.R., Dennis, C., Bronowska, A., Paesen, G.C. and Homans, S.W. (2010). Comparison of entropic contributions to binding in a "hydrophilic" versus "hydrophobic" ligand-protein interaction. *J. Am. Chem. Soc.*, **132**: 8682-8689.

Tonini, R., Ferroni, A., Valenzuela, S.M., Warton, K., Campbell, T.J., Breit, S.N. and Mazzanti, M. (2000). Functional characterisation of the NCC27 nuclear protein in stable transfected CHO-K1 cells. *FASEB J.*, **14**: 1171-1178.

Tulk, B.M., Kapadia, S. and Edwards, J.C. (2002). CLIC1 inserts from the aqueous phase into phospholipid membranes, where it functions as an anion channel. *Am. J. Physiol. Cell Physiol.*, **282**: C1103-C1112.

Tulk, B.M., Schlesinger, P.H. and Edwards, J.C. (1998). NCC27, a homolog of intracellular Cl channel p64, is expressed in brush border of renal proximal tubule. *Am. J. Physiol. Renal Physiol.*, **274**: F1140-F1149.

Ulmasov, B., Bruno, J., Woost, P.G. and Edwards, J.C. (2007). Tissue and subcellular distribution of CLIC1. *BMC Cell Biol.*, **8**: 1-18.

Upadhyay, A.K., Murmu, A., Singh, A. and Panda, A.K. (2012). Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in *Escherichia coli*. *PLoS ONE*, **7**: 1-11.

Uversky, V.N., Winter, S. and Lober, G. (1998). Self-association of 8-anilino-1-naphthalenesulfonate molecules: spectroscopic characterisation and application to the investigation of protein folding. *Biochim. Biophys. Acta*, **1388**: 133-142.

Valenzuela, S.M., Martin, D.K., Por, S.B., Robbins, J.M., Warton, K., Bootcov, M.R., Schofiel, P.R., Campbell, T.J. and Breit, S.N. (1997). Molecular cloning and expression of a chloride channel of cell nuclei. *J. Biol. Chem.*, **272**: 12575-12582.

Valenzuela, S.M., Mazzanti, M., Tonini, R., Qiu, M.R., Warton, K., Musgrove, E.A., Campbell, T.J. and Breit, S.N. (2000). The nuclear chloride ion channel NCC27 is involved in regulation of the cell cycle. *J. Physiol.*, **529**: 541-552.

Varela, L., Hendry, A.C., Cantoni, D. and Ortega-Roldan, J.L. (2019). The membrane insertion of soluble CLIC1 into active channels is triggered by specific divalent cations. *bioRxiv*, DOI: https://doi.org/10.1101/638080.

Vera, L. and Stura, E.A. (2014). Strategies of protein cryocrystallography. *Cryst. Growth Des.*, **14**: 427-435.

Verkman, A.S. and Galietta, L.J. (2009). Chloride channels as drug targets. *Nat. Rev. Drug Discov.*, **8**: 153-171.

Voet, D. and Voet, J.G. (2011). *Biochemistry*. 4th ed. New Jersey (USA): John Wiley and Sons, Inc.

Warton, K., Tonini, R. Fairlie, W.D., Matthews, J.M., Valenzuela, S.M., Qui, M.R., Wu, W.M., Pankhurst, S., Bauskkin, A.R., Harrop, S.J., Campbell, T.J., Curmi, P.M., Breit, S.N. and Mazzanti, M. (2002). Recombinant CLIC1 (NCC27) assembles in lipid bilayers via a pH-dependent two-state process to form chloride ion channels with identical characterisitcs to those observed in Chinese hamster ovary cells expressing CLIC1. *J. Biol. Chem.*, **277**: 26003-26011.

Weber, G. and Young, L. B. (1964). Fragmentation of bovine serum albumin by pepsin. i. the origin of the acid expansion of the albumin molecule. *J. Biol. Chem.*, **239**: 1415-1423.

Weidinger, A. and Kozlov, A.V. (2015). Biological activities of reactive oxygen and nitogen species: oxidative stress versus signal transduction. *Biomolecules*, **5**: 472-484.

Whitmore, L. and Wallace, B.A. (2007). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, **89**: 392-400.

Williams, D.L.H. (1998). The chemistry of *S*-nitrosothiols. *Acc. Chem. Res.*, **32**: 869-876.

Williams, E.A. and Rowley, C.N. (2016). Evaluation of methods for the calculation of the p*K*a of cysteine residues in proteins. *J. Chem. Theory Comput.*, **12**: 4662-4673

Wilson, K. and Walker, J. (2010). *Principles and techniques of biochemistry and molecular biology*. 7th ed. England (UK): Cambridge University Press.

Winter, G. (2010). Xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.*, **43**: 186-190.

Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.*, **179**: 131-137.

Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J*., **275**: 1-21.

Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J*., **280**: 5705-5736.

Woody, R.W. (1995). Circular dichroism. *Methods Enzymol.*, **246**: 34-71.

Wu, B. and Dong, D. (2012). Human cytosolic glutathione transferase: structure, function and drug discovery. *Trends Pharmacol. Sci.*, **33**: 656-668.

Xu., M. and Lill, M.A. (2013). Induced fit docking, and the use of QM/MM methods in docking. *Drug. Discov. Today. Technol*., **10**: 1-14.

Xue, Y., Liu, Z., Gao, X., Jin, C., Wen, L., Yao, X. and Ren, J. (2010). GPS-SNO: computational prediction of protein *S*-nitrosylation with a modified GPS algorithm. *PLoS ONE*, DOI:10.1371/journal.pone.0011290.

Yap, L.P., Sancheti, H., Ybanez, M.D., Garcia, J., Cadenas, E. and Han, D. (2010). Determination of GSH, GSSG, and GSNO using HPLC with electrochemical detection. *Methods Enzymol.*, **473**: 137-147.

Yildirim, M.A., Goh, K.I., Cusick, M.E., Barabasi, A.L. and Vidal, M. (2007). Drug-target network. *Nat. Biotechnol*., **25**: 1119-1126.

Youle, R.J. and Strasser, A. (2008). The BCL-2 protein family: opposing actiites that mediate cell death. *Nat. Rev. Mol. Cell Biol.*, **9**: 47-59.

Zhang, H., Andrekopoulos, C, Xu, Y., Joseph, J., Hogg, N., Feix, J. and Kalyanaraman, B. (2009). Decreases *S*-nitrosation of peptide thiols in the membrane interior. *Free Radic. Biol. Med.*, **47**: 962-968.

Zhang, L., Eugeni, E.E., Parthun, M.R. and Freitas, M.A. (2003). Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma*, **112**: 77-86.

Zhao, Y.L. and Houk, K.N. (2006). Thionitroxides, RSNHO: the structure of the SNO moiety in "*S*-nitrosohemoglobin", a possible NO reservoir and transporter. *J. Am. Chem. Soc.*, **128**: 1422-1423.

Zitka, O., Skalickova, S., Gumulec, J., Masarik, M., Adam, V., Hubalek, J., Trnkova, L., Kruseova, J., Eckschlager, T. and Kizek, R. (2012). Redox status expressed as GSH: GSSG ration as a marker for oxidative stress in paediatric tumour patients. *Oncol. Lett.*, **4**: 1247-1253.

