THE ROLE OF THE DOMAIN INTERFACE IN THE STABILITY, FOLDING AND FUNCTION OF CLIC1

STOYAN HRISTOV STOYCHEV

A thesis submitted to the Faculty of Science, University of Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, 2008

DECLARATION

I declare this thesis my own, unaided work. It is being submitted for the degree of Doctor of Philosophy to the University of Witwatersrand, Johannesburg. It has not been submitted before for any other degree or examination at any other University.

_____

Stoyan Stoychev

_____ day of _____ 2008

DEDICATIONS

THIS WORK IS DEDICATED TO MY MOM AND DAD. THE PEOPLE THAT HAVE ALWAYS SUPPORTED ME AND STOOD BY MY SIDE.

ABSTRACT

Chloride intracellular channel protein 1 (CLIC1) is a dual-state protein existing in both soluble monomeric conformation as well as integral-membrane form. The role of the domain interface in the conversion between these species was investigated. Bioinformatics-based analysis was undertaken to compare and contrast the domain interfaces of dimeric GSTs with their monomeric homologues CLIC1 and CLIC4. The mutants CLIC1-M32A and CLIC1-E81M were used as experimental case studies on the role of domain-domain interactions in the stability and folding of CLIC family proteins. A consensus interface was revealed with the prominent interaction being a conserved inter-domain lock-and-key type motif previously studied in class Alpha GSTs (Wallace *et al.,* 2000). A number of domain-interface interactions were found to be unique to the CLIC family and as such thought to play a role in the conversion of these proteins from their soluble form to an integral membrane form. Overall the domain interfaces of monomeric CLIC1 and CLIC4 did not differ significantly from the domain interfaces of dimeric GSTs. The removal of the unique CLIC family salt-bridges between Arg29 and Glu81 and the cavity forming domain interface mutation Met32Ala did not induce significant changes in the conformational flexibility of the native state. The true role of the Arg29-Glu81 salt-bridges was masked by the introduction of stabilizing hydrophobic contacts. Removal of the inter-domain lock-and-key interaction destabilized CLIC1 significantly with concomitant loss in cooperative folding that resulted in the stabilization of a molten globule-like species. This intermediate state was less stable and less structured than the equilibrium intermediate of wtCLIC1 at pH 5.5. However the bulk of the structures found to unfold during intermediate-species formation was the same in mutant and wild-type proteins. It was concluded that formation of the membrane-competent form of CLIC1 involves re-structuring of the N-terminal thioredoxin domain that takes place after destabilization of the salt bridges connecting h1 and h3 and uncoupling of the inter-domain lock-and-key motif.

ACKNOWLEDGEMENTS

I acknowledge my supervisors, Prof. Heini Dirr and Dr. Jonathan Burke, and thank them for the support and guidance they have provided. I also thank my colleagues and friends in the Protein Structure-Function Research Unit. I would like to recognize the contribution of Dr. Sylvia McIntyre, Dr. Louise Wallace, Dr. Chris Nathaniel and Darren Legg from the Protein Structure-Function Research Unit, allowing me access to their research findings. I also acknowledge Dr. Virgil Woods, School of Medicine, University of California, and Dr. Marietjie Stander, Stellenbosch University, South Africa, for their excellent advice and supervision in relation to Mass Spectrometry experiments. Special thanks to Anwar Rasool, ABSA bank systems developer, for developing the Biochem-MFC and ColorSol-MFC software applications.

I acknowledge the University of Witwatersrand and the National Research Foundation for financial support.

TABLE OF CONTENTS

ABBREVIATIONS

ACN	acetonitrile
$\lambda_{em\ max}$	emission wavelength maximum
AFU	autonomous folding unit
aMb	apomyoglobin
ANS	8-Anilino-1-napthalene sulfonate
Arg29	arginine 29 in CLIC1
ASA^A	average accessible surface area
A280	absorbance at 280nm
bb-bb	hydrogen bond between peptide backbone atoms
bb-sc	hydrogen bond between peptide backbone donor/acceptor atom and side chain acceptor/donor atom
Cα	alpha carbon
CID	collision induced dissociation
C-interface	c-terminal domain interface
CLIC1	chloride intracellular channel 1
C_m	transition midpoint
Cri	conservation ratio
DTT	dithiothreitol
DXMS	deuterium exchange mass spectrometry
E222	ellipticity at 222 nm
E. coli	*escherichia coli*
ESI-MS	electrospray mass spectrometry
FD	fully-deuterarted
F280	fluorescence emission intensity at 280 nm
F347	fluorescence intensity at 347 nm
F358	fluorescence intensity at 358 nm
F390	fluorescence emission intensity at 390 nm
F470	fluorescence emission intensity at 470 nm
F358/F347	the ratio of the fluorescence emission intensity at 358 nm to the fluorescence emission intensity at 347 nm

ESI	electrospray ionization
Far-UV CD	far ultraviolet circular dichroism
ΔG	gibbs free energy
ΔG_{H20}	change in free energy of unfolding in the absence of denaturant
Glu81	glutamic acid 81 in CLIC1
Grx2	glutaredoxin 2
GSH	reduced glutathione
GST	glutahtione transferase
hGSTA1-1	human class α glutathione transferase
ΔH	enthalpy change
H^+	hydrogen ion
I	intermediate conformation
iASA	interface accessible surface area
IPTG	isopropyl-D-thiogalactoside
LDW	low density water
LUVs	large unilamellar vesicles
***m*-value**	dependence of free energy as a function of denaturant concentration
M32A	methionine 32 mutation to alanine in CLIC1
MG	molten globule
$[M-H]^+$	protonated protein
$[M\text{-adduct}]^+$	protein-adduct complex
$[M+CH_3CN]^+$	protein-acetonitrile adduct complex
$[M+K]^+$	protein-potassium adduct complex
$[M+Na]^+$	protein-sodium adduct complex
$[M-H+2Na]^+$	protein double-sodium adduct complex
$[M+Na+CH_3CN]^+$	protein-sodium-acetonitrile adduct complex
$[M+NH_4]^+$	protein-ammonium adduct complex
MS	mass spectrometry
N	native conformation
$[N+Na]^+$	native protein-sodium adduct complex
$[N+NH_4]^+$	native protein-ammonium adduct complex
ND	non-deuterated
N-interface	n-terminal domain interface

kDa	kilodalton
ORF	open reading frame
PDB	protein data bank
pI	isoelelectric point
rGSTK1-1	rat class Kappa glutathione transferase
Q-Tof	quadropole time-of-flight
Rms	root mean square
Rmsd	root mean square of deviation
sc-sc	hydrogen bond between residues' side chain atoms
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SCOP	structural classification of proteins
Sj	schistosomal
ΔS	entropy change
ΔS^{conf}	change in conformational entropy
ΔS^{solv}	change in solvation entropy
TFA	trifluoroacetic acid
TMD	transmembrane domain
Trp 35	tryptophan 35 in CLIC1
U	unfolded conformation
Uf	uniqueness factor
wtCLIC1	wild-type chloride intracellular channel 1

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In order to fully comprehend the mechanisms behind protein folding and function one needs to account for both intra-molecular interactions as the polypeptide chain associates with itself and inter-molecular interactions as it associates with other domains and/or subunits. The model system used in this study, CLIC1, has been shown to exist in dual form. A soluble conformation where the protein takes on a fold homologous to the cytosolic glutathione transferases (GSTs) and a membrane-integral state where the protein functions as an anion channel. This study intends to investigate the role and contribution of domain-domain contacts in protein stability and folding with the specific aim of identifying domain-interface interaction/s that may be responsible for the conversion of CLIC1 from soluble to membrane-inserted form.

1.1 Domains and Domain interfaces: defining the terms

1.1.1 Domains

Domains have been shown to act as units of protein function by forming protein substructures with distinct functional properties or by completing active sites through domain interactions. Domains also act as units of protein evolution through domain stealing, swapping, and addition, as well as units of protein structure by acting as the building blocks of oligomeric proteins (Jaenicke, 1999). Based on these different properties, there are seven definitions used to classify domains (Jaenicke, 1999). Part of this study will address the role of the domain-domain interface in protein stability and folding. Therefore, the following definitions are based on the folding and structural features of domains.

According to Wetlaufer (1973), domains are stable units of protein structure that fold autonomously, thereby playing a central role as intermediates during folding. Note, however, that not all domains are autonomous folding units (AFUs). A more accurate definition is given by Richardson who describes domains as compact, local, semi-independent structural units (Richardson, 1981). In terms of folding, domains have been defined as cooperative thermodynamic units, detectable by distinct

folding/unfolding transitions and are separable by hydrodynamic and spectroscopic measurements (Privalov, 1979).

There are a number of algorithms used to identify domains from three-dimensional structures of proteins. The algorithm DOMAK (Siddiqui and Barton, 1995) locates domains on the basis that a domain has shorter residue-to-residue distances within itself than with the rest of the protein. DETECTIVE (Swindells, 1995) is based on the concept that each domain has a hydrophobic core or cluster. The domain assignment algorithm developed by Zehfus (1997) is based on domain compactness as well as hydrophobic cores.

Combining Wetlaufers, Richardsons and Privalovs definitions of a domain, and the different criteria used by algorithms for domain assignment, domains can be described as compact, semi-independent substructures of proteins that fold cooperatively and in some cases autonomously. Domains have hydrophobic cores that make more contacts with themselves than with the rest of the protein. Domains can be considered as the building blocks of proteins. They allow folding to occur at multiple sites along the polypeptide chain (folding-by-parts) and thus enhance the folding rate.

1.1.2 Domain interfaces

The definition of domains allows for two possibilities in bi-/multi-domain proteins: (i) domains may be independent units or, (ii) they may strongly interact with each other. In the second case all interaction would occur through domain interface residues. The domain interface describes the surface area buried upon domain association as well as the contacts formed between an interacting pair of domains. Therefore, in order to fully comprehend the role of domain interfaces in stability and folding one needs to classify the types of interactions that occur there.

1.2. Forces at the domain interface

The overall stabilizing effect induced from domain interface contacts, similarly to protein stability, is a balance between interactions that favour domain association and interactions that favour domain dissociation. This difference is represented by the change in Gibbs free energy ($\Delta G°$) upon total unfolding. Under constant pressure, $\Delta G°$ is made up of two contributions:

$$\Delta G° = \Delta H° - T\Delta S° \quad (1)$$

where ΔH° is the enthalpic (bond formation) and ΔS° is the entropic (freedom of a system to explore conformational space) contribution to ΔG°. Since the association of domains results in a more compact state than their dissociation, conformational energy is lost resulting in a force that opposes folding. On the other hand, inter- and intra-molecular interactions formed upon domain association favour folding. The dominant energy contributors to ΔH° are van der Waals interactions, electrostatic interactions and the hydrophobic effect (discussed in sections **1.2.1** to **1.2.3**). The effect of ΔS° on domain association is discussed in section **1.2.4**. The overall balance between stabilizing and destabilizing forces results in a favourable ΔG° of 5 - 20 kcal/mol. This small energy difference between the native and unfolded states is advantageous to proteins since the acquisition of different conformations only requires a minute energy change possibly induced by changes in the protein's environment i.e. differences in cytosolic and membrane pH. This allows proteins to be flexible and perform a number of different functions.

This study deals with inter-chain interactions between domains. Hence, it is important to distinguish between local and non-local contacts. Local contacts are formed between connected, in sequence, neighbouring or near neighbouring residues (Dill, 1990). Non-local interactions, like those found in domain and dimer interfaces, occur between amino acids that are significantly apart in sequence (Dill, 1990). It is also essential to differentiate between local/non-local and short-/long-range interactions. The short/long-range contacts refer to the dependence of bond energies on distance/radius (r^{-p}). Thus, contacts are defined as short-range if $p > 3$, i.e., van der Waals contacts, and long-range if $p \leq 3$ (Dill, 1990) i.e. hydrogen bonds, electrostatic interactions and hydrophobic contacts.

1.2.1 Van der Waals interactions and hydrogen bonds

Van der Waals interactions are non-covalent and occur between all atoms, both polar or non-polar (Lins and Brasseur, 1995). These interactions form due to the asymmetric distribution of electronic charges, resulting in fixed or induced dipoles (Dill, 1990). The strength of van der Waals interactions is directly proportional to the polarizabilities of the atoms and inversely proportional to the sixth power of the distance between them (Pace, 2001). Hence, van der Waals interactions are short-

ranged and are only functional at very small distances. Therefore, steric complementarity in the interior, domain and/or dimer interface of proteins is needed to maximise the stabilizing effects of these forces. In fact, recent studies have proposed that the loss of packing interactions rather then the hydrophobic effect dominates protein stability (Ratnaparkhi and Varadarajan, 2000).

Hydrogen bonds form between polar molecules when a hydrogen atom is shared between two electronegative atoms such as oxygen, nitrogen or sulphur. The strength of a hydrogen bond is dependent on the electronegativity and orientation of the bonding atoms and varies between 2-10 kcal/mol (Dill, 1990). Furthermore, hydrogen bonds have been shown to be cooperative (Stickle *et al*., 1992). A hydrogen bond to a peptide backbone carbonyl group will strengthen a second hydrogen bond to the amide group of the same peptide bond (Stickle *et al*., 1992). Hydrogen bond contribution toward protein stability is further enhanced by the existence of cooperative networks where a donor or acceptor participates in a number of hydrogen bonds (Stickle *et al*., 1992). In addition to the peptide backbone, polar amino acids have the potential to form hydrogen bonds due to the presence of amine, carbonyl, thiol, and hydroxyl groups. In terms of domain interfaces, most hydrogen bond contributions are due to polar residues, since these contacts are mostly non-local, while backbone hydrogen bonds are mainly local. The fact that the peptide backbone and polar residues can form hydrogen bonds once again highlights the importance of correct packing in the interior, domain and dimer interfaces of proteins. In the folded conformation, specific intra-molecular hydrogen bonds must form to replace the fluctuating intra- and inter-molecular bonds that form in the unfolded protein so that the native state is enthalpically favoured. Although hydrogen bonds can provide a significant contribution toward protein stability, they are not the dominant folding force. If they were, solvents that form strong hydrogen bonds with the protein should unfold it, while solvents that do not form or form weak hydrogen bonds with the protein should not affect or should not stabilize the native state (Dill, 1990). No such correlation has been observed.

1.2.2 Electrostatic interactions

Electrostatic contacts, in proteins, occur between charged residues with the strength of interactions related by Coulomb's law:

$$F = (k*q_1*q_2)/Dr^2 \qquad (2)$$

where ***F*** is the force between two electrical charges q_1 and q_2 that are separated by a distance ***r***. ***k*** is the proportionality constant (***k*** = 2.14×10^9 cal.m.C^{-2}) while, ***D*** is the dielectric constant of the medium. From equation **(2)** one can see that an increase in dielectric constant will decrease the strength of the electrostatic interactions. Hence, a more polar environment, such as at a proteins' surface, will result in weaker interactions. On the other hand, a more non-polar environment such as that in the protein interior, at domain and dimer interfaces of proteins will strengthen electrostatic interactions.

Charge contacts, in proteins, can be classified into two groups. 'Classical' electrostatic effects occur due to non-specific repulsions between surface residues with the same charge destabilizing proteins (Dill, 1990). The extent of destabilization is affected by ionic strength and pH. An increase in ionic strength in the protein's environment results in better shielding of opposite charges, which decreases repulsions and stabilizes the protein. A pH increase or decrease will result in an increase of the protein's net charge, leading to more charge repulsions along its surface and destabilization. Therefore, the majority of proteins are most stable at a pH that is close to their pI, where their net charge will be zero.

The second group of electrostatic interactions is the specific charge contacts known as ion pairing or salt-bridges. Salt-bridges are formed between the acidic/negatively charged aspartic acid (Asp) and glutamic acid (Glu) residues and the basic/positively charged amino acids arginine (Arg), lysine (Lys) and histidine (His). Most salt-bridges are stabilizing although there have been cases where destabilizing ion pairing has been reported (Kumar and Nussinov, 1999). The energy contribution of salt-bridges varies between 5-15 kcal/mol/ion pair according to their geometry, location in the protein, whether they are isolated or networked, if they are hydrogen bonded or not (Kumar and Nussinov, 1999). In spite of this energy gain, the number of charged residues and hence the number of ion pairs in the interior, domain and dimer interfaces of proteins is relatively low compared to the number of hydrophobic residues. The reason probably lies in the large energy cost (~19 kcal/mol), known as Born energy, needed for the transfer of a charged ion from a polar to a non-polar

environment (Dill, 1990). As a result, salt-bridges in protein interiors, domains and dimer interfaces are thought to play a role in packing specificity. The correct packing in non-polar environments is essential since burial of salt-bridges with bad geometries and uncompensated/unpaired ionisable groups will result in protein destabilization (Kumar and Nussinov, 1999; Hendsch and Tidor, 1994).

1.2.3 Hydrophobic effect

Hydrophobicity has been defined as the non-polar transfer from an aqueous to a non-aqueous media (Dill, 1990). It is widely believed that the hydrophobic effect is the major force behind protein folding and stability. This is based on the fact that non-polar molecules contribute favourable $\Delta G°$ values upon their transfer from water to organic solvent, with the magnitude of $\Delta G°$ increasing with an increase in the buried hydrophobic surface area (Ratnaparkhi and Varadarajan, 2000). It is thought that protein folding is driven by the unfavourable solvation entropy of hydrated non-polar residues in the unfolded state rather than favourable interactions between non-polar groups in the folded conformation (Lins and Brasseur, 1995). The unfavourable solvation entropy arises from the 'structuring' of water molecules around hydrophobic side chains into ordered lattices to form strong water-water hydrogen bonds. Upon folding, most hydrophobic side chains are buried in the non-polar protein interior, domain or dimer interface, while most polar residues are at the protein's surface. In addition, most soluble proteins assume a globular conformation that gives the lowest surface area to volume ratio, thus minimizing the protein's surface exposed to solvent. Therefore, in the native conformation water is less structured around the protein than in the unfolded conformation resulting in positive solvation entropy that favours folding and stabilises the native state.

A study of the affinity of amino acids for aqueous solvents showed that the side chain of arginine is the most hydrophilic followed by that of histidine and then the side chains of the basic and acidic amino acids (Wolfenden *et al*., 1981). Threonine, serine, tryptophan and tyrosine had similar hydrophilic character. The side chains of methionine, cysteine, phenylalanine, alanine, valine, isoleucine, leucine and glycine were the least hydrophilic/most hydrophobic (Wolfenden *et al*., 1981).

1.2.4 Entropic effect

Entropy can be generally defined as a measure of a system to explore its available conformational space (Brady and Sharp, 1997). Entropy can be divided into a number of components of which the solvation (ΔS^{solv}) and conformational entropy (ΔS^{conf}) play a role in protein folding and stability. The ΔS^{solv} describes the freedom of the solute around a molecule and as mentioned earlier ΔS^{solv} favours and thus stabilizes the native state. On the other hand, the ΔS^{conf} in terms of proteins is dependent on the rotational freedom around residue side chains and the peptide backbone ϕ and ψ angle space (Brady and Sharp, 1997). Due to the reduction of conformational freedom in the native-state ΔS^{conf} opposes folding and destabilizes proteins.

1.3 Analysis of domain-domain interfaces

How do we determine the role of the domain interface in protein stability and folding? First, a structural alignment of domain interface residues will identify conserved amino acids. Such an analysis of protein-protein interfaces showed that there is a good correlation between structurally conserved residues and experimentally-identified amino acids that are important in stability and folding (Keskin *et al.,* 2005). Thus, the structural alignment of domain interfaces can be consequently used as the basis for protein engineering experiments that look to target critical residues involved in protein structure maintenance. Second, it will be useful to analyze the common structural features of domain-domain interfaces using computationally-based methods. Previous bioinformatics-based studies of domain-domain and protein-protein interface anatomy (Jones *et al.,* 2000; Stites, 1997; Jones and Thornton, 1996; Tsai *et al*., 1996; Jones and Thornton, 1995) have shown that the size of the interface is directly related to the extent of interface contacts and thus, to protein stability. Hence, a computational analysis of domain-domain interfaces will give a theoretical estimate of their contribution to protein stability. In addition, domain interfaces can be compared to subunit interfaces throwing light on the similarities and differences between intra- and inter-protein interactions.

The bioinformatics-based approach to studying domain-domain interfaces and their contribution toward protein stability and folding needs to be supplemented by an experimental analysis of domain interfaces. In order to quantitate the contribution of

the domain interface toward protein stability one needs to measure the loss of free energy resulting from the removal of domain-domain contacts. Two approaches can be employed: (i) a dissection or (ii) a site-directed mutagenesis analysis. Briefly, dissection studies involve the removal of protein regions and measuring the difference in stability between the truncated and wild-type proteins. On the other hand site-directed mutagenesis, the approach used in this study, removes specific contacts and so the role of these interactions at the domain interface can be easier identified. The choice of amino acid to be mutated is based on analysis of the crystal structure of the wild-type protein. Analysis of dimer interfaces has shown that proteins often utilise residues with large hydrophobic side chains to anchor the two subunits together (Jones and Thornton, 1995). Hence, residues whose side chains are completely buried in the domain interface and protrude from one domain into a pocket of the interacting domain can be singled out as possible targets for site-directed mutagenesis. In addition, one needs to consider the structural conservation of such amino acids since highly conserved residues have been shown to play an important role in protein stability and folding (Keskin *et al.,* 2005). Once a residue has been singled out, its substitution needs to be non-disruptive. A non-disruptive mutation is one that removes a set of defined contacts without introducing new or unfavourable interactions (Kellis *et al.,* 1988). An example of a non-disruptive mutation is isoleucine (Iso) to alanine (Ala). This is a like-for-like substitution where the hydrophobic side chain of Iso ($CHCH_3CH_2CH_3$) is replaced by the shorter hydrophobic group of Ala (CH_3). Glycine is not a good candidate to replace a wild-type residue. It has the smallest possible side chain, a hydrogen atom, that does not impose much steric interference on the Cα-N bond (ϕ angle). As a result, there is a great deal of rotation around the Cα-N bond of glycine that may lead to significant packing rearrangements in the mutant's structure. Mutations that may introduce (i) hydrogen bonds, (ii) salt-bridges, and (iii) electrostatic or steric repulsions are also not advisable. Electrostatic repulsions occur between residues with the same charge i.e. two basic or two acidic side chains (see section **1.2.2**). Steric repulsions can be generated if the newly introduced residue has a side chain that is larger then side chain of the wild-type amino acid. It is especially important to introduce non-disruptive mutations when studying protein folding. The native and unfolded conformations need to be perturbed as little as possible by the

mutation so that the folding mechanisms of the mutant and the wild-type protein are isomorphous hence, comparisons can be made.

Proteins in solution do not exist as rigid structures but are in constant motion. In order to fully understand how they carry out their functions one needs to closely examine the structural dynamics of these three-dimensional molecules. A powerful technique that can be employed to do just that is deuterium exchange detected using mass spectrometry (DXMS) (for recent reviews see Woods and Hamuro, 2001; Hoofnagle *et al*., 2003; Busenelehner and Armstrong, 2004; Krishna *et al*., 2004; Xiao *et al*., 2004; see section **2.2.12** for experimental details). The technique is based on the ability of hydrogen atoms within a protein to continuously and reversibly exchange with hydrogen in the surrounding environment. We are able to observe this exchange once solvent-hydrogen is replaced by its heavier isotope deuterium and the increase mass taken up by the protein is measured using mass spectrometry (Deng and Smith, 1993). In addition to being able to measure the conformational dynamics of proteins under specific conditions, i.e. pH, temperature etc. one can also use DXMS to probe the folding/unfolding of proteins (Clarke and Itzhaki, 1998; Deng and Smith, 1999; Englander, 2000). In general, DXMS experiments label proteins with deuterium in one of two ways. In continuous labelling the conformational properties of native structures are probed. Proteins are incubated with deuterated buffer in a series of increasing time constraints with the aim of determining range of fast (unstructured and highly flexible) to slow (structured and buried and/or involved in stable hydrogen bond network) exchanging regions. On the other hand, during pulse-labelling DXMS the protein is exposed to deuterium in a short time scale only. When this method is used in combination with equilibrium unfolding one is able to take a snap-shot of all species present under these conditions and measure their relative populations (Deng *et al*., 1999). In this study, site-directed mutagenesis was combined with continuous- as well as pulse-labelling DXMS. Residues thought to be important in maintaining the correct packing at the domain interface were identified using sequence- and structure-based alignments. Next these amino acids were mutated and DXMS was used to compare and contrast the structural dynamics of the wild-type and mutant proteins as well as any intermediate states present at equilibrium.

1.4 Dual state proteins: the soluble membrane-inserted transition

Most trans-membrane proteins are permanently attached to the membrane. However, a small number of proteins are able to exist in stable water-soluble state as well as in integral membrane form. An example of such amphitropic proteins is the CLIC family that will be dealt with later in sections **1.5.3** to **1.5.5**. For now the general mechanisms responsible for soluble to membrane-inserted conversion will be described.

The general soluble structure of amphitropic proteins is made of a polar/charge exterior that interacts favourably with the relatively high di-electric constant of the cytosol. The membrane spanning regions, which are mostly hydropohobic, are buried inside the protein and only become exposed in the vicinity of the membrane. The exposure of the buried trans-membrane regions requires a major structural re-arrangement. The 'umbrella' hypothesis, mechanism common to the toxin protein family, states that the outer polar layers peel-away exposing hydrophobic helices or sheets that than drive membrane insertion (Parker and Pattus, 1993; Chenal *et al.*, 2002). In most cases, the re-structuring occurs when the protein in solution approaches the membrane and re-folds forming less structured molten-globule like state with exposed hydrophobic surfaces. The main driving force between the two states is pH. As the protein approaches the membrane surface it encounters a lower pH induced by the negative charge of the polar side chain of phospholipids attracting hydrogen ions (H^+). This acidic pH destabilizes the protein and so lowers the energy barrier between the soluble and membrane-competent states resulting in acidic denaturation and exposure of the hydrophobic trans-membrane region/s (Bychcova *et al.*, 1996). Low pH plays another important role in membrane insertion. Charged residues are neutralized at the low pH micro-environment of the membrane. This makes the polar side chains of these amino acids more hydrophobic and so the energy penalty for insertion of a charge group in a hydrophobic environment is reduced. Acidic pH is not the only factor that drives the soluble/membrane conversion of amphitropic proteins. The formation of a molten-globule state has also been shown to be dependent on redox effects, membrane curvature, membrane charge and ion concentration (Chenal *et al*., 2005).

A molten globule-like state that forms in solution as a direct result of acid destabilization is not always the prerequisite for membrane insertion. In the case of Bcl-x_L, an all alpha helical protein belonging to the Bcl-2 family, no pH-dependent change was observed in the thermodynamic stability of the protein, i.e. Bcl-x_L exists in its soluble form even in the micro-environment of the membrane surface (Thuduppathy and Hill, 2006). Due to this lack of acid unfolding the main driving force for membrane insertion is thought to be the free energy of binding to the membrane. This energy is derived from increase in the number of acidic side chains that partition into the membrane upon protonation at lower pH (Thuduppathy and Hill, 2006). Despite the fact that the free energy of unfolding ($\Delta\Delta G°$) of Bcl-x_L between pH 7.4 and 4.9 is negligible, the *m*-value[η] was significantly decreased at pH 4.9 (Thuduppathy and Hill, 2006). This means that although there are no major structural re-arrangements induced by pH, the distribution of ensemble of species available at equilibrium has changed. The decrease in *m*-value points to the stabilization of less compact states. Therefore acidic pH primes Bcl-x_L for membrane insertion. It lowers the energy barrier between soluble and membrane-integral states so that once binding to the membrane occurs through protonated acidic residues the protein can easily expose its hydrophobic helices and insert into the membrane. The advantage of this mechanism is that conversion to a molten globule-like state only occurs once the protein is bound to the membrane and as such protease digestion is minimised (Thuduppathy and Hill, 2006).

Once inserted into the membrane the trans-membrane region/s is stabilized through hydrophobic interactions with the non-polar fatty-acid tails in the membrane interior. Often oligomerization is needed to form a fully functional channel. This process can occur either before or after membrane insertion Proteins that use helices to transverse the membrane can do so using a single helix (Parker and Feil, 2005). On the other hand trans-membrane β-strand structures need to form β-barrels for stabilization (Parker and Feil, 2005).

[η]: *The m-value reflects the amount of newly exposed surface area upon denaturation (Baskakov and Bolen, 1998).*

1.5 GST family and their structural homologues the CLICs

As previously mentioned proteins belonging to the CLIC family are found in both soluble and membrane-integral form. The crystal structures of soluble CLIC1 and CLIC4 have been solved (Harrop *et al.,* 2001; Littler *et al*., 2005). It was shown that in solution CLIC1, CLIC4, and in most likelihood, due to high sequence identity, the rest of the CLICs assume a fold common to the GST super-family. Hence the following section briefly introduces this super-family as well as describes what is known about the role of the domain interface in their stability and folding.

1.5.1 The GST super-family

The GSTs (EC 2.5.1.18) are a family of multi–functional enzymes involved in the cellular detoxification and excretion of many physiological and xenobiotic substances (Wilce and Parker, 1994). They are also known as ligandins (ligand–binding proteins) that are involved in the intracellular storage and transport of a variety of hydrophobic, non-substrate compounds such as hormones, metabolites and drugs (Sinning *et al.,* 1993).

The GSTs exist in a widespread range of organisms. They are divided into an ever-increasing number of classes based on nucleotide sequence, immunological, kinetic, and tertiary/quaternary structural properties (Sheenan *et al.,* 2001). Because this study focuses on the anatomy of domain interfaces only those members with solved crystal structures will be further discussed. The Alpha, Mu, Pi, Theta, Kappa, Zeta and Omega, to which the CLICs are most closely related with about 16 % sequence identity, are the mammalian classes (Figure 1). The schistosomal (Sj), Sigma and Beta classes are non-mammalian. Members of the Phi and Tau classes are found in plants. The GST proteins are dimers with the dimerization process being highly specific and occurring only between subunits within the same gene class. The GSTs share little sequence similarity but their overall structural fold has been preserved. The exception is Kappa GST whose C-terminal domain is inserted into the N-terminal domain (Ladner *et al.,* 2004). Subunit-subunit contacts are vital in dimer assembly and protein stability but the composition of the dimer interface varies between classes. Noticeably there is an inter-subunit lock-and-key motif, found in many GSTs, that has been shown to be crucial in subunit association (Hornby *et al.,* 2002; Sayed *et al.,* 2000).

Each subunit consists of two domains. The N-terminal domain is a mixed α/β domain topologically similar to the thioredoxin fold. This domain provides most of the glutathione (GSH) binding site and as a result is better conserved then the C-terminal domain. The C-terminal domain is all α-helical with the number of helices varying between different classes. It is believed that the differences in the structure of this domain are responsible for the differences in substrate specificity between the GST classes (Wilce and Parker, 1994). The two domains are connected via a short linker. The domain-domain interfaces of GST proteins have not been analysed in detail. However, in the case of human α-class GST (hGSTA1-1) an inter-domain lock-and-key motif, similar to the one found in most GST dimer interfaces, was shown to play an important role in stabilizing the domain-domain interface (Wallace *et al.,* 2000) (see section **1.5.2**).

1.5.2 Stability and folding of GSTs: role of the domain interface

The N- and C-terminal domains of GST proteins have been shown to unfold cooperatively in the case of class Alpha (Wallace *et al.,* 1998), Sigma (Stevens *et al.,* 1998), Pi (Erhardt and Dirr, 1995) and Sj (Kaplan *et al.,* 1997) GSTs. In addition, there is no evidence to suggest that the individual domains of GST proteins unfold independently as indicated by the monophasic, coincident transitions of monomeric class Mu GST (Hornby *et al.,* 2000). Thus, the burial of a significant amount of hydrophobic surface area upon domain association indicates that the domain interfaces of GSTs play an important role in the stability and folding of these proteins. Luo and co-workers (2002) investigated the conformational stability and equilibrium unfolding of two domain-exchanged chimeric isoenzymes. The domains of the class Mu isozymes M1-1 and M2-2 were exchanged resulting in chimeras (M12 and M21) with one domain from M1 and one domain from M2 (Luo *et al.,* 2002). It was shown that the M12 and M21 monomers are less stable then the wild-type monomers (Luo *et al.,* 2002). This indicates that domain interface complementarity is critical for correct domain-domain packing which in turn plays a role in protein stability. The effect of domain packing on stability and function was also investigated in the case of hGSTA1-1 (Wallace *et al.,* 2000). The residue chosen for mutagenesis was tryptophan 20 (Trp20), a conserved amino acid in the class Alpha GST proteins (Wallace *et al.,* 2000). Similarly to the lock-and-key motif found in the dimer interface of Alpha/Pi/Mu/Sj class proteins, the side chain of Trp20 protrudes from the N-terminal

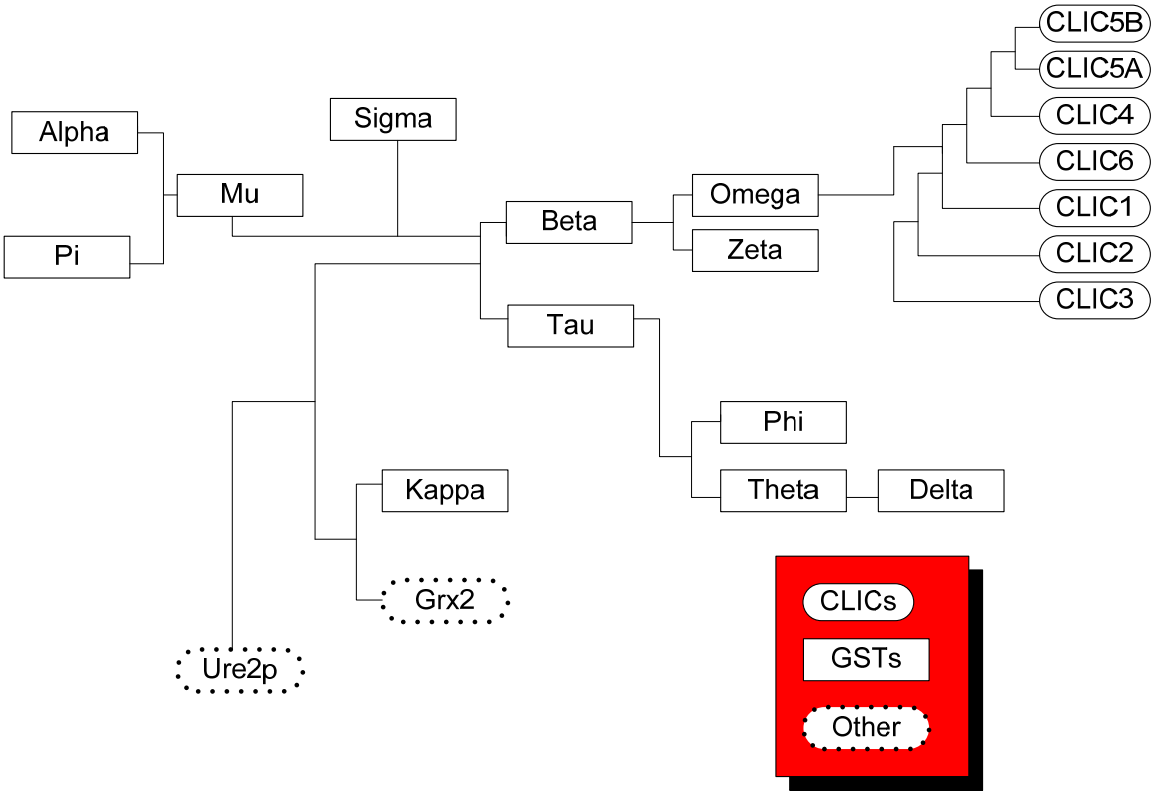


Figure 1: Dendogram of the GST super-family

Sequence relationship between CLIC (adapted from Berryman and Bretsher, 2000) and GST family proteins (adapted from Sheehan *et al*., 2001).

domain into a hydrophobic pocket of the C-terminal domain where it is completely buried (Wallace *et al.,* 2000). Trp20 was replaced with either phenylalanine (Trp20Phe) or alanine (Trp20Ala). Phenylalanines side chain is similar in size and hydrophobicity to that of tryptophan and as expected the Trp20Phe mutation did not impact on the structure and function of hGSTA1-1 but it did destabilize it slightly (Wallace *et al.,* 2000). The Trp20Ala mutation, a cavity forming mutation, was both disruptive and destabilizing with the equilibrium unfolding results pointing to the accumulation of one or more intermediate species (Wallace *et al.,* 2000). These results show that domain-domain contacts and their correct packing contribute significantly toward protein stability and function.

1.5.3 The CLIC protein family

The CLIC family is one of four major classes of anion-transport channels. The other being:

(1) Chloride channel proteins (CLC), which are homodimers with each monomer, made of 10 – 12 trans-membrane domains, forming its own pore (Dutzler *et al.*, 2002).

(2) Cystic fibrosis trans-membrane conductance regulators (CFTR), which form channels from a single subunit with twelve trans-membrane domains (Jentsch *et al*., 2001).

(3) Ligand gated anion channels, which appear to be oligomers where the individual subunits surround and contribute to a single pore (Jentsch *et al*., 2001).

There are six members that make up the CLIC1 family, namely: CLIC1, CLIC2, CLIC3, CLIC4, CLIC5A, p64/CLIC5B and parchorin (Harrop *et al.,* 2001). In comparison to other anion channels, the CLICs are relatively small composed of approximately 240 residues with the exclusion of p64 and parchorin, which have two additional amino-terminal domains. Another characteristic of CLIC proteins that is rare for ion channels is that they are also found in the cytoplasm and/or intracellular space. As a result, questions have been raised as to whether the CLICs are actual chloride channels, or if they form one part of a much larger channel or act as modulators/regulators of ion transport proteins (Duncan *et al*., 1997; Valenzuela *et al*., 1997; Tonini *et al*., 2000; Proutski *et al*., 2002; Tulk *et al*., 2002; Ashley, 2003;

Friedli *et al*., 2003). However, in the case of CLIC5A, p64/CLIC5B, CLIC4 and CLIC1 it has been shown that these proteins can facilitate ion transport on their own without the aid of additional subunits and/or accessory proteins (Landry *et al*., 1993; Chuang *et al*., 1999; Tulk *et al*., 2000; Harrop *et al*., 2001; Tulk *et al*., 2002; Berryman *et al*., 2004; Littler *et al*., 2005).

Apart from the two evolutionary non-conserved domains of p64/CLIC5B and parchorin, CLICs are highly conserved and share between 47% and 76% pairwise sequence identity (Cromer *et al.,* 2002). They are found in a number of cellular compartments including: the plasma membrane (Cromer *et al.,* 2002), mitochondria (Fernandez *et al.,* 1999), nuclear membrane (Valenzuela *et al.,* 1997), large dense core vesicles, trans-golgi vesicles (Edwards, 1999), secretory vesicles (Redhead *et al.,* 1997) and the endoplasmic recticular membrane (Duncan *et al*., 1997). CLIC channels have diverse physiological roles such as control of membrane potential, regulation of transepithelial ion absorption and secretion (Landry *et al.,* 1989) and facilitating the acidification of the ruffled border in bone resorption (Schlesinger *et al.,* 1997).

CLIC1, also known as NCC27 (nuclear chloride channel 27), was the first chloride channel found to localize to the nuclear membrane. However, it has a wide tissue distribution and while it is most abundant at the nucleoplasm and nuclear membrane it is also found, amongst others, in the kidney, heart and placental cells (Tulk *et al*., 2000). The latest electrophysiological studies have actually shown CLIC1 to be a non-selective channel allowing a number of anions to pass through the membrane (Singh and Ashley, 2006). In terms of function CLIC1 has been implicated in cell division where it controls the size of dividing cells through adjusting chloride and/or other anion concentrations (Valenzuela *et al*., 2000; Warton *et al*., 2002; Tonini *et al*., 2000). In addition the blockade of CLIC1 channel activity by its specific inhibitor, indanyloxiacetic acid 94 (IAA-94), has been shown to inhibit proliferation of microglial cells and reduce production of tumour necrosis factor-α and nitrite, which are up-regulated by the presence of β-amyloid. This protein is a major factor in Alzheimer's disease (Novarino *et al*., 2004).

1.5.4 Structure of soluble CLIC1

Soluble, reduced CLIC1 is a bi-domain monomer that assumes a fold similar to the GSTs. The protein is 26.9 kDa with a pI of 4.85. CLIC1 is mostly helical with only 4 β-sheets that compromise approximately 8 % of the total secondary structural content. The N-terminal domain (residues 1 – 90) mirrors the thioredoxin fold consisting of a four-stranded mixed β-sheet and three α-helices (Harrop *et al.,* 2001). The C-terminal domain (residues 90-240) is all α-helical most closely resembling the C-terminal domain of Omega class GST (Harrop *et al.,* 2001). The two domains of CLIC1 are linked via a proline-rich loop spanning residues 89-100 (Harrop *et al.,*2001). Proline side-chains are the most constrained of all amino acids and as such these residues may act as hinges keeping the two domains of soluble and/or membrane-competent CLIC1 in correct orientation (also see section **1.5.5.1**). CLIC1 has one tryptophan (Trp35) and eight tyrosine residues, spread throughout the tertiary structure of the protein. Fortunately for this study, the lone Trp35 is located at the domain interface and as such can be used as a local reporter of changes that may occur there (Figure 2). Overall there are 35 acidic and 27 basic residues resulting in a net negative imbalance in CLIC1. This is most clearly seen in the tertiary structure of the protein where the there are clusters of negative charged amino acids in close proximity to each other. There are two particularly negatively charged regions in CLIC1. The first is the negatively charged loop (Pro^{147} – Gln^{168}) found at the foot of the C-terminal domain (Figures 2). This fragment is also highly flexible as suggested by the lack of well defined electron density in the crystal structure of CLIC1 (Harrop *et al.,* 2001). The second region that has an asymmetric charge distribution is helix 9 (h9). The roles of the loop and h9 are unclear but they may be involved in some sort of protein interaction and/or ensuring that CLIC1 is in the correct orientation so that membrane insertion can occur. Finally, CLIC1 contains a glutathione (GSH) binding site found at the beginning of the N-terminal h2 (Figure 2). Binding occurs through the formation of a mixed disulfide bond between Cys24 and GSH. The affinity of Cys24 for GSH is lower than the rest of the GSTs except Kappa GST (Ladner *et al*., 2004). It has been suggested that CLIC1 can use its GSH-binding site to localize to specific targets in the cell (Harrop *et al*., 2001) as well as a mechanism for opening/closing the channel once inserted into membranes (Singh and Ashley, 2006).

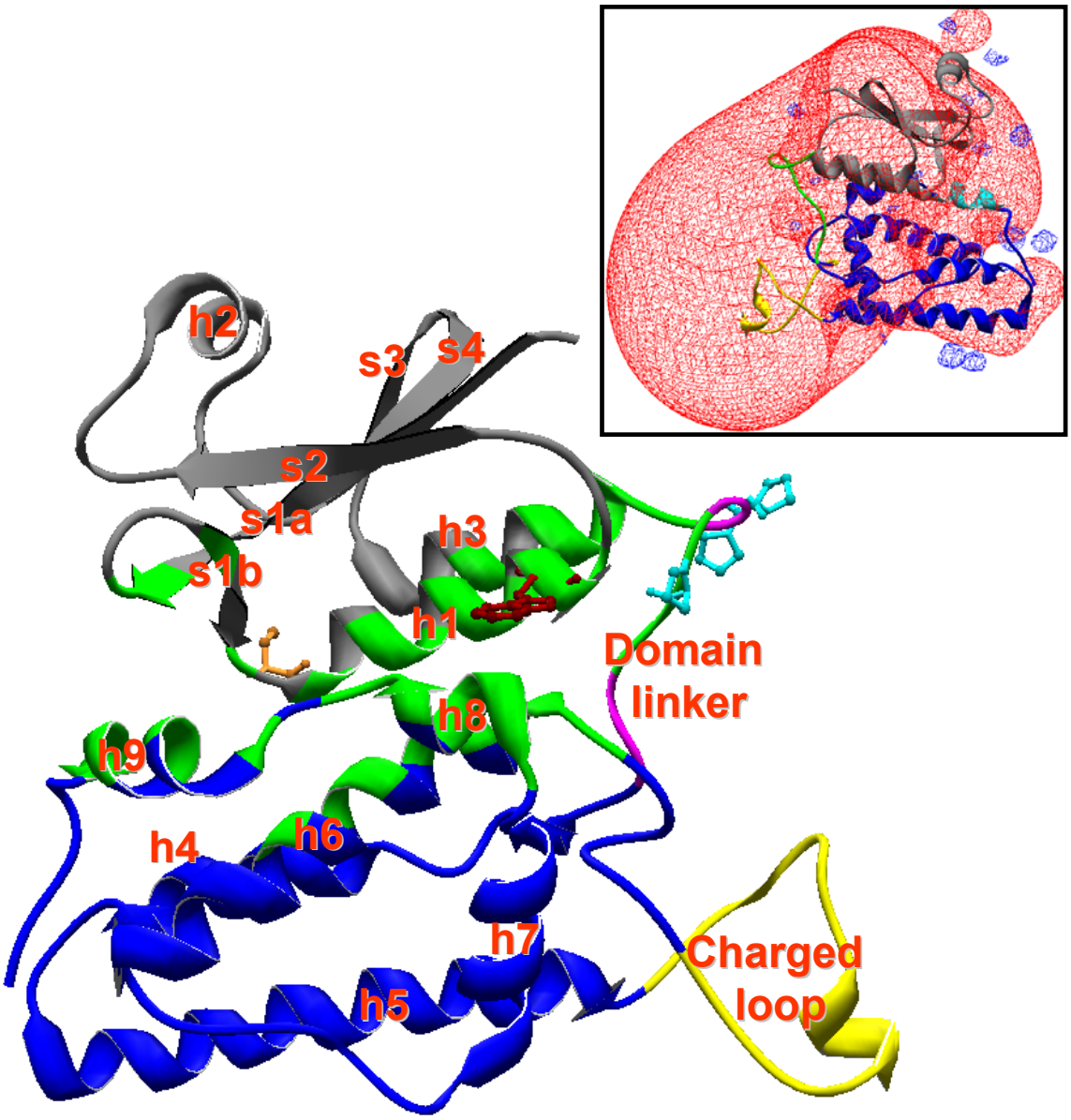


Figure 2: Structure of soluble, monomeric CLIC1

Structure of soluble CLIC1 (pdb code: 1k0m). N-terminal/thioredoxin-like domain is shown in grey, C-terminal/all-helical domain is blue, while the domain interface is marked in green. The side chains of Trp35 (brown) and Cys24 (orange) as well as Pro90, Pro91 and Pro94 (all in cyan) are also shown. The inset depicts the surface electrostatic potential of CLIC1 with negative charge in red and positive charge in blue. The figure was generated using SwissPDB-viewer (Guex and Peitsch, 1997).

1.5.5 Conversion of soluble CLIC1 to membrane-competent form

The first indication that CLIC1 is an integral membrane protein was provided by the fact that the protein was resistant to alkaline extraction (Tulk and Edwards, 1998). Next, it was shown that CLIC1 inserts into membranes with the amino terminus projecting extracellularly while the carboxyl-terminal is in the inside of the membrane (Tonini *et al*., 2000). Finally, CLIC1 expressed in bacteria and purified to homogeneity was shown to conduct chloride ions in reconstituted phospholipid vesicles indicating that the protein was able to function as a anion channel on its own in the absence of ancillary subunits or accessory proteins (Tulk *et al*., 2000).

To date, there are two theories of which CLIC1 region/s is involved in membrane insertion. Both proposals are based on clues derived from the crystal structure of soluble CLIC1. Cromer and his colleagues proposed that the C-terminal domain of CLIC1 may be involved in membrane insertion due to its similarity to pore-forming toxins (Cromer *et al.,* 2002). According to their model the hydrophobic helix 6 (h6) of CLIC1, buried at the centre of the C-terminal domain (see Figure 2), inserts into the membrane. This occurs after h6 is solvent-exposed through a structural re-arrangement based on the 'umbrella' hypothesis (see section **1.4**). On the other hand, the Harrop group proposed that the putative trans-membrane domain (TMD) is formed by residues Cys^{24} – Val^{46}, helix 1 and beta-stand 2 (h1-s2) in Figure 2 (Harrop *et al*., 2000).

The following evidence gives credence to Harrop's theory. Hydrophobicity plots show that both h6 and h1-s2 are significantly hydrophobic regions. However, h6 is mostly hydrophobic in all GSTs. On the other hand, h1-s2 is more hydrophobic and has a higher tendency to form a continuous helix, that may span the membrane, than any of the proteins in the GST super-family (Nathaniel, 2006). Further, h1 has both N-capping (Ser27) and C-capping (Lys37) motifs that act to stabilise helical structures. In addition, there is a positive motif positioned immediately after the h1-s2 putative TMD. Lys^{49} to Arg^{51} may prevent CLIC1 from inserting any further into the negatively charged membrane. Both Cromers' and Harrops' theories propose that CLIC1 membrane insertion requires a major structural re-arrangement (Cromer *et al*., 2002; Harrop *et al*., 2001). This is due to the globular shape of soluble CLIC1

ensuring that most non–polar regions are buried in the proteins' interior and as a result have minimal contact with the hydrophilic solvent. Hence, in order for the highly hydrophobic h6 or h1-s2 segments to be able to form a continuous helix that can span the membrane there needs to be a series of unfolding/refolding event/s. Hydrogen–deuterium exchange studies, performed by Nathaniel (unpublished data), show that the N-terminal domain of CLIC1 is less stable than the C-terminal domain and as such it is more likely to be re-arranged during conversion from soluble to membrane-competent form. Furthermore, an acidic pH destabilizes the N-domain of CLIC1. In particular, two peptides spanning regions 11 – 31 and 68 – 82 become significantly more flexible at pH 5.5, a common condition found at the micro-environment of membranes. Segment 11 – 31 forms part of the proposed N-terminal TMD (Harrop *et al*., 2001) CLIC1 also forms a dimer under strong oxidizing conditions (Littler *et al*., 2004). In this structure a intra-molecular disulfide bond is formed between Cys24 and Cys59 and the N-terminal domain is re-arranged so that the 4 β-sheets disappeared exposing a large hydrophobic surface that becomes the dimer interface (Figure 3). This dimer is proposed to only appear in the absence of lipids while in their presence the newly exposed hydrophobic surfaces of CLIC1 monomers can interact with the membrane (Littler *et al*., 2004). Finally, *in vivo* functional studies involving a CLIC protein from nematode *Caenorhabditis elegans*, EXC4, showed that an N-terminal sequence of 66 amino acids is crucial for correct membrane localization and channel activity of this protein (Berry and Hobert, 2006). This region made up of s1-h1-s2 encompasses the h1-s2 TMD proposed by Harrop and co-workers. An introduction of a helix-braking proline into h1 of EXC4 resulted in lack of membrane localization as well as subsequent channel activity (Berry and Hobert, 2006). Furthermore, the deletion of the s1-h1-s2 region, which leaves an intact C-terminal domain, also abolished membrane insertion. This contradicts Cromers' hypothesis that the C-terminal domain, h6, forms the trans-membrane helix of CLIC1.

What triggers the re-arrangement of the N-terminal, thioredoxin-like domain upon transition of CLIC1 from soluble to membrane-competent form? The low pI of CLIC1 is the first indication that pH plays a role in the conversion of this protein from soluble to membrane-inserted form. Electrophysiologically, it was shown that an

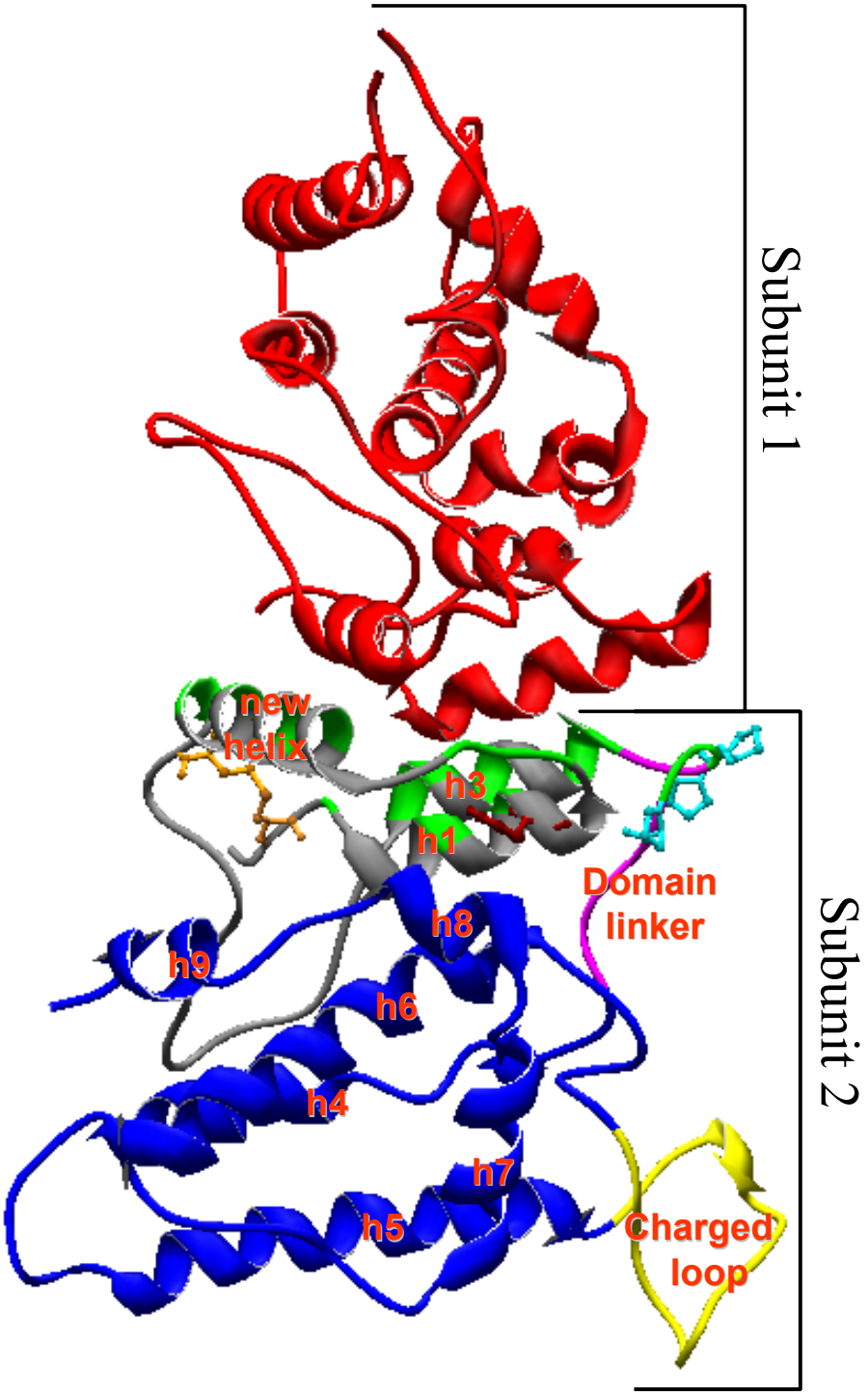


Figure 3: Structure of soluble dimeric CLIC1

Conformation of soluble homo-dimeric CLIC1 (pdb code: 1rky). The structure is formed under strongly-oxidizing conditions (Littler *et al*., 2004). The secondary structural elements are marked as per Figure 2. N-terminal/thioredoxin-like domain is shown in grey, C-terminal/all-helical domain is blue, while the dimer interface is green. The side chains of Trp35 (brown) as well as Pro90, Pro91 and Pro94 (all in cyan) are also shown. The intramollecular disulfide bond between Cys24 and Cys59 is in orange. The figure was generated using SwissPDB-viewer (Guex and Peitsch, 1997).

acidic pH increases the ability of CLIC1 to transport chloride ions (Tulk *et al.,* 2002; Warton *et al.,* 2002). Structurally, CLIC1 does not under-go a significant acid-induced conformational change. However, decrease in pH results in increased conformational flexibility of the N-terminal domain specifically at the s1-h1 and s4-h3 regions (Nathaniel, 2006). Therefore, the soluble structure of CLIC1 is primed for conversion to a membrane-competent form by acidic pH that lowers the activation barrier between the two states. This is confirmed by equilibrium unfolding studies where a decrease in pH significantly destabilized native CLIC1 while at the same time, at low denaturant concentrations, stabilized a molten globule-like state with exposed hydrophobic surfaces (McIntyre, 2006). When CLIC1 was exposed to a combination of conditions that mimic the membrane micro-environment such as reducing conditions, low pH, 37 °C and low-dielectric environment a less compact native state with exposed hydrophobic surfaces and re-arranged helical content was detected (McIntyre, 2006). Hence, the combination of these conditions reduced the activation barrier separating the soluble and membrane-competent states.

Considering all the above, a model for CLIC1 membrane insertion and chloride conductivity can be suggested. When CLIC1 comes into contact with lower pH environment, lower dielectric constant and the negative charge of the membrane, the energy barrier separating the soluble and membrane-integral forms is diminished. The destabilized thioredoxin domain is re-structured so that a TMD is formed between Pro^{25} – Val^{46}. The protein can than insert into the membrane as a monomer or oligomer. The majority of CLIC1 is on the cytosolic side of the membrane including the whole of the C-terminal domain. The only residues on the extracellular side of the membrane are 1 – 24. The exposed Cys24 can then possibly form disulphide bonds with neighbouring subunits depending upon the level of reducing/oxidizing conditions resulting in opening or closing of the channel (Singh and Ashley, 2006).

1.5.5.1 Role of domain interface in stability and function of CLIC1

Given the conformational switching of CLIC1, a question arises as to what is the involvement of the domain interface? Unlike the charged exterior of CLIC1, the CLIC1 domain interface is made up of mostly non-polar residues. This is expected since burial of charged/polar side chains is a highly destabilizing process due to a large desolvation penalty. There is only one, fairly exposed, inter-domain salt-bridge

between Lys20 and Asp225. Compared to Grx2, the only other monomeric GST homologue, the total domain interface buried upon domain association of CLIC1 is smaller by approximately 200 $Å^2$. In addition, the domain interface of CLIC1 (formed by helices h1 and h3) is malleable and susceptible to structural rearrangements (Harrop *et al.,* 2001). The movement is due to the *cis* or *trans* orientation of Pro91 (part of the inter-domain linker region) resulting in displacement and tilting of h1 and h3 (Harrop *et al.,* 2001). The relatively small surface area buried upon domain association as well as the plasticity of the domain interface indicates that it may not play as significant a role in protein stability as do the domain interfaces of other GSTs (see section **1.5.2**).

Visual inspection of the domain interface of CLIC1 reveals the presence of an inter-domain lock-and-key motif similar to the one found in the domain interface of hGSTA1-1 (see section **1.5.2**). The key residue in CLIC1 is Met32 (Figure 4). The side chain of Met32, like the one of tryptophan, is bulky and hydrophobic protruding from the N-terminal domain into a hydrophobic pocket in the C-terminal domain. Thus, it is possible that this lock-and-key motif plays an analogous role to the one found at the domain interface of hGSTA1-1 in interface packing and protein stability. However, keeping in mind that CLIC1 can exist in two states, soluble and membrane-competent, there may be a mechanism that allows the uncoupling of the lock-and-key interaction allowing the dissociation of the N- and C-terminal domains and consequent conversion of soluble CLIC1 to membrane-integral form.

1.6 Objectives

So far, a general sequence of events that take place as CLIC1 moves from conditions that stabilize its soluble form to an environment that promotes the membrane-competent state have been constructed. The main objective of this study is to zoom-in on the CLIC1 structure and provide details of the mechanism/s responsible for conversion of the soluble form to a membrane-competent form. Specifically, the role of the domain interface in this process will be analyzed. Attempts will be made to answer the following questions:

(1) Do the N- and C-terminal domains of CLIC1 dissociate before the thioredoxin fold can be re-arranged forming the TMD?

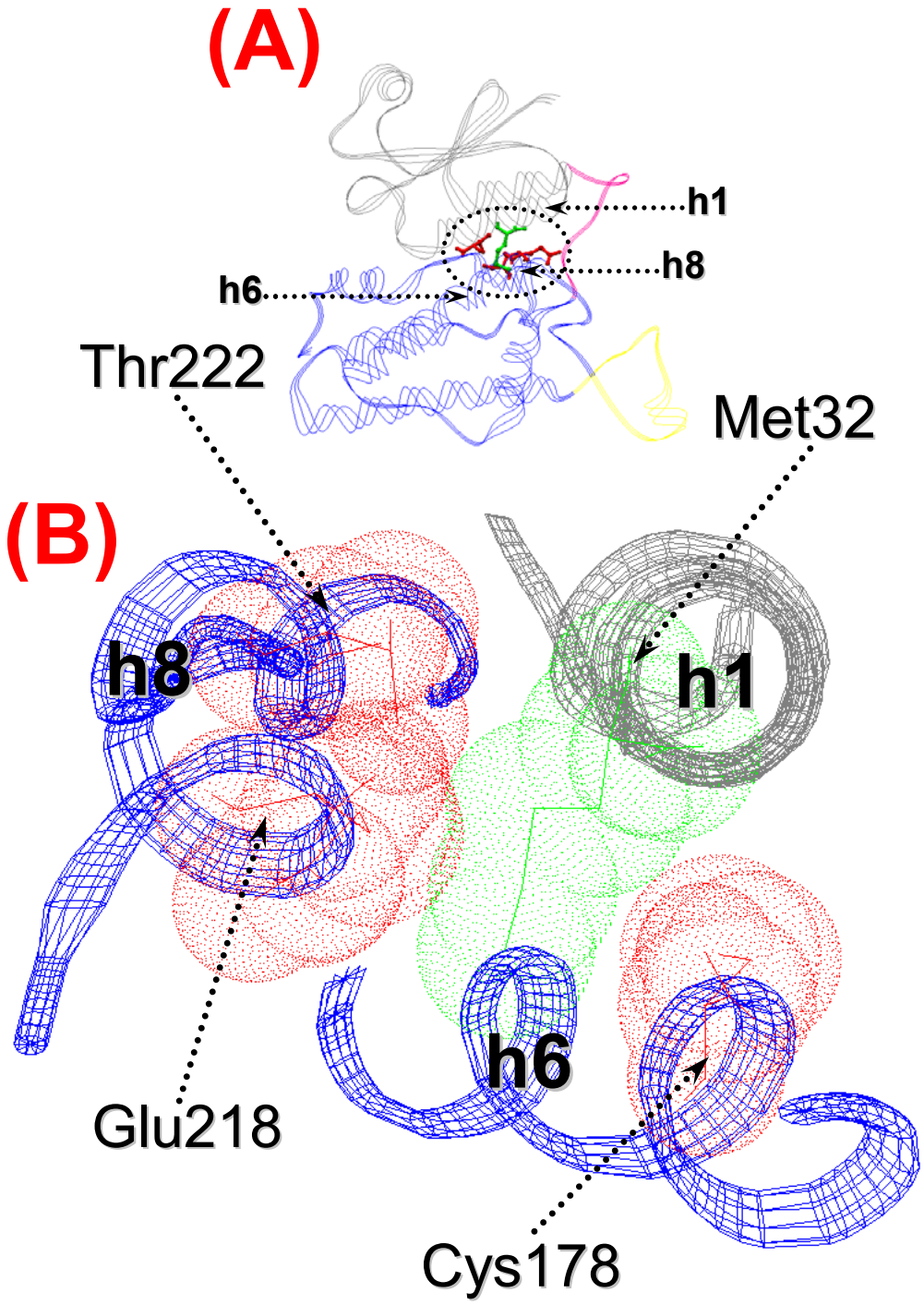


Figure 4: Inter-domain lock-and-key interaction of CLIC1

(A) The lock-and-key motif of the CLIC1 domain interface is shown in relation to the whole structure of native CLIC1 (pdb code 1k0m). **(B)** Interacting residues of the lock-and-key motif. The lock if formed by Cys178, Glu218 and Thr222 shown in red, while the key Met32 residue is in green. The figure was generated using SwissPDB-viewer (Guex and Peitsch, 1997).

(2) What domain interface interactions, if any, are destabilized and/or stabilized as CLIC1 approaches the membrane and prepares for the transition from soluble to membrane-inserted forms?

The study can be separated into two main parts: a bioinformatics-based investigation and an experimental analysis. In the bioinformatics section, domain interface interactions conserved in the GST super-family and those unique in the CLIC family will be identified through structural and sequence alignments. In addition, the domain interface architecture of CLICs and GSTs will be compared and contrasted in terms of surface area buried upon domain association, polarity, shape and complementarity as well as the number of inter-domain hydrogen bonds and salt-bridges.

In the experimental section, domain interface interactions that are deemed to be important in the stability and function of CLIC1, will be analysed. This will be achieved by generating mutants of wild-type CLIC1 (wtCLIC1) that attempt to isolate the effects of a particular interaction. In particular, the role of the inter-domain lock-and-key motif of CLIC1 in stability and folding will be probed.

CHAPTER 2

EXPERIMENTAL PROCEDURES

2.1 Materials

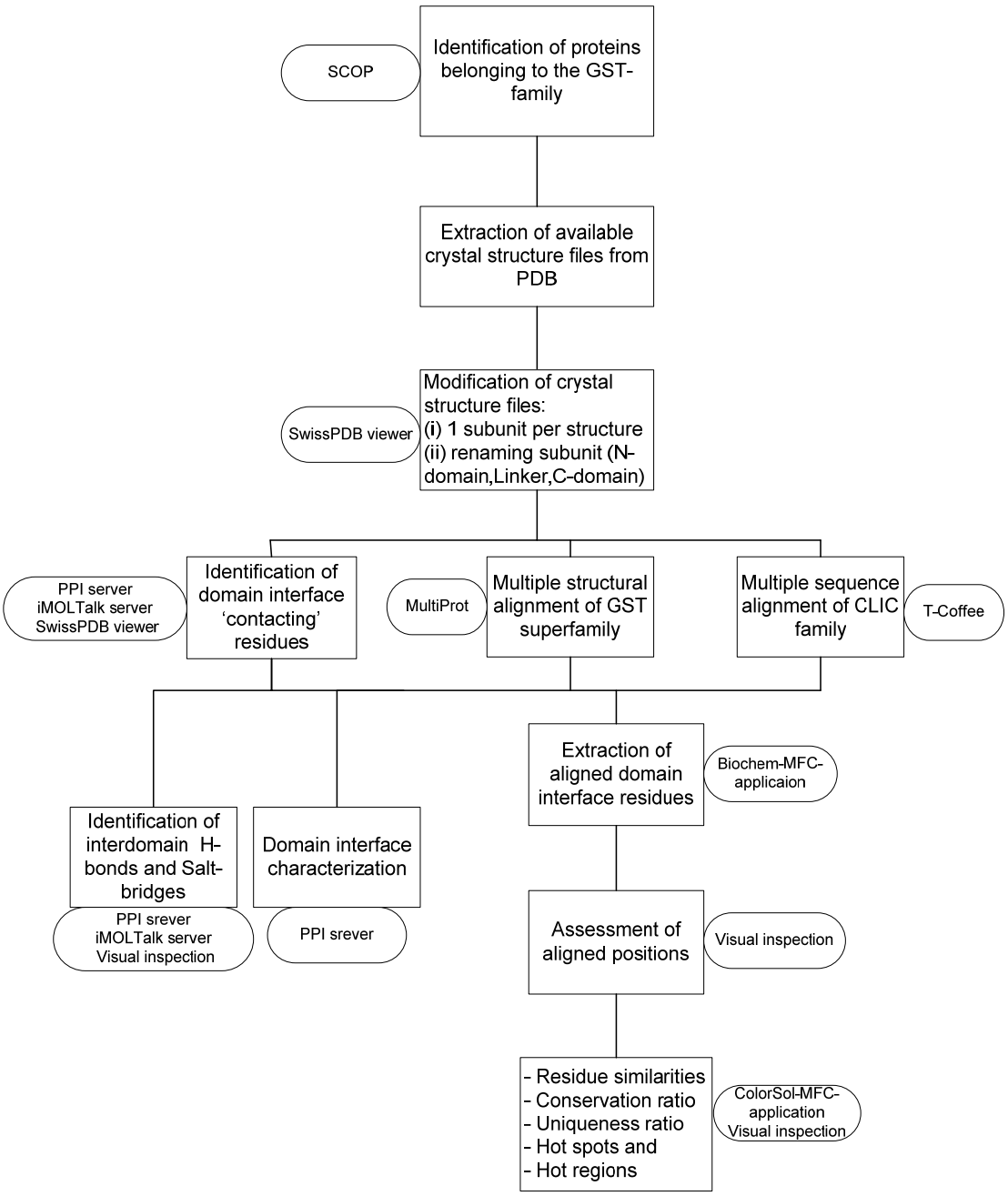
The cDNA encoding wild-type GST-CLIC1 fusion protein cloned into the pGEX-4T-1 vector was a gift from Dr. S. N. Breit, Centre of Immunology, St. Vincent's' hospital and University of New South Wales, Sydney, Australia. Deuterium oxide (99.9%), 8-anilino-1-naphthalenesulfonic acid (ANS), reduced glutathione (GSH), dithiothreitol (DTT), thrombin from bovine plasma, glutathione-agarose, porcine pepsin (3,200 – 4,500 units/mg protein; cat. No. P6887), ALD coupling solution (cat. No. 156159) and ethanolamine (cat. No. E9508) were purchased from Sigma-Aldrich (USA). The SDS-PAGE molecular weight marker (SM0431) was acquired from Fermentas. The QuikChange® Site-Directed Mutagenesis kit as well as *Escherichia coli* (*E. coli*) BL21 (DE3) competent cells were obtained from Stratagene (USA). All oligonucleotide primers were purchased from Inqaba-biotec, (South Africa). Peek tubing, pre-filters (cat. No. CPF10) and inline filters (cat. No. A.43) were attained from Upchurch Scientific. The C18 (1 x 50 mm, Vydac, cat. No. 218MS5105) reverse phase column, acetonitirle (ACN) and trifluoroacetic acid (TFA) solvents (HPLC grade) were acquired from Anatech, RSA. The peptide trap (cat No. 11-02872-TA) was purchased from OptiLynx. Ultrapure urea was obtained from Merck. POROS-20AL support matrix (cat. No. 1-6028-02) was attained from Applied-Biosystems. All other chemicals used were of analytical grade.

2.2 Methods

2.2.1 Structural alignment and characterization of domain interfaces in the GST super-family

The GST protein family domain interfaces were structurally aligned and analyzed in order to establish common trends and characteristics (see Figure 5 for an overview). Proteins belonging to the GST family were identified with the help of the Structural Classification of Proteins database (SCOP) (Murzin, 1995). The Protein Data Bank (PDB) (http://www.rcsb.org.pdb) was then searched in order to establish the number of proteins in the GST family with solved crystal structures. Forty, unique protein crystal structures belonging to 18 GST classes were downloaded. The crystal structure

Figure 5: Flow chart indicating the analyses of the GST-family domain interfaces



- The steps involved are shown in the rectangular boxes while the tools/programmes used to perform a particular step are indicated in the ellipsoid boxes.
- Diagram was generated using Microsoft Visio software

files belonging to dimeric proteins were modified, using SwissPDB viewer v3.7 (Guex and Peitsch, 1997), so that each file consisted of only one subunit. The sequences of the two domains that make up each subunit were identified and renamed as the N-terminal domain (N) and the C-terminal domain (C). The sequence linking the two domains was named (L) The domain interfaces were named according to which domain they belonged (N-domain/N-interface and C-domain/C-interface). In this study, domain interfaces consisted only of contacting residues and excluded any nearby residues that provide the interface scaffold. Two residues, one from an N-interface and one from a C-interface, whose atoms are between 4 and 6 Å from each other were identified as nearby residues. Two residues, from a pair of interacting interfaces, were defined to be contacting when any of their atoms were within a distance of 4 Å or less. The distance of 4 Å was chosen since it corresponds to the sum of the van der Waal's radii of two carbon atoms plus a tolerance of 0.5 Å. The contacting residues were identified using SwissPDB viewer (Guex and Peitsch, 1997) and verified by the PPI (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html) and iMOLTalk (http://i.moltalk.org) servers. A database of N- and C-interface residues was created from 39 members belonging to the GST-family.

Inter-domain hydrogen bonds and salt bridges were identified using the iMOLTalk server (http://i.moltalk.org) and verified by the PPI server (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html) as well as by visual inspection of the individual domain interfaces using SwissPDB (Guex and Peitsch**,** 1997). There were three selection criteria for hydrogen bonds. Firstly they had to form between an N-interface residue and a C-interface residue, secondly the bond distance had to be less than or equal to 3.9 Å and thirdly the bond angle had to be equal to greater than 90 °. In the case of salt-bridges, there were two selection criteria. Firstly salt bridges had to form between an N-interface residue and a C-interface residue and secondly the bond distance had to be less than or equal to 4.0 Å.

Interface residues separated by less than ten residues in sequence were grouped together into segments. The segments making up each interface were counted and inter-segment contacts recorded. The GST-family domain interfaces were further analyzed in terms of their hydrophobicity, size, shape and complementarity. All of these parameters were generated using the PPI server

(http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html). The sizes of the domain interfaces were reported in terms of interface accessible surface area (iASA), and percentage interface accessible surface area (%iASA) buried upon domain association. The %iASA was calculated using equation (**3)**:

$$\%iASA = iASA/dSA \qquad (3)$$

where **dSA** is the surface area of the N-domain or C-domain. To calculate the accessible surface areas, the PPI server uses an implementation of the Lee and Richards's algorithm (Lee and Richards, 1971), with a probe sphere of radius 1.4 Å. The shapes of the interfaces were analyzed in terms of planarity and length to breadth ratio. The planarity of the interfaces is analysed by calculating the best-fit plane through the 3-dimensional co-ordinates of the atoms at the interface using principal component analysis (Laskowski, 1991). The root mean square (rms) deviation of the atoms form the plane is calculated and used as the measure of planarity. Therefore, planarity values give an indication of protrusions and hollows at the interfaces. The length and breadth are measured through the standard deviations in the *x* and *y* dimensions, from the best-fit plane. The length/breadth ratio is the standard deviations in the *y* dimension divided by the standard deviations in the *x* dimension (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html). The length/breadth ratio indicates whether shape of the interface is spherical or extended. The gap-volume represents the complementarity/fit of two interacting interfaces. The PPI server uses the programme SURFNET (Laskowski, 1991) to determine the gap-volume between an N- and a C-interface. A sphere, of maximum radius 5.0 Å, is placed halfway between the surfaces of an N- and C-interface atom, such that its surface touches the surfaces of the atoms in the pair. Checks are made to test if any other atoms intercept this sphere and each time an intercept is detected the size of the sphere is reduced accordingly. If at any time the size of the sphere falls below 1.0 Å, it is discarded. The sizes of all the allowable gap-spheres are then used to calculate the gap-volume between the two interfaces. In this study, the gap-volumes were normalised and reported as gap-volume indices:

$$\text{Gap-volume index} = \text{Gap-volume}/(\text{N-iASA} + \text{C-iASA}) \qquad (4)$$

Forty subunits belonging to members of the GST-family were aligned using MulitiProt (Shatsky *et al.,* 2004). MultiProt is fully automated software used to align

simultaneously multiple structures of proteins. The alignments were performed independently of sequence order. The maximum rmsd (root mean square deviation) used for matching two fragments was set at 3 Å. The setting OnlyRefMol = 0 allows multiple alignments, not pair wise alignments. By default, MultiProt lists the best five solutions. These solutions were screened so that aligned domain interface residues were separated from aligned non-interface residues. The extraction of the aligned interface residues from the MultiProt solutions was done using the programme Biochem-MFC-application. This resulted in a database of structurally aligned interface residues from GST super-family proteins. The integrity of the aligned residues was randomly tested by visual inspection of the interfaces.

In addition to the structural alignment of proteins from the GST super-family, the sequences of twelve proteins belonging to the CLIC family were aligned using T-Coffee (Notredame *et al*., 2000). Positions belonging to the domain interface were identified and highlighted from domain interface residues of CLIC1 and CLIC4 since these are the only two CLIC family proteins whose crystal structures have been solved (Harrop *et al.,* 2001; Littler *et al*., 2005).

Residues were colour-coded according to their side chain chemistry using the ColorSol-MFC-application. Non-polar (Ala, Val, Leu, Iso, Met, Pro, Cys) blue, polar uncharged (Ser, Thr, Asn, Gln) red, polar positively charged (Lys, Arg, His) grey, polar negatively charged (Asp, Glu) green, aromatic (Phe, Tyr, Trp) pink and glycine (Gly) cyan. Residues belonging to the same group were named 'similar' residues. A number of conservation ratios were calculated at each aligned position along the GST and CLIC family interface:

(1) $\mathbf{Cr_{x\text{-}GST}}$: refers to the conservation ratio of the most prominent residue at each structurally aligned position at the domain interface of the GST super-family

(2) $\mathbf{Cr_{clic\text{-}GST}}$: refers to the conservation ratio of the CLIC1 residue at each structurally aligned position at the domain interface of the GST super-family

(3) $\mathbf{Cr_{z\text{-}GST}}$: refers to the conservation ratio of residues other than the most prominent and the CLIC1 residue at each structurally aligned position at the domain interface of the GST super-family

(4) $\mathbf{Cr_{x\text{-}CLIC}}$: refers to the conservation ratio of the most prominent residue at each sequence aligned position at the domain interface of the CLIC family

The general equation used to calculate each of the various conservation ratios was:

$$Cr_i = (\Sigma\delta_i + \Sigma\Delta_i)/m \qquad (5)$$

where $\mathbf{Cr_i}$ is the conservation ratio of residue **i**. $\boldsymbol{\delta_i}$ is one if residue **i** exists at the specific conserved position, and zero otherwise. $\boldsymbol{\Delta_i}$ is 1 if there is a similar residue, and zero otherwise. The number of members in the aligned position is represented by **m**. Positions along the interface that had a conservation ratio greater than or equal to 0.5 were considered as hot-spots. Hot-regions were identified where three or more hot-spot neighbours were found.

A uniqueness factor (Uf) of CLIC family domain interface positions was calculated using the following equation:

$$Uf = (\mathbf{Cr_{x\text{-}CLIC}} \text{ X } \mathbf{Cr_{x\text{-}GST}})/\mathbf{Cr_{clic\text{-}GST}} \qquad (6)$$

Where $\mathbf{Cr_{x\text{-}CLIC}}$, $\mathbf{Cr_{x\text{-}GST}}$, and $\mathbf{Cr_{clic\text{-}GST}}$ are the various conservation ratios obtained from the sequence alignment of the CLIC family and the structural alignment of the GST superfamily as described in the previous paragraph. A high Uf value indicates that at position 'X' of the consensus domain interface certain side chain chemistry is conserved in the CLIC family but not in the GST family.

2.2.2 Construction of CLIC1 mutants

Oligonucleotide primers were designed in accordance with the published wild-type nucleotide sequence of CLIC1 with the aid of the Primer-X software (http://bioinformatics.orf/primerx). The primers were created for use with the QuikChange® Site-Directed Mutagenesis kit (Stratagene; Papworth *et al.,* 1996). The sequences of the oligonucleotide primers used to generate the mutants CLIC1-M32A and CLIC1-E81M are shown in Table 1. The mutant proteins were generated by following the protocol described by Papworth (1996) in conjunction with the QuikChange® Site-Directed Mutagenesis kit (Stratagene). The sample reaction had a final volume of 50 µl. It was made up of 5 µl (10x) reaction buffer (100 mM KCL, 100 mM $(NH_4)_2SO_4$, 200 mM Tris-HCl pH 8.8, 20 mM $MgSO_4$, 1 % Triton X-100, 1 mg/ml nuclease-free bovine serum albumin), 1 µl (50 ng) double stranded DNA template, 1 µl (125 ng) forward primer, 1 µl (125 ng) reverse primer, 1 µl dNTP mix,

Table 1: Oligonucleotide primer sequences used for site-directed mutagenesis

Protein name	Fw/Rev	Sequence
CLIC1-	Fw	5’ CACAGACACCAACAAGATT**ATG**GAATTTCTGGAGGCAGTG 3’
E81M	Rev	5’ CACTGCCTCCAGAAATTC**CAT**AATCTTGTTGGTGTCTGTG 3’
CLIC1-	Fw	5’ CATTCTCCCAGAGACTGTTC**GCG**GTACTGTGGCTCAAGGGAG 3’
M32A	Rev	5’ CTCCCTTGAGCCACAGTAC**CGC**GAACAGTCTCTGGGAGAATG 3’

The numbers in column 1 following the proteins name indicate the codon position where the mutation was incorporated. The engineered mutations are shown in red in the sequences.

31 µl milli-Q water and 1 µl (2.5 U/µl) *Pfu Turbo* DNA polymerase. The products were generated by 16 amplification cycles of 30 seconds at 95 °C to denature the wild-type dsDNA, 60 seconds at 55 °C to anneal the mutant primers and 60 seconds at 68 °C for DNA extension. Parental DNA template was digested with 1 µl (10 U/µl) *Dpn I* for 1 hour at 37 °C and one hour at 20 °C. The reaction products were then transformed (see section **2.2.3**) into *E. coli* XL1-Blue Super-competent cells supplied with the mutagenesis kit. The cells were plated onto LB agar plates containing 100 µg/µl ampicillin. The LB-agar plates were incubated for 16 hours at 37 °C. Colonies were chosen at random and sent for DNA extraction and sequencing to Inqaba-biotec in order to confirm the presence of the engineered mutations and to ensure that no other mutations were generated during the mutagenesis amplification reaction.

2.2.3 Transformation of the mutant-plasmids into *Escherichia coli* BL21 (DE3) cells

The mutant plasmids were transformed into *E. coli* BL21 (DE3) cells by a one-step transformation method described by Chung *et al.* (1989). 50 µl of competent *Escherichia coli* BL21 (DE3) cells were thawed on ice. 1 µl (1-100 ng) mutant dsDNA was added and the reaction mixture was stored on ice for 30 minutes. The cells were then heat-shocked for 45 seconds at 42 °C on a heating block**,** followed by a rapid transfer to ice for 2 minutes. 950 µl 2xYT (1.6 g tryptone, 1 g yeast, 0.5 g NaCl per 100 ml dH_2O) was added to the reaction mixture and incubated in a shaker-incubator at 250 rpm for 90 minutes at 37 °C. The cells were plated on LB-agar plates containing 100 µg/µl ampicillin. Transformants were selected after the LB-agar plates were incubated for 16-18 hours at 37 °C.

2.2.4 CLIC1-E81M and CLIC1-M32A heterologous overexpression and purification

CLIC1-E81M and CLIC1-M32A were purified according to the method described by Tulk *et al.,* (2002). 100 µl *Escherichia coli* BL21 (DE3) cells transformed with the pGEX-4T-1 plasmid containing the cDNA sequence encoding GST-CLIC1-E81M or GST-CLIC1-M32A were added to 100 ml 2xYT medium containing 100 µg/µl ampicillin. The cells were grown on a shaker-incubator at 250 rpm for 16-18 hours at 37 °C. A 50-fold dilution of the over-night culture was added to 100 ml fresh 2xYT supplemented with 100 µg/µl ampicillin. Cells were grown with shaking at 37 °C until

an OD at 600 nm of 0.6 was reached. The culture was allowed to cool to 20 °C and over-expression of GST-CLIC1-E81M or GST-CLIC1-M32A was induced by the addition of 0.8 and 1 mM IPTG, respectively. The cells were grown for a further 6-8 hours at 20 °C in order to achieve optimum protein expression. The cells were harvested via centrifugation (5000 rpm, 15 min) and resuspended in approximately 15 ml resuspension buffer (10 mM Tris, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.02 % NaN_3, pH 7.5) per 1000 ml of original cell culture. The cells were aliquoted (1 ml per eppendorf) and stored over-night at - 20 °C to promote cell lysis.

The frozen cells were thawed at 4 °C on a rotator. 1 µl of 100 µg/µl DNAse, 10 µl of 1 M $MgCl_2$ and 100 µl of 10 mg/ml lysozyme were added per aliquot and the eppendorfs were rotated at 4 °C for 15-20 minutes. The cells were sonicated on ice by two cycles of 30 sec, duty cycle 40 % and power output 4 using Heat systems Ultrasonics cell disruptor model W-225R. The lysed cells were centrifuged at 9, 500 g for 20 minutes at 4 °C. The supernatant was loaded onto a GSH-agarose column pre-equilibrated with 10-column volumes equilibration buffer (10 mM Tris, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.02 % NaN_3, pH 8.0). The GST moiety of the GST-CLIC1 fusion protein binds the glutathione of the GSH-agarose column. Due to the slow binding kinetics between GST and GSH, samples were loaded at a flow-rate of 1ml/min to ensure maximum retention of the mutant fusion proteins. Unbound GST-CLIC1-E81M or GST-CLIC1-M32A and bacterial proteins were washed away with 10-column volumes of equilibration buffer. The GSH-agarose column was then equilibrated with 5-column volumes thrombin-cleavage buffer (200 mM Tris, 150 mM NaCl, 0.5 mM DTT, 0.02 % NaN_3, pH 8.4). 100 µl (1 U/ml) thrombin per 1 litre of original cell culture was added to 15 ml thrombin cleavage buffer. This solution was added to the GSH-agarose column containing the bound GST-CLIC1-E81M or GST-CLIC1-M32A. The fusion protein and thrombin were incubated on a rotator for 16 hours at 20 °C resulting in the release of CLIC1-E81M or CLIC1-M32A from the column-bound GST moiety.

The CLIC1-E81M or CLIC1-M32A thrombin mixture was collected from the GSH-agarose column and separated using a DEAE-anion exchange column. The DEAE-anion exchange column was connected to an Äktaprime system and equilibrated with 10-column volumes DEAE-equilibration buffer (20 mM Tris, 1 mM DTT, 0.02 %

NaN_3, pH 6.5). Thrombin has an estimated pI of 8 while the pI values of CLIC1-E81M and CLIC1-M32A are 5.2 and 5.1, respectively (the pI values of thrombin, CLIC1-E81M and CLIC1-M32A were calculated using the ProtParam tool found in EXPASY, http://www.expasy.org). Therefore, at pH 6.5 thrombin is positively charged and does not bind the DEAE-column. In contrast, at pH 6.5 the mutant CLIC1 proteins are negatively charged and bind the DEAE-column. The purified CLIC1-E81M or CLIC1-M32A were eluted from the DEAE-anion exchange column using a DEAE-elution buffer (20 mM Tris, 300 mM NaCl, 0.02 % NaN_3, pH 6.5). The mutant proteins were collected in 2 ml fractions and their purity assessed on 15 % acrylamide SDS-PAGE gels (see section **2.2.6**). Every 10 ml of purified CLIC1E81M/CLIC1-M32A was dialysed into 1 L CLIC1-storage buffer (50 mM $NaHPO_4$, 1 mM DTT, 0.02 % NaN_3, pH 7.0). For all ensuing experiments the protein was dialysed (10 ml protein/l buffer) every two weeks against CLIC1-storage buffer containing fresh, reduced DTT.

The GSH-agarose column needed to be regenerated after CLIC1 purification. Thus, the column-bound GST as well as any undigested GST-CLIC1-E81M or GST-CLIC1-M32A had to be eluted. This was accomplished by washing the affinity column with 5-column volumes GSH-agarose elution buffer (50 mM Tris, 10 mM GSH, 0.02 % NaN_3, pH 8.0).

In addition to the engineered point mutations, CLIC1-E81M and CLIC1-M32A contain two additional N-terminal residues, glycine and serine. These amino acids are part of the thrombin recognition sequence and remain with the mutant proteins after cleavage from the GST moiety (Harrop *et al*., 2001).

2.2.5 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

The solubility, homogeneity and purity of the expressed mutant proteins were assessed by separation on 12-15 % SDS-PAGE (Laemmli, 1970). The discontinuous gel system was constructed from a 4 % acrylamide/bis-acrylamide (w/v) stacking gel (0.1 % (w/v) SDS, 0.05 % (w/v) ammonium persulphate, 0.1 % (w/v) TEMED and 0.125 M Tris/HCl, pH 6.8) and a 12-15 % acrylamide/bis (w/v) separating gel (0.1 % (w/v) SDS, 0.05 % (w/v) ammonium persulphate, 0.1 % (w/v) TEMED and 0.375 M Tris/HCl, pH 8.8). Protein samples were diluted 5X with sample buffer (10 % (w/v)

glycerol, 2 % (w/v) SDS, 5 % (w/v) 2-mercaptoethanol, 0.05 % (w/v) bromophenol blue and 0.0625 M Tris/HCl, pH 6.8). Samples were vortexed briefly and boiled for 5 minutes to ensure that the proteins were denatured. The electrolyte buffer used contained 1 % (w/v) SDS, 0.192 M glycine and 0.025 M Tris/HCl, pH 8.3. 20-30 µl of the protein samples were applied to the SDS-PAGE wells and electrophoresed at 120-140 V for 2-3 hours. The molecular weight marker used contained a mixture of seven proteins: β-galactosidase (116k Da), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), lactate dehydrogenase (35 kDa), restriction endonuclease Bsp98I (25 kDa), β-lactoglobulin (18.4 kDa) and lysozyme (14.4 kDa). The gels were stained in 2 % (w/v) Coomassie Blue R250 staining solution containing 13.5 % (v/v) glacial acetic acid and 18.75 % (v/v) ethanol and destained with 40 % (v/v) ethanol and 10 % (v/v) glacial acetic acid for 2-3 hours until the background was clear.

2.2.6 Protein concentration determination

The concentrations of CLIC1-E81M and CLIC1-M32A were determined spectrophotometrically using the Beer-Lambert law:

$$A = \varepsilon_\lambda c l \qquad (7)$$

where **A** is the absorbance at 280 nm, $\boldsymbol{\varepsilon_\lambda}$ is the molar extinction of the absorber at wavelength $\boldsymbol{\lambda}$, **c** is the concentration of the absorbing solution and ***l*** is the path length of light through the solution (cuvette). The molar extinction coefficient (ε_λ) of CLIC1-E81M and CLIC1-M32A at 280 nm were established by using the extinction coefficients of tryptophan, tyrosine and cysteine residues (Mach *et al.,* 1995).

$$\varepsilon_{(280)}\,(M^{-1}cm^{-1}) = 5550\Sigma Trp + 1340\Sigma Tyr + 150\Sigma Cys \qquad (8)$$
$$= 5550(1) + 1340(8) + 150(6)$$
$$= 17170\ M^{-1}cm^{-1}$$

The absorbance (**A**) at 280 nm was determined by fitting a linear regression to five or more points from a serial dilution. All readings were buffer corrected with the appropriate buffer used for the concentration determination.

2.2.7 Secondary and tertiary structural characterization

2.2.7.1 Circular dichroism spectroscopy

Circular dichroism (CD) is a technique that measures the differential absorption of left- and right-handed circularly polarised light by optically active molecules. Optical activity in proteins arises from disulphide groups, aromatic side chains, and the peptide backbone (Woody, 1995). Disulphide groups and aromatic amino acids have characteristic absorption bands in the near-UV range (250-300 nm). In the far-UV region (170-250 nm) the predominant signal arises from the peptide backbone. The adoption of different secondary structures by the peptide backbone results in distinctive CD spectra (Woody, 1995). As a result, this wavelength range gives a good indication of the secondary structural content of proteins. Proteins with a high α-helical content display characteristic minima at 208 and 222 nm, as well as a stronger positive band near 190 nm (Woody, 1995). Due to the high noise to signal ratio of some buffers it is impossible to record readable spectra below 210 nm. Therefore, measurement of the ellipticity at 222 nm (E222) was used as a secondary structural probe of CLIC1-M32A, which is predominantly α-helical with 4 β-strands and 10 α-helices.

Far-UV-CD spectra (190-250 nm) were recorded using 2-5 μM CLIC1-M32A. The protein was in CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.02 % NaN_3) pH 7.0 or pH 5.5. In some cases the CLIC1 storage buffer was diluted 10X in order to reduce the noise signal. All CD spectra were recorded at 20 °C and represent an average of 10 accumulations, at a scan speed of 100 nm/min. The bandwidth used was 1 nm and the data pitch 0.2 nm. All readings were recorded in a 2 mm cuvette using a Jasco J-810 spectropolarimeter and the Spectra Manager software v1.5.00. All spectra were buffer corrected. The spectra were normalised by calculating the mean residue ellipticity **[θ]** $deg.cm^2dmol^{-1}residue^{-1}$ using the following equation:

$$[\theta] = (100x\theta)/cnl \qquad (9)$$

where (**θ**) is the ellipticity signal in mdeg, **c** (mM) is the protein concentration, **n** is the number of residues in the protein chain and ***l*** is the path length in cm. All CD spectra were plotted and smoothed using the negative exponential technique of SigmaPlot v9.0.

2.2.7.2 Fluorescence spectroscopy

Fluorescence is the emission that results from the return of an unpaired electron from the excited to the ground state (Lakowicz, 1983). The energy lost between excitation and emission, known as Stokes' shift, results in the shift of emission spectra to lower wavelengths. In proteins, the naturally occurring fluorophores are tryptophan, tyrosine and phenylalanine. Due to the small quantum yield of phenylalanine in proteins, its emission is rarely observed (Lakowicz, 1983). The fluorescence of most proteins is dominated by tryptophan, with its quantum yield being more than double that of tyrosine. In the native state, tyrosine emission is quenched by energy transfer to tryptophan, quenching due to nearby charged carboxyl and uncharged amino groups and the formation of bonds on either the carboxy or amino groups (Lakowicz, 1983). Hence, protein unfolding can result in an increased emission due to tyrosine. The indole ring of tryptophan is highly sensitive to solvent polarity (Lakowicz, 1983). Emission spectra of this residue reflect the polarity of its surrounding environment. Therefore, tryptophan fluorescence is used to monitor tertiary structural changes in proteins.

CLIC1 contains one tryptophan residue at position 35 (Trp35) and eight tyrosine (Tyr) residues. One tyrosine is situated in the N-terminal domain, one at the inter-domain linker and six in the C-terminal domain. The single Trp35 residue was selectively excited at 295 nm. Trp35 acts as a local tertiary structural reporter, since it is situated only three residues from the engineered mutation Met32Ala. Excitation at 280 nm resulted in combined Trp35/Tyr excitation which is used as a global tertiary-structural probe, although it must be remembered that the Trp35/Tyr emission spectra are dominated by tryptophan fluorescence. Fluorescence emission spectra were recorded using 1-3 μM CLIC1-M32A in the range 280-450 nm. The excitation and emission slit widths were kept at 5 nm. The buffer used was CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.02 % NaN_3) pH 7.0 or pH 5.5. The spectra were recorded at 20 °C, buffer corrected, and are an average of three accumulations at a scan speed of 200 nm/min. Readings were taken in a quartz cuvette with a 1 mm path-length using a Perkin-Elmer luminescence spectrometer LS50B and FLwinlab v4.0 software. The data was plotted and smoothed using the negative exponential technique of SigmaPlot v9.0.

2.2.8 Refolding studies

Determination of the conformational stability parameters ΔG_{H20} and *m*-value are dependent on the reversibility of the observed reaction. CLIC1-M32A reversibility was followed via fluorescence spectroscopy (see section **2.2.7.2**) and far-UV-CD (see section **2.2.7.1**). 10 μM CLIC1-M32A was unfolded in 8 M urea for one hour. The protein was then refolded for one hour via a 10X dilution into the appropriate CLIC1 storage buffer. The control used to calculate the percentage refolding contained 1 μM protein in 0.8 M urea. The buffer used was CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.02 % NaN_3) pH 7.0 or pH 5.5. The data was plotted and smoothed using the negative exponential technique of SigmaPlot v9.0.

2.2.9 Fluorescence- and CD-monitored urea-induced equilibrium unfolding

Protein unfolding transitions are a convenient way of estimating the stability of a protein under varying conditions. The effect of an engineered mutation on the stability of a protein can be studied by comparison of the wild-type and mutant equilibrium unfolding curves. A denaturant is used to shift the equilibrium from the native to the unfolded state. The equilibrium constant (K_{eq}) can be calculated and hence the conformational stability parameters ΔG_{H20} and *m*-value can be determined, provided that refolding is reversible (see section **2.2.8**).

During this study, urea was used as a denaturant. 10 M stock urea was prepared as described by Pace *et al.,* (1986), in the appropriate CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.0 2% NaN_3 pH 7.0 or pH 5.5). 2 μM CLIC1-M32A was incubated in 0 to 8 M urea at 20 °C for 60 to 90 minutes to reach equilibrium. The extent of urea-induced unfolding on CLIC1-M32A was monitored using CD (see section **2.2.7.1**) and fluorescence (see section **2.2.7.2**) probes. In the case of the CD-monitored unfolding the change of ellipticity at 222 nm (E222) with increasing urea concentrations were recorded. The fluorescence-monitored equilibrium unfolding recorded the fluorescence emission spectra of Trp 35 (excitation at 295 nm) and Trp 35/Tyr (excitation at 280 nm) of CLIC1-M32A with increasing urea concentrations (0-8 M). The Rayleigh scatter at 280 and 295 nm with increasing urea concentrations was used as a reporter for monitoring the presence of protein aggregates.

2.2.10 Urea-induced equilibrium unfolding in the presence of ANS

8-Anilino-1-naphthalene-sulfonate (ANS) is a hydrophobic dye used as an extrinsic fluorescence probe (Engelhard and Evans, 1995). It binds to hydrophobic patches in proteins. In an aqueous environment ANS fluorescence is quenched, but upon binding to a hydrophobic surface its fluorescence quantum yield increases and emission wavelength is blue shifted (Engelhard and Evans, 1995).

2 mM stock ANS was prepared in CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.02 % NaN_3) pH 7.0 or pH 5.5. The concentration of ANS was checked by recording the absorbance at 350 nm and using extinction coefficient of $\varepsilon_{350} = 4950\ M^{-1}cm^{-1}$ (see section **2.2.6**). 2 μM CLIC1-M32A was incubated for 60 minutes, with varying urea concentrations (0-8 M). ANS was added to the protein/urea mixture to a final concentration of 200 μM. Binding of ANS to the protein was allowed for at least an hour. A series of blanks were generated each containing 200 μM ANS with the appropriate urea concentration (0-8 M). The samples were excited at 390 nm and emission spectra were recorded from 390 to 600 nm. Spectra were produced from an average of three accumulations at 300 nm/min scan speed. The excitation and emission slit widths were at 5 nm. The buffer used was CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.02 % NaN_3) pH 7.0 or pH 5.5. Readings were taken at 20 °C using a PerkinElmer luminescence spectrometer LS50B and FLwinlab v4.0 software. The emissions at 470 nm were plotted as a function of urea concentrations. Recording the fluorescence emission at 390 nm with increasing urea concentrations monitored the presence of aggregates. The data was plotted using SigmaPlot v9.0.

2.2.11 Equilibrium unfolding data fitting

The CD-monitored unfolding transitions of CLIC-M32A were generated by plotting the E222 as a function of urea. The fluorescence-monitored unfolding data was analysed in three ways. These involved plotting:

(i) the fluorescence intensity at 347 nm (F347) at each urea concentration. 347 nm is the emission wavelength maximum ($\lambda_{em\ max}$) of the fluorescence emission spectrum of native CLIC1-M32A excited at either 280 nm or 295 nm.

(ii) the ratio of the fluorescence intensity at 358 nm and 347 nm (F358/F347) with increasing urea concentrations. 358 nm is the $\lambda_{em\ max}$ of the fluorescence emission spectrum of unfolded CLIC1-M32A.

(iii) $\lambda_{em\ max}$ at each urea concentration.

Values of C_m (the denaturant concentration at which half of the population of protein molecules are unfolded) were obtained from the mid-points of the unfolding transitions. The urea-induced equilibrium unfolding transitions were analysed with both two-state and three-state monomer models as described by Pace *et al.,* (1986).

2.2.11.1 Two-state monomer fit

The calculation of ΔG_{H2O}, using the two-state monomer model (N ↔ U), was based on the assumptions that:

(i) No intermediate states were present.

(ii) ΔG is related linearly to the denaturant concentration.

For a two-state monomer transition, a protein can exist only in the native (**N)** or unfolded **(U)** forms.

$$N \leftrightarrow U \qquad (10)$$

$$F_N + F_U = 1 \qquad (11)$$

Where $\mathbf{F_N}$ is the fraction native protein and $\mathbf{F_U}$ is the fraction of unfolded protein. $\mathbf{Y_{obs}}$ represents the physically observed values recorded in this study, using fluorescence and CD probes. $\mathbf{Y_{obs}}$ can be represented by the following equation:

$$Y_{obs} = Y_N F_N + Y_U F_U \qquad (12)$$

where $\mathbf{Y_N}$ and $\mathbf{Y_U}$ represent the measured properties of the native and unfolded states respectively. In the transition region $\mathbf{Y_N}$ and $\mathbf{Y_U}$ are estimated from the linear extrapolation of the pre- and post - transition baselines, respectively. Combining equations **(11)** and **(12)** results in:

$$F_U = (Y_N - Y_{obs})/(Y_N - Y_U) \qquad (13)$$

The free energy change upon unfolding (ΔG) is a measure of the conformational stability of a protein. It relates to the equilibrium constant (K_{eq}) by the equation:

$$\Delta G = -RT\ln K_{eq} \qquad (14)$$

where ***R*** is the universal gas constant (1.987 cal.mol^{-1}K^{-1}) and ***T*** is the temperature measured in Kelvin. In terms of a two-state monomer transition:

$$K_{eq} = U/N = F_U/F_N = F_U/(1 - F_U) = (Y_N - Y_{obs})/(Y_{obs} - Y_u) \qquad (15)$$

Combining equations **(14)** and **(15)** results in:

$$\Delta G = -RT\ln(Y_N - Y_{obs})/(Y_{obs} - Y_u) \qquad (16)$$

In this study the free energy change in the absence of denaturant (ΔG_{H2O}) and the dependence of free energy on denaturant concentration (*m*-value) were reported. These two equilibrium parameters provide thermodynamic information about a protein. They were calculated using the linear extrapolation method first described by Pace *et al.,* (1986).

$$\Delta G = \Delta G_{H2O} - m*(\text{denaturant}) \qquad (17)$$

The method assumes that ΔG is linearly related to the denaturant concentration and can therefore be extrapolated to zero denaturant concentration in order to obtain ΔG_{H2O}. The *m*-value is determined from the slope of ΔG versus denaturant concentration plot.

2.2.11.2 Three-state monomer fit

In a three-state monomer transition involving one intermediate (**I**), three species exist:

$$N \leftrightarrow I \leftrightarrow U \qquad (18)$$

$$F_N + F_I + F_U = 1 \qquad (19)$$

$$K_1 = F_I/F_N \qquad K_2 = F_U/F_I \qquad K_U = F_U/F_N = K_1*K_2 \qquad (20)$$

Where $\mathbf{F_I}$ is the fraction of intermediate present. The equilibrium constant for **N** ↔ **I** is represented by $\mathbf{K_1}$. The equilibrium constant for **I** ↔ **U** is represented by $\mathbf{K_2}$. The equilibrium constant for **N** ↔ **U** is represented by $\mathbf{K_1*K_2}$. $\mathbf{Y_{obs}}$ can be represented as:

$$Y_{obs} = Y_N F_N + Y_I F_I + Y_U F_U \qquad (21)$$

Where $\mathbf{Y_I}$ depicts the measured properties of the intermediate. Next we need to solve $\mathbf{F_U}$ in terms of $\mathbf{K_1}$ and $\mathbf{K_2}$ by combining equations **(19)** and **(20)**

$$F_U = (K_1*K_2)/(K_1*K_2 + 1 + K_1) \qquad (22)$$

Rearranging and consequently substituting equation **(20)** into equation **(21)** we can solve for $\mathbf{F_N}$ and $\mathbf{F_I}$ in terms of $\mathbf{K_1}$ and $\mathbf{K_2}$.

$$F_N = F_U/K_1*K_2$$

Therefore: $$F_N = [(K_1*K_2)/(K_1*K_2 + 1 + K_1)]/K_1*K_2$$

$$F_N = 1/[(K_1*K_2 + 1 + K_1)/K_1*K_2] \quad (23)$$

Similarly: $$F_I = K_1/[(K_1*K_2 + 1 + K_1)/K_1*K_2] \quad (24)$$

Substituting equations **(22)**, **(23)** and **(24)** into equation **(21)** we get:

$$Y_{obs} = (Y_N + Y_I*K_1 + Y_U*(K_1*K_2))/(1 + K_1 + K_1*K_2) \quad (25)$$

We know that: $\Delta G_1 = -RT\ln K_1$ hence $K_1 = e^{\Delta G_1/RT}$ (26)

and $$\Delta G_1 = \Delta G_{2H2O} - m_1*(\text{denaturant}) \quad (27)$$

Therefore, combining equations **(26)** and **(27)** we get:

$$K_1 = e^{[\Delta G_{1H2O} - m1*(\text{denaturant})]/RT} \quad (28)$$

Similarly: $$K_2 = e^{[\Delta G_{2H2O} - m2*(\text{denaturant})]/RT} \quad (29)$$

Substituting equations **(28)** and **(29)** into **(25)** we get the final model to fit the data.

$$Y_{obs} = [Y_N + Y_I*(e^{[\Delta G_{1H2O} - m1*(\text{denaturant})]/RT}) + Y_U*(e^{[\Delta G_{1H2O} - m1*(\text{denaturant})]/RT})*$$
$$(e^{[\Delta G_{2H2O} - m2*(\text{denaturant})]/RT})]/[1 + (e^{[\Delta G_{1H2O} - m1*(\text{denaturant})]/RT}) +$$
$$(e^{[\Delta G_{1H2O} - m1*(\text{denaturant})]/RT})*(e^{[\Delta G_{2H2O} - m2*(\text{denaturant})]/RT})] \quad (30)$$

2.2.12 Hydrogen-deuterium exchange monitored using mass spectrometry

As mentioned in section **1.3** continuous- and pulse-labelling DXMS are techniques used to probe structural dynamics as well as study protein folding. Local and global changes, induced by amongst others, ligand binding, protein-protein interactions, protein modification/s and addition of denaturants can be monitored using DXMS.

2.2.12.1 Continious labelling DXMS

2.2.12.1.1 Pepsin immobilization and packing

Pepsin digestion prior mass analyses of deuterated protein samples increases the spatial resolution of DXMS. In addition, immobilized pepsin packed into a stainless steal column increases the extent of protein digestion, decreases the digestion time and increases the reproducilbility of the DXMS experiment (Wang *et al*., 2002). It is important to note that certain support matixes cause substantial deuterium back-exhange resulting in unusable data (Wu *et al*., 2006).

Pepsin was immobilized using the protocol described by Wang and co-workers (2002). 200 mg porcine pepsin was dissolved in 2 ml of 50 mM sodium citrate buffer pH 4.4. The solution was loaded on a pre-equilibrated G-25 buffer exchange column and eluted with approximately 8 ml of the same buffer. 660 µl ALD coupling solution, 3 ml pepsin solution and 0.1 g POROS-20AL support matrix were mixed in a 25 ml glass beaker to form a homogeneous suspension. Next, 2.3 ml of 1.5 M Na_2SO_4 was added dropwise to the reaction mixture in a period of 5 minutes. Another 4.6 ml of 1.5 M Na_2SO_4 was added dropwise in a time span of 2 hours with the reaction being stirred gently. The suspension was than trasferred to a centrifuge tube and incubated for 18 hours at 4 °C. Throughout this time the tube was placed on a rotator to ensure gentle mixing. To quench the reaction 1 ml of 1.0 M ethanolamine was added and the solution was gently mixed for 5 hrs at 4 °C. Finally, free and immobilized pepsin were separated by washing the support in a sintered glass funnel with 50 ml of 50 mM citrate buffer pH 4.4, 50 ml of 1.0 M NaCl (in 50 mM citrate buffer pH 4.4) and 50 ml of 50 mM citrate buffer pH 4.4. The clean immobilized pepsin was resuspended in 50 mM citrate buffer pH 4.4 in a 1 : 1 ratio and stored at 4 °C.

Immobilized pepsin was packed into a stainless steel guard column with bed volume of 400 µl. The column was assembled such that the frit on the one end was removed. The immobilized pepsin was than sucked into PEEK tubing that in turn was connected to the frit-free end of the guard column. An HPLC pump was used to flow 0.1 % formic acid through the PEEK tubing, thus packing the immobilized pepsin slurry into the column. The flow rate was gradually increased from 1 ml/min until the back pressure reached 2,500 p.s.i. (the maximum pressure recomended for the POROS support matrix). The flow was mantianed for 10 – 15 min before decreasing it slowly to 0 p.s.i. The frit was replaced and the packed column was washed with 0.1 % formic acid at 9 ml/min for 5min.

2.2.12.1.2 Sample preparation

The DXMS experiment starts with determing conditions that produce the highest number of peptides providing the best seqeunce coverage. Due to time constraints, conditions that resulted in the best fragmentation pattern for wtCLIC1 (Nathaniel, PhD unpublished data) were also empolyed in the case of CLIC1-E81M and CLIC-

M32A. These consisted of a buffer containing final concentrations of 0.8 % formic acid, 17 % glycerol, 0.5 M (in the case of CLIC1-M32A) and 1.0 M (in the case of CLIC1-E81M) GuHCl. The samples were run at 0.1 ml/min through a pepsin colunm with bed volume of 400 µl so that the total contact time between protein and pepsin was 4 min.

Next, on-exchange samples were prepared where CLIC1-E81M and CLIC1-M32A were incubated with deuterium. As a rule of thumb each sample needs to contain an average of 50 µg protein and a ratio of protein to deuterated buffer to quench buffer of 1 : 3 : 6. In addition all samples, with the exception of the fully-deutearted control, and buffers were pre-chilled on ice and prepared at 4 °C in cold room. 15 µl of ~ 3.5 mg/ml (~ 120 µM) CLIC1-M32A/CLIC1-E81M in CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.02 % NaN_3 pH 7.0) were incubated with 45 µl deuterated buffer (10 mM Na_2HPO_4, 150 mM NaCl in 99.9 % D_2O, pD 7.0) for varying time periods[Ψ] (10, 30, 100, 300, 1000 and 3000 sec). After the allocated time, the hydrogen-deuterium exchange reaction was stopped by adding 90 µl quench buffer (0.8 % formic acid, 17 % glycerol, 0.5 M GuHCl or 1.0 M GuHCl for CLIC1-M32A and CLIC1-E81M respectively, pH 2.3) and incubating the reaction mixture for 1 min with gentle mixing. The samples, final volume of 150 µl, were aliquoted in triplicates and stored at – 70 °C until further analysis.

In addition to the on-exchange samples a non-deuterated and fully-deuterated controls were also prepared. For the non-deuterated control 15 µl of ~ 3.5 mg/ml (~ 120 µM) CLIC1-M32A/CLIC1-E81M in CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.02 % NaN_3 pH 7.0) were incubated with a mixture of 45 µl non-deuterated buffer (10 mM Na_2HPO_4, 150 mM NaCl, pH 7.0) and 90 µl quench buffer (0.8 % formic acid, 17 % glycerol, 0.5 M GuHCl or 1.0 M GuHCl for CLIC1-M32A and CLIC1-E81M respectively, pH 2.3) to make up a final volume of 150 µl. The samples were aliquoted in triplicate and stored at – 70 °C. The fully-deuterated control was prepared the night before the rest of the samples. 15 µl of ~ 3.5 mg/ml (~ 120 µM) CLIC1-M32A/CLIC1-E81M in CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.02 %

[Ψ]: *In the case of CLIC1-M32A it was determined that samples prepared using deuterium incubation for 100 and 3000 seconds were corrupted. Hence, these time points could not be used in the final data interpretation.*

NaN_3 pH 7.0) were incubated with 45 µl fully-deuterated buffer (0.8 % formic acid in 99.9 % D_2O) for 16 – 18 hours at 25 °C. After the allocated time the exchange reaciton was stopped through the addition of pre-chilled 90 µl quench buffer (0.8 % formic acid, 17 % glycerol, 0.5 M GuHCl or 1.0 M GuHCl for CLIC1-M32A and CLIC1-E81M respectively, pH 2.3) followed by 1 min incubation on ice, aliquoting in triplicate and storing at – 70 °C.

2.2.12.1.3 Equipment configuration and sample analysis

Sample analysis was performed at the School of Medicine, University of California under the supervision of Dr. V. Woods. The system used for analysis of continous deuterium labelled CLIC1-M32A and CLIC1-E81M was set up so that protein digestion was done on-line by connecting a column packed with immobilized pepsin (see Figure 6 for details). This reduces back-exchange by decreasing the anaylsis time. Other meassures taken to ensure minimal back-exchange were keeping samples, columns, valves and the bulk of the connecting PEEK tubing immersed in ice within a refrigirator at 4 °C. The highly-automated configuration consists of two HPLC systems (Schimadzu LC-10AD, operated by Schimadzu SCL-10A pump controller) each with two pumps. The first HPLC delivers the thawed samples at 0.1 ml/min through the pepsin column to the C18 column using 0.05 % TFA. It also uses the same solvent to backflush the immobiilzed pepsin at 1 ml/min after sample digestion (Figure 6). Once pepsin-digested protein was bound to the C18 column the second HPLC delivered buffers at 0.05 ml/min for gradient elution (**Pump A**: 0.05 % TFA; **Pump B**: 80 % ACN, 0.01 % TFA). The peptides were eluted using a linear ACN gradient (**Pump B**: 5 – 45 % at 0.05 ml/min). Initial peptide identification was done using collision induced dissociation (CID) on a Finigan LCQ electrospray ion trap mass spectrometer in data-dependent MS2 (tandem MS) mode with capillary temperature at 200 °C. All subsequent samples, including the non-deuterared and fully-deutarated controls as well as 10 – 3000 sec time points, were analysed using a Waters Quadropole Time-Of-Flight (Q-Tof) Premier electrospray mass spectrometer in MS mode and W-optics.

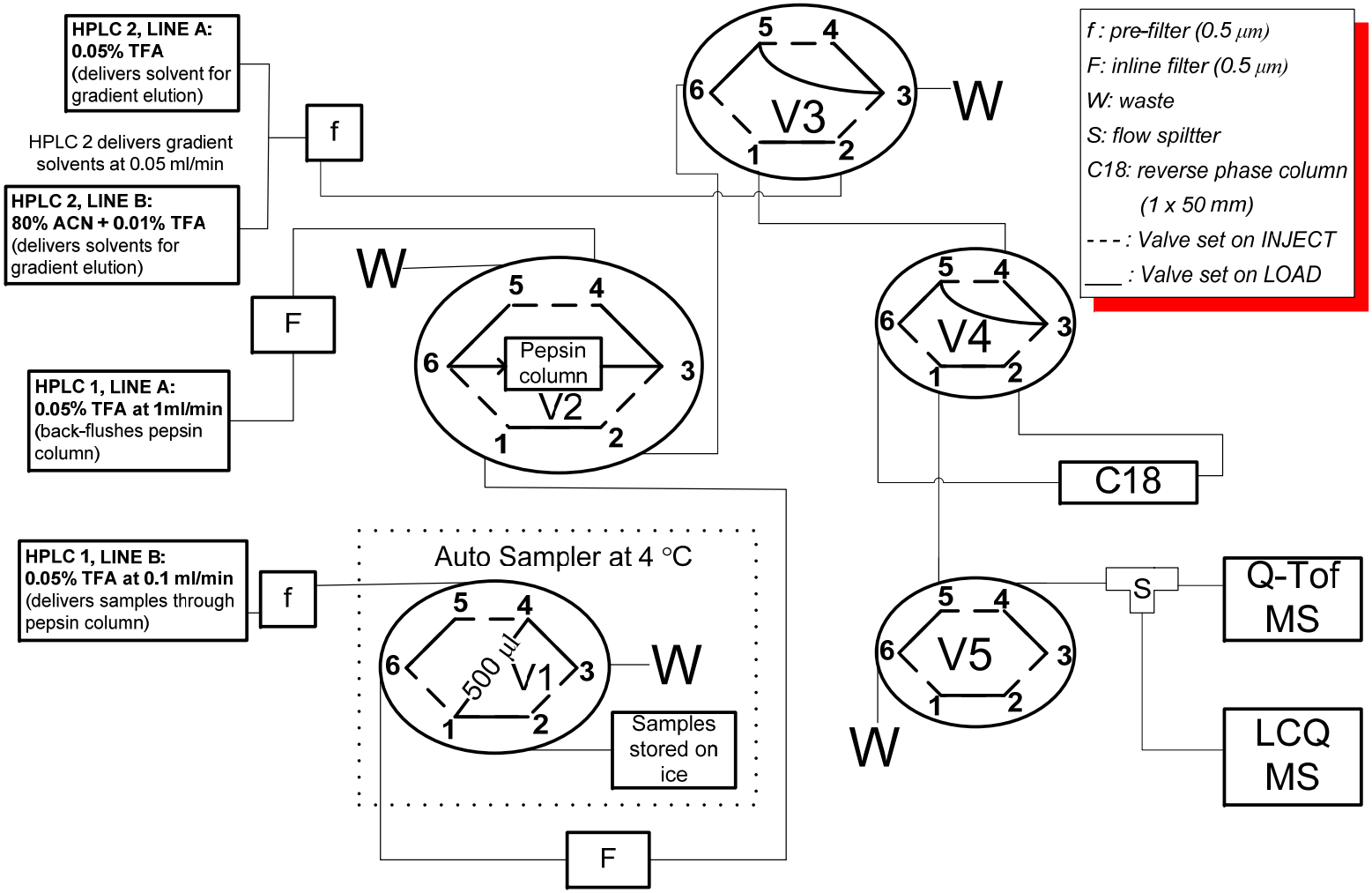


Figure 6: Equipment configuration for analysis of continuous-labelled DXMS samples

The system is set up so that protein samples are pepsin-digested on-line by connecting a column packed with immobilized pepsin at switch-valve 2 (V2). It must be noted that samples, columns, valves and the bulk of the connecting PEEK tubing were kept on ice within a refrigerator at 4 °C in order to minimize deuterium back-exchange. There are three major valve settings (1) **Sample loading on auto-sampler loop**: V1 = load, V2 = load, V3 = load, V4 = inject, V5 = load; (2) **Loading sample from loop to C18 column**: V1 = inject, V2 = inject, V3 = inject, V4 = load, V5 = load; (3) **Sample elution from C18 and analysis via mass spectrometer (MS)**: V1 = inject, V2 = load, V3 = load, V4 = load, V5 = inject. The diagram was generated using Microsoft Visio software

2.2.12.1.4 Data manipulation

As mentioned in **2.2.12.1.2**, spectral data from pepsin digested peptides were collected using MS2 mode. The sequence of each peptide was then identified using the Sequest software programme (Thermo Finigan Inc) which maps the raw spectral data to the sequences of CLIC1-M32A or CLIC1-E81M. The resulting peptide pool was quaility checked using specialized software (Sierra Analytics, LLC, Modesto, CA). Parametrs such as retention time, m/z range, centroid value and mapping score were scrutinized.

The level of deuterium incoroporation for peptide X was calculated by subtracting the centroid value of molecular isotope of partially deuterated peptide X from the ceontroid value of natural abundance isotope of non-deuterated peptide X as per the method devised by Zhang and Smith (1993). The procedure was automated through the use of the specialized software used to quality-check the peptide pool (Sierra Analytics, LLC, Modesto, CA). Sub-localization of deuterium was performed next, where the partially deuterated peptides (10 – 3000 sec) were mappped on to the primary sequence of CLIC1-M32A or CLIC1-E81M. The level of peptide overlap determines the resolution of this step i.e. multiple overlaping fragmets can narrow the position of deuterium localization within a single residue. Corrections for back-exhange were made by employing the methods used by Zheng and Smith (1993):

$$D_O = (m - m_{0\%})/(m_{100\%} - m_{0\%}) \text{ X } N \qquad (31)$$

where $\mathbf{D_O}$ is the average number of deuteriums per peptide at the time of the analysis, **m** is the average mass of partially deuterated peptide X, $\mathbf{m_{0\%}}$ is the average mass of non-deuterated peptide X, $\mathbf{m_{100\%}}$ is the average mass of fully-deuterated peptide X and **N** is the number of peptide amide linkages in peptide X.

$$N = T_N - 2 - T_{Pro} \qquad (32)$$

$\mathbf{T_N - 2}$ is the number of residues of peptide X minus the first two amino acid that can not retain deuterium, while $\mathbf{T_{Pro}}$ is the number of proline residues found in peptide X. Equation **(31)** was shown to introduce an error in $\mathbf{D_O}$. For three thousand peptides of random sequence and size this error was calculated to be at an averge of 5.5 % with a standard deviation of 5.5 % (Zeng and Smith, 1993; supplementary material). Consequently, only differences of 10 % or higher were deemed as significant when comparisons were drawn.

2.2.12.2 Pulse labelling DXMS

2.2.12.2.1 Sample preparation

Similarly to continious labelling, each sample prepared for pulse labelling DXMS needs to contain an average of 50 µg protein and a ratio of protein to deuterated buffer to quench buffer of 1 : 3 : 6 (see section **2.2.12.1.2**). During this study, urea was used as a denaturant. 10 M stock urea was prepared as described by Pace *et al.,* 1986, in the appropriate CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.0 2% NaN_3 pH 7.0 or pH 5.5). 120 µl of ~ 3.5 mg/ml (~ 120 µM) wtCLIC1 or CLIC1-M32A in CLIC1 storage buffer (50 mM Na_2HPO_4, 1 mM DTT, 0.02 % NaN_3 pH 7.0 or 5.5) were unfolded in 0 – 8 M urea for 60 – 90 min to reach equilibrium. The samples were then pulse-labelled through the addition of 360 µl deuterated buffer (50 mM Na_2HPO_4, 0 – 8 M urea in 99.9 % D_2O, pH 7.0 or 5.5; urea concentrations for each sample was pre-checked using a refractometer). The exchange reaction was stopped after 10 sec.[9] by 1 min. incubation with quench buffer (0.8 M GuHCl; 0.8 % formic acid, 17 % glycerol, pH 2.3). Samples were aliquoted in triplicate and stored at – 70 °C for further analysis. The absorbance at 340 nm of approximately 100 µl of each reaction mixture was recorded using a Beckman DU-600 spectrophotometer to check for possible presence of protein aggregates. Non-deuterated and fully-deuterated controls were made as per **2.2.12.1.2**.

2.2.12.2.2 Equipment configuration and sample analysis

Sample analysis was performed at the Mass Spectrometry Unit, Stellenbosch University, South Africa, under the supervision of Dr. M. Stander. Details of the configuration used to analyse deuteriun pulse-labeled samples are shown in Figure 7. Samples, peptide trap, valves and the bulk of the connecting PEEK tubing were buried in ice in order to minimize back-exchange. The configuration consists of two HPLC systems. HPLC1 (Agilent 1100 series) delivered 50 – 100 µl samples to the peptide trap at 0.2 ml/min using 0.05 % TFA. Trap-bound wtCLIC1 and CLIC1-M32A were desalted for 2 min. using 0.05 % TFA. HPLC2 (Acquity UPLC) was used to elute bound samples using a flow rate of 0.2 ml/min with solvent containng 80 % ACN,

[9]: *In the case of wtCLIC1 at pH 5.5 the hydrogen-deuterium exchange reaction was quenched after 300 seconds in order to account for the ~ 30 fold decrease in exchange rate as the pH decreases from 7.0 to 5.5 (*Bai *et al.,* 1993*).*

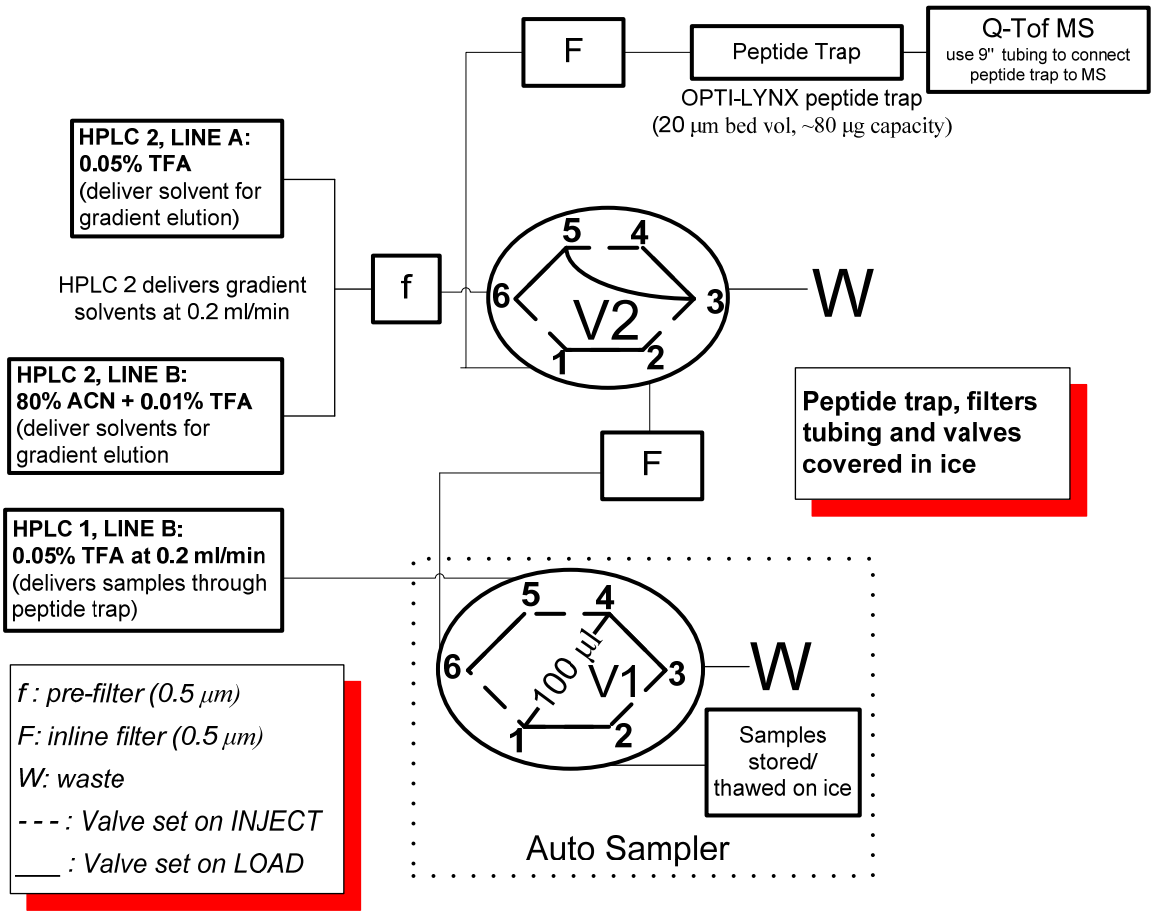


Figure 7: Equipment configuration for analysis of pulse-labeled DXMS samples

No pepsin digestion was performed using this set-up. Intact protein was retained by a peptide trap so that buffer salts, which interfere with the MS signal, can be washed away. Advantage of the trap over a C18 column is that the analysis time is reduced by a shorter elution time. The peptide trap, valves and the bulk of the connecting PEEK tubing were kept buried in ice in order to minimize deuterium back-exchange. There are three major valve settings (1) **Sample loading on auto-sampler loop**: V1 = load, V2 = inject; (2) Loading sample from loop to peptide trap: V1 = inject, V2 = load; (3) **Sample elution from peptide trap and analysis via MS**: V1 = inject, V2 = inject. The figure was generated using Microsoft Visio software.

0.01 % TFA over a period of 4 min. Spectral data was collected using a Waters Q-Tof Premier electrospray mass spectrometer in MS mode and V-optics. The cone- and capillary-voltages in the case of wtCLIC1 samples were set at 35 V and 80 V, respectively. On the other hand, for CLIC1-M32A these values were at 100 V and 150 V, respectively. This was done unintentionally and in future studies the higher values should be used since a better signal intensity was obtained for CLIC1-M32A samples. Initial analysis used wide scans (700 – 1999 m/z) to record all populated charge states of wtCLIC1 and CLIC1-M32A. Consequently, narrow scans (1120 – 1150 m/z) were used to record the evolution of the 24^+ charge state, which gave the best signal-to-noise ratio in the case of CLIC1-M32A, with increasing urea concentrations.

2.2.12.2.3 Data manipulation

The multiple envelopes of isotope peaks in the mass spectra of deuterium pulse-labelled wtCLIC1 and CLIC1-M32A in 0 – 8M urea were used to identify all species present at equilibrium. The relative areas of these peaks were used to quantify the populations of the various states. The isotopic envelopes were fitted to Gaussian areas using the software PeakFit (AISN Software. Inc.). The level of deuterium incoroporation for the range of species present at equilibrium was determined as per the Zheng and Smith (1993) method with centroid values obtained using Peakfit (AISN Software. Inc.). Corrections for back-exchange that occurred during HPLC analysis were performed as described in **2.2.12.1.4** The maximum number of exchangable amides (**N**) for wtCLIC1 and CLIC1-M32A was determined to be 229, considering that CLIC1 has 243 residues of which 14 are prolines.

CHAPTER 3

RESULTS

As previously mentioned, in this study, two approaches were undertaken to examine the role of domain-domain interfaces in the stability, folding and functioning of CLIC1: a bioinformatics-based and an experimental investigation. Section **3.1** deals with the structural alignment of domain interface residues belonging to members of the GST protein family. In addition, the domain interface composition of GSTs was analysed using computational techniques. The aim was to compare and contrast the domain interface architecture of CLIC1 to those of other GST proteins in order to extract structural clues that point to the dual functioning of CLICs as cytosolic and membrane proteins. Sections **3.2** to **3.5** deal with the experimental approaches used in determining the structural dynamics of CLIC1 as well as the thermodynamic significance of inter-domain contacts.

3.1 Domain interfaces in the GST super-family

3.1.1 Multiple structural alignment of domain interfaces

The domain interfaces of proteins belonging to the GST super-family, including CLIC1 and CLIC4, were analyzed. In this study, the thioredoxin-like domain was named N-domain and the all α-helical domain was referred to as C-domain. The interfaces belonging to the N- and C-domains were labeled as the N- and C-interface. The domain-interface residues of 40 GST super-family proteins were structurally aligned using MultiProt (Shatsky *et al.,* 2004). Each aligned spatial point along the consensus GST-family interface was given a position number. Positions with $Cr_{x\text{-}GST}$ higher than 0.5 were defined as highly conserved and labeled as computational hot-spots (see section **2.2.1** equation **3** for Cri calculation). Positions where the $Cr_{x\text{-}GST}$ values were calculated to be between 0.3 and 0.5 were defined as moderately conserved. Domain interface computational hot-spots and moderately conserved interface residues are shown in the Appendix **Table A**. The majority of inter-domain contacts were found to be non-polar with approximately 60% of the hot-spots being hydrophobic in character. The N-interface was found 70% more conserved than the C-interface. From the 40 structurally aligned domain interface positions, 28 belonged to the N-interface, 1 was found in the domain-linker and only 10 were found in the C-interface.

Three or more neighboring/contacting hot-spots were classified as a hot-region. Figure 8A shows a ribbon diagram depicting three hot-regions found in 40 GST members that were structurally aligned using MultiProt (Shatsky *et al.,* 2004). The three hot-regions form a consensus GST-family N-/C- interface. Hot-regions 1 and 2 belong to the N-interface, while hot-region 3 is found in the C-interface. A number of the hot-spots in the consensus GST domain interface are complementary i.e. the side chain of hot-spot 6 packs against the side chain of hot-spot 24 and the side chain of hot-spot 33 digitates between the side chains of hot-spots 28 and 24 (see Figure 8A). In addition, structurally conserved domain-interface residues are not randomly spread out along the interface but assemble into three densely packed areas. This can be seen in Figure 8B, a 3D-scatter plot generated by using the Cα co-ordinates of the hot-spot residues shown in Figure8A. Hot-spots whose side chains protrude out of instead of into the domain interface are included as part of the domain interface because their backbones provide the domain interface scaffold. In addition, positions 10, 26, 27, 32 and 34 have very high $Cr_{x\text{-}GST}$ values of 0.84, 0.87, 1, 0.92 and 0.95 respectively (see Appendix **Table A**). It is likely that these residues are structurally important and possibly interact with other conserved residues in the protein's interior.

Particular attention must be drawn to hot-spot number 8 in the consensus GST-domain interface (see Figure 9). This position forms part of a conserved lock-and-key type motif previously studied in Alpha class GSTs (Wallace *et al.,* 2000). The $Cr_{x\text{-}GST}$ of hot-spot 8 is 0.67 (see Appendix **Table A**). The majority of the 40 aligned residues, found at position 8, have bulky, hydrophobic side chains (see Appendix **Table A**) that extend from the N-domain into a preserved hydrophobic pocket found in the C-domain. The C-domain pocket is formed via the side-chains of residues at positions 31, 35, 37, 38 and 39 (see Figure 9 and Appendix **Table A**). These expel the bulk of the solvent resulting in 0 to 5 % solvent accessible surface area of hot-spot 8 residues. It must be noted that the orientation of the side chains of position 8 amino acids varies within the different GST classes and in some cases it does not form contacts with residues at position 31 or 38. In addition, the make up of the C-domain pocket fluctuate within the various GST classes in particular the CLICs (see section **3.1.2**).

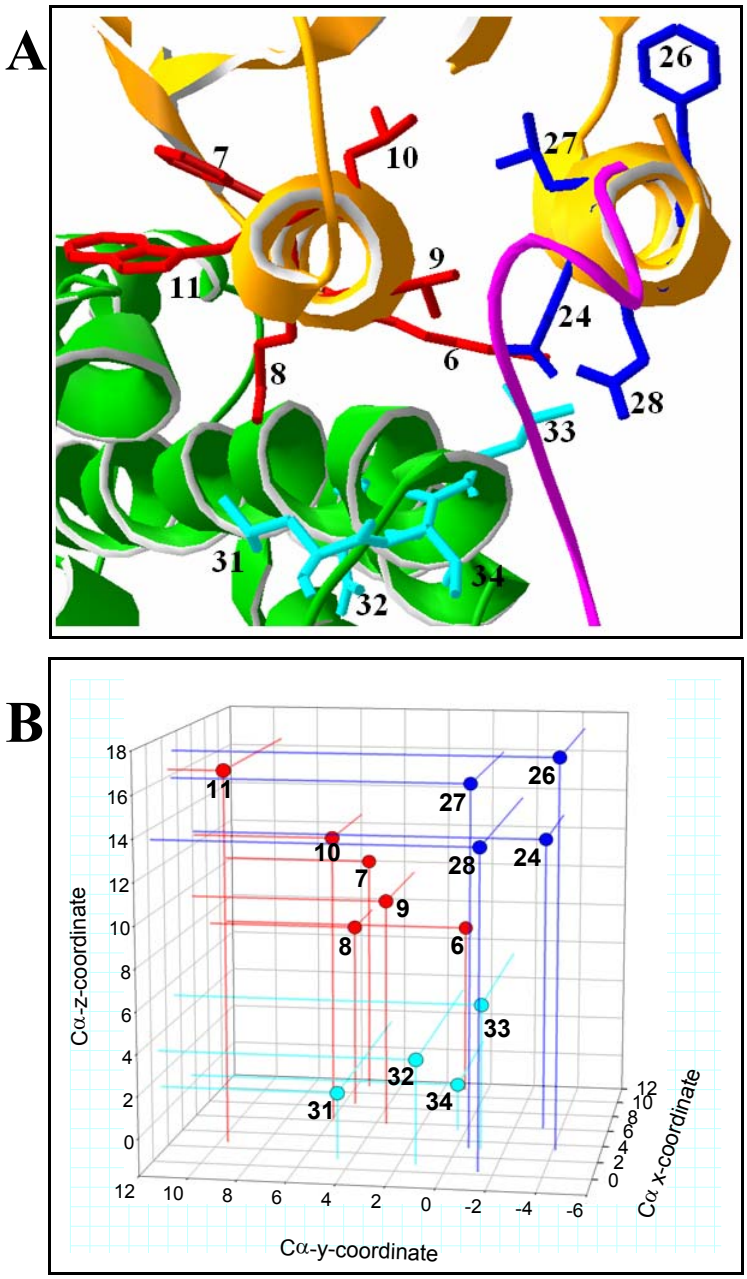


Figure 8: Hot regions in the domain interface of GST proteins

(A) Ribbon diagram illustrating three hot-regions (red, blue, cyan) in the domain interface common to 40 members of the GST family. A hot-region is made up of three or more neighbouring/contacting computational hot spots. The individual hot spots are numbered according to their position along the consensus interface (see Appendix **Table A**). The N-terminal domain is coloured orange while the C-terminal domain is green with the inter-domain linker in pink. The structure used to represent the positions of the consensus hot-regions is that of CLIC1 (pdb code: 1k0m). The diagram was generated using SwissPdb viewer (Guex and Peitsch, 1997).