Chapter 6: Supplementary Data

6.1 List of Figures

6.2 List of Tables

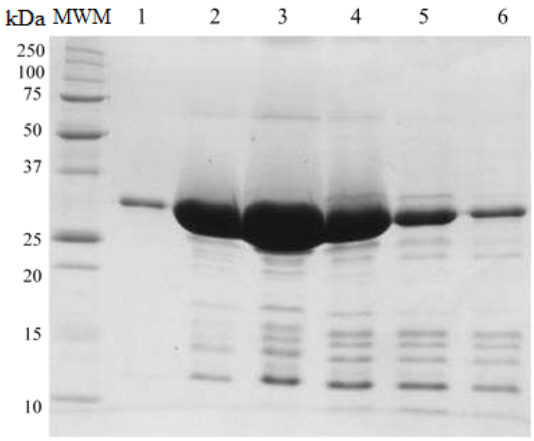
Table S1: List of non-standard reagents and their suppliers

Product name	Catalogue Number	International/local supplier
Isopropyl-β-D-1-thiogalactopyranoside (IPTG)	I56000	Separations (Randburg, SA)
Chloramphenicol	C0378	Sigma Aldrich (St. Louis, Missouri, USA)
Ampicillin sodium salt	A95118	Sigma Aldrich (St. Louis, Missouri, USA)
Kanamycin disulfate salt	K1876	Sigma Aldrich (St. Louis, Missouri, USA)
Lysozyme (chicken egg white)	62971	Sigma Aldrich (St. Louis, Missouri, USA)
DNase I (bovine pancrease)	10104159001	Sigma Aldrich (St. Louis, Missouri, USA)
Phenylmethylsulfonyl fluoride (PMSF)	P7626	Sigma Aldrich (St. Louis, Missouri, USA)
Benzamidine	B6506	Sigma Aldrich (St. Louis, Missouri, USA)
Thrombin (human plasma)	T7009	Sigma Aldrich (St. Louis, Missouri, USA)

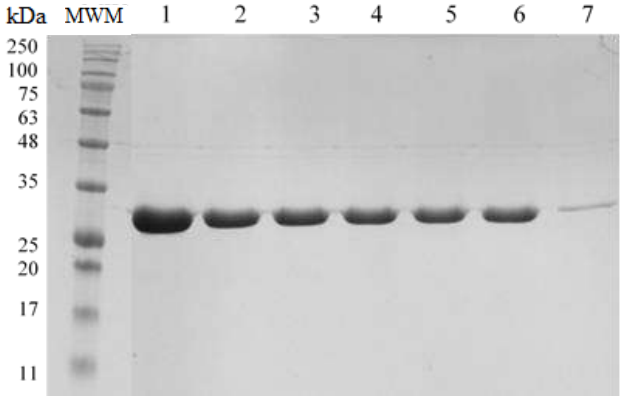
Molecular weight marker	94964	Sigma Aldrich (St. Louis, Missouri, USA)
SEC-HPLC Gel Filtration Standard	1511901	Biorad-Laboratories Hercules, USA (Pretoria, SA)
SEC Gel Filtration Markers Kit (12-200 kDa).	MWGF200	Sigma Aldrich (St. Louis, Missouri, USA)
Tris (2-carboxyethyl) phosphine hydrochloride (TCEP)	GLS GE1860	Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, SA)
Dithiotheitol (DTT)	GLS GC1199	Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, SA)
Dimethyl sulfoxide (DMSO)	C0378	Sigma Aldrich (St. Louis, Missouri, USA)
Indanyloxyacetic acid-94 (IAA-94)	I117	Sigma Aldrich (St. Louis, Missouri, USA)
Ethacrynic acid (EA)	SML1083	Sigma Aldrich (St. Louis, Missouri, USA)
6-iodoacetamide fluorescein (6-IAF)	SML1083	Sigma Aldrich (St. Louis, Missouri, USA)
8-anilinonaphthalene-1-sulfonic acid ammonium salt (ANS)	10417	Sigma Aldrich (St. Louis, Missouri, USA)

2,4-dinitrochlorobenzene (CDNB)	237329	Sigma Aldrich (St. Louis, Missouri, USA)
Glutathione (GSH)	G4251	Sigma Aldrich (St. Louis, Missouri, USA)
MagReSyn® HILIC (hydrophilic interaction chromatography) micro-particles	Purchased by the CSIR	ReSyn Biosciences® (Edenvale, South Africa)
Polyethylene glycol 6000 (PEG-6000)	HR2-513	Hampton Research (Aliso Viejo,California, USA)

A.



B.



C.

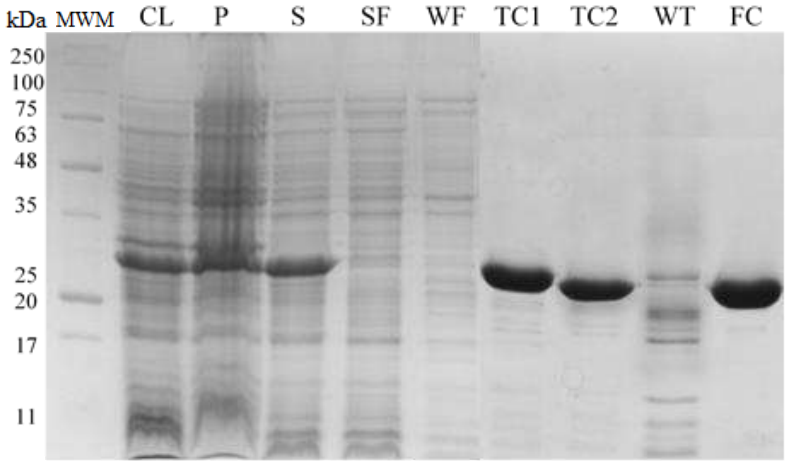
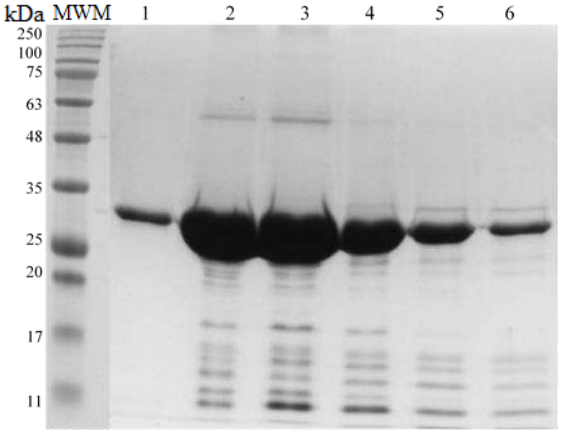
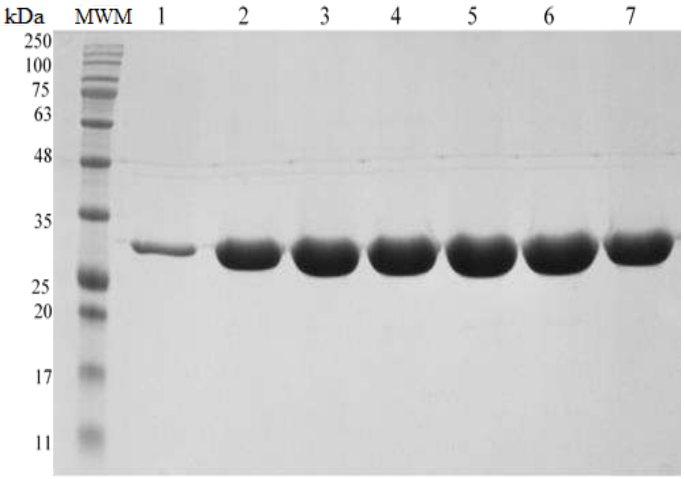


Figure S1. Monitoring the complete purification of CLIC1-WT by SDS-PAGE. (A) The hexa-histidine-CLIC1-WT fusion protein was eluted from the Ni^{2+}-IMAC column and a total of six fractions were collected. (B) A total of seven fractions were collected from the DEAE column. (C) Indicated are the various samples collected throughout the process of obtaining a purified product of CLIC1-WT: (CL) cell lysate containing the fusion protein, (P) pellet of the cell lysate, (S) supernatant of the cell lysate containing the overexpressed fusion protein, (SF) flow-through of the supernatant that was loaded onto the IMAC column, (WF) flow-through after the unbound contaminant and fusion proteins were removed from the IMAC column with the IMAC equilibration buffer, (TC1) fusion protein sample prior to thrombin cleavage, (TC2) fusion protein sample subsequent to thrombin cleavage, (WT) contaminant proteins which bound to the IMAC column subsequent to thrombin cleavage and (FC) final concentrated and purified product of CLIC1-WT. *See section 2.2.4 for the experimental details pertaining to this method.*

A.



B.



C.