(B) 3D-scatter plot generated using the Cα-coordinates of the hot-spot residues shown in **(A)**. The hot-regions and hot spots are coloured and numbered as in **(A)**. The data was plotted using SigmaPlot v9.0.

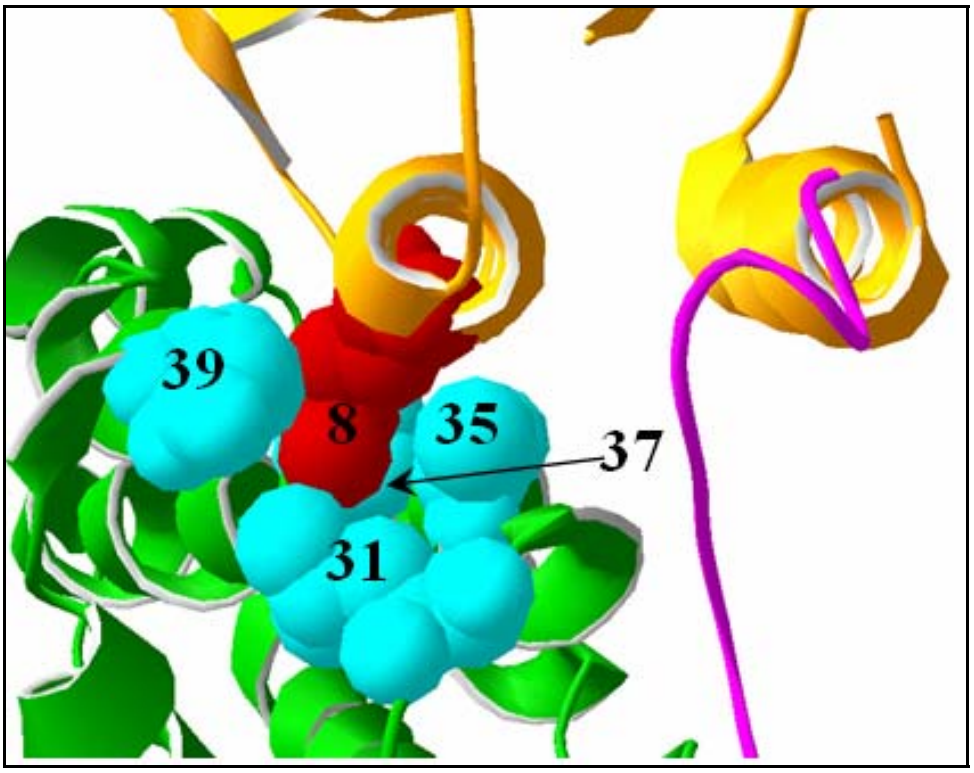


Figure 9: Conserved domain interface lock-and-key motif in the GST family

Ribbon diagram illustrating the conserved lock-and-key type motif found in the domain interfaces of GST proteins. The N-terminal domain is coloured orange while the C-terminal domain is green with the inter-domain linker in pink. The residues are numbered according to the structural alignment shown in Appendix **Table A**. The orientation of the side chain of residues at position 8 and the make up of the C-domain pocket may fluctuate for the different GST classes. The structure used to represent the inter-domain lock-and-key motif is that of CLIC1 (pdb code: 1k0m). The diagram was generated using SwissPdb viewer (Guex and Peitsch, 1997).

3.1.2 Domain interface interactions unique to CLIC family

Soluble CLIC1 and CLIC4 are structural homologues of GST family proteins (Harrop *et al.,* 2000; Littler *et al*., 2005). However, the ability of CLIC proteins to insert into membranes indicates that their structures contain a specific mechanism/s that, under the correct conditions, enables these proteins to convert from soluble to membrane-competent form. In order to identify the residues involved in the soluble-membrane transition one needs to contrast the sequences of CLIC proteins to the rest of the GST family. In theory, amino acids that are conserved in the CLICs and in the GST super-family but differ in their side chain chemistry may have diverse functions in CLICs as opposed to the rest of the GSTs.

Twelve amino acid sequences belonging to proteins from the CLIC family were aligned with the aid of T-Coffee (Notredame *et al*., 2000). Residues that formed part of the domain interface were identified using the crystal structures of CLIC1 and CLIC4. Next, using the conservation ratios at these positions and the conservation ratios at equivalent positions in the domain interface of GST family proteins a uniqueness factor (Uf) was calculated. Figure 10 is a scatter plot of Uf values as a function of CLIC1 domain residues. The Uf values give an indication of conserved domain interface positions that are unique to the CLIC family. In order to resolve specific from non-specific positions, the mean and standard deviation of the domain interface amino acid population were calculated. Positions with Uf values below the mean were considered as non-specific. Those with an Uf value falling between the mean and standard deviation were deemed as moderately specific. Positions with Uf values above the standard deviation were defined as highly specific to the CLIC family. The iMOLTalk server (http://i.moltalk.org) was used to identify interactions that moderately and highly specific CLIC1 residues made at the domain interface. Figure 11 illustrates these unique CLIC family interactions. Attention is drawn to the network of contacts formed between Glutamic acid 81 (Glu81) and Arginine 29 (Arg29) as well as Glutamic acid 85 (Glu85) and Lysine 37 (Lys37). These interactions were deemed particularly important since pH change, and therefore charge, has been shown to be a key factor in the functioning of CLIC proteins as membrane ion channels (Warton *et al.,* 2002). Glu81-Arg29 and Glu85-Lys37 form inter-domain contacts with residues from the C-terminal h5 as well as domain linker

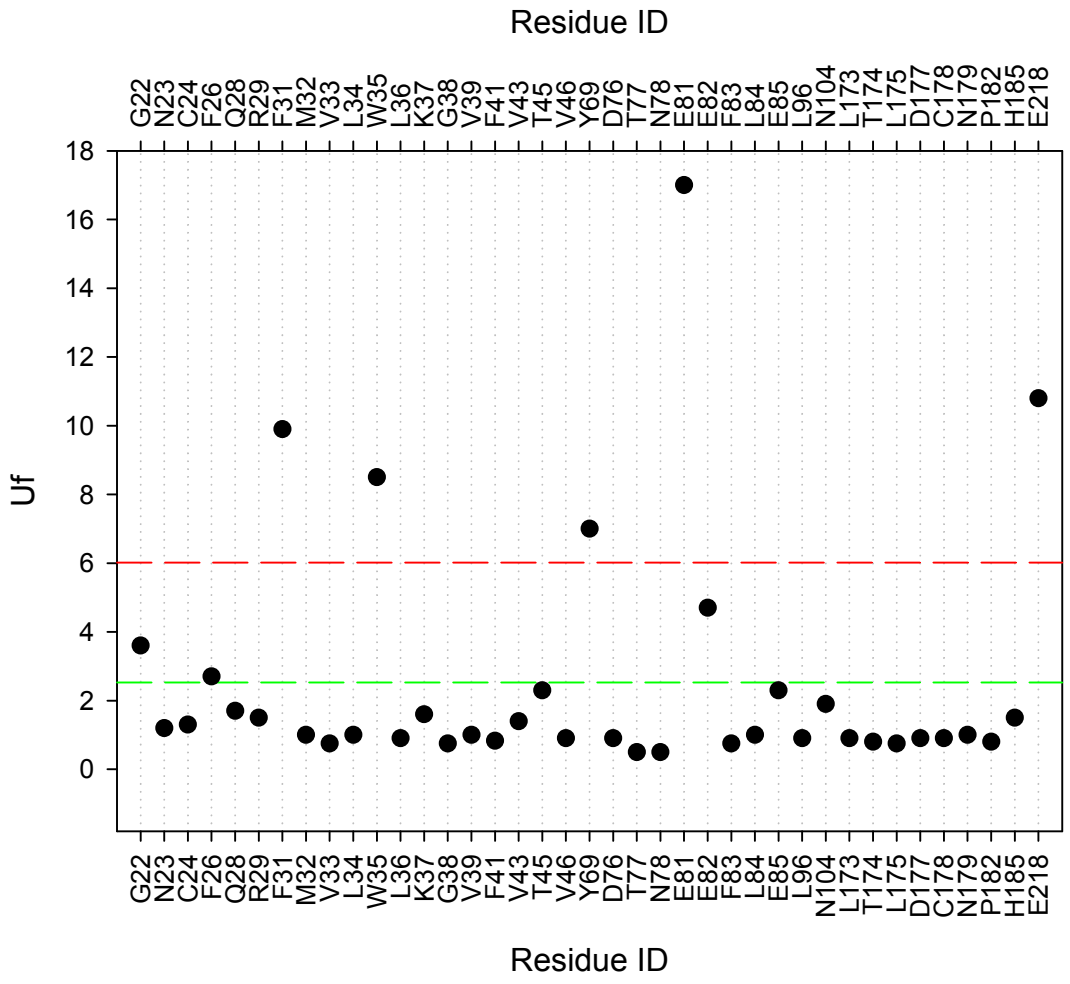


Figure 10: Uniqueness factor (Uf) of CLIC1 residues that form part of the GST family consensus domain interface

CLIC1 residues that form part of the GST family consensus domain interface are plotted as a function of Uniqueness factor (Uf). Uf, calculated as per the method shown in section **2.2.1** and equation **6**, indicates whether a particular conserved position is specific to the CLIC family and thus may play a role in structural rearrangements responsible for transforming CLIC proteins from soluble to membrane bound form. Positions with Uf values below the mean (green dash) are considered as non-specific. Positions with Uf values found between the mean and standard deviation (red dash) are moderately specific. Uf values above the standard deviation represent positions that are highly specific to the CLIC family. Residue numbering is according to that of CLIC1 (pdb code: 1k0m).

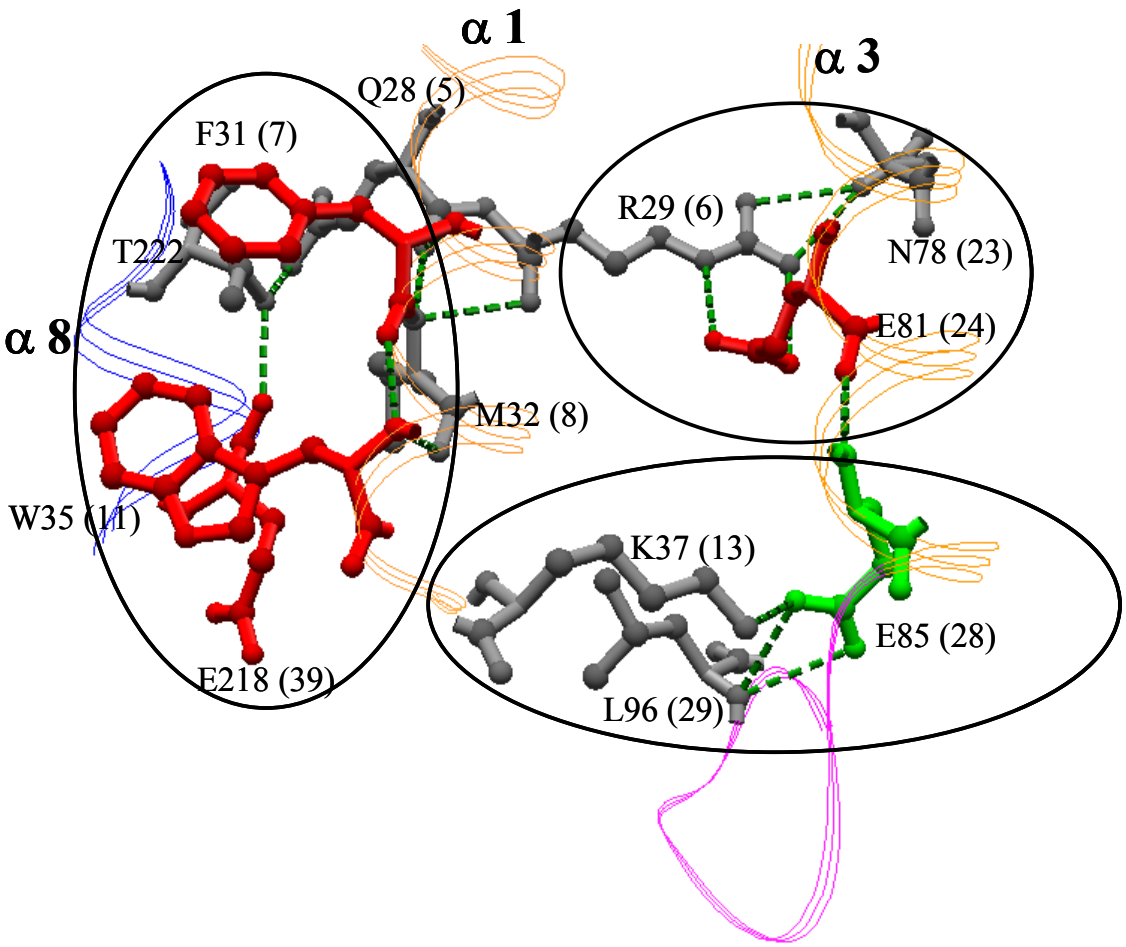


Figure 11: Domain interface network unique to the CLIC family

SwissPdb generated diagram of contacting CLIC1 domain interface residues with Uf values lower than 2 (grey), between 2 and 6 (green) and higher than 6 (red). The N-terminal domain alpha helices 1and 3 (orange), C-terminal domain alpha 8 (blue) and the domain linker (pink) are indicated. Residue numbering follows that of CLIC1 (pdb code: 1k0m) with numbers in brackets indicating positions at the consensus GST-family domain-interface (see Appendix Table A). The hydrogen-bond network is shown in green. The circled residues form interacting networks unique to the CLIC family and thus may form part of the mechanism responsible for the soluble-membrane-bound conversion.

amino acids. In addition, they link h1 and h3, which in turn form the N-domain interface of CLIC1. A buried salt-bridge is classified as such when the average accessible surface area (ASA^A) of contacting amino acids is less than 20 %. The ASA^A of both Glu81 – Arg29 and Glu85 – Lys37 was found to be ~ 14 %.

The environment surrounding salt-bridges play a role in the overall stabilizing/destabilizing effects of these interactions. In the case of the Glu81 – Arg29 salt-bridge the majority of residues within 4 Å contained polar side chains. These included Thr77, Asn78, Lys79, Glu82 and Asn179. On the other hand, the interaction between Glu85 – Lys37 was found to be surrounded mainly by hydrophobic amino acids. The only polar side chains within 4 Å of Glu85 – Lys37 belonged to Lys95 and Glu82. Two hydrogen bonds were formed between side-chain charged groups of Glu81 – Arg29, while only one hydrogen bond was detected between the charged side-chain groups of Glu85 – Lys37. Salt-bridge geometry is another important parameter in determining the stability of these interactions. This factor is characterized by the distance between side-chains of residues involved in a salt-bridge as well as the orientation of these atoms in relation to each other (Kumar and Nussinov, 1999). In both Glu81 – Arg29 and Glu85 – Lys37, the charged side chains were less than 3 Å apart. In the case of Glu81 – Arg29, the orientation of the atoms involved in the formation of the two salt-bridges was approximately perpendicular. On the other hand, side chain charged groups that formed the interaction between Glu85 – Lys37 were oriented more linearly to each other at ~ 125 °.

The network of contacts that couple the N-domain h1 to the C-domain h8 of CLIC1 (N1 and C3 segments in Figure 12) are also shown in Figure 11. This group of contacts is important due to the presence of the conserved lock-and-key type motif (see section **3.1.1** and Figure 9) that couples the N- and C-domains of CLIC and GST family proteins. As previously mentioned, the C-domain pocket of the lock-and-key motif is made of residues at positions 31, 35, 37, 38 and 39 (see Figure 9) of the consensus GST domain interface. In the majority of GST proteins these positions have residues with non-polar side chains (see Appendix **Table A**). However, in the CLIC family, part of the C-domain pocket is formed by a conserved Glutamic acid (Glu218 in CLIC1; position 39 of consensus GST domain interface). Glu218 forms a charged

interaction with the side chain of Methionine 32 (Met32; the CLIC1 key residue that forms part of the conserved lock-and-key interaction).

3.1.3 Domain interface characterization

3.1.3.1 Interface segmentation and inter-domain contacts

The discontinuous nature of the GST domain interfaces was analyzed by grouping amino acids into segments. Amino acids separated by less than 10 residues were allocated to a segment. Figure 12 (A) and (B) shows the segments that make up the N- and C-interfaces of a typical GST-family protein. It was found that the N-C interface residues could be grouped into five segments. The N1 and N2 segments are found in the N-interface. The C-interface is divided into the C1, C2 and C3 segments. Visual inspection of the interface segments shows that over 90 % of the N-C interface is α-helical with only one β-strand found in the N1 segment (see Figure 12 A and B). This classifies the domain interfaces of GST-family proteins as α-class interfaces (Jones and Thornton, 1995). Interestingly, in terms of CLIC1 and CLIC4 the bulk of the inter-domain contacts are formed through the N1 segment (h1 in case of CLIC1 and CLIC4). On the other hand the N2 segment (h3 in CLIC1 and CLIC4) forms only one inter-domain contact with segment C2 (h6 in CLIC1 and CLIC4).

Hydrogen bonds and/or salt bridges connecting an N-interface residue and a C-interface residue were referred to as N-C hydrogen bonds and/or salt bridges (see Appendix **Tables C** and **D**). Only 20 % of the inter-domain hydrogen bonds were formed between non-polar residues. Of the 80 % of polar residues involved in N-C hydrogen bonds, 60 % were charged (basic or acidic). Thirty four of the 40 pairs of GST-interfaces, including that of CLIC1, were found to form hydrogen bond networks. This is a common phenomenon in proteins where each donor/acceptor is bonded to multiple acceptors/donors (Stickle *et al.*, 1992). The 40 analysed GST family proteins had a total of 2289 domain-interface residues that formed 316 inter-domain hydrogen bonds (0.14 bonds/residue), 194 inter-domain salt-bridges (0.08 bonds/residue), and in total, 510 inter-domain hydrogen bonds/salt bridges (0.22 bonds/residue). Figure 13 illustrates the break down of the above-mentioned contacts in terms of the various GST classes of proteins. The classes are arranged according to ascending number of total inter-domain hydrogen bonds/salt bridges. The monomeric

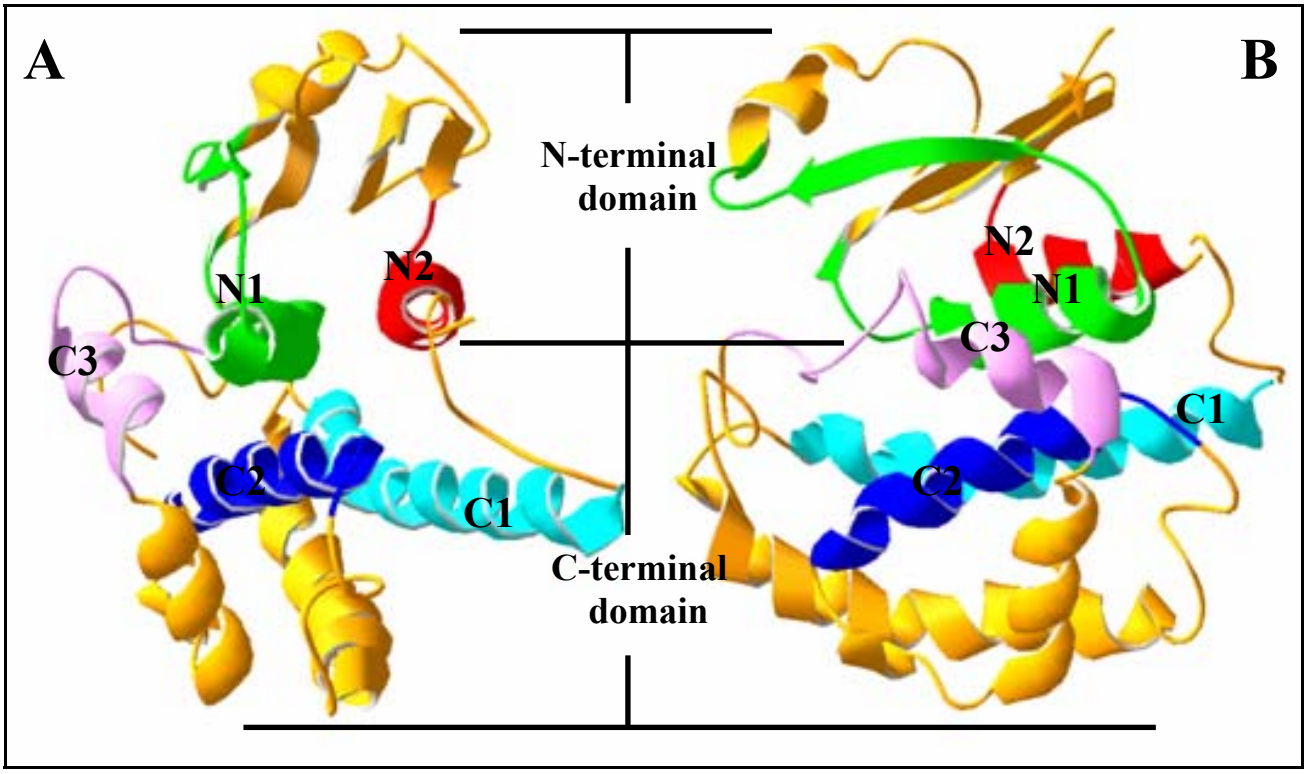


Figure 12: GST-family domain interface segmentation

Ribbon diagram showing a **(A)** front and **(B)** side view of the segments that make up the domain interface of a typical GST-family protein. The interface of the N-terminal domain is made up of two segments N1 (green) and N2 (red). The domain interface of the C-terminal domain consists of three segments C1 (cyan), C2 (blue) and C3 (pink). Interface residues separated by more than 10 residues were allocated to different segments. The structure used to represent the interface segmentation within the GST-family was picked randomly. It is that of Sj. GST (pdb file: 1gta). The model was generated using the Swiss-PDB viewer (Guex and Peitsch, 1997).

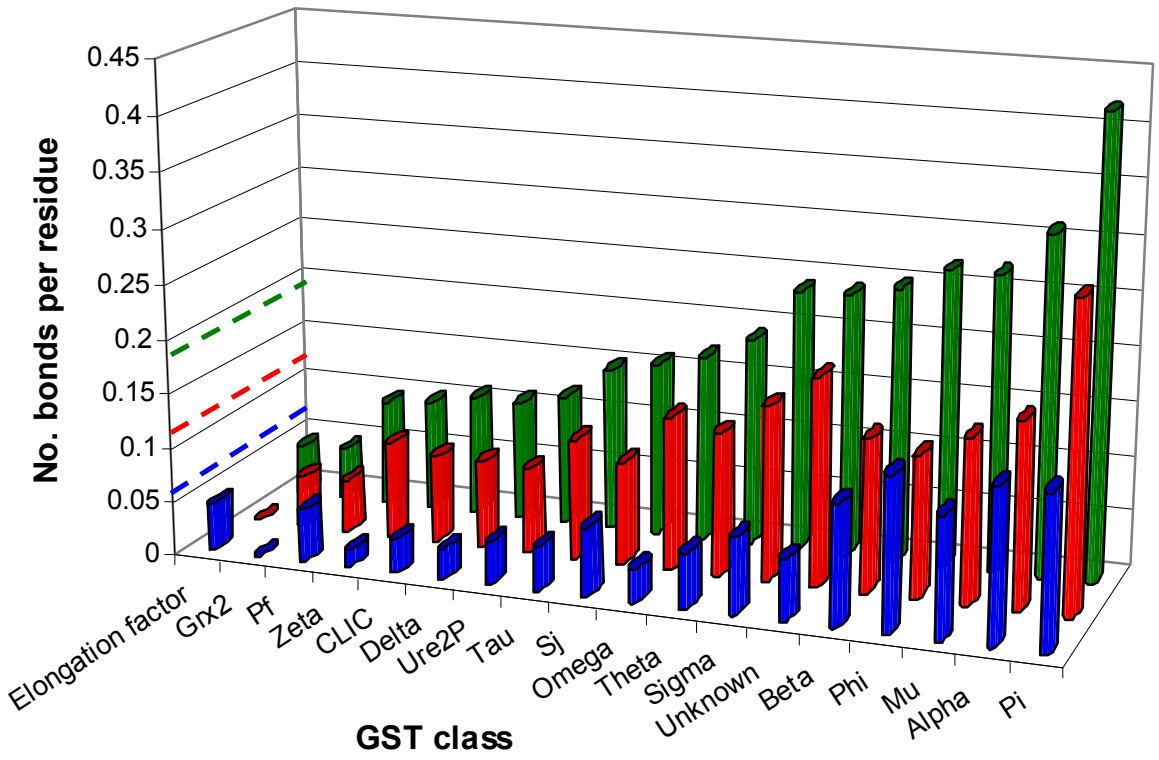


Figure 13: Inter-domain hydrogen bonds and salt-bridges per amino acid of GST class proteins

Bar chart depicting number of inter-domain salt-bridges (blue), hydrogen bonds (red) and sum of hydrogen bonds and salt bridges (green) per domain-interface residue. The x-axis shows the different GST classes in ascending order according to the sum of hydrogen bonds and salt bridges per residue. In cases where GST classes were represented by more than one protein member the average of the number of domain-interface residues, hydrogen bonds and salt bridges was used to plot the bar chart. The average number of inter-domain salt-bridges, hydrogen bonds and sum of salt bridges and hydrogen bonds for the GST family are shown as blue, red and green dashed lines respectively. Inter-domain contacts were identified using iMOL server (http://i.moltalk.org) and confirmed by PPI server (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html) as well as visual inspection of the GSTs crystal structures. The cut-off distance for H-bonds was 3.9Å (<90°), while that for salt bridges was 4.0 Å. The chart was generated using Microsoft Excel.

Grx2 and CLICs (CLIC1 and CLIC4 used for analysis) form far fewer inter-domain contacts than the average for the GST family (see Figure 13). It seems that the number of hydrogen bonds between the N- and C-interfaces is proportional to the interface size although, the correlation coefficient was very small ($R^2 = 0.27$, result not shown). The statistical significance of this correlation is supported by the observation by Jones and Thornton that the number of hydrogen bonds in dimeric interfaces is roughly proportional to the interface size (Jones and Thornton, 1995). In addition to the fact that, the average number of GST domain interface H-bonds (0.8 ± 0.38 per 100 $Å^2$ of iASA) is comparable to the average number of dimer interface hydrogen bonds (0.88 ± 0.4 per 100 $Å^2$ of iASA) as reported by Jones and Thornton (1995). No relationship was found between the number of N-/C-interface salt bridges and interface size (R^2 = 0.15, result not shown). Interface ionic interactions were less common than hydrogen bonds with an average of 5 ± 3.4 salt-bridges compared to an average of 8 ± 4.0 hydrogen bonds per domain interface. This can be attributed to the relatively few number of charged residues found in the N-/C-interfaces of GST proteins.

The N-C hydrogen bonds and salt bridges were assigned to an interface segment (Figure 14). The highest numbers of contacts exist between N1 and C3, followed by N2-C1 and N1-C2. No contacts were found between N2-C3. Hot-spot residues were not favored to form N-C hydrogen bonds and salt-bridges and no apparent conservation of these contacts was established. Visual inspection of the segments revealed that most inter-domain hydrogen bonds and salt bridges are found on the opposite side of the domain linker. This suggests that these non-local contacts act as pegs clamping the N- and C-domain segments together and possibly forming during the later stages of folding.

The percentages of polar and non-polar residues in the N- and C-interfaces of 40 GST proteins were recorded (see Appendix **Table B**). Figure 15 illustrates the mean percentage hydrophobic and hydrophilic residues found in the N- and C-interfaces. The percentage non-polar residue means of the N- and C-interfaces were found to be 65.4 % ± 7.1 and 67.5 % ± 4.5, respectively. The percentage polar residue means of the N-and C-interfaces were 34.6 % ± 7.2 and 32.4 % ± 4.2, respectively. Thus on average, the number of non-polar atoms are approximately double that of the polar

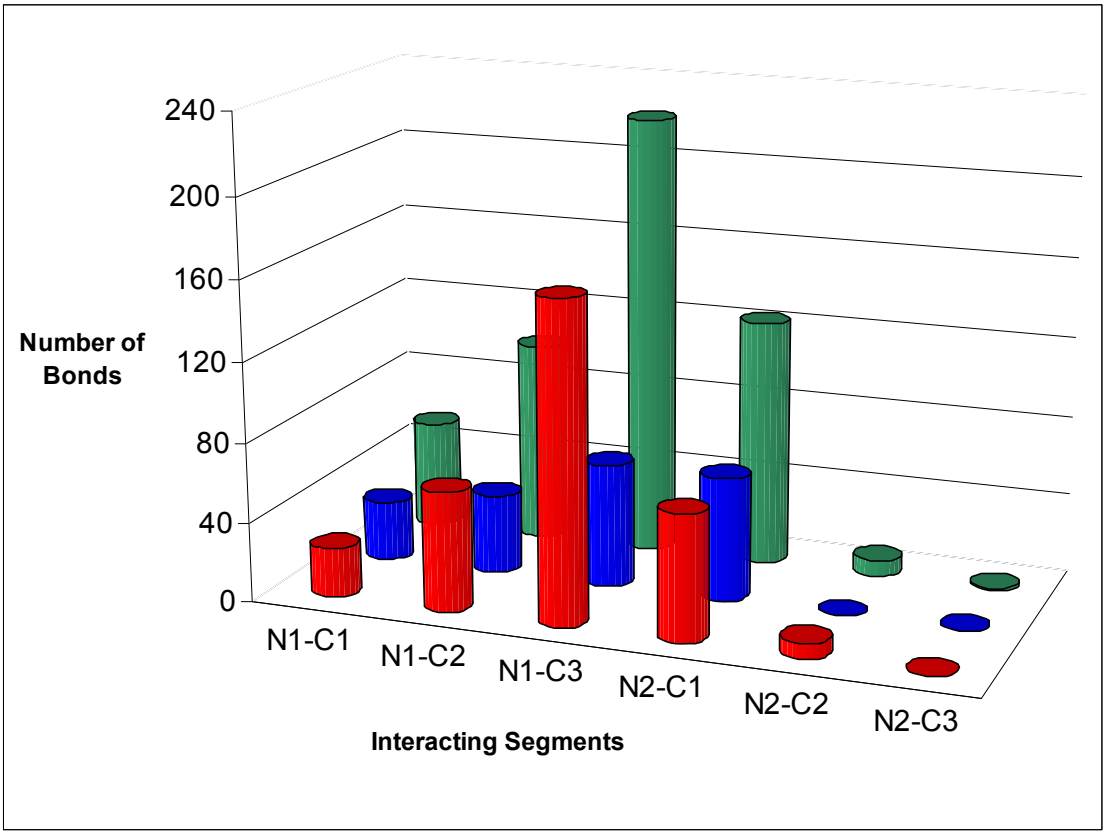


Figure 14: Total number of N-C interface hydrogen bonds and salt bridges in the GST-family

Bar chart depicting the total number of inter-domain hydrogen bonds (red), salt bridges (blue), and sum of hydrogen bonds and salt bridges (green) (see Appendix **Tables C** and **D** for full data). The x-axis, labelled interacting segments, shows the hydrogen bonds and salt bridges between the N-terminal (N1, N2) and C-terminal interface segments (C1, C2, C3). Inter-domain contacts were identified using iMOL server (http://i.moltalk.org) and confirmed by PPI server (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html) as well as visual inspection of the GSTs crystal structures. The cut-off distance for H-bonds was 3.9 Å (<90 °), while that for salt bridges was 4.0 Å. The chart was generated using Microsoft Excel.

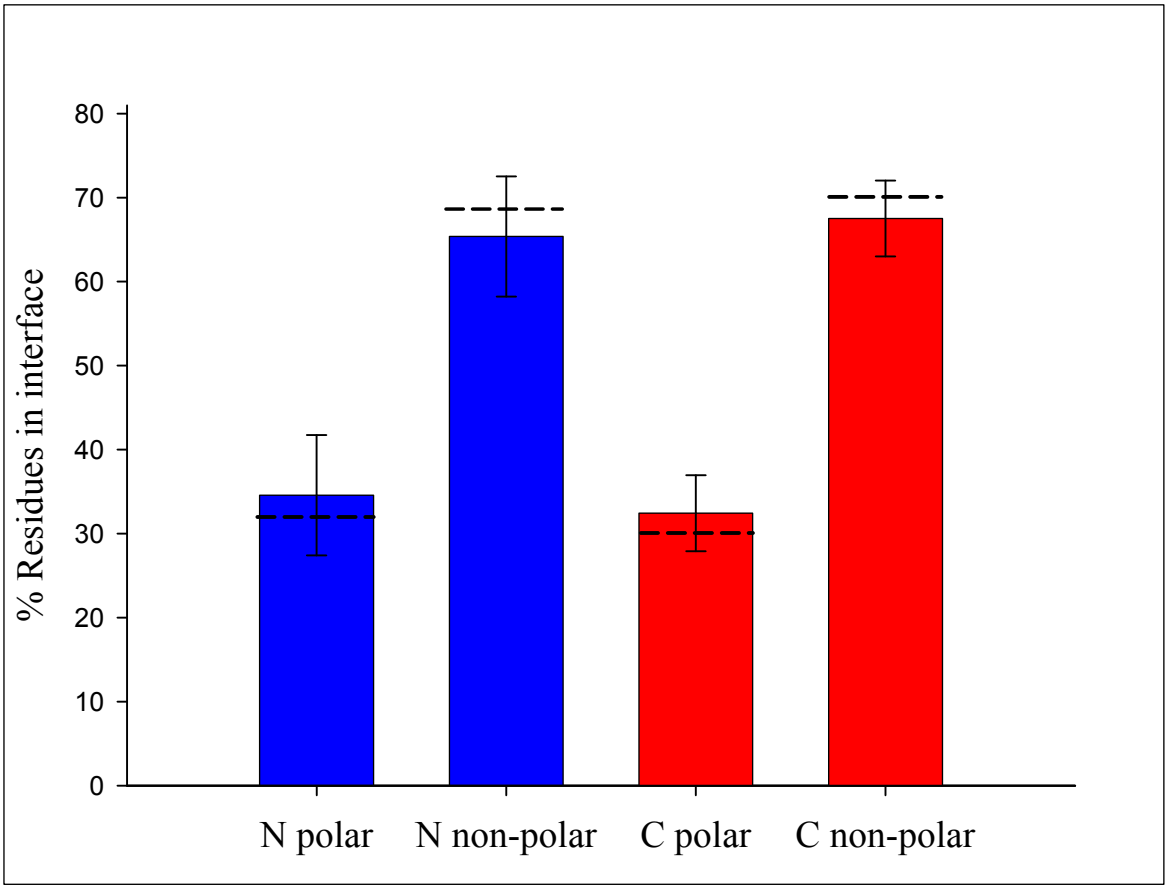


Figure 15: Hydrophobicity of the N- and C- interfaces belonging to GST-family members

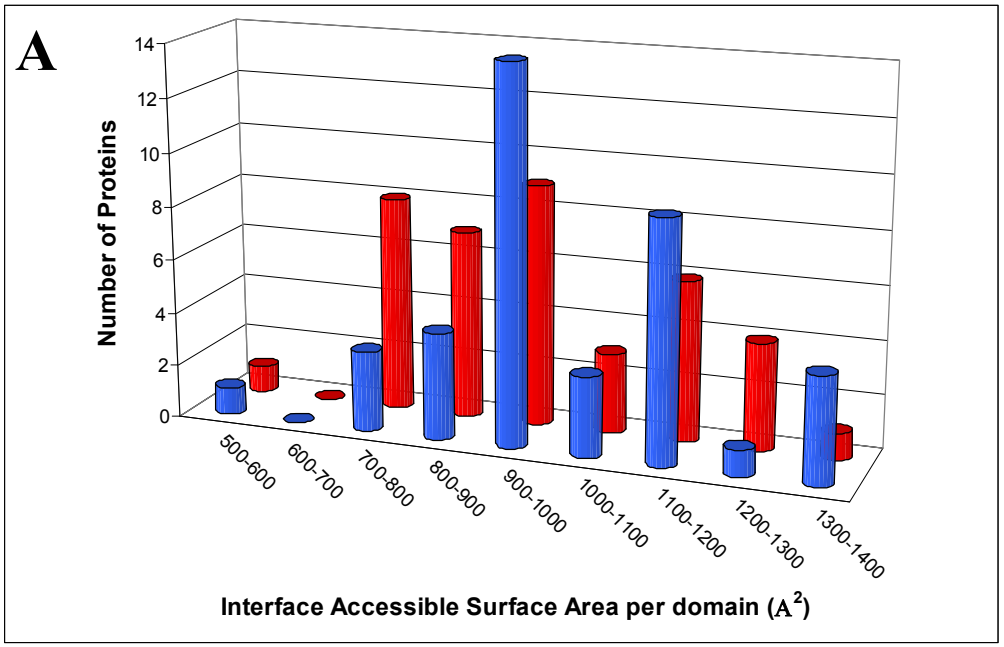
Bar chart depicting the percentage of polar and non-polar residues found at the N- (blue) and the C- (red) interfaces belonging to GST-family members. The graph was generated by plotting the means of the N- and C- percentage polar and non-polar interface residues (see Appendix **Table E** for full data). The error bars represent the samples standard deviation. The dashed lines indicate the averaged CLIC1 and CLIC4 values of percentage polar and non-polar residues. For each member belonging to the GST-family the percentage polar and non-polar N and C-terminal interface residues were calculated using PPI server (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html). The chart was plotted using SigmaPlot v9.0

atoms. The percentage of non-polar amino acids present at the domain interfaces of CLIC1 and CLIC4 is higher than the average percentage for the GST family (see Figure 15). Hence, the domain interfaces of CLIC1 and CLIC4 are somewhat more hydrophobic than the rest of the GSTs. This is also true for Grx2, the other monomeric GST structural homologue (74.9 % and 71.9 % non-polar residues for the N- and C-domains of Grx2, respectively).

3.1.3.2 Interface size, shape and complementarity

Figure 16A shows the interface accessible surface areas (iASA), per domain, which becomes buried upon domain association. Not surprisingly, the N-interface iASA correlates well with the C-interfaces iASA at a ratio of 1:1 (Figure 16B). Therefore, iASA values further mentioned in the text will represent the mean iASA buried by each domain. The 40 GST domain interfaces analysed showed an average iASA of 990.3 $Å^2$ (± 177.5 $Å^2$). CLIC1 and CLIC4 have lower iASA values at 892.9 $Å^2$ and 873.8 $Å^2$, respectively. These values are similar to the average ASA buried upon protein dimerization of 848 $Å^2$ ± 248 $Å^2$ but higher than the corresponding values for enzyme-inhibitor (ΔASA = 785 $Å^2$ ± 74 $Å^2$) and antibody- antigen (ΔASA = 777 $Å^2$ ± 135 $Å^2$) complex formation (Jones and Thornton, 1996). In comparison to the CLICs, monomeric Grx2 has a much more extensive iASA (994.1 $Å^2$). The iASA of analysed GST proteins were found to be proportionally related to the number of interface residues (R^2 = 0.61, result not shown). A similar linear relationship has been established in the case of other monomeric and oligomeric proteins (Jones and Thornton 1995; Stites 1997).

Figure 17 shows the N-and C-percentage interface accessible surface area buried upon domain association (%iASA) (see Appendix **Table E** for full data). The %iASA gives an indication of what fraction of the total domain surface is used in inter-domain contacts. The N-interface %iASA range between 11.5 % and 26.3 % and have an average of 20.0 % ± 3.2 %. On the other hand, the C-interface %iASA vary between 6.7 % and 15.8 % with a mean value of 12.3 % ± 1.9 %. Further, the N-%iASA correlates well with the C-%iASA at a ratio just below 2:1 (Figure 17B). Therefore, the fraction of N-domain surface involved in inter-domain contacts is about two-fold larger than the corresponding fraction of C-domain surface. This ratio is also



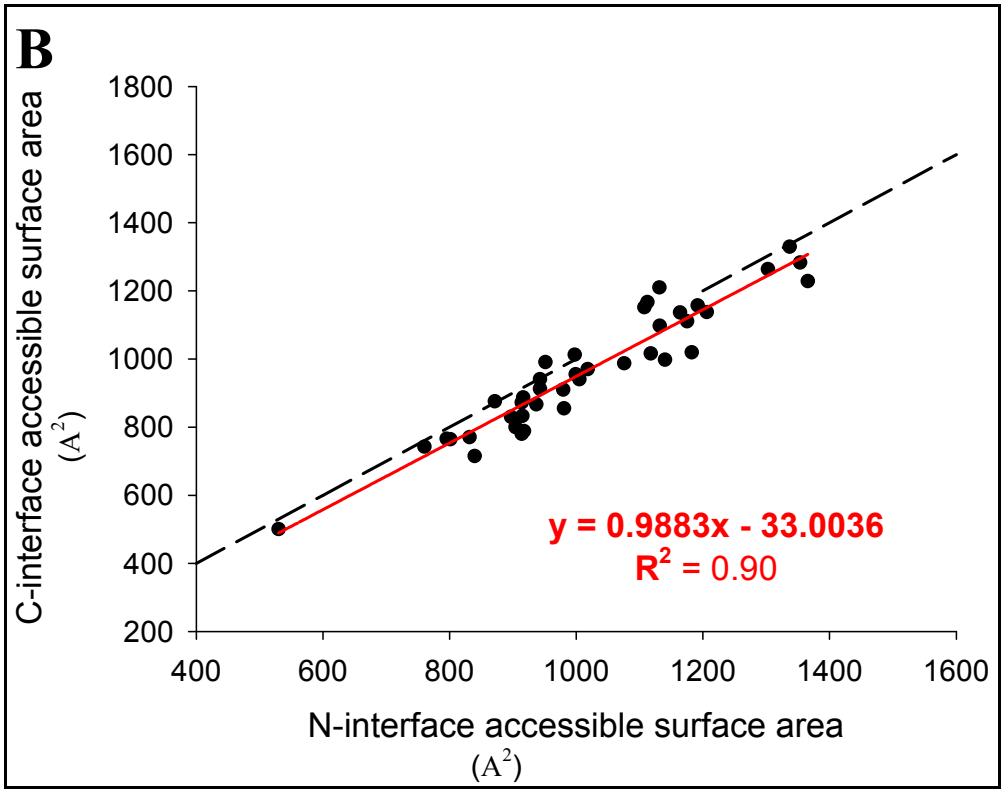


Figure 16: Distribution of the interface accessible surface area in the GST family

(A) Bar chart depicting the interface accessible surface area (iASA), per domain, upon domain association. The N-interface iASA is shown in blue, while the C-interface iASA is red. The iASA was calculated using PPI server (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html), where a probe sphere of 1.4 Å was rolled around the van der Waals surface of each protein. See Appendix **Table E** for full data.

(B) Correlation between the N-interface iASA and C-interface iASA. The dotted line traces N-iASA/C-iASA ratio of 1:1. The data was plotted using SigmaPlot v9.0 and fitted using a liner regression analysis.

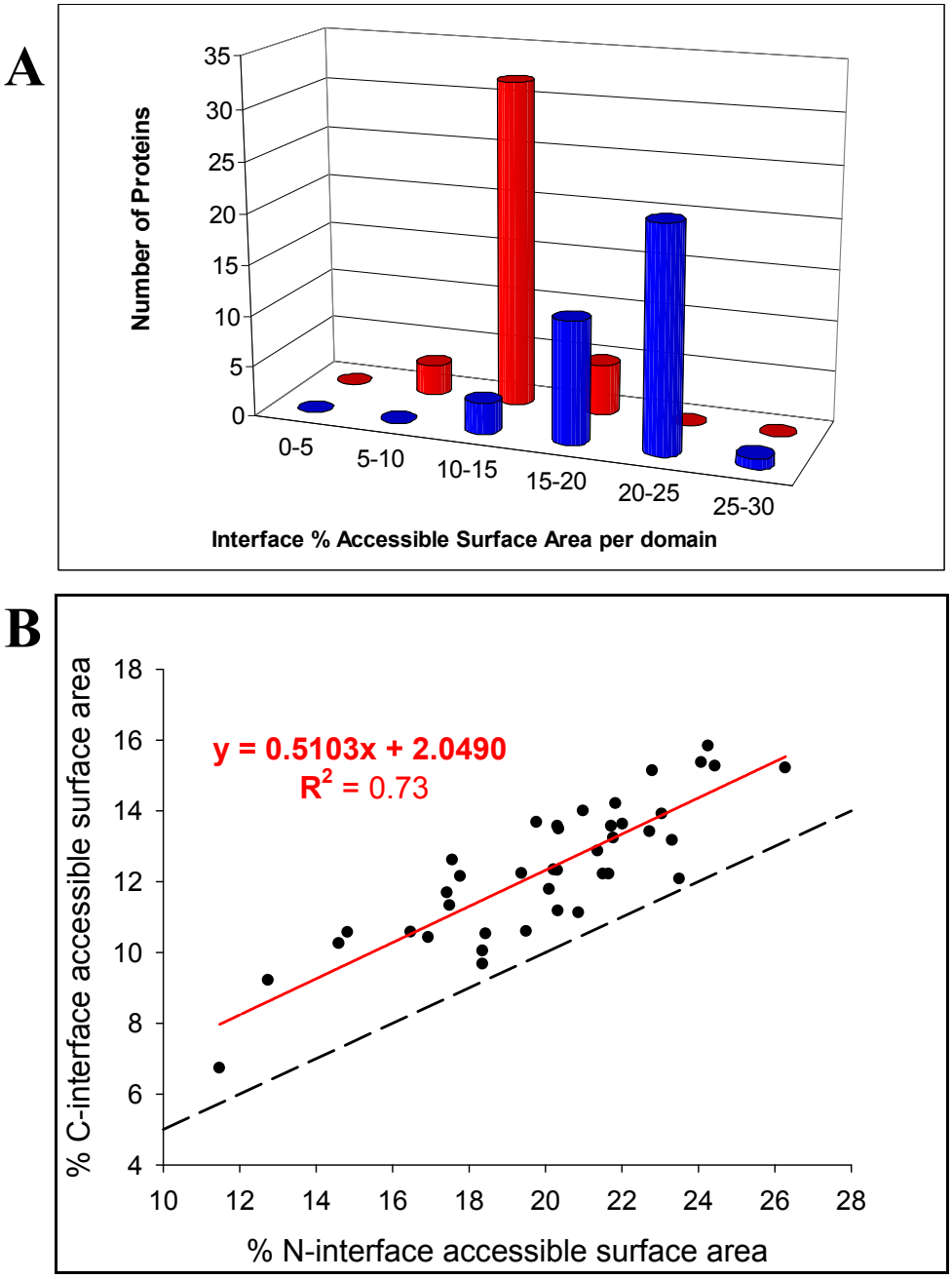


Figure 17: Distribution of the percentage interface accessible surface in the GST-family proteins

(A) Distribution of percentage interface accessible surface area (%iASA) in 40 members of the GST-family. The N-interface %iASA is shown in blue, while the C-interface %iASA is in red. The %iASA was calculated by dividing the N- or C-iASA by the corresponding total domain surface area. See Appendix **Table E** for full data.

(B) Correlation between the N-interface %iASA and C-interface %iASA. The dashed lines represent a ratio of 2:1. The data was fitted using SigmaPlot v9.0 and fitted using linear regression analysis.

maintained in the CLICs. However, both the N-%iASA and C-%iASA of CLIC1 (16.5 % N-%iASA and 10.6 % C-%iASA) and CLIC4 (16.9% N-%iASA and 10.4 % C-%iASA) are smaller than the corresponding average values for the GST family. In contrast, Grx2 the only other monomeric GST protein, has N-%iASA (22.7 %) and C-%iASA (13.4 %) values slightly higher than the GST family average (20.0 % N-%iASA and 12.3 % C-%iASA). The range of %iASA of the GST domain interfaces is within the range of the %iASA of dimer interfaces which were found to vary between 6.5 % and 29.4 % (Jones and Thornton, 1995; Stites, 1997).

Planarity gauges how far interface residues deviate form a plane, indicating how flat or how twisted an interface surface is. It is measured as the Root Mean Square (RMS) of deviation. The planarity of the interface surfaces were estimated by fitting a best-fit plane through the three-dimensional co-ordinates of the interface atoms. The deviation from this plane was plotted as the RMS (Å). Thus, the larger the RMS value the less planar the interface surface. Interfaces that deviate by more than 6 Å are classified as twisted (Jones and Thornton, 1995). The 40 GST-interfaces analysed were planar varying between 1.95 Å and 4.43 Å with an average value of 2.79 ± 0.09 Å for the N-interface and 2.89 ± 0.10 Å for the C-interface (Figure 18A and Appendix **Table E**). It is interesting to note that the CLICs (CLIC1 and CLIC4) have one of the plainest domain interfaces in the GST-family, second only to Elongation Factor protein (see Figure 18A). The other monomeric GST protein, Grx2, has significantly higher RMS values at 2.75 Å for the N-interface and 3.15 Å for the C-interface. Figure 18B is a scatter plot depicting the correlation between the N- and C-interface planarity values. The directly proportional relationship indicates the presence of symmetry between a pair of interacting interfaces. Hence, an N-interface protrusion is complemented by C-interface hollow resulting in lock-and-key type motifs along the domain interface. It was also found that the planarity values increase with increasing iASA (R^2 = 0.54; result not shown). This can be attributed to the fact that protruding side chains from the domain interface considerably increase the surface area buried upon domain association. Therefore, GST proteins with larger iASA are likely to have more lock-and-key contacts providing added stability.

The Length/Breadth ratio is a measure of the sphericity/circularity of an interface. A

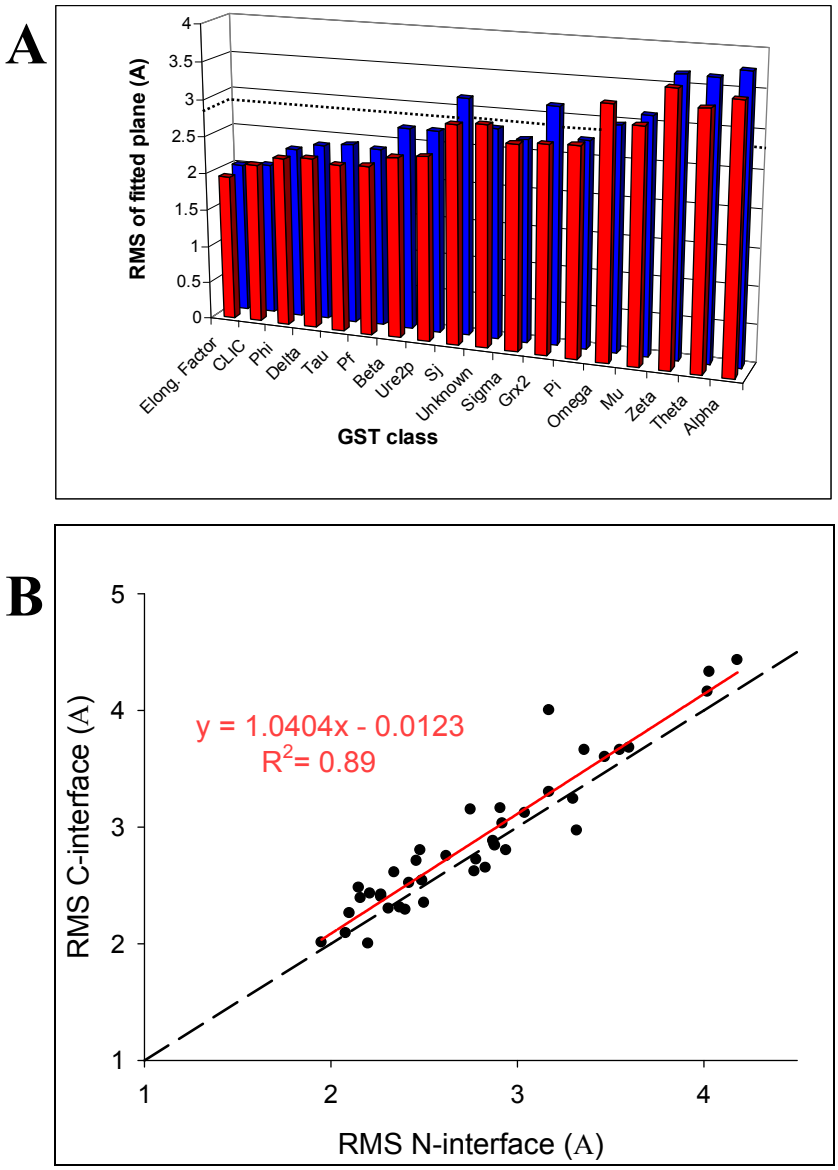


Figure 18: Planarity of the N- and C-domain interfaces in the GST-family

(A) Bar plot illustrating the root mean square (RMS) of best-fitted planes through the N- (blue) and C- (red) domain interfaces of 40 proteins belonging to the GST-family. The x-axis shows the different GST protein classes arranged according to ascending order of average N-/C-interface RMS. The dotted line represents the mean value for the plotted data. The larger the RMS value the less planar the interface surface. The interfaces are flat with approximately 70 % of them having an RMS between 2 and 3 Å. See Appendix **Table E** for full data.

(B) Correlation between planarity of the N- and C- domain interfaces. The dashed line traces a ratio of 1:1. The data was plotted using SigmaPlot v9.0 and fitted using a liner regression analysis.

ratio of 1 indicates an approximately circular interface. Figure 19 shows the length/breadth ratios of the GST domain interfaces (see Appendix **Table E** for full data). More than 90 % of the interfaces are relatively circular having length/breadth ratios above 0.5 with an average of 0.67 ± 0.15 for N-interfaces and 0.70 ± 0.16 for C-interfaces.

Figure 20 reports the gap-volume indexes of 40 GST-family interfaces. The Gap-volume index (see section **2.2.1** equation **4**) gives a measure of the complementarity of interacting surfaces. The smaller the gap-volume index the more complementary the interacting surfaces. The GST domain interfaces have an average gap-volume index of 2.14 ± 0.52. CLIC1 and CLIC4 have slightly less complementary domain interfaces than the majority of GST proteins with values of 1.89 and 1.70, respectively (see Appendix **Table E** for full data).

An additional factor used to assess the complementarity of GST domain interfaces was the number of bridging water molecules. A bridging molecule was defined as any water that formed a hydrogen bond with an N-interface residue and a second hydrogen bond with a C-interface residue. The domain interfaces are closely packed since no bridging water molecules were detected.

3.2 Verification of wild-type and mutant plasmid DNA

The structural and sequence alignments of GST and CLIC family proteins highlighted a number of domain interface interactions that may be critical in maintaining the structure as well as dual-form function of CLIC1 (see sections **3.1.1** and **3.1.2**). The conserved lock-and-key motif (see Figure 9) and the unique salt-bridge between Glu81-Arg29 (see Figures 10 and 11) were chosen as experimental case studies. Hence, two mutants CLIC1-M32A and CLIC1-E81M were designed in order to observe the effects of the removal of these domain interface interactions on the thermodynamic stability and structure of CLIC1.

The pGEX-4T-1 plasmids containing the open reading frames (ORF) encoding wtCLIC1 and the mutants CLIC1-M32A and CLIC1-E81M were sequenced. Figure 21 (A) and (B) shows portion of the wild-type CLIC1 ORF as well as fragments of the mutated ORF obtained from DNA sequencing. The numbers following the proteins'

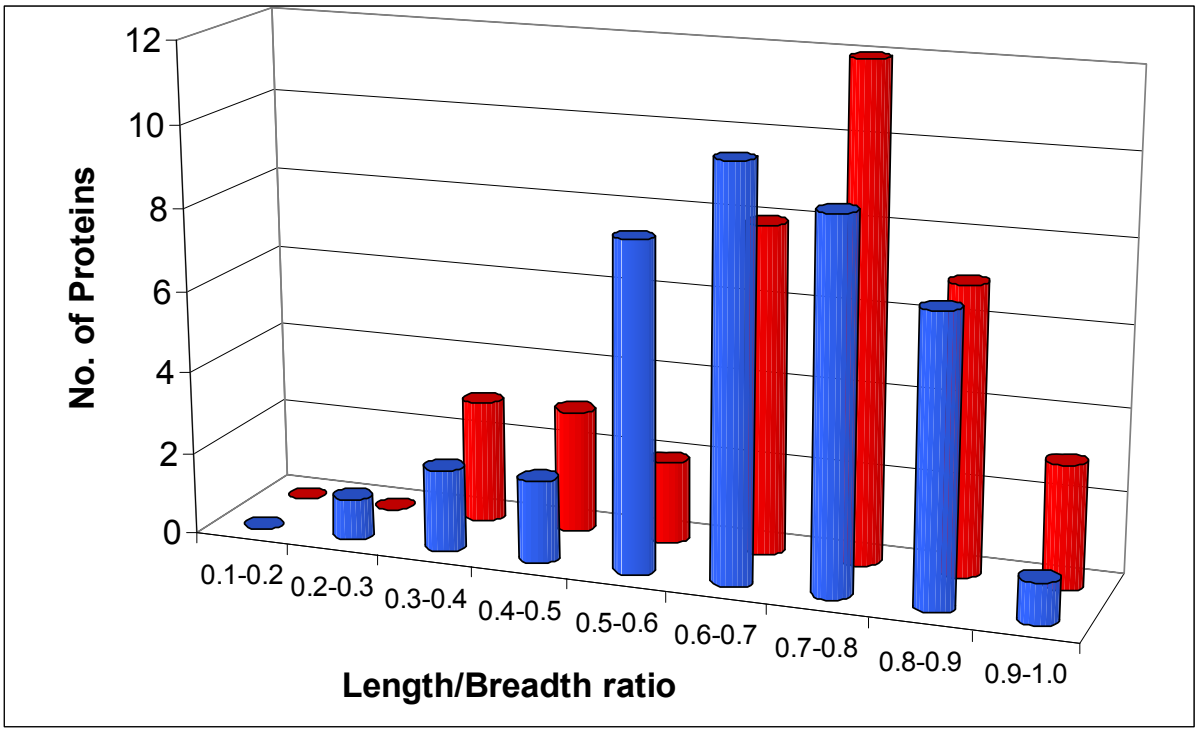


Figure 19: Length/Breadth ratio of the N- and C- interfaces in the GST-family

Bar plot illustrating the Length/Breadth ratios of the N- (blue) and C- (red) interfaces in the GST-family. The standard deviations in the *x* and *y* dimensions from the best-fitted plane are used as a measure of the length and breadth of the interface. The length/breadth ratio is the standard deviations in the *y* dimension divided by the standard deviations in the *x* dimension. The closer the Length/Breadth ratio is to one the more circular the interface. The smaller the Length/Breadth ratio the more extended the interface. The data was obtained from the PPI server (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html). See Appendix **Table E** for full data.

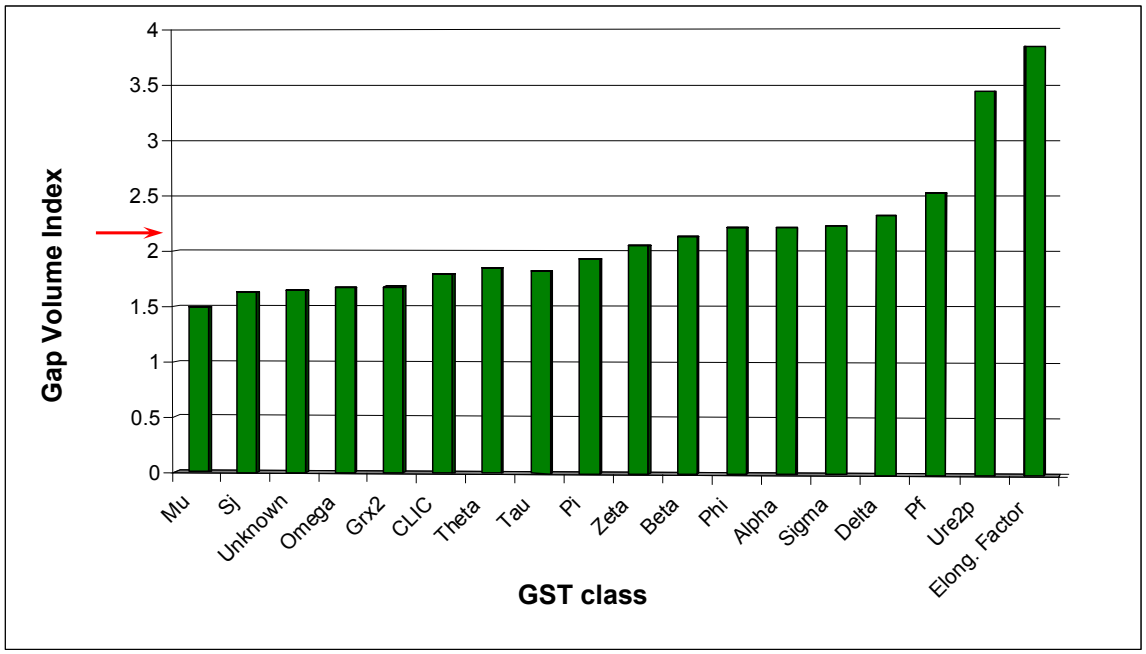


Figure 20: Gap volume index of the GST-family interfaces

Bar chart depicting the gap volume indexes of GST-family interfaces arranged in ascending protein-class order. The gap volume index is a measure of the complementarity of interacting surfaces. The smaller the gap volume index the closer the interacting pair of interfaces is to each other. The arrow indicates the mean value for the plotted population The PPI server (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html) that uses the programme SURFNET (Laskowski, 1991) calculated the gap volume indexes. See Appendix **Table E** for full data.

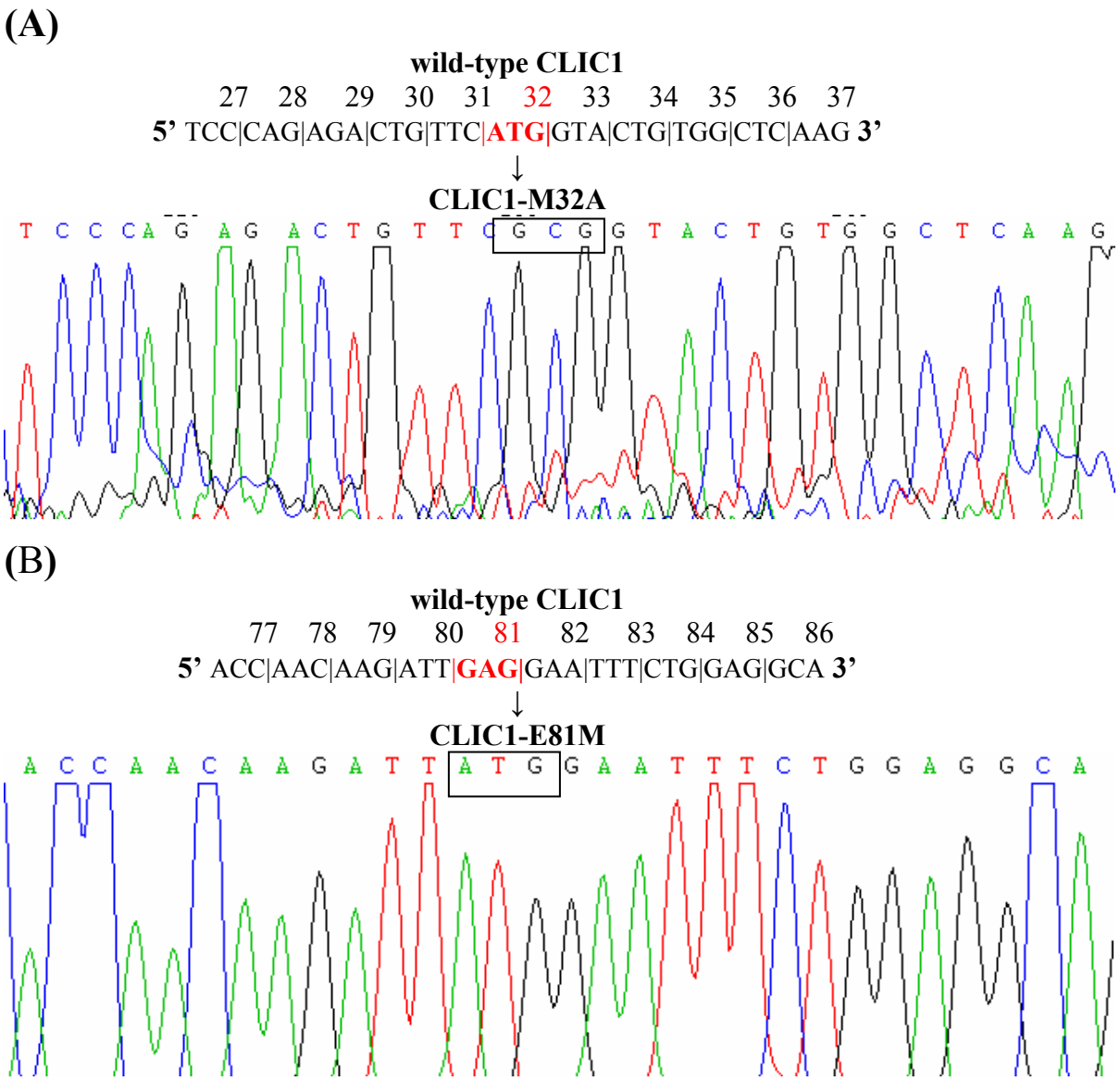


Figure 21: Mutant CLIC1 plasmid sequencing results

Fragments of the wild-type nucleotide sequences indicating the codon/s to be mutated (red). The vertical lines within each sequence designate the reading frame. The codon numbers are indicated on top of each nucleotide sequence. Part of the pGEX-4T-1 ORFs encoding the mutant proteins obtained from DNA sequencing which confirm the presence of the engineered mutation/s (boxed) are also shown. **(A)** CLIC1-M32A, **(B)** CLIC1-E81M.

names designate the codon number where the mutation was incorporated. The presence of the engineered mutation/s was confirmed and no other mutations were found to have been introduced by the thermal cycling reactions.

3.3 Protein over-expression

3.3.1 CLIC1-M32A

The solubility of GST-CLIC1-M32A in *E. coli* was assessed by SDS-PAGE using 15 % acrylamide gels. Figure 22 shows an intense band at approximately 49 kDa, corresponding in size to GST-CLIC1-M32A when the bacterial cultures were grown in the presence of 1 mM IPTG. In the absence of IPTG no over-expressed band was observed at 49.0 kDa (results not shown) The ProtParam tool at the EXPASY website (http://www.expasy.org), calculated the molecular weight of GST-CLIC1-M32A to be 52 kDa which is consistent with the induced band in Figure 22 lanes 1and 5. Further analysis of the SDS-PAGE gel reveals the presence of the over-expressed 49 kDa band in the insoluble/pellet fraction and the absence of the same band in the soluble/supernatant fraction (Figure 22 lanes 2 and 3). This shows that the 49 kDa fusion protein is insoluble when expressed at 37 °C. When the incubation temperature was lowered to 20 °C most of GST-CLIC1-M32A was found to be soluble, as indicated by the intense 49k Da band in the supernatant fraction (Figure 22, lane 7). At 20 °C expression, a prominent 35 kDa band is observed in the insoluble/pellet fraction (Figure 22, lane 6). This band is not present in the corresponding fraction when protein induction was performed at 37 °C (Figure 22, lane 2). Hence, the gene encoding the unknown 35 kDa, *E. coli* protein is temperature regulated.

3.3.2 CLIC1-E81M

GST-CLIC1-E81M was found to be soluble when expressed at either 37 °C or 20 °C. This is shown by the concentrated 49 kDa band, equivalent in size to GST-CLIC-E81M, present in the soluble fraction (Figure 23 (A) and (B) lanes 7 and 10). However, comparison of Figure 23 (A) and (B) reveals that the 49 kDa band present in the soluble fraction derived at 20 °C is more intense than the corresponding band derived at the 37 °C. This indicates that the yield of GST-CLIC1-E81M is higher when the protein is expressed at 20 °C due to superior solubility.

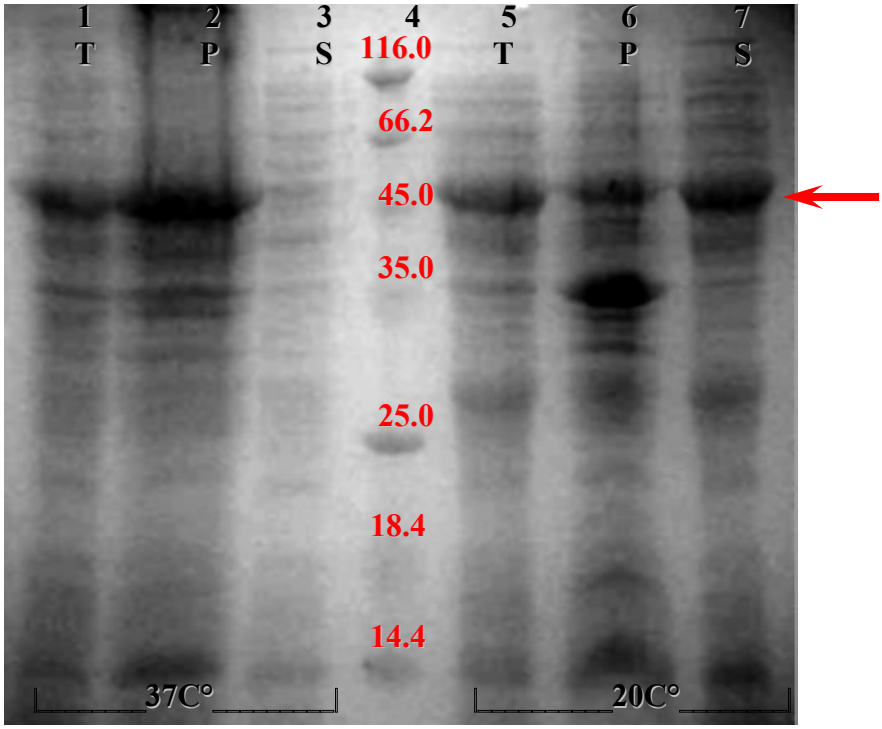


Figure 22: Solubility study of GST-CLIC1-M32A

15 % polyacrylamide SDS-PAGE gel showing the solubility of heterologously expressed GST-CLIC1-M32A at 37 °C and 20 °C. Lanes 1-3 depict the total lysate (T), insoluble/pellet fraction (P), and soluble/supernatant fraction (S) when the bacterial cells were grown and induced with 1 mM IPTG at 37 °C. Lane 4 shows the molecular weight marker (sizes in, kDa, are indicated in red). Lanes 5-7 illustrate total lysate (T), insoluble/pellet fraction (P), and soluble/supernatant fraction (S) when the bacterial cells were grown and induced with 1 mM IPTG at 20 °C. The arrow indicates the overexpressed, soluble GST-CLIC1-M32A fusion protein.

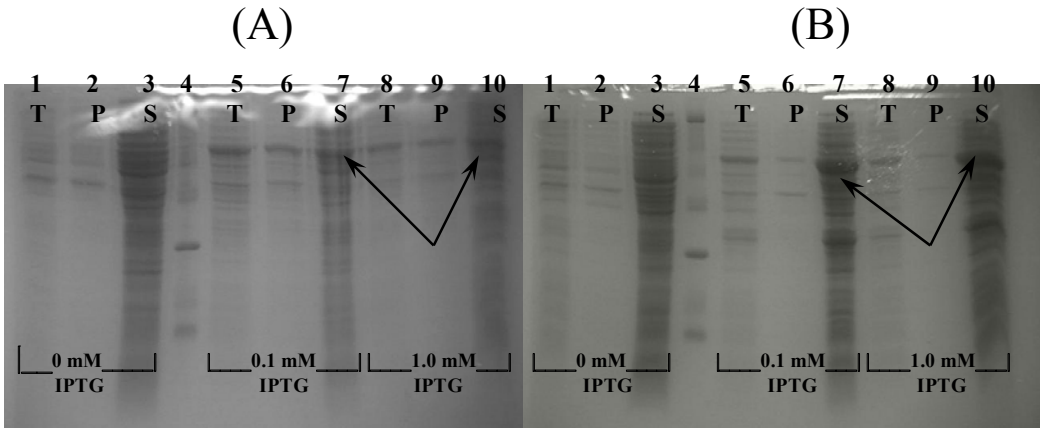


Figure 23: Expression study of GST-CLIC1-E81M

15 % polyacrylamide SDS-PAGE gel showing the solubility of heterologously expressed GST-CLIC1-E81M at **(A)** 37 °C and **(B)** 20 °C. Lanes 1-3 depict the total lysate (T), insoluble/pellet fraction (P), and soluble/supernatant fraction (S) when the bacterial cells were grown at 37 °C and 20 °C respectively. Lane 4 shows the molecular weight marker (116, 66.2, 45, 35, 25, 18.4 and 14.4 kDa). Lanes 5-7 illustrate total lysate (T), insoluble/pellet fraction (P), and soluble/supernatant fraction (S) when the bacterial cells were grown and induced with 0.1 mM IPTG at 37 °C and 20 °C respectively. Lanes 8 – 10 represent the total lysate (T), insoluble/pellet fraction (P), and soluble/cytosolic fraction (S) when the bacterial cells were grown and induced with 1.0 mM IPTG at 37 °C and 20 °C, respectively. The arrows indicate the over-expressed, soluble GST-CLIC1-E81M fusion protein.

3.4 CLIC1-M32A and CLIC-E81M purification

CLIC1-M32A was purified using the protocol of Tulk *et al.,* (2002). The protein was over-expressed at 20 °C, as described in **3.3.1** and **2.2.4**, and purified via GSH-agarose affinity chromatography followed by DEAE-anion exchange chromatography. The homogeneity and purity of CLIC1-M32A was assessed on 15 % acrylamide SDS-PAGE gels. Figure 24A lanes 1 and 2 show the insoluble/pellet and soluble/supernatant fractions. The fusion protein was bound to the GSH-agarose column allowing bacterial proteins to be washed away. This is shown by the lack of the 49 kDa band in the flow-through collected from the affinity column (Figure 24A lanes 4 and 5). Thrombin digestion resulted in the cleavage of CLIC1-M32A from the GSH-agarose bound GST moiety. Lane 7 of Figure 24A contains eluted material after thrombin cleavage. CLIC1-M32A was successfully cleaved from the GSH-agarose column as illustrated by the single band of 29 kDa corresponding in size to CLIC1-M32A. The ProtParam tool (Gasteiger *et al.,* 2005) calculated the molecular weight of CLIC1-M32A to be 28 kDa which is consistent with the eluted band in Figure 24A lane 7. CLIC1-M32A was separated from the thrombin using DEAE anion-exchange chromatography. The different pI values of the two proteins (thrombin pI = 8, CLIC1-M32A pI = 5.1) allows for their separation. The DEAE-column was equilibrated with a buffer of pH 6.5 resulting in positively charged thrombin and negatively charged CLIC1-M32A. Figure 24D shows the elution of thrombin and CLIC1-M32A from the DEAE column. The positively charged thrombin can not bind the DEAE-column and was eluted first. The negatively charged CLIC1-M32A bound the column and was eluted using a high-salt concentration buffer in a single homogenous peak. CLIC1-M32A, eluted from the DEAE-column, was collected in 2 ml fractions that were analyzed on a 15 % acrylamide SDS-PAGE gel (Figure 24B).

There is a slight band (approximate size of 26 kDa) just below the bands corresponding in size to CLIC1-M32A (Figure 24B). The faint band was found to comprise less than 10 % of the total protein in the sample. Hence, CLIC1-M32A was of sufficient purity in terms of the experimental procedures described in the remainder of this chapter.

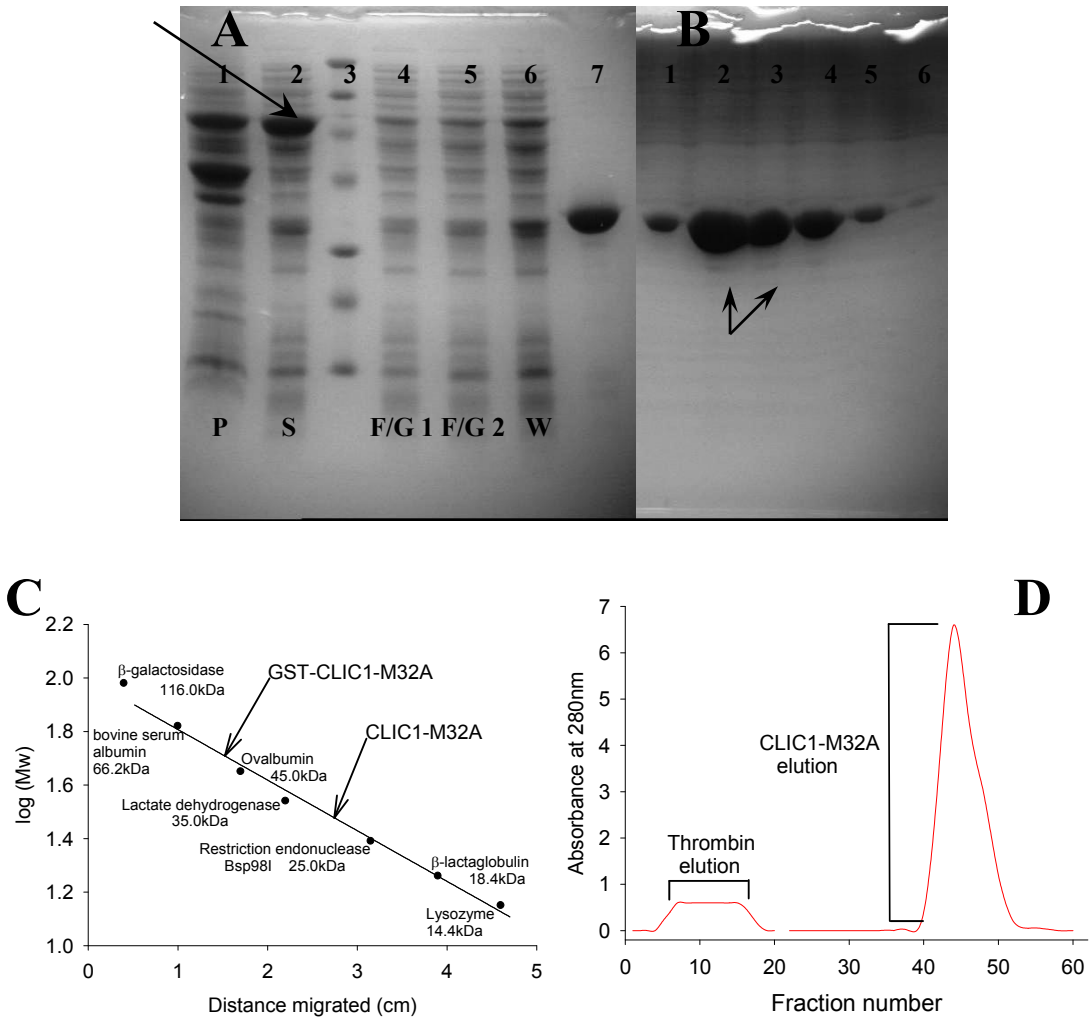


Figure 24: CLIC1-M32A purification

Purification of CLIC-M32A using GSH-agarose followed by DEAE anion chromatography. 15 % polyacrylamide SDS-PAGE gels depicting **(A)** Lane 1: insoluble/pellet fraction (P), and soluble/cytosolic fraction (S) of GST-CLIC1-M32A (arrow indicates the overexpressed, at 20 °C, soluble fusion protein). Lane 3: shows the molecular weight marker (sizes, in kDa, 116, 66.2, 45, 35, 25, 18.4, and 14.4). Lanes 5-7: indicate the flow-through (F/G1 and F/G2) and wash (W) collected from the GSH-agarose affinity column. **(B)** Fractions 42 – 52 of CLIC1-M32A collected from the DEAE-anion exchange column. The arrows point out a possible protein contaminant. **(C)** Calibration curve constructed using molecular weight standards indicating the positions of GST-CLIC1-M32A and CLIC1-M32A. **(D)** CLIC1-M32A elution profile off DEAE-Sepharose column. CLIC1-M32A was eluted using 300 mM NaCl elution buffer pH 6.5.

Over-expressed GST-CLIC-E81M was purified using the above-described procedures used for the purification of CLIC-M32A. Figure 25 depicts a 15 % acrylamide SDS-PAGE gel with fractions of CLIC-E81M eluted from the DEAE-column. The single bands, matching in size with CLIC1-E81M, point to pure homogeneous protein.

Figure 26 shows absorbance spectra of purified CLIC1-M32A and CLIC1-E81M. The low absorbance at 340 nm indicates lack of protein aggregates. The ratio of A280:A260 suggests that the proteins were free of nucleic acids. The yields, per 1 l culture, of CLIC1-M32A and CLIC1-E81M were calculated to be in the region of 0.9 – 3 mg/ml.

3.5 Verification of the M32A mutation using Electrospray-Ionization Mass Spectrometry (ESI-MS)

ESI-MS was used to resolve the exact *average masses*[*] of purified wtCLIC1 and CLIC-M32A. In addition, the difference between the masses of wtCLIC and CLIC-M32A, determined via ESI-MS, was used to validate the Methionine to Alanine mutation at position 32 (Met32Ala) of CLIC-M32A. Figure 27 shows the multiple charge spectrum of CLIC1-M32A where the different peaks represent the various charge states of the protein. Using the deconvolution software, MagTran, the *average masses*[*] of CLIC1-M32A and wtCLIC1 were determined to be 27,005.8 amu and 27,066.6 amu respectively (Figure 27, insets A and B). These values agree with the ProtParam sequence-derived CLIC1-M32A and wtCLIC1 molecular weights of 27,006.7 and 27,066.8 amu respectively (Gasteiger *et al.,* 2005). The mass difference of 60.8 amu between wild-type and mutant proteins, corroborates the presence of the M32A mutation since the *average mass*[*] difference between Methionine (131.20 amu) and Alanine (71.08 amu) is 60.12 amu.

A series of multiply charged peaks, other than those belonging to CLIC-M32A, suggest the presence of a protein contaminant (see Figure 27). Using MagTran, the contaminant *average mass*[*] was calculated to be 26.877.7 amu. Dividing the relative abundance of contaminant charge peaks by the relative abundance of CLIC1-M32A

[*]: *The average mass was calculated from the entire isotopic distribution of each charged peak belonging to wtCLIC1 and CLIC-M32A. Average mass values are closely related to molecular weight, sum of the atomic weights of all atoms in a molecule (*Kaltashov and Eyles, 2005*).*

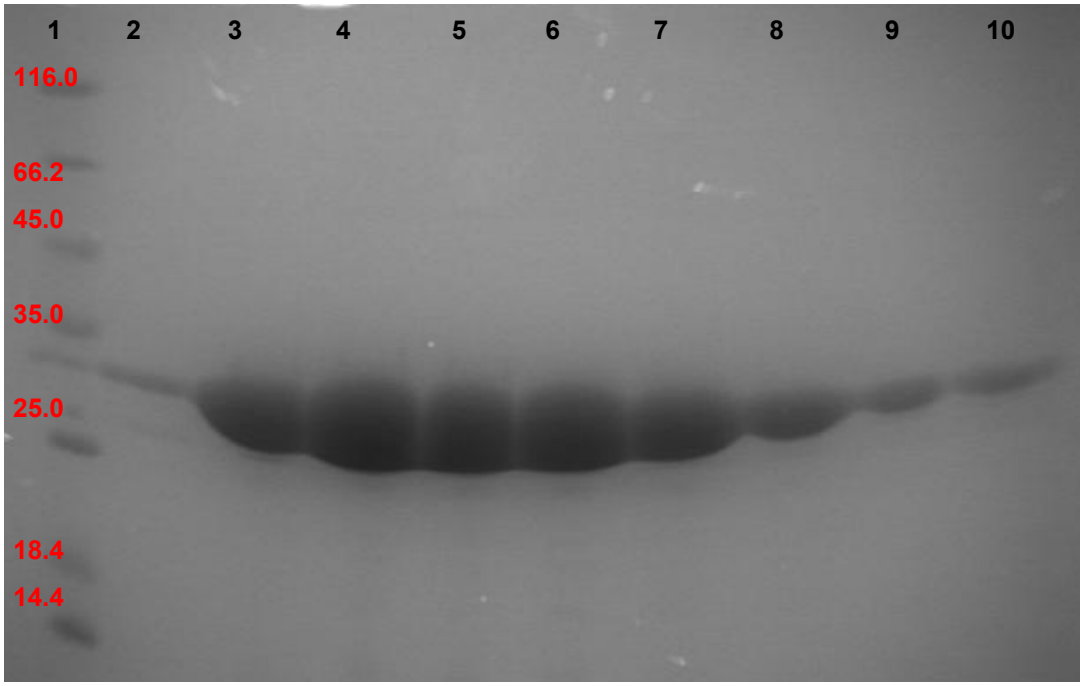


Figure 25: SDS-PAGE separation of CLIC1-E81M

15 % acrylamide SDS-PAGE gels depicting fractions of CLIC1-E81M collected from DEAE anion-exchange column (lanes 2 – 10). Lane 1 shows the molecular weight marker (sizes, in kDa, are shown in red).

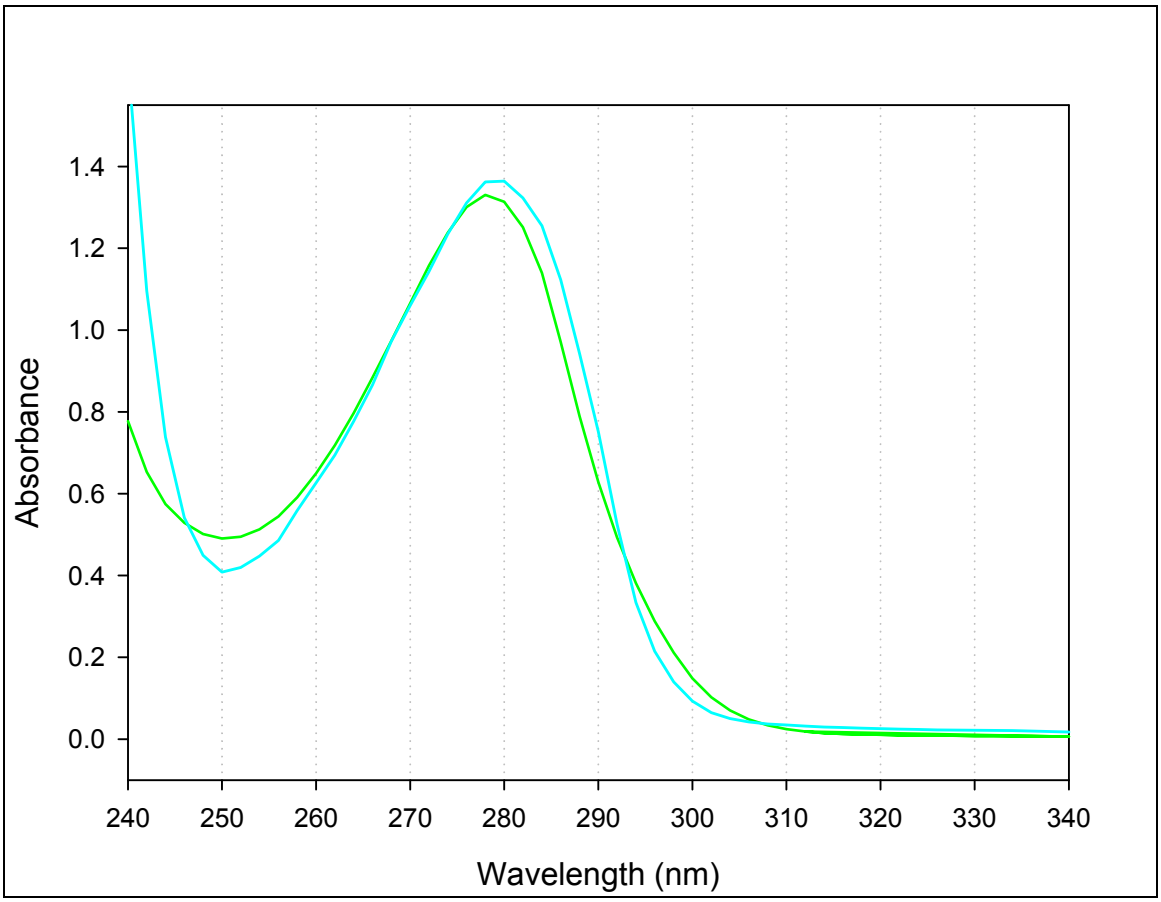


Figure 26: Absorbance spectra of purified CLIC1-M32A and CLIC1-E81M

Purified CLIC1-M32A (green) and CLIC1-E81M (cyan). Both proteins are free of DNA contamination (Abs 280 nm/Abs 260 nm = 2 and 2.6 respectively). The low absorbance at 340 nm (0.006 and 0.02, respectively) points to absence of aggregates. The spectra were recorded at 20 °C using a Jasco 550 UV/VIS Spectrophotometer. The data was plotted using SigmaPlot v9.0

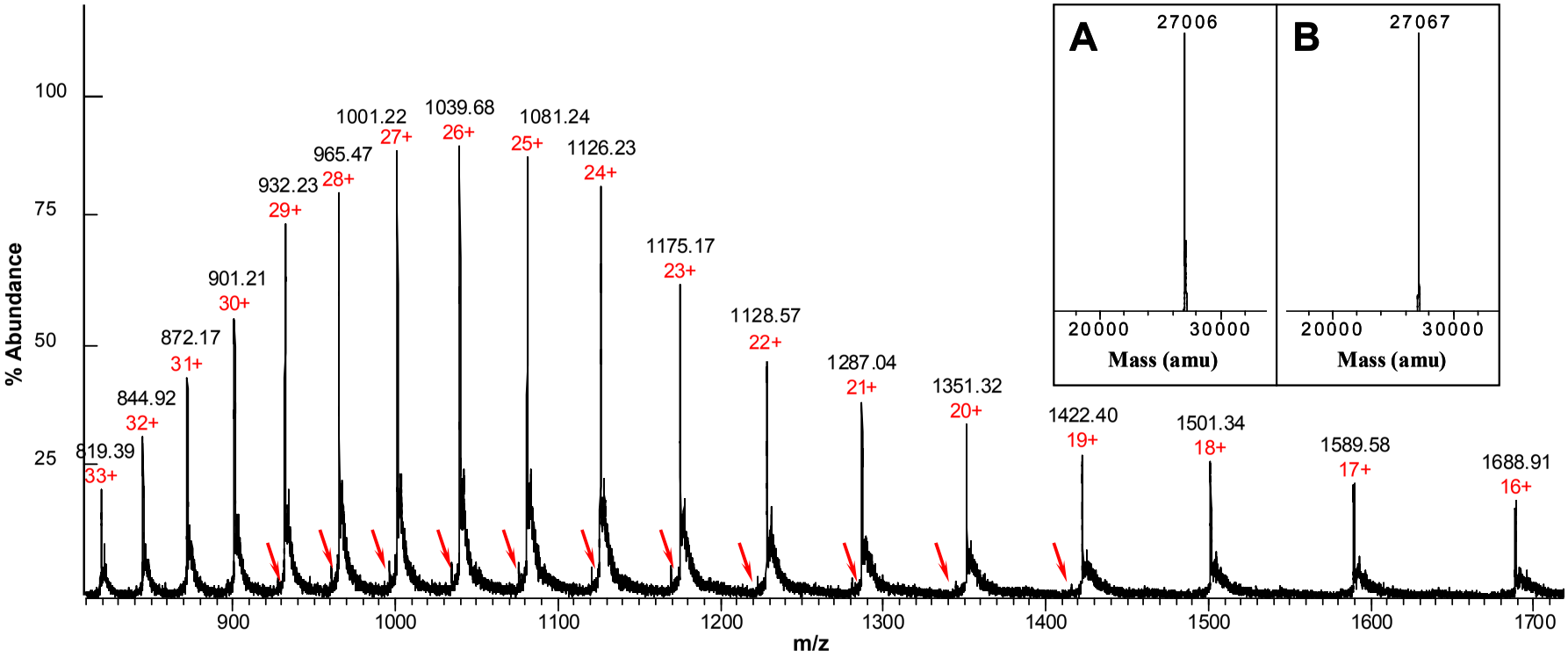


Figure 27: ESI-MS of CLIC1-M32A

Electospray ionization mass spectrum (ESI-MS) of intact CLIC1-M32A in 50 mM Na_2HPO_4, 1mM DTT, 0.02% NaN_3 buffer pH 7.0. Above each peak the different charge states of CLIC-M32A are indicated in red while the m/z ratios are in black. The data was collected on a Q-TOF-MS, in positive ion mode, coupled to an ULITIMA UPLC system. The inset shows the molecular weight of **(A)** CLIC1-M32A and **(B)** wtCLIC1 calculated using the deconvolution software MagTran (Zhongqi and Marshall, 1998). The arrows point to the multiple charge envelope of a possible protein contaminant whose molecular weight was calculated to be 26,878 amu.

charge peaks, it was calculated that the contaminant comprised approximately 8 % of the total protein in the sample. This agrees with the SDS-PAGE analysis of CLIC1-M32A purity (see section **3.4** and Figure 24B).

3.6 Effect of M32A and E81M mutations on the structural dynamics of native CLIC1

Continuous labelling DXMS is a highly efficient technique used in studying local and global changes in the native conformation of proteins induced by, amongst others, structural modification. The comparison of the data obtained form such experiments between wild-type and mutant protein can reveal structural rearrangement induced by the mutation. These changes will manifest through altered hydrogen for deuterium exchange along the peptide backbone. Hence, a region that becomes unfolded in the mutant protein will exchange faster than the corresponding fragment in the wild-type species, and vica versa. The aim of the continuous labelling DXMS analysis of CLIC-M32A and CLIC-E81M was two fold. First, this technique was used to establish whether the engineered mutations brought about any conformational changes of the native state. This was achieved by comparing the mutant data to wtCLIC1 data, previously obtained by Nathaniel (2006) at identical conditions (pH 7.0; 4 °C). Second, continuous labelling DXMS was used to determine whether any structural changes induced by the Met32Ala and Glu81Met mutations mimicked the effects of pH on the native conformation of wtCLIC1. Hence, data obtained for CLIC1-M32A and CLIC1-E81M at pH 7.0 was compared to wtCLIC1 at pH 5.5 data (Nathaniel, 2006).

Two important aspects of DXMS need to be mentioned before further data description is undertaken. The first refers to the resolution of the hydrogen/deuterium exchange technique in terms of determining the exact location of deuterium incorporation. In theory, single amino acid resolution can be achieved by progressive proteolysis where pepsin digestion is optimized so that multiple overlapping fragments are generated The production of sub-fragments will eventually narrow down the position of an exchanged deuterium to a precise amide (Woods and Hamuro, 2001). Although, complete sequence coverage was achieved in both CLIC-M32A and CLIC1-E81M, sufficient fragment overlap needed for single-amide resolution could not be attained in most parts of the mutant sequences. Thus, the deuterium-localization peptide maps

of CLIC-M32A and CLIC1-E81M were used to identify regions of deuterium incorporation rather than specific exchange sites. The second aspect that needs to be considered when contrasting DXMS data from two proteins relates to calculation of the average number of deuteriums per peptide (see section **2.2.12.1.4**). Smith and Zhang (1993) showed that adjustment made for back-exchange during digestion and HPLC analysis introduced an average error in deuterium incorporation of 5.5 % ± 5.5. In the case of CLIC1 on-line proteolysis was used reducing the reaction time and possibly the back-exchange error. However, only regions that differed by 10 % or more, between wild-type and mutant, were considered as significantly different.

3.6.1 Peptide evaluation

Once a peptide pool was generated by correlating the tandem raw mass spectra with the sequence of CLIC1-M32A or CLIC1-E81M the isotopic profiles of the peptides were validated. Figure 28 shows the non-deuterarted theoretical and experimental isotopic envelopes for peptide/fragment GSMAEEQPQVEL (residues 1 – 12, molecular formula: $C_{54}H_{88}N_{14}O_{22}$) of CLIC1-M32A. The authenticity of the singly charged species of peptide GSMAEEQPQVEL was validated by the good fit between its theoretical and experimental isotopic envelopes (Figure 28A). On the other hand, the experimental scan of the triply charged species of fragment GSMAEEQPQVEL did not agree with its theoretic profile (Figure 28B). Consequently, this peptide was deleted from the peptide pool and hence from taking part in further analysis. The validation process was repeated for the fully-deuterated as well as the various time scans (10, 30, 300 and 1000 s in the case of CLIC1-M32A; 10, 30, 100, 300, 1000 and 3000 s for CLIC1-E81M) ensuring that the correct peptide was used in all runs.

The next step in peptide evaluation involved generating consensus peptide-maps for each time point. The authenticated fragments were plotted against the amino-acid sequence of CLIC1-M32A or CLIC1-E81M. This map had a two-fold function. Firstly, it illustrated the amount of sequence coverage provided by the peptide pool. Secondly, the consensus peptide maps were used to identify regions of deuterium incorporation along the proteins' sequence. As a result, overlapping fragments that differed by more than two deuteriums were identified as outliers and deleted from further analysis. Figure 29 depicts the consensus peptide-map of CLIC1-M32A incubated for 10 s with deuterated buffer. The sequence numbering spans 243 amino

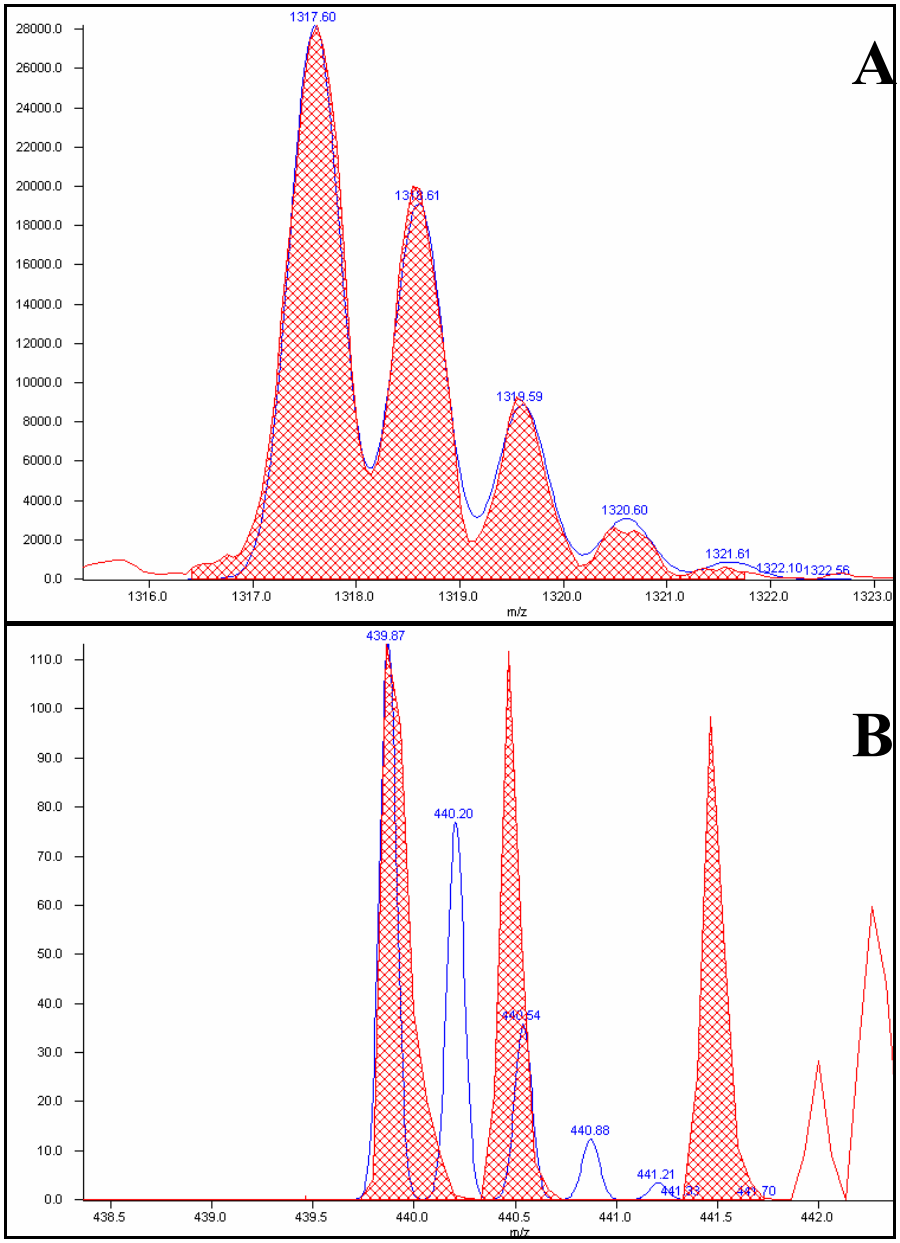


Figure 28: Peptide quality check

Theoretical (blue outline) and experimental (red fill) isotopic envelopes of **(A)** singly and **(B)** triply charged species for peptide GSMAEEQPQVEL (residues 1 – 12, molecular formula $C_{54}H_{88}N_{14}O_{22}S$) of CLIC1-M32A. The triply charged species of fragment 1 - 12 shown in **(B)** does not fit the theoretical profile and is deleted from the peptide pool. The two species of the peptide shown above were generated from a non-deuterarted control run. The quality check is repeated for the fully-deuterated as well as varying time point runs (10, 30, 300 and 1000 s in the case of CLIC1-M32A; 10, 30, 100, 300, 1000 and 3000 s for CLIC1-E81M). In addition to comparing theoretical and experimental isotopic envelopes, a number of other criteria such as retention time and centroid value were to validate the fragments. Peptide authentication was done using DXMS Explorer software (Sierra Analytics, LLC. Modesto, CA).

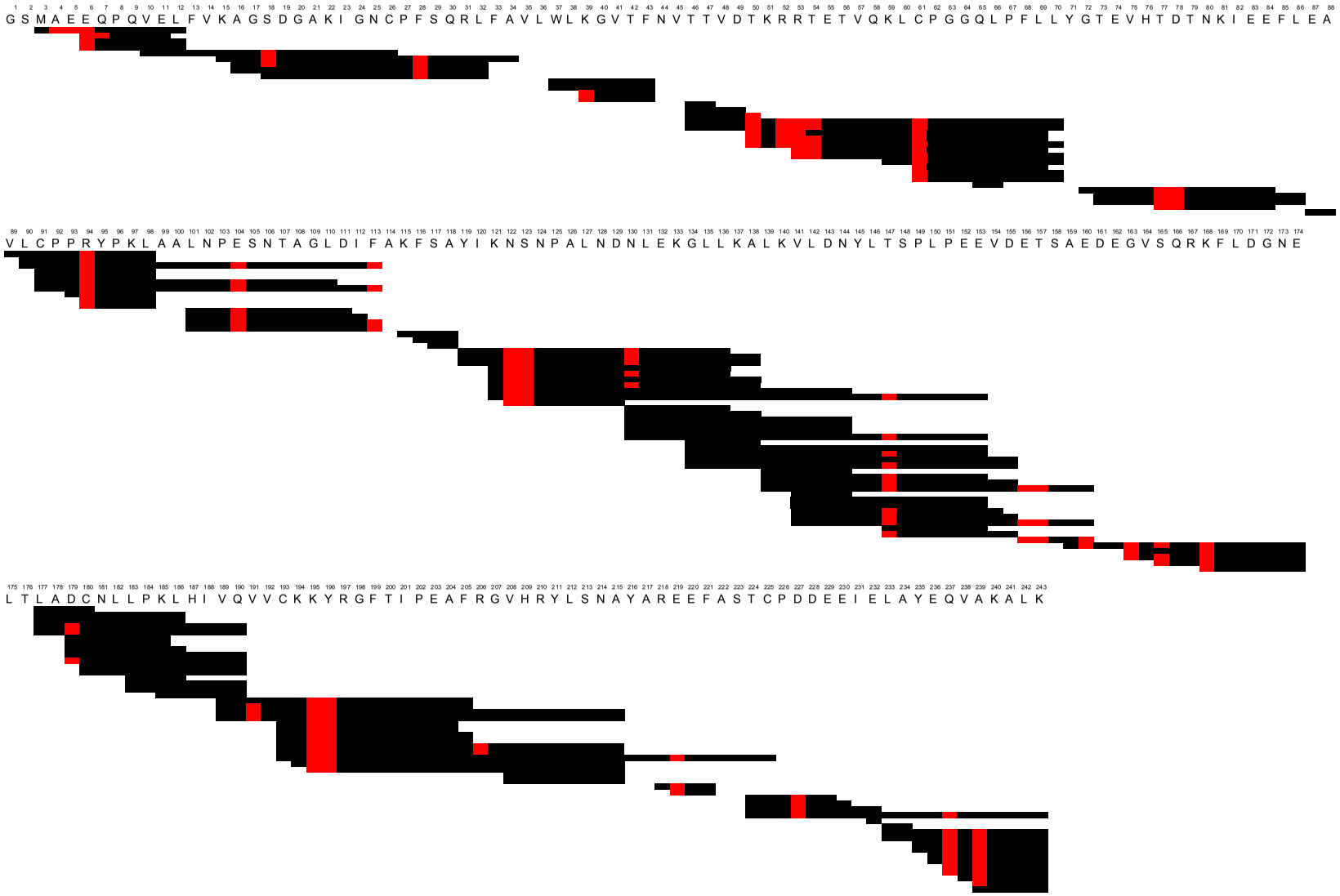


Figure 29: CLIC1-M32A consensus peptide map - rapid exchange

Consensus map of CLIC1-M32A incubated for 10 s with deuterated buffer, The peptides (black), quality checked with DXMS Explorer (Sierra Analytics, LLC. Modesto, CA), are mapped against the amino acid sequence of CLIC1-M32A. The possible locations of deuterium are shown in red. Overlapping peptides that lack consensus in terms of deuterium incorporation are deleted from the peptide pool and further analysis. The first two amino acids of each peptide are not shown on the map. Deuterium incorporation that takes place at those positions is back-exchanged duirng the chromatographic separation of the peptides. The sequence and numbering of CLIC1-M32A includes two additional residues (Glycine and Serine) at the N-terminus of the protein. These amino acids form part of the thrombin cleavage site of the GST-CLIC1 fusion construct.

acids (ORF of wtCLC1, CLIC1-M32A and CLIC1-E81M encodes 241 amino acids). The two extra residues are accounted for by the N-terminal GlySer sequence that form part of the thrombin cleavage site of GST-CLIC1 fusion proteins (Harrop *et al.,* 2001). The consensus peptide-map of CLIC1-M32A is made of 139 overlapping peptides that provided full sequence coverage (Figure 29). In addition, all overlapping fragments demonstrated consensus in terms of number of incorporated deuteriums, hence corroborating their quality. The peptide pool obtained for CLIC1-E81M also provided full sequence coverage with 163 positively identified peptides. The consensus peptide-maps of CLIC1-M32A incubated with deuterated buffer for 30, 300 and 1000 s as well as CLIC1-E81M incubated for 10, 30, 100, 300, 1000 and 3000 s are shown in Appendix **Tables F1 – F3** and **G1 – G6**, respectively.

3.6.2 Structural changes induced by the M32A mutation

Regions that exhibit greater than 45 – 50 % deuterium exchange within 10 s, also referred to as rapid exchange, are highly solvent exposed and not involved in strong hydrogen bonding. Hence, such fragments are deemed as unstructured or highly flexible. Figure 30A depicts the rapid hydrogen exchange quality-checked peptides of CLIC1-M32A plotted as a function of the proteins amino-acid sequence. Two fragments display deuterium localization above 45 % (Figure 30A). These fast-exchanging regions span residues Met^{1} – Leu^{10}§ and Val^{152} – Glu^{158}§. In the case of Met^{1} – Leu^{10}, the corresponding peptide in wtCLIC1 at pH 7 also showed fast deuterium exchange at 54 % after 10 s deuterium incubation (Nathaniel, 2006). The other unstructured fragment Val^{152} – Glu^{158} forms part of the negatively charged loop (Pro^{147} – Gln^{164}) found between helices h5 and h6 of CLIC1 (Figure 30B). DXMS studies using wtCLIC1 at pH 7.0 indicated that peptide Val^{152} – Glu^{172}, part of the negatively charged loop, was the fastest exchanging regions in the protein at 55 % after 10 s deuterium incubation (Nathaniel, 2006).

Figure 30B reveals that the majority of slow exchanging regions, after 10 s deuterium incubation, (≥ 10 % exchange) are found in the C-terminal domain of CLIC1-M32A. This trend was also seen in longer-exchanging time points such as 1000 s (longest

§: *The amino acid numbering follows that of the crystal structure of CLIC1 (*Harrop *et al.,* 2001*) and does not include the N-terminal Gly-Ser that form part of the thrombin cleavage site. This numbering format is followed in the remainder of the text.*

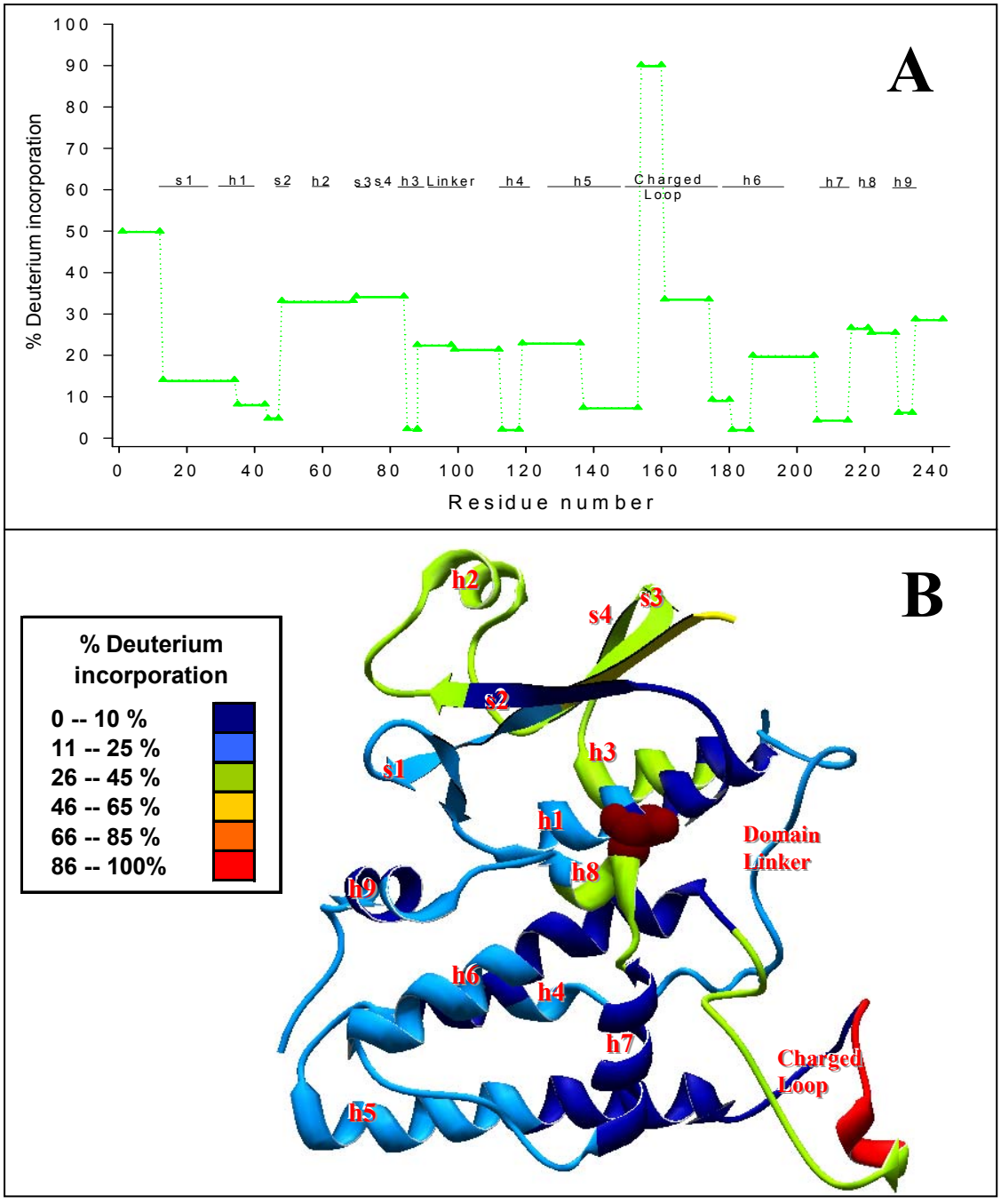


Figure 30: Rapid deuterium exchanging regions in CLIC1-M32A

(A) Peptides of CLIC1-M32A at pH 7.0 incubated in deuterated buffer for 10 s plotted as function of percentage deuterium incorporation. Fragments with greater than 45 – 50 % deuterium localization are rapidly-exchanging and deemed as floppy/unstructured regions.

(B) Percentage deuterium incorporation from **(A)** is mapped on the crystal structure of CLIC1 (pdb code: 1k0m). The engineered mutation Met32Ala is shown in brown. The secondary structural elements, marked as per **(A)**, are depicted in red. The inset illustrates the colour-coding scheme. The diagram was generated using SwissPdb viewer (Guex and Peitsch, 1997).

incubation time in the case of CLIC1-M32A). Figure 31 depicts peptides of CLIC1-M32A at pH 7 that experience minimal deuterium exchange after 1000 s. In addition, these fragments were matched with the corresponding ones from wtCLIC1 at pH 7.0. In some cases, identical peptides could not be found between CLIC1-M32A and wtCLIC1 and so the closest fragment in terms of sequence was used. Phe^{83} – Ala^{86} (h3, N-domain), Ala^{112} – Ala^{116} (h4, C-domain), Asn^{126} – Asn^{142} (h5, C-domain) and Leu^{173} – Gln^{188} (h6, C-domain), belonging to CLIC1-M32A, are the most solvent-protected regions in the mutant protein (~ 10 % exchange after 1000 s). All four fragments consist of residues that have minute backbone accessible surface areas [0 – 8 %, calculated by GETEREA 1.1 (Fraczkiewicz and Braun, 1998)] and low amide B-factor values. In addition, Glu85 of Phe^{83} – Ala^{86} forms part of an extensive hydrogen bonded network that is conserved in the CLIC family (see section **3.1.1** and Figure 11). In comparison to CLIC1-M32A at pH 7.0, the corresponding fragments from wtCLIC1 at pH 7.0 also show minimal deuterium exchange after 1000 s (Figure 31A).

Significant differences, induced by the Met32Ala mutation, will be enhanced at longer deuterium incubation times. Hence, the 1000 s time point (longest incubation time in the case of CLIC1-M32A) was used to compare the deuterium localization levels in CLIC1-M32A at pH 7.0, wtCLIC1 at pH 7.0 and wtCLIC1 at pH 5.5. It has been shown that a 1 unit decrease in pH results in 10 fold reduction in deuterium exchange rate (Bai *et al*., 1993). In order to directly compare CLIC1-M32A at pH 7.0 and wtCLIC1 pH 5.5, a 30 fold adjustment factor ($10^{1.5\ pH\ units}$) was taken into consideration. Therefore, CLIC1-M32A at pH 7.0 deuterated for 10 s was compared with wtCLIC1 at pH 5.5 deuterated for 300 s. Similarly, 1000 s deuterium exchange at pH 7.0 was contrasted with 30,000 s deuterium exchange at pH 5.5, and so on. Figure 32A compares the amide deuterium exchange patterns of CLIC1-M32A at pH 7.0, wtCLIC1 at pH 7.0 and wtCLIC1 at pH 5.5. Only one region of CLIC1-M32A exhibited a significant difference (> 10 % and/or > 1 deuterium) in deuterium localization when compared to wtCLIC1 at pH 7.0. It must be noted that the difference in deuterium incorporation was not noted during fast-exchange i.e. 10 s (Figure 32B inset). Peptide Ala^{220} – Glu^{228} showed a ~ 20 % increase in deuterium incorporation after 1000 s incubation with deuterated buffer (Figure 32A). This fragment, which forms part of h8 and h9 of CLIC1, is a distance away from the engineered Met32Ala mutation in terms of the protein's primary structure. However,

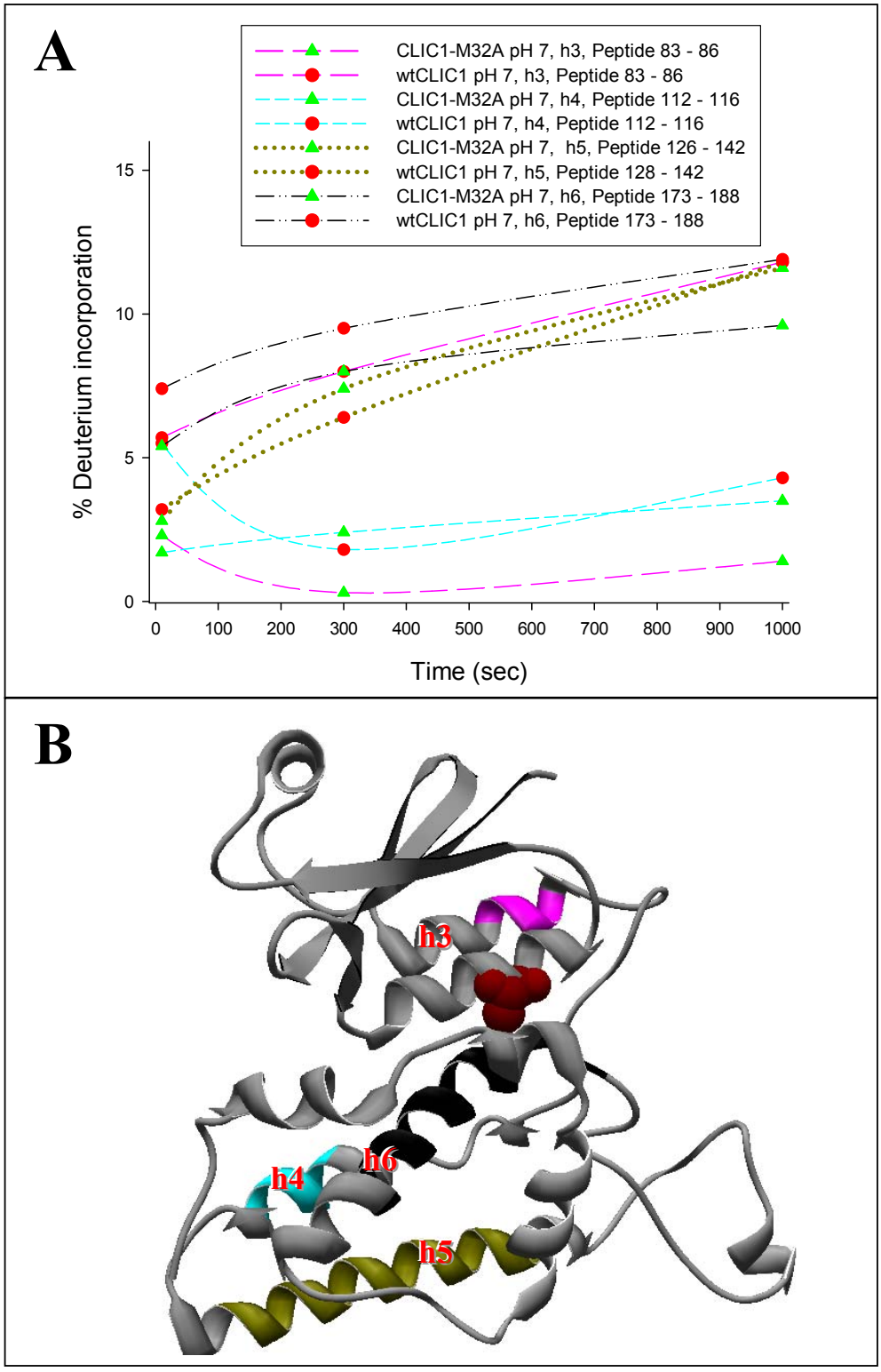


Figure 31: Slow-exchanging regions in CLIC1-M32A

(A) Peptides belonging to CLIC1-M32A (▲) and wtCLIC1 (●) at pH 7.0 plotted as function of percentage deuterium incorporation. Even after 1000 s the depicted fragments show slow-exchange with no more than ~ 10 % deuterium incorporation.

(B) Slow-exchanging peptides, colour-coded as per **(A)**, mapped on the structure of CLIC1 (pdb code: 1k0m). The engineered mutation, Met32Ala, is show in brown. The diagram was generated using SwissPdb viewer (Guex and Peitsch, 1997).

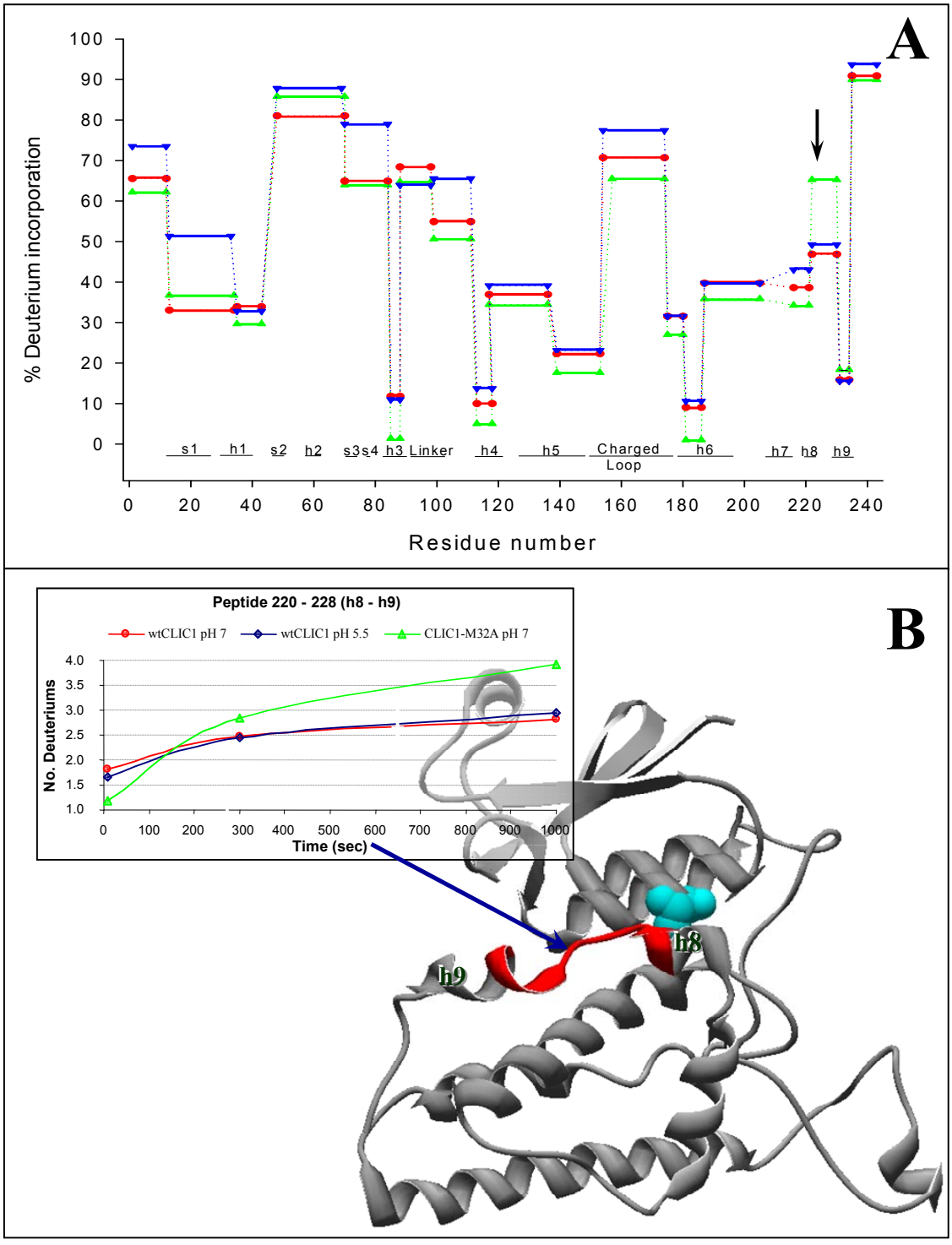


Figure 32: Effect of M32A on the structural dynamics of CLIC1

(A) Bar chart depicting deuterium exchange of CLIC1-M32A at pH 7.0 (green; 1000 s deuterium incubation), wtCLIC1 pH 7.0 (red; 1000 s deuterium incubation) and wt CLIC1 at pH 5.5 (blue; 30,000 s deuterium incubation). The arrow points out a region (peptide Ala^{220} – Glu^{228}) that exhibits ~ 20 % higher deuterium localization in CLIC1-M32A than wtCLIC1.

(B) The position of the fragment Ala^{220} – Glu^{228} (red) is shown in relation to the Met32Ala mutation (cyan). The inset illustrates the number of exchanged deuteriums as a function of time for the peptide Ala^{220} – Glu^{228}. In the case of wtCLIC1 pH 5.5, the plot has been adjusted for the ~ 30 fold decrease in exchange due to pH difference. The structure shown is that of CLIC1 (pdb code: 1k0m).

in its tertiary structure, Ala^{220} – Glu^{228} is in close proximity to position 32 (Figure 32B). Thr222 is within 4 Å of Met32 and form hydrophobic contacts with this residue in wtCLIC1.

The effects of pH on the dynamics of wtCLIC1 were also subtle (Nathaniel, 2006). Peptides Phe^{11} – Met^{32} (s1 – h1) and Leu^{68} - Glu^{82} (s3 – h4) were shown to be more flexible at pH 5.5 with a 20 % increase in deuterium exchange (Figure 32A). In those two regions, CLIC1-M32A exhibited identical deuterium exchange as wtCLIC1 at pH 7.0 (Figure 32A).

3.6.3 Structural changes induced by the E81M mutation

Figure 33A depicts the rapid hydrogen exchange quality-checked peptides of CLIC1-E81M (10 s deuterium incubation) plotted as function of the proteins amino-acid sequence. Three fragments were found to have more than 45 % of their amide hydrogens replaced with deuterium after 10 s exchange. Two of them, Met^{1} – Leu^{10} and Val^{152} – Glu^{158}, were also identified as fast-exchanging in the case of CLIC1-M32A (see section **3.6.2** and Figure 30). The third peptide, Val^{87} – Leu^{96} (part of the domain linker loop), is in close proximity to the engineered E81M mutation (Figure 33B). Significantly, this fragment exchanged ~ 20 % and ~ 15 % more deuterium at 10 s than the corresponding region in wtCLIC1 at pH 7.0 and pH 5.5, respectively.

Figure 34 depicts peptides, obtained from CLIC1-E81M at pH 7.0, which experienced minimal deuterium exchange after 3000 s (≥ 15 % deuterium incorporation). In addition, these fragments were matched with the corresponding ones from wtCLIC1 at pH 7.0. Phe^{31} – Leu^{34} (h2, N-domain), Phe^{83} – Ala^{86} (h3, N-domain), Ala^{112} – Ala^{116} (h4, C-domain), Asn^{126} – Asn^{142} (h5, C-domain) and Leu^{173} – Gln^{188} (h6, C-domain) were the slowest exchanging peptides in both CLIC1-E81M and wtCLIC1 at pH 7.0. As was the case with CLIC1-M32A (section **3.6.2**, Figure 31) the bulk of these solvent-protected regions were found in the C-terminal domain (Figure 34B). Peptide Phe^{203} – Ala^{213}, belonging to CLIC1-E81M, showed less than 10 % deuterium localization after 3000 s exchange (Figure 34A). However, under matching conditions (pH 7.0 and 3000 s exchange) the equivalent wtCLIC1 peptide exhibited more than double deuterium incorporation (~ 20 % of amide hydrogens were exchanged after 3000 s; Figure 34A). A second significantly different region found between CLIC1-

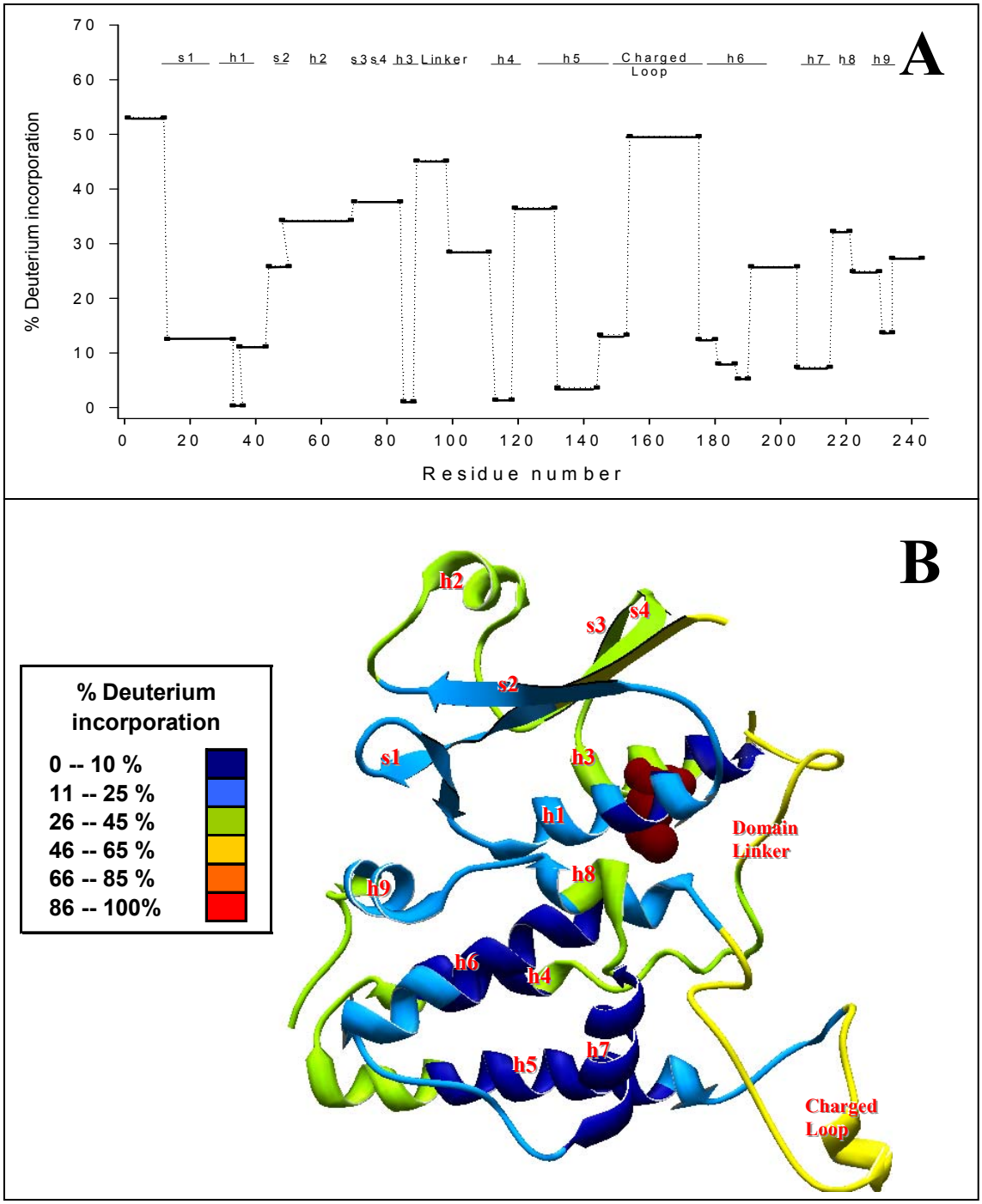


Figure 33: Rapid deuterium exchanging regions in CLIC1-E81M

(A) Peptides of CLIC1-E81M at pH 7.0 incubated in deuterated buffer for 10 s plotted as function of percentage deuterium incorporation. Fragments with greater than 45 – 50 % deuterium localization were identified as rapidly-exchanging and deemed as floppy/unstructured regions.

(B) Percentage deuterium incorporation from **(A)** is mapped on the crystal structure of CLIC1 (pdb code: 1k0m). The engineered mutation Glu81Met is shown in brown. The secondary structural elements, marked as per **(A)**, are depicted in red. The inset illustrates the colour-coding scheme. The diagram was generated using SwissPdb viewer (Guex and Peitsch, 1997)

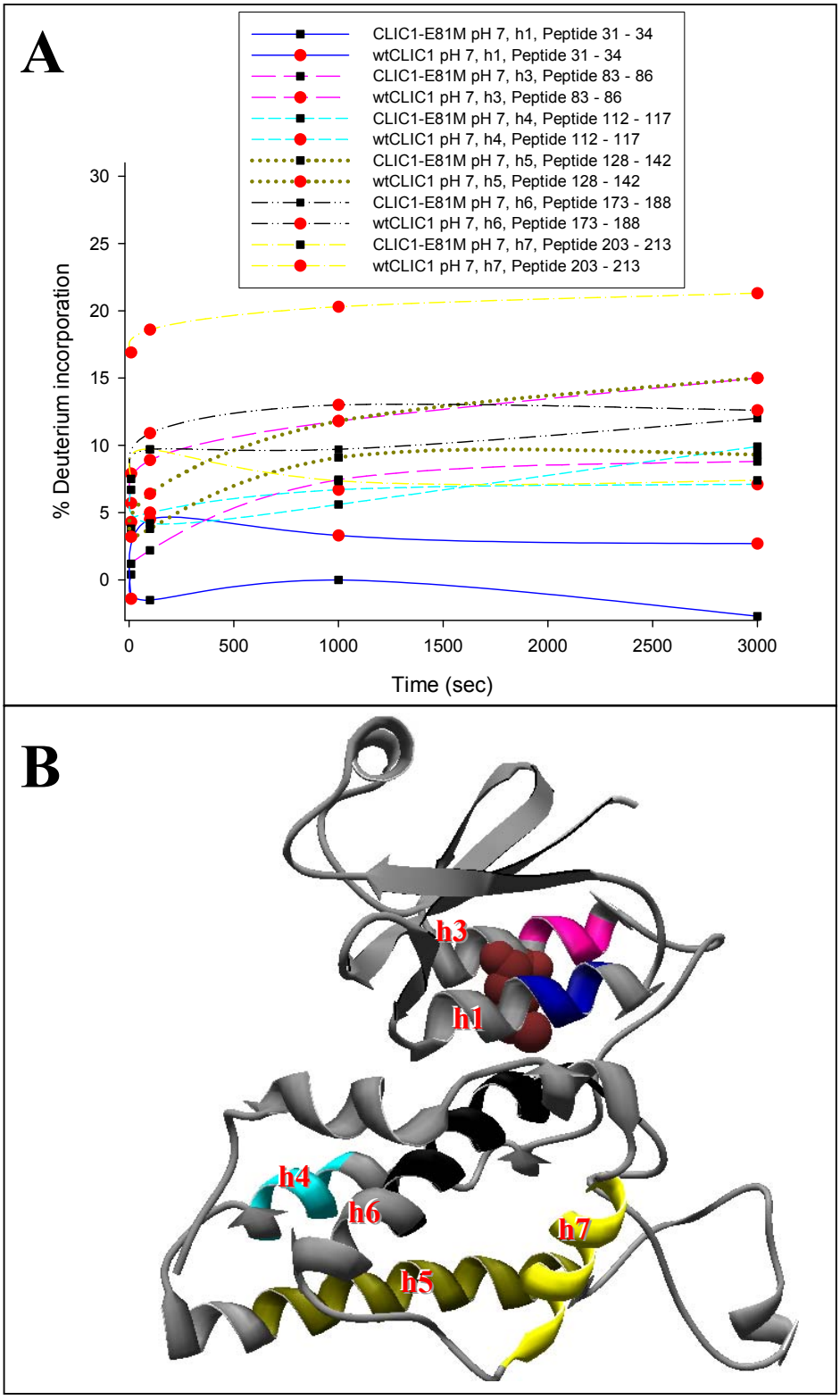


Figure 34: Slow-exchanging regions in CLIC1-E81M

(A) Peptides belonging to CLIC1-E81M (■) and wtCLIC1 (●) at pH 7.0 plotted as function of percentage deuterium incorporation. Even after 3000 s depicted fragments, belonging to CLIC1-E81M, show slow-exchange with no more than ~ 15 % deuterium incorporation.

(B) Slow-exchanging peptides, colour-coded as per **(A)**, mapped on the structure of CLIC1 (pdb code: 1k0m). The engineered mutation, Glu81Met, is show in brown. The diagram was generated using SwissPdb viewer (Guex and Peitsch, 1997).