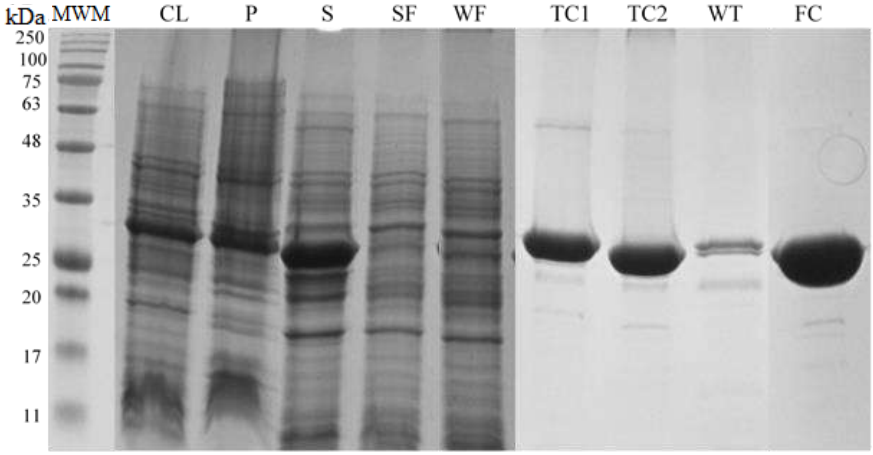
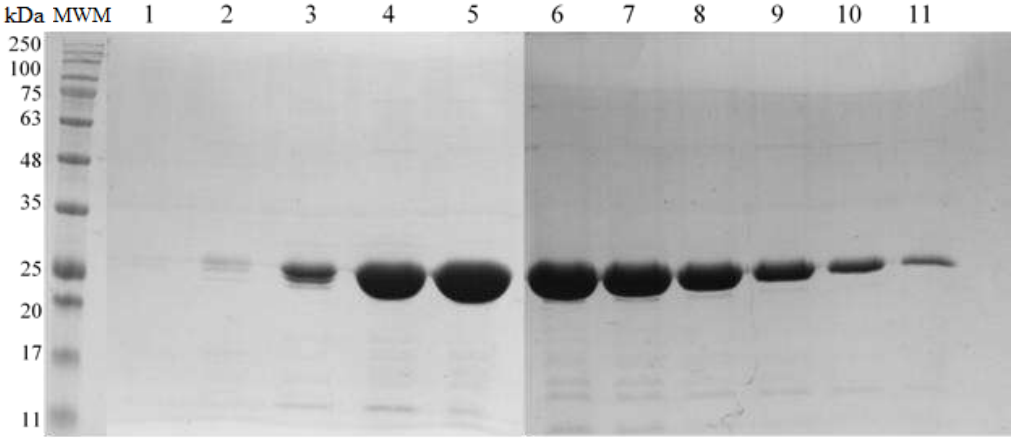
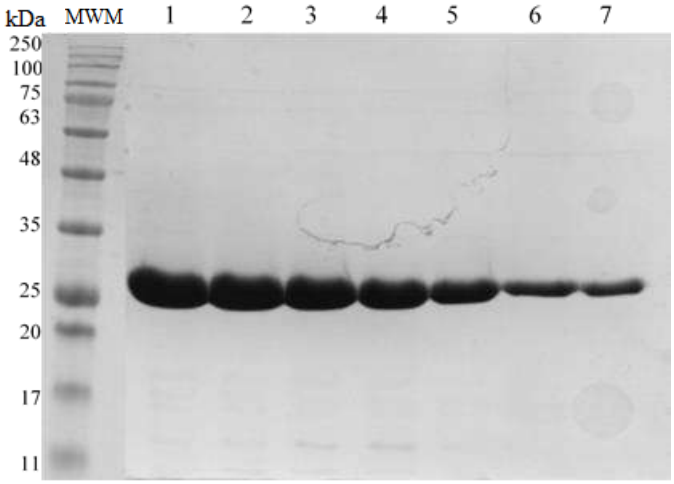


Figure S2. Monitoring the complete purification of CLIC1-C24A by SDS-PAGE. (A) The hexa-histidine-CLIC1-C24A fusion protein was eluted from the Ni^{2+}-IMAC column and a total of six fractions were collected. (B) A total of seven fractions were collected from the DEAE column. (C) Indicated are the various samples collected throughout the process of obtaining a purified product of CLIC1-C24A: (CL) cell lysate containing the fusion protein, (P) pellet of the cell lysate, (S) supernatant of the cell lysate containing the overexpressed fusion protein, (SF) flow-through of the supernatant that was loaded onto the IMAC column, (WF) flow-through after the unbound contaminant and fusion proteins were removed from the IMAC column with the IMAC equilibration buffer, (TC1) fusion protein sample prior to thrombin cleavage, (TC2) fusion protein sample subsequent to thrombin cleavage, (WT) contaminant proteins which bound to the IMAC column subsequent to thrombin cleavage and (FC) final concentrated and purified product of CLIC1-C24A. *See section 2.2.4 for the experimental details pertaining to this method.*

A.



B.



C.

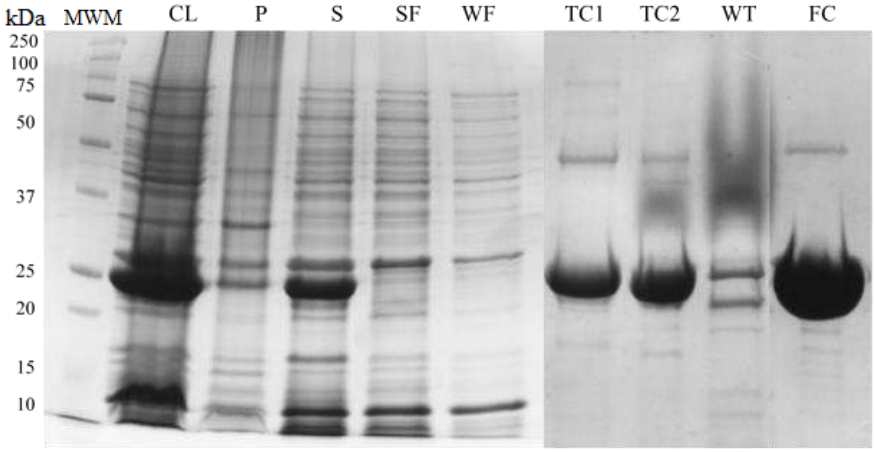
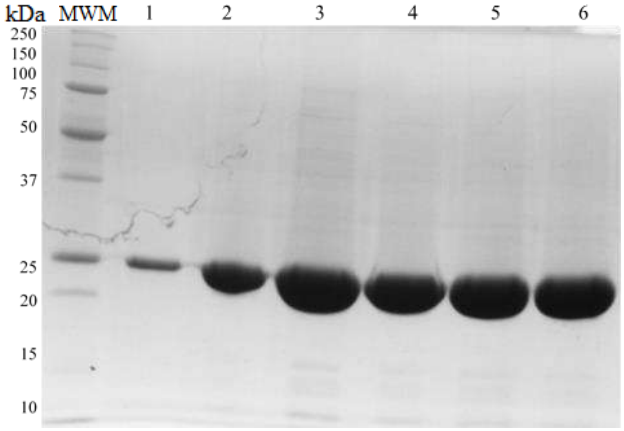
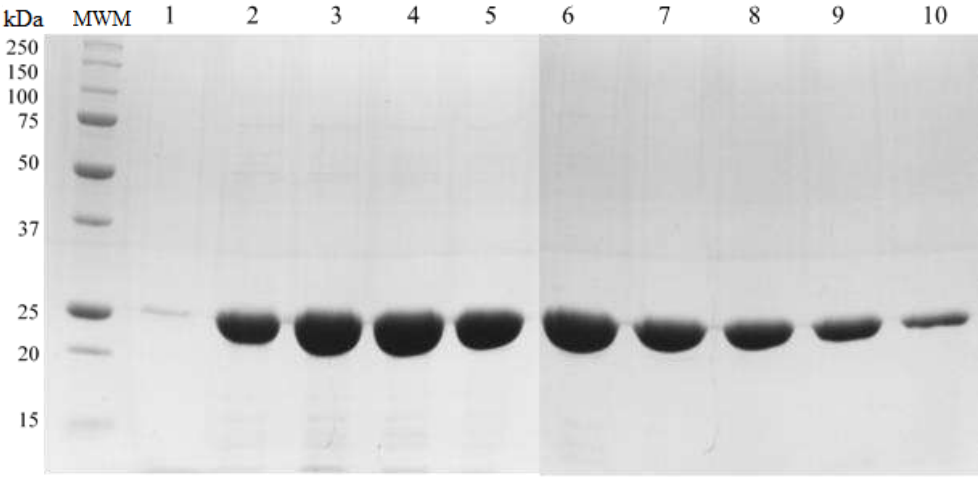


Figure S3. Monitoring the complete purification of hGSTP1-1 by SDS-PAGE. (A) The hexa-histidine-hGSTP1-1 fusion protein was eluted from the Co^{2+}-IMAC column and a total of six fractions were collected. (B) A total of seven fractions were collected from the benzamidine column. (C) Indicated are the various samples collected throughout the process of obtaining a purified product of hGSTP1-1: (CL) cell lysate containing the fusion protein, (P) pellet of the cell lysate, (S) supernatant of the cell lysate containing the overexpressed fusion protein, (SF) flow-through of the supernatant that was loaded onto the IMAC column, (WF) flow-through after the unbound contaminant and fusion protein were removed from the IMAC column with the IMAC equilibration buffer, (TC1) fusion protein sample prior to thrombin cleavage, (TC2) fusion protein sample subsequent to thrombin cleavage, (WT) contaminant proteins which bound to the IMAC column subsequent to thrombin cleavage and (FC) final concentrated and purified product of hGSTP1-1. *See section 2.2.4 for the experimental details pertaining to this method.*

A.