E81M and wtCLIC1 at pH 7.0 was Tyr^{69} – Phe^{83}. This fragment incorporates the engineered E81M mutation (Figure 35B). After 3000 s exchange, Tyr^{69} – Phe^{83} exhibited ~ 20 % higher deuterium localization in CLIC1-E81M as compared to wtCLIC1 at pH 7.0 (Figure 35A). Noticeably, Tyr^{69} – Phe^{83} from CLIC1-E81M at pH 7.0 mimicked the equivalent fragment in wtCLIC1 at pH 5.5 in terms of hydrogen-deuterium exchange (Figure 35A). As was the case with CLIC1-M32A, 3000 s exchange in CLIC1-E81M at pH 7.0 was compared with 90,000 s exchange in wtCLIC1 at pH 5.5 in order to correct for the decrease in deuterium incorporation at lower pH (see section **3.6.2**).

3.7 Characterization of CLIC1-M32A

3.7.1 Secondary and tertiary structure analyses

Changes, induced by the Met32Ala mutation, in the secondary structure of CLIC1-M32A were detected using far-UV CD spectroscopy. In the far-UV range, 190 – 250 nm, this technique is sensitive to the secondary structural conformations of proteins (Woody, 1995). Figure 36 compares the far-UV CD spectra of native CLIC1-M32A at pH 7.0, wtCLIC1 at pH 7.0 and wtCLIC1 at pH 5.5. The spectra of CLIC1-M32A display two minima at 208 and 222 nm characteristic of a predominantly alpha helical protein. The far-UV CD spectra of CLIC1-M32A and wtCLIC1 at pH 7.0 are non-imposable. The mutant protein exhibited ~ 16 % decrease in signal compared to wtCLIC1 at pH 7.0. On the other hand, the spectra of CLIC1-M32A at pH 7.0 as well as pH 5.5 and wtCLIC1 at pH 5.5 overlap indicating that, under these conditions, the secondary structural content of the two proteins was the same.

The tertiary structure of CLIC1-M32A was analyzed using fluorescence spectroscopy. As previously mentioned, CLIC1 has one tryptophan and 8 tyrosine residues. Trp35 is found in the domain interface of CLIC1 with its indole chain partially exposed to the solvent (±27 % ASA in the native conformation). Trp35, whose side chain is sensitive to the polarity of its environment, is in close proximity to Met32 and was used as a local reporter of tertiary structural changes at the domain interface of CLIC1. The single tryptophan at position 35 was selectively excited at 295 nm. Excitation at 280 nm resulted in the excitation of the 8 CLIC1 tyrosines in addition to Trp35. Figure 37 depicts the fluorescence spectra of native and denatured CLIC1-M32A at pH 7.0 and

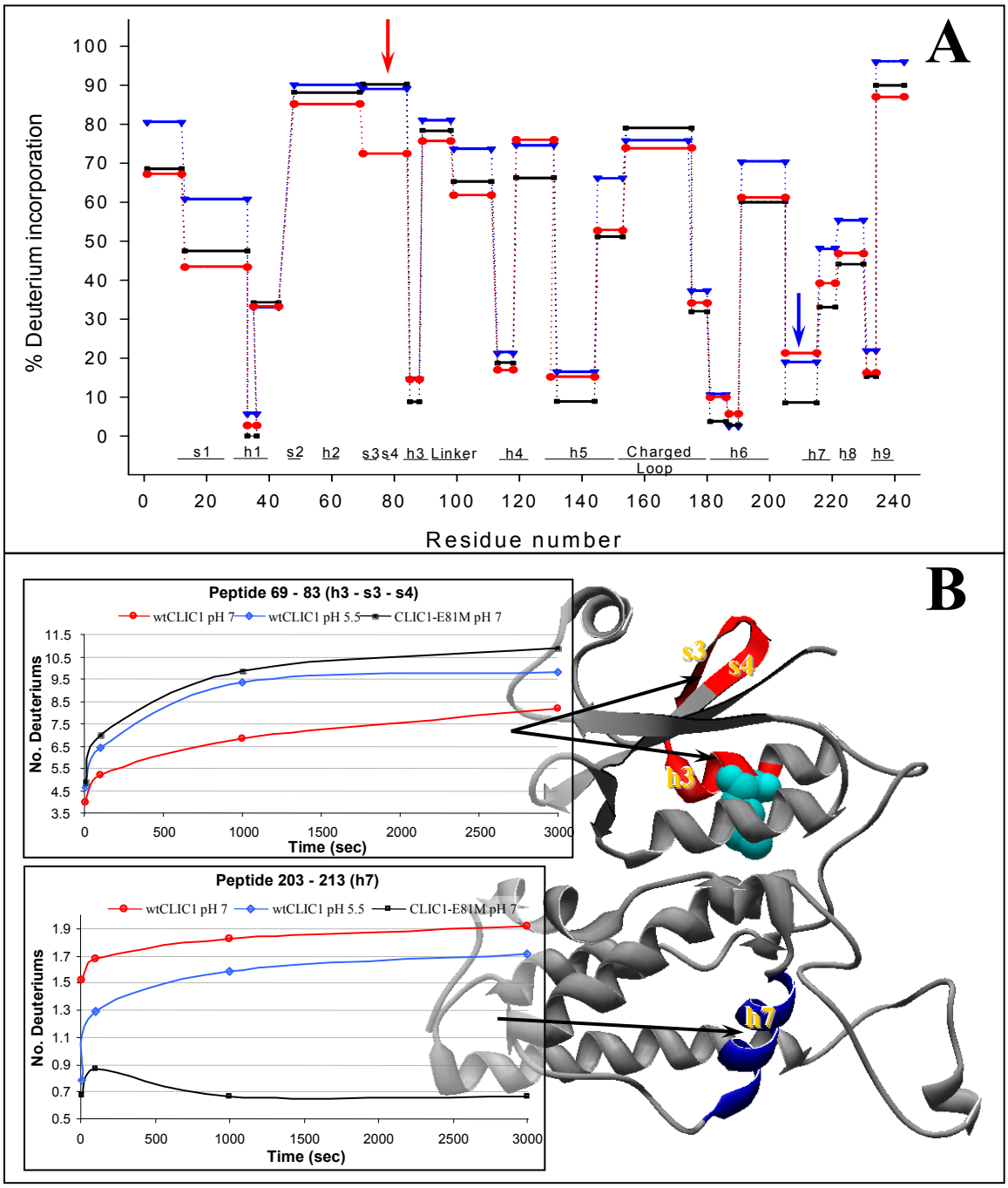


Figure 35: Effect of E81M on the structural dynamics of CLIC1

(A) Bar chart depicting deuterium exchange of CLIC1-E81M at pH 7.0 (black; 3000 s deuterium incubation), wtCLIC1 pH 7.0 (red; 3000 s deuterium incubation) and wt CLIC1 at pH 5.5 (blue; 90,000 s deuterium incubation). Peptide Tyr^{69} – Phe^{83} (red arrow) exhibits ~ 20 % higher deuterium localization in CLIC1-E81M than wtCLIC1 at pH 7.0. Fragment Phe^{203} – Ala^{213} in CLIC1-E81M shows ~ 10 % lower deuterium incorporation than wtCLIC1 at pH 7.0.

(B) The positions of fragment Tyr^{69} – Phe^{83} (red) and Phe^{203} – Ala^{213} (blue) are shown in relation to the Glu81Met mutation (cyan). The inset illustrates a plot of the number of exchanged deuteriums as a function of time for the two peptides. In the case of wtCLIC1 pH 5.5, the plots have been adjusted for the ~ 30 fold decrease in exchange due to the pH difference The structure shown is that of CLIC1 (pdb code: 1k0m).

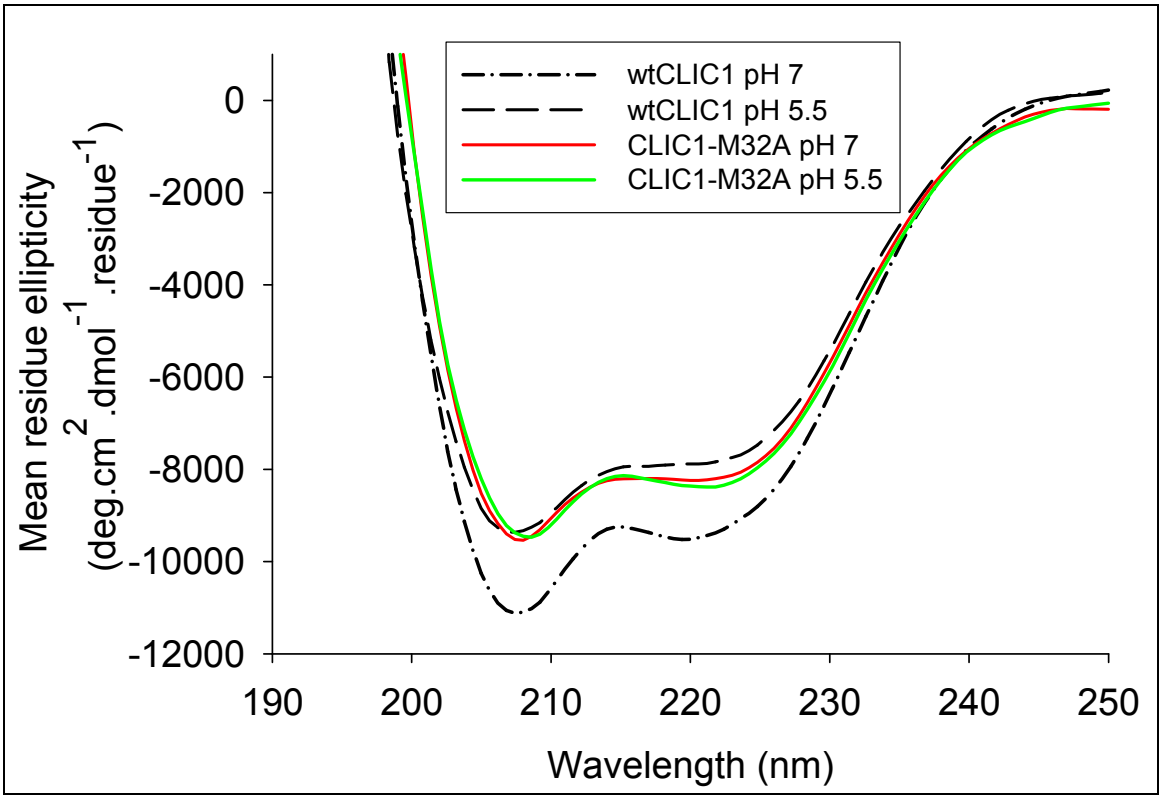


Figure 36: Far-UV CD spectra of native CLIC1-M32A and wtCLIC1

Far-UV CD spectra of native CLIC1-M32A at pH 7.0 (red) and pH 5.5 (green). The native CD spectra of wtCLIC1 at pH 7.0 (dashed-dot) and pH 5.5 (dashed) are also shown. wtCLIC1 data taken from McIntyre (2006). The difference between the wtCLIC1 at pH 7.0 signal and the rest of the illustrated spectra is ~ 16 %. The buffer used contained 5 mM sodium phosphate, 0.1 mM DTT, 0.02 % NaN_3. Readings were taken in a 2 mm cell at 20 °C. The spectra are an average of 10 readings at 100 nm/sec scan speed. The data were plotted using SigmaPlot v9.0 and smoothed using the negative exponential method.

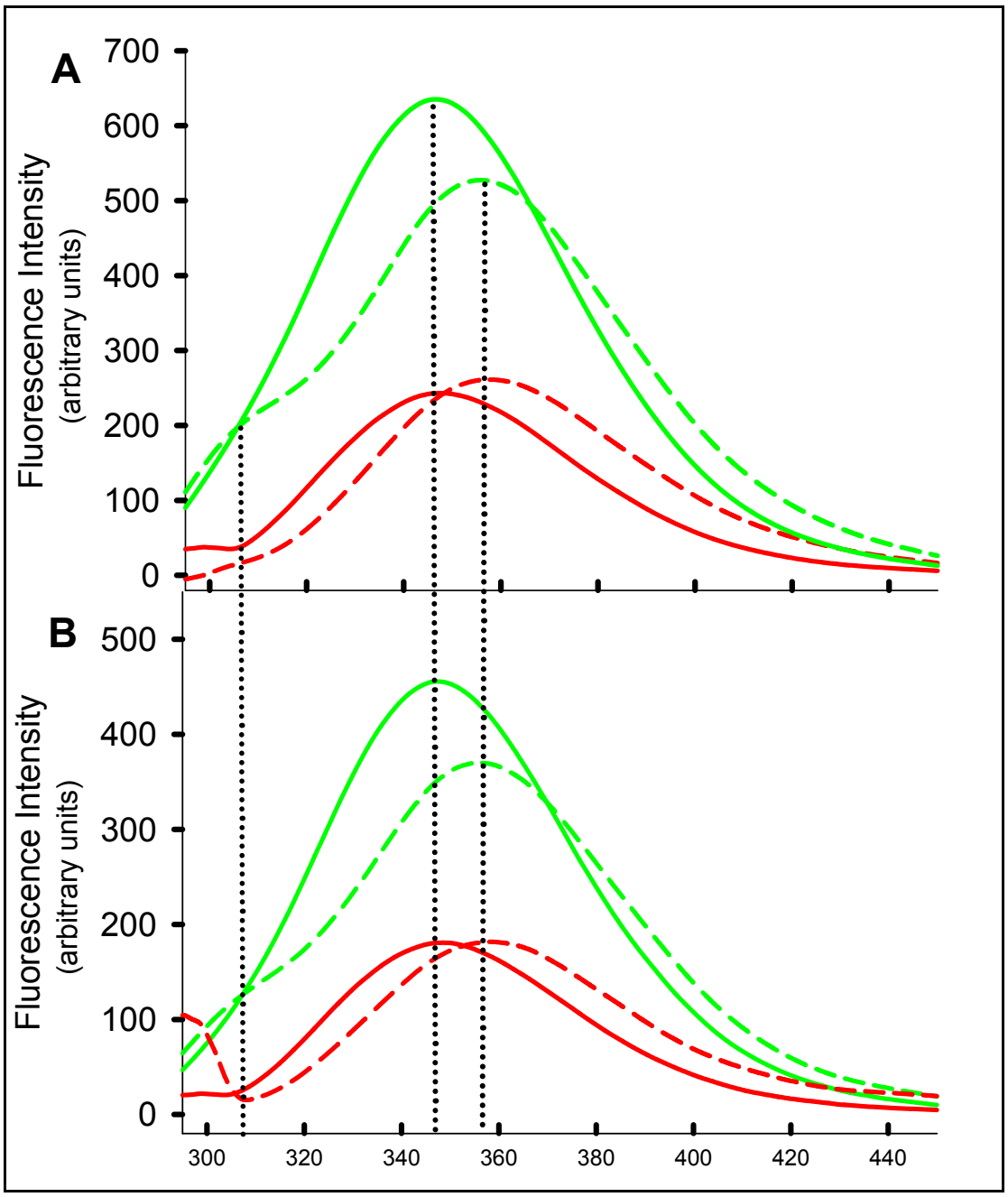


Figure 37: Fluorescence emission spectra of native and unfolded CLIC1-M32A

Fluorescence emission spectra of 2 μM CLIC1-M32A at **(A)** pH 7.0 and **(B)** pH 5.5. The protein was excited at 295 nm (red) and 280 nm (green) and the emission spectra were recorded. The native protein (solid lines) emitted at 347 nm while the denatured protein (dashed lines) emited at 358 nm. The denatured, 280 nm excitation, spectra exhibited an additional peak at 310 nm. The buffer used was 50 mM sodium phosphate, 1 mM DTT, 0.02 % NaN_3. Readings were taken using a 1ml cell, 5 nm slit width using a Perkin Elmer LS50B luminescence spectrometer. The spectra are an average of three accumulations at 200 nm/min scan speed. The data were plotted using SigmaPlot v9.0 and smoothed using the negative exponential method

pH 5.5. The fluorescence emission maxima, under varying pH environments, were the same suggesting that pH did not affect the tertiary structure of CLIC1-M32A. However, one has to keep in mind that fluorescence spectra are dominated by tryptophans' emission and as a result Trp35 emission may not be representative of the global structure of CLIC1. The native emission spectra, at 280 and 295 nm excitation, exhibited an emission wavelength maximum ($\lambda_{em\ max}$) at 347 nm (Figure 37). The corresponding $\lambda_{em\ max}$ of wtCLIC1 was at 345 nm as reported by McIntyre (2006). The denatured emission spectra of CLIC1-M32A display a red shift ($\lambda_{em\ max}$ = 358 nm) compared to the native spectra. In addition, the spectra resulting from excitation of denatured CLIC1-M32A at 280 nm, showed an additional peak at 310 nm (Figure 37). This peak resulted due to the uncoupling of energy transfer between the tryptophan and tyrosine residues upon unfolding.

3.7.2 Recovery of CLIC1-M32A

To be able to determine the conformational stability parameters, ΔG_{H20} and *m*-value, of CLIC1-M32A the native to unfolded reaction has to be fully reversible. Refolding of secondary and tertiary structure was monitored using far-UV CD and fluorescence spectroscopy. Figure 38 shows the fluorescence and far-UV CD spectra of native and refolded CLIC1-M32A at pH 7.0 and pH 5.5. At pH 7.0 (Figure 38A), the native and denatured spectra of the mutant protein were super-imposable indicating 100 % refolding (see section **2.2.8**). At pH 5.5 (Figure 38B), the fluorescence refolded spectra exhibited an increase in fluorescence intensity accompanied by a blue shift ($\lambda_{em\ max}$ 341 nm) compared to the native spectra ($\lambda_{em\ max}$ 347 nm). Samples containing refolded CLIC1-M32A showed a 50 % increase in scatter compared to samples containing native protein (results not shown). The CD spectra for native and refolded protein were not super-imposable at pH 5.5 (Figure 38B), thus not allowing analysis of equilibrium-unfolding curves and the calculation of ΔG_{H20} and *m*-value.

3.7.3 Effect of M32A on the conformational stability and folding cooperativity of CLIC1

The conformational stability of CLIC-M32A was analyzed using urea-induced equilibrium unfolding transitions. The unfolding transitions were monitored via Trp35 (295 nm excitation) and combined Trp35/Tyr (280 nm excitation) fluorescence, as well as ellipticity at 222 nm (E222) using far-UV CD. Figure 39 illustrates the

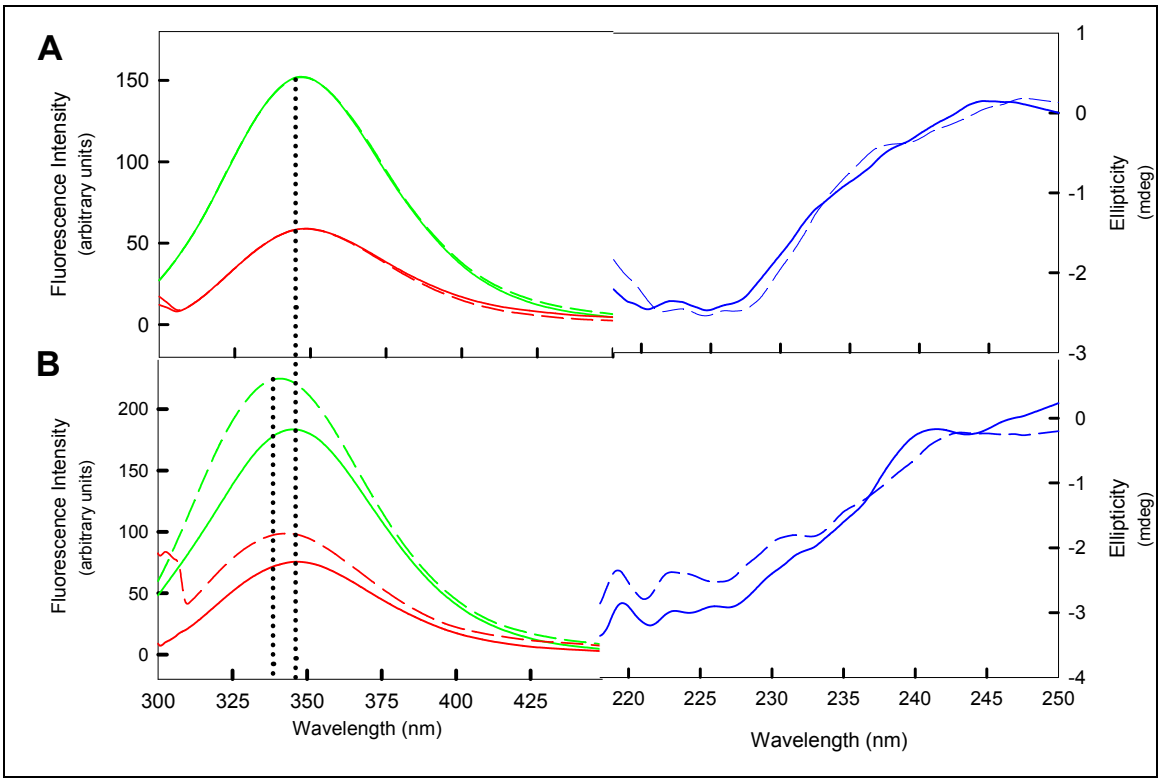


Figure 38: Reversibility of CLIC1-M32A unfolding

Recovery of tertiary and secondary structure of CLIC1-M32A after urea denaturation at **(A)** pH 7.0 and **(B)** pH 5.5. Tertiary structural changes were monitored via Trp35 emission (red) and Trp35/Tyr emission (green). Secondary structural change were detected via variations in far-UV-CD spectra (blue) Native spectra are shown as solid lines while the refolded spectra are indicated as dashed lines. 10 µM protein was unfolded in 7.5 M urea for one hour and then refolded for another hour via a ten-fold dilution into 50 mM sodium phosphate, 1 mM DTT, 0.02 % NaN_3. The data were plotted using SigmaPlot v9.0 and smoothed using the negative exponential method.

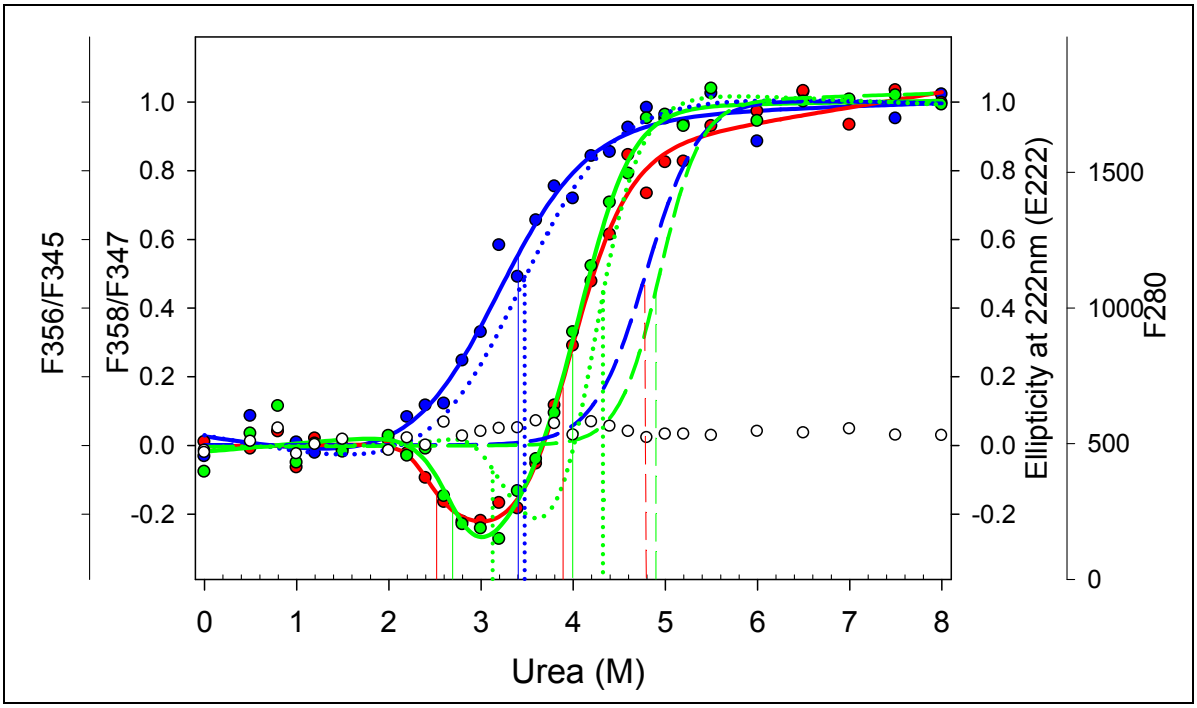


Figure 39: Equilibrium unfolding transitions of CLIC1-M32A

Equilibrium unfolding curves of 2 µM CLIC1-M32A at pH 7.0, wtCLIC1 at pH 7.0 and wtCLIC1 pH 5.5. The proteins were unfolded in 50 mM sodium phosphate, 0.1 mM DTT, 0.02 % NaN_3 with varying urea concentrations. F358/F347 (CLIC1-M32A) and F356/F345 (wtCLIC1) represent the ratios of the fluorescence intensities of the denatured to the native with increasing urea concentration (0-8M). The fluorescence signal was generated by either Trp35 excitation (red) or combined Trp35/Tyr excitation (green). The CD-monitored unfolding curves (blue) were generated by plotting the change in signal at 222 nm (E222). The dashed lines represent the unfolding transitions of wtCLIC1 at pH 7.0. The dotted lines illustrate the unfolding curves of wtCLIC1 at pH 5.5. All wtCLIC1 data were taken from McIntyre (2006). The C_m values for each curve are indicated by the drop lines. The Raleigh scatter at 280 nm (F280) is shown as (○). CLIC1-M32A fluorescence curves were fitted using a three-state monomer fit, while the CD curves were fitted to a two-state monomer model (see sections **2.2.11.1** and **2.2.11.2** respectively). The data were plotted using SigmaPlot v9.0.

equilibrium unfolding curves of CLIC1-M32A at pH 7.0. The mutants' Trp35 and Trp35/Tyr unfolding curves were super-imposable. Both transitions were non-sigmoidal and biphasic indicative of a three-state unfolding process. The first fluorescence transition was manifested as a minimum that occurred between 2.2 and 3.6 M urea with a lowest value at 3.4 M urea. The second fluorescence transition occurred between 3.6 M and 6 M urea as the protein unfolded completely. The Raleigh scatter plot (F280) depicted constant light scattering between 0 and 8 M urea indicating the absence of aggregates over the unfolding range. The fluorescence unfolding data was fitted using a three-state monomer model (N ↔ I ↔ U). This fit produced ΔG_{H2O} and *m*-values with large standard deviations making it impossible to compare these parameters to those obtained for wtCLIC1 (see Table 2).

Figure 39 also shows the CD-monitored unfolding transition (E222) of CLIC-M32A. The fluorescence and E222 curves were non-coincident confirming the presence of equilibrium intermediate/s. Unlike the fluorescence-generated curves, the CD-generated curve exhibits a single sigmoidal transition. The transition midpoint (C_m), point at which half the protein is unfolded, of the CD curve coincided with the first fluorescence phase of the Trp35 and Trp35/Tyr unfolding curves (Figure 39). The CD unfolding data was fitted using a two-state monomer model (N ↔ U) that generated very low ΔG_{H2O} (3.7 ± 0.8 kcal/mol) and *m*-value (1.2 ± 0.2 kcal/mol/M). The ΔG_{H2O} and *m*-value of other monomers such as Grx2 and wild-type CLIC1 at pH 7 are in the region of 12 kcal/mol and 2.5 kcal/mol/M urea, respectively (McIntyre, 2006).

The fluorescence and CD-monitored unfolding transitions of wtCLIC1 at pH 7.0 were monophasic and nearly super-imposable (Figure 39). The large standard deviations of the ΔG_{H2O} and *m*-values obtained from fluorescence-monitored unfolding and the underestimation of the ΔG_{H2O} and *m*-values obtained from the CD-monitored unfolding of CLIC1-M32A did not allow for a meaningful comparison to the corresponding ΔG_{H2O} and *m*-values of wtCLIC1. However, C_m values of CLIC-M32A denaturation curves were shifted to lower urea concentrations compared to the C_m values of wtCLIC1 at pH 7.0. The unfolding transitions of CLIC1-M32A at pH 7.0 closely resemble those of wtCLIC1 at pH 5.5 (see Figure 39). However, it must be noted that C_m values of CLIC1-M32A at pH 7.0 unfolding transitions are lower than the C_m values of wtCLIC1 at pH 5.5 (Figure 39).

Table 2: ΔG_{H_2O}, *m*-value and C_m values obtained from the equilibrium unfolding transitions of CLIC1-M32A

2-state fit			3-state fit					
ΔG_{H2O} (kcal/mol)	***m*** (kcal/mol/M)	C_m (M)	$\Delta G_{H2O,1}$ (kcal/mol)	m_1 (kcal/mol/M)	Cm_1 (M)	$\Delta G_{H20,2}$ (kcal/mol)	m_2 (kcal/mol/M)	Cm_2 (M)
Trp 35 Excitation								
-	-	-	7.8 (±10.3)	3.4 (±4.8)	2.5	7.8 (±1.9)	2.0 (±0.4)	3.9
Trp 35/Tyr. Excitation								
-	-	-	8.5 (±3.9)	3.1 ±1.6	2.7	8.5 (±2.5)	2.0 (±0.5)	4.0
Ellipticity at 222nm (E222)								
3.7 (±0.8)	1.2 (±0.2)	3.4	-	-	-	-	-	-

The ΔG_{H2O} and *m*-values were obtained by plotting F358/F347 and E222 as a function of urea concentration and fitting the data to either a two-state (N ↔ U) or a three-state monomer (N ↔ I ↔ U) models. The values in brackets represent the standard deviations obtained from the fits. The buffer used was 50 mM sodium phosphate, 0.1 mM DTT, 0.02 % NaN_3, pH 7.0.

3.7.4 Characterization of the stable equilibrium intermediate

Analysis of the equilibrium unfolding curves of CLIC1-M32A showed the presence of a thermodynamically stable equilibrium intermediate (see section **3.7.3**). The following section describes the results aimed at characterizing the structure of this intermediate.

3.7.4.1 Probing tertiary and secondary structure

CLIC1-M32A fluorescence emission spectra in the presence of increasing urea concentrations were recorded. The protein was excited at 295 nm (Trp35 excitation) and at 280 nm (Trp35/Tyr excitation). The fluorescence emission spectra of native CLIC1-M32A and CLIC1-M32A in 3.4 M urea at pH 7.0 are shown in Figure 40. The wavelength emission maximum ($\lambda_{em\ max}$) of the native spectra was at 347 nm. Between 2.2 and 3.6 M urea fluorescence emission spectra showed gradual blue shift compared to the native emission spectra (results not shown). The shift to lower wavelengths was largest at 3.4 M urea with an $\lambda_{em\ max}$ at 342 nm (Figure 40). The urea range within which the fluorescence spectra were blue shifted coincided with the first fluorescence unfolding phase of CLIC1-M32A (see section **3.7.3**, Figure 39).

Figure 41 shows far-UV CD spectra of native, 3.4 M urea and unfolded CLIC1-M32A at pH 7.0. The 3.4 M spectrum exhibited approximately 50 % decrease in the ellipticity signal compared to the corresponding native spectrum. In addition, the 222 nm minima of the 3.4 M spectrum was less pronounced. Since 108 amino acids make up the 9 helices of CLIC1, a 50 % decrease in helical content translates in 54 residues, involved in helical structures, being unfolded. Comparison of the unfolded to the 3.4 M urea ellipticity signals showed that the protein retained a considerable degree of secondary structure in the presence of 3.4 M urea.

3.7.4.2 ANS binding

CLIC1-M32A was exposed to the hydrophobic dye ANS in the presence of increasing urea concentrations (0-8M). Figure 42A shows fluorescence emission spectra of CLIC1-M32A at pH 7.0 in the presence of ANS. Excitation of protein-free ANS at 390 nm resulted in a fluorescence emission spectrum with an $\lambda_{em\ max}$ in the region of 525 nm. Consequently, all protein emission spectra were corrected for free ANS

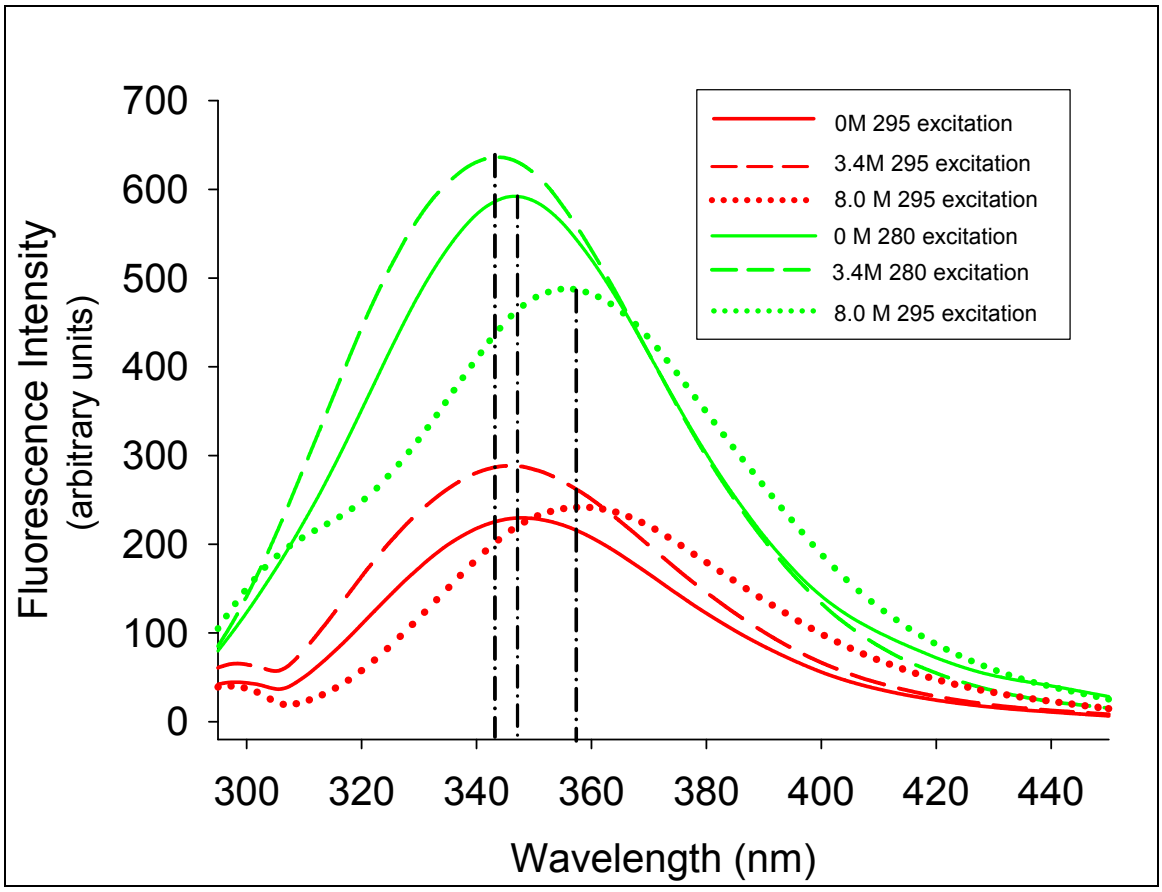


Figure 40: Fluorescence spectra of CLIC1-M32A with increasing urea concentration

Fluorescence spectra of 2 µM CLIC1-M32A. The protein was excited at 280 nm (green) and 295 nm (red) and the wavelength emissions recorded. Spectra of native CLIC1-M32A are shown as solid lines. The dashed lines represent spectra of CLIC1-M32A in the presence of 3.4 M urea. The wavelength emission maxima of low-urea concentration spectra are blue shifted compared the native spectra. The buffer used was 50 mM sodium phosphate, 1 mM DTT, 0.02 % NaN_3, pH 7.0. Readings were taken at 20 °C using a 1 ml cell, 5 nm slit width using a PerkinElmer LS50B luminescence spectrometer. The spectra are an average of three accumulations at 200 nm/min scan speed. The data were plotted using SigmaPlot v9.0 and smoothed using the negative exponential method.

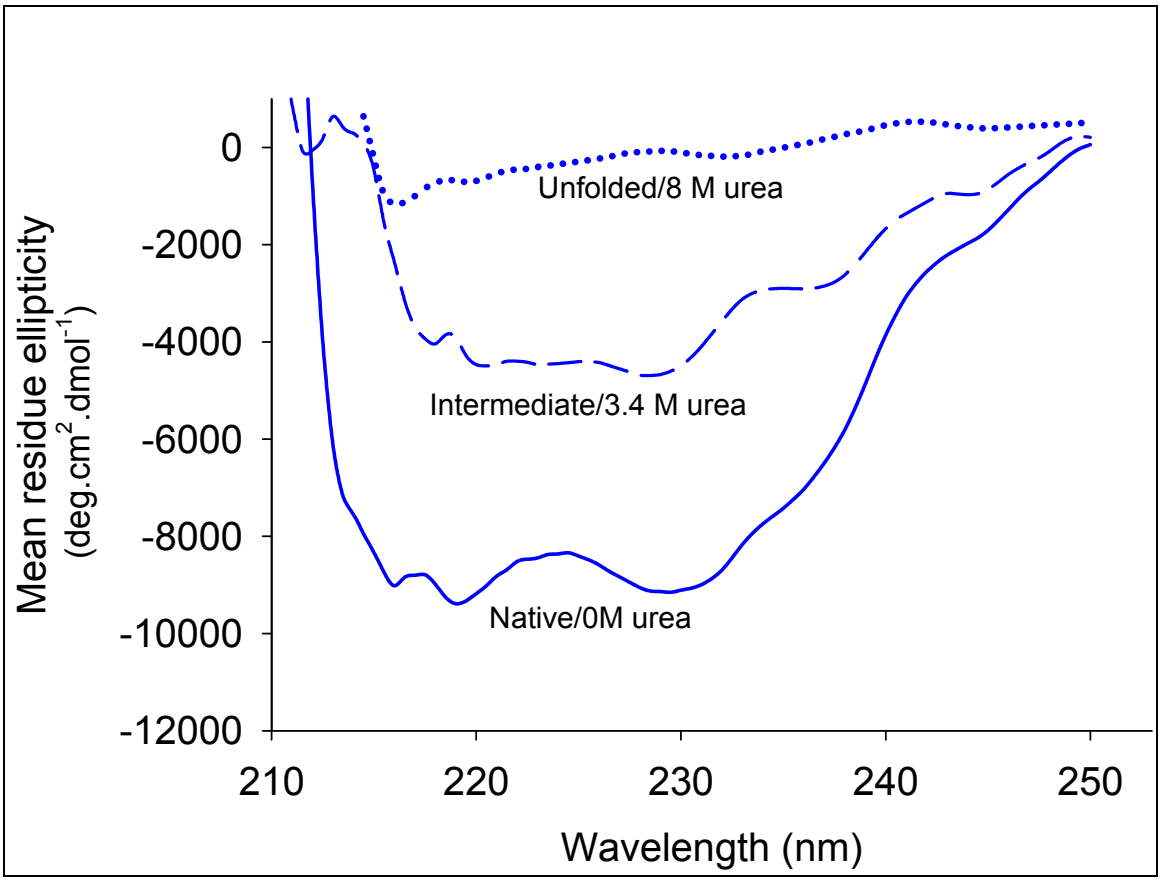


Figure 41: Far-UV CD spectra of CLIC1-M32A with increasing urea concentration

Far-UV-CD spectra of 2 μM CLIC1-M32A. There is ~ 50 % decrease in negativity of the ellipticity values of the 3.4 M spectrum compared to the native spectrum. The buffer used was 50 mM sodium phosphate, 1 mM DTT, 0.02 % NaN_3, pH 7.0. Readings were taken in a 2 mm cell at 20 °C using a Jasco J-810 spectropolarimeter. The spectra are an average of 10 readings at 100 nm/sec scan speed. The data were plotted using SigmaPlot v9.0 and smoothed using the negative exponential method.

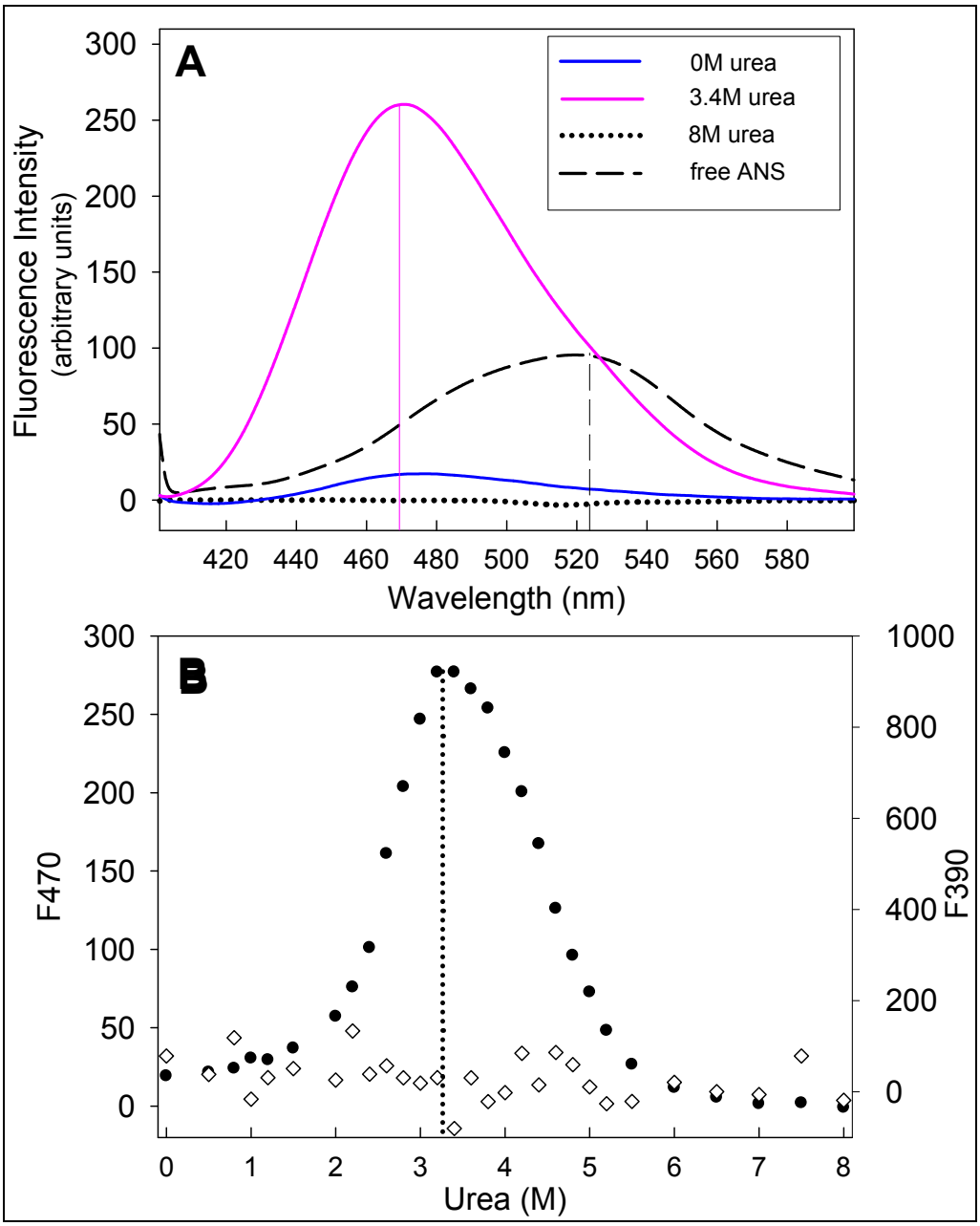


Figure 42: Binding of ANS to CLIC1-M32A

(A) Fluorescence emission spectra of 2 µM CLIC1-M32A in the presence of 200 µM ANS. The CLIC1-M32A spectrum in the presence of 3.4 M urea (dashed line) shows an increase in fluorescence intensity as well as a blue wavelength shift ($\lambda_{em\ max}$ = 470 nm) as compared to the native, and denatured spectra. All spectra are corrected for free ANS.

(B) Fluorescence intensities at 470 nm (●) and Raleigh scatter at 390 nm (◊) of 2 µM CLIC1-M32A in the presence of 200 µM ANS as a function of urea. The F470 plot indicates that CLIC1-M32A binds ANS between 2.2 and 5.5 M urea with the peak occurring around 3.4 M urea The buffer used in both **(A)** and **(B)** was 50 mM sodium phosphate, 1 mM DTT, 0.02 % NaN_3, pH 7.0. Readings were taken at 20 °C using a 1 ml cell, 5 nm slit width using a PerkinElmer LS50B luminescence spectrometer. The spectra are an average of three accumulations at 200 nm/min scan speed. The data were plotted using SigmaPlot v9.0 and smoothed using the negative exponential method descried.

emission. CLIC1-M32A at pH 7.0 did not bind ANS in the native and fully unfolded conformations. This was indicated by the lack of fluorescence emission of the native and denatured spectra along the 0-8 M urea range (Figure 42A). Upon incubating CLIC1-M32A with increasing urea concentrations ANS binding was observed as indicated by an increase in fluorescence emission of the corresponding spectra. Maximum binding of ANS to CLIC1-M32A occurred at 3.4 M urea. At 3.4 M urea, the fluorescence emission spectrum exhibited an $\lambda_{em\ max}$ at 470 nm with a fluorescence intensity of 250 arbitrary units (Figure 42A). The fluorescence intensities at 470 nm were plotted as function of urea (Figure 42B). There was an increase in fluorescence intensity between 2.2 and 5.5 M urea indicative of ANS binding. The Raleigh scatter plot (F390) showed no increase in fluorescence intensity along the 0 to 8 M urea range (Figure 42B). Hence, the binding of ANS to CLIC1-M32A between 2.2 and 5.5 M urea was not a scattering artefact. The increase in fluorescence emission between 2.2 and 3.4 M urea corresponded with the accumulation of the equilibrium intermediate (see section **3.7.3** and Figure 39).

3.5.4.3 Pulse-labelling DXMS under equilibrium conditions

As was shown in section **3.6**, native continuous-labelling DXMS provides information on changes in local as well as global conformational dynamics, of the folded state, induced by alterations in the proteins environment (wtCLIC1 at pH 7.0 versus wtCLIC1 at pH 5.5) or modifications due to mutation/s (CLIC1-M32A and CLIC1-E81M). On the other hand, equilibrium pulse-labelling DXMS uses denaturants such as urea to stabilize partially unfolded states and hence provide information on the folding/unfolding pathways of proteins (Deng *et al*., 1999). Under equilibrium pulse-labelling conditions, the populations of various folded/partially-folded/unfolded forms is constant and the short exposure of the protein to D_2O (10 s in this study) provides a snap-shot of all populations present under various urea-induced equilibrium unfolding conditions (Deng *et al*., 1999).

The equilibrium unfolding of wtCLIC1 at pH 7.0 and pH 5.5 as well as CLIC1-M32A at pH 7.0 were monitored via pulse-labelling DXMS. Initial analysis involved using wide scan ranges (700 – 1999 m/z) to record all charge states of the wild-type and mutant proteins (see section **3.5**, Figure 27). Later, narrow scans (1120 – 1150 m/z) were applied to record only the 24^+ charge state since, in the case of CLIC1-M32A, it

gave the best signal-to-noise ratio. The first step in the interpretation of the pulse-labelling DXMS experiments was to identify possible adduct/pseudomolecular ion peaks. These artificial peaks often obscure the data in terms of the number of protein conformers present. Table 3 lists common adducts encountered in positive polarity ESI. Pseudomolecular sodium ions are very stable and hence most numerously observed. The jump from $[M-H]^+$ (protonated protein) to $[M\text{-adduct}]^+$ (protein- adduct complex) was used to identify adduct ion peaks present in the m/z spectra of deuterium pulse-labelled wtCLIC1 and CLIC1-M32A. Figure 43 shows the 24^+ charge state of native wtCLIC1 at pH 7.0 and pH 5.5, as well as native CLIC1-M32A at pH 7.0. The proteins were pulse-labelled with deuterium for 10 s at pH 7.0 and 300 s at pH 5.5[ʒ] before MS analysis. In the case of wtCLIC1 at pH 7.0 and CLIC1-M32A at pH 7.0, the average mass difference between the two peaks was 22.1 and 22.6 amu, respectively (Figure 43 A and C). Hence, the split-peak appearance was most likely due to the presence of CLIC1-sodium adducts ($[N\text{-}Na]^+$). In the case of wtCLIC1 at pH 5.5, the mass difference between the two peaks was 18.5 amu and so the second peak was resulted from the possible formation of CLIC1-ammonium ion ($[N\text{-}NH_4]^+$) complexes (Figure 43B). The differences between the experimental and theoretical jump values, from $[M-H]^+$ to $[M\text{-adduct}]^+$ (Figure 43 versus Table 3), were due to errors in calculating the precise centroids of the full isotopic envelopes. These inaccuracies possibly resulted from the overlapping distributions of the $[M-H]^+$ and $[M\text{-adduct}]^+$ peaks. Adduct-induced peaks were also detected in the m/z spectra of non-deuterated controls, as well as partially- and fully-unfolded wtCLIC1 and CLIC1-M32A pulse-labelled with deuterium (see Figures 44 and 45).

Figure 44 is a line plot showing the evolution of the 24^+ charge state of CLIC1-M32A, pulse-labelled with deuterium for 10 s, as function of increasing urea concentration. Between 0 and 2 M urea the prominent signal was due to native conformation ($[N\text{-}Na]^+$) of CLIC1-M32A. However, a peak representing the intermediate (I) state was also detected in the pre-transition unfolding region (0 – 3 M urea). Between 3 – 3.6 M urea, the I state was progressively stabilized as shown by the increase in amplitude of the I isotopic envelope (Figure 44). The stabilization of the I state was also corroborated by the first fluorescence-detected transition in the

[ʒ] *Increase in labeling time accounts for the ~ 30 fold decrease in exchange rate as the pH decreases from 7.0 to 5.5 (*Bai *et al.,*1993*).*

Table 3: Common adducts encountered in positive polarity electrospray ionization (ESI^+) mass spectrometry (MS)

Adduct	Mass (amu)	Jump relative to $[M-H]^+$ [a] (amu)
$[M -H]^+$ [a]	1.0078	-
$[M + NH_4]^+$	18.0344	17
$[M + Na]^+$	22.9898	22
$[M + K]^+$	38.9637	38
$[M + CH_3CN]^+$	41.0265	41
$[M-H + 2Na]^+$	45.9795	45
$[M + Na + CH_3CN]^+$	64.0163	63

a: $[M-H]^+$ represents a protonated protein molecule

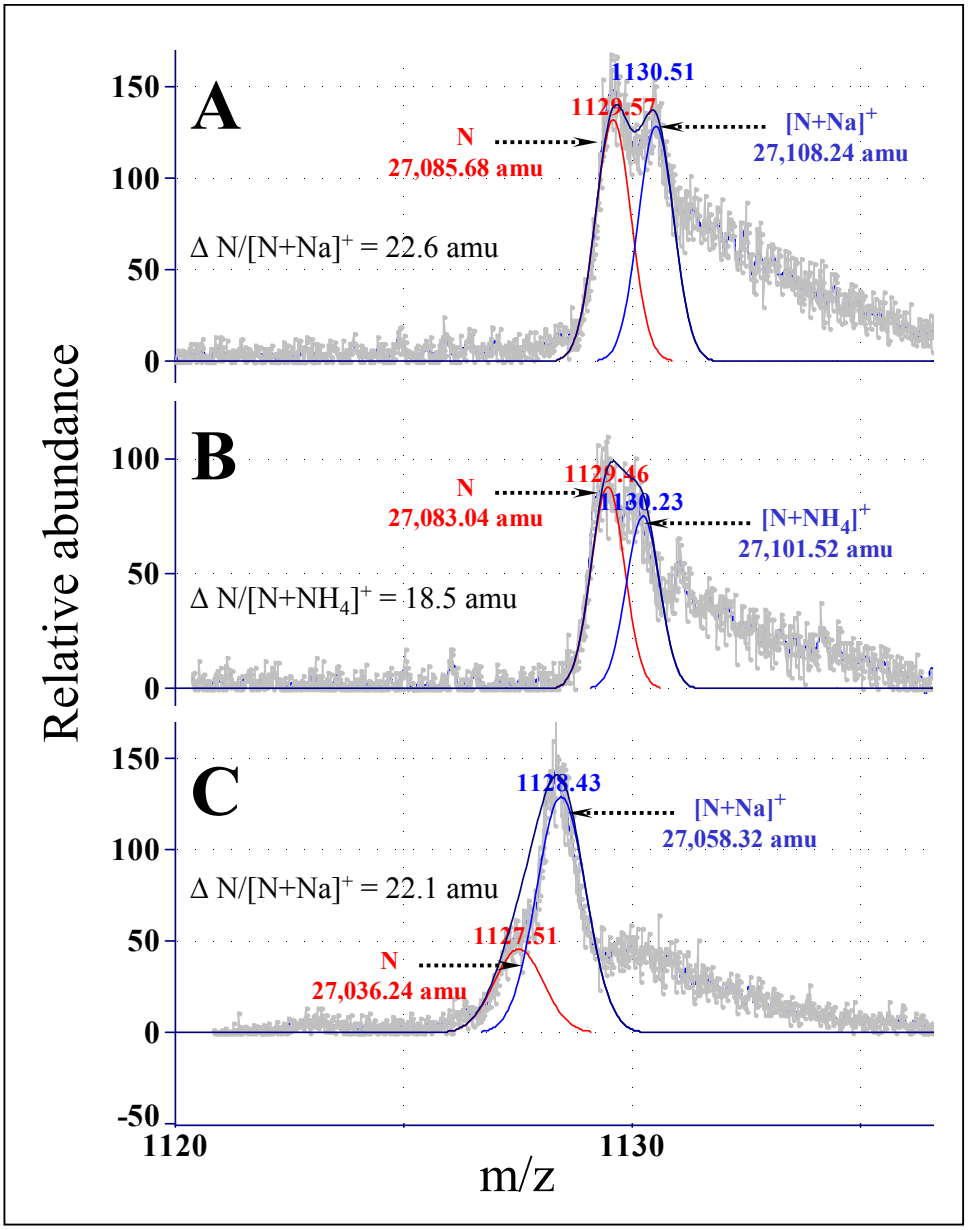


Figure 43: Adduct identification in the ESI+ spectra of deuterium pulse-labelled wtCLIC1 and CLIC1-M32A

Mass spectra representing the 24^+ charge state of deuterated **(A)** wtCLIC1 at pH 7.0, **(B)** wtCLIC1 at pH 5.5 and **(C)** CLIC1-M32A at pH 7.0. The raw data is shown in grey. The centroids of each peak are depicted above the corresponding isotopic envelopes. Below the arrows the average masses, calculated from centroid values, of the native (N) and adduct peaks are shown. In the case of wtCLIC at pH 7.0 and CLIC1-M32A at pH 7.0, the difference between the two peaks was 22.6 and 22.1 amu, respectively. Therefore the second envelope was most likely that of the sodium adduct ($[N+Na]^+$). In the case of wtCLIC1 at pH 5.5, the mass difference between the two peaks was 18.5 amu. Hence the second peak was identified as the ammonium adduct ($[N+NH_4]^+$). The data were fitted to Gausian peaks using the program PeakFit (AISN Software, Inc.). The relative population of each species present was calculated by measuring the area under each isotopic envelope. The proteins, pulsed-labelled with deuterium for 10 s, were in 50 mM Na_2HPO_4, 1mM DTT, 0.02 % NaN_3.

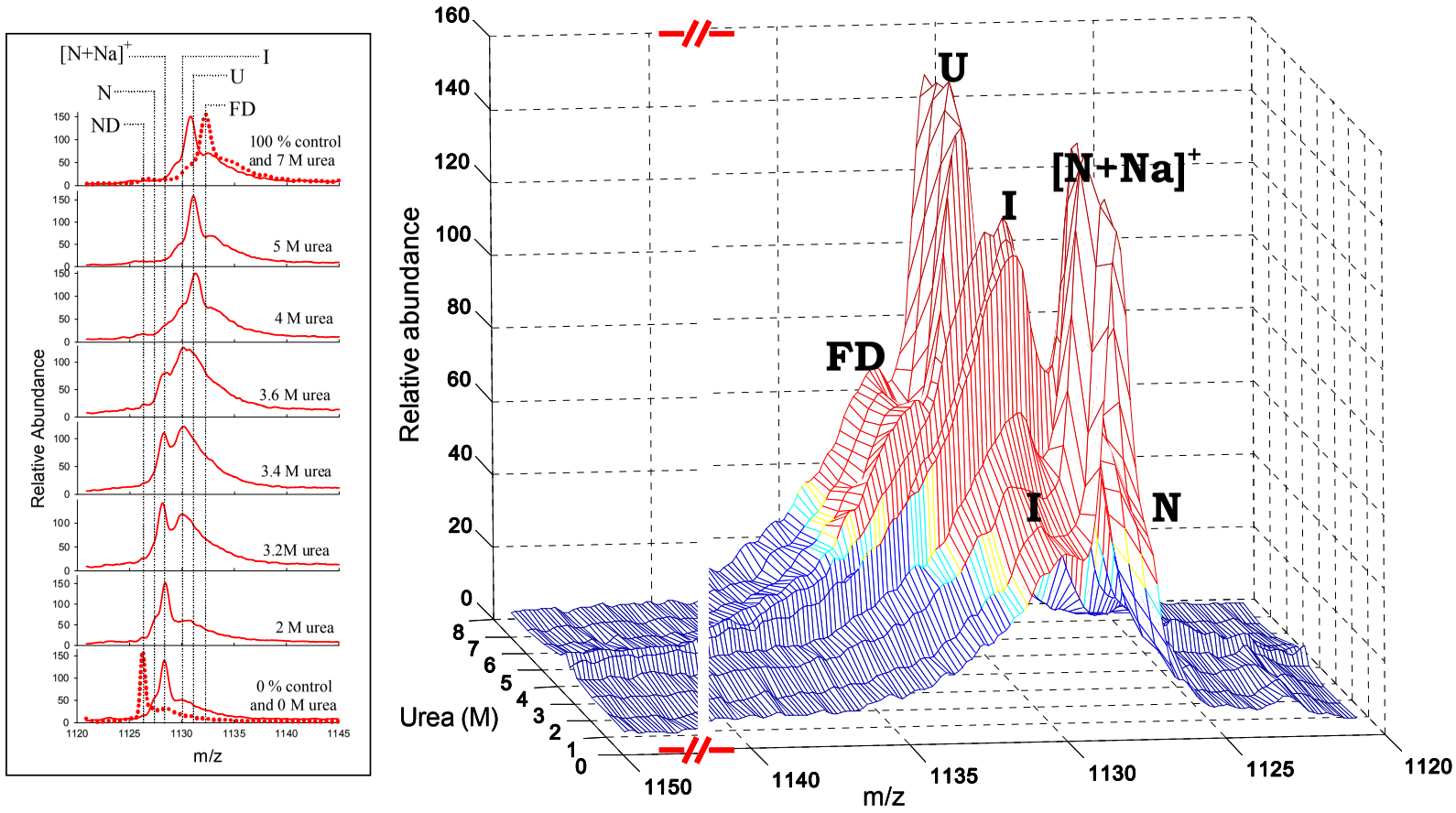


Figure 44: Equilibrium unfolding of CLIC1-M32A monitored by deuterium exchange mass spectrometry

The shift in m/z of the 24^+ charge state of CLIC1-M32A is shown as function of urea concentration. The Native (N), Native-adduct ($[N+Na]^+$), Intermediate (I), Unfolded (U) and Fully-deuterated (FD) peaks are marked on main figure. Inset depicts 0 % (Non-deuterated, ND) and 100 % (Fully deuterated, FD) controls (dashed lines in top and bottom panels) in addition to mass spectra of the 24^+ charge state of CLIC1-M32A in select urea concentrations. CLIC1-M32A was unfolded in 0 – 7 M urea for one hour followed by 10 s deuterium pulse-labelling. The protein was in 50 mM Na_2HPO_4, 1mM DTT, 0.02 % NaN_3 buffer, pH 7.0. The reaction was performed at 25 °C. Measurements were acquired on a Quadropole Time-Of-Flight (Q-TOF) Mass Spectrometer (MS) interfaced with a ULTIMA Ultra-Pressure-Liquid-Chromatography (UPLC) system.

equilibrium unfolding of CLIC1-M32A (section **3.7.3** and Figure 39). On the other hand, the relative abundance of the native conformation, represented by $[N\text{-}Na]^+$, gradually decreased as the protein was destabilized between 0 and 4 M urea. Between 4 and 7 M urea the unfolded (U) conformation became gradually more populated as indicated by the amplitude increase of the U isotopic envelope. Furthermore a peak synonymous with the I state was also detected in the 4 – 7 M urea range. Therefore, the so-called I state was present throughout the unfolding transition (0 – 7 M urea) of CLIC1-M32A. On the other hand, the I state was not detected in the pre- and post-unfolding transitions of CLIC1-M32A (see section **3.7.3** and Figure 39) when fluorescence and far-UV CD were used as a probes for equilibrium unfolding. The spectra of CLIC1-M32A that contained no deuterium (ND, 0 % control) and CLIC1-M32A whose amide hydrogens were completely exchanged (FD, 100 % control) are shown in Figure 44 inset. Noticeably, the centroid of the ND peak was smaller than the centroid of the N peak. This indicates that the native conformation of CLIC1-M32A contained a number of unprotected/unfolded residues whose amide hydrogens were involved in deuterium exchange. Similarly, the m/z centroid of the 100 % control peak was slightly higher than the m/z centroid of the 7 M urea peak. The difference suggests that part of CLIC1-M32A remained protected from deuterium in the post-unfolding transition and hence the U state possesses some form of folded structure/s.

The unfolding transition of CLIC1-M32A at pH 7.0, monitored by pulse-labelling DXMS, resembled that of wtCLIC1 at pH 5.5 (Figure 45). An I conformer was prominent between 3 and 4 M urea for both proteins. However, in the case of wtCLIC1 the I isotopic envelope could not be completely resolved (Figure 45) and as a result the I conformation was not clearly observed. This could be due to the fact that lower cone- and capillary-voltage values were accidentally set when data for wtCLIC1 was collected as compared to CLIC1-M32A (see section **2.2.12.2.2**). The cone- and capillary-voltage affect the overall analyte ion abundance as well as the level of sample fragmentation (Xian *et al.*, 2005). Too low settings result in signal decrease while too high values will produce excessive fragmentation (Xian *et al.*, 2005). wtCLIC1 at pH 7.0 exhibited a two-state unfolding when probed using pulse-labelling DXMS. Between 0 and 4 M urea the prominent signal was that of the native state (Figure 45).

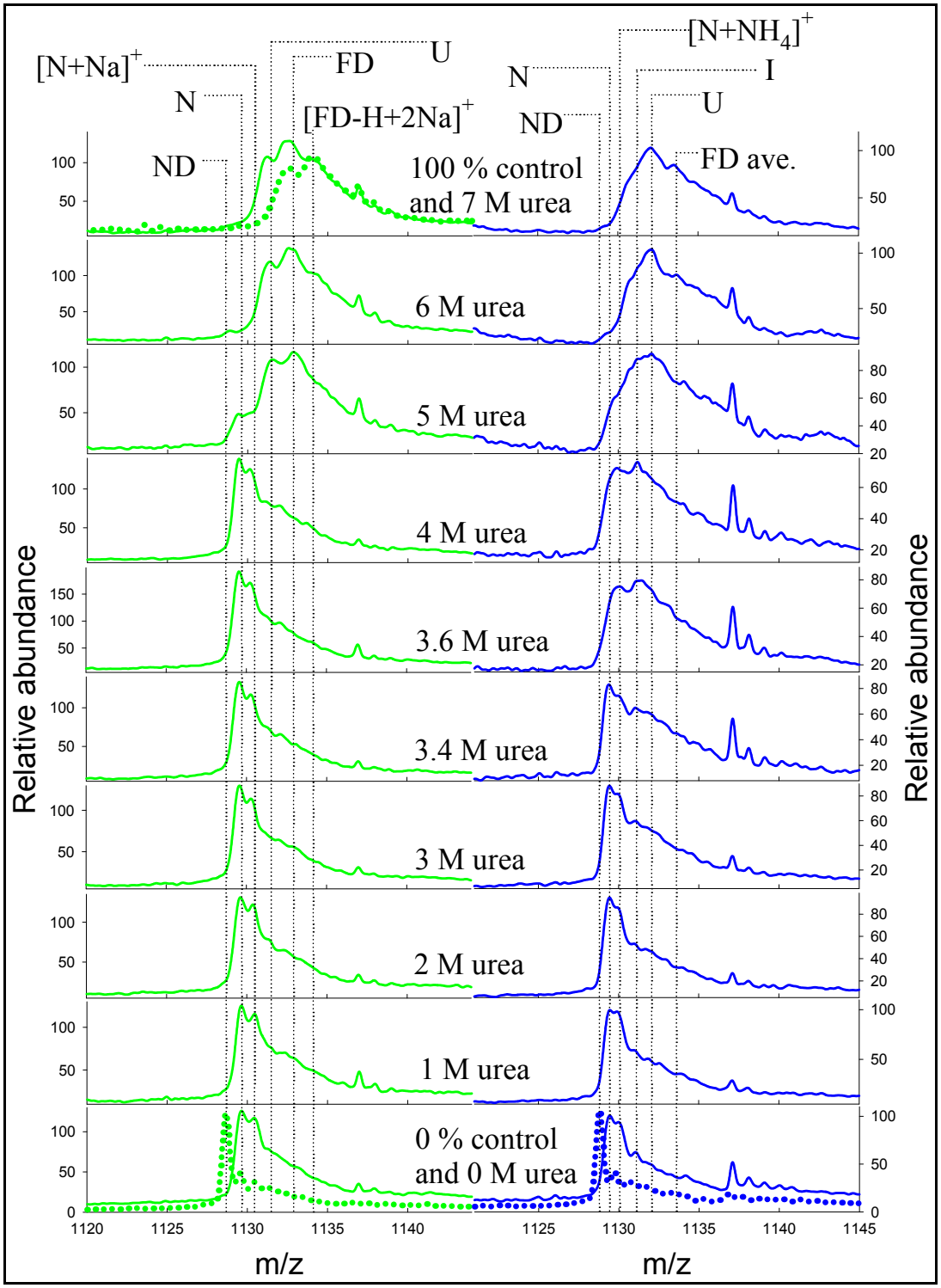


Figure 45: Equilibrium unfolding of wtCLIC1 monitored by deuterium exchange mass spectrometry

Evolution of the 24+ charge state of wtCLIC1 at pH 7.0 (green) and wtCLIC1 at pH 5.5 (blue) plotted as function of urea concentration. In the case of wtCLIC1 at pH 7.0, the predominant peak between 0 and 4 M urea was that of the native state (N and ($[N+Na]^+$). From 5 – 8 M urea the unfolded state (U) was most highly populated. In the case of wtCLIC1 at pH 5.5, an intermediate state (I) was detected between 3 and 4 M urea. wtCLIC1 at pH 5.5 was deuterated for 300 s (10s deuteration in case of wtCLIC1 at pH 7.0) in order to account for the slower exchange at pH 5.5. FD ave. was obtained by averaging the centroids of fully-deuterated isotopic envelopes detected in m/z scans of wtCLIC1 at pH 5.5 incubated in 0 – 7 M urea.

At high denaturant concentrations, 5 – 7 M urea, the major isotopic envelope was that of the unfolded state. Non-deuterated and fully deuterated centroids of wtCLIC1 at pH 7.0 and pH 5.5 are also marked in Figure 45. At both pH environments, the centroids of the ND peaks were lower than the centroids of the N envelopes. This, similarly to CLIC1-M32A, reflects the exchange of backbone amide protons that were weakly protected (not involved in hydrogen bonds) or unprotected (unfolded) in the native conformation ofwtCLIC1. In addition, the centroids of the FD envelopes of wtCLIC1 at pH 7.0 and pH 5.5[ψ] did not coincide with the relevant centroids belonging to the unfolded peaks of wtCLIC1 (see Figure 45). Thus, at 7.0 M urea, part of the backbone of wtCLIC1 at pH 7.0 and pH 5.5 remained folded and prevented a number of amide hydrogens from exchanging with deuterium.

The raw pulse-labelling DXMS scans of CLIC1-M32A and wtCLIC1, shown in Figures 44 and 45 were fitted to Gaussian peaks using PeakFit (AISN Software, Inc.). As previously mentioned, the area under the N, I and U isotopic envelopes were used to determine the relative populations of equilibrium-species (section **2.2.12.2.3**). In addition, the level of deuterium incorporation was calculated using the centroids of folded/partially-folded/unfolded isotopic envelopes. This fraction of deuterium incorporation was used as a guide to the loss/gain of secondary/tertiary content between the N, I and U conformers of wtCLIC1 and CLIC1-M32A. It must be noted at this point that calculations involving number of exchanged deuteriums as well as percentage deuterium incorporation excluded the 14 proline residues of CLIC1[f]. Figure 46 illustrates 3D line plots depicting the relative populations and the percentage deuterium incorporation of the N, I and U conformers of wtCLIC1 (Figure 46A and B) and CLIC1-M32A (Figure 46C) as a function of urea concentration (full numerical data can be found in Appendix, **Table H**). As previously noted, the N conformation of CLIC1-M32A and wtCLIC1 at pH 7.0 and pH 5.5 contained a number of unprotected amide protons. In the case of the mutant protein, the average number of localized deuteriums, after back-exchange correction, was 39 (Table 4). In comparison, the N state of wtCLIC1 at pH 5.5 was shown to incorporate an average of

---

[ψ] *The signal to noise ratio of the fully-deuterated wtCLIC1 at pH 5.5 sample was very low. Hence, an FD ave. was obtained by averaging the centroids of fully-deuterated isotopic envelopes detected in m/z scans of wtCLIC1 at pH 5.5 incubated in 0 – 7 M urea.*

[f]: *There are no amide hydrogen atoms at proline peptide linkages.*

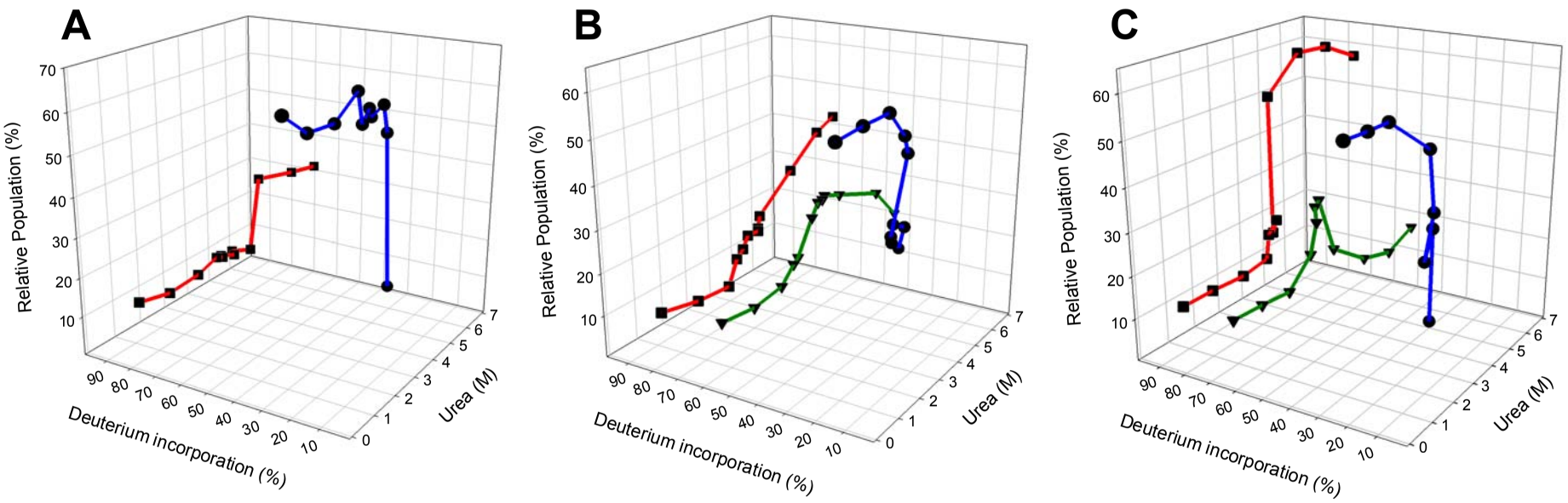


Figure 46: Relative populations and higher-order structural content of wtCLIC1 and CLIC1-M32A equilibrium species

Relative populations of the native N (●/blue), intermediate I (▼/green) and unfolded U (■/red) equilibrium conformations of **(A)** wCLIC1 at pH 7.0, **(B)** wtCLIC1 at pH 5.5 and **(C)** CLIC1-M32A at pH 7.0, plotted as a function of urea and deuterium incorporation. The data was obtained by monitoring the 24^{+} charge state of deuterium-pulse-labelled CLIC1. Subsequently, the resulting m/z spectra were fitted using Peakfit (AISN Software, Inc.). The areas under the N, I and U isotopic envelopes were used to calculate the conformers relative populations. The deuterium incorporation was calculated from the centroids of folded/partially-folded/unfolded isotopic envelopes as per the method used by Zheng and Smith (1993). Next, the number of back-exchange-corrected protein-assimilated deuteriums were divided by the total number of exchangeable amides to obtain the fraction/percentage of deuterium localization (see section **2.2.12.2.3** for calculations as well as Appendix Table H for full data). The percentage deuterium incorporation was used as a measure of the gain/loss of secondary and/or tertiary structural content between the N, I and U equilibrium-state species of wtCLIC1 and CLIC1-M32A.

Table 4: Number of unprotected amides and fraction deuterium incorporation of wtCLIC1 and CLIC1-M32A equlibrium unfoldding species detected via pulse-labelling DXMS

		Native	Intermediate	Unfolded
wtCLIC1 pH 7	Unprotected amides [a]	50 ± 4		167 ± 6
	ave. % D_2O incorp. [b]	22% ± 2%		74% ± 3%
wtCLIC1 pH 5.5	Unprotected amides [a]	42 ± 12	114 ± 7	166 ± 6
	ave. % D_2O incorp. [b]	18% ± 5%	50% ± 3%	73% ± 3%
CLIC1-M32A pH 7	Unprotected amides [a]	39 ± 7	133 ± 7	179 ± 9
	ave. % D_2O incorp. [b]	17% ± 3%	59% ± 3%	79% ± 4%

***** All values exclude proline residues with the total number of CLIC1 exchangable amides equal to 229

a: The number of unprotected amides was calculated by averaging the back-exchnage corrected number of localized deuteriums [**# D_2O (Adj.)**] form Appendix, Table H

b: The average % D_2O incoroporation was calculated by averaging the **% D2O (adj.)** column for each state from Appendix, Table H **% D2O (adj.)**

42 deuteriums, while N-wtCLIC1 at pH 7.0 averaged 50 unprotected amide hydrogens. The higher number of exchanged deuteriums, in the case of wtCLIC1 at pH 7.0, was surprising since continuous pulse-labelling DXMS experiments of native CLIC1 showed that the protein was slightly less flexible, and therefore more structured, at pH 7.0 than pH 5.5 (Nathaniel, 2006). Interestingly, the number of unprotected/unfolded amides of N-CLIC1-M32A decreased significantly between 3.0and 3.4 M urea (Figure 46C and Appendix, **Table H**). At this urea range the fluorescence emission spectra of CLIC1-M32A were blue shifted compared to the native emission spectra (see section **3.7.4.1** and Figure 40). In the case of wtCLIC1 at pH 5.5, the N species gradually unfolded between 0 and 5.0 M urea as shown by the congruent increase in deuterium incorporation with urea concentration (Figure 46B and Appendix, **Table H**). The N state of wtCLIC1 at pH 7.0 was not significantly altered, in terms of unstructured/unprotected amide hydrogens, with increasing urea concentration from 0 – 5.0 M (Figure 46A and Appendix, **Table H**).

As illustrated in Figure 46A, only two species, N and U, were detected in the case of wtCLIC1 at pH 7.0, confirming the two-state behaviour shown by the fluorescence and CD-monitored unfolding transitions of the wild-type protein at pH 7.0, 20 C° (see section **3.7.3** and Figure 39; McIntyre, 2006). On the other hand, pulse-labelling DXMS analysis of CLIC1-M32A at pH 7.0 and wtCLIC1 at pH 5.5 confirmed the presence of I conformers at equilibrium. In the case of CLIC1-M32A, the I state was most abundant between 3.2 and 3.6 M urea with a population of ~ 30 % (Figure 46A and Appendix, **Table H**). In comparison, I-wtCLIC1 at pH 5.5 was most highly populated between 3.6 and 4.0 M urea with ~ 32 % abundance (Figure 46B and Appendix, **Table H**). Interestingly, the partially-unfolded species of CLIC1-M32A and wtCLIC1 at pH 5.5 were also detected in the pre- and post-unfolding transitions of the two proteins, with an average population of ~15 %. In fact, the I species of wtCLIC1 at pH 5.5 was still significantly populated at 6.0 M urea with 28 % abundance (Figure 46B and Appendix, **Table H**). The I-CLIC1-M32A conformation exhibited an average of 144 unprotected amides, while the average number of deuterium exchange sites observed in I-wtCLIC1 at pH 5.5 was 114 (Table 4). This implied that in total, in the intermediate conformation, ~ 60 % of CLIC1-M32A and ~ 50 % of wtCLIC1 at pH 5.5 were unstructured. The average number of residues that unfolded from N to I were 94, in the case of CLIC1-M32A, and 72 in the case of

wtCLIC1 at pH 5.5. This translated to ~ 41 % loss in structural content in the conversion from N to I for CLIC1-M32A. On the other hand in the case of wtCLIC1 at pH 5.5, the N to I transition resulted in ~ 31 % decrease of ordered structures.

The U conformations of CLIC1-M32A and wtCLIC1 at pH 7.0 and pH 5.5 contained a significant degree of secondary and/or tertiary structures. On average ~ 27 %, 26 % and 20 % of amide hydrogens remained protected in the U state of wtCLIC1 at pH 7.0, wtCLIC1 pH 5.5 and CLIC1-M32A at pH 7.0, respectively (Figure 46 and Table 4). In the case of the mutant protein, this meant that approximately 50 residues remained folded (protected from hydrogen exchange) in the U species. In comparison, the number of wtCLIC1 residues involved in higher-order structures in the U conformation was in the region of 60. Noticeably, in the case of wtCLIC1 at pH 7.0 and pH 5.5 the relative population of the U state only reached 32 and 42 % at 7.0 M urea, respectively (Figure 46 A and B and Appendix, **Table H**). The remainder of the total population was made up by the fully-deuterated species of wtCLIC1.

CHAPTER 4

DISCUSSION

Most studies performed on CLIC proteins are based on functional analyses where their ability to form ion channels and subsequently conduct anions are tested physiologically (Landry *et al*., 1993; Chuang *et al*., 1999; Tulk *et al*., 2000; Harrop *et al*., 2001; Tulk *et al*., 2002; Berryman *et al*., 2004; Littler *et al*., 2005). Although, these studies reveal critical information about their ability to conduct Cl^- ions without the aid of other proteins and/or co-factors (Tulk *et al*., 2000), relatively little is known about the transformation of CLICs from a soluble to a membrane-inserted state. This has recently been addressed through two studies involving wtCLIC1 and the effects of the environment on its structural stability, folding and interaction with membranes (Nathaniel, 2006; McIntyre, 2006). As a result, a detailed pathway was established describing a series of structural changes that affect the native conformation of CLIC1 as the protein goes from conditions likely to be found in the cytoplasm to conditions that are prevalent at the membrane surface. The aim of the present study is to zoom in on the structure of CLIC1 and attempt to uncover the mechanism/s responsible for the structural alterations that transform the protein from a soluble form to a membrane form. Specifically, the role of the domain interface in the stability, folding and function of CLIC1 were examined using both bioinformatics and experimentally-based techniques.

4.1 Domain interface components involved in CLIC1 stability, folding and function

4.1.1 Primary structure anatomy

Similarly to most domain and dimer interfaces, non-polar contacts play a principal role in the association and stabilization of the N- and C-domains of GST and CLIC proteins (section **3.1.1** and Appendix, **Table A**). Hydrophobic interactions are nonspecific thus allowing sequence variation. Therefore, a number of different hydrophobic residues are found at non-polar hot-spot. On the other hand, polar charged contacts are more specific and primarily one type of residue is found at polar hot-spots. The proposed domain-addition evolutionary pathway of GST proteins (Ladner *et al.*, 2004) provides an insight into why the N-domain was found to be

approximately 70 % more conserved than the C-domain. The Ladner study suggested that GST proteins diverged from a single-domain protein with a thioredoxin-like fold whose role was to recognize glutathione (GSH) and catalyze the addition of GSHs thiol group to electrophilic substrates. The subsequent need to extend this function resulted in the addition of varying C-terminal domains to the thioredoxin-like domain forming monomeric CLIC1 and Grx2, and the association of monomers to form dimeric GSTs. Hence, the structural fold of the thioredoxin-like N-domain has been generally preserved while the structure of the subsequently added C-domain is much more variable.

Hot-regions preserve highly stabilising contacts and form a continuous, cooperative network of interactions. Hence, domain interface residues within the 3 hot-regions of GST family proteins (Figure 8) were deemed as critical components in the stabilization and the association of the domain interface. One particular set of interactions that stood out was an inter-domain lock-and-key motif, first identified in class Alpha GSTs (Wallace *et al*., 2000). Removal of the hydrophobic side chain locking the N- and C-terminal domains of hGSTA1-1 resulted in significant destabilization and loss in cooperative folding (Wallace *et al*., 2000). In the present study the lock-and-key interaction (Figure 9) is investigated because the lock residue is structurally conserved in the GST family (see position 8 in Appendix, **Table A**). The anatomy of this inter-domain interaction fits the O-ring proposition (Bogan and Thorn, 1998), which states that in order for an amino acid to have a large impact on the free energy of binding it needs to be protected from contact with the bulk solvent (Bogan and Thorn, 1998). In the case of the GST family the side chain of the lock residue is buried in the domain interface, protruding from the N-domain into a hydrophobic pocket found at the C-domain (Figure 9). Solvent exclusion is achieved by tight packing and surrounding the side chain of the lock residue with energetically unimportant contacts formed by moderately conserved amino acids (Figure 9). In terms of the CLICs, the importance of the lock-and-key motif, based on visual inspection, is further highlighted due to the fact that the majority of the domain interface is formed through h3 contacts with the insertion of the hydrophobic side chain of Met32 (lock residue in the case of CLIC1) in the C-domain being the prominent interaction.

Although the CLICs and GSTs are structural homologues, the ability of CLIC proteins to exist in a soluble form as well as a membrane form implies that their amino acid sequences contain a unique set of contacts that, under the correct conditions, allow the soluble to membrane metamorphosis to take place. Three sets of domain-interface interactions were identified as unique to the CLIC1 family (Figure 11). The network of contacts formed by Glu81 – Arg29 and Glu85 – Lys37 spans the domain interface joining the N-terminal h1 and h3 with the C-terminal h5 as well as the domain linker. Interestingly, h1 and h3, which form the N-terminal domain interface, are malleable with shifts of up to 1.5 Å (Harrop *et al.,* 2001). The displacement of these helices means that the domain interface of CLIC1, CLIC4, and possibly other CLIC proteins is flexible. This plasticity allows for contacts to be broken and/or formed under various environmental conditions. Consequently, the N- and C-domains can uncouple possibly leading to the formation of the membrane-competent conformation.

The contribution toward CLIC1 stability by the salt-bridges formed through Glu81 – Arg29 and Glu85 – Lys37 interactions is difficult to predict. In general buried salt-bridges, as was the case here, are found to have a stabilizing $\Delta\Delta G_{TOT}$[ℵ] of 4.53 ± 5.13 kcal/mol (Kumar and Nussinov, 1999). This is a huge stabilizing effect especially in view of the fact that wtCLIC1 at its most stable has a ΔG_{H2O} of ~ 10 kcal/mol (McIntyre, 2006). However, the large standard deviation of $\Delta\Delta G_{TOT}$ for buried salt-bridges means that the strength of these types of interactions is protein specific. As previously mentioned, the salt bridges formed by both Glu81 – Arg29 and Glu85 – Lys37 are found to be buried. This implies that the desolvation penalty for the burial of the charge groups upon folding will be high. This penalty will probably be less in the case of Glu81 – Arg29 because the immediate surroundings of this salt-bridge are mostly made up of polar residues. On the other hand, the ionic interaction formed via Glu85 – Lys37 is in a more hydrophobic environment. The desolvation penalties will be counteracted by a strong electrostatic attraction force because salt-bridges in the interior of proteins are better screened against the solvent (Kumar and Nussinov,

---

ℵ: $\Delta\Delta G_{TOT} = \Delta\Delta G_{DSLV} + \Delta\Delta G_{BRD} + \Delta\Delta G_{PRT}$. *Where* $\Delta\Delta G_{DSLV}$ *represents the sum of the unfavorable desolvation incurred by the burial of a charged group from a polar to a relatively non-polar environment* $\Delta\Delta G_{BRD}$ *represents the favorable energy from the electrostatic interactions between the charged groups of side chain atoms.* $\Delta\Delta G_{PRT}$ *represents the electrostatic interactions between the side chains of the salt-bridge forming residues and the side chains of the surrounding amino acids in the folded conformation. Equation was taken from Kumar and Nussinov (1999).*

1999). In addition, the presence of hydrogen bonds between the oppositely charged side-chains of Glu81 – Arg29 and Glu85 – Lys37 brings them closer to one other hence further increasing the strength of the salt-bridge. In terms of the salt-bridge geometry, the oppositely charged side-chains of Glu81 – Arg29 are more favorably oriented at approximately right angles to each other. Therefore, overall both set of interactions formed between Glu81 – Arg29 and Glu85 – Lys37 are expected to have a significant stabilizing contribution toward the soluble CLIC1 conformation, with the former probably forming the stronger contact.

Glu218, position 39 in the consensus GST-interface, forms part of the lock-and-key interaction. The charged side chain chemistry of this residue is unique to the CLICs (Figures 9, 10 and 11). At the corresponding position, most GST-family proteins possess hydrophobic amino acids (Appendix, **Table A**). Significantly all of the above-mentioned unique inter-domain interactions form, amongst others, charged contacts. This is important since the strength of ionic interaction is highly dependent on variations in pH. Lowering the pH has recently been shown to destabilize the conformation of CLIC1 significantly resulting in loss of helical content and the formation of an equilibrium unfolding intermediate (McIntyre, 2006). Due to the cooperative nature of protein folding it is unlikely that one set of contacts is solely responsible for the transition of CLIC1 from a soluble to a membrane-competent form. It is more likely that a set of communicating contacts, as those found at the domain interface (Figure 11), are weakened thus lowering the energy barrier between the soluble and membrane conformations of CLIC proteins.

4.1.2 Domain interface architecture

The size of the interface-accessible surface area, number of inter-domain hydrogen bonds and salt-bridges, the planarity and shape complementarity of domain interfaces are all related to the energy contribution of domain association and the intrinsic stability of the individual domains. For easier comparison these domain-architectural characteristics are summarized, in terms of the various GST classes, in Figure 47.

The domain interfaces of CLIC1 and CLIC4 contribute significantly toward keeping the native, soluble conformation of these proteins in tact. This is indicated by the fact that the iASAs of the analysed CLICs are much larger than the theoretical minimum

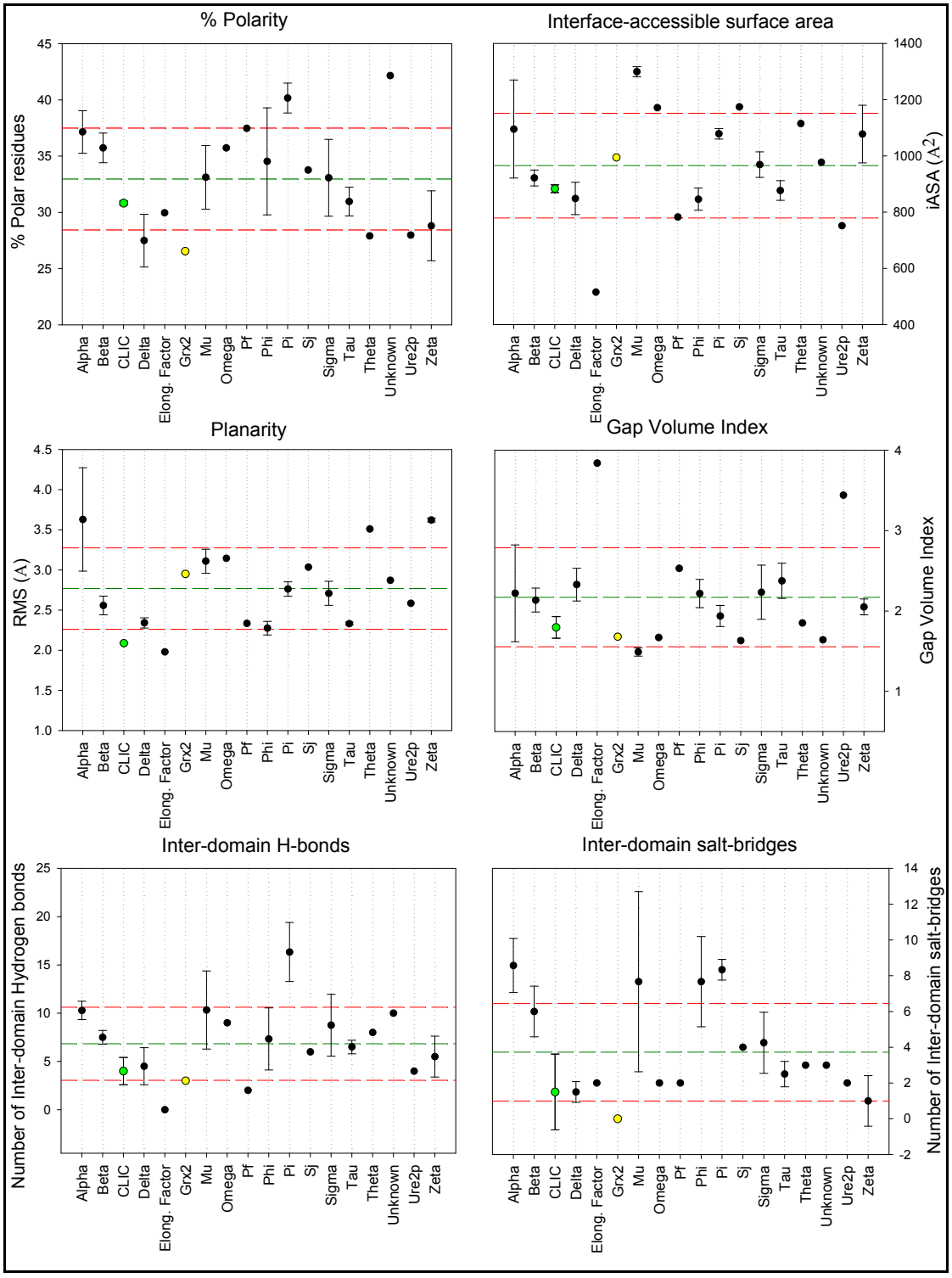


Figure 47: Summary of domain interface characteristics

Scatter plots showing six domain interface characteristics. The green line indicates the mean of each category, while the red lines represent the standard deviation. The errors bars depict standard deviations calculated from various members in each GST class.

for stable dimer formation (600 Å^2) (Chothis and Janin, 1975). This value corresponds to the minimal area required to exclude bulk solvent around highly stabilizing interactions (Bogan and Thorn, 1998). In addition, in terms of the area buried upon folding, the domain interfaces of CLIC1 and CLIC4 are more similar to dimer interfaces than to antibody-antigen interfaces. The former are designed to ensure stronger inter-subunit binding that stabilize proteins tertiary and quaternary structures, while the later have evolved to optimize rapid binding/release mechanisms rather than permanent stabilizing associations. However, in comparison to Grx2, the only other monomeric GST homologue, the CLICs have a smaller, less hydrophobic domain interface (Figure 47). This is reflected in the overall stability of the two monomeric proteins. CLIC1 at its most stable, pH 7.0 and 20 °C, has a ΔG_{H2O} value that was ~ 22 % smaller than the corresponding ΔG_{H2O} for Grx2 (McIntyre, 2006). Interestingly, the planarity values for the CLIC1 and CLIC4 domain interfaces are significantly smaller than those for dimeric proteins as well as the monomeric Grx2 (Figures 18 and 47). This implies that the domain interfaces of the CLICs are relatively smooth and as a result the conserved lock-and-key motif (Figure 9) is likely to play a crucial role in anchoring the N- and C-domains.

Another interface characteristic that is directly related to protein stability is the gap-volume index. As mentioned in sections **2.21** and **3.1.3.2**, the gap-volume index gives an indication of the complementarity of interacting surfaces. In terms of domain interfaces, the average strength of interactions between interfaces has been related to their complementarity (Jones and Thornton, 1996). For example, the poorly correlating antibody-antigen surfaces have smaller binding constants than the more complementary dimer interfaces. Therefore, a small gap-volume index points to strong binding between a pair of interacting interfaces and hence, a large contribution toward the overall stability of the protein. CLIC1 and CLIC4 have more complementary domain interfaces than most GST proteins (Figures 20 and 47). This reinforces the idea that inter-domain contacts are critical in the preservation of the native conformation of CLIC1, CLIC4 and possibly other CLICs in solution.

Inter-domain hydrogen bonds have an important role in protein stability for two reasons. Firstly, they are cooperative, a hydrogen bond to a backbone >C=O will strengthen a second hydrogen bond to a backbone >NH of the same peptide by further

polarizing the peptide unit (Stickle *et al.*, 1992), thus increasing their strength and enhancing the hydrogen bonding contribution toward conformational stability. Secondly, regions where networks of hydrogen bonds are found are more tightly packed. Therefore, the strength of van der Waals interactions in such regions increases leading to greater contribution to protein stability. CLIC1 has only one while CLIC4 has no networked inter-domain hydrogen bonds (see Appendix **Table C**). CLIC1 and CLIC4 are also found to possess fewer than average inter-domain hydrogen bonds and salt-bridges (Figure 47). These contacts are relatively solvent exposed possibly clamping the N- and C-domain during the later stages of folding. Thus, the energy contribution of inter-domain hydrogen bonds and salt-bridges, in the case of the CLICs, seems to be lower than most GSTs. As a result, under favourable conditions, like low pH, these contacts can be disturbed possibly initiating domain uncoupling and soluble to membrane-competent transformation.

Overall, in spite of the above-mentioned differences between the domain interface architectures of CLICs and the rest of the GSTs, there are no definitive clues pointing to a specific mechanism that may drive the conversion of CLIC proteins from a soluble to a membrane-competent form. If there is some kind of rearrangement taking place at the domain interface it is likely to be the result of cooperative reactions involving a number of interactions (as shown in Figure 11), that may spread well beyond the domain interface, rather than a single switch involving one specific contact.

4.2 Structural dynamics of native CLIC1 mutants

The role of two inter-domain contacts, namely the conserved lock-and-key motif (see Figure 9) and the unique salt-bridge between Arg29 and Glu81 (see Figures 10 and 11), were experimentally investigated using the engineered mutants CLIC1-M32A and CLIC1-E81M.

4.2.1 The bulk of the native conformation is unaffected by the domain interface mutations

4.2.1.1 Global structural probes

The overall secondary and tertiary structural content of CLIC1 is unaffected by the removal of the Arg29 – Glu81 salt-bridge, as shown by indistinguishable far-UV-CD and fluorescence spectra of CLIC1-E81M and wtCLIC1 obtained under identical conditions (Legg, PhD unpublished data).

The removal of the key residue, Met32, seems to significantly affect the secondary structure of CLIC1 with CLIC1-M32A showing an approximately 16 % decrease in helical content as compared to wtCLIC1 (Figure 36). However, the other global probes, fluorescence and pulse-labelling DXMS, used in this study indicate that the effects of the removal of the lock-and-key motif on native CLIC1 are not as significant as suggested by far-UV-CD. The $\lambda_{em\ max}$ obtained from fluorescence emission spectra of CLIC1-M32A is red shifted by 2 nm in comparison to the $\lambda_{em\ max}$ of wtCLIC1 (section **3.7.1**). The small change to higher wavelengths demonstrates that the scale of this mutation-induced rearrangement/s is much smaller than what is implied by the CD data. In fact, it is more accurate to explain the 2 nm red shift in terms of loss in local packing interactions in the region of the lone tryptophan (Trp35) rather than a more global unfolding/refolding event. In this case the removal of the large, hydrophobic side chain of Met32 by the introduction of an alanine eliminates a number of stabilizing contacts between h1 and h8 (see Figure 4). Consequently these helices will uncouple resulting in increase in solvent exposure of residues in the immediate environment of Met32, such as Trp35.

Pulse-labelling DXMS supports the fluorescence data in that the overall native conformation of CLIC1 is not significantly altered by the Met32Ala mutation. In the absence of denaturants, wtCLIC1 at pH 7.0 and pH 5.5 and CLIC1-M32A at pH 7.0 are shown to exchange on average 50, 42 and 39 (22 %, 18 % and 17 %) of their respective amide hydrogens with deuterium (Table 4 and Figure 46). The higher number of deuterated amides in the native state of wtCLIC1 at pH 7.0 suggests more unprotected/unstructured residues. However, one needs to consider that the data was

fitted to Gaussian peaks where deviations in m/z result in errors in mass and so in level of deuterium incorporation. In addition, calculations involving deuterium incorporation do not take into account any proline residues possibly introducing further inaccuracies. For these reasons, the differences in the number of exchanged deuteriums between the native conformations of wtCLIC1 at pH 7.0 and pH 5.5 as well as CLIC1-M32A at pH 7.0 were not considered as significant.

4.2.1.2 Local structural probe

The local structural probe used in this study, namely continuous-labelling DXMS, also shows that the overall native structure of CLIC1 is unaffected by the Met32Ala and Glu81Met mutations. Similarly to wtCLIC1, the C-terminal domain of the two mutants is more stable than the N-terminal domain with the majority of slow exchanging fragments present in the all helical domain 2 (Figures 31 and 34). As is the case with wtCLIC1, helices 4, 5, and 6 most likely form the folding-cores of CLIC1-M32A and CLIC1-E81M due to the fact that these secondary structures are the slowest-exchanging regions in both wild-type and mutant proteins (Figures 31 and 34). Hence, correct packing of helices 4 – 6 must be vital in maintaining the three-dimensional conformation of soluble CLIC1. This weakens Cromers' hypothesis (Cromer *et al*., 2002) that the C-terminal domain is involved in pore-forming activity and that, after solvent-exposure, h6 forms the trans-membrane helix (see section **1.5.5**). Structural re-arrangements involving this region will probably lead to total unfolding of CLIC1.

The rapidly exchanging fragments (first 10 N-terminal amino acids/Met^{1} – Leu^{10} and the bulk of the negatively charged loop/Pro^{147} – Gln^{164}) detected in wtCLIC1 at pH 7.0 also behave similarly in CLIC1-M32A and CLIC1-E81M (Nathaniel, 2006; Figures 30 and 33). A number of additional factors indicate that the lack of higher-order structure displayed by these two flexible regions did not result from the engineered mutations. Both fragments contain residues with high amide B-factor[℘] values. In addition, no clear electron density for residues Met^{1} – Gln^{5} could be obtained from crystals of wtCLIC1 (Harrop *et al.,* 2001). The negatively charged loop has slight structural alterations in two crystal forms of wtCLIC1, while Leu^{148} –

[℘]: *B-factor values for amide groups taken from crystal structure of CLIC1 (*Harrop *et al.,* 2001*)*

Arg^{165} which forms part of the loop, is disordered in one of the analysed wtCLIC1 crystals (Harrop *et al.,* 2001). Incidentally, total of 22 residues make up the two so-called unstructured regions (Met^{1} – Gln^{5}, Leu^{148} – Arg^{165}). In addition, another 11 amino acids have a SAS of 90 % or higher. Therefore in theory, 33 positions in the native conformation of CLIC1 could exchange with deuterium since they are unprotected. This number correlates relatively well with the average number of unprotected amide hydrogens determined via pulse-labelling DXMS for wtCLIC1 at pH 7.0 and pH 5.5 and CLIC1-M32A at pH 7.0 (Table 4), especially when considering that the later calculations do not include proline residues.

4.2.2 Met32Ala and Glu81Met induce local changes in native CLIC1

4.2.2.1 CLIC1-E81M

At pH 7.0, native CLIC1-E81M has three regions whose deuterium exchange patterns are significantly different than those of native wtCLIC1. One region (Val^{87} – Leu^{96}), exchanged faster than the corresponding peptide in wtCLIC1 only during the fast (10 s) and medium (100 s) exchange times (see section **3.6.3**). On the other hand, at longer exchange (1000 s to 3000 s) times no difference in deuterium localization is observed between Val^{87} – Leu^{96} of CLIC1-E81M and wtCLIC1 (Figure 35A). This observation implies that although the oscillations that this region experiences in solution are significantly faster in the case of CLIC1-E81M, the total number of unprotected amide hydrogens in the peptide are the same for the mutant and wild-type proteins. The Val^{87} – Leu^{96} fragment is situated in close proximity to the engineered mutation forming part of the domain linker. Leu96 is a highly conserved amino-acid in the GST family (position 29 in consensus GST-domain interface, Appendix, **Table A**). It forms part of a hydrogen bond network that involves Arg29 – Glu81 and Ly37 – Glu85 (Figure 11). The increased flexibility in the domain linker peptide could be the result of steric hindrance and/or disruption of the salt-bridge between Agr29 and Glu81 caused by the engineered mutation. Although there are no direct contacts between Glu81 and Lue96, the disturbance of the Val^{87} – Leu^{96} interaction is a good example of how a modification at one amino acid site can be manifested at a relatively distant site through a cooperative set of interactions.

The second fragment in CLIC1-E81M that exchanged deuterium significantly faster than in wtCLIC1 at pH 7 is Tyr^{69} – Phe^{83} (Figure 35). This region consists of s3-h4-s4 and includes the engineered mutation Glu81Met. Hence, the increased flexibility in this fragment is most likely directly connected to the removal of the Arg29 – Glu81 salt bridge. Interestingly, the level of deuterium exchange of Tyr^{69} – Phe^{83} in wtCLIC1 at pH 5.5 mimics that of Tyr^{69} – Phe^{83} in CLIC1-E81M at pH 7.0 (Figure 35B inset). Hence, the Arg29 – Glu81 salt-bridge, a unique CLIC interaction (Figures 10 and 11), is indeed a critical part of the mechanism responsible for the transformation of CLIC1 from soluble to membrane-bound form. It seems that the weakening of this salt-bridge, by a drop in pH, is needed to prime native CLIC1 for the conversion. It must be noted that the other region destabilized by a decrease in pH, s1- h1 (Phe^{11} – Met^{32}) is not affected by the Glu81Met mutation (Figure 35A). It is possible that the additional hydrophobic contacts introduced by the side chain of methionine compensated for the full effects of the removal of Arg29 – Glu81 salt bridge on the s1-h1 region. This observation is supported by the fact that the engineered mutation results in one region of native CLIC1-E81M, namely Phe^{203} – Ala^{213} (forms part of h7), becoming less flexible than the corresponding fragment in wtCLIC1 at pH 7 (Figure 35). The exact mechanism that induced this stabilization is unclear since this peptide does not form part of the domain interface and is 20 – 30 Å away from Glu81. Similarly to Val^{87} – Leu^{96}, the effects of the mutation are most likely conveyed through a network of cooperative interactions from position 81 to amino acids found in h7. The increased protection of h7 could be due to the additive stabilizing effects conducted through hydrophobic interactions between Met81 (h3) – Leu175 (h5) and Leu181 (h5) – Tyr209 (h7).

4.2.2.2 CLIC1-M32A

Only one region is significantly different, in terms of structural dynamics, between native conformations of CLIC1-M32A and wtCLIC1 at pH 7.0 (Figure 32). The fragment Ala^{220} – Glu^{228} extends from h8 to h9 with part of it packing against h1. At pH 7.0, the number of solvent-accessible/unprotected amide hydrogens of h8-h9 is higher in the mutant protein, while the amplitude of the fluctuation of h8-h9 is higher in wtCLIC1. This is indicated by increased deuterium localization during intermediate and slow-exchange times, but decreased deuterium incorporation during fast-exchange in Ala^{220} – Glu^{228} of CLIC1-M32A as compared to the corresponding

peptide in wtCLIC1 (Figure 32B inset). In the wild-type protein, Ala^{220} – Glu^{228} contacts directly Met32 through interaction with Thr222. The cavity forming Met32Ala mutation results in the loss of hydrophobic and charged contacts. This probably results in the uncoupling of the C-terminal h8 – h9 region from the N-terminal h1 as shown by the higher number of solvent-exposed amide-hydrogens in Ala^{220} – Glu^{228} of CLIC1.M32A. The loss in packing contacts between h8 – h9 and h1 is in most likelihood the reason behind the slightly more solvent-exposed environment of Trp35 in CLIC1-M32A as suggested by fluorescence spectroscopy (sections **3.7.1** and **4.2.1.1**).

It is surprising that h1 (residues 25 – 36), or at least part of it, is not affected by the engineered Met32Ala mutation. However, the results shown in Figure 30 may be slightly misleading since only one peptide, which covers that region, could be compared between CLIC1-M32A and wtCLIC1. The lack of equivalent fragments is a common phenomenon when contrasting wild-type with mutant proteins, since an engineered mutation will modify the digestion pattern of pepsin resulting in varying fragments in the region of the mutation. Phe^{11} – Met^{32} is a relatively long peptide made up of 21 amino acids. As such it may not represent changes in deuterium exchanges in the immediate environment of Met32. Nevertheless, the lack of significant difference in exchange between Phe^{11} – Met^{32} of CLIC1-M32A and wtCLIC1 at pH 7.0 confirms that any changes induced by the Met32Ala mutation are only localized to the direct location of Met32. In addition, the cavity-forming mutation does not disrupt any cooperative set of interactions since, in the native state, amino acids that are more than 4 Å apart from Met32 are not affected in terms of an increase or decrease of solvent-exposure.

Decrease in the pH of the environment is also shown to induce local changes in the structural dynamics wtCLIC1 (Figure 30A; Nathaniel, 2006). Only two fragments encompassing s1- h1 and s2-h2-s3-s4, which form part of the proposed N-terminal TMD (Harrop *et al*., 2001), spanning residues 24 – 46 (see section **1.5.5**), became more flexible/unstructured as the pH is lowered from 7.0 to 5.5 (Nathaniel, 2006). In the case of CLIC1-M32A at pH 7.0, the above-mentioned regions behave similarly to the corresponding peptides in wtCLIC1 at pH 7.0 (Figure 32). Hence, the Met32Ala mutation does not mimic the effects of pH on the structural dynamics of native

CLIC1. However, having said that, the conserved lock-and-key motif still plays a critical role in the stability, folding and function of CLIC1 as shown by the equilibrium unfolding studies discussed in section **4.3.2**. The small local changes induced by Met32Ala, in the case of CLIC1-M32A, and drop in pH from 7.0 to 5.5, in the case of wtCLIC1, suggest that the native conformation of CLIC1 is only primed for membrane insertion instead of being converted to a molten-globule conformation thought to be a prerequisite for membrane insertion. This view is supported by studies of apomyoglobin (aMb) whose membrane insertion was shown to be a two step pH-dependent process where a molten-globule state, required for membrane insertion, was only observed at pH 4.0 (Man *et al*., 2007). At pH 5.5, first step, aMb weakly binds to large unilamellar vesicles (LUVs) but remains native-like with no changes in secondary and tertiary structure detected by circular dichroism and fluorescence spectroscopy. At pH 4.0, second step, aMb forms a molten-globule state that binds tighter to and penetrates the LUVs with the binding being related to the stabilization of the partially folded conformation (Man *et al*., 2007).

4.3 Changes in the stability and unfolding of CLIC1 induced by Glu81Met and Met32Ala substitutions

4.3.1 Hydrophobic interactions compensate for removal of Glu81 – Met81 salt-bridges

The initial indications were that the Glu81Met mutation destabilizes CLIC1. When fusion GST-CLIC1-E81M is expressed at 20 °C, instead of 37 °C, more soluble protein is obtained (Figure 23). This is confirmed by the equilibrium unfolding studies of CLIC1-E81M performed by Legg (PhD, unpublished data). However, the extent of the destabilization is surprising in that the ΔG_{H2O} and C_m values obtained for CLIC1-E81M are only marginally smaller than those obtained for wtCLIC1 under identical conditions. As discussed in **4.1.1**, the salt-bridges formed via Glu81 – Arg29 in most likelihood contribute significantly toward the stability of CLIC1. Thus, one expects the removal of these buried charged interactions to be manifest in significantly larger destabilization of CLIC1-E81M than that suggested by the unfolding transitions. It is most likely that the true effects of the salt-bridge removal are compensated by newly introduced stabilizing contacts. The ΔG_{H2O} of proteins represents the net sum of stabilizing and destabilizing forces such as favourable energy due to enthalpic

interactions and unfavourable energy due to decrease in entropy as the protein folds. In the case of CLIC1-E81M the destabilizing effects induced by the removal of the salt-bridges between Glu81 – Arg29 will be counteracted by stabilizing hydrophobic interactions introduced by the Glu81 → Met81 substitution. The fact that the overall effect of the Glu81 → Met81 substitution is a net decrease in CLIC1 stability indicates that the salt-bridges destabilizing energy is larger than the stabilizing hydrophobic contacts introduced by the non-polar side chain of methionine. This statement will only be accurate if the engineered mutation does not remove any additional contacts apart from the 2 salt-bridges. This is possible for CLIC1-E81M since the modelled orientation of the side chain of Met81 shows that the two hydrogen bonds formed between Glu81 – Arg29 in wtCLIC1 are preserved in CLIC1-E81M. Waldburger and co-workers (1995) removed a buried salt-bridge triad from arc repressor, the most stabilizing type of salt-bridge interaction (Kumar and Nussinov, 1999), by introducing a well packed hydrophobic amino acid. The result was that the protein was stabilized by ~ 4.5 kcal/mol. Since the overall effect of Glu81Met is destabilizing, the charged interactions formed between the polar side-chain groups of Arg29 and Glu81 contribute even more than initially thought from the visual inspection of the salt-bridge characteristics (see section **4.1.1**).

In spite of the stabilization induced by the replacement of Glu81 with methionine, the CD and fluorescence monitored unfolding transitions of CLIC1-E81M are not indistinguishable (Legg PhD, unpublished data). This suggests that the equilibrium unfolding of CLIC1-E81M at pH 7.0 is not true two-state as is the case with wtCLIC1 at pH 7.0 (McIntyre, 2006). This is interesting because studies involving wtCLIC1 and the decrease in pH from 7.0 to 5.5 result in the accumulation of a molten globule-like state that is thought to represent the membrane-competent species of CLIC1 (McIntyre, 2006). This conformational state is formed by the partial unfolding of the thioredoxin domain, a process most likely initiated by the weakening of the domain interface network involving amongst others the charged contacts formed by Glu81 – Arg29. Therefore, the accumulation of a CLIC1-E81M equilibrium intermediate at pH 7.0 corroborates the involvement of the Glu81 – Arg29 in the transition of CLIC1 from soluble to membrane-competent form.

4.3.2 Met32Ala mutation destabilizes CLIC1 resulting in 3-state unfolding at pH 7

The over-expression of GST-CLIC1-M32A at 37 °C resulted in the formation of insoluble protein aggregates (Figure 22). This is one of the first indications that the protein is destabilized by the Met32Ala mutation. Structural alterations at the domain interface probably result in slower domain association upon protein expression. Thus, the exposure of hydrophobic surfaces is prolonged promoting non-specific interactions between different polypeptide chains and so forming inclusion bodies. Since aggregation is a 2nd order reaction, the rate of inclusion body formation is directly proportional to protein concentration (Goldberg *et al.,* 1999). Thus, lowering the incubation temperature to 20 °C reduces the amount of over-expressed GST-CLIC1-M32A, which in turn reduces aggregation while formation of the native state prevails (Figure 22). Therefore, a temperature decrease during protein over-expression will reduce the rate of aggregate formation (± 3 fold per 10 °C) and possibly prevent inclusion body formation (Atkins, 1998).

The lowered stability of CLIC1-M32A is confirmed via equilibrium unfolding studies. The ΔG_{H2O} and *m*-value derived from the far-UV-CD-monitored transitions of CLIC1-M32A at pH 7.0 are significantly lower than the corresponding values obtained for wtCLIC1 at pH 7.0. In addition, the transition midpoints of CLIC1-M32A unfolding curves at pH 7.0 are shifted to lower urea concentrations as compared to those obtained for wtCLIC1 (Figure 39 and Table 2). The origin of CLIC1-M32A destabilization can be attributed to two factors. Firstly, the Met32Ala mutation results in the removal of the $>CH_2$-S-CH_3 group of methionine that leads to a loss of hydrophobic contacts and a decrease in ΔG_{H2O} of approximately ± 1.1 kcal/mol per $>CH_2$ group (Pace, 2001; Kellis, 1988). This loss in stability is probably greater due to the cooperative contribution of hot spot residues, such as Met32, toward protein stability (section **3.1.1** and Figures 8 and 9). Secondly, the Met32Ala mutation is a cavity forming substitution that results in a looser packing at the domain interface. This is supported by two facts. First, fluorescence emission spectra show Trp35 to be slightly more solvent-exposed in CLIC1-M32A as compared to wtCLIC1 (section **3.7.1**). Secondly, the h8-h9 region, which packs onto Met32 and the rest of h1, is more flexible in the mutant protein as detected by continuous pulse-labelling DXMS

(Figure 32). Loose packing results in the loss of packing interactions in particular short-range van der Waals contacts (Ratnaparkhi and Varadarajan, 2000). Hence, the folded state of CLIC1-M32A is destabilized due to an unfavourable enthalpic contribution toward ΔG. It must be mentioned that the looser packing at the domain interface of the mutant results in favourable entropic contribution toward ΔG. However, the entropic energy gain is relatively small because only one fragment is shown to be more flexible in CLIC1-M32A as compared to wtCLIC1 at pH 7.0 (Figure 32). This is counteracted by a much a larger energy loss due to the generation of a cavity and the removal of inter-domain hydrophobic as well as van der Waals interactions.

The equilibrium unfolding of wtCLIC1 at pH 7.0/20 °C is a two-state process where the only species detected are the native and unfolded conformations (McIntyre, 2006). In the case of CLIC1-M32A unfolding under identical conditions is a three-state event (Figure 37). A number of factors are in support of a multi-state process:

(1) the fluorescence and CD-monitored transitions are non-coincident and hence the ΔG_{H2O} and *m*-value obtained by the two probes are different (Figure 39, Table 4)

(2) the fluorescence unfolding curves are biphasic (Figure 39).

(3) Raleigh scatter indicates lack of protein aggregates throughout the unfolding transition of CLIC1-M32A (Figure 39).

(4) deuterium pulse-labelling DXMS used to follow changes in the CLIC1-M32A structure with increasing urea concentration, detects a molten globule like intermediate state (Figures 44 and 46).

(5) in retrospect, the very low ΔG_{H2O} and *m*-value obtained from the CD-monitored unfolding of CLIC1-M32A are attributed to the fact that the presence of significant amounts of equilibrium intermediate/s will always translate into the underestimation of these parameters when they are obtained through a two-state analysis of a three-state unfolding process (Soulages, 1998).

The lack of equilibrium intermediate/s in the two-state unfolding of wild-type CLIC1 at pH 7.0 indicates a highly cooperative process. On the other hand, the three-state unfolding of CLIC-M32A and the accumulation of equilibrium intermediate/s

signifies a less cooperative unfolding. This can be attributed to the loss in packing interactions at the domain interface as well as the loss of hydrophobic driving force caused by the replacement of Met32 with alanine.

Intriguingly, the equilibrium unfolding of wtCLIC1 at pH 5.5 is also a three-state process. Molten globule formation, as in the case of wtCLIC1 as well as other amphitropic proteins, is thought to be a vital prerequisite for membrane insertion (McIntyre, 2006; Kleinschmidt and Tamm, 1996). Therefore, it is important to establish whether the intermediate species induced by the Met32Ala mutation is structurally similar to the molten globule-like state of CLIC1 induced by low pH. If so, we can deduce that the inter-domain lock-and-key motif forms part of the mechanism that drives CLIC1 from a soluble to a membrane-competent form.

4.3.3 The molten-globule states of wtCLIC1 and CLIC1-M32A

The intermediate state detected at low denaturant concentration in the case of wtCLIC1 at pH 5.5 is more structured and stable than the molten-globule state detected in the unfolding transitions of CLIC1-M32A at pH 7.0. In terms of secondary structural content, wtCLIC1 at pH 5.5 in 3.8 M^{Σ} urea shows a ~ 25 % decrease in its far-UV CD signal as compared to the spectrum of wtCLIC1 at pH 5.5 in the absence of denaturant (McIntyre, 2006). On the other hand, the far-UV-CD spectrum of CLIC1-M32A at pH 7.0 in 3.4 M^{Σ} urea displays ~ 50 % loss in ellipticity relative to the spectrum of native CLIC1-M32A at pH 7.0 (Figure 41). Therefore, the molten-globule state of CLIC1-M32A has lost approximately double the helical content as compared to the molten-globule state of wtCLIC1. This is confirmed by pulse-labelling DXMS where the conversion from native to intermediate state involves ~ 31 and 41 % loss in structural content, i.e. helices/sheets/coils, in the case of wtCLIC1 and CLIC1-M32A, respectively. In addition, ANS binding studies show that the wild-type intermediate has fewer exposed surfaces than the CLIC1-M32A intermediate since its bound-ANS signal is approximately 2.5 times lower than in the case of the mutant. Furthermore, the $\lambda_{em\ max}$ of ANS bound to wtCLIC1 is blue shifted compared to the $\lambda_{em\ max}$ of ANS bound to CLIC1-M32A (460 nm and 470 nm, respectively). This means that the exposed surfaces are more hydrophobic in the wild-type protein

$^{\Sigma}$: *Denaturant concentration where intermediate was most prominent*

suggesting a more structured intermediate. In summary, the molten-globule state of wtCLIC1 in comparison to the CLIC1-M32A intermediate:

(1) is more stable since it exists at higher urea concentration.
(2) has preserved more of its secondary and tertiary structural content.
(3) has fewer exposed hydrophobic surfaces.
(4) is more structured.

Similarly to wtCLIC1, the fluorescence spectrum of the molten-globule state of CLIC1-M32A exhibits a blue shift in its $\lambda_{em\ max}$ relative to the spectrum of native CLIC1-M32A (McIntyre, 2006; section **3.7.4.1** and Figure 40). This suggests that Trp35 is more buried in the intermediate state than in the native state. In the case of wtCLIC1, the burial of tryptophan was attributed to the formation of an oligomeric structure (McIntyre, 2006). If this is true than pulse-labelling DXMS experiments, performed in this study, should show an increase in the number of residues protected from deuterium exchange as an oligomeric state forms at low denaturant concentrations. However, no such evidence could be found both for wtCLIC1 and CLIC1-M32A (Figure 46 B and C, Appendix **Table H**). Hence, the burial of Trp35 is attributed to local structural changes as the protein unfolds instead of global rearrangement and/or complex formation.

Which regions of wtCLIC1 and CLIC1-M32A unfold during the conversion from a native state to a molten-globule state (N → I)? This is a difficult question to answer since most of the probes used in this study provide global information and as such they do not present direct evidence on local structural changes. However, there are a number of clues to which structures participate in the N → I conversion. Continuous labelling DXMS indicates that the C-terminal domain of wtCLIC1 and CLIC1-M32A is more stable than the N-terminal domain (Nathaniel, 2006; Figure 31). Hence, the first structures to be perturbed during the N → I conversion will be the helices/sheets/coils in the N-domain. Far-UV-CD showed that 25 % and 50 % helical content is lost in the N → I conversion for wtCLIC1 and CLIC1-M32A, respectively (McIntyre, 2006; Figure 41). The N-terminal domain of CLIC1 contains 3 helices that make up approximately a quarter of the total helical content. Therefore, in the case of wtCLIC1 the N → I conversion possibly involves the re-structuring of the N-domain

helices. This is probably also true for CLIC1-M32A since continuous labelling DXMS showed that the Met32Ala mutation did not affect the N-domain of CLIC1 (Figure 32). If so, approximately another 25 % helical structures unfold upon N → I conversion of CLIC1-M32A. Continuous labelling DXMS shows that h8 – h9 region becomes more flexible as a result of the Met32Ala mutation (Figure 32). The destabilization of this region in the N state, due to removal of tertiary contacts, implies that h8, h9 and possibly h7 will be one of the first structures to be disturbed as the protein starts unfolding. These 3 helices form roughly 25 % of the total helical content. Therefore, it seems that in addition to three N-domain helices (h1, h2, h3) three C-domain helices (h7, h8, h9), in total half the helical content of CLIC1, unfold upon conversion of CLIC1-M32A from N → I. This is supported by the fact that the C_m value of the far-UV-CD unfolding curve coincides with the first phase of the fluorescence unfolding curves of CLIC1-M32A (Figure 39). Hence, the equilibrium intermediate has approximately half of the secondary structural content of the native state. A third line of evidence is provided by pulse-labelling DXMS experiments where the number of incorporated deuteriums is directly related to the number of residues that become unstructured as the protein unfolds. The transition from N to I of wtCLIC1 results in approximately 30 % increase in unprotected residues (Table 4). On the other hand, upon conversion of CLIC1-M32A from N → I, 40 % of the amide hydrogens became unprotected (Table 4). The ~ 10 % difference in structures that unfold between the I state of wtCLIC1 and the I state of CLIC1-M32A can be accounted by destabilization of the region encompassing helices 7 – 9, which in total forms roughly 10 % of the total structural content of CLIC1. Hence, apart from the additional h7 – h9 region, the same structures unfold during the N → I conversion in wtCLIC1 and CLIC1-M32A. In must be noted that the above proposal is oversimplified in that it does not take into account any structures that may refold in the conversion and hence generate new helices and/or sheets. This is very possible especially in the case of the proposed TMD-s2 that has a high helical propensity which upon N-domain unfolding may convert to a helix (Nathaniel, 2006). However, as a rough guideline, this rearrangement model indicates that the Met32Ala mutation and acidic pH affect the conversion of CLIC1 from N to I in the same way. And so, the unfolding of the N-terminal domain of wtCLIC1 probably involves the uncoupling of the conserved inter-domain lock-and-key motif. In order to confirm this future work should make use of continuous- as well as pulse-labelling DXMS followed by

pepsin digestion to compare the native and intermediate states of wtCLIC1 and CLIC1-M32A. As performed in this study on native CLIC1-M32A and CLIC1-E81M (sections 2.2.12.1 and 3.6), pepsin digestion will provide more local information through the generation of peptide fragments. Hence, local structural changes will be detected during the conversion of CLIC1 from native to intermediate state.

According to data obtained from pulse-labelling DXMS, the I state is stabilized at low denaturant concentrations but present throughout the unfolding transition of wtCLIC1 at pH 5.5 and CLIC1-M32A at pH 7.0 (Figure 46). On the other hand, ANS binding studies indicate that no binding of the hydrophobic dye to these proteins occurs in the pre- and post-unfolding transition (McIntyre, 2006; Figure 42). A possible explanation for this discrepancy may be that the difference in protein concentration used in the two techniques is approximately 50 fold and so the I state is probably undetected at low and high denaturant concentrations by ANS fluorescence due to the low protein concentration used. If this is the case the detection of the I state in the absence of denaturant indicates that the drop in pH from 7.0 to 5.5 and the Met32Ala substitution have lowered the energy barrier between the N and I conformations so that the I state becomes more populated and hence detected via pulse-labelling DXMS. The decrease in energy gap between N and I is due to destabilization of the N conformation and/or stabilization of the I state. As the conditions shift toward the I conformation (i.e. mild denaturing conditions, drop in pH, reducing environment, increase in temperature) it becomes progressively more populated until a point where it is the predominant state [i.e native wtCLIC1 at pH 5.5/37 °C has exposed hydrophobic surfaces and looser packing with ~ 16 % less helical content than native wtCLIC1 at pH 7.0/20 °C (McIntyre, 2006)]. Some mutants such a CLIC1-M32A mimic the effects of the environment by destabilizing the N state and/or stabilizing the I conformation. In fact the replacement of Lys37, which is part of the h1 N-capping motif and forms contacts between h1 and h3 through a salt-bridge with Glu85, with threonine resulted in I being the pre-dominant state (McIntyre, 2006). This is an indication that interactions such as the inter-domain lock-and-key and the h1 N-capping motifs form part of the mechanism responsible for the transition of CLIC1 from a soluble state to a membrane-competent state.

4.4 Possible mechanism for soluble to membrane transition of CLIC1

In view of all the information provided thus far, the following mechanism responsible for the conversion of CLIC1 from a soluble conformation to a molten-globule like state, that is possibly a prerequisite for membrane insertion, is proposed (Figure 48). Acidic pH, low dielectric constant and negative charge at the membrane surface, prime CLIC1 for membrane insertion by destabilizing charged interactions. Amongst those are the buried salt-bridges formed between Arg29 – Glu81 and Lys37 – Glu85. These high-energy contacts link h1 and h3 that in turn form the N-terminal domain interface of CLIC1. The destabilization of interactions linking h1 and h3 results in increased flexibility in this region that is translated to the rest of the N-terminal domain and across the domain interface. This in turn allows the uncoupling of the inter-domain lock-and-key motif to occur with the concomitant dissociation of the N- and C-terminal domains. The loss of domain interface contacts allows the thioredoxin domain and specifically s1 – h1 to re-structure and form the TMD while the more stable C-terminal domain remains largely intact. At this stage either monomeric CLIC1 inserts into the membrane and associates with other CLIC1 monomers forming a functional channel or it first forms an oligomeric complex followed by membrane insertion.

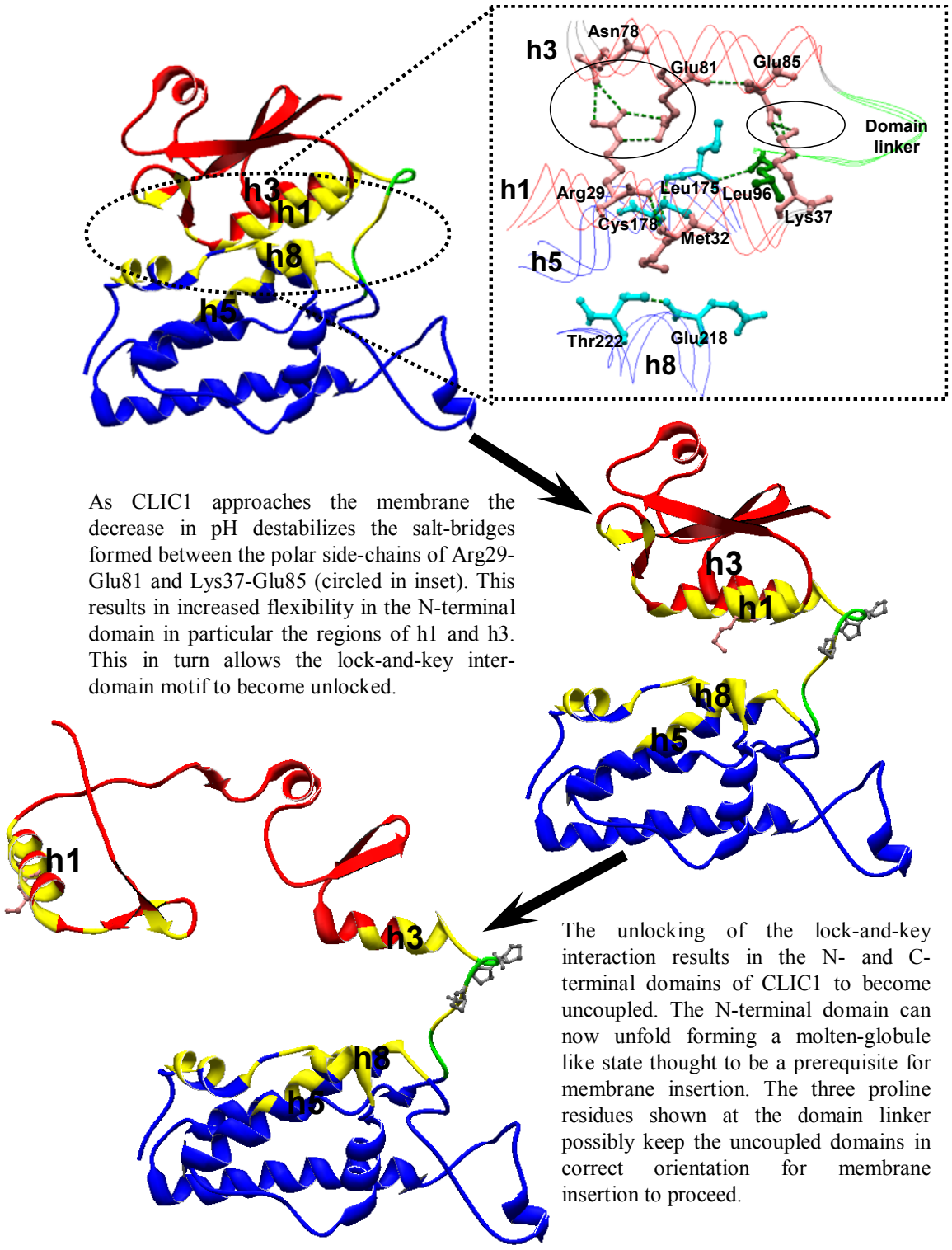


Figure 48: Possible domain interface involvement in soluble/membrane transition of CLIC1

Various structures of CLIC1 from soluble to membrane-competent form are depicted. The N-terminal domain is shown in red and C-terminal domain in blue. Domain interface residues are coloured yellow. The side chain of Met32 is shown in pink. The inset shows details of the domain interface network of interactions responsible for the conversion of CLIC1 from soluble to membrane-competent form.

CHAPTER 5

REFRENCES

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

Argos, P. (1988). An investigation of protein subunit and domain interfaces. *Protein Eng.* **2**, 101-113.

Ashley, H. R. (2003). Challenging accepted ion channel biology: p64 and the CLIC family putative intracellular anion channel proteins. *Mol. Memb. Biol.* **20,** 1 – 11.

Atkins, P.W. (1998). *Physical Chemistry*, in *Physical Chemistry*. Oxford University press: Oxford New York p.776.

Bai, Y., Milne, S. J. and Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 75 – 86.

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

Berry, L. K. and Hobert, O. (2006). Mapping functional domains of Cholride Intracellular Channel (CLIC) proteins *in vivo*. *J. Mol. Biol.* **359**, 1316 – 1333.

Berryman, M. A., Bruno, J., Price, J. and Edwards, J. C. (2004). CLIC-5A functions as a chloride channel *in vitro* and associates with the cortical actin cytoskeleton *in vitro* and *in vivo*. *J. Biol. Chem.* **279**, 34794 – 34801.

Baskakov, V. I. and Bolen, W. D. (1998). Monitoring the sizes of denatured ensembles of staphylococcal nuclease proteins: implications regarding *m* values, intermediates and thermodynamics. *Biochemistry.* **37**, 18010 – 18017.

Bogan, A. A. and Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1-9.

Brady, P. G. and Sharp, A. K. (1997). Entropy in protein folding and in protein-protein interactions. *Curr. Opin. Struct. Biol.* **7**, 215-221.

Brooks, L. C., Gruebelle, M., Onuchic, N. J., and Wolynes, G.P. (1998). Chemical physics of protein folding. *Proc. Natl. Acad. Sci. U.S.A* **95**, 11037-11038.

Busenlehner, S. L. and Armstrong, N. R. (2004). Insights into enzyme structure and dynamics elucidated by amide H/D exchange mass spectrometry. *Arch. Biochem. Biophys.* **433**, 34 – 46.

Bychcova, V. E., Dujsekina, A. E., Kleni, S. I., Tiktopulo, E. I., Uversky, V. N. and Ptitsyn, O. B. (1996). Molten globule-like state of cytochrome under conditions simulating those near the membrane surface. *Biochemistry.* **35**, 6058 – 6063.

Caccace, M. G., Landan, E. M., and Ramsten, J. J. (1997). The Hofmeister series: salt and solvent effects on interfacial phenomena. *Q. Rev. Biophys.* **30**, 241-277.

Chotia, C., and Janin, J. (1975). Principles of protein-protein interactions. *Nature* **256**, 705-708.

Chenal, A., Savarin, P., Nizard, P., Guillain, F., Gillet, D. and Forge, V. (2002). Membrane protein insertion regulated by bringing electrostatic and hydrophobic interactions into play. A case study for the translocation domain of diphtheria toxin. *J. Biol. Chem.* **277**, 43425 – 43432.

Chenal, A., Vernier, G., Savarin, P., Bushmarina, N. A., Gèze, A., Guillain, F., Gillet, D. and Forge, V. (2005). Conformational states and thermodynamic of α-lactalbumin bound to membranes: a case study for the effects of pH, calcium, lipid membrane curvature and charge. *J. Mol. Biol.* **349**, 890 – 905.

Chrunyk, B. A., Perregaux, D. E., Gabel, C. A., Geoghegan, K. F., and Pandir, J. (2000). Identification, characterization, and crystal structure of the omega class glutathione transferases. *J. Biol. Chem.* **275**, 24798-24806

Chuang, J. Z., Milner, T. A., Zhu, M. and Sung, C. H. (1999). A 29 kDa intracellular chloride channel p64H1 is associated with large dense core vesicles in rat hippocampal neurons. *J. Neurosci.* **19**, 2919 – 2928.

Chung, C.T., Niemela, S.L. and Miller, R.H. (1989) One-step preparation of competent Escherichia coli: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. U.S.A* **86**, 2172-2175.

Clarke, J. and Itzhaki, L. S. (1998). Hydrogen exchange and protein folding. *Curr. Opin. Struct. Biol.* **8**, 112 – 118.

Codreanu, G. S., Thompson, C. L., Hackey, L. D., Dirr, W. H. and Armstrong, N. R. (2005). *Biochemistry.* **44,** 10605-10612.

Collins, K. D., and Washabaugh, M, W. (1985). The Hofmeister effect and the behaviour of water at interfaces. *Q. Rev. Biophys.* **18**, 323-422.

Creighton, E. T. (1990). Protein folding. *Biochem. J.* **270,** 1-16.

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

Demmers, A. A. J., Rijkers, S. T. D., Haverkamp, J., Killian, A. J. and Heck, R. J. A. (2001). Factors affecting gas phase deuterium scrambling in peptide ions and their implication for protein structure determination. *J. Am. Chem. Soc.* **124**, 1191 – 1198.

Deng, Z. and Smith, L. D. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **2**, 522 – 531.

Deng, Y. and Smith, L. D. (1999). Hydrogen exchange demonstrates three domains in aldolase unfold sequentially. *J. Mol. Biol.* **294**, 247 – 258.

Deng, Y., Zhang, Z., Smith, L. D. (1999). Comparison of continuous and pulse labelling amide hydrogen exchange/mass spectrometry for studies of protein dynamics. *J. Am. Soc. Mass Spectrom.* **10**, 675 – 684.

Dill, A. K. (1990). Dominant forces in protein folding. *Biochemistry* **29**, 7133-7151.

Dill, K. A. and Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10-19.

Dill, A. K., Trusket, M.T., Vlachy, V., and Mribar-Lee. B. (2005). Modelling water, the hydrophobic effect and ion solubility. *Annu. Rev. Biophys. Struct.* **34**, 173-199.

Dirr, H., Reinemer, P., and Huber, R. (1994). X – ray structures of cytosolic glutathione S – transferases. *Eur. J. Biochem.* **220,** 645–661.

Dirr, H. W. (2001). Folding and assembly of glutathione transferases. *Chem.-Biol. Interact.* **333,** 19-23.

Duncan, R. R., Westwood, P. K., Boyde, 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.

Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. and MacKinnon, R. (2002). X-ray structure of CLC chloride channel at 3 Å reveals the molecular basis for anion selectivity. *Nature*. **415**, 287 – 294.

Edwards, J.C. (1999). A novel p64-related Cl channel: subcellular distribution and nephron segment-specific expression. *Am J. Physiol*. **276**, 398-408.

Engelhard, M., and Evans, A.P. (1995). Kinetics of interaction of partially folded proteins with a hydrophobic dye: Evidence that molten globule character is maximal in early folding intermediates. *Protein Sci.* **4**, 1553-1562.

Engelhard, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol.* **29**, 213 – 238.