B.



C.

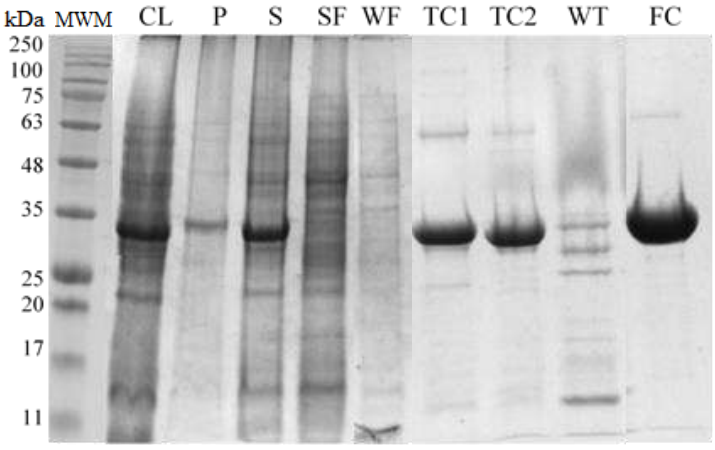


Figure S4. Monitoring the complete purification of EcSspA by SDS-PAGE. (A) The hexa-histidine-EcSspA fusion protein was eluted from the Ni^{2+}-IMAC column and a total of six fractions were collected. (B) A total of seven fractions were collected from the benzamidine column. (C) Indicated are the various samples collected throughout the process of obtaining a purified product of EcSspA: (CL) cell lysate containing the fusion protein, (P) pellet of the cell lysate, (S) supernatant of the cell lysate containing the overexpressed fusion protein, (SF) flow-through of the supernatant that was loaded onto the IMAC column, (WF) flow-through after the unbound contaminant and fusion proteins were removed from the IMAC column with the IMAC equilibration buffer, (TC1) fusion protein sample prior to thrombin cleavage, (TC2) fusion protein sample subsequent to thrombin cleavage, (WT) contaminant proteins which bound to the IMAC column subsequent to thrombin cleavage and (FC) final concentrated and purified product of EcSspA. *See section 2.2.4 for the experimental details pertaining to this method.*

Table S2: Summary of the purity of the final concentrated and purified target proteins that were verified by SDS-PAGE

Protein type	Band No.	Protein band intensity (Int.) (AU)	Relative band abundance (%)	Molecular weight (kDa)	
				Predicted	Expected
CLIC1-WT	1	21 813 273	98.4	29.9	26.2
	2	358 986	1.6	25.5	-
CLIC1-C24A	1	23 093 400	97.3	31.1	26.2
	2	373 164	1.6	21.7	-
	3	260 238	1.1	17.1	-
hGSTP1-1	1	2 111 760	3.8	56.4	-
	2	51 800 805	94.0	26.5	23.4
	3	645 246	1.2	20.1	-
	4	260 631	0.5	17.1	-
	5	268 065	0.5	14.6	-
EcSspA	1	422 527	1.6	71.5	-
	2	25 144 056	97.2	30.5	24.3
	3	314 384	1.2	23.5	-

Table S3: Summary of the separation parameters associated with the SE-HPLC profile of the purified target proteins

Parameters	Protein type	Peak 1	Peak 2	Peak 3	Peak 4
Retention time (min)	CLIC1-WT/C24A	14.23 ± 0.07	15.93 ± 0.01	17.23 ± 0.01	23.93 ± 0.01
	hGSTP1-1	16.92 ± 0.00	23.89 ± 0.02	-	-
	EcSspA	17.78 ± 0.01	24.72 ± 0.08	-	-
Predicted size (kDa)	CLIC1-WT/C24A	186.95 ± 0.07	78.13 ± 0.01	39.48 ± 0.01	0.06 ± 0.01
	hGSTP1-1	47 ± 0.00	1.31 ± 0.02	-	-
	EcSspA	30.23 ± 0.01	0.86 ± 0.08	-	-
Integrated area	CLIC1-WT/C24A	125 083.33 ± 10 591.19	157 108.00 ± 6 897.59	5 686 782.67 ± 66 021.80	531 149.67 ± 94 161.37
	hGSTP1-1	35 506 964.67 ± 130 721.91	280 099.67 ± 60 073.14	-	-
	EcSspA	19 939 143.67 ± 10 795 485.02	165 756.67 ± 56 940.95	-	-
Relative abundance (%)	CLIC1-WT/C24A	1.92 ± 0.14	2.42 ± 0.06	87.51 ± 1.27	8.15 ± 1.28
	hGSTP1-1	99.20 ± 0.19	0.80 ± 0.19	-	-
	EcSspA	98.77 ± 0.70	1.23 ± 0.70	-	-
Tailing factor (*S*)	CLIC1-WT/C24A	0.00 ± 0.00	0.00 ± 0.00	1.65 ± 0.03	1.47 ± 0.11
	hGSTP1-1	2.20 ± 0.01	1.35 ± 0.00	-	-
	EcSspA	3.01 ± 0.05	2.16 ± 0.62	-	-
Resolution (R_S)	CLIC1-WT/C24A	0.00 ± 0.00	0.44 ± 0.06	0.58 ± 0.04	3.95 ± 0.03
	hGSTP1-1	0.00 ± 0.00	4.01 ± 0.01	-	-
	EcSspA	2.39 ± 0.01	2.59 ± 0.21	-	-

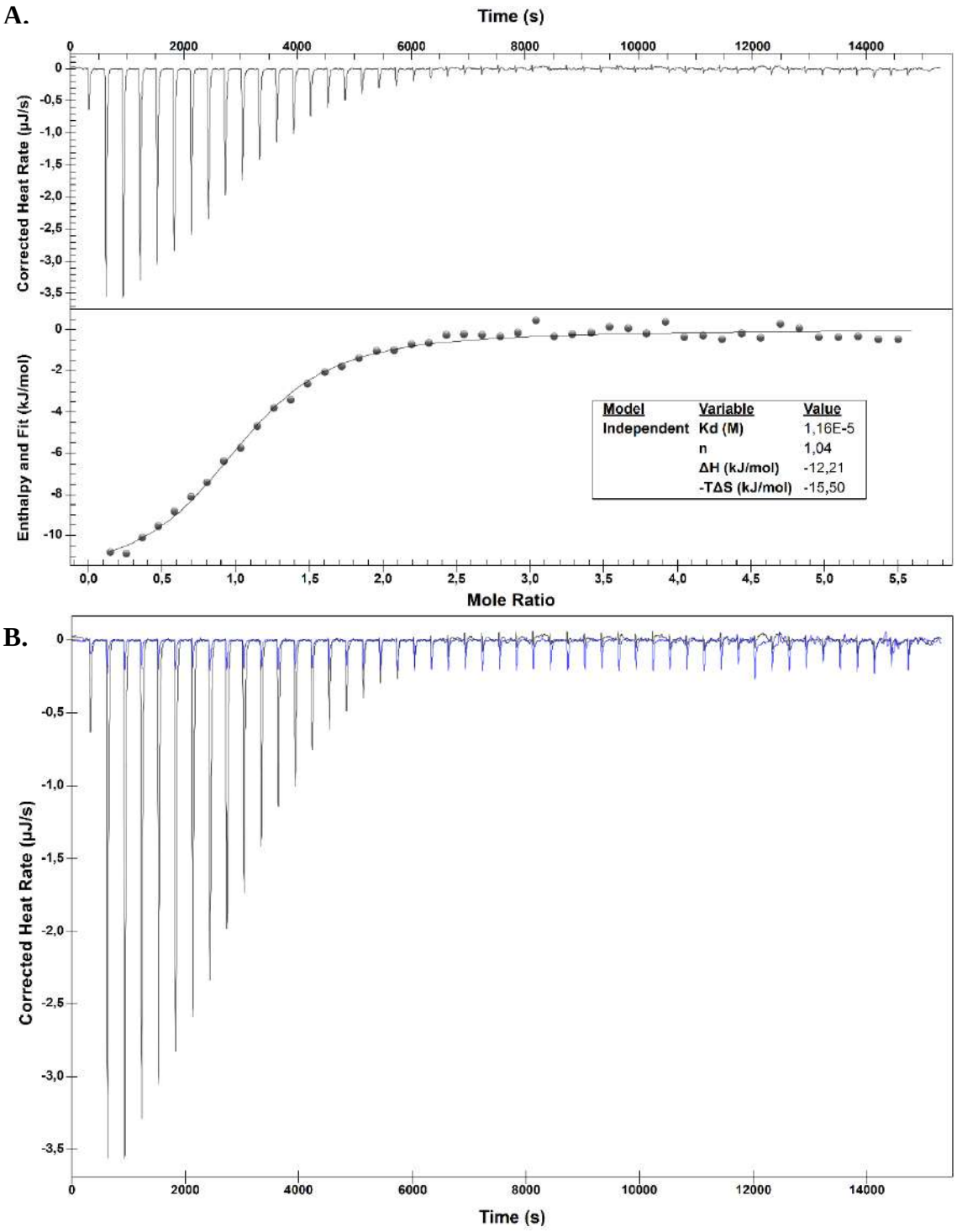
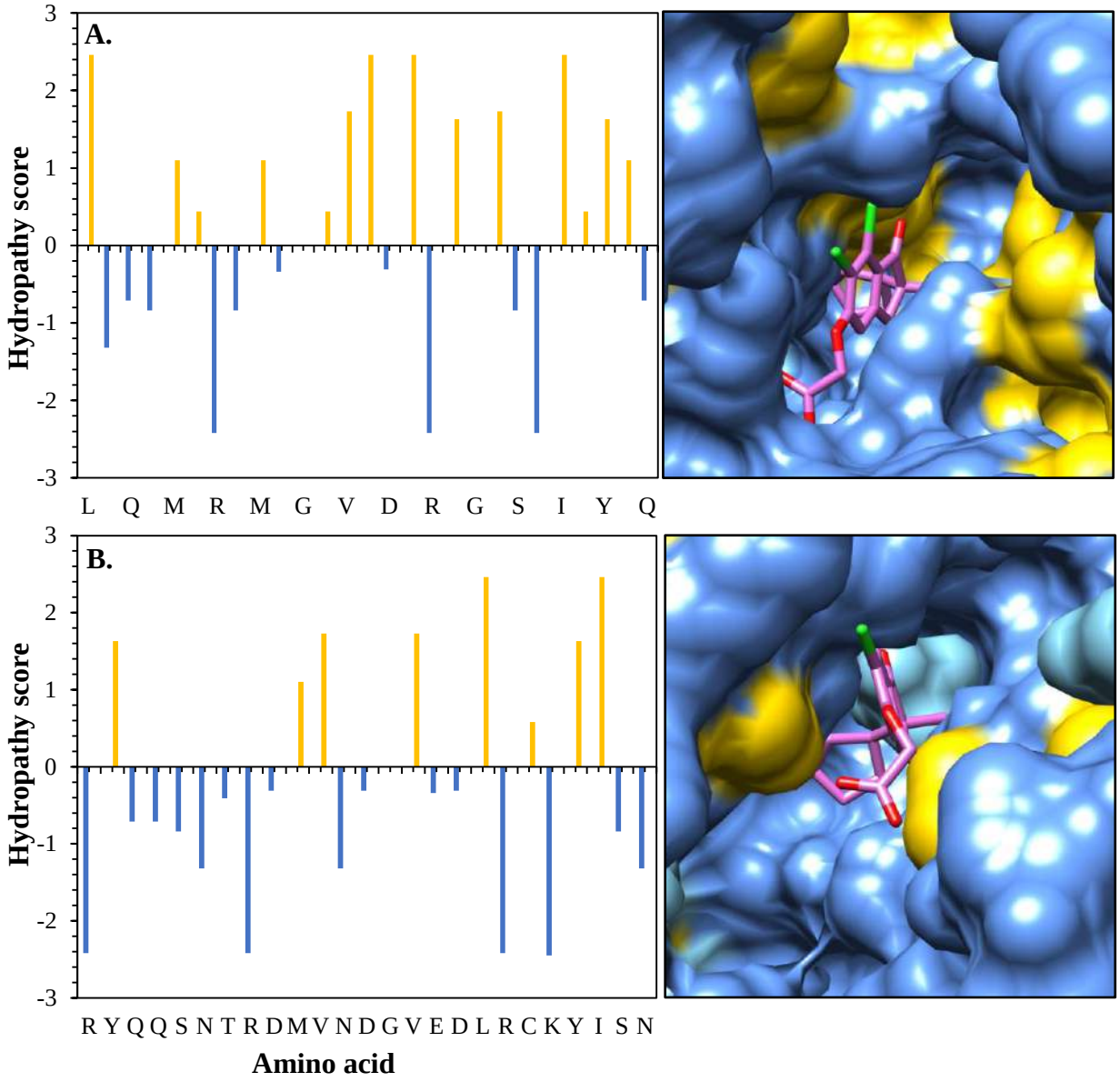


Figure S5. Monitoring the thermodynamic parameters for the interaction between IAA-94 and SjGST. (A) Using ITC, the interaction between IAA-94 and SjGST was monitored by titrating 2.1 mM IAA-94 into 100 µM SjGST in sample buffer (Section 2.2.3.3) (protein provided by Dr Ikechukwu Achilonu). The titration was performed at 20 °C with a stir rate of 250 rpm, an initial and final baseline of 300 s and a duration period of 300 s between each injection. The volume of the ITC syringe used was 250 µl and during each run the initial injection volume was 2 µl, but for all subsequent injections it was 5 µl and therefore a total of 49 injections were performed. The interaction occurred with a stoichiometry of $n \sim 1$, was entropically driven and had a K_d value in the micromolar range. (B) The heat generated from the titration of IAA-94 into SjGST (●) and buffer (●) (HOD) have all been overlaid. *See section 2.2.9 for the experimental details pertaining to this method.*



Protein type	Polar residue hydropathy score	Non-polar residue hydropathy score	Ratio polar/non-polar residue
A. SjGST	- 9.20	13.82	0.67
B. hGSTP1-1	- 18.45	13.32	1.39

Figure S6. Hydropathy plot of residues that are within 8 Å of the dimer interface of SjGST and hGSTP1-1. The polarity of the dimer interface of (A) SjGST and (B) hGSTP1-1 was assessed by an Abraham & Leo hydropathy plot (https://web.expasy.org/protscale/). The polar and non-polar residue hydropathy scores represents the sum of the hydropathy scores for all the individual polar and non-polar residues that were within 8 Å of the dimer interface. The polarity of the site was then assessed by taking a ratio of the polar residue hydropathy score to the non-polar residue hydropathy score. The dimer interface of SjGST was shown to be less polar compared to the dimer interface of hGSTP1-1. Indicated next to each hydropathy plot is the dimer interface which shows IAA-94 (●) and its surrounding polar (●) and non-polar (●) residues. These images were generated by using Chimera V1.11.2.