Erhardt, J. and Dirr, H. (1995). Native dimer stabilizes the subunit tertiary structure of porcine class pi glutathione S-transferase. *Eur. J. Biochem.* **230,** 614-620.

Ferguson, L. P., Pan, J., Wilson, J. D., Dempsey, B., Lajoie, J., Shilton, B. and Konermann, L. (2006). Hydrogen/deuterium scrambling during quadropole time of flight MS/MS analysis of zinc-binding protein domain. *Anal. Chem.* **79**, 153 – 160.

Fernandez-Salas, E., Sagar, M., Cheng, C., Yuspa, S., and Weinber, W. C. H. (1999). p53 and tumour necrosis factor alpha regulate expression of a mitochondrial chloride channel protein. *J. Biol. Chem*. **274**, 36448-36497.

Frangioni, V.J., and Neel, G.B. (1993). Solubilization and purification of enzymatically active Glutathione S-Transferase (pGEX) fusion proteins. *Anal. Biochem.* **210**, 179-187.

Fraczkiewicz, R. and Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface area and their gradient for macromolecules. *J. Comp. Chem*. **19**, 319.

Friedli, M., Guipponi, M., Bertrand, S., Bertrand, D., Neerman-Arbez, M., Scott, H. S. and Antonarakis, S. E. (2003). Identification of novel member of the CLIC family, CLIC6, mapping to 21q.22.12. *Gene*. **320**, 31 – 40.

Griko, Y. V., Privalov, P. L., Venyamizov, S. Y. and Kutyshenko, V. P. (1988). Thermodynamic structure of the apomyoglobin structure. *J. Mol. Biol.* **202,** 127-138.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook, John M. Walker (ed) Humana Press. pp. 571-607.

Goldberg, M.E., R. Rudolph, and R. Jaenicke, *A.* (1991). Kinetic study of the competition between renaturation and aggregation during the refolding of denatured-reduced egg white lysozyme. *Biochemistry* **30**, 2790-2797.

Guex, N. and Peitsch, M. C. (1997). Swiss-Model and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723.

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

Hendsch, S. Z. and Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211-226.

Hilser, J. V., Dowdy, D., Terrence, G. O., and Freire, E. (1998). The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proc. Natl. Acad. Sci. U.S.A* **95**, 9903-9908.

Hoofnagle, N. A., Resing, A. K. and Ahn, G. N. (2003). Protein analysis by hydrogen exchange mass spectrometry. *Annu. Rev. Biomol. Struct.* **32**, 1 – 25.

Honing, B. R., and Levinthal, C. (1976). Conformational flexibility and protein folding: rigid structural fragments connected by flexible joints in substlisin. *Proc. Natl. Acad. Sci. U.S.A* **73**, 1974-1978

Hornby, T. A. J., Luo, J-K., Stevens, M. J., Wallace, A. S., Kaplan, W., Armstrong, N. R., and Dirr, W. H. (2000). Equilibrium folding of dimeric class μ glutathione transferases involves a stable monomeric intermediate. *Biochemistry* **39**, 12336-12344.

Hu, Z., Ma, B., Wolfson, H., and Nussinov, R. (2000). Conservation of polar residues as hot spots at protein interfaces. *Proteins: Struct. Funct. Genet.* **39**, 331-342.

Hubbard, S.J., (1992). Analysis of protein-protein molecular recognition. University of London, PhD. Thesis.

Jaenicke, R. (1999). Stability and folding of domain proteins. *Prog. Biophys. molec. Biol.* **71**, 155-241.

Jentsch, T. J., Stein, V., Weineirch, F. and Zdebik, A. A. (2001). Molecular structure and physiological function of chloride channels. *Physiol. Rev.* **82**, 503 – 568.

Jewett, I. A., Pande, S. V., and Plaxco. W. K. (2003). Cooperativity, smooth energy landscapes and the origins of topology-dependant protein folding rates. *J. Mol. Biol.* **326**, 247-253.

Jones, S. and Thornton, M.J. (1995). Protein-protein interfaces: a review of protein dimmer structures. *Prog. Biophys. molec. Biol.* **63**, 31-65.

Jones, S. and Thornton, M.J. (1996). Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A* **93**. 13-20

Jones, S., Marin, A., and Thornton, M. J. (2000). Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. *Protein Eng.* **13**, 77-82

Kaltashov, A. I. and Eyles, J. S. (2005). Mass spectrometry in biophysics. Conformation and dynamics of biomolecules. pp 96 – 97, John Wiley & Sons, Inc., Hoboken, New Jersey.

Kaltashov, A. I. and Eyles, J. S. (2005). Mass spectrometry in biophysics: conformation and dynamics of molecules. pp 210 – 212, John Wiley & Sons, Inc., Hoboken, New Jersey.

Kaplan, W., Husler, P., Erhardt, J., Sluis-Cremer, N. and Dirr, W. H. (1997). Conformational stability of pGEX-expressed *Schistosoma japonicium* glutathione S-transferase. *Protein Sci.* **6,** 399-406.

Kellis, T. J., Nyberg, K., Šali, D., and Fersht, R. A. (1988). Contribution of hydrophobic interactions to protein stability. *Nature* **333**, 784-786.

Keskin, O., Ma, B., and Nussinov, R. (2005). Hot-regions in protein-protein interactions: the organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* **345**, 1281-1294.

Kleinschmidt, J. H. and Tamm, L. K. (1996). Folding intermediates of a β-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism. *Biochemistry.* **35**, 12993 – 13000.

Krishna, M. G. M., Hoang, L., Ling, Y and Englander, W. S. (2004). Hydrogen exchange methods to study protein folding. *Methods*. **34,** 51 – 64.

Kumar, S. and Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **293**, 1241-1255.

Laemmli, K.U. (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227,** 680–685

Ladner, J. E., Parsons, F. J., Rife, L. C., Gilliland, L. G., and Armstring, N. R. (2004). Parallel Evolutionary Pathways for Glutathione Transferases: Structure and Mechanism of the Mitochondrial Class Kappa Enzyme rGSTK1-1. *Biochemistry* **43**, 352-361.

Lakowicz, J. R. (1983) *Principles of fluorescence spectroscopy*. Plenum Press, New York.

Landry, D. W., Sullivan, S., Nicolaides, M., Redhead, C., Edelman, A., Field, A., Al-Awqati, Q., and Edwards, J. (1989). Purification and reconstitution of chloride channels from kidney and trachea. *Science* **244**, 1469-1472.

Landry, D., Sullivan, S., Nicolaides, M., Redhead, C., Edelman, A., Field, M., Al-Awqati, Q. and Edwards, J. (1993). Molecular cloning and characterization of p64, chloride channel protein from kidney microsomes. *J. Biol. Chem.* **268**, 14948 – 14955.

Laskowski, R.A. (1995). SURFNET: A program for visualizing molecular surfaces, cavities and intermolecular interactions. *J. Mol. Graph.*, **13**, 323-330.

Lawrence, C.M., and Colman, M.P. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946-950.

Lee B., and Richards, F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

Levinthal, C. (1968). Are there pathways for protein folding? *J. Chem. Phys.* **65**, 44-45.

Lins, L. and Brasseur, R. (1995). The hydrophobic effect in protein folding. *FASEB J.* **9,** 535-540.

Lins, L. and Brasseur, R. (1995). The hydrophobic effect in protein folding. *FASEB J.* **9,** 535-540.

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

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

Looman, A. C., Bodlaender, J., de Gruyter, M., Vogelaar, M. and Knippenberg, P. H. (1986). Secondary structure as primary determinant of the efficiency of ribosomal binding sites in *Eschericihia coli*. *Nucl. Acids Res.* **14,** 5481-5497.

Luo, J-K., Hornby, A. T. J., Wallace, A. W., Chien, J., Armstrong, N. R. and Dirr, W. H. (2002). Impact of domain interchange on conformational stability and equilibrium folding of chimeric class μ gluthione transferases. *Protein Sci.* **11,** 2208-2217.

Man, P., Montagner, C., Vernier, G., Dublet, B., Chenal, A., Forest, E. and Forge, V. (2007). Defining the interacting regions between apomyoglobin and lipid membrane by hydrogen/deuterium exchange coupled to mass spectrometry. *J. Mol. Biol.* **368**, 464 – 472.

Mach, H., Volkin, D. B., Burke, C. J. and Middaugh, C. R. (1995) Ultraviolet absorption spectroscopy in *Methods in Molecular Biology, Vol. 40: Protein stability and Folding: Theory and Practice* (Shirley, B. A. ed), pp 91-114, Humana Press Inc., Totowa, New Jersey

Martin, J. L. (1995). Thioredoxin - a fold for all reasons. *Structure* **3**, 245-250.

McIntyre, S. (2006). Effects of the environment on the conformational stability of the chloride intracellular channel protein CLIC1. PhD thesis unpublished results. University of Witwatersrand, Johannesburg.

Murzin, A. G. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol*. **247**, 536-540.

Nathaniel, C. (2006). Structural dynamics of the chloride intracellular channel protein CLIC1. PhD thesis unpublished results. University of Witwatersrand, Johannesburg.

Notradame, C., Higgins, G. D. and Heringa J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205-207.

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

Pace, C.N. (1986) Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves. *Methods Enzymol.* **131,** 266-280

Pace, C. N. (2001). Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry* **40,** 310-313.

Papworth, C., Bauer, J.C., Braman, J. and Wright, D.A. (1996) Site-directed mutagenesis using double-stranded plasmid DNA templates. *Strategies* **9**, 3-4.

Parker, W. M., Pattus, F., Tucker, A. D. and Tsernoglou, D. (1989). Structure of membrane pore-forming fragment of colicin. *Nature*. **337**, 93 – 96.

Parker, W. M. and Pattus, F. (1993). Rendering a membrane protein soluble in water: a common packing motif in bacterial toxins. *Trends in Biochem. Sci.* **18**, 391 – 395.

Parker, M. J., Spencer, J., Jackson, G. S., Burston, S. B., Hosszu, L. L. P., Craven, C. J., Waltho, J. P., and Clarke, A. R. (1996). Domain Behavior during the folding of a thermostable phosphoglycerate kinase. *Biochemistry* **35**, 15740-15752.

Parker, W. M. (2003). Cryptic clues as to how water-soluble protein toxins form pores in membranes. *Toxicon.* **42**, 1 – 6.

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

Preibisch, G., Ishihara, H., Tripier, D. and Leineweber, M. (1988). Unexpected translational mutation within the coding region of eukaryotic genes expressed in *Escherichia coli*. *Gene* **72,** 179-186.

Privalov, P.L. (1979). Stability of proteins: small globular proteins. *Adv. Protein Chem* **33**, 167-241.

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

Ratnaparkhi, S. G., and Varadrajan, R. (2000). Thermodynamics and structural studies of cavity formation in proteins suggests that loss of packing interactions rather than the hydrophobic effect dominates the observed energetics. *Biochemistry* **39**, 12365-12374.

Redfield, C., Smith, R. A. G. and Dobson, C. M. (1994). Structural characterization of a highly ordered 'molten globule' at low pH. *Nature Struct. Biol.* **1,** 23-29.

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

Richardson, J.S (1981). The anatomy and taxonomy of protein structure. *Adv. Prot. Chem*. **34**, 246-253.

Sayed, Y., Wallace, A. L. and Dirr, W. H. (2000). The hydrophobic lock-and-key intersubunit motif of glutathione transferase A1-1: implications for catalysis ligandin function and stability. *FEBS lett.* **465,** 169-172.

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

Schueler, O., and Margalit, H. (1995). Conservation of salt bridges in protein families. *J. Mol. Biol.* **248**, 125-135.

Shatsky, M., Nussinov, R., and Wolfson, J.H. (2004) A method for simultaneous alignment of multiple protein structures. *Proteins* **56,** 143-156.

Sheehan, D., Meade, G., Foley, M. N., and Dowd, A. C. (2001). Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem. J.* **360**, 1-16.

Siddiqui, A.S. and Barton., G. J. (1995). Continious and discontinious domains: ad algorithnm for the automatic generation of reliable domain definitions. *Protein Sci.* **4**, 872-884.

Singh, H. and Ashley, R. H. (2006). Redox regulation of CLIC1 by cysteine residues associated with the putative membrane channel pore. *Biophys. J.* **90**, 1628 – 1638.

Sining, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. I., Armstrong, R. N., Ji, X., Board, P. G., Olin, B., Mannevrik, B., and Jones, T. A. (1993). Structure Determination and Refinement of Human Alpha Class Glutathione Transferase A1-1, and a comparison with the Mu and Pi Class Enzymes. *J.Mol.Biol.* **232,** 192-212.

Soulages, L. J. (1998). Chemical denaturation: potential impact of undetected intermediates in the free energy of unfolding and *m*-values obtained from a two-state assumption. *Biophys. J.* **75**, 484-492.

Stevens, J. M., Hornby, J. A. T., Armstrong, R. N., and Dirr H. W. (1998). Class sigma glutathione transferase unfolds via a dimeric and a monomeric intermediate: impact of subunit interface on conformational stability in the superfamily. *Biochemistry* **37,** 15534-15541.

Stickle, F.D., Presta, G.L., Dill, A.K., and Rose, D.G. (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**, 1143-1159.

Stinger, D., and Dill A. K. (1990). Charge effects on folded and unfolded proteins. *Biochemistry* **29**, 1262-1271.

Stites, E.W. (1997). Protein-protein interactions: interface structure, binding thermodynamics, and mutational analysis. *Chem Rev.* **97**, 1223-1250.

Swindells, M.B. (1995). A procedure for detecting structural domains in proteins. *Protein Sci.* **4**, 103-112.

Tanford, C. (1968). Protein denaturation. *Adv. Prot. Chem.* **23,** 121-282.

Tessier, L. H., Sondermeyer, P., Faure, T., Dreyer, D., Benavente, A., Villeval, D., Courtney, M. and Lecocq, J. P. (1984). The influence of mRNA primary and secondary strucuture on human IFN-gamma gene expression in *Escherichia coli*. *Nucl. Acids Res.* **12,** 7663-7675.

Thuduppathy, R. G. and Hill, B. R. (2006). Acid destabilization of the solution conformation of Bcl-x_L does not drive its pH-dependent insertion into membranes. *Protein Sci.* **15**, 248 - 257.

Tobias, J. W., Shrader, T. E., Rocap, G. and Varshavsky, A. (1991). The N-rule in bacteria. *Science* **254,** 1374-1377.

Tonnin, R., Ferroni, A., Valenzuela, S. M., Warton, K., Campbell, T. J., Breit, S. N. and Mazzanti, M. (2000). Functional characterization of the NCC27 nuclear protein in stable transfected CHO-K1 cells. *FASEB J.* **14**, 1171 – 1178.

Tsai, C-J., Lin, L. S., Wolfson, J. H., and Nussinov, R. (1996). Protein-protein interfaces architectures and intercations in protein-protein interfaces and in protein cores. Their similarities and differences. *Crit. Rev. in Biochem. and Mol. Biol.* **31,** 127-152.

Tulk, B.M. and Edwards J. C. (1998). NCC27, a homologue of intracellular chloride channel p64, is expressed in brush border of renal proximal tubule. *Am. J. Physiol.* **274**, F1140 – F1149.

Tulk, B.M., Schlesinger, P. H., Kapadia, S. A. and Edwards J. C. (2000). CLIC1 functions as a chloride channel when expressed and purified from bacteria. *J. Biol. Chem.* **275**, 26986 – 26998.

Tulk, B.M., Kapadia, S., and Edwards, C.J. (2002). CLIC1 inserts from the aqueous hase into phospholipid membranes, where it functions as an anion channel. *Am. J. Physiol. Cell Physiol.* **282**, 1103-1112.

Uversky, V. N. and Ptitsyn, O. B. (1994). 'Partly folded' state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding β-lactamase at low temeperature. *Biochemistry.* **33,** 2782-2791.

Valenzuela, S. M., Martin, K. D., Por, B. S., Robbins, M. J., Warton, K., Bootcov, R. S., Schofield, R. P., Campbell, J. T., and Breit, N. S. (1997). Molecular cloning and expression of a chloride ion channel of cell nuclei. *J. Biol. Chem*. **272**, 12575-12582.

Vlamis-Gardikas, A., Åslund, F., Spyrou, G., Bergman, T., and Holmgren, A. (1997). Cloning, overexpression, and chracterization of Glutaredoxin 2, an atypical Glutaredoxin from Escherichia coli. *J. Biol. Chem*. **272**, 11236-11243.

Wang, L., Pan, H. and Smith, L. D. (2002). Hydrogen exchange-mass spectrometry optimization of digestion conditions. *Molecular and Cellular Proteomics.* **1**, 132 – 138.

Waldburger, C., Schildbach, J. and Sauer, R. (1995). Are buried salt-bridges important in protein stability and conformational specificity? *Nat. Struct. Biol.* **2**, 122 – 128.

Warton, K., Tonini, R., Fairlie, D. W., Matthews, M. J., Valenzuela, M. S., Qiu, R. M., Wu, M. W., Pankhurst, S., Bauskin, R. A., Harrop, J. S., Campbell, J. T., Curmi, G. M. P., Breit, N. S., and Mazzanti, M. (2002). Recombinant Clic1 (NCC27) assembles in lipid bilayers via a pH-dependant two-state process to form chloride ion channels with identical characteristics to those observed in Chinese hamster ovary cell expressing CLIC1. *J. Biol. Chem.* **277**, 26003-26011.

Wallace, A. L., Sluis-Cremer, N. and Dirr, W. H. (1998). Equilibrium and kinetic unfolding properties of dimeric human glutathione transferase A1-1. *Biochemistry* **37,** 5320-5328.

Wallace, A. L., Burke, J., and Dirr, W. H. (2000). Domain-domain interface packing at conserved Trp-20 in class α glutathione transferase impacts on protein stability. *Biochim. Biophys. Acta* **1478,** 325-332.

Wenk, M., Jaenicke, R., and Mayr, E-M. (1998). Kinetic stabilisation of a modular protein by domain interactions. *FEBS lett.* **438**, 127-130.

Wetlaufer., D.B. (1973). Nucleation, Rapid folding, and Globular Intrachain regions in Proteins. *Proc. Nat. Acad. Sci. U.S.A* **70**, 697-701.

Whitten, T. S. and Garcia-Moreno, B., (2000). pH dependence of stability of Staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. *Biochemistry* **39**, 14929-14304.

Wilce, M. C. J. and Parker, M. W. (1994). Structure and function of glutathione S-transferases. *Biochim. Biophys. Acta.* **1205**, 1-18.

Woods, L. V. and Hamuro, Y. (2001). High-resolution, high-throughput amide deuterium exchange-mass spectrometry (DXMS) determination of protein binding site structure and dynamics: utility in pharmaceutical design. *J. Cellular Biochemistry Supplement.* **37**, 89 – 98.

Woody, R.W. (1995) Circular dichroism. *Methods Enzymol.* **246**, 34-71.

Wolfenden, R., Anderssson, L., Cullis P. M., and Southgate, B. C. C. (1981). Affinities of amino acid side chains for solvent water. *Biochemistry* **20**, 849-855.

Wu, Y., Kaveti, S. and Engen, R. J. (2006). Extensive deuterium back-exchange in certain immobilized pepsin columns used for H/D exchange mass spectrometry. *Anal. Chem.* **78**, 1719 – 1723.

Xia, B., Vlamis-Gardikas, A., Holmgren, A., Wright, E. P., and Dyson, J.H. (2001). Solution structure of Escherichia coli Glutaredoxin 2 shows similarity to mammalian glutathione S-transferases. *J. Mol. Biol*. **310**, 907-918.

Xiao, H., Hoerner, K. J., Eyles, J. H., Dobo, A., Voigtman, E., Cuk, M. I. A. and Kaltashov, A. I. (2004). Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I a generalized model for a two-state protein and comparison with experiment. *Protein Sci.* **14**, 543 – 557.

Xiao, W., Chen, B., Yao, S., Cheng, Z. (2005). Simultaneous determination of benazepril hydrochloride and benazeprilat in plasma by high performance liquid chromatography/electrospray mass spectrometry. *J. Chromatorg. B*. **814**, 303 – 308.

Zhang, Z. and Smith, L. D. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **2**, 522 – 531.
Zehfus, M. H. (1997) Identificatioon of compact, hydrophobically stabilised domains and modules containing multiple peptide chains. *Protein Sci.* **6**, 1210-1219.

Zhongqi, Z. and Marshall, A. G. (1998). A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra, *J. Am. Soc. Mass Spectrom.* **9**, 225-233.

Appendix

Key to Tables A and B: structural alignment of GST-family domain-interface residues and sequence alignment of CLIC-family proteins

Table A: Multiple structural alignment of domain interface residues belonging to GST family proteins

- Alignment was performed using MultiProt (Shatsky *et al.,* 2004).
- The notation used is: domain/residue name/residue number i.e. N.I.9 stands for N-terminal domain isoleucine 9.
- Capital letters are used to indicate that the residue is within the 4 Å cut-off used to detect domain interface amino acids (see section **2.2.1**). Residues shown in small letters are outside the 4 Å cut-off and were included after the structural alignment was performed.
- Residues are coloured according to the properties of their side chains: non-polar (Ala, Val, Iso, Leu, Met, Pro and Cys) blue, polar uncharged (Ser, Thr, Asn and Gln) red, polar positively charged (Lys, Arg and His) grey, polar negatively charged (D and E) green, aromatic (Phe, Tyr and Trp) pink and glycine is shown in cyan.
- The conservation ratios ($\mathbf{Cr_{x\text{-}GST}}$, $\mathbf{Cr_{clic\text{-}GST}}$, and $\mathbf{Cr_{z\text{-}GST}}$) at each aligned position are calculated according to side chain property (see section **2.2.1**). Conservation ratios are only calculated when three or more amino acids with the same side chain chemistry are present at an aligned position.

Table B: Sequence-based alignment of CLIC family proteins

- Alignment was performed using T-Coffee (Notredame *et al*., 2000).
- Residue numbering and secondary structural content follow that of CLIC1 (pdb code: 1k0m).
- Domain interface residues are coloured as per **Table A**.
- Consensus GST-family domain interface positions obtained from **Table A** are indicated above the alignment.
- CLIC domain interface positions not found in the consensus GST-family interface are marked with arrows.
- The conservation ratio ($\mathbf{Cr_{x\text{-}CLIC}}$) and uniqueness factor (**Uf**) are calculated as shown in section **2.2.1**.

Table A: Multiple structural alignment of domain-interfaces belonging to the GST protein family

		Consensus Domain-Inteface Position									
Class	**PDB code**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Alpha	**2fhe**	N.I.9	N.R.10	N.G.11	N.L.12	N.Q.14	N.P.15	N.R.17	N.L.18	N.L.19	N.L.20
	1k3o	n.a.12	N.R.13	n.g.14	N.R.15	N.E.17	N.S.18	N.R.20	N.W.20	n.l.22	N.L.23
	1f3b	N.A.11	N.R.12	n.g.13	n.r.14	N.E.16	n.c.17	N.R.19	N.W.20	N.L.21	N.L.22
	1ml6	N.A.11	N.R.12	n.g.13	N.R.14	N.E.16	N.C.17	N.R.19	N.W.20	n.l.21	N.L.22
	1guk	N.G.12	N.R.13	N.G.14	N.R.15	N.E.17	n.s.18	N.R.20	N.W.21	N.L.22	N.L.23
	1ev9	N.A.12	N.R.13	N.G.14	n.r.15	N.E.17	N.C.18	N.R.20	N.F.21	N.L.22	N.L.23
	1oe8	n.g.13	N.R.14	n.g.15	N.R.16	N.E.18	n.s.19	N.R.21	N.M.22	N.T.23	N.L.22
Beta	**1pmt**	N.P.7	N.S.9	N.S.9	N.C.10	N.L.12	N.S.13	N.H.15	N.I.16	n.v.17	n.l.18
	1a0f	N.P.7	N.A.9	N.A.9	N.C.10	n.l.12	n.a.13	N.H.15	N.I.16	N.T.17	n.l.18
CLIC1	**1k0m**	n.g.22	N.N.23	n.c.24	n.f.26	N.Q.28	N.R.29	N.F.31	N.M.32	N.V.33	n.l.34
CLIC4	**2ahe**	n.g.33	N.N.34	n.c.35	n.f.37	N.Q.39	N.R.40	N.F.42	N.M.43	N.I.44	n.l.45
Delta	**1jlv**	N.G.8	N.S.9	N.A.10	n.p.11	N.R.13	N.A.14	N.Q.16	N.M.17	N.T.18	n.a.19
	1jlw	N.G.8	N.S.9	N.A.10	n.p.11	N.R.13	N.A.14	N.Q.16	N.M.17	N.T.18	n.a.19
	1r5a	N.A.10	N.A.10	N.P.12	N.P.13	N.R.15	N.S.16	N.L.18	N.L.19	N.L.20	n.a.21
	1v2a	N.L.7	N.S.9	N.P.10	N.P.11	N.Q.13	N.S.14	n.i.16	N.L.17	N.L.18	n.a.19
ElonFac.	**1nhy**	n.f.10	N.R.11	n.r.11	n.r.13	N.W.15	n.v.16	N.R.18	N.G.19	n.l.20	n.v.21
Grx2	**1g7o**	N.H.8	N.H.8	N.P.10	n.y.11	N.L.13	N.K.14	N.R.16	N.M.17	N.I.18	n.f.19
Mu	**1gsu**	N.I.9	N.R.10	N.G.11	N.L.12	N.H.14	N.A.15	N.R.17	N.L.18	N.L.19	n.l.20
	1gtu	n.i.9	N.R.10	N.G.11	N.L.12	N.H.14	N.A.15	N.R.17	N.L.18	N.L.19	n.l.20
	2gst	N.V.9	N.R.10	N.G.11	N.L.12	N.H.14	N.P.15	N.R.17	N.L.18	N.L.19	n.l.20
Omega	**1eem**	N.F.31	N.C.32	N.P.33	N.F.34	N.E.36	N.R.37	N.R.39	N.L.40	N.V.41	n.l.42
Pf	**1pa3**	n.a.12	N.R.13	n.g.14	n.k.15	N.E.17	N.L.18	N.R.20	N.L.21	N.I.22	n.f.23
Phi	**1gnw**	N.A.10	N.A.10	N.I.12	n.a.13	N.R.15	N.R.16	N.L.18	N.I.19	N.A.20	N.L.21
	1axd	n.m.10	N.S.11	N.W.12	n.n.13	N.T.15	N.R.16	N.A.18	N.T.19	n.a.20	N.L.21
	1aw9	N.L.11	N.P.13	n.s.12	N.N.14	N.V.16	N.R.17	n.a.19	N.T.20	N.V.21	N.L.21
Pi	**1gss**	N.V.10	N.R.11	N.G.12	N.R.13	N.A.15	n.a.16	N.R.18	N.M.19	N.L.20	N.L.21
	1glp	N.V.10	N.R.11	N.G.12	N.R.13	N.E.15	N.A.16	n.r.18	N.M.19	N.L.20	N.L.21
	2gsr	N.V.10	N.R.11	N.G.12	N.R.13	N.E.15	N.A.16	N.R.18	N.M.19	N.L.20	N.L.21
Sj	**1gta**	N.I.10	N.K.11	N.G.12	N.L.13	N.Q.15	N.P.16	N.R.18	N.L.19	n.l.20	N.L.21
Sigma	**1m0u**	N.V.57	N.A.59	N.A.59	N.L.60	N.E.62	N.P.63	N.R.65	N.Y.66	N.L.67	n.f.68
	1iyi	N.M.11	N.R.12	N.G.13	N.R.14	N.E.16	N.I.17	N.R.19	N.Y.20	n.i.21	n.f.22
	1pd2	N.M.11	N.R.12	N.G.13	N.R.14	N.E.16	N.I.17	N.R.19	N.Y.20	n.i.21	n.f.22
	2gsq	N.L.10	N.M.11	N.G.12	N.R.13	N.E.15	N.L.16	N.R.18	N.F.19	N.V.20	N.L.21
Tau	**1oyj**	N.V.14		N.P.16	N.F.17	N.Q.19	N.R.20	N.R.22	N.I.23	N.A.24	
	1gwc	N.P.14	N.S.15	n.s.15	N.F.17	N.T.19	N.R.20	N.K.22	N.L.23	n.a.24	n.l.25
Theta	**1ljr**	N.V.10	N.L.9	N.Q.12	n.p.13	N.R.15	N.A.16	N.Y.18	N.I.19	N.F.20	n.a.21
Unknow	**1tw9**	N.G.11	N.R.12	N.G.13	n.a.14	N.E.16	N.C.17	N.R.19	N.Q.20	N.V.21	n.f.22
Ure2p	**1g6w**	n.s.121	N.A.122	n.a.122	N.N.124	N.F.126	N.K.127	N.A.129	N.I.130	n.v.131	n.l.132
Zeta	**1e6b**	N.R.16	N.S.17	N.S.18	n.c.19	N.H.21	N.R.22	N.R.24	N.I.25	N.A.26	n.l.27
	1fw1	N.R.13	N.S.14	N.S.15	N.C.16	N.W.18	N.R.19	N.R.21	N.I.22	N.A.23	n.l.24
$Cr_{x\text{-}GST}$		0.67	0.51	0.48	0.43	0.38	0.51	0.74	0.67	0.87	0.87
$Cr_{clic\text{-}GST}$		0.17	0.25	0.31	0.13	0.2	0.28	0.075	0.67	0.87	0.87
$Cr_{z\text{-}GST}$		n/a	n/a	0.15	0.36	0.18 0.13	0.15	0.18	0.22 0.075	0.1	0.15

See page 162 for key.

Table A: Multiple structural alignment of domain-interfaces belonging to the GST protein family

		Consensus Domain-Inteface Position									
Class	**PDB code**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Alpha	**2fhe**	N.E.21	N.Y.22	N.L.23	N.G.24	N.E.25	N.Y.27	N.E.29	N.I.31	N.Y.32	N.D.59
	1k3o	N.A.24	N.A.25	n.a.26	N.G.27	N.V.28	N.F.30	N.E.32	n.f.34	n.i.35	n.i.60
	1f3b	n.a.23	N.A.24	N.A.25	N.G.26	n.v.27	N.F.29	N.E.31	N.F.33	N.I.34	n.i.59
	1ml6	N.A.23	N.A.24	N.A.25	N.G.26	N.V.27	N.F.29	N.E.31	N.F.33	n.i.34	n.i.59
	1guk	N.A.24	N.A.25	N.A.26	N.G.27	N.V.28	N.F.30	N.E.32	N.F.34	n.I.35	N.I.60
	1ev9	N.A.24	N.A.25	N.A.26	N.G.27	N.V.28	N.F.30	N.E.32	N.F.34	n.i.35	n.i.60
	1oe8	N.V.25	N.A.26	N.A.27	n.g.28	N.V.29	N.Y.31	n.d.33	n.r.35	n.i.36	n.i.58
Beta	**1pmt**	N.R.19	N.E.20	N.T.21	n.g.22	N.L.23	n.f.25	n.i.27	n.r.29	n.i.30	n.I.57
	1a0f	N.R.19	n.e.20	N.S.21	N.G.22	N.K.23	n.f.25	n.I.27	n.s.29	n.v.30	n.I.57
CLIC1	**1k0m**	N.W.35	N.L.36	N.K.37	n.g.38	N.V.39	n.f.41	n.v.43	n.t.45	n.v.46	n.y.69
CLIC4	**2ahe**	N.W.46	N.L.47	N.K.48	n.g.49	N.V.50	n.f.52	n.v.54	n.t.56	n.v.57	n.f.80
Delta	**1jlv**	N.A.20	N.A.21	N.V.22	n.g.23	n.v.24	n.I.26	n.I.28	n.I.30	n.t.31	n.d.57
	1jlw	n.a.20	N.A.21	N.V.22	n.g.23	N.V.24	n.I.26	n.I.28	n.I.30	n.t.31	N.D.57
	1r5a	N.K.22	N.M.23	N.I.24	n.g.25	N.V.26	n.I.28	n.I.30	n.v.32	n.I.33	N.D.59
	1v2a	N.K.20	N.K.21	n.I.22	n.g.23	N.I.24	n.I.26	n.I.28	n.k.30	n.t.31	N.D.56
ElonFac.	**1nhy**	N.K.22	N.A.23	N.L.24	n.k.25	N.L.26	n.v.28	n.v.30	n.t.32	n.p.33	n.g.54
Grx2	**1g7o**	N.G.20	N.L.21	N.K.22	n.n.23	n.i.24	n.v.26	n.I.28	N.V.30	N.L.31	n.k.53
Mu	**1gsu**	N.E.21	N.Y.22	N.T.23	N.E.24	N.T.25	n.y.27	n.e.29	N.R.31	n.y.32	N.D.64
	1gtu	N.E.21	N.Y.22	n.t.23	N.D.24	N.S.25	n.y.27	n.e.29	n.k.31	n.y.32	N.D.64
	2gst	N.E.21	N.Y.22	N.T.23	N.D.24	N.S.25	N.Y.27	n.e.29	N.R.31	n.y.32	n.d.64
Omega	**1eem**	N.K.43	N.A.44	N.K.45	n.g.46	N.I.47	n.h.49	n.v.51	n.n.53	n.i.54	n.n.77
Pf	**1pa3**	N.A.24	N.Y.25	N.L.26	n.g.27	n.i.28	n.y.30	n.d.32	n.r.34	n.f.35	n.i.64
Phi	**1gnw**	N.H.22	N.E.23	N.K.24	N.N.25	N.L.26	n.f.28	n.I.30	n.h.32	n.v.33	N.D.59
	1axd	N.E.22	N.E.23	N.A.24	N.G.25	N.S.26	n.y.28	n.i.30	n.p.32	n.i.33	n.d.59
	1aw9	N.N.23	N.E.24	N.K.25	n.g.26	n.I.27	n.f.29	n.i.31	n.p.33	n.v.34	N.D.60
Pi	**1gss**	N.A.22	N.D.23	N.Q.24	N.G.25	n.q.26	N.W.28	N.E.30	n.v.32	n.v.33	n.d.55
	1glp	N.A.22	N.D.23	n.q.24	N.G.25	n.q.26	n.w.28	n.e.30	n.v.32	n.v.33	n.d.57
	2gsr	N.A.22	N.D.23	N.Q.24	n.d.25	n.q.26	n.w.28	n.e.30	n.v.32	n.v.33	n.d.55
Sj	**1gta**	N.E.22	N.Y.23	N.L.24	N.E.25	N.E.26	n.y.28	n.e.30	n.I.32	N.Y.33	n.d.60
Sigma	**1m0u**	n.a.69	N.Y.70	n.g.71	n.n.72	n.q.73	N.Y.75	N.D.77	N.R.79	n.v.80	n.v.102
	1iyi	N.A.23	N.Y.24	n.I.25	n.d.26	n.i.27	N.Y.29	N.D.31	N.R.33	n.i.34	n.v.56
	1pd2	N.A.23	N.Y.24	n.I.25	n.d.26	n.i.27	N.Y.29	n.d.31	N.R.33	n.i.34	n.v.56
	2gsq	N.A.22	N.A.23	n.h.24	N.G.25	n.e.26	n.f.28	n.d.30	n.v.32	n.v.33	n.i.55
Tau	**1oyj**	N.A.26	N.E.27	N.K.28	n.g.29	N.L.30	n.f.32	n.y.34	n.e.36	n.e.37	N.H.61
	1gwc	N.A.26	N.L.27	N.K.28	n.g.29	N.L.30	n.y.32	n.d.34	n.e.36	n.e.37	N.H.61
Theta	**1ljr**	N.K.22	N.K.23	N.N.24	n.g.25	N.I.26	n.I.28	n.I.30	n.t.32	n.v.33	n.d.59
Unknow	**1tw9**	N.A.23	N.L.24	N.A.25	N.D.26	n.q.27	N.Y.29	n.d.31	N.R.33	n.I.34	n.v.56
Ure2p	**1g6w**	N.S.133	N.E.134	N.L.135	n.g.136	n.f.137	N.Y.139	n.t.141	n.f.143	n.I.144	n.d.170
Zeta	**1e6b**	N.A.28	N.L.29	N.K.30	n.g.31	N.L.32	n.y.34	N.Y.36	N.P.38	n.v.39	n.d.65
	1fw1	N.A.25	N.L.26	N.K.27	n.g.28	N.I.29	n.y.31	N.T.33	N.P.35	n.i.36	N.I.64
$Cr_{x\text{-}GST}$		0.57	0.5	0.47	0.7	0.65	0.8	0.5	0.35	0.67	0.47
$Cr_{clic\text{-}GST}$		0.05	0.5	0.25	0.7	0.65	0.8	0.35	0.15	0.67	0.05
$Cr_{z\text{-}GST}$		0.2 0.15	0.22 0.25	0.22	0.22 0.075	0.075 0.22	0.17	n/a	0.15 0.3	0.15 0.075	0.37

See page 162 for key.

Table A: Multiple structural alignment of domain-interfaces belonging to the GST protein family

		Consensus Domain-Inteface Position								
Class	**PDB code**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**
Alpha	**2fhe**	N.Q.66	N.S.67	N.L.68	N.L.71	N.R.72	N.Y.73	N.I.74	N.A.75	L.M.80
	1k3o	n.q.67	n.t.68	N.R.69	N.L.72	N.N.73	n.y.74	N.I.75	N.A.76	L.M.81
	1f3b	n.q.66	n.t.67	N.R.68	N.L.71	N.N.72	n.y.73	N.I.74	N.A.75	L.M.80
	1ml6	n.q.66	n.t.67	N.R.68	N.L.71	N.N.72	n.y.73	N.I.74	N.A.75	L.M.80
	1guk	n.q.67	n.t.68	n.r.69	N.L.72	N.S.73	N.Y.74	N.L.75	N.A.76	L.M.81
	1ev9	n.q.67	n.t.68	N.R.69	N.L.72	N.N.73	n.y.74	N.I.75	N.A.76	L.M.81
	1oe8	n.e.70	n.s.71	n.l.72	N.A.75	N.R.76	n.y.77	N.M.78	N.A.79	L.M.84
Beta	**1pmt**	n.e.65	n.g.66	N.V.67	N.V.70	N.Q.71	N.Y.72	N.L.73	N.A.74	L.L.82
	1a0f	n.e.65	n.g.66	N.V.67	N.M.70	N.Q.71	N.Y.72	N.L.73	N.A.74	L.L.82
CLIC1	**1k0m**	n.d.76	n.t.77	n.n.78	N.E.81	n.e.82	n.f.83	N.L.84	N.E.85	L.L.96
CLIC4	**2ahe**	n.d.87	n.v.88	n.n.89	N.E.92	N.E.93	n.f.94	N.L.95	N.E.96	L.L.107
Delta	**1jlv**	N.E.64	n.s.65	N.R.66	n.c.69	N.T.70	n.y.71	N.L.72	N.A.73	L.L.82
	1jlw	N.E.65	n.s.66	N.R.67	N.Q.70	N.I.71	n.y.72	N.L.73	N.V.74	L.L.88
	1r5a	N.E.66	n.s.67	N.R.68	N.L.71	N.S.72	n.y.73	N.L.74	N.V.75	I.I.84
	1v2a	N.E.63	n.s.64	N.Y.65	n.v.68	n.l.69	n.y.70	n.l.71	N.V.72	L.L.81
ElonFac.	**1nhy**	n.e.62	n.a.63		n.n.67	N.Y.68	N.Y.69	N.L.70	N.V.71	L.L.83
Grx2	**1g7o**	n.e.61	n.s.62	n.m.63	N.V.66	N.H.67	n.y.68	n.v.69	N.D.70	L.L.78
Mu	**1gsu**	n.q.71	n.s.72	N.N.73	N.L.76	N.R.77	N.Y.78	N.I.79	N.A.80	L.M.85
	1gtu	n.q.71	n.s.72	n.n.73	N.L.76	N.C.77	N.Y.78	N.I.79	N.A.80	L.L.85
	2gst	n.q.71	N.S.72	n.n.73	N.M.76	N.R.77	n.y.78	N.L.79	N.A.80	L.L.85
Omega	**1eem**	n.e.85	n.s.86	n.a.87	N.C.90	n.e.91	n.y.92	N.L.93	N.D.94	L.L.102
Pf	**1pa3**	n.q.71	n.s.72	N.Q.73	N.V.76	N.R.77	n.y.78	N.L.79	N.S.80	L.I.85
Phi	**1gnw**	n.e.66	n.s.67	N.R.68	N.T.71	N.Q.72	n.y.73	N.I.74	N.A.75	L.L.85
	1axd	n.e.66	n.s.67	N.R.68	N.C.71	N.K.72	n.y.73	N.A.74	N.A.75	L.L.82
	1aw9	n.e.67	n.s.68	N.R.69	N.N.72	N.R.73	n.y.74	N.I.75	N.A.76	L.L.86
Pi	**1gss**	n.q.62	n.s.63	N.N.64	N.L.67	N.R.68	n.h.69	N.L.70	N.G.71	L.L.76
	1glp	n.q.64	n.s.65	N.N.66	N.L.69	N.R.70	n.h.71	n.l.72	N.G.73	I.I.78
	2gsr	n.q.62	n.s.63	N.N.64	N.L.67	N.R.68	n.h.69	N.L.70	N.G.71	L.L.76
Sj	**1gta**	n.q.67	n.s.68	n.m.69	N.I.72	N.R.73	N.Y.74	N.I.75	N.A.76	L.M.81
Sigma	**1m0u**	n.q.109	n.s.110	N.I.111	N.A.114	N.R.115	n.f.116	n.l.117	N.A.118	L.L.123
	1iyi	n.q.63	n.s.64	N.L.65	N.A.68	N.R.69	N.Y.70	N.L.71	N.T.72	L.L.77
	1pd2	n.q.63	n.s.64	N.L.65	N.A.68	N.R.69	N.Y.70	N.L.71	N.T.72	L.L.77
	2gsq	n.q.62	n.s.63	N.M.64	N.A.67	N.R.68	n.h.69	n.l.70	N.A.71	L.L.76
Tau	**1oyj**	n.e.68	n.s.69	N.L.70	N.L.73	N.Q.74	N.Y.75	N.L.76	N.D.77	L.L.86
	1gwc	n.e.68	n.s.69	N.M.70	N.L.73	N.Q.74	n.y.75	N.I.76	N.D.77	L.L.87
Theta	**1ljr**	n.e.66	N.S.67	n.s.68	N.L.71	N.I.72	n.y.73	n.l.74	N.S.75	L.W.84
Unknow	**1tw9**	n.q.63	n.s.64	N.Q.65	N.C.68	N.R.69	n.y.70	n.l.71	N.A.72	L.F.77
Ure2p	**1g6w**	n.e.180	n.s.181	N.G.182	N.L.185	N.L.186	n.h.187	n.l.188	N.V.189	L.L.201
Zeta	**1e6b**	n.d.72	N.S.73	N.F.74	N.I.77	N.M.78	n.y.79	N.L.80	N.D.81	L.L.89
	1fw1	n.q.71	n.s.72	N.L.73	N.I.76	N.E.77	N.Y.78	N.L.79	N.E.80	L.L.88
$Cr_{x\text{-}GST}$		0.5	0.9	0.36	0.85	0.42	0.87	1	0.62	0.95
$Cr_{clic\text{-}GST}$		0.5	0.9	0.28	0.05	0.075	0.87	1	0.2	0.95
$Cr_{z\text{-}GST}$		n/a	n/a	0.28	0.1	0.3	0.12	n/a	0.075 0.1	n/a

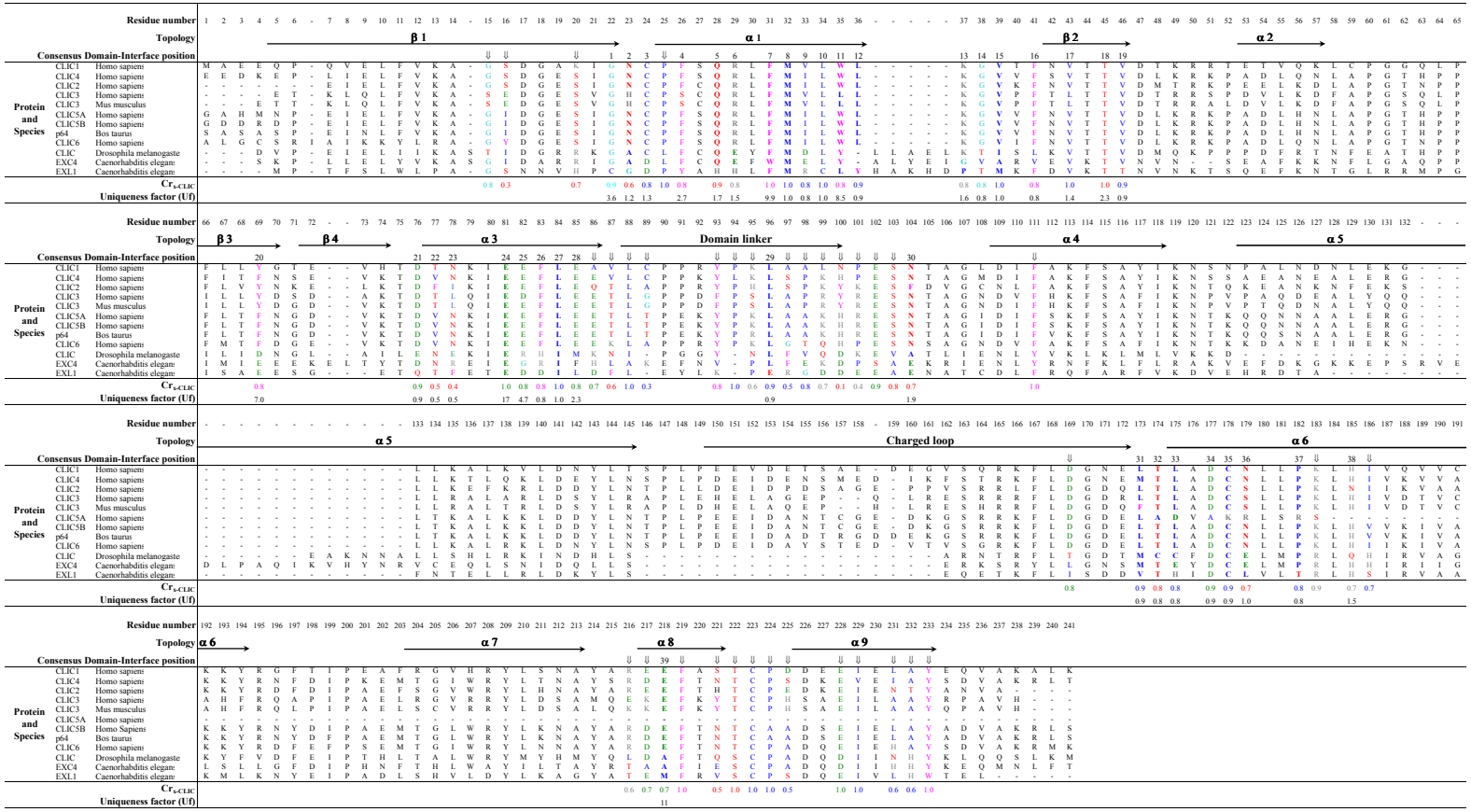
See page 162 for key.

Table A: Multiple structural alignment of domain-interfaces belonging to the GST protein family

		Consensus Domain-Inteface Position									
Class	**PDB code**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
Alpha	**2fhe**	C.S.92	C.V.147	C.S.148	C.H.149	C.D.151	C.F.152	C.M.153	C.E.156	C.D.159	C.I.188
	1k3o	C.D.93	C.L.153	C.S.154	C.R.155	C.D.157	C.I.158	c.h.159	C.E.162	C.Y.165	c.v.194
	1f3b	C.D.92	C.L.152	C.T.153	c.r.154	C.D.156	C.I.157	c.h.158	C.E.161	C.L.164	C.V.193
	1ml6	C.D.92	C.L.152	C.T.153	C.R.154	C.D.156	C.I.157	c.h.158	C.E.161	C.L.164	C.V.193
	1guk	C.D.93	C.L.153	C.S.154	C.W.155	C.D.157	C.I.158	c.q.159	C.E.162	c.l.165	C.I.194
	1ev9	C.D.93	C.L.153	C.T.154	C.R.155	C.D.157	C.I.158	c.h.159	C.E.162	c.l.165	c.v.194
	1oe8	C.E.96	C.V.156	C.T.157	C.L.158	c.d.160	C.L.161	c.v.162	c.a.165	C.D.168	C.L.199
Beta	**1pmt**	c.i.95	C.F.151	c.t.152	c.v.153	c.d.155	C.A.156	C.Y.157	C.T.160	c.q.163	C.V.191
	1a0f	C.I.95	c.f.151	c.t.152	c.i.153	c.d.155	C.A.156	C.Y.157	C.T.160	C.R.163	c.v.191
CLIC1	**1k0m**	C.N.104	C.L.173	C.T.174	C.L.175	c.d.177	C.C.178	C.N.179	C.P.182	c.h.185	C.E.218
CLIC4	**2ahe**	C.N.115	C.M.184	C.T.185	C.L.186	c.d.188	C.C.189	C.N.190	C.P.193	c.h.196	C.E.229
Delta	**1jlv**	C.N.94	C.L.152	C.T.153	C.I.154	c.d.156	C.L.157	C.T.158	C.A.161	C.S.164	C.G.192
	1jlw	C.H.100	C.P.161	C.T.162	C.I.163	c.d.165	C.L.166	C.S.167	C.A.170	c.a.173	C.G.201
	1r5a	C.D.96	C.F.154	C.T.155	C.I.156	c.d.158	C.I.159	C.A.160	C.V.163	c.s.166	c.l.192
	1v2a	C.N.93	C.L.150	C.T.151	C.V.152	c.d.154	C.I.155	C.C.156	C.G.159	c.t.162	C.M.188
ElonFac.	**1nhy**	c.i.97	C.I.158	C.S.159	C.L.160	c.d.162	C.L.163	c.v.164	C.S.167	C.T.170	c.k.202
Grx2	**1g7o**	C.E.87	C.L.168	C.S.169	C.E.170	C.D.172	C.I.173	c.q.174	C.P.177	C.R.180	C.T.205
Mu	**1gsu**	C.D.97	C.L.152	C.T.153	C.F.154	c.d.156	C.F.157	C.L.158	C.D.161	C.D.164	C.I.193
	1gtu	c.d.97	C.I.152	C.T.153	C.F.154	c.d.156	C.F.157	C.L.158	C.D.161	C.D.164	c.i.193
	2gst	C.D.97	C.V.152	C.T.153	C.Y.154	c.d.156	C.F.157	C.L.158	C.D.161	C.D.164	C.I.193
Omega	**1eem**	c.k.114	c.i.170	C.S.171	C.M.172	c.d.174	C.Y.175	C.L.176	C.P.179	C.E.182	C.V.212
Pf	**1pa3**	C.D.97	C.L.158	C.T.159	C.Y.160	c.d.162	C.L.163	C.A.164	C.N.167	C.D.170	c.i.199
Phi	**1gnw**	c.m.103	C.F.165	C.T.166	C.L.167	c.d.169	C.L.170	C.H.171	C.P.174	c.q.177	C.A.206
	1axd	c.d.96	c.l.161	C.S.162	C.L.163	c.d.165	C.L.166	C.N.167	C.V.169	c.l.173	c.s.202
	1aw9	C.L.100	C.F.162	C.T.163	C.L.164	c.d.166	C.A.167	C.N.168	C.S.171	c.l.174	C.F.204
Pi	**1gss**	C.D.88	C.I.146	C.S.147	C.F.148	c.d.150	C.Y.151	C.N.152	C.D.155	c.l.158	C.L.187
	1glp	C.D.90	C.I.148	C.S.149	C.F.150	c.d.152	C.Y.153	C.N.154	C.D.157	c.l.160	C.I.189
	2gsr	C.D.88	C.I.146	C.S.147	C.F.148	c.d.150	C.Y.151	C.N.152	C.D.155	c.r.158	C.I.187
Sj	**1gta**	C.S.93	C.V.148	C.T.149	C.H.150	c.d.152	C.F.153	C.M.154	C.D.157	C.D.160	C.I.189
Sigma	**1m0u**	C.D.135	C.L.195	C.T.196	C.W.197	c.d.199	C.V.200	C.Y.201	C.G.204	c.d.207	C.I.237
	1iyi	C.D.89	C.V.146	C.T.147	C.W.148	c.d.150	C.F.151	C.Y.152	C.I.155	C.T.158	C.V.187
	1pd2	C.D.89	C.V.146	C.T.147	C.W.148	c.d.150	C.F.151	C.Y.152	C.I.155	C.T.158	C.I.187
	2gsq	C.D.88	C.M.149	C.T.150	C.L.151	c.d.153	C.L.154	C.H.155	C.V.158	C.E.161	C.I.190
Tau	**1oyj**			C.G.163		c.d.166	C.V.167				
	1gwc	C.R.99	c.g.157	C.G.159	C.L.160	c.d.162	C.V.163	C.A.164	C.G.167	c.l.169	C.A.205
Theta	**1ljr**	C.H.96	C.V.162	C.T.163	C.L.164	C.T.163	C.L.167	C.M.168	C.E.171	c.m.174	C.L.201
Unknow	**1tw9**	C.D.89	C.I.153	C.S.154	C.W.155	c.d.157	C.L.158	c.l.159	C.E.162	c.a.165	C.I.194
Ure2p	**1g6w**	C.N.213	C.L.306	C.T.307	c.i.308	c.d.310	C.L.311	C.A.312	C.P.315	c.n.318	C.V.347
Zeta	**1e6b**	C.Y.101	C.I.164	C.Y.165	C.L.166	C.Y.165	c.d.168	C.F.170	C.P.173	c.h.176	C.F.205
	1fw1	C.R.100	c.v.159	C.T.160	C.M.161	C.T.160	c.d.163	C.C.165	C.P.168	C.A.171	C.F.199
$Cr_{x\text{-}GST}$		0.53	0.87	0.92	0.51	0.92	0.67	0.41	0.38	0.31	0.74
$Cr_{clic\text{-}GST}$		0.18	0.87	0.92	0.51	0.92	0.67	0.28	0.38	0.15	0.05
$Cr_{z\text{-}GST}$		0.13 0.13	0.13	n/a	0.13 0.32	n/a	0.27	0.15 0.15	0.13	0.26	0.1

See page 162 for key.

Table B: Sequence-based alignment of CLIC family proteins



Key to Tables C - E: domain interface characteristics

Table C: Inter-domain hydrogen bonds in the GST family

- Inter-domain hydrogen bonds are calculated using software provided by the iMOLTalk server (http://i.moltalk.org) using a cut-off distance of 3.9 Å and verified using the PPI server as well as visual inspection of the of the individual domain interfaces using SwissPDB viewer (Guex and Peitsch**,** 1997).
- The notation used is: residue name/residue number/number of hydrogen bonds/ interface segment i.e. R13/2_1/1 stands for arginine 13 found in interface segment 1 forms 2 inter-domain and 1 intra-domain hydrogen bonds. The interface to which each residue belongs is indicated below the pdb codes of the proteins.
- Residues are coloured according to the properties of their side chains: non-polar (Ala, Val, Iso, Leu, Met, Pro and Cys) blue, polar uncharged (Ser, Thr, Asn and Gln) red, polar positively charged (Lys, Arg and His) grey, polar negatively charged (D and E) green, aromatic (Phe, Tyr and Trp) pink and glycine is shown in cyan..

Table D: Inter-domain salt-bridges in the GST family

- Inter-domain salt bridges are calculated using software provided by the iMOLTalk server (http://i.moltalk.org) using a cut-off distance of 4.0 Å and verified using the PPI server.
- The notation used is: residue name/residue number/number of salt bridges/ interface segment i.e. R10/2/1 stands for arginine 10 found in interface segment 1 forms 2 inter-domain salt-bridges. The interface to which each residue belongs is indicated below the pdb codes of the proteins.

Table E: Domain interface properties of the GST protein family

- All parameters are generated using software provided by the PPI server (http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html) (see section **2.2.2**).
- A cut-off of 4 Å is used between interacting domain-domain interfaces.

Table C: Inter-domain hydrogen bonds in the GST family

Class	Alpha															
Origin	Fasciola hepatica		Human A1-1		Mouse A1-1		Mouse A2-2		Mouse A4-4		Rat A1-1		Schistosoma Haematobium	
PDB code	2fhe		1k30		1f3b		1ml6		1guk		1ev9		1oe8	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C	N	C
Interdom.H-bonds	12		10		10		11		9		10		10	
Residue ID/ No.Bonds/ Segment no.	R10/6/1	E156/1/2	R13/2**1**/1	E169/1**1**/2	R12/3**1**/1	E168/2**2**/2	R12/3**2**/1	E168/2**2**/2	Y9/1/1	Y212/1/3	R13/3**3**/1	E169/2**2**/2	R14/2/1	H169/1**1**/2
		D159/2/2		K205/3/3		K204/3/3		K204/3/3	R13/3**1**/1	R169/2**1**/2		K205/3/3		D172/1**2**/2
		K199/2/3	E17/1**2**/1		E16/1**3**/1		E16/1**3**/1			K205/3/3	E17/1**3**/1		R16/3/1	E103/1**1**/1
		P201/1/3	E32/1**1**/1		E31/1**1**/1		E31/1**1**/1		E17/1**2**/1		E32/1**1**/1			Y110/2/1
	G11/1/1	L202/1/3	R15/2/1	E104/2**1**/1	A24/1/1	N192/1**1**/3	R14/1/1	E103/1**1**/1	E32/1/1		A25/1**2**/1	N193/1**1**/3	E18/23/1	Y202/1**3**/3
	E14/1**2**/1	I198/1/3	A25/1**1**/1	T193/1**1**/3	F29/1/1	Q202/1/3	A24/1/1	N192/1**1**/3	A25/1**1**/1	T193/1**1**/3	F30/1/1	N203/1**1**/3		R206/1**2**/3
	E21/1/1	R196/1/3	R69/2**1**/2	E97/2/1	R68/2**1**/2	E96/2**2**/1	F29/1/1	Q202/1/3	R38/1/1	I218/1**1**/3	R69/2**1**/2	E97/2**1**/1	A26/1**3**/1	R198/1**1**/3
	R34/1/1	D213/1/3	N73/1**1**/2	R155/1**1**/2	N72/1**2**/2	R154/1**1**/2	R68/2**1**/2	E96/2**2**/1	R69/1**1**/2	D97/1**1**/1	N73/1**2**/2	R155/1**1**/2	R76/2**1**/2	E96/2**2**/1
	R72/2/2	E95/2/1					N72/1**2**/2	R154/1**1**/2						

See page 168 for key.

Table C: Inter-domain hydrogen bonds in the GST family

Class	Beta GST				CLIC1		CLIC4		Delta GST							
Origin	Proteus Mirabilus		E.coli		Human		Human		Mosqito 1-3		Mosqito 1-4		Mosqito 1-5		Mosqito 1-6	
PDB code	1pmt		1aof		1k0m		2ahe		1jlv		1jlw		1r5a		1v2a	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C	N	C	N	C
Interdom.H-bonds	7		8		5		3		6		6		2		4	
Residue ID/ No.Bonds/ Segment no.	G8/2/1	W164/1/2	G8/2/1	W164/1/2	S16/2/1	E228/2$_2$/3	S27/2/1	E239/2$_1$/3	R13/3$_3$/1	A161/1$_4$/2	R13/3$_4$/1	A170/1$_4$/2	R15/1$_4$/1	Q167/1$_1$/2	S14/1$_3$/1	C156/1$_2$/2
		E198/2/3		E198/2/3	K20/1$_1$/1	D225/1$_3$/3	Q39/1$_2$/1	T233/1$_3$/3		T165/1$_1$/2		T174/1$_1$/2	R68/1/2	D102/1$_1$/2	K20/1/1	D190/1$_1$/3
	H15/1$_2$/1		H15/1$_3$/1		Q28/2$_3$/1	T222/1$_3$/3				S164/1$_3$/2		N205/1$_3$/3			K21/1$_2$/1	E187/1/3
	R19/2$_2$/1	S193/1$_2$/3	E20/2$_3$/1	R188/2/3		C223/1/3			R66/2$_2$/2	D100/2$_4$/1	R67/2$_2$/2	D106/2$_3$/1			Y65/1$_1$/2	D99/1$_2$/1
		T197/1$_2$/3	D75/3/2	R91/35/1					E74/1$_1$/2	R90/1$_4$/1	Q70/1/2	S167/12/2				
	D75/2$_1$/2	R91/2$_2$/1														

See page 168 for key.

Table C: Inter-domain hydrogen bonds in the GST family

Class	Elongation factor		Grx2		Mu GST						Omega GST		Pf GST	
Origin	saccharomyces cerevisiae		E.coli		Chicken		Human		Rat		Human		Plasnodium Falcipar	
PDB code	1nhy		1g7o		1gsu		1gtu		2gst		1eem		1pa3	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C	N	C
Interdom.H-bonds	0		3		11		7		14		9		2	
Residue ID/ No.Bonds/ Segment no.			H8/1/1	E112/2/1	R10/5/1	D161/16/2	R10/41/1	D161/13/2	R10/61/1	D161/27/2	S8/2/1	D238/2/3	E17/14/1	Y202/1/3
			N33/1/1			D164/14/2		D164/12/2		D164/23/2	R30/2/1	Y239/1/3	Q74/13/2	Q104/11/1
			D34/11/1	T115/1/1		K204/21/3		P204/1/3		S204/1/3		C237/1/3		
						P206/1/3		P206/1/3		P206/1/3	F31/1/1	R183/23/2		
					G11/1/1	I207/1/3	G11/1/1	V207/1/3	G11/1/1	Q165/14/2	E36/14/1			
					E21/22/1	K192/11/3	D24/11/1	K192/11/3	E21/22/1	R201/32/3	R37/11/1	L176/11/2		
						R201/1/3	M34/1/1	K210/1/3	R31/1/1		N55/11/1	P234/1/3		
					R77/31/2	E100/21/1			M34/1/1	K210/2/3	L56/1/1	C237/1/3		
						D97/11/1			D36/11/1					
									R77/22/2	D97/11/1				
										E100/11/1				

See page 168 for key.

Table C: Inter-domain hydrogen bonds in the GST family

Class	Phi GST						Pi GST					
Origin	Arabidopsis Thaliana		Maize Type 1		Maize Type 3		Human		Mouse		Pig	
PDB code	1gnw		1axd		1aw9		1gss		1glp		2gsr	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C
Interdom.H-bonds	5		6		11		13		19		16	
Residue ID/ No.Bonds/ Segment no.	A10/1/1	Y178/1$_1$/2	W12/1/1	S170/1$_1$/2	L11/1$_1$/1	Y175/1$_1$/2	R11/5$_2$/1	P200/1/3	E15/3$_4$/1	H198/1/3	E15/2$_4$/1	R199/1$_1$/3
	R16/1/1	H171/1$_1$/2	E23/3/1	R200/2/3	P13/1$_2$/1	S171/1$_1$/2		Y196/2$_1$/3		R201/3$_1$/3		H196/3$_1$/3
	H22/12/1	K209/1$_1$/3		S202/1$_1$/3	N14/1$_1$/1	Y108/1$_1$/1		L199/1$_1$/3	R11/5/1		R11/5/1	
	R68/22/2	E106/2/1	R68/2$_1$/2	E102/2/1	R17/1$_3$/1	N168/1$_2$/2		I201/2/3		H198/2/3		P200/1/3
					N23/1/1	T207/1$_2$/3	G12/1/1			I203/2/3		V197/1/3
					E24/1$_2$/1	R201/1$_1$/3	R13/1$_1$/1	E95/1$_3$/1	G12/1$_1$/1			R199/1$_1$/3
					E69/2$_1$/2	E103/2$_1$/1	D23/3$_2$/1	R184/1$_2$/3	R13/2/1	E97/2$_2$/1	G12/1/1	I201/1/3
					N72/1$_2$/2	N168/1$_2$/2		Y151/1$_3$/2	D23/3$_2$/1	R186/2$_2$/3	R13/2$_1$/1	E95/2$_3$/1
					R73/2$_4$/2	E97/2$_1$/1		L187/1$_3$/3		Y153/2$_2$/2	D23/3$_1$/1	Y151/1$_1$/2
							N64/1$_1$/2	N91/2$_4$/1	Q24/1$_1$/1			R184/2$_1$/3
							R68/2$_3$/2		N66/2$_1$/2	N93/1$_3$/1	N64/2$_2$/2	N91/1$_2$/1
								D92/1$_2$/1		E97/1$_4$/1		E95/1$_4$/1
									R70/3/2	D90/2$_3$/1	R68/1$_3$/2	D88/1$_4$/1
										D94/1$_3$/1		

See page 168 for key.

Table C: Inter-domain hydrogen bonds in the GST family

Class	Sj GST		Sigma GST								Tau GST			
Origin	Schistosomal		Fruit Fly		Human		Rat		Squid		Rice		Wheat	
PDB code	1gta		1m0u		1iyi		1pd2		2gsq		1oyj		1qwc	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C	N	C
Interdom.H-bonds	6		6		11		12		6		6		7	
Residue ID/ No.Bonds/ Segment no.	K11/3$_1$/1	D160/1$_1$/2	K58/3$_2$/1	V249/2/3	R12/3/1	T159/1/2	N10/1$_1$/1	T197/2/3	M11/1$_1$/1	F202/2/3	R20/3$_3$/1	D112/1$_1$/1	P14/1/1	W171/1$_1$/2
		Y198/1/3		T247/1/3		T197/1/3	R12/3/1		G12/1/1			A168/2$_1$/2	R20/4$_1$/1	D106/2$_1$/1
		P202/1/3	E62/2$_4$/1	W240/1$_1$/3		K198/1/3		K198/1/3	R13/2$_1$/1	N99/2$_1$/1	E27/1$_2$/1	S206/1$_1$/3		A164/2$_1$/2
	G12/1/1	L203/1/3		R244/1$_1$/3	G13/1/1	L199/1$_3$/3		L199/2/3	E15/2$_3$/1	R197/1$_3$/3	Q74/2$_2$/2	R105/2$_1$/1	Q74/2$_2$/2	R99/2/1
	R73/2/2	S93/1$_1$/1	R115/1$_3$/2	D135/1$_2$/1	R14/2$_1$/1	D96/2$_5$/1	G13/1/1			Y193/1$_1$/3				
		E96/1$_2$/1			E16/2$_3$/1	W190/1$_2$/3	R14/1/1	D96/1$_2$/1						
						R194/2$_1$/3	E16/3$_1$/1	W190/1$_3$/3						
					Y20/2$_2$/1			R194/3$_1$/3						
						E154/1$_3$/2	W20/2$_2$/1							
					R69/1$_1$/2	D89/1$_2$/1		D154/1$_3$/2						
							R69/1$_2$/2	D89/1$_2$/1						

See page 168 for key.

Table C: Inter-domain hydrogen bonds in the GST family

Class	Theta GST		Unknown		Ure2p		Zeta GST			
Origin	Human		Nematode		Yeast		Arabidopsis Thaliana		Human	
PDB code	1ljr		1tw9		1g6w		1e6b		1fw1	
Interface ID	N	C	N	C	N	C	N	C	N	C
Interdom.H-bonds	8		10		5		4		7	
Residue ID/ No.Bonds/ Segment no.	V10/1$_1$/1	Q171/2$_1$/2	R12/1$_1$/1	T204/1$_1$/3	R120/1/1	G353/1/3	W15/1/1	D215/2/3	F12/1/1	D209/1/3
	R15/3$_2$/1		G13/1$_1$/1	F206/1/3	N124/11/1	W316/1/2	Y36/1/1		S15/1$_2$/1	N172/1$_2$/2
		Q175/2/2	E16/2$_2$/1	W197/1$_3$/3	E134/21/1	R344/2$_1$/3	R24/22/1	A208/2$_1$/3	W18/1$_3$/1	S202/3$_1$/3
	K22/2$_2$/1	E208/2$_2$/3		R201/1$_3$/3	Y139/11/1	K349/1/3			R21/2$_2$/1	
	K23/1/1	F200/1/3	Q20/1$_6$/1	E162/1$_4$/2					E77/2$_3$/2	R100/2$_2$/1
	K37/1/1	S229/1$_2$/3	L24/1$_2$/1	R193/1$_1$/3						
			Q65/2$_1$/2	R69/1$_3$/1						
				A92/1$_4$/1						
			R69/1$_3$/2	D89/1$_3$/1						
			K73/1$_1$/2	S85/1$_4$/1						

See page 168 for key.