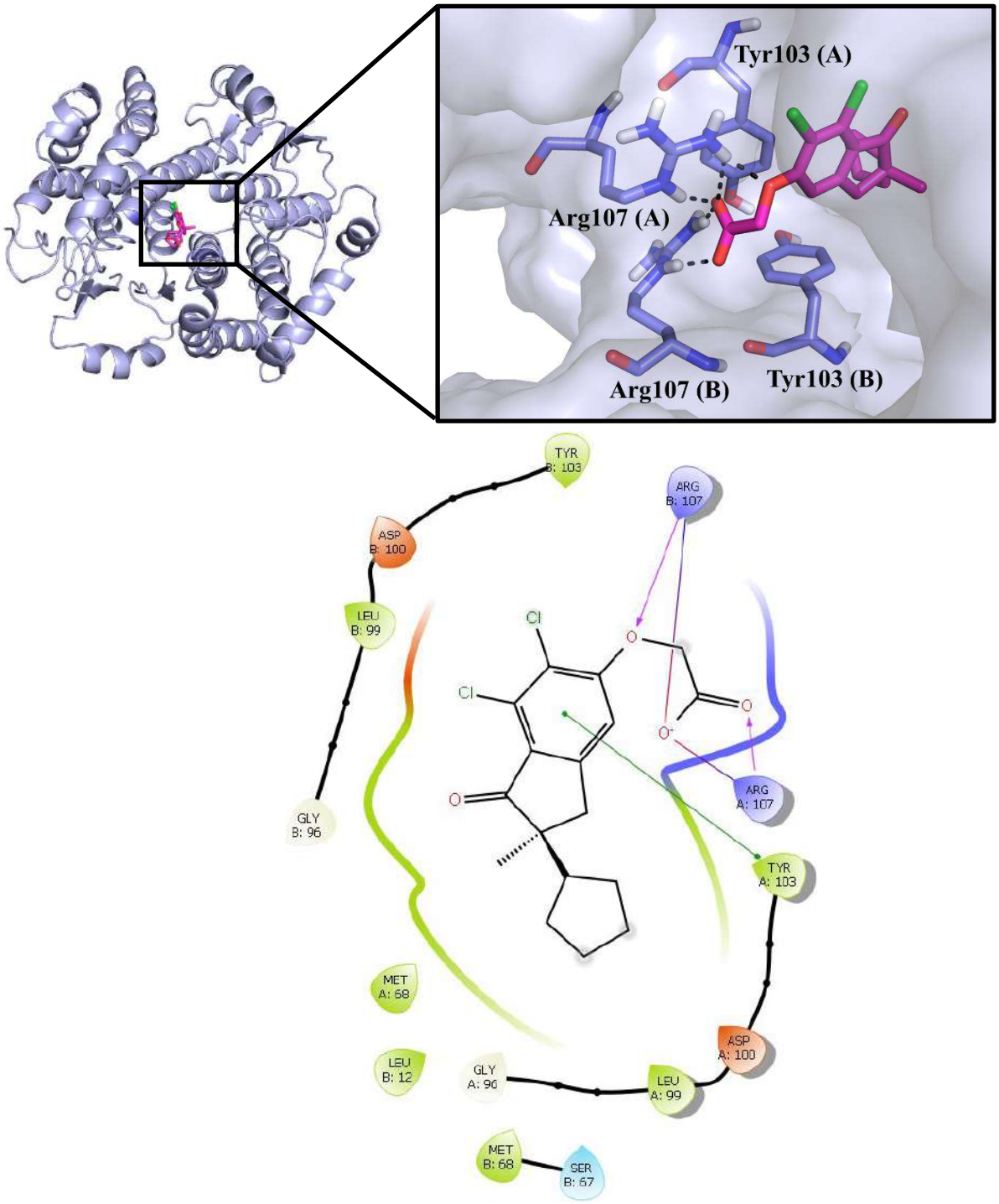


Figure S7. IFMD of IAA-94 at the dimer interface of SjGST. The IFMD of IAA-94 to the dimer interface of SjGST (PDB: 6RWD) was validated by selecting the docked pose which had the lowest IFMD score. IFMD of SjGST in complex with IAA-94 suggested that the acetate moiety of IAA-94 was coordinated by a hydrogen bond with Arg107, while the indenyl moiety of IAA-94 was involved in a π-π stacking interaction with Tyr103. The Glide E_{model}, docking score and components for the G_{score} for the top pose can be found in Table S4. Indicated are positively charged residues (●), negatively charged residues (●), hydrophobic residues (●), polar residues (●), hydrogen bonds (→), salt bridges (—) and π-π interactions (—). The top (3D diagram) and bottom (2D diagram) images for the protein-ligand interaction were generated by using Pymol V1.3 and the 2D protein-ligand interaction diagram tool in Mastero 11.2, respectively. *See section 2.2.13 for the experimental details pertaining to this method.*

Table S4: Summary of the docking functions obtained from IFMD

Protein-ligand interaction	GEmodel	IFMD score	Gscore	Glipo	Ghbond	Gvdw	Gcoul	GRotB	Gsite	Gligand efficiency	No. of rotatable bonds
	kcal.mol^{-1}										
○ CLIC1 (C24) + IAA-94	- 47.942	- 494.816	- 6.241	- 2.215	-0.320	- 28.117	- 6.399	0.452	- 0.004	- 0.271	4
◇ CLIC1 (F41) + IAA-94	- 37.602	- 495.025	- 4.581	- 2.140	-0.200	- 27.888	- 1.872	0.452	- 0.038	- 0.199	4
○ CLIC1 (C24) + EAA	- 41.089	- 492.702	- 5.064	- 1.580	- 0.452	- 18.596	- 9.299	1.008	- 0.049	- 0.267	6
◇ CLIC1 (F41) + EAA	- 29.728	- 492.891	- 3.783	- 1.422	- 0.278	- 22.425	- 1.429	1.008	- 0.094	- 0.199	6
○ CLIC1 (C24) + 6-IAF	- 42.807	-	- 3.524	- 0.587	- 0.253	- 25.176	- 9.766	0.109	0	- 0.110	4
○ hGSTP1-1 (Y108) + IAA-94	- 56.483	- 893.026	- 6.919	- 2.631	- 0.673	- 26.872	10.163	0.452	- 0.045	- 0.301	4
◇ hGSTP1-1 (W28) + IAA-94	- 35.445	- 888.708	- 4.341	- 1.172	- 0.498	- 20.367	- 7.288	0.452	- 0.031	- 0.189	4
○ hGSTP1-1 (Y108) + EAA	- 57.673	- 891.573	- 6.310	- 1.967	-0.680	- 26.318	11.886	1.008	- 0.085	- 0.332	6
◇ hGSTP1-1 (W28) + EAA	- 38.987	- 887.289	- 4.545	- 0.657	- 0.947	- 19.170	10.449	1.008	- 0.073	- 0.239	6
○ hGSTP1-1 (Y108) + MES	- 41.773	- 891.122	- 6.173	- 0.944	- 0.450	- 13.088	15.358	0.429	- 0.043	- 0.494	3
◇ hGSTP1-1 (E197) + MES	- 27.726	- 889.292	- 4.292	- 0.142	- 0.480	- 12.061	- 8.950	0.425	- 0.071	- 0.695	3
○ hGSTP1-1 (W28) + IAA-94 + MES [1]	- 43.672	- 892.568	- 5.367	- 2.240	- 0.347	- 27.673	- 4.153	0.452	- 0.022	- 0.233	4
○ hGSTP1-1 (W28) + IAA-94 + MES [2]	- 32.461	- 897.228	- 4.346	- 2.002	- 0.160	- 21.219	- 3.535	0.452	0	- 0.189	4
○ hGSTP1-1 (W28) + ANS	- 49.705	- 892.138	- 5.766	- 2.603	- 0.200	- 30.622	- 5.764	0.412	- 0.045	- 0.275	3
◇ hGSTP1-1 (Y108) + ANS	- 38.345	- 888.296	- 4.489	- 1.288	- 0.160	- 23.887	- 4.942	0.412	- 0.142	- 0.214	3
○ EcSspA (Y21) + IAA-94	- 40.672	-851.818	- 4.480	- 1.218	- 0.720	- 26.204	- 3.704	0.452	- 0.100	- 0.195	4
◇ EcSspA (F36) + IAA-94: Pose 1	- 28.752	- 849.642	- 4.153	- 1.362	- 0.329	- 20.438	- 4.009	0.452	0	- 0.181	4
Pose 3	- 28.920	- 849.449	- 3.823	- 1.035	- 0.360	- 18.365	- 5.557	0.452	0	-0.166	4
○ EcSspA (Y21) + EAA	- 39.357	- 851.320	- 4.682	- 0.771	- 0.530	- 27.421	- 3.129	0.706	- 0.077	- 0.246	6
◇ EcSspA (F36) + EAA	- 53.239	- 853.064	- 5.901	- 1.912	- 0.720	- 27.997	- 5.239	1.008	- 0.069	- 0.311	6
○ EcSspA (Y21) + ANS	- 43.882	- 852.454	- 5.536	- 1.974	- 0.160	- 26.354	- 6.347	0.412	- 0.169	- 0.264	3
◇ EcSspA(F36) + ANS	- 45.086	- 853.443	- 5.338	- 2.038	- 0.124	- 29.116	- 4.148	0.412	- 0.136	- 0.254	3
□ SjGST (Arg107) + IAA-94	- 71.068	-921.551	- 7.562	- 2.562	- 0.820	- 32.851	- 7.722	0.452	- 0.100	- 0.329	4

(x): The amino acid residue which was selected to define the site where IFMD will occur.

○: The ligand was docked to the EAA (putative EAA) site of the target protein.

◇: The ligand was docked to the MES (putative IAA-94) site of the target protein.

□: The ligand was docked to the dimer interface of the target protein.

[1]: MES which had previously been docked to the target protein.

[2]: MES which was from the crystal structure of the target protein.

Table S5: List of additional conditions used during the crystal trials of hGSTP1-1 in complex with IAA-94

Condition 1	Condition 2	Condition 3
100 mM MES (pH 6.0) 350 mM calcium acetate 22% (w/v) PEG-6000	100 mM MES (pH 6.0) 20 mM calcium chloride 22 % (w/v) PEG-6000	100 mM MES (pH 6.0) 200 mM calcium acetate 20 % (w/v) PEG-6000
Condition 4	**Condition 5**	**Condition 6**
100 mM MES (pH 5.4) 270 mM calcium acetate 20 % (w/v) PEG-6000	50 mM MES (pH 5.5) 20 % (w/v) PEG-6000	100 mM MES (pH 6.5) 16 % (w/v) PEG-6000
Condition 7	**Condition 8**	**Condition 9**
100 mM MES (pH 6.0) 270 mM calcium acetate 20 % (w/v) PEG-6000	100 mM MES (pH 5.4) 200 mM ammonium sulfate 10 % (v/v) glycerol	100 mM MES (pH 5.4) 270 mM calcium acetate 20 % (w/v) PEG-6000
Condition 10	**Condition 11**	**Condition 12**
100 mM MES (pH 5.5) 20 mM calcium chloride 28 % (w/v) PEG-6000	100 mM MES (pH 6.0) 22 % (w/v) PEG-6000 20 mM calcium chloride	100 mM MES (pH 6.5) 200 mM sodium acetate 30 % (w/v) PEG-6000
Condition 13	**Condition 14**	**Condition 15**
100 mM MES (pH 5.8) 20 % (w/v) PEG-6000 20 mM calcium chloride	100 mM HEPES (pH 7.5) 200 mM sodium chloride 25% (w/v) PEG-6000	100 mM MES (pH 5.5) 25 % (w/v) PEG-6000

Condition 16	Condition 17	Condition 18
100 mM MES (pH 6.4) 1.9 M ammonium sulfate 5 % (v/v) glycerol	100 mM MES (pH 6.4) 2.1 M ammonium sulfate 10 % (w/v) PEG-6000	100 mM MES (pH 6.4) 2.1 M ammonium sulfate 20 % (w/v) PEG-6000

Condition 19	Condition 20	Condition 21
100 mM MES (pH 6.4) 200 mM calcium acetate 20 % (w/v) PEG-6000	50 mM imidazole (pH 7.0) 20 % (w/v) PEG-6000	100 mM MES (pH 6.0) 22 % (w/v) PEG-6000 20 mM calcium chloride

Condition 22	Condition 23	Condition 24
100 mM MES (pH 6.0) 20 % (w/v) PEG-6000 20 mM calcium chloride	100 mM MES (pH 5.8) 20 % (w/v) PEG-6000 20 mM calcium chloride	100 mM MES (pH 5.5) 20 % (w/v) PEG-6000

Condition 25	Condition 26	Condition 27
100 mM MES (pH 5.5) 25 % (w/v) PEG-6000	100 mM MES (pH 6.4) 200 mM calcium acetate 20 % (w/v) PEG-6000	100 mM MES (pH 6.5) 200 mM sodium acetate 30 % (w/v) PEG-6000

Condition 28	Condition 29	Condition 30
100 mM K_2HPO_4 (pH 6.6) 100 mM ammonium sulfate	100 mM Bis-tris (pH 6.5) 20 mM calcium chloride 30% (v/v) PEG-6000	50 mM imidazole (pH 7.0) 20 % (w/v) PEG-6000

Table S6: Molprobity statistics for the structure of hGSTP1-1 in complex with IAA-94

All-atom contacts	Clashscore, all atoms[a]	3.31	97th percentile[b] (N = 474, 1.402 Å ± 0.25 Å)	
Protein geometry	Poor rotamers	2	0.28 %	Goal: < 0.3 %
	Favoured rotamers	692	96.92 %	Goal: > 98 %
	Ramachandran outliers	0	0.00 %	Goal: < 0.05 %
	Ramachandran favoured	807	97.58 %	Goal: > 98 %
	MolProbity score[c]	1.19	97th percentile[b] (N = 3 307, 1.402 Å ± 0.25 Å)	
	Cβ deviations > 0.25 Å	7	0.90 %	Goal: 0 %
	Bad bonds	102/6843	1.49 %	Goal: <0 %
	Bad angles	76/9300	0.82 %	Goal: < 0.1 %
Peptide Omega	Cis Prolines	5/39	12.82 %	Expected: ≤ 1 per chain, or ≤ 5%

[a]: The clashscore metric indicates how many steric clashes per 1000 atoms are present in the model. Therefore, the lower the clashscore, the better the model.

[b]: 100th percentile is the best among structures of comparable resolution while the 0th percentile is the worst.