Table D: Inter-domain salt-bridges in the GST-family

Class	**Alpha**													
Origin	***Fasciola hepatica***		**Human A1-1**		**Mouse A1-1**		**Mouse A2-2**		**Mouse A4-4**		**Rat A1-1**		***Schistosoma haematobium***	
PDB code	**2fhe**		**1k3o**		**1f3b**		**1ml6**		**1guk**		**1ev9**		**1oe8**	
Interface ID	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**
Total Interdom. Salt bridges	9		11		8		10		7		8		7	
Residue ID/ No.Bonds/ Segment no.	R10/2/1	E156/2/2	R13/3/1	E169/3/2	R12/4/1	E168/4/2	R12/4/1	E168/4/2	R13/3/1	E169/3/2	R13/4/1	E169/42	R14/3/1	D172/3/2
	R10/2/1	D159/2/2	R15/4/1	E104/4/1	E31/2/1	K204/2/3	R14/2/1	E103/2/1	E32/2/1	K205/2/3	E32/2/1	K205/2/3	R16/1/1	E103/1/1
	R34/1/1	D213/1/3	E32/2/1	K205/2/3	R68/2/2	E96/2/1	E31/2/1	K204/2/3	R69/2/2	E93/2/1	R69/2/2	E97/2/1	E18/1/1	R206/1/3
	E21/2/1	R196/2/3	R69/2/2	E97/2/1			R68/2/2	E96/2/1					R76/2/2	E96/2/1
	R34/2/1	D213/2/3												

See page 168 for key.

Table D: Inter-domain salt-bridges in the GST-family

Class	**Beta GST**				**CLIC1**		**CLIC4**		**Delta GST**							
Origin	***Proteus Mirabilus***		***E.coli***		**Human**		**Human**		**Mosquito 1-3**		**Mosquito 1-4**		**Mosquito 1-5**		**Mosquito 1-6**	
PDB code	**1pmt**		**1aof**		**1k0m**		**2ahe**		**1jlv**		**1jlw**		**1r5a**		**1v2a**	
Interface ID	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**
Total Interdom. Salt bridges	7		5		3		0		2		2		1		1	
Residue ID/ No.Bonds/ Segment no.	E20/3/1	R188/3/3	*E20/2/1*	R188/2/3	K20/1/1	E228/1/3			R66/1/2	D100/1/1	R67/1/2	D106/1/1	R68/1/2	D102/1/1	K20/1/1	D190/1/3
	D75/4/2	R91/4/1	D75/3/2	R91/3/1	K20/2/1	D225/2/3			E74/1/2	R90/1/1	E75/1/2	R96/1/1				

See page 168 for key.

Table D: Inter-domain salt-bridges in the GST-family

Class	**Elongation factor**		**Grx2**		**Mu GST**						**Omega GST**	
Origin	***Saccharomyces cerevisiae***		***E.coli***		**Chicken**		**Human**		**Rat**		**Human**	
PDB code	**1nhy**		**1g7o**		**1gsu**		**1gtu**		**2gst**		**1eem**	
Interface ID	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**
Total Interdom. Salt bridges	2		0		7		3		13		2	
Residue ID/ No.Bonds/ Segment no.	R18/1/1	E204/1/3			R10/3/1	D164/2/2	R10/2/1	D161/1/2	R10/3/1	D161/1/2	E36/2/1	R183/2/2
	K22/1/1	D203/1/3				D161/1/2		D164/1/2		D164/2/2		
					E21/1/1	R201/1/3	D24/1/1	K192/1/3	E21/2/1	R201/2/3		
					R77/3/2	D97/1/1			E21/2/1	R201/2/3		
						E100/2/1			D24/1/1	D192/1/3		
									D36/1/1	K210/1/3		
									R77/4/2	D97/2/1		
										E100/2/1		

See page 168 for key.

Table D: Inter-domain salt-bridges in the GST-family

Class	**Pf GST**		**Phi GST**						**Pi GST**					
Origin	***Plasnodium falcipar***		***Arabidopsis thaliana***		**Maize Type 1**		**Maize Type 3**		**Human**		**Mouse**		**Pig**	
PDB code	**1pa3**		**1gnw**		**1axd**		**1aw9**		**1gss**		**1glp**		**2gsr**	
Interface ID	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**
Total Interdom. Salt bridges	2		8		5		9		9		8		9	
Residue ID/ No.Bonds/ Segment no.	R13/1/1	D170/1/2	E23/2/1	R204/2/3	R16/1/1	E102/1/1	R17/1/1	E103/1/1	R13/4/1	E95/4/1	R13/2/1	E97/2/1	R13/2/1	E95/2/1
	R77/1/2	D97/1/1	R16/3/1	D111/1/2	E23/2/1	R200/2/3	E24/2/1	R201/2/3	*D23/1/1*	R184/1/3	E15/2/1	R201/2/3	E15/2/1	R199/2/3
				E106/3/1	R68/2/2	E102/2/1	E69/2/2	E103/2/1	R68/4/2	D88/2/1	D23/2/1	R186/2/3	D25/2/1	K186/23
			R68/5/2				R73/4/2	E97/4/1		D92/2/1	R70/2/2	D90/1/1	R68/3/2	D88/1/1
				D107/2/1								D94/1/1		D92/2/1
				D111/1/1										

See page 168 for key.

Table D: Inter-domain salt-bridges in the GST-family

Class	**Sj GST**		**Sigma GST**								**Tau GST**			
Origin	***Schistosomal***		**Fruit Fly**		**Human**		**Rat**		**Squid**		**Rice**		**Wheat**	
PDB code	**1gta**		**1m0u**		**1iyi**		**1pd2**		**2gsq**		**1oyj**		**1gwc**	
Interface ID	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**
Total Interdom. Salt bridges	4		2		5		6		2		3		2	
Residue ID/ No.Bonds/ Segment no.	K11/1/1	D160/1/2	*E62/1/1*	R244/1/3	R14/3/1	D96/3/1	R14/4/1	D96/4/1	E15/1/1	R197/1/3	R20/3/1	D112/3/1	R20/2/1	D106/2/1
	R35/1/1	D214/1/3	R115/1/2	D135/1/1	E16/1/1	R194/1/3	E16/1/1	R194/1/3	R68/1/1	D88/1/1				
	R73/2/2	E96/2/1			R69/1/2	D89/1/1	R69/1/2	D89/1/1						

See page 168 for key.

Table D: Inter-domain salt-bridges in the GST-family

Class	**Theta GST**		**Unknown**		**Ure2p**		**Zeta GST**			
Origin	**Human**		**Nematode**		**Yeast**		***Arabidopsis thaliana***		**Human**	
PDB code	**1ljr**		**1tw9**		**1g6w**		**1e6b**		**1fw1**	
Interface ID	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**
Total Interdom. Salt bridges	3		3		2		0		2	
Residue ID/ No.Bonds/ Segment no.	R15/1/1	Q171/1/2	R12/1/1	D166/1/2	E134/2/1	R344/2/3			E77/2/2	R100/2/1
	K22/2/1	E208/2/3	D26/1/1	R193/1/3						
			R69/1/2	D89/1/1						

See page 168 for key.

Table E: Domain interface properties of the GST protein family

Class	Alpha													
Origin	*Fasciola hepatica*		Human A1-1		Mouse A1-1		Mouse A2-2		Mouse A4-4		Rat A1-1		*Schistosoma Haematobium*	
PDB code	2fhe		1k3o		1f3b		1ml6		1guk		1ev9		1oe8	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C	N	C
Number of residues in subunit	216		198		222		219		215		219		204	
Number of residues in domain	77	132	78	114	77	137	76	136	71	134	77	135	81	119
Interface Accessible surface area per domain	1175.0	1110.6	951.3	990.9	1112.5	1167.3	1107.6	1151.6	1337.0	1329.7	1131.3	1209.9	795.3	765.7
% Interface Accesible surface area per domain	21.7	13.6	17.6	12.6	21.7	12.2	21.5	12.2	26.3	15.2	21.8	13.2	14.6	10.3
Planarity	3.2	3.3	3.5	3.6	4.2	4.4	4.0	4.3	4.0	4.2	3.2	4.0	2.4	2.5
Length	35.2	41.8	29.3	31.4	33.3	38.7	33.8	37.5	37.9	37.9	33.8	37.2	28.9	33.0
Breath	21.9	24.5	25.2	30.4	22.1	24.4	20.7	25.1	26.7	28.4	22.4	24.1	19.4	24.1
Length/Breath ratio	0.5	0.6	0.9	0.9	0.7	0.7	0.6	0.8	0.6	0.4	0.6	0.7	0.7	0.7
Interface residue segments	2	3	2	3	2	3	2	3	2	3	2	3	2	3
% Polar residues at interface	38	30	42	30	44	35	39	36	42	31	38	40	38	37
% Non-polar residues at interface	62	70	58	70	56	65	61	64	58	69	62	60	62	63
Gap volume	3936.6		5044.7		5359.6		5482.1		3732.5		4344.9		4968.9	
Gap volume index	1.7		2.6		2.4		2.4		1.4		1.9		3.2	
Bridging water molecules	0		0		0		0		0		0		0	

See page 168 for key.

Table E: Domain interface properties of the GST protein family

Class	Beta GST				CLIC1		CLIC4		Delta GST								Elongation factor		Grx2	
Origin	*Proteus mirabilus*		*E.coli*		Human		Human		Mosqito 1-3		Mosquito 1-4		Mosquito 1-5		Mosquito 1-6		*Saccharomyces cerevisiae*		*E.coli*	
PDB code	1pmt		1aof		1k0m		2ahe		1jlv		1jlw		1r5a		1v2a		1nhy		1g7o	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C	N	C	N	C	N	C	N	C
Number of residues in subunit	201		201		235		257		207		217		214		208		215		215	
Number of residues in domain	75	114	75	111	87	139	98	145	76	133	76	125	76	127	74	123	70	129	72	132
Interface Accessible surface area	942.6	940.8	916.2	887.1	913.8	872.1	914.9	832.7	913.9	780.0	839.6	714.9	917.3	788.2	980.6	855.0	530.2	500.6	1018.1	970.2
% Interface Accesible surface area	20.3	13.6	20.2	12.3	16.5	10.6	16.9	10.4	20.3	11.2	18.4	9.7	19.5	10.6	20.9	11.1	11.5	6.7	22.7	13.4
Planarity	2.5	2.8	2.3	2.6	2.1	2.1	2.2	2.0	2.4	2.3	2.5	2.4	2.2	2.4	2.2	2.5	2.0	2.0	2.8	3.2
Length	27.9	28.8	27.7	28.1	33.9	31.2	34.9	30.8	33.7	34.5	33.2	35.0	31.8	34.9	31.6	35.7	27.0	30.6	31.6	39.9
Breath	25.6	23.6	25.2	23.7	21.6	21.6	21.5	22.9	22.2	23.1	19.9	21.7	22.4	25.2	22.0	25.4	17.6	21.2	17.3	19.9
Length/Breath ratio	0.8	0.7	0.8	0.7	0.7	0.7	0.3	0.7	0.7	0.8	0.7	0.8	0.8	0.9	0.3	0.8	0.7	0.8	0.5	0.5
Interface residue segments	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
% Polar residues at interface	38	35	35	34	31	31	32	29	27	27	29	32	25	26	26	29	17	43	25	28
% Non-polar residues at interface	62	65	65	66	69	69	68	71	73	73	71	68	75	74	74	71	83	57	75	72
Gap volume	4228.1		3668.3		3382.6		3032.7		3470.1		3948.3		4064.9		4292.4		3959.9		3248.4	
Gap volume index	2.2		2.0		1.9		1.7		2.1		2.5		2.4		2.3		3.8		1.7	
Bridging water molecules	0		0		0		0		0		0		0		0		0		0	

See page 168 for key.

Table E: Domain interface properties of the GST protein family

Class	Mu GST						Omega GST		Pf GST		Phi GST					
Origin	Chicken		Human		Rat		Human		*Plasnodium falcipar*		*Arabidopsis thaliana*		Maize Type 1		Maize Type 3	
PDB code	1gsu		1gtu		2gst		1eem		1pa3		1gnw		1axd		1aw9	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C	N	C	N	C
Number of residues in subunit	217		217		217		237		196		217		209		216	
Number of residues in domain	81	128	81	128	82	128	92	135	78	111	76	119	77	122	77	125
Interface Accessible surface area	1365.6	1228.3	1302.5	1263.8	1353.4	1283.3	1206.1	1137.7	801.2	764.3	871.5	876.0	831.5	770.7	896.9	830.0
% Interface Accesible surface area	24.4	15.3	24.1	15.4	24.3	15.8	21.0	14.0	14.8	10.6	17.8	12.1	17.5	11.3	17.4	11.7
Planarity	3.3	3.2	3.0	3.1	2.9	3.0	3.3	3.0	2.3	2.4	2.4	2.3	2.3	2.3	2.1	2.3
Length	38.0	35.4	38.4	35.1	39.3	34.9	35.9	36.9	26.9	28.0	35.7	29.8	25.2	32.9	31.1	32.1
Breath	24.2	29.6	21.9	28.8	24.3	29.2	22.4	23.0	23.6	28.6	23.9	26.1	19.8	24.8	21.3	22.6
Length/Breath ratio	0.6	0.8	0.5	0.7	0.5	0.8	0.6	0.6	0.9	0.8	0.8	0.8	0.8	0.3	0.8	0.8
Interface residue segments	2	3	2	3	2	3	2	3	2	3	3	3	2	3	2	3
% Polar residues at interface	38	27	37	25	42	30	39	32	36	39	39	39	31	28	39	31
% Non-polar residues at interface	62	73	63	75	58	70	61	68	64	61	61	61	69	72	61	69
Gap volume	3901.6		3919.1		3761.6		3924.9		3959.9		3557.5		3551.7		4123.2	
Gap volume index	1.5		1.5		1.4		1.7		2.5		2.0		2.2		2.4	
Bridging water molecules	0		0		0		0		0		0		0		0	

See page 168 for key.

Table E: Domain interface properties of the GST protein family

Class	Pi GST						Sj GST		Sigma GST							
Origin	Human		Mouse		Pig		*Schistosomal*		Fruit Fly		Human		Rat		Squid	
PDB code	1gss		1glp		2gsr		1gta		1m0u		1iyi		1pd2		2gsq	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C	N	C	N	C
Number of residues in subunit	209		209		207		218		203		198		199		202	
Number of residues in domain	76	127	76	127	73	127	77	133	75	122	71	118	72	118	74	122
Interface Accessible surface area	1182.4	1019.3	1140.1	997.6	1117.6	1016.1	1191.6	1157.0	1075.6	987.3	1004.9	940.3	979.4	910.3	942.7	912.6
% Interface Accesible surface area	23.1	13.9	23.3	13.2	22.0	13.6	21.8	14.2	21.4	12.9	20.3	12.3	19.4	12.2	20.1	11.8
Planarity	2.6	2.8	2.8	2.7	2.9	2.8	2.9	3.2	2.8	2.6	2.9	2.9	2.8	2.7	2.5	2.5
Length	26.6	32.6	31.0	30.8	32.9	32.1	35.7	37.7	30.3	31.8	26.9	28.9	26.5	29.1	24.8	29.0
Breath	23.2	29.3	22.6	29.8	24.4	30.8	25.0	25.1	23.0	25.2	21.7	25.9	20.5	26.0	22.0	31.2
Length/Breath ratio	0.9	0.3	0.9	0.9	0.8	0.9	0.6	0.7	0.7	0.7	0.8	0.4	0.7	0.8	0.4	0.9
Interface residue segments	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
% Polar residues at interface	44	34	42	39	44	39	34	33	28	35	40	30	42	31	23	35
% Non-polar residues at interface	56	66	58	61	56	61	66	67	72	65	60	70	58	69	77	65
Gap volume	3965.8		4169.9		4403.2		3837.5		3843.3		4578.1		4992.2		3860.8	
Gap volume index	1.8		2.0		2.1		1.6		1.9		2.4		2.6		2.1	
Bridging water molecules	0		0		0		0		0		0		0		0	

See page 168 for key.

Table E: Domain interface properties of the GST protein family

Class	Tau GST				Theta GST		Unknown		Ure2p		Zeta GST			
Origin	Rice		Wheat		Human		Nematode		Yeast		*Arabidopsis thaliana*		Human	
PDB code	1oyj		1gwc		1ljr		1tw9		1g6w		1e6b		1fw1	
Interface ID	N	C	N	C	N	C	N	C	N	C	N	C	N	C
Number of residues in subunit	228		221		244		191		234		194		208	
Number of residues in domain	77	134	76	133	77	156	75	110	97	128	76	108	78	120
Interface Accessible surface area	936.9	866.7	903.9	800.0	1131.7	1097.8	999.0	955.4	760.4	742.6	997.5	1012.7	1163.8	1136.5
% Interface Accesible surface area	18.4	10.5	18.4	10.0	23.5	12.1	20.4	13.5	12.8	9.2	19.8	13.7	22.8	15.1
Planarity	2.3	2.4	2.2	2.4	3.4	3.7	2.9	2.8	2.5	2.7	3.6	3.7	3.6	3.7
Length	31.4	35.0	33.4	35.3	34.7	39.1	31.3	32.3	34.6	28.1	27.0	33.4	37.3	35.9
Breath	21.2	24.2	24.4	23.6	22.2	26.1	25.8	27.5	18.3	24.8	20.1	23.3	20.1	27.1
Length/Breath ratio	0.7	0.8	0.6	0.7	0.5	0.6	0.8	0.4	0.7	0.8	0.6	0.7	0.6	0.8
Interface residue segments	2	3	2	3	2	3	2	3	2	3	2	3	2	3
% Polar residues at interface	32	28	28	36	29	26	48	37	26	30	27	26	30	32
% Non-polar residues at interface	68	72	72	64	71	74	52	63	74	70	73	73	70	68
Gap volume	4560.6		3779.2		4117.4		3195.9		5167.2		3971.6		4887.2	
Gap volume index	2.5		2.2		1.9		1.6		3.4		2.0		2.1	
Bridging water molecules	0		0		0		0		0		0		0	

See page 168 for key.

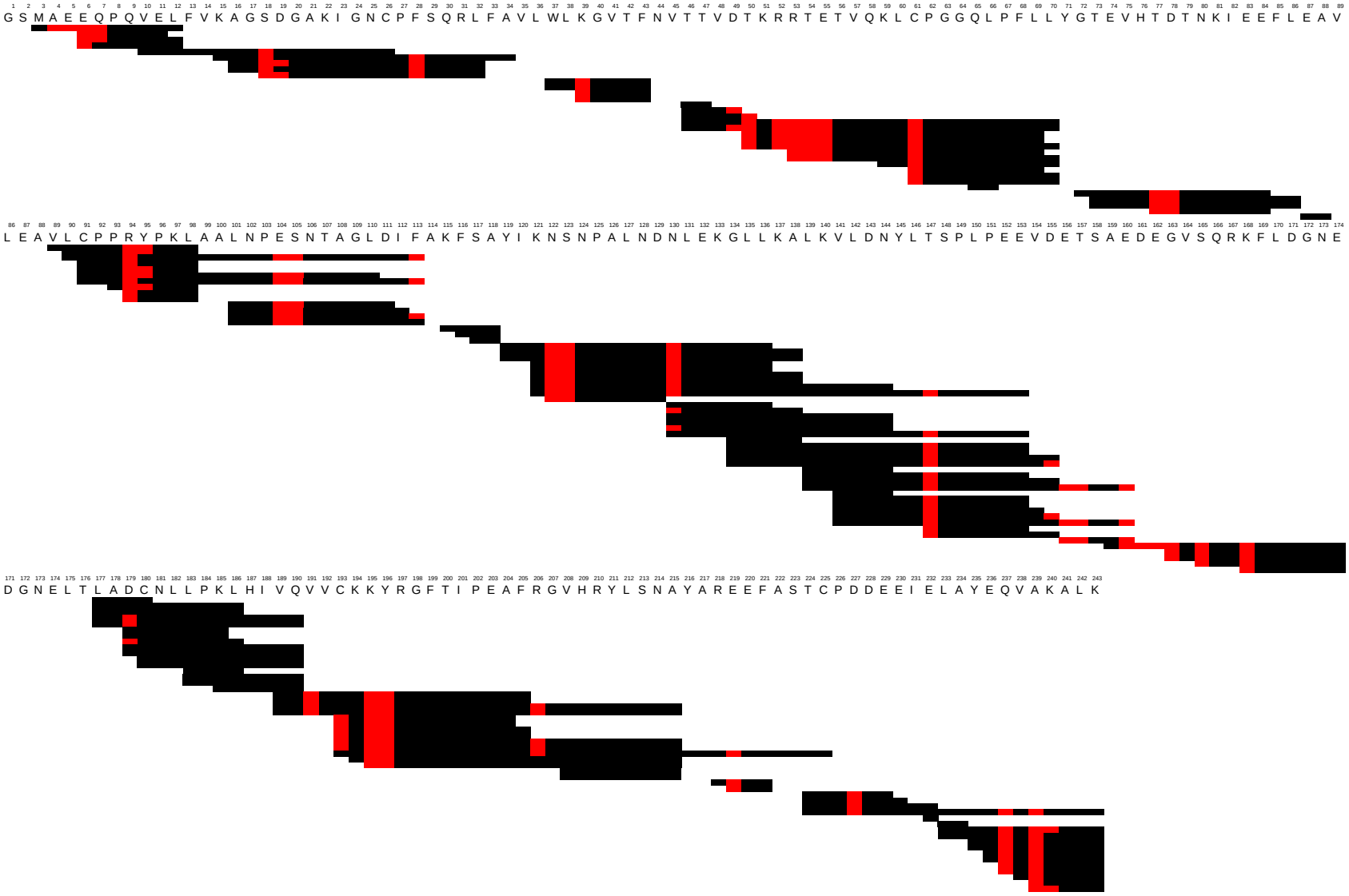
Key to Tables F1 – F3 and G1 – G6: CLIC1-M32A and CLIC1-E81M consensus peptide maps and deuterium localization

Tables F1 – F3: CLIC1-M32A consensus peptide maps and deuterium localization.

Tables G1 – G6: CLIC1-E81M consensus peptide maps and deuterium localization.

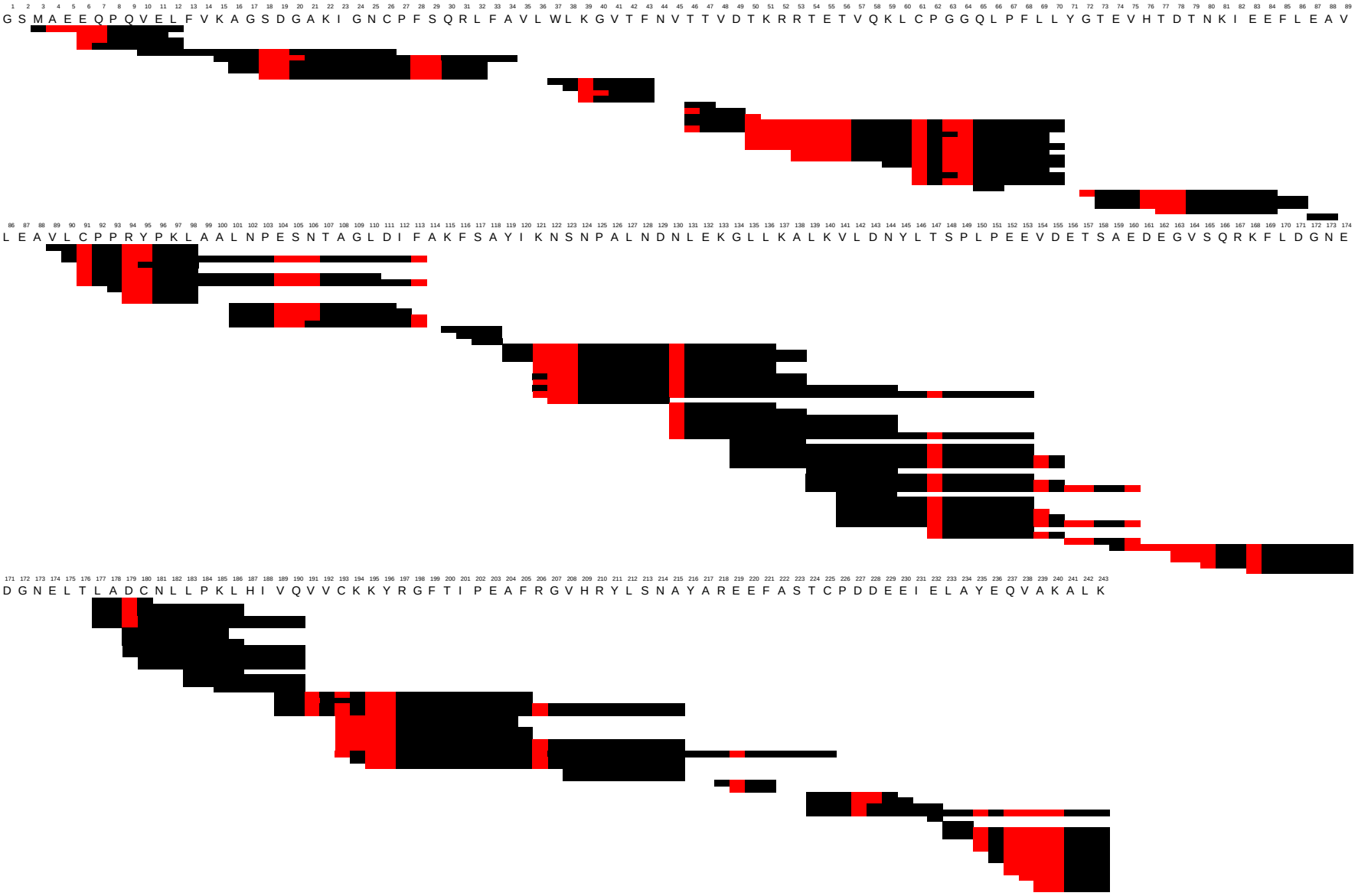
- The sequence and numbering of CLIC1-M32A and CLIC1-E81M include two additional residues (Glycine and Serine) at the N-terminus of the protein. These amino acids form part of the thrombin cleavage site of the GST-CLIC1 fusion construct.
- Peptides, shown in black, are mapped against the sequence of CLIC1-M32A or CLIC1-E81M.
- The first two amino-acids of each peptide are not shown on the maps since deuterium incorporated at these positions is always lost during the chromatographic separation of the peptides.
- Possible sites of deuterium localization are indicated in red.

Table F1: CLIC1-M32A peptide map showing deuterium sub-localization at 30 seconds



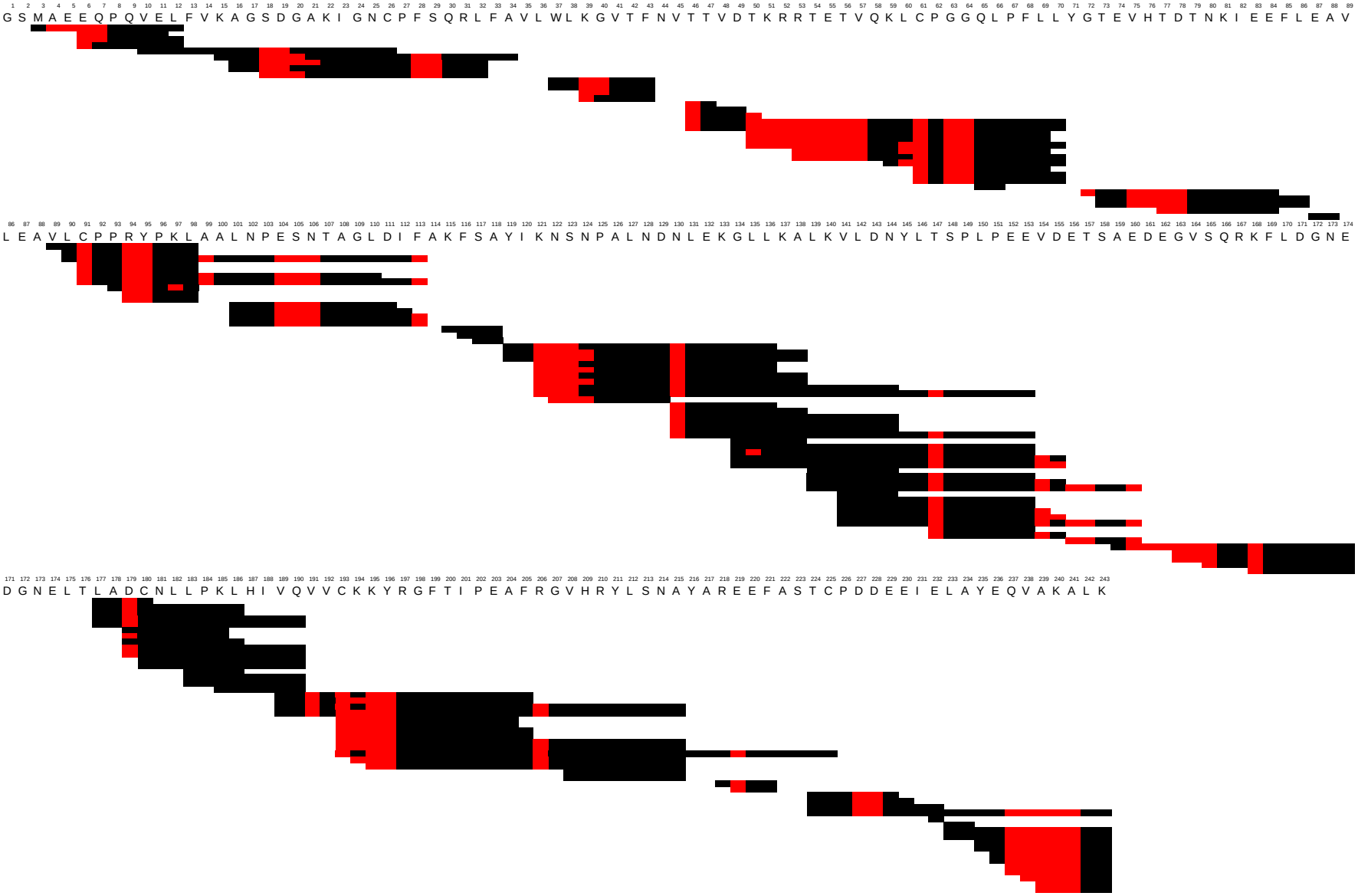
See page 186 for key.

Table F2: CLIC1-M32A peptide map showing deuterium sub-localization at 300 seconds



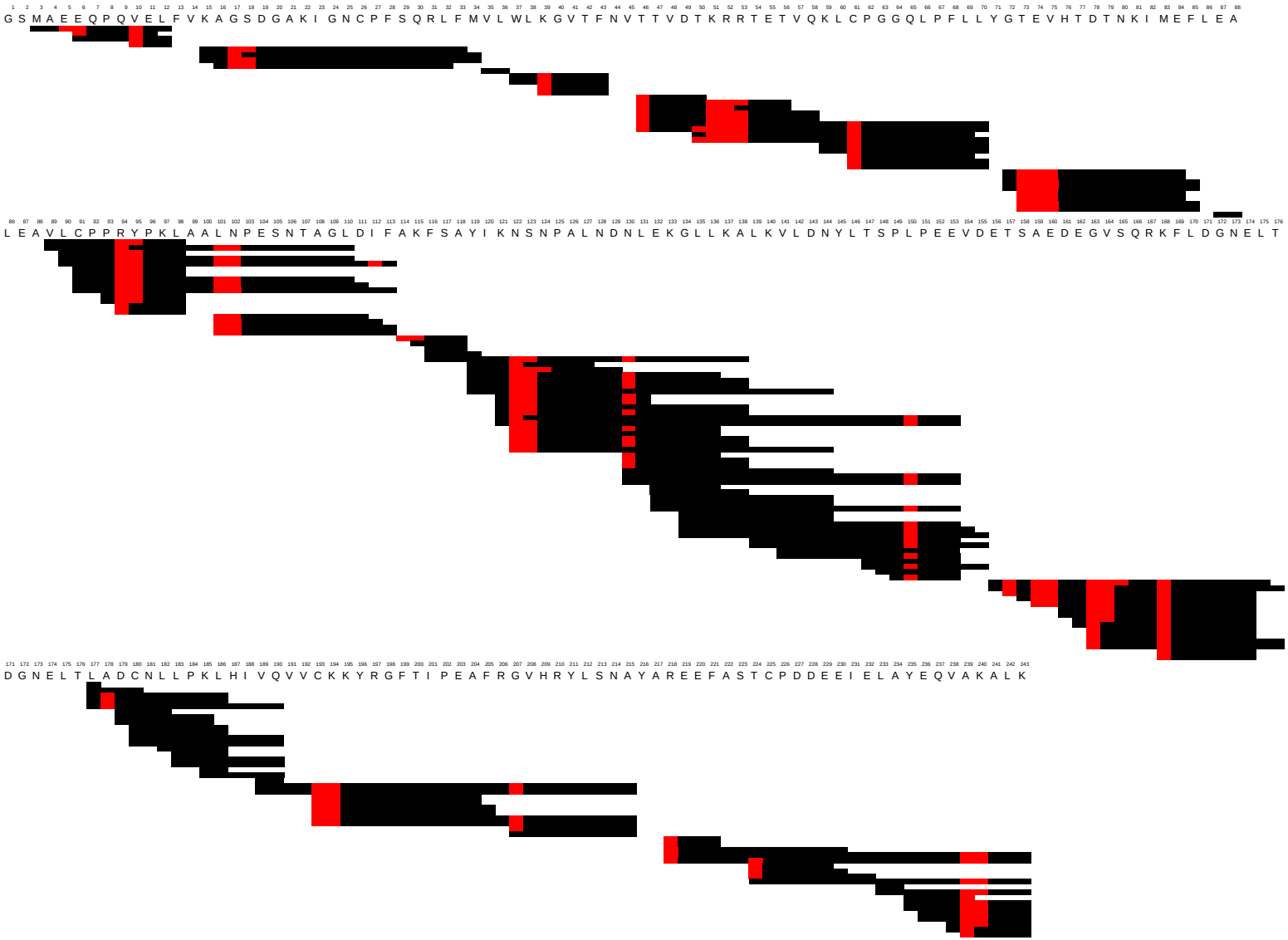
See page 186 for key.

Table F3: CLIC1-M32A peptide map showing deuterium sub-localization at 1000 seconds



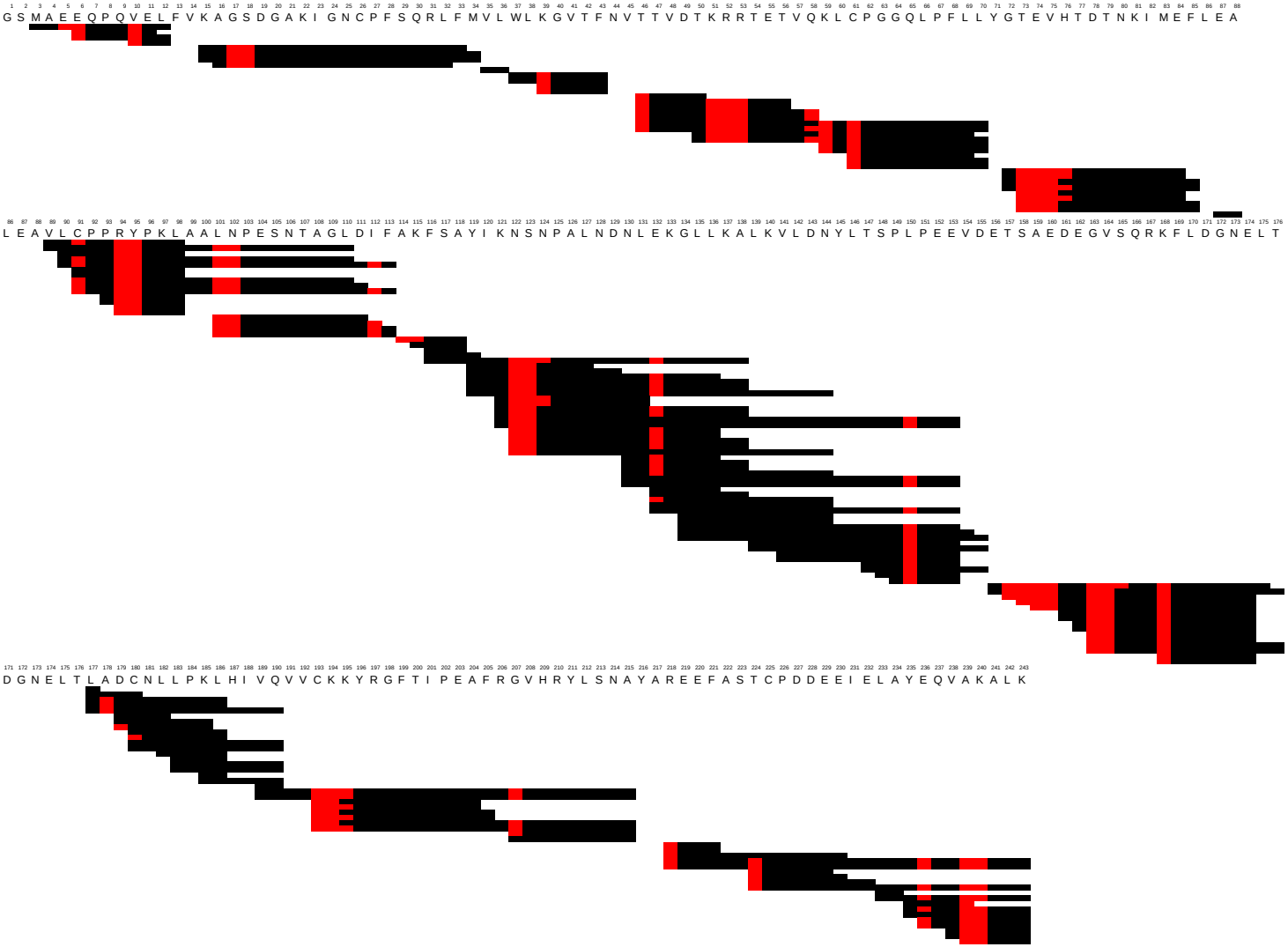
See page 186 for key.

Table G1: CLIC1-E81M peptide map showing deuterium sub-localization at 10 seconds



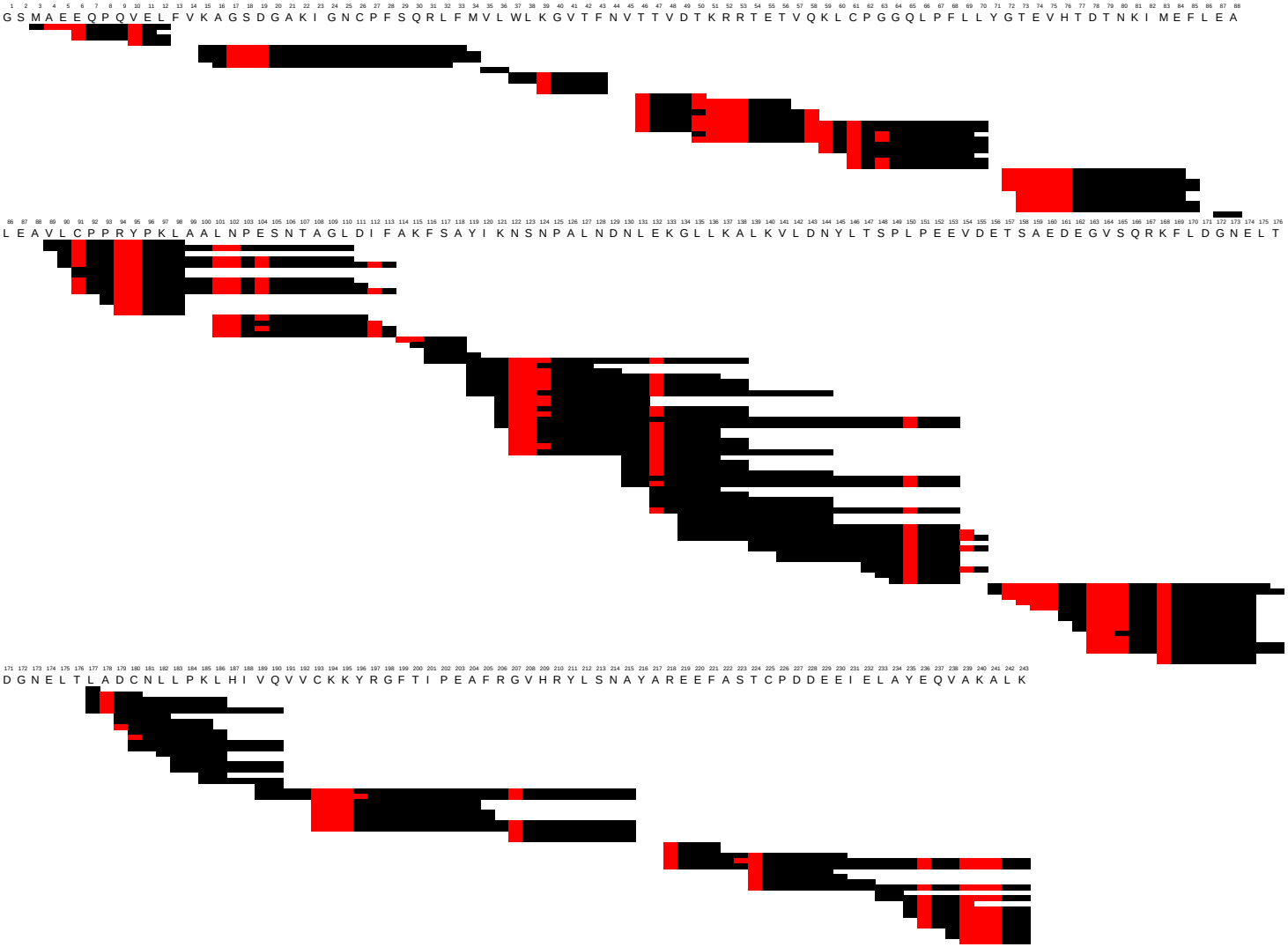
See page 186 for key.

Table G2: CLIC1-E81M peptide map showing deuterium sub-localization at 30 seconds



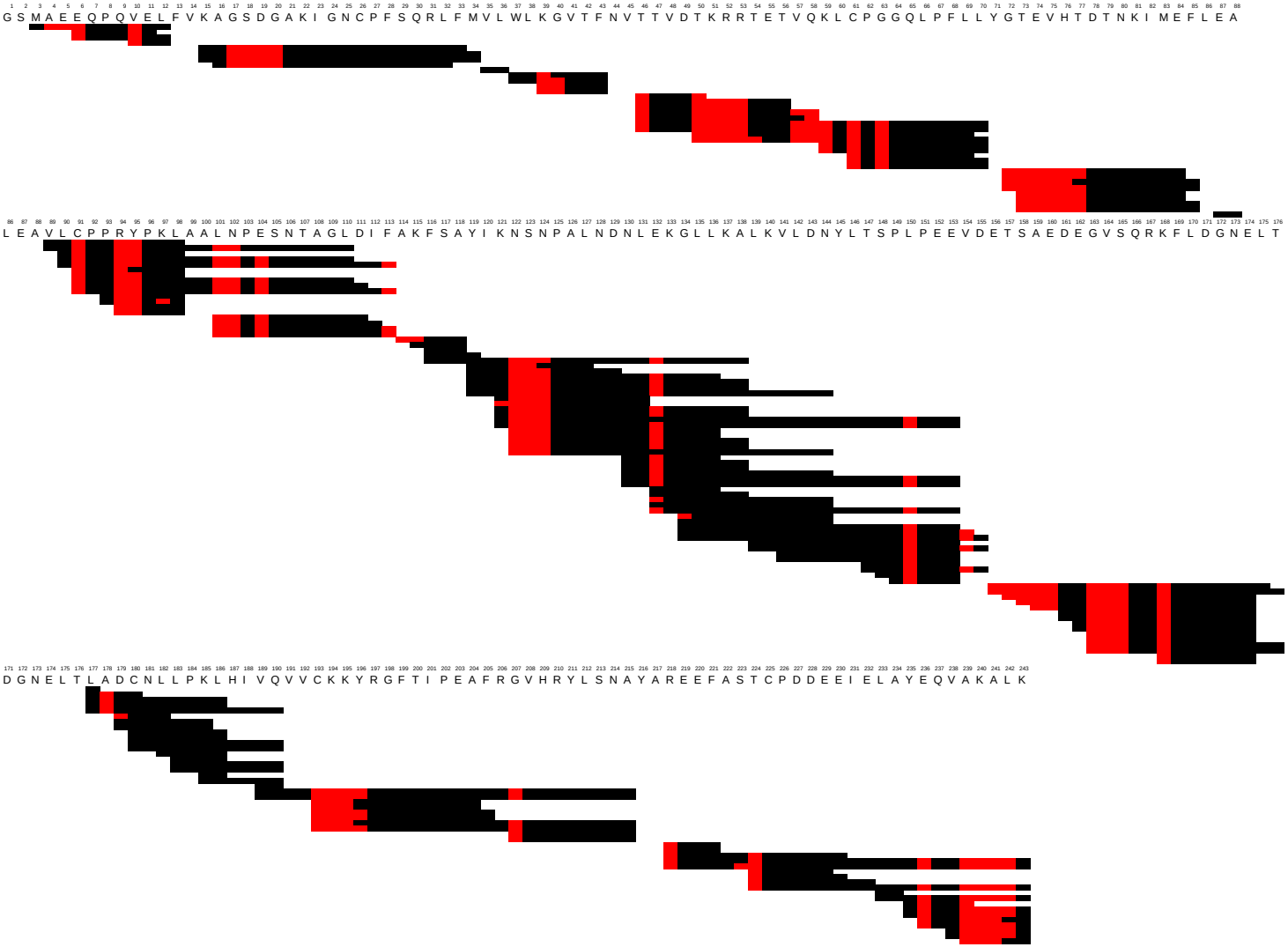
See page 186 for key.

Table G3: CLIC1-E81M peptide map showing deuterium sub-localization at 100 seconds



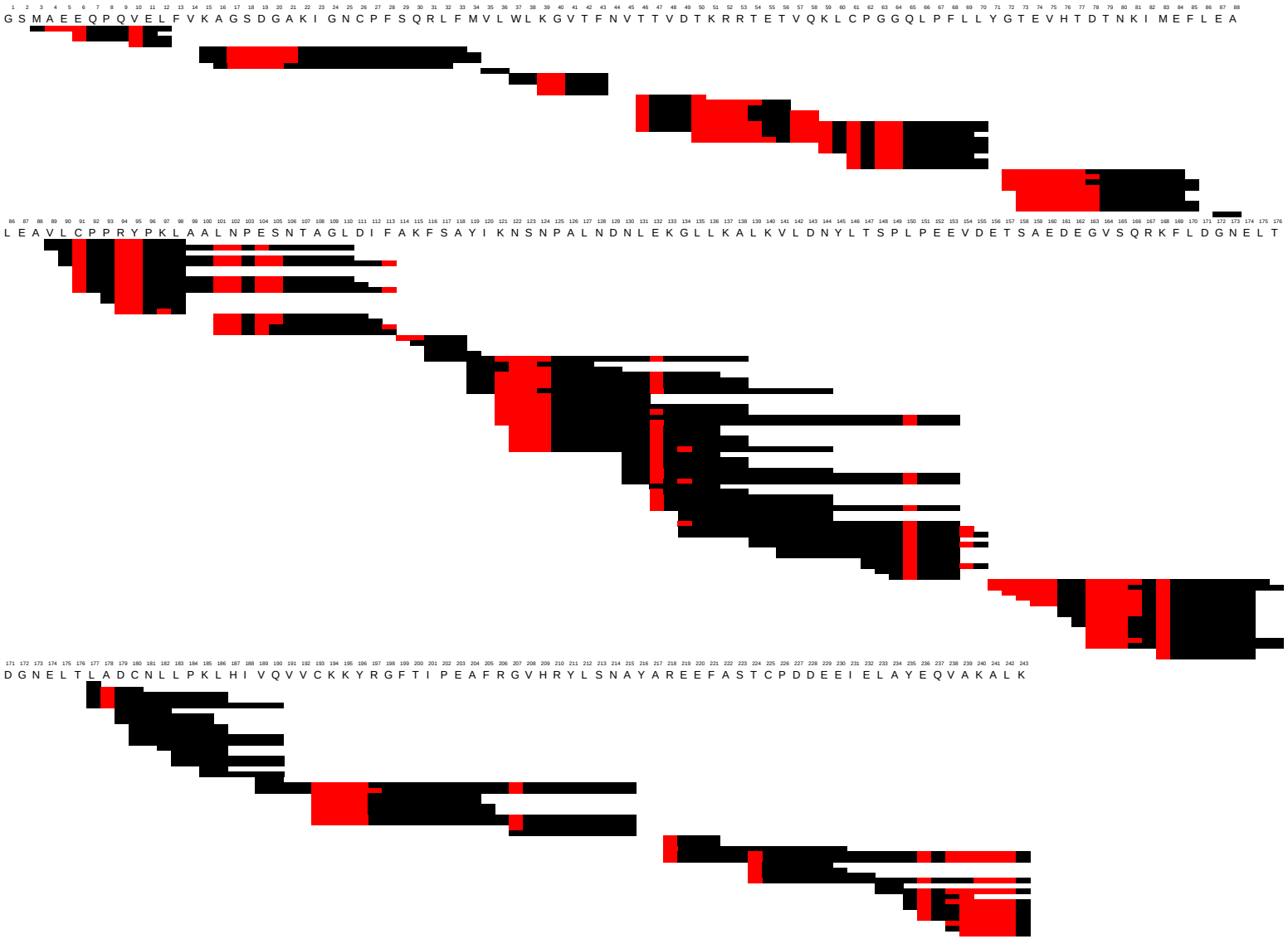
See page 186 for key.

Table G4: CLIC1-E81M peptide map showing deuterium sub-localization at 300 seconds



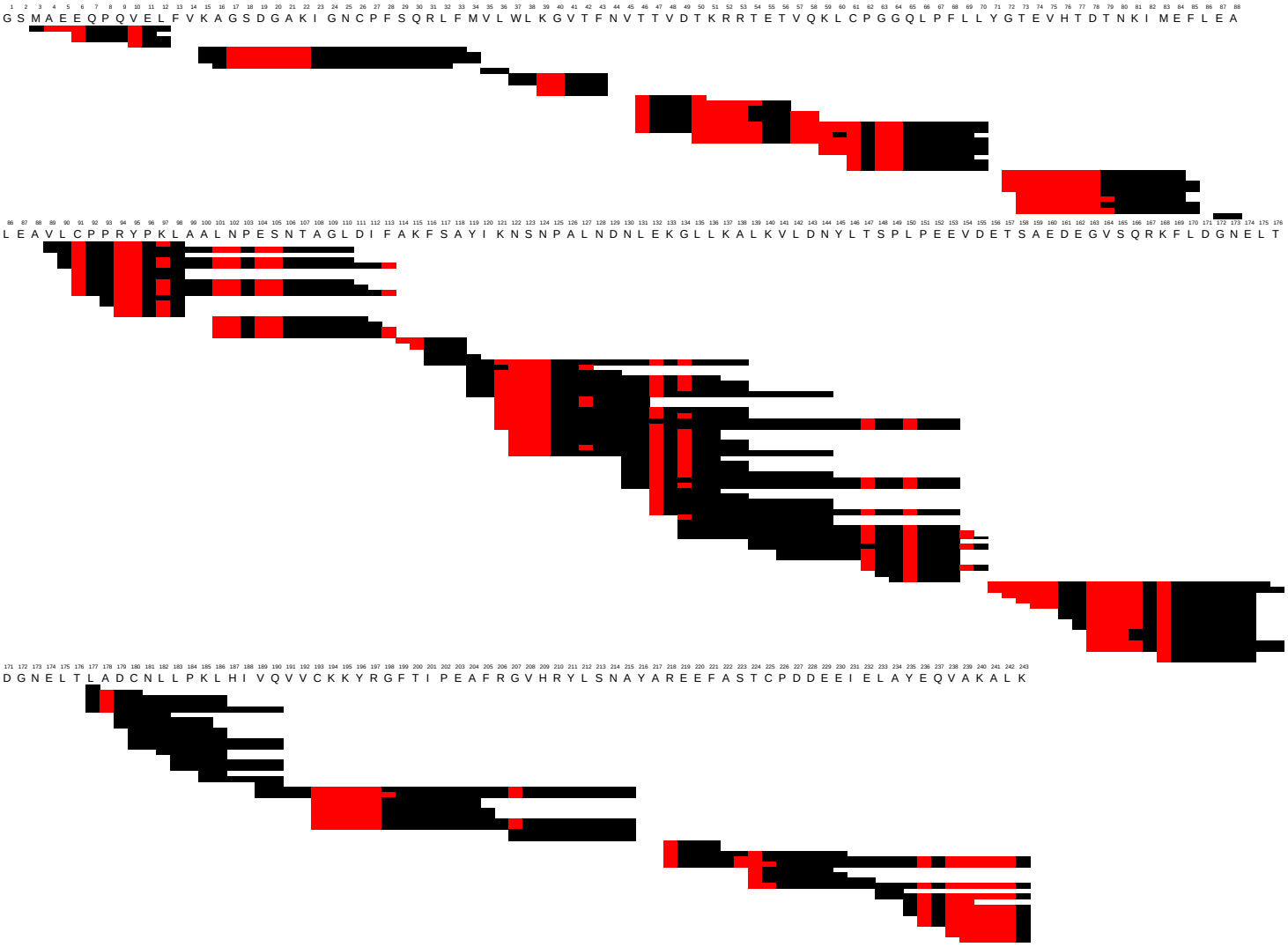
See page 186 for key.

Table G5: CLIC1-E81M peptide map showing deuterium sub-localization at 1000 seconds



See page 186 for key.

Table G6: CLIC1-E81M peptide map showing deuterium sub-localization at 3000 seconds



See page 186 for key.

Key to Table H: Equilibrium pulse-labeling DXMS parameters

Table H: Summary of parameters obtained from PeakFit-deconvoluted mass spectra of deuterium-pulse-labeled wtCLIC1 and CLIC1-M32A.

- Values in brackets indicate standard deviation of replicates

(a) m/z values are determined from centroids of deuterium-labeled CLIC1 spectra deconvoluted using PeakFit (AISN Software. Inc.).

(b) The average mass of each conformation is calculated using the following formula: [**(m/z*24+) - 24+**], where **24+** represents the 24th charge state of CLIC1 and CLIC1-M32A.

(c) **# D2O (adj.) = (Mpd - Mnd)/(Mfd-Mnd)*N**, where **Mpd** is the mass of the partially deuterated states (Native/Intermediate/Unfolded), **Mnd** is the mass of the non-deuterated control and is **Mfd** the mass of the fully-deuterated control (also see section **2.2.12.2.3**). **N** represents the total number of backbone amides (243 in case of CLIC1) minus total number of proline residues (14 in case of CLIC1). **Mnd** values for wCLIC1 pH 7.0, wtCLIC1 pH 5.5 and CLIC1-M32A pH 7.0 are 27,065.38 Da, 27,067.20 Da and 27,006.83 Da, respectively. **Mfd** values for wtCLIC1 pH 7.0, wtCLIC1 pH 5.5 and CLIC1-M32A pH 7.0 were 27,153.66 Da, 27,180.43 Da and 27,150.38 Da, respectively.

(d) **% D2O (adj.) = # D2O (adj.)/229**. The calculation does not include Proline residues.

(e) The % population is calculated from the Gaussian areas of deconvoluted peaks from deuterium-labeled CLIC1 spectra. Spectra were deconvoluted using PeakFit (AISN Software. Inc.).

Table H: Summary of parameters obtained from PeakFit-deconvoluted mass spectra of deuterium-pulse labelled wtCLIC1 and CLIC1-M32A

	Native					Intermediate					Unfolded					Fully-deuterated
Urea (M)	m/z [a]	Mass (Da) [b]	# D_2O (adj.) [c]	% D_2O (adj.) [d]	% population [e]	m/z [a]	Mass (Da) [b]	# D_2O (adj.) [c]	% D_2O (adj.) [d]	% population [e]	m/z [a]	Mass (Da) [b]	# D_2O (adj.) [c]	% D_2O (adj.) [d]	% population [e]	% population [e]
								wild-type CLIC1 pH 7.0								
0 M	1129.57 (0.04)	27085.8 (1.0)	53 (2.3)	23.4% (1.0%)	67.0% (3.6%)						1131.48 (0.04)	27131.6 (1.0)	172 (1.2)	76.0% (0.5%)	17.5% (0.7%)	15.5% (0.5%)
1 M	1129.59 (0.03)	27086.0 (0.7)	54 (1.5)	23.7% (0.7%)	60.9% (1.6%)						1131.43 (0.05)	27130.4 (1.3)	169 (1.2)	74.6% (0.5%)	16.2% (0.7%)	16.4% (0.5%)
2 M	1129.55 (0.03)	27085.1 (0.7)	51 (1.5)	22.6% (0.7%)	60.6% (1.6%)						1131.39 (0.05)	27129.3 (1.3)	166 (1.2)	73.4% (0.5%)	17.4% (0.7%)	14.2% (0.5%)
3 M	1129.54 (0.03)	27084.9 (0.7)	51 (1.5)	22.4% (0.7%)	65.5% (1.6%)						1131.47 (0.05)	27131.3 (1.3)	171 (1.2)	75.6% (0.5%)	18.0% (0.7%)	16.5% (0.5%)
3.2 M	1129.53 (0.03)	27084.8 (0.7)	50 (1.5)	22.2% (0.7%)	57.6% (1.6%)						1131.45 (0.05)	27130.7 (1.3)	169 (1.2)	75.0% (0.5%)	17.4% (0.7%)	15.2% (0.5%)
3.4 M	1129.50 (0.03)	27084.0 (0.7)	48 (1.5)	21.4% (0.7%)	60.9% (1.6%)						1131.54 (0.05)	27133.0 (1.3)	175 (1.2)	77.6% (0.5%)	16.8% (0.7%)	13.7% (0.5%)
3.6 M	1129.52 (0.01)	27084.6 (0.2)	50 (0.6)	22.0% (0.3%)	58.3% (0.6%)						1131.41 (0.04)	27129.8 (1.0)	167 (2.6)	74.0% (1.2%)	16.9% (1.5%)	15.5% (0.3%)
3.8 M	1129.42 (0.03)	27082.1 (0.7)	43 (1.5)	19.2% (0.7%)	61.0% (1.6%)						1131.51 (0.05)	27132.3 (1.3)	174 (1.2)	76.8% (0.5%)	16.8% (0.7%)	13.4% (0.5%)
4 M	1129.42 (0.03)	27082.2 (0.7)	44 (1.5)	19.3% (0.7%)	54.0% (1.6%)						1131.30 (0.05)	27127.1 (1.3)	160 (1.2)	70.9% (0.5%)	17.4% (0.7%)	17.3% (0.5%)
5 M	1129.61 (0.01)	27086.5 (0.3)	55 (0.8)	24.3% (0.4%)	11.3% (1.1%)						1131.49 (0.01)	27131.8 (0.3)	172 (0.7)	76.2% (0.3%)	32.7% (0.7%)	34.5% (0.7%)
6 M											1131.30 (0.02)	27127.2 (0.4)	160 (1.0)	71.0% (0.4%)	32.3% (0.0%)	39.4% (0.7%)
7 M											1131.25 (0.01)	27126.0 (0.2)	157 (0.5)	69.6% (0.2%)	31.4% (0.5%)	39.8% (1.0%)
								wild-type CLIC1 pH 5.5								
0 M	1129.46 (0.01)	27083.0 (0.2)	32 (0.3)	14.2% (0.2%)	58.3% (0.3%)	1131.27 (0.02)	27126.5 (0.5)	120 (1.0)	53.0% (0.5%)	15.9% (0.8%)	1132.35 (0.05)	27152.4 (1.2)	172 (2.4)	76.2% (1.1%)	14.3% (0.6%)	10.3% (0.1%)
1 M	1129.43 (0.00)	27082.3 (0.0)	31 (0.0)	13.5% (0.0%)	59.1% (0.2%)	1131.16 (0.01)	27123.8 (0.3)	115 (0.7)	50.7% (0.3%)	15.9% (0.6%)	1132.17 (0.02)	27148.1 (0.5)	164 (1.0)	72.4% (0.5%)	13.9% (0.8%)	10.1% (0.1%)
2 M	1129.39 (0.03)	27081.4 (0.7)	29 (1.4)	12.7% (0.6%)	59.5% (0.8%)	1131.11 (0.14)	27122.6 (3.4)	112 (6.9)	49.6% (3.0%)	17.4% (0.7%)	1132.07 (0.23)	27145.7 (5.4)	159 (11.0)	70.2% (4.9%)	14.2% (0.0%)	10.1% (0.2%)
3 M	1129.48 (0.04)	27083.5 (1.0)	33 (2.1)	14.6% (0.9%)	52.4% (0.7%)	1131.31 (0.13)	27127.4 (3.1)	122 (6.2)	53.9% (2.7%)	18.6% (0.6%)	1132.37 (0.21)	27152.9 (4.9)	173 (10.0)	76.7% (4.4%)	16.4% (0.6%)	12.9% (0.8%)
3.2 M	1129.49 (0.04)	27083.8 (1.0)	33 (2.1)	14.8% (0.9%)	48.3% (0.4%)	1131.31 (0.01)	27127.4 (0.3)	122 (0.7)	53.9% (0.3%)	19.6% (0.7%)	1132.34 (0.06)	27152.2 (1.5)	172 (3.1)	76.0% (1.4%)	18.4% (0.4%)	15.0% (1.5%)
3.4 M	1129.79 (0.04)	27091.0 (1.0)	48 (2.1)	21.3% (0.9%)	28.5% (0.9%)	1131.14 (0.09)	27123.4 (2.2)	114 (4.5)	50.3% (2.0%)	28.9% (0.1%)	1132.33 (0.19)	27151.9 (4.6)	171 (9.3)	75.8% (4.1%)	21.1% (0.6%)	21.0% (0.2%)
3.6 M	1129.84 (0.02)	27092.2 (0.5)	50 (1.0)	22.3% (0.5%)	26.3% (1.1%)	1131.11 (0.02)	27122.6 (0.5)	112 (1.0)	49.6% (0.5%)	32.1% (3.6%)	1132.22 (0.05)	27149.3 (1.2)	166 (2.4)	73.4% (1.1%)	22.0% (0.3%)	22.8% (2.2%)
3.8 M	1129.89 (0.06)	27093.4 (1.4)	53 (2.7)	23.4% (1.2%)	29.8% (0.4%)	1131.32 (0.15)	27127.7 (3.6)	122 (7.2)	54.1% (3.2%)	32.2% (1.5%)	1132.31 (0.09)	27151.4 (2.2)	170 (4.5)	75.4% (2.0%)	21.8% (0.2%)	18.8% (1.8%)
4 M	1129.77 (0.03)	27090.5 (0.7)	47 (1.4)	20.8% (0.6%)	29.0% (1.5%)	1131.15 (0.00)	27123.6 (0.0)	114 (0.0)	50.5% (0.0%)	32.4% (2.7%)	1132.36 (0.04)	27152.6 (0.8)	173 (1.7)	76.5% (0.8%)	24.1% (1.6%)	18.5% (2.8%)
5 M	1130.18 (0.12)	27100.3 (2.9)	67 (5.8)	29.6% (2.6%)	19.4% (0.8%)	1131.26 (0.05)	27126.2 (1.2)	119 (2.4)	52.8% (1.1%)	29.5% (3.8%)	1132.18 (0.03)	27148.3 (0.7)	164 (1.4)	72.6% (0.6%)	33.1% (2.2%)	22.2% (0.8%)
6 M						1130.94 (0.05)	27118.6 (1.1)	104 (2.2)	46.0% (1.0%)	28.0% (1.7%)	1132.07 (0.03)	27145.8 (0.8)	159 (1.5)	70.3% (0.7%)	40.4% (0.6%)	34.8% (2.3%)
7 M						1130.93 (0.08)	27118.3 (1.9)	103 (3.9)	45.7% (1.7%)	19.9% (3.8%)	1132.15 (0.09)	27147.5 (2.1)	162 (4.3)	71.9% (1.9%)	42.0% (2.2%)	35.0% (1.2%)
								CLIC1-M32A pH 7.0								
0 M	1127.50 (0.02)	27036.0 (0.5)	47 (0.8)	21.9% (0.3%)	58.1% (4.2%)	1129.87 (0.00)	27092.8 (0.0)	137 (0.0)	60.7% (0.0%)	15.9% (1.0%)	1131.00 (0.00)	27120.0 (0.1)	181 (0.2)	79.9% (0.1%)	15.8% (3.5%)	10.2% (1.8%)
1 M	1127.41 (0.04)	27033.9 (0.9)	43 (1.4)	22.5% (0.6%)	57.4% (1.6%)	1129.82 (0.04)	27091.7 (0.9)	135 (1.5)	59.9% (0.7%)	15.8% (0.6%)	1130.96 (0.06)	27119.0 (1.4)	179 (2.2)	79.2% (1.0%)	16.0% (1.8%)	10.0% (1.8%)
2 M	1127.56 (0.04)	27037.5 (0.9)	49 (1.4)	23.6% (0.6%)	56.9% (1.6%)	1129.77 (0.07)	27090.5 (0.9)	134 (1.5)	59.1% (0.7%)	15.4% (0.6%)	1130.84 (0.06)	27116.3 (1.4)	175 (2.2)	77.2% (1.0%)	16.3% (1.8%)	11.3% (1.8%)
3 M	1127.07 (0.03)	27025.6 (0.7)	30 (1.4)	17.3% (0.5%)	49.7% (0.8%)	1129.83 (0.01)	27091.8 (0.2)	136 (0.3)	60.0% (0.1%)	20.7% (0.5%)	1130.87 (0.06)	27116.8 (1.5)	175 (2.4)	77.6% (1.1%)	17.1% (0.4%)	12.5% (1.7%)
3.2 M	1127.05 (0.04)	27025.1 (0.9)	29 (1.4)	16.9% (0.6%)	35.7% (1.6%)	1129.79 (0.07)	27091.0 (0.9)	134 (1.5)	59.4% (0.7%)	27.8% (0.6%)	1130.92 (0.06)	27118.1 (1.4)	177 (2.2)	78.5% (1.0%)	22.2% (1.8%)	14.4% (1.8%)
3.4 M	1127.24 (0.04)	27029.6 (0.9)	36 (1.4)	21.1% (0.6%)	23.3% (1.6%)	1129.93 (0.07)	27094.3 (0.9)	140 (1.5)	61.8% (0.7%)	30.5% (0.6%)	1130.84 (0.06)	27116.2 (1.4)	174 (2.2)	77.2% (1.0%)	25.4% (1.8%)	20.8% (1.8%)
3.6 M	1127.35 (0.06)	27032.3 (1.4)	41 (2.3)	19.9% (1.0%)	30.5% (0.1%)	1129.95 (0.01)	27094.7 (0.2)	140 (0.3)	62.0% (0.1%)	31.6% (0.7%)	1131.05 (0.16)	27121.1 (3.9)	182 (6.2)	80.6% (2.8%)	21.2% (3.0%)	16.7% (0.3%)
4 M	1127.33 (0.04)	27031.8 (1.0)	40 (1.6)	23.1% (0.7%)	6.8% (1.5%)	1129.81 (0.03)	27091.4 (0.7)	135 (1.1)	59.7% (0.5%)	18.9% (0.2%)	1131.33 (0.03)	27127.9 (0.7)	193 (1.1)	85.5% (0.5%)	52.3% (2.7%)	22.1% (1.0%)
5 M						1129.59 (0.08)	27086.2 (2.0)	127 (3.2)	56.0% (1.4%)	13.8% (0.1%)	1131.16 (0.06)	27123.8 (1.4)	187 (2.2)	82.6% (1.0%)	60.8% (1.4%)	25.5% (1.4%)
6 M						1129.47 (0.11)	27083.2 (2.5)	122 (4.1)	53.9% (1.8%)	12.4% (1.3%)	1131.02 (0.04)	27120.4 (0.8)	181 (1.4)	80.1% (0.6%)	60.7% (0.1%)	26.9% (0.6%)
7 M						1129.38 (0.07)	27081.1 (0.9)	119 (1.5)	52.4% (0.7%)	16.0% (0.6%)	1130.83 (0.06)	27115.9 (1.4)	174 (2.2)	77.0% (1.0%)	57.0% (1.8%)	26.9% (1.8%)

See page 196 for key.