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

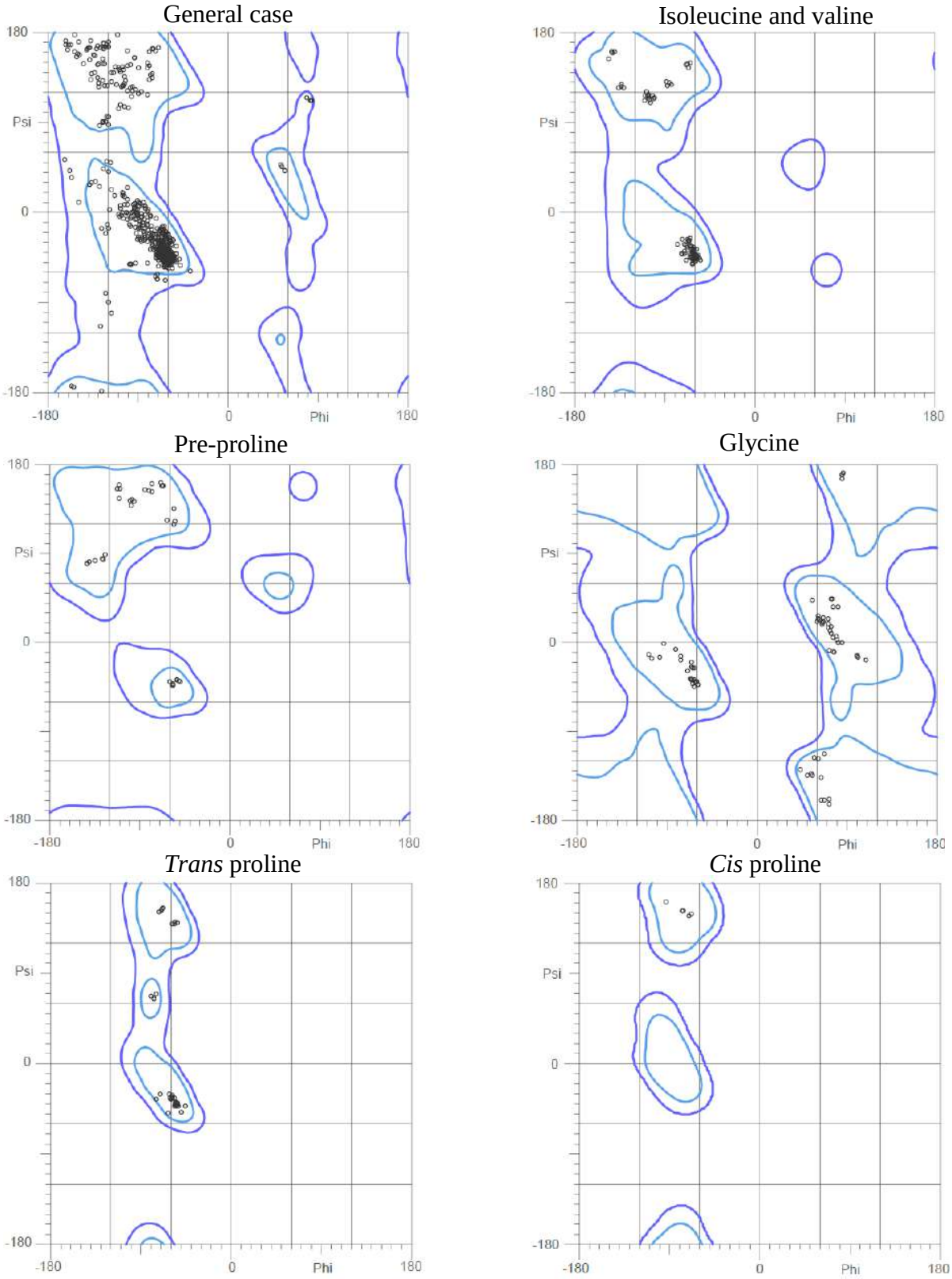


Figure S8. Analysis of the structural geometry of hGSTP1-1 in complex with IAA-94. The Ramachandran plot for the refined structure of hGSTP1-1 in complex with IAA-94 indicated that only 0.24 % of the amino acids were found in the disallowed or unfavourable regions on the Ramachandran plot and therefore there were no outliers. Instead, 97.70 % of the residues were found in the favourable regions and 2.06 % of the residues were found in this generously allowed regions. The Ramachandran plot was obtained from Molprobity (http://molprobity.biochem.duke.edu/).

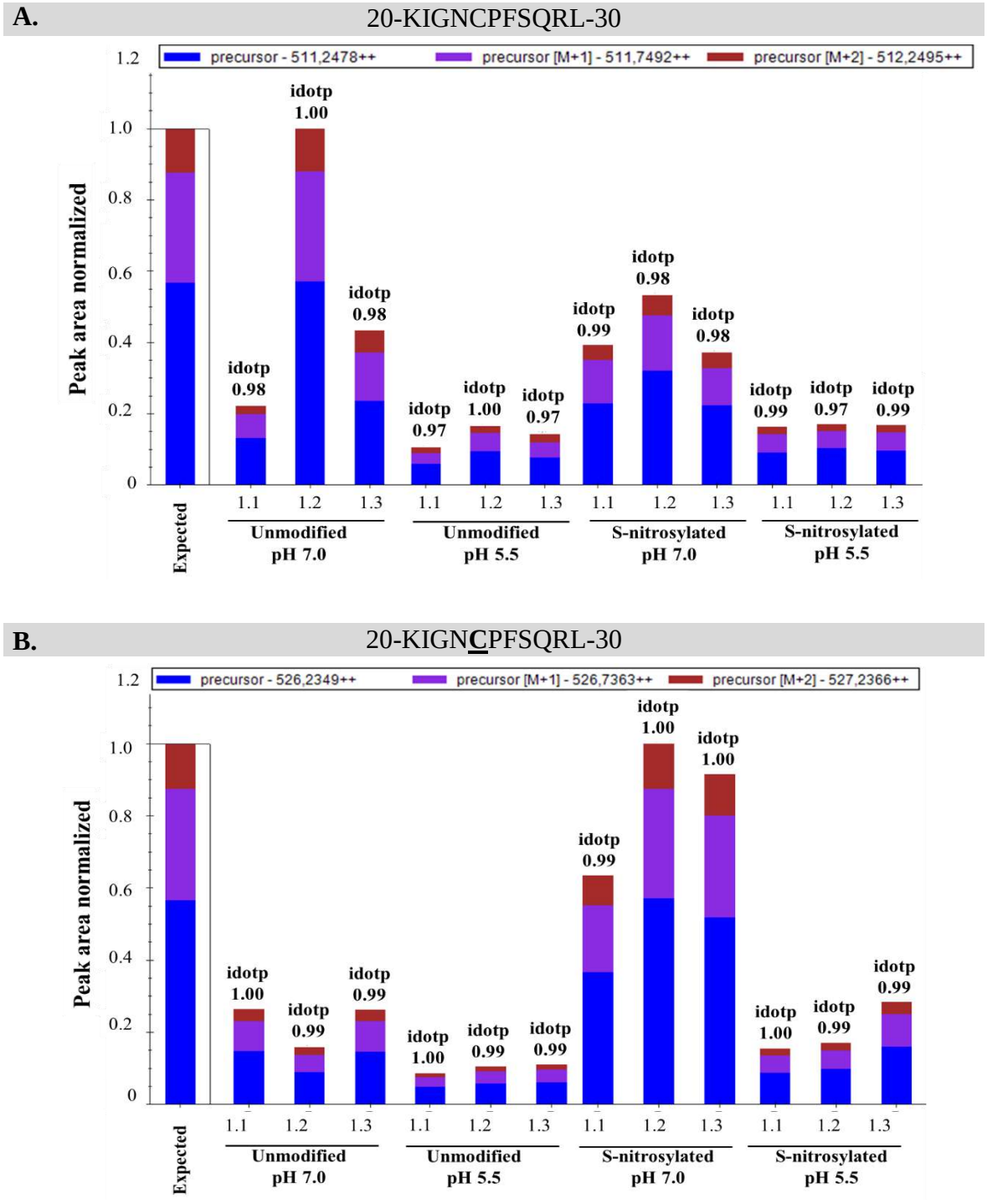


Figure S9. The IDOTP scores reported for the *S*-nitrosylated peptide from CLIC1-WT under both pH conditions. Depending on the chemical formula for a peptide, the IDOTP score will compare the predicted and observed isotope distribution based on the integrated peak area for each isotope in the chromatogram. The IDOTP score for both the (A) unmodified and (B) *S*-nitrosylated peptide from every sample was ≥ 0.97, demonstrating that both the observed and predicted isotope distribution correlated with one another. Indicated are the first three isotope peaks for the precursor M (●), M+1 (●) and M+2 (●) from three individual samples of unmodified and *S*-nitrosylated CLIC1-WT under both pH conditions. All LC-MS/MS data were analysed using the Skyline-daily 3.7.1.11357 MacCoss Lab software ©.

598272:Roland_Worth-_Complete_PhD_Thesis(Turnitin).pdf

ORIGINALITY REPORT

18%	8%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Witwatersrand Student Paper	3%
2	Submitted to Stourbridge College Student Paper	1%
3	"Award Winners and Abstracts of the 33rd Annual Symposium of The Protein Society; Seattle, Washington, June 30-July 3, 2019", Protein Science, 2019 Publication	1%
4	d-nb.info Internet Source	<1%
5	www.uni-giessen.de Internet Source	<1%
6	www.hgmd.cf.ac.uk Internet Source	<1%
7	etheses.whiterose.ac.uk Internet Source	<1%
8	hdl.handle.net Internet Source	<1%

28 April 2020

Re: Turn-it-in report for final PhD thesis (R Worth 598272)

The similarity index value of 18% obtained for the Turn-It-In analysis of Roland's corrected PhD thesis has a 9% component based on the work he submitted to Wits previously when he also worked on a CLIC1 project. I am satisfied that when this and that there is a <1% similarity for the individual references he cited are taken into account, the similarity index value is acceptable.

Regards

HW Dirr
Supervisor
Retired Professor Emeritus
Protein Structure-Function Research Unit
School of Molecular and Cell Biology
University of the Witwatersrand
Johannesburg 2050
South Africa
E-mail: heinidirr@gmail.com